NIST IR 8351
December 2024

DNA Mixture Interpretation:

A NIST Scientific Foundation Review

NIST
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

These opinions, recommendations, findings, and conclusions do not necessarily reflect the views or policies of NIST or the United States Government. This report and its supplemental documents are not intended to imply any criticism of any past, pending, or future legal proceedings involving forensic science or evidence based on forensic science.

NIST Interagency Report
NIST IR 8351

DNA Mixture Interpretation:
A NIST Scientific Foundation Review

John M. Butler
Hari Iyer
Rich Press
Melissa K. Taylor
Peter M. Vallone
Sheila Willis*
*former NIST International Associate under contract; retired director of Forensic Science Ireland

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8351

NIST NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

NIST Interagency Report
NIST IR 8351

DNA Mixture Interpretation:
A NIST Scientific Foundation Review

John M. Butler
Melissa K. Taylor
Sheila Willis*
Special Programs Office
Laboratory Programs
National Institute of Standards and Technology

Hari Iyer
Statistical Engineering Division
Information Technology Laboratory

Peter M. Vallone
Biomolecular Measurement Division
Material Measurement Laboratory

Rich Press
Public Affairs Office
Director's Office

*former NIST International Associate under contract; retired director of Forensic Science Ireland

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8351

December 2024

U.S. Department of Commerce
Gina M. Raimondo, Secretary

National Institute of Standards and Technology
Laurie E. Locascio, NIST Director and Under Secretary of Commerce for Standards and Technology

NISTIR 8351
December 2024

Disclaimer

Certain commercial entities, equipment, or materials may be identified in this document in order to describe an experimental procedure or concept adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

These opinions, recommendations, findings, and conclusions do not necessarily reflect the views or policies of NIST or the United States Government. This report and its supplemental documents are not intended to imply any criticism of any past, pending, or future legal proceedings involving forensic science or evidence based on forensic science.

Publication History
Approved by the NIST Editorial Review Board on 2024-12-17

How to Cite this NIST Technical Series Publication
Butler, J.M., et al. (2024) DNA Mixture Interpretation: A NIST Scientific Foundation Review. (National Institute of Standards and Technology, Gaithersburg, MD), NIST IR 8351. https://doi.org/10.6028/NIST.IR.8351

NIST Author ORCID iDs
John M. Butler: 0000-0001-6472-9157
Hari Iyer: 0000-0001-5198-9945
Rich Press: 0009-0006-5953-4404
Melissa K. Taylor: 0000-0002-4204-3864
Peter M. Vallone: 0000-0002-8019-6204
Sheila Willis: 0000-0002-6213-6486

Acknowledgments

Members of the DNA Mixture Resource Group (listed in Table 1.2) contributed helpful feedback and assistance in the early stages of drafting this report. Katherine Gettings, Nikola Osborne, and Sarah Riman provided valuable input, including the data summaries now in NISTIR 8351sup2. Sandy Koch checked revisions against public comments, and Jason Weixelbaum, Susan Ballou, Christina Reed, and Katherine Sharpless assisted with copy editing. Kathryn Miller from the NIST Library helped finalize the document for public release. This document was published initially as a draft on June 9, 2021. Public comment periods were held from June 9 to August 23 and October 22 to November 19, 2021. Comments received were compiled and shared at https://www.nist.gov/dna-mixture-interpretation-nist-scientific-foundation-review. Revisions resulting from public comment and updates were incorporated into the final version. The significant time and effort of those who carefully read and provided valuable written feedback on the draft report are appreciated. These contributions and input were an important part of the process to finalize this NISTIR 8351 report. Cover image credit: Svetlaborovko via Adobe Stock.

Author Contributions

JMB: conceptualization, data curation, project administration, writing (original draft of chapters 2, 3, and 4), writing (review & editing); HI: conceptualization, writing (original draft of chapter 4), writing (review & editing); RP: conceptualization, writing (original draft of executive summary and chapter 1), writing (review & editing); MKT: conceptualization, writing (review & editing); PMV: conceptualization, writing (original draft of chapter 6), writing (review & editing); SW: conceptualization, writing (original draft of chapter 5), writing (review & editing)

Abstract

Improvements in DNA testing and interpretation methods have allowed forensic scientists to reduce the quantity of DNA required for profiling an individual. Today, DNA profiles can be generated from a few skin cells. This increased sensitivity has extended the usefulness of DNA analysis into new areas of criminal activity beyond homicides and sexual assaults into the complex DNA mixtures often seen in casework. Distinguishing one person's DNA from another's in these mixtures, estimating how many individuals contributed DNA, determining whether the DNA is even relevant or is from contamination, or whether there is a trace amount of suspect or victim DNA makes DNA mixture interpretation inherently more challenging than examining single-source samples. These issues, if not properly considered and communicated, can lead to misunderstandings regarding the strength and relevance of the DNA evidence in a case.

This report explores DNA mixture interpretation in six chapters, a bibliography of cited references, an appendix with glossary and acronyms, and two supplemental documents. Chapter 1 introduces the topic of DNA mixtures, the difficulties behind their interpretations, and discusses the relevance of issues explored in the other chapters of this scientific foundation review. Chapter 2 provides background information on the use of DNA testing in forensic casework and describes principles and practices underlying mixture measurement and interpretation. The likelihood ratio (LR) framework and probabilistic genotyping software (PGS) are also discussed. Chapter 3 lists data sources used in this study and strategies used to locate them. Chapter 4 and Chapter 5 cover the report's core concepts: reliability and relevance issues in DNA mixture interpretation when seeking to answer "who" and "how" questions at the sub-source and activity levels of the hierarchy of propositions, respectively. Chapter 6 explores the potential of new technologies to assist in mixture interpretation and considerations for implementation. Two supplemental documents provide context on how the field has progressed (NISTIR 8351sup1) and summarized information from publicly accessible validation and proficiency test results covering DNA mixture interpretation (NISTIR 8351sup2). There are 498 references cited in this report.

Keywords

activity level propositions; case assessment and interpretation; continuous (fully continuous) models; discrete (semi-continuous) models; DNA; DNA mixture; DNA mixture interpretation; DNA transfer and persistence; forensic science; hierarchy of propositions; interlaboratory studies; internal validation studies; interpretation; likelihood ratio; probabilistic genotyping software; proficiency tests; scientific foundation review; technical merit evaluation.

Table of Contents

List of Tables

List of Figures

List of Boxes

Preface

Forensic science plays a vital role in the criminal justice system by providing scientifically based information through the analysis of physical evidence. The National Institute of Standards and Technology (NIST) is a non-regulatory scientific research agency within the U.S. Department of Commerce with a mission to advance national measurement science, standards, and technology. NIST has been working to strengthen forensic science methods for almost a century. Several scientific advisory bodies have expressed the need for a review of the scientific bases of forensic methods and identified NIST as an appropriate agency for conducting them. A scientific foundation review, also referred to as a technical merit evaluation, is a study that documents and assesses the foundations of a scientific discipline, that is, the trusted and established knowledge that supports and underpins the discipline's methods. Congress has appropriated funds for NIST to conduct scientific foundation reviews in forensic science. These reviews seek to answer the question: "What established scientific laws and principles as well as empirical data exist to support the methods that forensic science practitioners use to analyze evidence?" Background information on NIST scientific foundation reviews was previously published (NISTIR 8225).

A draft version of this report was released for public comment on June 9, 2021. Public comment periods were held from June 9, 2021 to August 23, 2021 and from October 22, 2021 to November 19, 2021. In addition, a public webinar was held on July 21, 2021. The authors of this report are immensely grateful for the detailed feedback provided during the public comment periods and have thoughtfully considered each submission.

A total of 63 sets of comments were received across the two periods in addition to 83 questions or comments submitted during the webinar. All comments and questions submitted are available in a 446-page PDF file (NISTIR8351-draft PCs). We recognize and acknowledge the significant time and effort of those who carefully read and provided valuable written feedback on the draft report. These contributions and input were an important part of the process to finalize the NISTIR 8351 report.

NIST is committed to maintaining a high level of quality in the information it disseminates. In preparing the draft and final versions of NISTIR 8351, we have made every effort to follow the NIST Information Quality Standards[1] which value utility, integrity, and objectivity.

As a result of the feedback provided, Appendix 1: History of DNA Mixture Interpretation in the draft version of the report was moved into a separate supplemental document (NISTIR 8351sup1). Information in the draft Appendix 2: Training and Continuing Education has been removed from this final report. A new supplemental document (NISTIR 8351sup2) was developed to summarize information from publicly accessible validation data and proficiency test results covering DNA mixture interpretation.

While a line-by-line adjudication of all comments is not provided, a few clarifying points are offered in response to feedback received:

[1] See https://www.nist.gov/director/nist-information-quality-standards (accessed November 1, 2024)

- As previously stated (NISTIR 8225), foundational scientific data should be *publicly accessible for independent review* so that interested parties can judge for themselves the soundness of the underpinning information.
    - In their public comments, several forensic laboratories indicated their validation studies have been reviewed as part of their accreditation process (so they have had outside review but not continuous open access to the data), and they invited NIST team members to privately review their validation studies in a similar manner. We do not believe that this would be a useful exercise in the context of our efforts. NIST evaluation of validation information would only move the needle from "trust the laboratory" to "trust the NIST authors who have looked at the laboratory data" rather than sharing publicly accessible data to facilitate independent review.
- Reliability statements based on aggregate performance across many types of samples and many different probabilistic genotyping software (PGS) systems do not provide the information needed to judge the degree of reliability of the measurement and interpretation in a particular case of interest. We believe it is inappropriate to transfer any global reliability statements to a specific case because of the number of variables that affect DNA mixture interpretation. What is needed in the context of a specific case is information concerning the performance of these methods when applied in casework-similar scenarios. Even when methods have foundational validity, application in an individual case may or may not be reliable. NIST is not making any statements about the degree of reliability for any individual forensic case or scenario.
- When considering the reliability of the entire DNA mixture interpretation process (including judgments from the analyst prior to using any software), it is important to understand and consider the assumptions around DNA transfer, persistence, prevalence, and recovery (TPPR). If this report had focused solely on likelihood ratios with sub-source level propositions assigned by probabilistic genotyping software systems, without considering questions of TPPR, then the resulting assessment would have been incomplete. Chapter 5 provides a discussion of DNA TPPR.
- Given that DNA mixture interpretation is an active area of research and practice, we recognize that new information has been published since the draft report was released in June 2021. The final version of this report maintains the scope of the original draft and therefore may not be fully up-to-date. Where applicable and as described below, new information has been added to respond to public comments and inform the final version of this report.

Primary revisions made since the June 2021 draft report:

- **Chapter 1**: Some minor revisions were made based on public comments received. For example, to the list of factors contributing to increased complexity, we added "assignment of allele pairs to specific contributors when similar contributions exist (e.g., 1:1 or 1:1:1 mixtures)" and moved "degree of overlapping alleles across contributors" to its own factor.

- **Chapter 2**: A new section (2.6) was inserted prior to the DNA principles (now Section 2.7) to introduce the hierarchy of propositions earlier in the report. This topic was covered originally in Section 5.4.2.5 of the draft report. Draft Figure 5.3 was revised and changed to new Table 2.5. No new principles were added nor were any removed. Some of the text for Principles 1, 7, 8, 9, 10, 11, 12, 14, and 16 were revised based on public comment received. Text was revised for Key Takeaways 2.1, 2.2, 2.3, 2.4, 2.5, and 2.6 and a new Key Takeaway 2.7 added. Section 2.5.2 was changed from "LR Results, Transposed Conditionals, and Verbal Scales" to "Communicating LR Results." Figure 2.3 was removed that provided an illustration of likelihood ratio as a ratio of two likelihoods and tipping of scales.
- **Chapter 3**: Draft Table 3.2 was moved to a supplemental document (NISTIR 8351sup2) as part of "Summarized Information from Publicly Accessible Validation Data and Proficiency Test Results Covering DNA Mixture Interpretation." Validation studies have been added that were provided with public comments or listed on the Brooklyn Public Defender's website after the initial draft report was released in June 2021. Collaborative Testing Services began offering a probabilistic genotyping proficiency test in 2022 and results from these tests are summarized in NISTIR 8351sup2 (see Tables S2.8 to S2.11).
- **Chapter 4**: The title was changed from "Reliability of DNA Mixture Measurements and Interpretation" to "DNA Interpretation at the Sub-Source Level." The chapter was restructured after consideration of public comments received. Tables of summarized information from publicly accessible data (draft report Tables 4.2, 4.3, 4.5, 4.6, 4.7, and 4.8) were moved to a supplemental document (NISTIR 8351sup2). Textual revisions were made to Key Takeaways 4.1, 4.2, 4.3, 4.4, 4.5, and 4.7 and some were renumbered compared to the draft report. Key Takeaways 4.6 and 4.8 were removed, and a new Key Takeaway 4.7 was created.
- **Chapter 5**: The title was changed from "Context and Relevance Related to DNA Mixture Interpretation" to "DNA Interpretation at the Activity Level." The word "relevance" was changed in many places to "appropriate" or other phrases to improve clarity. The chapter was restructured somewhat with Section 5.3.1.9 moved to Section 5.1.1 to introduce an example near the beginning. Text was revised for Key Takeaways 5.1, 5.2, 5.3, 5.4, 5.5, and 5.6. Figure 5.3 was replaced with Table 2.5 as information introducing the hierarchy of propositions was moved to Chapter 2. Sections 5.4.2.6 and 5.4.2.7 on activity-level propositions and case assessment and interpretation (CAI) were expanded. While additional information from several publications was included (e.g., Lapointe et al. 2015, Yang et al. 2022, Prinz et al. 2024), the growing literature on DNA transfer, persistence, prevalence, and recovery (TPPR) has not been fully captured. DNA TPPR could be explored in greater detail as suggested in other recent reports (see EWG 2024, recommendation 7.3; TFSC 2024, pp. 62-67).
- **Chapter 6**: The wording of Key Takeaway 6.2 was revised based on public comments. As mentioned in Section 6.3, some examples of using single-cell analysis to reduce mixed-sample complexities have been published since the draft report was released (Duffy et al. 2023, Grgicak et al. 2024, Huffman & Ballantyne 2023a, Huffman & Ballantyne 2023b, Huffman et al. 2023, Kulhankova et al. 2023, Kulhankova et al. 2024, Schulte et al. 2023, Schulte et al. 2024). Sections 6.4 and 6.4.1 discuss the probabilistic genotyping software

systems that have added modules to assist with NGS data of DNA mixtures (e.g., Bleka et al. 2022, Cheng et al. 2023), and an ISFG DNA Commission has offered recommendation on nomenclature for STR allele sequences (Gettings et al. 2024).

- **Appendix 1**: This material was updated and moved to supplemental document "History of DNA Mixture Interpretation" (NISTIR 8351sup1).
- **Appendix 2**: This material on training and education was removed as much of the information is now available in Chapter 9 "Education, Training, and Professional Credentialing" in *Forensic DNA Interpretation and Human Factors: Improving Practice Through a Systems Approach* (EWG 2024, pp. 241-274).
- **Bibliography**: Now renamed "References," this section contains 498 references cited in the final report. Digital object identifier (DOI) links were added to references to improve accessibility.

Executive Summary

All scientific methods have limits. To use a method appropriately, one must understand those limits, which are inevitably tied to the risk one is willing to accept either as an individual or as a society. This is especially important in forensic science, as critical decisions impacting life and liberty are often based on the results of forensic analysis.

Forensic DNA technology brings immense benefits to society, and new tools and techniques can increase those benefits further. But as new technologies are implemented with increased detection capabilities, it is important to periodically assess their impacts on the scientific discipline. We do that in this scientific foundation review, which explores the capabilities and limitations of DNA mixture interpretation methods, including probabilistic genotyping software systems.

To conduct this review, we identified scientific principles, reviewed the scientific literature, gathered other empirical evidence from publicly available sources, and collected input from a group of forensic DNA practitioners and researchers. This final version of the report was also informed by public comments received in response to an initial draft of this report.

Information contained in this report comes from the authors' technical and scientific perspectives and review of information available during the time of this study. Where the findings identify opportunities for additional research and improvements to practices, researchers and practitioners are encouraged to take action toward strengthening methods used to move the field forward.

As with any field, the scientific process (research, results, publication, additional research, etc.) continues to lead to advancements and better understanding. For example, improvements in DNA testing methods have allowed forensic scientists to reduce the quantity of DNA required for profiling an individual. In the 1990s, an evidence sample needed to contain thousands of cells, such as from a visible blood or semen stain. Today, analysts can extract a DNA profile from the few skin cells that someone might leave behind when handling an object.

This increased sensitivity extended the usefulness of DNA analysis into new areas of criminal activity beyond homicides and sexual assaults. For example, DNA on bullets or cartridge casings can reveal clues to crimes involving firearms. Swabbing objects that a perpetrator handled can yield evidence in property crimes. Cold-case evidence previously analyzed with less discriminating methods can be re-opened and researched again to find new insights.

However, people constantly shed small amounts of DNA into the environment. For instance, by touching objects, people can potentially transfer small amounts of DNA, including someone else's DNA, from one surface to another. Therefore, analyzing small quantities of DNA can create new interpretation challenges.

Highly sensitive methods, now universally used across the forensic DNA community, often detect DNA from more than one individual in a sample. But it can be difficult to distinguish one person's DNA from another in these mixtures, to estimate how many individuals contributed DNA, to determine whether the DNA is relevant to the crime being investigated or is from contamination, and to determine whether there is a trace amount of DNA from the suspect or

victim. These issues make DNA mixtures inherently more challenging to interpret than single-source samples.

In addition, when laboratories analyze high-quality, single-source samples, decision-makers (e.g., jury members) often have high confidence in DNA test results in part because it has been demonstrated that different laboratories will arrive at the same result. This is true regardless of the specific instruments, tests, and software used. However, multiple interlaboratory studies conducted by different groups over the past two decades have demonstrated that different laboratories can produce a wide range of results when interpreting the same *DNA mixtures*.

These challenges posed by DNA mixtures need to be carefully considered throughout the forensic science process and clearly communicated when describing forensic results. Failure to do this can lead to misunderstandings regarding the strength and relevance of the DNA evidence in a case.

This report is arranged into six chapters and two supplemental documents. Chapter 1 introduces the topic of DNA mixtures (samples that contain DNA from more than one individual), the challenges behind their interpretations, and the importance of the issues explored in the other chapters of this scientific foundation review. Chapter 2 provides background information on DNA and describes principles and practices underlying mixture measurement and interpretation. The likelihood ratio (LR) framework and probabilistic genotyping software (PGS) are also discussed. Chapter 3 lists data sources used in this study and the strategies used to identify them. Chapters 4 and 5 cover the report's core concepts: the issues of reliability and relevance in DNA mixture interpretation. Chapter 6 explores the potential of new technologies to assist mixture interpretation and considerations for implementation. Supplement 1 presents the history of DNA mixture interpretation, while Supplement 2 provides summarized information from publicly accessible validation data and proficiency test results covering DNA mixture interpretation. Key takeaways, which are provided in the text and summarized in this executive summary, should be evaluated within the context of the entire report.

A DNA Mixture Resource Group (see Table 1.2), with extensive experience in public and private forensic DNA laboratories, reviewed an early version of the published draft report and provided valuable feedback, insights, and suggestions. However, they were not asked to sign off on the final report nor endorse its conclusions. The NIST team is grateful for their dedication and contributions to these efforts.

Chapter 1: Introduction

New tools and techniques for analyzing and interpreting DNA mixtures are now routinely employed in everyday casework in the United States and around the world. These tools include DNA profiling kits, genetic analyzer instruments, and probabilistic genotyping software.

DNA mixtures can be partly understood by analogy to latent print examination. If multiple fingerprints are deposited on top of one another, it would be difficult to tease apart the individual fingerprints because it may not be clear which ridge lines belong to which print. In a

DNA mixture, it may not be clear which genetic components, called alleles, belong to which contributor. Interpreting the mixture requires an assessment of weighted possibilities for which alleles go together to form the DNA profiles of the individual contributors, which are then compared to a person of interest (POI).

Forensic scientists interpret DNA mixtures with the assistance of statistical models and expert judgment. Interpretation becomes more complex when contributors to the mixture share common alleles and/or when random variations, also known as stochastic effects, occur during the polymerase chain reaction (PCR) amplification process that make it more difficult to confidently interpret the resulting DNA profile.

Not all DNA mixtures present these types of challenges. This review focuses on methods and practices for interpreting data from complex DNA mixtures, which are defined as samples that contain comingled DNA from two or more contributors in which stochastic effects or allele sharing increase uncertainty in determining contributor genotypes. Chapter 2 details factors that contribute to the complexity of DNA mixtures including:

- Number of contributors
- Low-quantity DNA from one or more minor contributors
- Assignment of alleles or allele pairs to specific contributors when similar contributions exist (e.g., 1:1 or 1:1:1 mixtures)
- Degree of degradation or presence of PCR inhibitors in the DNA sample
- Degree of overlapping alleles across contributors.

It is important that users of forensic DNA test results understand that DNA evidence can vary greatly in complexity based on these factors, and that more complex samples result in greater uncertainty with respect to measurement and interpretation of results.

Chapter 2: DNA Mixture Interpretation: Principles and Practices

Successful analysis and interpretation of DNA results depends on crime scene evidence (the “Q” or questioned sample) being of suitable quality and quantity, and the availability of a reference sample (the “K” or known sample). When appropriate Q and K DNA profiles are available, forensic scientists can perform a Q-to-K comparison to make an assessment of whether or not K is a contributor to Q. This assessment is often made in the form of a likelihood ratio (LR) that is an evaluative interpretation of the strength of the results of this comparison using specific assumptions (competing propositions) and usually one of several statistical approaches.

The process of DNA evidence analysis can be divided into two major steps: (1) the generation of and subsequent *measurements* of relative abundances of PCR products in an analyzed DNA sample that are displayed as an electropherogram (EPG), and (2) *interpretation* involving use of the EPG data to make a strength-of-evidence assessment when an evidentiary DNA profile is compared to the profile of a POI. The outcome of interpretation includes the assignment of an LR number that is dependent on the analyst’s assumptions, protocols, algorithms, tools, and other variables, such as case information. Forensic scientists interpret DNA mixtures with the assistance of statistical models and expert judgment. A framework known as the *hierarchy of*

propositions helps consider the types of questions being considered during DNA interpretation, e.g., whether at a sub-source (DNA) level or an activity level.

This chapter describes 16 principles of DNA mixture interpretation and includes seven key takeaways.

KEY TAKEAWAY #2.1: DNA mixtures, in which the DNA of more than one individual is present in a sample, are inherently more difficult to interpret than high-quality single-source DNA samples.

KEY TAKEAWAY #2.2: Generating a DNA profile involves measuring the physical properties of a sample. Interpreting a DNA profile involves the DNA analyst applying their judgment, training, tools (including computer software), and experience, and considers factors such as case context and laboratory protocols and policies.

KEY TAKEAWAY #2.3: The process of generating a DNA profile from low amounts of DNA template can produce stochastic or random variation and artifacts that contribute to the challenge of DNA mixture interpretation.

KEY TAKEAWAY #2.4: DNA mixtures vary in complexity. Factors that contribute to complexity include the number of contributors, the quantity of DNA from each contributor, contributor mixture ratios, sample quality, and the degree of allele sharing. These factors will impact the degree of variation in interpretation.

KEY TAKEAWAY #2.5: Continuous probabilistic genotyping software (PGS) methods utilize more information from a DNA profile than binary or discrete PGS approaches.

KEY TAKEAWAY #2.6: Likelihood ratios are assigned and not measured. Different individuals may assign different LR values, even when using PGS systems, when presented with the same evidence because they base their judgments on different collection protocols, quantification systems, STR kit results, interpretation protocols, models, assumptions, or computational algorithms. For any given sample, there is no single, true likelihood ratio.

KEY TAKEAWAY #2.7: A probabilistic genotyping framework offers the best available tool for DNA mixture interpretation at the sub-source or sub-sub-source levels within the hierarchy of propositions. However, its fitness for purpose in any specific casework application must be supported by validation data from known samples that are similar in complexity to those seen in casework. Continuous PGS systems have many advantages if they are used following suitable training, and if the decisions informing the LR assignment(s) are clearly stated.

Chapter 3: Data and Information Sources

This chapter contains sources of data and information that were used in conducting this review along with strategies that were used to locate them. These sources include (1) peer-reviewed articles appearing in scientific journals, (2) published interlaboratory studies, (3) laboratory internal validation study summaries that are accessible online, and (4) proficiency test data available on test provider websites. An accompanying supplemental document summarizes this publicly accessible information (see NISTIR 8351sup2).

Chapter 4: DNA Interpretation at the Sub-Source Level

In this report, the challenges presented by DNA mixtures are divided into two main categories. The first involves the *reliability* of mixture interpretation methods when used with DNA evidence of varying complexity at the sub-source level while attempting to answer questions about the findings, *considering the potential source(s) of the DNA profile*. (Chapter 5 deals with the second challenge: *relevance* or considering DNA evidence at the activity level while attempting to answer questions about the findings, *considering potential activities of the person(s) of interest*.) In this report, the "plain English" definition of reliability as a measure of trustworthiness is used. A highly reliable method is one that consistently produces accurate results. Reliability is not a yes or no question, but a matter of degree. Understanding the degree of reliability of a method can help the user of that information decide whether they should trust the results of that method in any specific situation when making important decisions.

This chapter considers foundational issues related to reliability of DNA mixture interpretation. Reliability centers on trustworthiness established through empirical assessments of available data to evaluate the degree of reliability of a system or its components. The term "factor space" is used to describe the variables (factors) that influence complexity, measurement, and interpretation reliability – these factors include the number of contributors, the degree of allele sharing, the ratios of mixture components, and the amount and quality of the DNA tested.

We note that the degree of reliability of a DNA mixture interpretation system, such as a DNA analyst using a probabilistic genotyping software program, depends on sample complexity. Results cannot be simply categorized as "reliable" or "unreliable" without considering context. In addition, reliability cannot be established without validation testing using known samples of similar complexity. The results of such tests provide data that are considered reliable when shown to be consistently accurate; only with such valid results can comparisons be made as to the reliability of a method for use with casework samples for which the ground truth is not known. This report also emphasizes that samples used in proficiency tests need to be representative of complex DNA mixtures seen in casework if these tests are intended to assess analysts' ability to conduct dependable DNA mixture interpretation.

Finally, the theme of reliability is discussed throughout this report. The report's initial aim was to assess the reliability of DNA mixture interpretation using information derived from publicly accessible data. Sources considered included published scientific articles including interlaboratory comparison studies, proficiency test results, and publicly accessible internal validation reports (see NISTIR 8351sup2). In this report, we define reliability (or degree of reliability) as an assigned value being consistently accurate.

Following public comments on the draft report, it was determined that an assessment of reliability (or even a degree of reliability) for global forensic cases was not feasible for LR values assigned by PGS systems given sub-source propositions, in large part because there is no true LR (one cannot determine ground-truth accuracy for a specific LR value). This would be the case with or without more detailed publicly available data. This is not to say that publicly available data is not useful as it provides the community with an indication of the performance of certain

aspects of mixture interpretation and how it evolves. Providing relevant metadata related to the findings will continue to benefit the overall transparency for the use of this information in the community.

As noted in Key Takeaway #2.1, there is a difference in the process of obtaining LR values resulting from the interpretation of single-source samples and those from contributors in DNA mixtures. Alleles from lower-level contributors in a DNA mixture are expected to exhibit stochastic variation during PCR, which results in variability in the resulting EPGs if subjected to replicate analyses (e.g., the number and intensity of detected alleles) and, therefore, the assigned LR value. The importance of clearly stating what goes into the entire DNA measurement and interpretation processes and the supporting validation data is what can inform the user of the information with what data and decisions have resulted in the reported assigned LR value.

This chapter includes seven key takeaways.

KEY TAKEAWAY #4.1: To assess the degree of reliability of a component or a system for any forensic method or practice, detailed empirical data are needed, such as data from validation experiments, interlaboratory studies, and proficiency tests.

KEY TAKEAWAY #4.2: There is a growing body of scientific literature on DNA mixture interpretation. However, supporting data provided in the scientific literature is not always sufficiently detailed for an independent review of claims. Such data and details, if required as part of the journal publication acceptance process, will assist with independent review of published articles.

KEY TAKEAWAY #4.3: Currently, publicly accessible validation data does not have the detail (including metadata, protocols, conditions, etc.) to enable an external and independent assessment of the degree of reliability of DNA mixture interpretation practices, including the use of probabilistic genotyping software (PGS) systems.

KEY TAKEAWAY #4.4: Current proficiency tests are primarily focused on single-source samples and simple two-person mixtures containing large quantities of DNA. To appropriately assess the ability of analysts to interpret complex DNA mixtures, tests of analysts should include the types of samples often seen in forensic casework, such as mixtures with low-template components and more than two contributors.

KEY TAKEAWAY #4.5: To enable effective use of any information, providers and users of the information both have responsibilities. While a provider explains the relevance and significance of the information and data, only an informed user can assess the degree of reliability and validity, and whether that information is fit-for-purpose for their application.

KEY TAKEAWAY #4.6: When assessing the degree of reliability of DNA mixture results for a specific case, the assessor (e.g., an expert user of the results) needs to have access to validation data from known samples that are similar in complexity to the sample in the case.

KEY TAKEAWAY #4.7: To improve data sharing across laboratories and support independent assessments, the Scientific Working Group on DNA Analysis Methods (SWGDAM) and the Organization of Scientific Area Committees for Forensic Science (OSAC) are encouraged to

develop minimum requirements and standard formats for data in validation studies and recommend that validation data be made publicly accessible.

Chapter 5: DNA Interpretation at the Activity Level

The second major challenge posed by DNA mixtures involves the *relevance*[2] of a DNA sample to the crime being investigated. The question of relevance arises because DNA can be transferred between surfaces, potentially more than once. This means that some of the DNA present at a crime scene may be irrelevant to the crime, and current DNA profiling methods increase the likelihood of detecting more DNA. Similarly, today's highly sensitive DNA methods increase the risk that very small amounts of contamination might affect DNA test results.

Chapter 5 focuses on questions of context and activity level considerations: How and when was the DNA deposited, and is that DNA relevant to the crime being investigated?

The question of relevance arises because people shed DNA into the environment, and they can potentially transfer DNA between surfaces when touching objects or other people. Therefore, the DNA present at a crime scene or on a piece of evidence may be irrelevant to any crime. To assess relevance, in addition to knowing specific details of the case, one would need information on what factors make DNA more or less likely to transfer and to persist in the environment. This chapter reviews the scientific literature on DNA transfer and persistence and presents strategies for assessing DNA relevance.

The fact that DNA can be transferred between surfaces upon contact is a foundational principle of forensic DNA analysis. This has several implications for DNA found at a crime scene. First, that DNA might have been deposited before or after the crime was committed and therefore may not be associated with the crime. Second, the DNA might have been deposited via indirect (secondary) transfer, which occurs when DNA is picked up from one surface and deposited on another. For instance, a person might pick up DNA from a second person during a handshake, then deposit the second person's DNA onto a surface.

These possibilities mean that the presence of a person's DNA in an evidence sample does not necessarily mean that the detected DNA is relevant to (i.e., associated with) the crime. When these possibilities are a concern, activity level evaluations and reporting provide an opportunity to consider the findings given specific propositions about how the observed DNA profile(s) may have arisen. If not, the evidence can be misleading.

By definition, highly sensitive methods are more likely to detect small quantities of DNA, including background DNA that may be present in the environment. In addition, highly sensitive methods are more likely to detect DNA mixtures, which by their nature may include irrelevant DNA. Therefore, when assessing evidence that involves very small quantities of DNA, it is especially important to carefully consider relevance.

This report uses the word contamination to describe the transfer of irrelevant DNA during an investigation. For example, a fingerprint brush can potentially transfer minute amounts of DNA

[2]Science does not assess legal relevance but rather provides information to those who do (see PC16 in NISTIR8351-draft PCs).

onto evidence at a crime scene. Such a small amount of DNA might have gone undetected in the past, but highly sensitive methods increase the likelihood that it might now be detected. This increases the likelihood that contamination might affect an investigation.

Forensic laboratories have long used procedures to avoid contamination of evidence. However, because the likelihood of detecting extraneous DNA has increased with the development of highly sensitive DNA methods, contamination avoidance in forensic laboratories is more important than ever. Furthermore, contamination avoidance procedures should be used during all stages of an investigation, including at the crime scene or the hospital when staff interact with a victim. Elimination databases that include DNA profiles of police and laboratory staff who go to crime scenes or analyze evidence items prior to DNA processing can help identify contamination and should be maintained.

Many interpretation methods, including probabilistic genotyping, address questions about who might have contributed DNA to a crime scene profile and express opinions for the strength of evidence in the form of a likelihood ratio given sub-source propositions. This statistic does not provide any information about how much DNA was present, or how or when the DNA was deposited. For instance, a large blood stain might produce a very similar likelihood ratio to a swab from a light switch, yet the two types of evidence might vary greatly in terms of their evidential value. Therefore, likelihood ratios should not be used in isolation. It is imperative that the likelihood ratio value(s) (assigned with sub-source or activity level propositions) be considered in the context of other evidence in the case.

The fact that DNA can be transferred does not mean that DNA is useless as evidence. To the contrary, this is what makes DNA useful to criminal investigations. However, the possibility of DNA transfer may raise questions of relevance that need to be addressed, especially in cases that involve very small amounts of DNA. These questions can be addressed by considering DNA evidence in the context of case circumstances, including other evidence in the case.

There is a need for more structured research and education around assigning probabilities from DNA transfer and persistence studies in activity level assessments and reporting. In addition, to make use of the studies that have been published, individual laboratories would need to know how the sensitivity of methods used in their laboratory compares to the sensitivity of methods employed in the studies being considered.

This chapter includes six key takeaways.

KEY TAKEAWAY #5.1: DNA can be transferred from one surface or person to another, and this can potentially occur multiple times. Therefore, the relevance of the DNA to the crime being investigated should be considered when evaluating the evidence.

KEY TAKEAWAY #5.2: Highly sensitive DNA methods increase the likelihood of detecting DNA mixtures that may or may not be related to the crime being investigated. In cases that involve very small quantities of DNA, it is especially important for the users of the information to consider context when determining the utility of the evidence.

KEY TAKEAWAY #5.3: Contamination avoidance procedures should be robust both at the crime scene and in the laboratory. These procedures should include the maintenance of

elimination databases containing samples from personnel who have access to crime scenes and evidence items.

KEY TAKEAWAY #5.4: DNA statistical results such as a likelihood ratio given sub-source propositions do not provide information about how or when DNA was transferred, or whether it is relevant to circumstances of a case. Therefore, reporting a likelihood ratio as a standalone number can be misleading without sharing the assumptions made in the LR assignments and the level in the hierarchy of propositions being addressed (i.e., considering what question is being answered).

KEY TAKEAWAY #5.5: The fact that DNA can transfer between objects does not negate the value of DNA evidence. However, the value of DNA evidence depends on the circumstances of the case.

KEY TAKEAWAY #5.6: There is a growing body of knowledge about DNA transfer, persistence, prevalence, and recovery (TPPR). However, significant knowledge gaps exist, including: (1) appropriate TPPR data for casework-like scenarios, and (2) education of and standardized approaches for users on how to apply the LR framework to activity-level questions in a specific case.

Chapter 6: New Technologies: Potential and Limitations

New technologies are often investigated to assess whether they can provide solutions to existing problems in the forensic community. The adoption and implementation of these technologies depends upon a cost/benefit analysis within forensic laboratories. Considering whether new approaches can bring desired improvements to mixture interpretation involves appreciating fundamental challenges with DNA mixture interpretation.

The ability to analyze short tandem repeat alleles by sequence in addition to length promises to bring some new capabilities to forensic DNA laboratories. Next-generation sequencing platforms also enable additional genetic markers to be examined, some of which, such as microhaplotypes, have been pursued with the potential to improve DNA mixture interpretation. Additionally, cell separation techniques offer the potential to separate contributors prior to DNA extraction.

The ultimate decision to implement new technologies in forensic laboratories should be driven by real use cases and by those responsible for producing and reporting the information. A vendor or members of the general public may encourage forensic DNA laboratories to adopt a new approach or technology, but adoption often requires significant investments. When making these decisions, forensic laboratories consider whether supporting factors and resources will be available upon implementation (e.g., allele frequencies, analysis software, interpretation methods, training, and support for potential admissibility hearings). An overall assessment of 1) how a new technology works, 2) its limitations, 3) how it might specifically help improve a process or solve a problem (e.g., DNA mixture interpretation), and 4) whether this new technology can be justly and equitably implemented is important and a key component of evaluating whether implementation will be worthwhile.

This chapter includes two key takeaways.

KEY TAKEAWAY #6.1: Fundamental measurement and interpretation issues surrounding DNA mixtures, as described in Chapter 2, should be understood before attempting to apply a new technology.

KEY TAKEAWAY #6.2: Implementation of new technologies requires a thorough understanding of the socio-technical benefits and limitations in addition to the required investment of time and effort put forth for its adoption by the laboratory.

1. Chapter 1: Introduction

All scientific methods have limitations. One must understand those limitations to use a method appropriately. This is especially important in forensic science as critical decisions impacting life and liberty are often based on the results of forensic analyses. This scientific foundation review explores what is known about the limitations of DNA mixture interpretation methods, including probabilistic genotyping software systems, by reviewing the scientific literature and other sources of information.

1.1. Advances in Forensic DNA

The field of forensic DNA analysis is constantly advancing. One important change involves the ability to detect and analyze very small quantities of DNA (Butler 2011, Butler 2014). During the early decades of forensic DNA analysis, an evidence sample containing thousands of cells, such as a visible blood or semen stain, was needed to produce a DNA profile. Today, analysts can extract a DNA profile from the few skin cells that someone might leave behind when handling an object.

This increased sensitivity extends the usefulness of DNA analysis into new areas of criminal activity beyond the homicides and sexual assaults that were once the primary focus. Crimes involving firearms can be investigated by testing for DNA on bullets or cartridge casings (e.g., Montpetit & O'Donnell 2015). Property crimes can be investigated by swabbing objects that a perpetrator might have handled (Mapes et al. 2016). Cold cases that were previously analyzed with less discriminating methods can now yield useful evidence.

However, people constantly shed small amounts of DNA into the environment, and by touching objects, people can potentially transfer small amounts of DNA – including someone else's DNA – from one surface to another. Analyzing small quantities of DNA can create challenges when interpreting the data. Highly sensitive methods, now universally used across the forensic DNA community (Gill et al. 2015), often detect DNA from more than one individual in a sample. Analysts know they are dealing with what is called a DNA mixture when they detect more than two alleles at multiple locations in a DNA profile. Because humans typically inherit one allele from each parent for every gene, finding more than two is one indication that more than one genotype, the variation in a DNA sequence that is specific to an individual organism, may be present in the sample. As a result, more than one genotype combination may be possible at each tested location in the DNA sequence.

Distinguishing one person's DNA from another's in these mixtures, estimating how many individuals contributed to the recovered DNA sample, not knowing whether the DNA is associated with a crime or is from contamination, or whether the findings support the presence of a trace amount of suspect or victim DNA make DNA mixtures inherently more challenging to interpret than single-source samples. These issues, if not properly considered and communicated, can lead to misunderstanding the strength and relevance of the DNA evidence in a case.

The ability to detect small amounts of DNA has been improving for decades (Butler 2011, Butler 2014). When forensic DNA analysis was first introduced in the mid-1980s (Gill et al. 1985), a

stain about the size of a quarter was needed to generate a DNA profile. In the early 1990s, forensic laboratories started using polymerase chain reaction (PCR), a method that leverages the natural tendency of DNA to produce copies of itself, to amplify DNA. This method allowed the analysis of much smaller amounts of starting material (e.g., Saiki et al. 1989, Blake et al. 1992), though a visible stain was still generally needed. In 1997, scientists demonstrated high-sensitivity methods that allowed for recovery of DNA information from touched objects (van Oorschot & Jones 1997) and even from single cells (Findlay et al. 1997).

Highly sensitive methods began moving from research centers into crime laboratories over the past two decades (see NISTIR 8351sup1), but the application of such methods to detect minor contributors in DNA mixtures has increased rapidly in recent years. New tools and techniques for analyzing and interpreting minor contributors to DNA mixtures are now routinely employed in everyday casework in the United States and around the world (Butler 2015, Gill et al. 2015, Buckleton et al. 2016). These tools include DNA profiling kits, genetic analyzer instruments, and probabilistic genotyping software (PGS).

Forensic DNA technology brings immense benefits to society, and these new tools and techniques can increase those benefits further. But as new technologies are implemented with increased detection capabilities, we believe it is important to periodically assess the impacts on the scientific discipline. We do so in this scientific foundation review by identifying scientific principles, reviewing the scientific literature, gathering other empirical evidence from unpublished sources, and collecting input from a group of forensic DNA practitioners and researchers.

As with any field, the scientific process (research, results, publication, additional research, etc.) continues to lead to advancements and better understanding. Information contained in this report comes from the authors' technical and scientific perspectives and review of information available during the time of this study. Where the findings identify opportunities for additional research and improvements to practices, researchers and practitioners are encouraged to take action toward strengthening methods used to move the field forward.

1.2. DNA Mixtures Vary in Complexity

DNA mixtures can be partly understood by analogy to latent print examination. If multiple fingerprints are deposited on top of one another, it would be difficult to tease apart the individual fingerprints because it may not be clear which ridge lines belong to which print. In a DNA mixture it may not be clear which genetic components, called alleles, belong to which contributor. Interpreting the mixture requires an assessment of which alleles go together to form the DNA profiles of the individual contributors.

Forensic scientists interpret DNA mixtures with the assistance of statistical models and expert judgment. Interpretation becomes more complicated when contributors to the mixture share common alleles (e.g., Clayton et al. 1998). Complications can also arise when reduced DNA template amounts are used in PCR, where random sampling, also known as stochastic effects, make it more difficult to confidently interpret the resulting DNA profile (e.g., Gill et al. 2005).

Not all DNA mixtures present these types of challenges. This review focuses on methods for interpreting data from complex DNA mixtures, which we define as samples that contain comingled DNA from two or more contributors in which stochastic effects or allele sharing may cause uncertainty in determining contributor genotypes. The following factors contribute to increased complexity (see Chapter 2):

- Number of contributors
- Low-quantity DNA from one or more minor contributors
- Assignment of alleles or allele pairs to specific contributors when similar contributions exist (e.g., 1:1 or 1:1:1 mixtures)
- Degree of degradation or presence of PCR inhibitors in the DNA sample
- Degree of overlapping alleles across contributors.

It is important that users of forensic DNA test results understand that DNA evidence can vary greatly in complexity based on these factors, and that more complex samples involve greater uncertainty with respect to measurement and interpretation of results.

1.3. Reliability

In this report, the challenges presented by DNA mixture interpretation are divided into two main categories. The first involves the *reliability* of mixture interpretation methods when used with DNA evidence of varying complexity. In this report, we use the "plain English" definition of reliability as a measure of trustworthiness. A highly reliable method is one that consistently produces accurate results. Reliability is not a yes or no question, but a matter of degree. Understanding the degree of reliability of a method can help the user of that information decide whether they should trust the results of that method when making important decisions. In addition, the degree of reliability of a method can often be demonstrated with appropriate empirical data.

Reliability issues are considered by surveying available validation studies, which are meant to demonstrate how a method performs under defined sets of circumstances (e.g., varying numbers of contributors, template amounts, mixture ratios). Interlaboratory studies and proficiency tests are also considered, which provide information on the variability in test results across laboratories and analysts (see NISTIR 8351sup2). In addition, the history of standards and guidelines for mixture interpretation is explored (see NISTIR 8351sup1).

Some performance assessments are also briefly discussed that are frequently used in other sectors, such as receiver operating characteristic (ROC) curves (Green & Swets 1966, Bleka et al. 2016b) and calibration of likelihood ratios (Zadora et al. 2014). When sufficient data are available, these assessments can be used to evaluate the reliability of DNA mixture interpretation methods and compare reliability across different PGS systems (e.g., Bleka et al. 2016b, You & Balding 2019). Laboratories might also use these assessments to set operational limits based on observations from their validation studies.

1.4. Relevance

The second major challenge posed by DNA mixtures involves the *relevance* of a DNA sample to the crime being investigated. The question of relevance arises because DNA can be transferred between surfaces, potentially more than once (van Oorschot et al. 2019). This means that some of the DNA present at a crime scene may be irrelevant to the crime, and current DNA profiling methods increase the likelihood of detecting more DNA. While relevance is always a critical assessment that depends on case circumstances, today's highly sensitive DNA methods increase the risk that very small amounts of contamination might affect DNA test results (e.g., Fonneløp et al. 2016, Szkuta et al. 2015a).

This report uses the word contamination to describe the transfer of irrelevant DNA during an investigation. For example, a fingerprint brush can potentially transfer minute amounts of DNA onto evidence at a crime scene. Such a small amount of DNA might have gone undetected in the past, but the use of highly sensitive methods increases the likelihood that it might now be detected. This increases the likelihood that contamination might affect an investigation.

Forensic laboratories have been using procedures to avoid contamination since the advent of DNA methods. However, because the likelihood of detecting contaminating DNA has increased with highly sensitive DNA methods, contamination avoidance in forensic laboratories is more important than ever. Furthermore, contamination avoidance procedures should be used during all stages of an investigation, including at the crime scene and at the hospital when staff interact with a victim. Elimination databases that include DNA profiles of police and laboratory staff who go to crime scenes or analyze evidence items prior to DNA processing can help identify contamination and should be maintained. Therefore, relevance should be carefully assessed and considered by both the DNA analyst and users of the DNA results, especially when an item of evidence contains very small amounts of DNA.

In this report, we address relevance issues by surveying the existing literature on DNA transfer and persistence, identifying what is known about these phenomena, and highlighting knowledge gaps. We discuss several ways in which DNA transfer might mislead an investigation if DNA evidence is not considered in the context of the facts and evidence in the case. We also suggest strategies for mitigating the risks presented by DNA transfer.

Mixture interpretation methods address questions about the source of a DNA sample (i.e., who the DNA came from) and provide statistical strength of evidence such as a likelihood ratio. The interpretation of a DNA profile can be useful by itself for generating leads in an investigation. However, the investigator or the trier of fact should consider not just the source of the DNA, but also what activity might have caused the DNA to be deposited as evidence (Gill et al. 2018, Taylor et al. 2018). Answering questions about activity generally requires consideration of contextual information, including other evidence in the case (Gill et al. 2020a). In Chapter 5, we argue that uncertainties about an activity are usually much greater than uncertainties about a source (e.g., Taylor et al. 2018), and it is therefore critical to consider DNA evidence in context. Focusing only on a statistic without considering context can mislead. This is especially so in cases involving very small quantities of DNA, such as when touch samples are collected from a

store counter or from a firearm that many people may have handled. There may be legitimate alternative activities that could explain the presence of the DNA of a person of interest.

1.5. Why Conduct This Scientific Foundation Review?

As described in an earlier publication (NISTIR 8225), a scientific foundation review is "a study that seeks to document and evaluate the foundations of a scientific discipline, that is, the trusted and established knowledge that supports and underpins the discipline's methods. These reviews seek to answer the question: 'What empirical data exist that speak to the reliability of the methods that forensic science practitioners use to analyze crime scene material?'"

Such a review can help identify knowledge gaps and provide guidance for future research. Understanding capabilities and limitations can aid in performing valid and reliable interpretations. In addition, documenting foundational studies and core principles in a written report can assist laboratories in identifying appropriate limits for interpretation and contribute to the training of forensic practitioners. This report can also help investigators, officers of the court, and other users of forensic science to consider DNA test results in context and with awareness of their limitations so they can make informed decisions.

There is abundant forensic DNA testing literature due to the large number of active researchers and a history of publishing that surpasses many other forensic disciplines. Thousands of articles pertaining to forensic DNA methods have been published in dozens of peer-reviewed scientific journals in the past 35 years. Similar review studies have been performed by other groups on forensic disciplines like fire investigations (Almirall et al. 2017) and latent fingerprints (Thompson et al. 2017). This foundation review seeks to explore DNA mixture interpretation in a similar manner.

When laboratories analyze high-quality, single-source samples, decision-makers often have confidence in DNA test results in part because it has been demonstrated that different laboratories will arrive at the same result; that is, obtain the same DNA profile at the tested loci regardless of the specific instruments, kits, and software used. However, multiple interlaboratory studies conducted by different groups over the past two decades have demonstrated a wide range of variation in interpretation of DNA *mixtures* (Duewer et al. 2001, Crespillo et al. 2014, Benschop et al. 2017, Barrio et al. 2018, Butler et al. 2018a, Brinkac et al. 2023, Hicklin et al. 2023). A scientific foundation review might shed light on the sources of variability observed.

1.6. Limitations of This Study

First, forensic genetics is an evolving field, and this study can only provide a snapshot of the state of the science at a particular moment in time. Therefore, the literature and empirical evidence discussed in this review will be dated as soon as it is published, as is the case with evidence reviews in other evolving fields such as medicine and public health.

Second, the data available for conducting this review were limited. For instance, most laboratories do not publish data from their validation studies. We find merit in the perspective

that “Dissemination is a critical part of the scientific process because it exposes our work to peer review and allows scientists to build upon the contributions of others. A study isn’t complete until it’s been published” (Martire & Kemp 2018). In addition, many published developmental validation studies do not include enough details for an independent assessment of performance. We believe that greater transparency through forensic laboratories openly sharing their supporting validation data, along with an independent review, would help strengthen the field of forensic DNA analysis. NIST researchers have set an example of transparency with details provided in supplemental files (e.g., Riman et al. 2021).

Third, we may not have succeeded in identifying all of the studies relevant to our research objectives.

1.7. NIST Review Team

The review team consisted of six individuals from the National Institute of Standards and Technology (NIST) whose diverse expertise allowed examination of issues from many perspectives and to use lessons learned in other fields. Table 1.1 lists members of the review team, their NIST operating unit, and their expertise. This team met regularly between September 2017 and July 2020 and then on an as-needed basis while conducting this review and developing the content of this report. Assistance in finalizing this report was also provided by several additional NIST employees or contractors as noted in the Acknowledgments.

Table 1.1. Members of the NIST review team and their areas of expertise.

Name	NIST Operating Unit	Areas of Expertise
John M. Butler	Special Programs Office	Forensic DNA methods and scientific literature
Hari K. Iyer	Statistical Engineering Division, Information Technology Laboratory	Mathematics and statistics
Rich Press	Public Affairs Office	Communication and science writing
Melissa K. Taylor	Special Programs Office	Human factors (previous efforts in latent fingerprints and handwriting analysis)
Peter M. Vallone	Applied Genetics Group, Material Measurement Laboratory	DNA technology, research, rapid DNA analysis, next-generation DNA sequencing
Sheila Willis	Special Programs Office (hired under contract as an International Research Associate)	Forensic laboratory management and trace evidence (retired director of Forensic Science Ireland)

1.8. DNA Mixture Resource Group

During the initial phases of this project, the NIST review team met regularly with a group of outside experts, the DNA Mixture Resource Group (Resource Group), which provided input and feedback that were vital to keeping this project focused on critical and relevant issues.

The Resource Group (Table 1.2) provided important perspectives based on their extensive experience in public and private forensic laboratories. This group included nine active practitioners, including five DNA technical leaders, from federal, state, and local jurisdictions in the United States and Canada, and four academics and consultants who have published in the forensic DNA literature.

Table 1.2. Members of the DNA Mixture Resource Group.

Name	Affiliation (at that time)
Jack Ballantyne	Professor of Chemistry, University of Central Florida
Todd Bille	Alcohol, Tobacco, Firearms, and Explosives (ATF) Laboratory, DNA Technical Leader
Jennifer Breaux	Montgomery County (MD) Police Crime Laboratory, DNA Technical Leader
Robin Cotton	Boston University School of Medicine (and former laboratory director of Cellmark Diagnostics)
Roger Frappier	Centre of Forensic Sciences (Toronto, Canada)
Bruce Heidebrecht	Maryland State Police, DNA Technical Leader
Keith Inman	California State University East Bay and Forensic DNA Consultant
Eugene Lien	New York City Office of Chief Medical Examiner, Department of Forensic Biology, DNA Technical Leader
Tamyra Moretti	Federal Bureau of Investigation Laboratory, DNA Support Unit
Lisa Schiermeier-Wood	Virginia Department of Forensic Sciences, DNA Supervisor
Joel Sutton	Defense Forensic Science Center, U.S. Army Criminal Investigation Laboratory, DNA Technical Leader
Ray Wickenheiser	New York State Police Laboratory Director (and president of the American Society of Crime Laboratory Directors, 2017–2018)
Charlotte Word	Independent Forensic DNA Consultant (and former laboratory director at Cellmark Diagnostics)

The Resource Group reviewed an early version of the published draft report and provided valuable feedback, insights, and suggestions during its development. However, they were not asked to provide consensus advice or recommendations, sign off on the final report, or endorse its conclusions. The NIST team is grateful for their dedication and contributions to these efforts.

Input from the Resource Group was requested to: (1) make sure we were addressing real-world problems faced by the community, (2) help define the scope and direction of this project, and (3) provide a sounding board for communications before sharing them with a wider community. This included a review of an early version of this report to ensure that the document was appropriate and helpful. The group met with the NIST team eight times in person and four times by teleconference over an 18-month period (December 2017 to June 2019).

Prior to the first meeting in December 2017, two questions were asked of the invited attendees to serve as a starting point: (1) What is your main concern in DNA mixture analysis today? (2) Where is there room for improvement in DNA testing?

Responses regarding concerns in DNA mixture analysis centered around the following areas, which are listed in no particular order:

- *Defining interpretation limits* so analysts know when to stop attempting to interpret a mixture, especially when only low-level data are available and when it is difficult to differentiate stutter from true alleles of another donor;
- *Delineating interpretation accuracy and reliable use of probabilistic genotyping software* (PGS) and ascertaining whether or not laboratories are adopting new approaches with proper foundation and training needed to create new interpretation protocols;
- *Estimating the number of contributors* and establishing a cutoff for mixtures in terms of the number of contributors who can reliably be distinguished in a particular case;
- *Addressing report writing and content,* including the difficulties of communicating results to law enforcement or attorneys;
- *Recognizing the need to increase consistency/reproducibility in interpretation and report writing* in some cases, within laboratories and across the community; and
- *Acknowledging the need to increase the scope of validation studies particularly for PGS systems* and in subsequent interpretation protocols to more accurately represent the meaning and value of DNA mixture results to law enforcement, attorneys, judges, and juries.

Responses to the question about room for improvement expressed a need for:

- *Standards with "teeth"* (impact or real influence)*,* rather than general guidelines;
- *More publication and dissemination of results* to the community, along with tools to improve;
- *More consistent training* that helps the analyst improve DNA mixture interpretation, as opposed to presentations on research projects that are years away from implementation;

- *More information on validation and implementation* of PGS tools, with training that is hands-on, interactive, and involves critical-thinking exercises;
- *Improved understanding of secondary transfer possibilities;* and
- *More training and continuing education* for analysts and stakeholders.

1.9. Informing Stakeholders

While conducting this scientific foundation review, the authors made several presentations to a wide range of stakeholders, including DNA analysts, technical leaders, academic researchers, students, prosecutors, defense attorneys, and judges. These public presentations enabled the NIST team to keep members of these communities informed about plans and progress being made as well as to receive input such as suggested topics for consideration and articles to add to the literature review.

After the first public presentation regarding this scientific foundation review at the January 2018 SWGDAM meeting, copies of slides and a draft reference list were provided to all known probabilistic genotyping software vendors or developers. Progress made after the first year was summarized in the *Proceedings of the 29th International Symposium on Human Identification* titled "DNA Mixture Interpretation Principles: Insights from the NIST Scientific Foundation Review" (Butler et al. 2018b). Progress after the second year was reported at the 2019 Congress of the International Society for Forensic Genetics (ISFG) (Butler et al. 2019).

Two of the NIST team members prepared an INTERPOL literature review covering forensic DNA articles published between 2016 and 2019, which included information on PGS and DNA mixture interpretation (Butler & Willis 2020). This effort also involved a presentation at the INTERPOL International Forensic Science Managers Symposium in October 2019. The literature from 2019 to 2022 on DNA mixture interpretation and other forensic DNA topics was also summarized in a subsequent INTERPOL review (Butler 2023).

Approximately 120 people attended a full-day workshop held in February 2019 at the American Academy of Forensic Sciences (AAFS) meeting in Baltimore, Maryland. This workshop, titled "DNA Mixture Interpretation Principles: Observations from a NIST Scientific Foundation Review," provided a detailed progress report of our findings and insights from Resource Group members about their experiences participating in the NIST review. A total of 19 presentations were given by the six NIST team members and 11 Resource Group members (AAFS 2019).

In September 2019, three authors of this report – John Butler, Hari Iyer, and Sheila Willis – gave a workshop titled "DNA Mixture Interpretation Principles and Best Practices" in Palm Springs, California as part of the 30th International Symposium on Human Identification (ISHI) (ISHI 2019). In November 2019, John Butler and Hari Iyer gave an hour-long webinar for the Center for Statistics and Applications in Forensic Science (CSAFE) (CSAFE 2019). Members of the NIST team have provided additional workshops and presentations on validation (ISHI 2020), useful literature regarding DNA measurement and interpretation (AAFS 2021, AAFS 2022), key principles involved in DNA mixture interpretation (ISHI 2021), and a brief history of the field (AAFS 2024). Further efforts to keep stakeholders informed include more than two dozen

presentations at various conferences between 2018 and 2024 on aspects of DNA mixture interpretation, as well as our efforts in collecting information and writing this report.

Plans for this DNA mixture interpretation review were announced to the general public in a NIST press release (Press 2017) on October 3, 2017, and through an interview and subsequent ProPublica news article (Kirchner 2017) shortly thereafter. A plain-language summary covering DNA mixtures and difficulties in their interpretation was also shared online during the course of this study (Press 2019). Finally, ongoing NIST research efforts have sought to fill gaps identified (e.g., Riman et al. 2021, Riman et al. 2024a, Riman et al. 2024b, Vallone et al. 2024, Romsos et al. 2024).

1.10. Structure of This Report

This report contains six chapters and two supplemental documents. Following this introductory chapter, Chapter 2 provides background information on DNA and describes principles and practices involved in mixture interpretation. Chapter 3 lists data sources used and strategies to locate them. Chapters 4 and 5, which are the core of the report, discuss issues in DNA mixture interpretation at the sub-source and activity levels. Chapter 6 explores the potential of new technologies to aid DNA mixture interpretation. Finally, two supplemental documents provide a brief history of DNA mixture interpretation (NISTIR 8351sup1) and a summary of publicly accessible validation data and proficiency test results covering DNA mixture interpretation (NISTIR 8351sup2).

2. Chapter 2: DNA Mixture Interpretation: Principles and Practices

DNA mixture interpretation principles and practices are introduced in this chapter. The DNA testing process involves measurement and interpretation. Measurements reflect the physical properties of the sample while interpretation is subjective and depends on the DNA analyst assigning values that based on personal judgments. Multiple statistical approaches are used to answer different questions. This includes strength-of-evidence interpretation, such as the random match probability (for major components of mixtures), the combined probability of inclusion, and the likelihood ratio approaches. DNA samples are not all equal in complexity, and some are more difficult to analyze than others. Factors influencing the complexity include the number of contributors, DNA quantities of components, mixture ratios, sample quality, and the degree of allele sharing. In addition, artifacts created during the process of generating the DNA profile contribute to the challenge of DNA mixture interpretation. Continuous probabilistic genotyping systems, which report a likelihood ratio based on a pair of selected propositions, utilize more information from a DNA profile than previous approaches. The theory and application of likelihood ratios are introduced here in the context of probabilistic genotyping software. The chapter concludes with 16 principles related to DNA mixture interpretation. This information is intended as a precursor to topics covered in other chapters on DNA interpretation at the sub-source level (Chapter 4), DNA interpretation at the activity level (Chapter 5), and the potential of new technology (Chapter 6).

2.1. Value of DNA Evidence to Forensic Science

Forensic science processes involve collection, analysis, interpretation, and reporting on properties of evidence. Since its introduction in the mid-1980s (Gill et al. 1985), DNA testing has been an important resource to forensic science and the criminal justice system. Forensic DNA results provide important capabilities to aid law enforcement investigations, strengthen prosecutions, and enable exoneration of the innocent. These capabilities include (1) the ability to assist in identification of an individual or association of a perpetrator with a crime scene, since DNA remains unchanged throughout life and across bodily cells, (2) high sensitivity with DNA amplification techniques, (3) well-established quality assurance measures, (4) ability to provide a numerical strength of the evidence based on established genetic principles with the help of statistical models, (5) use of close biological relatives as potential reference points through applying established characteristics of genetic inheritance, and (6) new technology development aided by biotechnology and genomics efforts (see Butler 2011, Butler 2014, Butler 2015).

Information about DNA left at a crime scene can assist both law enforcement (investigative) and prosecutorial (evaluative) aspects of the criminal justice system. DNA results may also assist the defense of an accused person with exculpatory information or provide associations of remains with biological relatives during disaster victim identification. Investigative leads may be generated when a crime scene profile or a deconvoluted mixture component of a DNA profile is searched against a local, state, or national DNA database to locate a potential person of interest (POI). When writing reports or providing court testimony, the evaluative strength of available DNA evidence can be assessed when comparing a POI to an evidentiary DNA profile. Investigative and evaluative examinations serve different purposes and answer different questions (Gill et al. 2018).

Successful DNA analysis and resulting interpretation depends on the quality and quantity of the crime scene evidence (the “Q” or questioned sample) and the availability of a reference sample (the “K” or known sample). When appropriate Q and K DNA profiles are available, forensic scientists can perform a Q-to-K comparison and report the strength of this association using specific assumptions and usually one of several statistical approaches. A range of DNA profile qualities and quantities can be observed in forensic casework samples.

2.1.1. DNA Basics

A biological sample collected directly from a single individual (i.e., a “single-source sample”) can be analyzed to generate a *DNA profile*. This profile identifies the genetic variants (termed *alleles*) found at tested locations (*loci* or when singular, *locus*) along the human *genome*. Usually fewer than two dozen loci, which are each in a size range of 100 to 400 *nucleotides* in length, are examined to generate a forensic DNA profile. Thus, information from only a few thousand nucleotides in total are examined in a forensic DNA test out of the approximately three billion nucleotides across 23 pairs of *chromosomes* that comprise the human genome.

Core sets of loci have been selected for use in national DNA databases (e.g., Budowle et al. 1998, Hares 2015). These tested loci, also termed *DNA markers*, were selected from non-

protein-coding regions of the genome occurring between genes. Thus, results from forensic DNA profiles are not expected to contain information on physical traits or susceptibility to genetic diseases (e.g., Katsanis & Wagner 2013).

The DNA markers used in most forensic applications include short genetic sequences that are repeated a variable number of times. These are called *short tandem repeat (STR)* markers. The number of repeats at each STR marker varies from person to person. This variability in STR alleles is what allows a DNA analyst to associate a DNA sample with an individual. A variety of commercially available STR kits have been used over the past 25 years. These kits have evolved and expanded over time using 6 to 10 markers in the mid- to late-1990s, 10 to 16 markers between 2000 and 2013, and 20 to 24 markers or more, presently (see Butler 2011, pp. 108-122 and Butler 2014, pp. 17-21).

Humans are *diploid,* i.e., they possess two copies of each non-sex-determining chromosome (*autosome*) with one allele at each locus coming from an individual's biological mother and the other from their biological father. Thus, alleles at each tested locus exist in pairs, which are termed *genotypes*. Allele pairs that are indistinguishable and cannot be differentiated with the technology used are termed *homozygous.* An analyst might label these 12,12 or A,A. Those genotypes that are distinguishable from one another, in other words, differing alleles that are inherited from each parent, are called *heterozygous.* These might be labeled 12,13 or A,B.

When analyzing the DNA sample, a technique called the *polymerase chain reaction* (PCR) is used to create millions of copies of each STR marker. The purpose of this step, called *amplification*, is to generate a quantity of STR alleles sufficient for laboratory analysis. The PCR process labels STR alleles with different colored fluorescent dyes to enable multiple markers to be examined in a single analysis.

The amplified and labeled STR alleles are then separated and detected using a technique called *capillary electrophoresis* (CE). CE instruments utilize four, five, or six dye channels to analyze many STR markers simultaneously. Peak positions, heights, and sizes relative to a calibration "allelic ladder" standard are visualized by dye-channel color and DNA size in a chart format called an *electropherogram* (EPG). The locations of peaks on the chart indicate which alleles (i.e., STR marker variants of different size) are present in the tested sample. The EPG is the data that must be interpreted to draw conclusions from the sample.

The amplification step using PCR and the separation and detection step using CE are important in the context of this report because they produce artifacts that can confound the EPG's interpretation. These artifacts are discussed in Section 2.2.1 (**Factors that Affect Measurement Reliability**). Analysis of samples containing very small quantities of DNA tends to produce EPGs with a higher proportion of artifacts due to *stochastic variation* or random sampling of DNA molecules (see Butler & Hill 2010).

The amount of DNA recovered from crime scene evidence depends on a number of factors including the amount of biological material deposited, DNA extraction efficiencies, and environmental conditions that can contribute to DNA degradation or PCR inhibition. When degraded, DNA molecules break into smaller pieces, such that some or all of the tested loci are no longer amplifiable by PCR and thus not detectable by CE. Loss of allele information from a

DNA profile is termed *allele drop-out* or, if both alleles are not present or detectable, *locus drop-out*. Swabs from so-called "touch evidence" samples, which typically have a relatively small quantity of biological material deposited (with perhaps tens of cells), are more likely to exhibit loss of allelic signal compared to visible blood or semen stains, which may contain hundreds to thousands of cells in pristine samples not exposed to harsh environmental conditions. Allele drop-out can occur due to allelic signal being below a set detection threshold or because of inadequate sampling of available DNA molecules (from stochastic variation) during the PCR amplification process.

Further details on DNA basics and the process for generating forensic DNA profiles are available in textbooks such as *Fundamentals of Forensic DNA Typing* (Butler 2009) or *An Introduction to Forensic Genetics, Second Edition* (Goodwin et al. 2010).

2.1.2. DNA Mixtures

A DNA mixture can occur when biological material from more than one individual is deposited on the same surface. In single-source samples, only a single genotype is possible at each locus. With DNA mixtures, however, more than one genotype combination may be possible at each locus. This ambiguity is an important reason why DNA mixture interpretation is more difficult than testing single-source samples. Interpretation of evidence, in the words of Peter Gill, a leader in the field, "continues to be the most difficult challenge that faces scientists, lawyers, and judges" (Gill 2019b).

DNA from multiple contributors cannot be physically separated once DNA molecules are extracted from their biological cells (see Chapter 6 and Figure 6.2). Instead, DNA mixture interpretation is an effort to (1) infer possible genotypes as detectable sample contributors (a process sometimes referred to as *deconvolution* of the mixture components) and (2) provide the strength of evidence for a POI to be included in or excluded from an evidentiary DNA profile.

DNA mixtures are common, and even expected, in many evidence types coming from criminal investigations. Person-on-person crimes, such as sexual assaults or homicides, may involve DNA mixtures of biological material (e.g., semen or blood) from the perpetrator and the victim. DNA mixtures may be detected in many property crimes where items in a house or a vehicle are handled by a burglar but also touched previously by the owner(s) or other people not associated with the crime in question.

In their 2016 report, the President's Council of Advisors on Science and Technology (PCAST) differentiated between single-source samples, simple mixtures, and complex mixtures (PCAST 2016). *DNA samples and mixtures in forensic casework exist on a continuum,* although artificial categories have been described (e.g., Wickenheiser 2006, Schneider et al. 2006b, Schneider et al. 2009) to explain where use of different approaches to mixture interpretation may be helpful.

An analogy involving mathematics may assist in explaining aspects of various categories that have been used for DNA profiles. If one considers that single-source DNA profiles are like basic arithmetic and simple mixtures are like algebra, then complex mixtures (e.g., profiles with three or more contributors, with low-level and/or degraded DNA where *uncertainty in assigning*

contributor genotypes increases) can be considered the equivalent of calculus. In a similar manner, calculus builds upon principles of arithmetic and algebra but requires more advanced training and perspective to fully appreciate; so does DNA interpretation of complex mixtures. Validation studies and training are required to develop the necessary expertise. However, the fundamental principles must be understood before approaching complex DNA mixture interpretation.

KEY TAKEAWAY #2.1: DNA mixtures, in which the DNA of more than one individual is present in a sample, are inherently more difficult to interpret than high-quality single-source DNA samples.

2.2. The DNA Testing Process

The general steps involved in forensic DNA testing are illustrated in Figure 2.1. Briefly, an item of evidence is collected, or a sample is obtained by swabbing a surface containing possible crime scene evidence, which has been determined to be of potential value to an investigation. DNA, which could be from one or more contributors, is extracted from the sample. Following DNA extraction, DNA quantitation (with adjustments for amount of human DNA present), and PCR amplification with predefined DNA marker sets of STR loci, the amplification products are separated and detected. Results are then interpreted, compared to reference sample profiles, assigned a statistical strength of evidence, and reported in a written summary. If a case goes to trial, then the analyst might be asked to provide testimony as an expert witness.

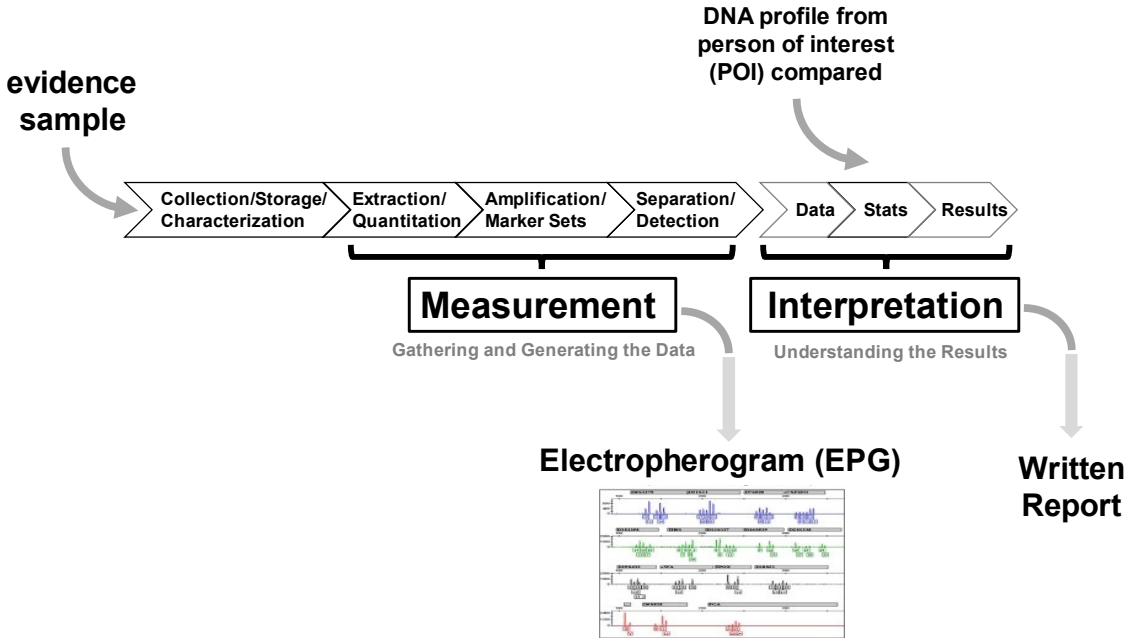

Figure 2.1. Illustration of the general steps involved in processing an evidence sample containing DNA (either single-source or mixture) after the sample is determined to be of potential value to an investigation. The output of the measurement steps is an electropherogram. The output of interpretation is a result in a written report.

This overall process can be divided into two parts (Figure 2.1): (1) ***measurement*** that involves a series of steps to generate a DNA profile and (2) ***interpretation*** of the DNA profile to help fact finders understand the value of the evidence. The measurement steps result in an *electropherogram* (EPG), which is a representation of the DNA profile observed from the test sample at specific DNA locations. Interpretation of the EPG concludes with a written report describing a strength-of-evidence statistic for Q-to-K comparison with the POI(s), and in some cases, court testimony. Statistics are typically not generated until after a Q-to-K comparison is made and usually only with an inclusionary opinion.

Figure 2.1 outlines general steps; however, the details of measurement and interpretation steps may vary between laboratories. For example, different STR kits, PCR cycle numbers, and CE instruments may be used in different laboratories after conducting validation experiments and deciding that a particular method is deemed trustworthy. Likewise, interpretation approaches may differ among analysts and, more often, laboratories. Therefore, general practices and principles involved in measurement and interpretation are discussed rather than one specific protocol. For more details about the variation that exists in current practice, see the 42-page Human Forensic DNA Analysis Process Map[3] created through collaboration among NIST, OSAC, and SWGDAM.

Measurements reflect the physical properties of the sample, such as the number of alleles observed, while interpretation involves the DNA analyst making decisions based on those measurements to, for example, assign the number of contributors. These interpretations are based on case context and their own training and experience in conjunction with laboratory interpretation guidelines and policies developed based on validation studies performed. In part, because interpretation of the same evidence may vary from person to person, it is described as an opinion (see Gill 2019b). Complex DNA mixtures are challenging because they require more interpretation than a high-quality, single-source sample.

When a POI is available for comparison to the evidence, DNA analysts render their opinions (often in the form of likelihood ratios) in written reports drawing upon (1) empirical data from the evidence sample compared to a POI's DNA profile, (2) available case information (e.g., location from which the sample originated, body fluid screening results, quantity of DNA extracted, and overall quality and context of the DNA profile), and (3) their training and experience (see SWGDAM 2017a).

Further details are available in textbooks such as *Interpreting DNA Evidence: Statistical Genetics for Forensic Scientists* (Evett & Weir 1998), *Forensic DNA Evidence Interpretation* (Buckleton et al. 2005, Buckleton et al. 2016), and *Forensic Practitioner's Guide to the Interpretation of Complex DNA Profiles* (Gill et al. 2020b).

[3] Available at https://www.nist.gov/system/files/documents/2022/05/05/OSAC%20Forensic%20Biology%20Process%20Map_5.5.22.pdf (accessed November 1, 2024). This Process Map provides a visual description of the various steps of the casework process performed by DNA analysts and is an attempt to represent all reasonable variations in current practice. NIST, OSAC, and SWGDAM do not explicitly support or endorse (as best practices) all of the different steps and/or paths as depicted on this current-practices process map.

KEY TAKEAWAY #2.2: Generating a DNA profile involves measuring the physical properties of a sample. Interpreting a DNA profile involves the DNA analyst applying their judgment, training, tools (including computer software), and experience, and considers factors such as case context and laboratory protocols and policies.

2.2.1. Factors that Affect Measurement Reliability

The measurement portion of the DNA testing process produces an EPG (see Figure 2.1). DNA mixture interpretation (as well as single-source DNA interpretation) is conducted in the presence of PCR amplification and CE analysis artifacts. These artifacts influence the complexity of the DNA profile to be interpreted, and validation studies are performed to characterize them.

Artifacts that may be observed in an EPG include the non-allelic products of the PCR amplification process (e.g., stutter products, non-templated nucleotide addition, or other non-specific products), anomalies of the detection process (e.g., single or multichannel voltage spikes or "pull-up" from spectral channel bleed-through), or by-products of primer synthesis (e.g., "dye blobs") (see Butler 2014, pp. 183-210).

There are several quantifiable factors that affect measurement variability and reliability.

The first is peak position. The DNA profile peaks observed in an EPG are fluorescently labeled PCR products (STR alleles) that differ in length due to variation in the number of STR repeats. Use of an internal size standard with each tested sample along with calibration to an allelic ladder enables accurate STR allele designations with electrophoresis separation and detection systems (Butler 2014, pp. 48-58). Peak positions are measured as migration time (raw data), nucleotides (against the size standard), and allele designations (against an allelic ladder). Accurate determination of peak locations is necessary for reliable STR allele designations.

Another measurable factor includes peak morphology or resolution. This is when wide peaks result in poor resolution and the inability to fully separate STR alleles that differ by as little as a single nucleotide. In general, capillaries fail and resolution is lost after many CE sample injections. Peak resolution can be monitored by examining separation of the alleles in an allelic ladder (Butler 2014, pp. 201-202). Failure to resolve similar length STR alleles may result in missing true contributor genotypes. Wide peaks may also size inaccurately.

Peak heights are measured in relative fluorescence units (RFUs) and are generally proportional to the amount of PCR product detected. While an RFU value does not necessarily correspond to a specific number of picograms of DNA, variation in peak heights matters because this information is used to deconvolute mixture components into contributor genotype possibilities. On-scale data are essential when calculating results impacted by peak heights, such as stutter percentages and peak height ratios (Butler 2014, pp. 30-33).

Artifacts are anything in the EPG that result from anything other than the DNA profile alleles in the sample. Artifacts introduced during the PCR and CE processes are referred to as amplification and detection artifacts. Understanding the nature of the artifacts is important

because when low quantities of DNA are tested, it can be challenging to differentiate true alleles from amplification or detection artifacts, such as stutter products and pull-up.

Stutter products are produced during PCR amplification from slippage of the DNA strands while being copied, and are typically one repeat shorter or longer than the originating STR allele (Walsh et al. 1996, Butler 2014, pp. 70-79). The relative heights of stutter products correlate in large measure with the length of sequence composed of the same repeat pattern of the corresponding STR allele (Brookes et al. 2012). Stutter products can be indistinguishable from true alleles of minor contributors and therefore can significantly impact DNA interpretation (Gill et al. 2006b).

Spectral artifacts are a measurable factor, as well. They are an anomaly of the detection process where fluorescent signal from one spectral channel "bleeds through" into an adjacent color channel (e.g., green into blue). Pull-up occurs from a saturating signal on the instrument detector (see Butler 2014, pp. 32, 200-201). Spectral artifacts may also signal off-scale data in an EPG that should be avoided, as the stutter ratio will not be accurate.

Relative peak heights of allele pairs within a locus are another measurable factor. Heterozygous STR loci possess two alleles that differ in overall PCR product size. The peak heights of these two "sister" alleles can be compared in single-source samples to enable genotype assumptions in samples containing more than one contributor (Butler 2014, pp. 87-93). This factor is important in determining the limits of pairing alleles into genotypes with binary approaches and also helps define parameters used for assigning potential genotypes and mixture ratios with PGS systems.

Assessing relative peak heights across loci in a DNA profile provides an indication of the quality of a sample. When analyzing degraded DNA, peak heights decrease from left to right across an EPG (small-size to large-size STR alleles) (Butler 2014, pp. 121-123). This factor is important because ratios between mixture components may differ across tested loci.

Finally, baseline noise is also a measurable factor in this context. Noise exists in all measuring systems. In a DNA profile EPG, noise is represented as jitter in the baseline signal (Butler 2014, p. 33). Characterizing the level of baseline noise enables a laboratory to set an analytical threshold and establish a lower limit of reliability for peak heights.

These measurable factors in DNA profile EPGs can affect measurement reliability. Table 2.1 lists validation experiments typically conducted and the purpose of each factor in DNA mixture interpretation. For foundational purposes, it is important to consider what is known about the degree of uncertainty associated with each of these measurements as well as other factors that can influence interpretation, including artifacts. For this reason, studies regarding stutter product variation (e.g., Bright & Curran 2014) and allele drop-in (e.g., Moore et al. 2020) are valuable.

Table 2.1. Measurable factors and features in a short tandem repeat (STR) DNA profile electropherogram (EPG) that influence DNA mixture interpretation with binary or probabilistic genotyping software (PGS) approaches. Assessments of some of these factors are more qualitative than quantitative. Validation experiments (SWGDAM 2016) to demonstrate measurement reliability are typically performed using single-source DNA samples (e.g., Moretti et al. 2001a, Moretti et al. 2001b, Butler et al. 2004, Rowan et al. 2016).

Measurable Factor (units)	**Validation Experiments to Demonstrate Reliability**	**Purpose in DNA Mixture Interpretation**
1a) Peak Position (nucleotides)[a]	Accuracy and precision studies to verify consistency in peak sizing and STR allele calls	To determine limits of peak sizing and accurate allele calls compared to an allelic ladder
1b) Peak Morphology or Resolution	Examination of peak height and width in allelic ladders and inspecting separation of similar length allelic ladder alleles (e.g., TH01 alleles 9.3 and 10) as quality control of kit and instrumentation	To examine CE separation resolution that can influence ability to accurately designate similar length STR alleles (e.g., Butler et al. 2004)
2a) Peak Height (RFU)[b]	Precision studies to verify consistency in allele calls; variability is typically studied in terms of presence or absence; repeatability of peak heights can be investigated with replicate injections and reproducibility of peak heights with replicate PCR amplifications of sample aliquots	To determine the presence of stochastic effects such as allele drop-out (only when examining ground-truth samples); to determine presence of contamination including allele drop-in (only when examining ground-truth samples); to help infer parameters used for assigning potential genotypes and mixture ratios with PGS systems
2b) Stutter Products	Calculation of stutter peak height to STR allele peak height ratio	To determine stutter thresholds applied in binary approaches or to develop and inform stutter models for PGS; multiple types of stutter (e.g., n−1, n−2, n+1) and approaches (e.g., allele-specific, locus-specific, or profile-wide) have been used
2c) Spectral Artifacts	Visual inspection of EPGs for signal "bleed-through" between dye channels (e.g., green into blue) with off-scale peaks; calculation of "bleed-through" to parent peak height ratio; quality control for spectral calibration of system	To determine upper limits of DNA quantities used to generate profile EPG; to help define parameters for distinguishing bleed-through from true peaks

Measurable Factor (units)	**Validation Experiments to Demonstrate Reliability**	**Purpose in DNA Mixture Interpretation**
2d) Relative Peak Heights of Allele Pairs within a Locus	Calculation of heterozygote balance or peak height ratios from heterozygous allele pairs in single-source samples	To determine the limits of pairing alleles into genotypes with binary approaches and to help infer parameters used for assigning potential genotypes and mixture ratios with PGS systems and calculating probability of allele drop-out
2e) Relative Peak Heights Across Loci in a DNA Profile	Calculation of interlocus balance to determine if peak heights are significantly reduced for longer length PCR products (on the right side of the EPG)	To estimate the level of DNA degradation or PCR inhibition (some new STR kits have quality sensors included in the STR profile) and to help infer parameters used for assigning potential genotypes and mixture ratios with PGS systems
2f) Baseline Noise (RFU)	Examination of variation in baseline noise from negative controls and extraction blank samples	To determine the analytical threshold so that noise can be distinguished from true peaks (that can be alleles or artifacts); multiple approaches have been used (e.g., Bregu et al. 2013)

[a] in nucleotides relative to an internal size standard with allele calls made in comparison to an allelic ladder run simultaneously or sequentially with the same internal size standard
[b] relative fluorescence units

A series of single-source samples and negative controls are commonly examined to assess observed variability of these measurable factors including artifact behavior. Greater variability in peak heights leads to greater uncertainty in the possible genotype combinations for contributors in mixture interpretation. These measurable factors are mathematically modeled to create probability distributions with probabilistic genotyping software (e.g., Taylor et al. 2013, Kelly et al. 2018).

2.2.2. Steps in the Interpretation Process

Following a decision regarding whether the profile is of value, interpretation begins with separate evaluations of EPGs from a Q (Question - the evidentiary DNA profile) and a K (Known - the DNA profile of a POI). Data interpretation decisions are made *separately* for Q and K EPGs, in accordance with validation-based interpretation protocols, which includes questions such as "is this a peak or part of baseline noise?," "is this an allele or an artifact or some of both?," "could this DNA profile have come from more than one contributor?," etc. Increasingly, these decisions, which respond to the above questions, are made with assistance from suitable

computer software. If the Q profile appears to be a mixture, then the DNA analyst assesses possible genotype combinations of contributors and compares these possible genotypes with one (or more) POIs.

In 2006, the DNA Commission of the International Society for Forensic Genetics (ISFG) published nine recommendations on DNA mixture interpretation (Gill et al. 2006b). These recommendations, which are summarized elsewhere (NISTIR 8351sup1, Box S1.4), serve as core fundamental principles for working with DNA mixtures. The ISFG recommendations build upon previous work (e.g., Weir et al. 1997, Clayton et al. 1998, Bill et al. 2005) and provide a framework around the steps shown in Figure 2.2. This framework was initially developed for manual interpretation methods with simple, two-person mixtures. However, the concepts also apply to software programs used for examining complex mixtures.

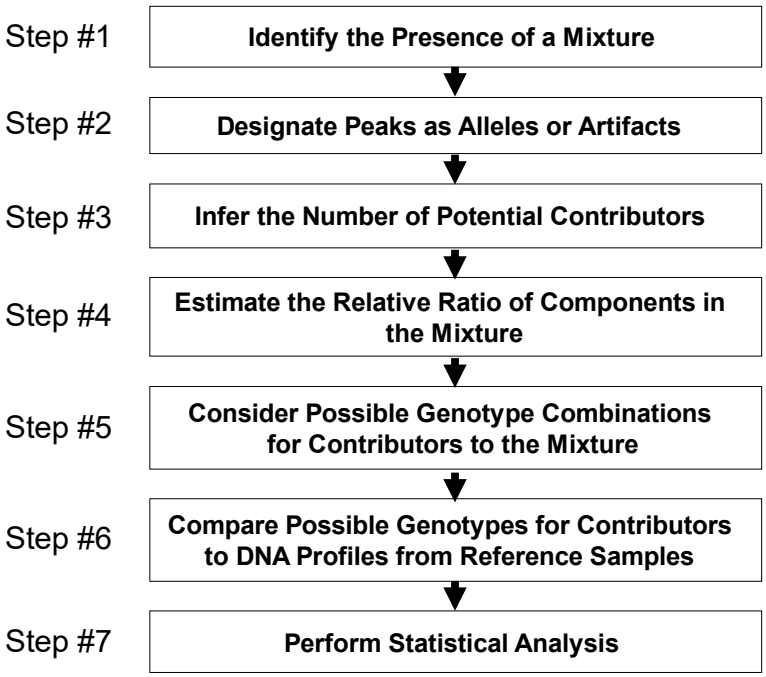

Figure 2.2. Steps in DNA mixture interpretation first outlined by the UK Forensic Science Service (Clayton et al. 1998) and endorsed by the ISFG DNA Commission (Gill et al. 2006b).

The 2006 ISFG DNA Commission noted that there are three kinds of alleles in a crime scene profile. There are alleles that (1) are unmistakable, (2) may be masked by an artifact such as stutter, and (3) have dropped out completely and are therefore not detected (Gill et al. 2006b). When assessing possible genotype combinations of contributors to a mixture, a DNA analyst may encounter any or all of three situations along with allele-sharing among contributors.

Alleles may contain components from more than one contributor that are shared and need to be deconvoluted (i.e., separated out into component genotypes). More possible contributors mean more possible genotype combinations with any of the observed set of alleles. The creation of computer software to explore possible genotype combinations has been an important development in DNA mixture interpretation (Coble & Bright 2019).

For a detailed analysis of these interpretation steps using an example DNA mixture and the various statistical approaches discussed later in this chapter, see *Advanced Topics in Forensic DNA Typing: Interpretation* (Butler 2014, pp. 129-158 and pp. 537-567).

2.3. Complexity and Ambiguity in Mixture Interpretation

DNA samples recovered from crime scenes vary in quality and may be challenging to analyze and interpret (Word 2011). The types of cases being submitted to a laboratory will impact the complexity of mixtures observed (e.g., Torres et al. 2003, Mapes et al. 2016). Over the past decades as DNA testing methods have become more sensitive (see NISTIR 8351sup1), more challenging evidence types (e.g., touch evidence with limited quantities of DNA and complex DNA mixtures) have been submitted to forensic laboratories (Mapes et al. 2016). A "complex" DNA mixture sample is one in which uncertainty exists in the genotype assignments at tested STR loci in a DNA profile, which is more likely to occur in samples with three or more contributors.

2.3.1. Factors that Contribute to Increased Complexity

There are at least three challenges that are fundamental to DNA mixture interpretation: (1) ***stochastic variation***, which impacts recovered quantities of alleles from contributors and can lead to uncertainty in assigning alleles to genotypes and uncertainty in assigning genotypes to contributor profiles when examining small amounts of DNA, (2) ***stutter products***, which create uncertainty when assigning peaks as minor contributor(s) with alleles in the stutter positions of major contributor(s) alleles, and (3) ***sharing of common alleles***, which influences the ability to estimate the number of contributors, particularly when combined with stochastic variation and the existence of stutter products that create uncertainty in deconvoluting mixture components.

Ambiguity in DNA mixture interpretation arises when (1) small quantities of DNA are tested that, when copied, may not fully represent the original sample (i.e., the recovered DNA profile is incomplete and missing information), (2) a mixture of DNA from more than one individual may make it hard to deconvolute or separate information from each individual contributor depending on the contributor ratios, their amounts, and degree of allele overlap, (3) the DNA molecules may be damaged or destroyed (i.e., the recovered DNA profile is incomplete and may be missing information), (4) environmental contamination may impact the ability to recover the original sample (DNA may come from a transfer not related to the crime or PCR inhibitors that lead to an incomplete recovered DNA profile), or (5) any combination of these four issues.

2.3.2. Improved Sensitivity Methods Can Result in Higher Complexity Profiles

As techniques for generating DNA profiles become more sensitive, smaller amounts of DNA can be detected, analyzed, and interpreted. DNA testing sensitivity has increased due to improvements in STR kits (e.g., Ensenberger et al. 2016, Ludeman et al. 2018), introduction of new CE instruments, use of higher PCR cycle numbers (e.g., Whitaker et al. 2001), reduced

volume PCR (e.g., Leclair et al. 2003), PCR product desalting (e.g., Smith & Ballantyne 2007), and higher CE injection conditions (e.g., Westen et al. 2009). "High-sensitivity" DNA testing has become the new normal (Gill et al. 2015).

When analyzing small quantities of DNA, stochastic (random sampling) effects can cause alleles that are present in the sample to "drop out" of the detected profile (e.g., Lohmueller & Rudin 2013). Stochastic effects can also cause alleles that are not present in the sample to "drop in" to the profile (e.g., Gill et al. 2000, Moore et al. 2020). In other words, with low-quantity DNA samples, the resulting profile and EPG may vary in how accurately they reflect the original sample, which can lead to loss of genotype information from a true contributor to the mixture.

Furthermore, in part due to stochastic variation, two low-quantity DNA samples collected from the same surface can produce DNA profiles with different peak heights and therefore different ratios of alleles and possible genotype combinations. Analyzing the same low-quantity DNA mixture two or more times can also produce dissimilar DNA profiles with different degrees of stochastic variation (e.g., Benschop et al. 2013).

KEY TAKEAWAY #2.3: The process of generating a DNA profile from low amounts of DNA template can produce stochastic or random variation and artifacts that contribute to the challenge of DNA mixture interpretation.

2.3.3. Mixture Complexity Increases as Number of Contributors Increase

The challenge of genotype assignment increases with the number of contributors in a mixture due to the possibility of allele sharing (Alfonse et al. 2017). In addition, estimating the number of contributors in a DNA mixture becomes more uncertain when there are more contributors as noted in several publications (Paoletti et al. 2005, Buckleton et al. 2007, Coble et al. 2015). The frequency of occurrence for an allele from population data correlates to the degree of allele sharing that is expected if that allele is present in the crime scene DNA mixture. If mixture contributors are close genetic relatives, then even more allele sharing between contributors is expected. Thus, with more contributors to a mixture, more allele sharing occurs, which increases the complexity and ambiguity of interpretation (e.g., Dembinski et al. 2018, Lynch & Cotton 2018).

KEY TAKEAWAY #2.4: DNA mixtures vary in complexity. Factors that contribute to complexity include the number of contributors, the quantity of DNA from each contributor, contributor mixture ratios, sample quality, and the degree of allele sharing. These factors will impact the degree of variation in interpretation.

2.4. Approaches and Models for Dealing with Complexity

DNA mixture interpretation can be divided into two general approaches: (1) binary (e.g., Budowle et al. 2009) or (2) probabilistic genotyping (e.g., Gill et al. 2012). Both approaches generally follow the seven steps outlined in Figure 2.2 with an important difference at step five, where possible genotype combinations of contributors are considered.

2.4.1. Binary Statistical Approaches

Statistical analysis provides a quantitative expression of the strength or value of the evidence when K is considered as a possible contributor to the evidence sample Q. When a DNA analyst evaluates a mixture and determines that a *major* component can be confidently separated from a *minor* component of a mixture, then a random match probability (RMP) or modified RMP (mRMP) method has been used on the major component – treating it statistically as a single-source sample (DAB 2000, Bille et al. 2013, SWGDAM 2017a). Likewise, conditioning on the donor of an intimate sample under the assumption of a defined number of contributors has been used to perform mRMP calculations on the foreign profile even if it is not the major component (see SWGDAM 2017a). For mixture contributors that cannot be confidently distinguished because of allele overlap or similar mixture ratios, then "manual" likelihood ratio (LR) methods have been used (e.g., Weir et al. 1997, Evett & Weir 1998, Gill et al. 2006b). Either of these approaches can be applied with simple, two-person mixtures, such as sexual assault intimate samples. These types of calculations become more challenging as the number of contributors exceeds two people.

A commonly used statistical approach in the United States has been the combined probability of inclusion (CPI), which is defined as the probability that a randomly chosen (unrelated) individual would be included as a possible contributor to the mixture (NRC 1992, Bieber et al. 2016). Once a K is included as a possible contributor to Q, the CPI, which is sometimes referred to as random man not excluded (RMNE), indicates the statistical value of all possible genotypes present in a mixture (giving them equal weight) based on observed alleles (NRC 1992, p. 59).

As seen in Table 2.2, different statistical approaches answer different questions using the data available. Each approach has strengths and weaknesses (e.g., Buckleton & Curran 2008). A trier of fact in a court of law is typically interested in what DNA results mean in a particular case, with regard to a specific POI and set of case circumstances. For this reason, likelihood ratio methods (Question 4 in Table 2.2), as will be discussed later in this chapter, have been considered a valuable tool in DNA mixture interpretation and recommended by the ISFG DNA Commission (Gill et al. 2006b, see also NISTIR 8351sup1).

Table 2.2. Different approaches used in statistical analysis of DNA and the questions addressed. RMP and MP are calculated for single-source DNA profiles (or deduced major profiles). CPI and LR are used for mixtures.

	Question	Approach (Reference)	Specific Requirements
1	What is the probability of observing this profile in the population? (i.e., what is the rarity of the profile?)	Profile Probability (or random match probability, RMP) (NRC 1996 for single-source samples; Bille et al. 2013 for mixtures)	For mixtures, an assumption that the major contributor can be distinguished from minor components so that specific genotypes in the major can be inferred
2	What is the probability of observing this profile in the population if one person with this profile has already been observed in this population?	Match Probability (MP) (Balding & Nichols 1994, Weir 2001)	Use of conditional probabilities and a subpopulation correction
3	What is the probability that a person selected randomly in the population would be included (or not excluded) as a possible donor of the DNA typing result?	Combined Probability of Inclusion (CPI) (Bieber et al. 2016)	All alleles for all contributors are all present at the reported loci (i.e., cannot cope with allele drop-out that is expected with low quantities of DNA); an estimate of the number of contributors is not required
4	By how much do the DNA typing results support the person of interest (POI) being the donor under specific assumptions and propositions?	Likelihood Ratio (LR) (Evett & Weir 1998)	Assumptions made in the selection of a statistical model, the number of contributors, and a specific pair of propositions

2.4.2. Limitations with Binary Methods

Traditional *binary* methods and approaches to DNA mixture interpretation (e.g., Clayton et al. 1998) work under the assumption that a specific genotype of interest is either present or absent. Statistical approaches include LR (e.g., Weir et al. 1997), CPI (e.g., Budowle et al. 2009), and mRMP (Bille et al. 2013). However, binary approaches cannot account for the possibility of missing information (i.e., allele drop-out) when testing small quantities of DNA, nor can they

account for the possibility of allele drop-in, which is more common with high-sensitivity methods (Balding & Buckleton 2009).

As noted in Peter Gill and colleagues' 2020 textbook:

> "These [CPI] calculations found favor and were widely used, because they were very easy to implement and assumptions about the number of contributors were not needed. There are two drawbacks however: (1) There is an implicit assumption that all of the contributors have all alleles fully represented in the EPG. There is no allele drop-out present, i.e., the calculation is not valid for minor contributors with drop-out that is or may be present. (2) The calculation exists by itself and is unchanged by the suspect's profile, i.e., the calculation is unmodified by the presence of a suspect who matches or does not match … When an RMNE is reported, then it is necessary to make a binary decision about whether a suspect could have contributed to a crime stain. Either he has (probability = 1) or he has not (probability = 0)" (Gill et al. 2020b, p. 386).

Thus, proper application of a CPI calculation is dependent on all possible alleles being present and therefore commonly involves use of a stochastic threshold to provide confidence that loci used in statistical calculations are not missing alleles (Moretti et al. 2001a, Moretti et al. 2001b, Budowle et al. 2009, SWGDAM 2017a). In addition to the CPI statistic not accounting for the possibility of allele drop-out when testing small quantities of DNA, this same limitation exists for minor components of complex mixtures, even when the total DNA input is optimal. Guidance on the appropriate application of CPI has been published (e.g., Bieber et al. 2016, Buckleton et al. 2016, pp. 238-247).

In a binary approach, measurement limitations and stochastic effects can make it difficult to identify which of the peaks in an EPG correspond to alleles, which are stutter products, and which are noise peaks. During the PCR amplification process, certain alleles present in the original sample may not have a corresponding peak in the EPG (failure to amplify) or may be judged as absent (below a predetermined analytical threshold), and certain peaks in the EPG that are artifacts may be judged to be real alleles from the original sample (e.g., stutter products, allele drop-ins, spectral pull-up peaks).

To address the complexity that comes with increased DNA sensitivity (Gill et al. 2000), leaders in the forensic DNA community have looked to probabilistic genotyping over the last two decades (see NISTIR 8351sup1).

2.4.3. Advantages of Probabilistic Genotyping Approaches

Probabilistic genotyping approaches can address complexity in DNA profiles. In their 2006 publication, the ISFG DNA Commission concluded:

> "A future approach would elaborate the combinatorial approaches by taking into account all aspects including stutter, contamination and other artefacts, allelic drop-out, such as using a probabilistic weighting for each possible genotype rather than just using a weighting of zero or one, as is inherent in the restricted combinatorial (binary) approach" (Gill et al. 2006b).

The first three authors of the 2006 ISFG DNA Commission article (Peter Gill, Charles Brenner, and John Buckleton) have been involved in developing probabilistic genotyping software systems over the past decade.

Probabilistic genotyping enables weighting (based on the probability of) specific genotype contributions through biological and statistical models informed by probabilities of missing alleles (Kelly et al. 2014, Gill et al. 2020b). These methods incorporate mathematical modeling that can reflect uncertainty in genotype combinations for the mixture interpretation. PGS uses assigned LR values to quantify the strength of the evidence, where the probabilities of the data being observed are compared under two hypotheses or propositions. Depending on the propositions used and probabilistic genotyping models applied, different LRs can be produced (see Gill et al. 2018).

Probabilistic genotyping considers possible genotype combinations for contributors when information may be missing in a crime scene DNA profile (Gill et al. 2012). Two different probabilistic genotyping approaches have been used: discrete or continuous (Kelly et al. 2014, Gill et al. 2015). Table 2.3 compares binary and probabilistic genotyping approaches to DNA mixture interpretation.

Table 2.3. Comparison of approaches used in DNA mixture interpretation. CPI = combined probability of inclusion, mRMP = modified random match probability, LR = likelihood ratio. Adapted from ISFG 2015 workshop by John Butler and Simone Gittelson available at https://strbase-archive.nist.gov/training/ISFG2015-Basic-STR-Interpretation-Workshop.pdf (accessed November 1, 2024).

	Takes into account		***Mathematically models***	
	Presence/ absence of alleles	Possible genotypes based on peak heights	Allele drop-out and allele drop-in	Peak heights
Binary Approaches				
CPI	X			
mRMP	X	X		
LR (binary)	X	X		
Probabilistic Genotyping				
LR (discrete)	X		X	
LR (continuous)	X	X	X	X

Discrete approaches (sometimes referred to as semi-continuous) require the analyst to determine the presence of alleles and artifacts prior to use in their models. Potential allele drop-out or allele drop-in are accommodated without considering parameters such as peak heights, peak height ratios, mixture ratios, or stutter percentages (e.g., Balding & Buckleton 2009, Inman et al. 2015).

Continuous approaches (sometimes called fully continuous) use all observed alleles and their corresponding peak height information and accommodate potential allele drop-out or allele drop-in, while also incorporating information regarding peak height ratios, mixture ratios, and stutter percentages. Some continuous models even consider amplification efficiencies, degradation, and other factors (e.g., Perlin et al. 2011, Taylor et al. 2013, Cowell et al. 2015). Probabilistic genotyping models utilize more information from the DNA profile and thus outperform CPI and other binary approaches.

KEY TAKEAWAY #2.5: Continuous probabilistic genotyping software (PGS) methods utilize more information from a DNA profile than binary or discrete PGS approaches.

2.5. Likelihood Ratios: Introduction to Theory and Application

Dennis Lindley, a modern pioneer in using Bayesian statistics, introduced the concept of likelihood ratios (LRs) to forensic science more than four decades ago (Lindley 1977). LRs were first applied to DNA mixture interpretation about 14 years later (Evett et al. 1991; see NISTIR 8351sup1). The LR involves a ratio of two conditional probabilities: the probability of the evidence given that one proposition (hypothesis or narrative) is true and the probability of the evidence given an alternative proposition is true. The magnitude of the LR value is commonly used to express a strength of the evidence in support of one proposition versus an alternative proposition.

Numerical results obtained from assigning LR values are dependent on the evidence available, statistical models applied, propositions selected based on case information, and the scientist making various judgments. LR results vary based on amount of information available and assumptions made. With less information (e.g., results from a partial DNA profile possessing fewer loci), a lower LR number should be obtained with a well-calibrated system (Meuwly et al. 2017).

2.5.1. Likelihood Ratio Framework

The LR framework or paradigm is linked to Bayes Theorem, which is attributed to an eighteenth-century clergyman named Thomas Bayes (Bayes 1763). Bayesian statisticians[4] define the probability of an event as *the degree of belief* in the truth of the proposition that asserts it will occur. An individual's degree of belief is updated, in light of any new information, by multiplying the individual's prior degree of belief the event will occur (expressed as odds) by their LR to obtain their posterior degree of belief (expressed as odds). The Bayesian framework is based on the philosophical viewpoint that all probabilities are *personal*, meaning[5] "of, relating to, or coming as from a particular person." Probabilities quantify a personal state of uncertainty regarding the truth of propositions (see Lindley 2014, pp. 1 and 19, Kadane 2011, p. 1).

[4] See https://en.wikipedia.org/wiki/Bayesian_statistics (accessed November 1, 2024)
[5] See https://www.dictionary.com/browse/personal (accessed November 1, 2024)

The term *assigning* is used when describing LR results (e.g., Bright & Coble 2020) rather than "calculating" to reflect dependence on subjective (personal, not arbitrary) judgments. That is, different people may assign different values to the same evidence. Concerns have been raised that the LR framework applies only to personal decision making and cannot automatically be used for the transfer of information from one expert to a separate decision maker (Lund & Iyer 2017, Lund & Iyer 2024) and that the Bayesian framework can violate presumptions of innocence in the forensic setting (Stiffelman 2019, Stiffelman 2021). Others have argued that the role of the forensic scientist in providing an LR value is to advise the fact finders and that any consideration of the prior and posterior odds of the propositions is left to these fact finders who can accept, reject, or adapt the scientist's LR (Aiken et al. 2018, Aiken & Nordgaard 2018, Gittelson et al. 2018, Buckleton et al. 2020).

The LR framework (Jackson et al. 2006) offers a way in DNA mixture interpretation (e.g., NRC 1996, Gill et al. 2006b) to report an expert's opinion regarding the strength of evidence (E), which comes from the comparison of Q and K, in support of one proposition (H_1) over an alternative proposition (H_2). For example, H_1 may be that the POI (and in some cases, specific other individuals) contributed to the crime sample vs. a chosen alternative proposition H_2 stating that the POI is a non-contributor to the mixture.

An LR is defined as the ratio of the probability of the findings given H_1 is true versus the probability of the findings given H_2 is true. Note that a reported LR value is *not* the odds that a particular proposition is true. The probabilities are assessed considering other relevant background information as well as various assumptions needed for chosen statistical models, often denoted as *I*.

Symbolically,

$$LR = \frac{\Pr(E|H_1, I)}{\Pr(E|H_2, I)}.$$

Different approaches and statistical models can be used within the LR framework. For DNA mixture interpretation, these include binary, discrete (semi-continuous), and continuous (fully continuous) models and approaches (e.g., Kelly et al. 2014, Bille et al. 2014).

2.5.2. Communicating LR Results

Likelihood ratios are sometimes thought of in terms of weight-comparing scales. Initially, the scale starts with the presumption of innocence or with a belief that H_2 is more likely to be true than H_1. When an LR is greater than one, the scale tips (from its initial position) in the direction favoring H_1. When an LR is less than one, the scale tips (from its initial position) in the direction favoring H_2. The magnitude of an LR is a reflection of how far the scale has tipped from its initial position in the direction favoring H_1 or in the direction favoring H_2. An LR numeric value is not a measurement of a physical quantity. Rather, it is a ratio of probabilities in the opinion of the expert assessing the LR and is dependent on the specific propositions used to formulate it and also on the individual making the assessment.

A common problem known as "transposing the conditional" (Evett 1995) or committing the "prosecutor's fallacy" (Thompson & Schumann 1987) can lead to a misunderstanding of the meaning of an LR. In these situations, a user confuses "the probability of the evidence given the propositions" with "the probability of the propositions given the evidence." This confusion comes from misinterpreting the conditional probabilities used: rather than Pr(E|H), or the probability of the evidence if (or given) the proposition is true, the terms are effectively reversed to Pr(H|E), or the probability of the proposition given the evidence.

A commonly used example illustrates the impact of transposing the conditional:

> "*The probability that an animal has four legs if it is a cow is one"* does not mean the same thing as "*the probability that an animal is a cow if it has four legs is one*." (Evett 1995).

If rewritten in symbols, Pr(four legs|cow) = 1 is not equivalent to Pr(cow|four legs) = 1. The second statement is false since horses, dogs, cats, and other animals also have four legs. Even the first statement, Pr(four legs|cow) = 1, assumes that rare situations of cows with missing limbs are not considered.

For DNA evidence, a statement such as "The DNA results are one million times more likely to have come from Person X than anyone else" transposes the conditional. This statement emphasizes the proposition rather than the evidence. An appropriate way to report this LR result would be "The DNA results are one million times more likely to be observed *if* the evidence came from Person X than *if* the evidence came from Person Y." The inclusion of the word "if" emphasizes the conditional probabilities and assumptions made in assigning the LR value. It is always the trier-of-fact's final decision whether the DNA originates from a specific person or not, and the relevance of this information. Additionally, as noted by a group of statisticians: "To update their prior odds to their posterior odds, a trier of fact must assign their own LR" (Gittelson et al. 2018).

In an effort to describe the relative significance of their results, some forensic scientists use a verbal scale in conjunction with the LR to communicate the probative value of the evidence (e.g., Marquis et al. 2016). A verbal scale recommended by the SWGDAM Ad Hoc Working Group on Genotyping Results Reported as Likelihood Ratios includes categories of *uninformative* (LR=1), *limited support* (LR = 2 to <100), *moderate support* (LR = 100 to <10,000), *strong support* (LR = 10,000 to < 1 million), and *very strong support* (LR > 1 million) (SWGDAM 2018). This SWGDAM verbal scale was adopted in September 2018 as part of the Department of Justice Uniform Language for Testimony and Reports for Forensic Autosomal DNA Examinations Using Probabilistic Genotyping Systems[6].

In their 2020 book *Forensic DNA Profiling: A Practical Guide to Assigning Likelihood Ratios*, authors Jo-Anne Bright and Michael Coble note (pp. 30-31):

> "There has been some justifiable criticism that LRs are not understood by our audience. The use of words to represent the strength of evidence has been proposed as a way to supplement numerical LR evidence. The assignment of words to a numerical LR scale is,

[6] https://www.justice.gov/olp/page/file/1095961/download (accessed November 1, 2024)

> of course, arbitrary…and there are a number of different scales used around the world for different jurisdictions" (Bright & Coble 2020; see also Thompson & Newman 2015).

2.5.3. Probabilistic Genotyping Software

A number of software programs have been developed to assist analysts in performing DNA mixture interpretation by computing LRs using discrete or continuous approaches (Coble & Bright 2019, Butler & Willis 2020). Probabilistic genotyping software (PGS) systems utilize statistical genetics, biological models, computer algorithms (e.g., Graversen & Lauritzen 2015), and probability distributions to infer possible genotypes and assign LRs using either discrete or continuous approaches. Examples of discrete PGS systems include LRmix (Gill & Haned 2013), likeLTD (Balding 2013), Lab Retriever (Inman et al. 2015), or LiRa (Puch-Solis & Clayton 2014). Examples of continuous models include EuroForMix (Bleka et al. 2016a), STRmix (Taylor et al. 2013), and TrueAllele (Perlin et al. 2011).

A PGS system assists a DNA analyst with deconvolution of information in mixtures and assigns a statistical value of the DNA comparison in the context of the case, contributing to the "stats" portion of the interpretation process illustrated in Figure 2.1. Some PGS systems use Markov chain Monte Carlo (MCMC) simulations to estimate weighted genotype possibilities and assess possible combinations of parameters considered in deconvoluting potential contributor genotypes (e.g., Curran 2008, Buckleton et al. 2016, p. 287-293).

A PGS system computes LR values based on the information provided (Figure 2.3), including (1) *modeling choices* made by the system architect(s), (2) *data input choices* made by the analyst regarding an analytical threshold[7] for calling peaks as alleles, selecting the number of contributors to the mixture for use in PGS calculations, and sometimes categorizing artifacts (e.g., pull-up peaks), (3) *proposition choices and assumptions* made by the analyst (e.g., use of unrelated individuals versus relatives, conditioning on a victim when analyzing an intimate sample (i.e., assuming that collecting a swab from an individual's body should contain that individual's DNA), and underestimating or overestimating the number of contributors), and (4) *population database choices* used by the laboratory to provide allele and genotype frequency estimates including using or not using subpopulation correction and if using, the value selected.

An increasing number of forensic laboratories are using PGS for DNA mixture interpretation. The UK Forensic Science Regulator shared seven benefits of PGS compared to manual calculations (UKFSR 2018b, p. 8): (1) increased consistency within and between organizations utilizing the same software, (2) information available in the profile is used more efficiently, (3) deconvolution of genotypes enabling database searches that would not otherwise be feasible, (4) improved reliability due to increased automation in processing, (5) reduced variability between analysts in deciding whether peaks are true alleles or artifacts, (6) increased range of DNA profiles suitable for interpretation, and (7) publication of statistical models in peer-reviewed journals as demonstrations of scientific acceptance.

[7] Some PGS systems do not advocate use of an analytical threshold and may utilize a lower peak detection threshold, such as 10 RFU.

While PGS can assist in interpretation of complex DNA mixtures, "a computer program does not replace the need to think carefully about the case" (Gill et al. 2015). Thinking carefully about a case involves assigning an LR using propositions that address case-relevant questions.

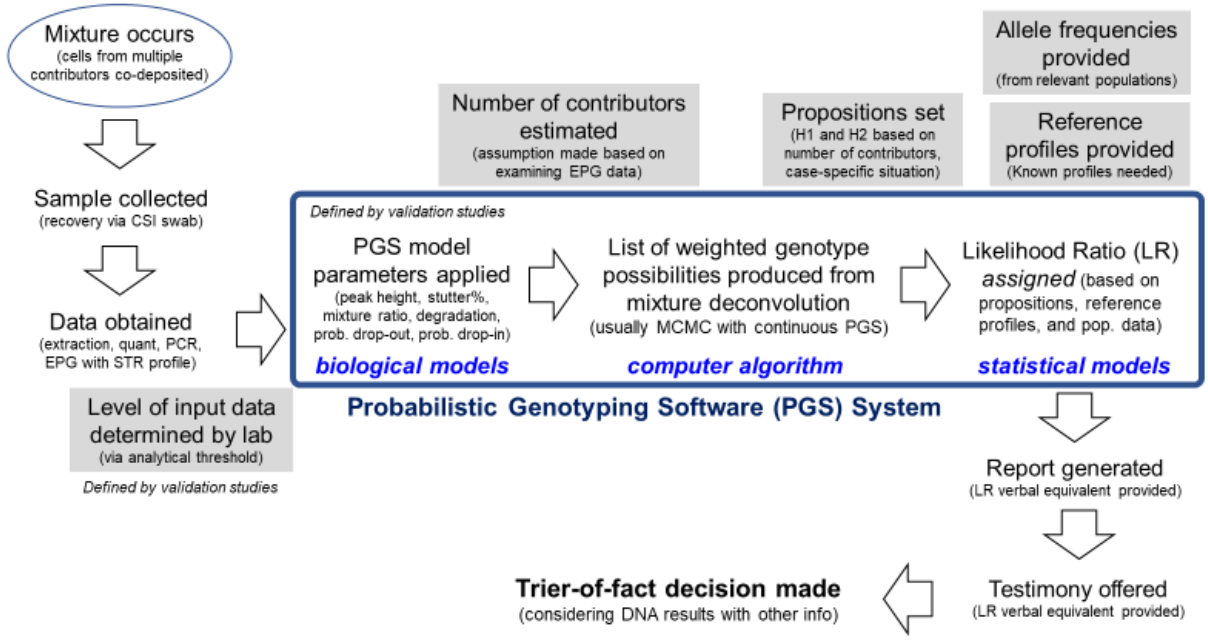

Figure 2.3. Illustration of aspects of a PGS system along with inputs needed (grey shaded boxes). Abbreviations: CSI = crime scene investigation, EPG = electropherogram, LR = likelihood ratio, MCMC = Markov chain Monte Carlo, PGS = probabilistic genotyping software, STR = short tandem repeat. Adapted from Butler & Willis 2020.

2.5.4. Propositions Impact LR Results

As noted by a group of statisticians and forensic scientists, selection of propositions is a vital part of LR assignment:

> "…the choice of these propositions depends on the case information and the allegations of each of the parties. This dependence is unavoidable for the forensic scientist to be able to accomplish his/her duty of presenting what the DNA results mean with regard to the issue of interest to the court" (Gittelson et al. 2016).

LR results vary when different propositions and assumptions are used. The guidance from the UK Forensic Science Regulator on DNA mixture interpretation emphasizes the need to record in the case file the reasoning used by the analyst to support the propositions selected (UKFSR 2018a). The magnitude of this variation can be observed with worked examples using the same data set (Table 2.4). With PGS, propositions are typically arranged as follows, assuming a number of contributors (*N*) who are unrelated to each other and to the POI:

H_1: POI + (*N*-1) unknown, unrelated contributors to the crime sample

H_2: *N* unknown, unrelated contributors to the crime sample

In Chapter 7 of the 2020 book *Forensic DNA Profiling: A Practical Guide to Assigning Likelihood Ratios* (Bright & Coble 2020), the authors provide detailed, worked examples using a two-locus DNA profile (involving D16S539 and D2S1338) with all observed alleles above the analytical threshold. Assuming two contributors, genotype weights were estimated using a PGS system. A person of interest was typed at these loci and could not be excluded as a possible contributor to the mixture. Caucasian allele frequencies from a published data set (Moretti et al. 2016) were used in calculations performed. The same EPG data were examined under four different sets of propositions and assumptions. The LRs varied from over 4,000 (*moderate support* on SWGDAM 2018 verbal scale) to less than 10 (*limited support*) depending on the propositions and assumptions made (Table 2.4). These LR results were all determined at the sub-source level on the *hierarchy of propositions* (see Gill 2001, Taylor et al. 2018).

The highest LR in Table 2.4 occurred when conditioning on the victim, meaning that the victim's genotypes are expected and assumed to be present at each locus in the EPG. This conditioning removes some ambiguity in the possible genotype combinations, which leads to a higher LR for the POI under consideration.

Another possible source of variation in LRs comprises the estimated degree of co-ancestry in observed alleles, which involves using a subpopulation correction factor, Fst, often symbolized by the Greek letter theta (Balding & Nichols 1994, NRC 1996). Using different assumptions in the genetic model (e.g., without or with a 1% subpopulation correction, $\theta = 0.01$), the LR changes from 2895 to 1144.

Finally, the lowest LR in Table 2.4 comes from considering a possible untested brother rather than an unrelated individual in the propositions (i.e., "the DNA evidence [is] eight times more likely if it had originated from the POI and one other individual, rather than if it had originated from a brother of the POI and one unknown individual selected at random from the population" (Bright & Coble 2020)). Even considering only two loci, LR assignments can differ by several orders of magnitude. *Differences in assigned LRs are expected to occur when propositions change and the questions under consideration differ.*

Table 2.4. Summary of LR results from worked examples with two STR loci using different propositions and assumptions (information from Bright & Coble 2020). For information on NRC II 4.2, see NRC 1996.

Pages in book with worked example details	Summary of Propositions and Assumptions Used	LR Result
pp. 160-161	Conditioning on the victim	4143
pp. 148-150	Using the product rule ($\theta = 0$)	2895
pp. 150-153	Using NRC II 4.2 ($\theta = 0.01$)	1144
pp. 151,154-160	With possible untested brother	8

Providing relevant answers depends on asking the right questions. In a review of the 1996 NRC II report (NRC 1996), several authors note:

> "At best DNA profiling can provide very strong evidence of association between people and places. It does not address ultimate questions of guilt or innocence" (Chambers et al. 1997).

Earlier in their article, these authors point out:

> "It should be accepted that there is now no dispute about the potential for DNA analysis to identify[8] individuals, *subject to the constraints imposed by the quality of the evidential samples*" (Chambers et al. 1997, emphasis added).

More recently the following suggestion has been provided by a group of statisticians and forensic scientists:

> "The need to formalize one's propositions for assigning an LR may act as a beneficial restraint. If it is simply not possible to form propositions, then maybe the situation is beyond interpretation" (Gittelson et al. 2016).

DNA mixture interpretation is performed in the face of uncertainty of possible genotype combinations for contributors. As noted by Ian Evett and Bruce Weir in their 1998 book:

> "The origins of crime scene stains are not known with certainty, although these stains may match samples from specific people. The language of probability is designed to allow numerical statements about uncertainty, and we need to recognize that *probabilities are assigned by people rather than being inherent physical quantities*" (Evett & Weir 1998, p. 21, emphasis added).

KEY TAKEAWAY #2.6: Likelihood ratios are assigned and not measured. Different individuals may assign different LR values, even when using PGS systems, when presented with the same evidence because they base their judgments on different collection protocols, quantification systems, STR kit results, interpretation protocols, models, assumptions, or computational algorithms. For any given sample, there is no single, true likelihood ratio.

2.6. Hierarchy of Propositions

Interpretation using an LR framework considers findings given a pair of competing propositions (e.g., the POI is in the mixture versus the POI is not in the mixture or the POI is the source of the bloodstain versus an unknown, unrelated person is the source of the bloodstain). As seen in Table 2.4, considering the same findings under different propositions leads to different LRs.

The type of propositions considered can be grouped into a hierarchy of propositions, where the different levels address different questions (Table 2.5). The initial levels of offense, activity, and source were first introduced by the UK Forensic Science Service in the late 1990s (Cook et al. 1998b) and have expanded to sub-source (DNA level) and sub-sub-source (DNA contributor level) with high-sensitivity methods and DNA mixtures (see Gill 2001, Taylor et al. 2014, Taylor et al. 2018).

[8] A public comment on the draft report correctly noted in response to this quote: "DNA contributes to [identification efforts] but one cannot identify a person only based on DNA" (PC30 in 8351-draft PCs).

The highly sensitive DNA testing methods used today are capable of producing results from small amounts of DNA coming from non-visible stains. This capability changes conclusions that can be drawn as noted in an article titled “A logical framework for forensic DNA interpretation”:

> “When the hierarchy of propositions was first suggested, it was only possible to obtain a DNA profile from biological fluids present in relatively large quantities. In such cases, one could reasonably assume that the DNA profile was derived from a known biological fluid (e.g., blood). This assumption became questionable with the advent of more sensitive techniques” (Hicks et al. 2022).

Chapter 4 discusses DNA interpretation issues at the sub-source level while Chapter 5 introduces DNA interpretation at the activity level. Both aspects are important. A general science guide on forensic DNA concludes: “As DNA profiling continues to grow more sensitive, and it is used in more investigations, the need for accurate communication between scientists and nonscientists only grows - both to ensure that their expectations of the technology are realistic, and its limits are properly understood” (Sense about Science 2017, p. 36).

Table 2.5. Levels in the hierarchy of propositions and questions that can be addressed when information is available (adapted from Gill et al. 2018 and Hicks et al. 2022). The questions addressed here focus on evaluative use of DNA results where a person of interest (POI) is considered rather than investigative use where information is being used to locate a potential POI (see Gill et al. 2018). Likelihood ratios (LRs) assigned by probabilistic genotyping software (PGS) systems primarily play a role at the sub-sub-source and sub-source levels.

Levels	Questions Addressed*	Results Used	Factors Considered
Sub-sub-source (DNA contributor level)	Is the POI the source of part of the mixture?	DNA profile comparison; LR value(s) assigned by a PGS system	Reliability of the DNA mixture interpretation protocol used in similar situations to the case in question; mixture factor space (e.g., potential presence of related people who are expected to share more alleles than unrelated contributors)
Discussed in Chapter 4 **Sub-source** (DNA level)	Is the POI the source of the DNA (single-source or mixture)?	DNA profile comparison; LR value(s) assigned by a PGS system with mixtures	(sub-sub-source factors when examining DNA mixtures) + occurrence of DNA profile genotypes in the relevant population; variability of results (e.g., presence or absence of peaks, peak heights, DNA quantities)
Source (cell level)	Is the POI the source of the body fluid?	DNA profile comparison; biological fluid presumptive tests	(Sub-source factors) + presumptive test false positive/ false negative rates (e.g., cross-reactivity)
Discussed in Chapter 5 **Activity**	Did the POI perform the given activity?	DNA profile comparison; biological fluid presumptive tests; relative quantity of DNA; where DNA was recovered; existence of multiple samples	(Source factors) + DNA transfer, persistence, prevalence, and recovery; DNA present for unknown reasons (i.e., background DNA)
Offense	Is the POI guilty or not guilty?	*The trier-of-fact, and never the scientist, addresses questions at this level using DNA findings plus additional information* (see Gill 2014, pp. 154-158)	

*Evaluations with likelihood ratios require examining the findings given specific pairs of propositions (e.g., see Hicks et al. 2022)

2.7. DNA Principles

This chapter concludes with a list of 16 important DNA principles. A principle is "a fundamental, primary, or general law or truth from which others are derived"[9]. An understanding of foundational principles can provide the basis for explaining a concept's importance and can assist in deciding what actions should be taken in specific situations. The principles and concepts described here, which are not necessarily exhaustive, have been distilled out of various publications and aspects of DNA mixture interpretation. They are grouped by theme and ordered arbitrarily. With each principle, which is numbered and displayed in bold font, additional information is provided concluding with a statement in italics that describes why that principle is important to DNA interpretation.

A shared understanding of fundamental principles described in this chapter benefits all stakeholders and helps users of DNA information appreciate the potential and the limitations of DNA mixture interpretation (see Schneider et al. 2006a, Morling et al. 2007, Stringer et al. 2009). Training and continuing education can assist in acquiring this understanding (see EWG 2024, pp. 241-274). These principles are not new but may need to be re-emphasized.

Principle 1 [Biology]: An individual's DNA generally remains unchanged across time and cell type.

Each cell of the human body contains DNA, except for mature red blood cells (Grasso & Woodard 1967). The DNA sequence and patterns found in the genome of an individual generally remain unchanged over time (Jeffreys 1987). Likewise, DNA samples originating from the same individual will yield, with very rare exceptions, the same DNA profile independent of the type of cells examined (e.g., sperm vs. epithelial) (e.g., Cotton et al. 2000, Holt et al. 2002). Thus, a sample from an individual collected at different times over their lifetime is expected to yield equivalent DNA profiles. *This principle enables meaningful comparison of DNA from a reference sample to an evidence sample deposited and/or collected at a different time and to verify identity in a "biometric" sense, where a previously analyzed DNA profile is checked against a new one for "authentication" purposes.*

Principle 2 [Biology]: DNA transfers and persists and can be collected and analyzed.

Human cells can be transferred to a surface through a variety of means, such as touching or coughing (van Oorschot & Jones 1997). DNA transfers and persists (e.g., van Oorschot et al. 2019) – and when collected and analyzed, can assist investigations. *This principle of direct or primary transfer enables results to be generated from evidentiary DNA profiles to assist in crime-to-crime and crime-to-individual associations.*

Principle 3 [Biology]: Forensic DNA profiles examine a limited number of specific sites in the human genome.

Current forensic DNA tests used in crime laboratories examine only a small portion of the human genome. A DNA profile comes from examining specific sites (*loci*) that are known to

[9] https://www.dictionary.com/browse/principle (accessed November 1, 2024)

vary between individuals and do not code for genetic traits (Katsanis & Wagner 2013). Short tandem repeat (STR) markers, which possess multiple (e.g., 10 to 20) alleles that vary in the number of repeats, are the primary loci used today in forensic DNA tests (Butler 2007). The ability to distinguish DNA profiles of two unrelated individuals increases as more loci are tested. *This principle is a reminder that the entire DNA sequence is not examined in forensic tests. Statistical assessments of profile rarity are used based on inheritance patterns and population genetics.*

Principle 4 [Genetics]: DNA passes from parent to offspring according to established genetic inheritance patterns.

Half of an individual's autosomal nuclear DNA comes from each of their biological parents. Each child can inherit different combinations of their parents' DNA (e.g., Roach et al. 2010). For this reason, the genetic characteristics shared among siblings can vary. Lineage markers, such those found on Y-chromosomes and mitochondrial DNA, typically pass from parent to offspring unchanged although an occasional mutation may occur (Kayser 2007). DNA profiles from biological relatives can be associated using the expected genetic inheritance patterns of various DNA markers. *This principle enables missing persons investigations, familial searching, relationship testing, and genetic genealogy.*

Principle 5 [Genetics]: Genetic inheritance patterns and population genetics enable strength of evidence statistical calculations.

A statistical strength of evidence can be calculated because of probabilities associated with genetic inheritance expectations. The statistical model for these population genetics calculations was described more than a century ago (Hardy 1908, Weinberg 1908) and is known as Hardy-Weinberg equilibrium (Crow 1999). The random match probability (RMP) is a measure of a DNA profile's rarity and reflects an estimate of the probability of drawing one individual with a specific DNA profile at random from a group of unrelated individuals in a population (NRC 1996). The rarity of a specific DNA profile can be calculated using allele frequency estimates for individual markers along with sub-population adjustments and combining genotype probability estimates across each marker deemed to be independent of other markers in the DNA profile (Balding & Nichols 1994). *This principle supports population frequency calculations made with RMP and LR approaches when a known is considered as a possible contributor to an evidence profile.*

Principle 6 [Genetics]: DNA profiles from close relatives are more similar than DNA from unrelated people.

DNA profiles from close relatives are expected to be more similar than DNA profiles from unrelated individuals (Li et al. 1993). There are a limited number of alleles at each locus, and even individuals who are not closely related will share alleles and genotypes. The frequency of occurrence of specific alleles varies. *This principle is a reminder that while statistical models typically assume individuals are unrelated, if case context suggests closely related individuals may have contributed to the sample in question, then performing calculations assuming individuals are related will be important to decision makers. Not accounting for*

relatedness can increase the risk of falsely including a non-contributor relative in the DNA mixture.

Principle 7 [Relevance]: Answers derived from DNA profiles depend on questions asked and circumstances of the evidence.

The FBI DNA Advisory Board stated: "Proper statistical inference requires careful formulation of the question to be answered. Inference must take into account how and what data were collected, which, in turn, determine how the data are analyzed and interpreted" (DAB 2000). DNA results typically address questions at the sub-source level of the hierarchy of propositions (i.e., who could be the source of the DNA or is the DNA from the person of interest, Taroni et al. 2013). *This principle is a reminder to users that DNA information by itself can only help answer "who" questions, that is, questions of source not activity.*

Principle 8 [Measurement]: Polymerase chain reaction (PCR) amplification is a process that enriches the initial DNA material into measurable amounts. However, when small amounts of DNA are amplified, the results may not exactly represent the original DNA sample, including the relative quantities of each allele and genotype. In addition, the PCR process with STR alleles introduces artifacts, such as stutter products, that complicate interpretation of the resulting DNA profile.

PCR relies on replicating specified areas of the available DNA template to generate a detectable DNA profile at multiple STR markers. This DNA profile, which is depicted as an EPG, is influenced by DNA template amount and degradation level, the presence of inhibitors, and the primer binding region sequence – all of which can influence the overall balance of the DNA profile. STR kits from different manufacturers may target slightly different regions of the same STR markers. PCR enables sensitive detection of even small amounts (e.g., 10 or fewer cells) of DNA, but also introduces artifacts such as stutter products into the test results that can influence the uncertainty of an interpretation (Gill et al. 2006b). *This principle is a reminder that STR results are a copy of the recovered DNA in a tested sample and depend on the accuracy and efficiency of the copying process. Sampling low amounts of DNA and PCR artifacts increases uncertainty for the genotype possibilities of contributors to complex DNA mixtures.*

Principle 9 [Measurement]: Peak positions are used to accurately designate alleles whereas peak heights are subject to stochastic effects and are variable.

Use of an internal size standard with each tested sample along with calibration to an allelic ladder enables accurate STR allele designations in electrophoresis separation and detection systems (e.g., Gill et al. 1997, Lazaruk et al. 1998). Peak heights and relative peaks heights, which do not use internal size standards to normalize stochastic variation, are not as reproducible as peak positions but do show trends by locus (e.g., Leclair et al. 2004, Debernardi et al. 2011). *This principle is a reminder that while alleles may be either present or absent (impacted by their peak heights and instrument detection thresholds), detected alleles are reproducible in terms of their designation (i.e., replicate testing does not show*

alleles shifting to a different allele, e.g., a "12" cannot become a "14" because peak position/sizing is stable).

Principle 10 [Measurement]: The variability of peak height ratios (and heterozygote imbalance) increases as peak height decreases.

Peak heights are a function of starting amount and quality of the DNA template. When sufficient quality and quantity of DNA template exist, reliable and unambiguous DNA profiles can be generated from crime scene evidence. However, PCR amplification of low amounts of DNA template result in stochastic variation including severe peak imbalance of paired alleles in a genotype, allele drop-out, high stutter, and allele drop-in (Butler & Hill 2010). The chance of failing to replicate alleles that are present in the original sample during the PCR process, referred to as the probability of drop-out, increases when attempting to copy small amounts of DNA or highly fragmented DNA. Replicate amplification from aliquots of the same DNA extract have been used to improve the degree of reliability (Taberlet et al. 1996, Gill et al. 2000, Benschop et al. 2011). More recent studies have explored advantages of amplifying the entire extract versus splitting it and interpreting replicates (Grisedale & van Daal 2012, Bille et al. 2022). *This principle relates particularly to minor contributors in DNA mixtures.*

Principle 11 [Interpretation]: Although there is a single physical mixture ratio created at the time of deposition, it may be manifested differently at each tested locus due to stochastic (i.e., random) variation in the PCR amplification process and potential variable DNA degradation across the contributors' genome sequences.

Stochastic variation in the PCR amplification process or sampling of template influences heterozygote balance and variation in mixture proportion (Bill et al. 2005). Assumptions are commonly made that allele peak heights are approximately linearly proportional to the amount of DNA prior to amplification and that contributions from two separate alleles are additive. Some studies have suggested that the estimated mixture proportion at each locus was highly variable at different loci within the same sample with variance at a locus from the overall profile estimate as high as 35% (Bill et al. 2005). *This principle emphasizes the need for interpretation methods or computer algorithms to account for variations in mixture ratios based on peak height variability and relative peak heights differences between loci in a DNA profile.*

Principle 12 [Interpretation]: When a peak is observed in a stutter position, interpretation should consider the possibility that the peak represents a minor contributor or a minor contributor plus stutter rather than a stutter product only.

STR allele stutter products can complicate DNA mixture interpretation particularly when estimating the number of contributors. Depending on the ratio of contributor amounts in the mixture, peaks in the stutter position may need to be considered as possible alleles from a minor contributor (Gill et al. 2006b, Budowle et al. 2009). Allele-specific stutter filters can aid mixture interpretation (Kalafut et al. 2018). *This principle recognizes the impact of artifacts, such as STR allele stutter products, on mixture interpretation.*

Principle 13 [Interpretation]: Accurate estimates of the number of contributors to a DNA mixture are impacted by and may be underestimated when (a) the number of contributors increases, (b) the amount of DNA tested decreases, or (c) the degree of allele overlap in mixture contributors increases, such as when the contributors are related.

Estimating the number of contributors in a DNA mixture becomes more uncertain when there are more contributors – and thus more potential alleles (Paoletti et al. 2005, Buckleton et al. 2007, Coble et al. 2015). The more alleles observed at a tested locus, the greater the chance for allele overlap among mixture contributors. As noted in Principle #6, biologically related contributors are expected to share alleles. When alleles overlap and are shared among contributors, it becomes more difficult to definitively estimate the number of contributors to the DNA mixture. Missing alleles from true contributors can also impact estimation of the number of contributors. Low-quantity and low-quality DNA templates are subject to allele drop-out as well as stochastic variation that can skew normal stutter product amounts and heterozygote balance (Butler & Hill 2010). *This principle emphasizes that factors impacting sample complexity, such as allele sharing and allele drop-out, influence reliable estimates of the number of contributors to a DNA mixture.*

Principle 14 [Interpretation]: Mathematical models in PGS systems can provide a list of possible genotype deconvolutions with associated probabilities for mixture components that cannot be physically separated. Continuous models use more information from a DNA profile than discrete or binary approaches.

A DNA mixture arises when cells from multiple contributors are present in a sample. Following the extraction process, DNA from these cells commingles and mixes – and this mixture cannot be chemically separated into its original components. Instead, mathematical models deconvolute EPG data or infer possible genotype combinations for detectable contributors. Then an assessment can be performed of the strength of evidence whether a person of interest contributed to a mixed DNA profile or not. The inclusion of peak height information in continuous models increases the strength of evidence for true donors especially for major contributors (Taylor 2014, Slooten 2018). *This principle recognizes that continuous models involving allele peak height information can discriminate better between true contributors and non-contributors than discrete or binary approaches only involving allele information.*

Principle 15 [Statistics]: Different statistical approaches can produce different numerical results as they utilize different information and/or models and may answer different questions.

Multiple statistical approaches have been used for DNA mixture interpretation. Questions addressed and information used by these approaches can differ (see Tables 2.2 and 2.3). For example, different LR approaches can yield different results because these approaches may utilize different information (e.g., modeling different types of stutter products) or process the same information differently (e.g., using a log normal model versus a gamma model). Thus, the 2018 ISFG DNA Commission concludes: "*There are no true likelihood ratios, just like there are no true models.* Depending on our assumptions, our knowledge and the results we want to assess, different models will be adopted, hence different values for the

LR will be obtained. It is therefore important to outline in our [reporting] statements what factors impact evaluation (propositions, information, assumptions, data, and choice of model)" (Gill et al. 2018, emphasis added). *This principle recognizes that answers obtained are dependent on information and statistical models utilized and questions asked (see also Principle #7).*

Principle 16 [Statistics]: Assessing the strength of evidence in support of a proposition (hypothesis) H_1 requires at least one other proposition (hypothesis) H_2. These propositions H_1 and H_2 are required to be mutually exclusive and reasonably exhaustive within the context of the case. Strength-of-evidence assessments depend on the framework of circumstances within which they are evaluated.

The three principles of evidence interpretation that were described in the 1998 book by Ian Evett and Bruce Weir (Evett & Weir 1998, pp. 23-29) and restated in the 2020 book by Jo-Anne Bright and Michael Coble (Bright & Coble 2020, pp. 23-24) are included here as the final DNA principle (Principle 16). *Evidence Interpretation Principle 1*: To evaluate the uncertainty of any given proposition, it is necessary to consider at least one alternative proposition. *Evidence Interpretation Principle 2*: Scientific interpretation is based on questions of the kind: "What is the probability of the evidence given the proposition?" *Evidence Interpretation Principle 3*: Scientific interpretation is conditioned not only by the competing propositions, but also by case context within which they are to be evaluated (Hicks et al. 2021). The hierarchy of propositions describes the types of questions that can be evaluated – offense, activity, source, sub-source, and sub-sub-source levels (Cook et al. 1998b, ENFSI 2015, Taylor et al. 2018, Gill et al. 2018, Gill et al. 2020a). *These three principles emphasize the foundational elements of the likelihood ratio framework.*

KEY TAKEAWAY #2.7: A probabilistic genotyping framework offers the best available tool for DNA mixture interpretation at the sub-source or sub-sub-source levels within the hierarchy of propositions. However, its fitness for purpose in any specific casework application must be supported by validation data from known samples that are similar in complexity to those seen in casework. Continuous PGS systems have many advantages if they are used following suitable training, and if the decisions informing the LR assignment(s) are clearly stated.

3. Chapter 3: Data and Information Sources

This scientific foundation review seeks to document and independently assess the empirical evidence that supports the reliable use of DNA mixture interpretation methods. The sources of data and information used in conducting this review are described in this chapter. These sources include (1) peer-reviewed articles appearing in scientific journals, (2) published interlaboratory studies, (3) laboratory internal validation studies that are accessible online, and (4) proficiency test data available on test-provider websites. A supplemental document (NISTIR 8351sup2) contains summarized information from publicly accessible validation data, interlaboratory studies, and proficiency test results covering DNA mixture interpretation.

3.1. Information Sources

This scientific foundation review focused on DNA mixture interpretation involving autosomal short tandem repeat (STR) markers. To assess reliability and relevance issues related to DNA mixture interpretation, empirical data and information were sought from a variety of publicly available sources.

Resources were examined that were available in the public forum during the time frame of this study which includes the public comment period. These included (1) publications in the peer-reviewed scientific literature and relevant books and (2) data or information located on the internet, such as proficiency test (PT) results from PT provider websites or publicly available internal validation data summaries from individual laboratories. PT data provide insights into how individual analysts performed on specific tests while internal validation studies offer insights into how laboratories performed when analyzing a range of DNA mixtures of varying complexity. Published interlaboratory studies enable an important assessment of analyst and laboratory performance. This is because the same samples and/or data are evaluated among the participants to examine reproducibility across methods.

By searching and studying the peer-reviewed literature on forensic DNA, the authors of this report collected and examined articles on DNA mixture interpretation and DNA transfer studies.

It is recognized that there are information and data collected in forensic laboratories that may not yet be publicly available or published. However, as stated previously (NISTIR 8225), the authors of this report believe for information to be considered foundational, it needs to be reasonably accessible to anyone who wishes to review it.

3.1.1. Peer-Reviewed Publications

Multiple literature searches were performed seeking articles related to DNA mixture interpretation using PubMed, Google Scholar, and Web of Science. Knowledge distilled from the examination of these articles informed the entire report.

As part of this review, the authors of this report examined titles and abstracts for articles published since 2009 in the following journals: *Journal of Forensic Sciences*, *Forensic Science International*, *Forensic Science International: Genetics*, *Science & Justice*, *Legal Medicine*, *Australian Journal of Forensic Sciences, Electrophoresis*, *International Journal of Legal Medicine*, and *Forensic Science Medicine and Pathology*. In addition, over 1500 extended abstracts published in the 2009, 2011, 2013, 2015, 2017, and 2019 *Forensic Science International: Genetics Supplement Series,* representing the proceedings of the biennial meetings of the International Society for Forensic Genetics, were considered.

Search parameters impact the number and types of articles that can be located on any particular topic. The challenge of locating applicable articles is illustrated in Table 3.1, which contains a summary of PubMed searches for articles containing the words “DNA” and “mixture” in the text. Table 3.1 is used as an example only. Many additional searches were performed during the course of this study to locate appropriate articles for citation purposes.

Table 3.1 illustrates a steady stream of new literature and is a reminder that information gathered to compile this report on DNA mixture interpretation represents a snapshot in time. In addition, these PubMed search results are missing some relevant publications (e.g., ones cited in this report's reference list) within these journals or in other journals not listed. Many of the search results provided articles that have "DNA" and "mixture" within the text but are not applicable to DNA mixture interpretation involving autosomal STR markers. This is the case with many of the *PLoS ONE* articles and other journals as well. For example, an examination of the three articles from 2009 published in *Forensic Science International: Genetics* finds only one that falls within the scope of this review (Cowell 2009), as the other two describe Y-chromosome STR analysis or tri-allelic single nucleotide markers (SNP) markers.

Table 3.1. Numbers of articles published with "DNA" and "mixture" in the text across the listed forensic science journals from 2009 to 2018 based on PubMed searches (https://www.ncbi.nlm.nih.gov/pubmed/) conducted May 10, 2019.

Journal	Total	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
PLoS ONE	**187**	7	7	15	26	40	30	20	11	13	18
Forensic Sci. Int. Genet.	**135**	3	3	7	12	4	22	16	15	26	27
Int. J. Legal Med.	**30**	2	2	0	3	3	2	2	4	5	7
J. Forensic Sci.	**27**	4	3	6	3	2	1	4	1	1	2
Electrophoresis	**25**	5	3	1	1	1	2	4	3	2	3
Sci Justice	**11**	1	1	2	2	2	0	0	0	1	2
Legal Med.	**9**	0	1	2	1	1	0	0	2	1	1
Forensic Sci. Int.	**4**	0	0	1	0	0	0	0	1	1	1
For. Sci. Med. Pathol.	**2**	0	1	0	0	0	0	0	0	1	0
TOTAL	**430**	**22**	**21**	**34**	**48**	**53**	**57**	**46**	**37**	**51**	**61**

By examining online search results, publications dealing specifically with DNA mixtures and aspects of DNA interpretation were identified. Each located article was first assessed by reviewing the title and abstract. Articles of interest were downloaded and studied further. Citation lists were inspected in the articles examined to see whether a relevant article may have been missed in initial searches. Information used in Chapter 5 regarding DNA transfer studies was located with similar types of search strategies. Hundreds of relevant articles were collected and are cited throughout this report. Every attempt has been made to develop a comprehensive, curated bibliography on DNA mixtures in the timeframe of this study. While many additional articles related to DNA mixture interpretation were published between the time that the draft report was released in June 2021 and this final version was completed (e.g., see Butler 2023), only a few of these have been cited in this final report.

3.1.2. Available Internal Laboratory Data

Forensic laboratories conduct internal validation experiments before implementing a new technique to assess method performance under specific conditions. Data from these studies are not typically shared outside the laboratory except in response to a discovery request connected to a specific legal proceeding. Auditors as part of an accreditation review examine validation

studies[10] and look for the types of experiments conducted as part of their approval process (QAS 2020 Audit). With an understandable focus on casework production in forensic laboratories, information from internal validation studies or related research experiments may not be prepared in a manner conducive to sharing with a wider community. Even if prepared, manuscripts reporting internal validation analysis are unlikely to be considered for publication in a peer-reviewed journal unless they provide a new insight that has not been previously reported. Google searches for data from internal validation studies were performed by searching for the state, city, and agency (if known) and the phrase "forensic DNA laboratory validation data." Forensic laboratories' public websites were also reviewed for available standard operating procedures (SOPs) and/or validation documents. Eight laboratory probabilistic genotyping software (PGS) internal validation summaries were located on John Buckleton's STRmix website.

Internal validation summaries from 16 U.S. forensic laboratories were located with online searches or made available as part of the public comment period (see NISTIR 8351sup2). Generally speaking, the authors of this report found that sufficient data of this sort are not publicly available for an independent assessment of reliability (see Chapter 4). Some laboratories provide summary information from their validation studies, but detailed data are often unavailable, in part because of privacy concerns around releasing genotype information from individuals. The same lack of detailed data is true for most peer-reviewed articles that describe validation experiments.

Information included in these 16 laboratories' summaries is related to the PGS system being validated and the types of DNA mixture samples being used. However, it is recognized that additional internal validation data exists within many individual laboratories, and these validation studies are reviewed by independent external auditors for FBI QAS and ISO 17025 accreditation requirements. The authors of this report did not examine data that would otherwise be inaccessible to others (e.g., obtained under a non-disclosure agreement to only be seen by the NIST reviewers). Information used was limited to publicly accessible information described in the accompanying supplemental document (NISTIR 8351sup2). Given that two public laboratories provided their STRmix validation summaries during the public comment period (see PC12 and PC63 in 8351-draft PCs), there was a recognized opportunity to share information by those who wished to do so after the draft report was released.

3.1.3. Available Proficiency Test Data

Proficiency test (PT) data has been thought to be useful when assessing the reliability of DNA mixture interpretation methods. The DNA Identification Act of 1994 and the FBI Quality Assurance Standards require successful completion of semiannual proficiency tests for all DNA analysts working in a U.S. laboratory that receives federal funding or supplies data to the national DNA database (DNA Identification Act 1994, QAS 2020). Over the years, a variety of DNA mixture tests have been provided to participating forensic DNA analysts. In the United States, PT providers offering DNA mixture tests include Collaborative Testing Services (Sterling,

[10] Review of validation studies by external accreditation assessors is typically only conducted once following the completion of a study (see QAS 2020, Standard 15.2.2).

VA), Bode Technology (Lorton, VA), and Forensic Assurance (Northville, MI). In addition, the German DNA Profiling Group (GEDNAP) provides DNA proficiency tests for many laboratories in Europe. PT provider websites were searched for available information.

3.1.3.1. CTS Forensics

Collaborative Testing Services, Inc. (CTS) offers several DNA mixture proficiency tests. Participants are rated on their ability to return results that agree with a consensus result.

CTS has reported about 80% return rates for their DNA PT exams from 2004 to 2015 (Kolowski et al. 2016). Currently, CTS offers a DNA mixture test twice a year (5801 and 5806 series) and a DNA interpretation test (588 and 589 series) twice a year. The CTS forensic biology tests (until 2017 the 571, 572, 573, 574, 575, and 576 series and since 2017, the 5701, 5702, 5703, 5704, 5705, and 5706 series) also contain mixtures of human whole blood and semen. In 2022, CTS also began offering a probabilistic genotyping PT, and several data sets were available for inspection before this report was finalized (see NISTIR 8351sup2, Tables S2.7 to S2.11).

The DNA mixture test samples contain two known bloodstains provided on Whatman FTA cards or clean white fabric, and two questioned stains where one or both contains a mixture of body fluids. This is typically blood and semen mixed in a 1:1 volume ratio before the mixture is applied to the substrate (see NISTIR 8351sup2, Table S2.5).

The CTS DNA interpretation tests are intended for the technical reviewers and consultants who may not have access to laboratory equipment or data analysis software. These tests are distributed via digital download in the form of electropherogram files (.pdf, .fsa, or .hid formats) with results from a variety of common autosomal and Y-STR typing kits. Participants with the DNA interpretation study evaluate and report DNA profiles of four samples, consisting of two known and two question samples, using their existing protocols. Mixtures present in question samples are usually two-person and sometimes three-person mixtures with components in the range of 1:1 to 1:4 or 2:1:1 or 3:1:1 (mixed by body fluid volume rather than predetermined DNA quantity).

A supplemental document to this report contains a summary of CTS DNA mixture data sets along with analysis of their contents (NISTIR 8351sup2).

3.1.3.2. Bode Technology

Bode Technology, formerly known as Bode Cellmark Forensics, offers International Quality Assessment Scheme (IQAS) PT kits. Two kits (IQAS-50 and IQAS-60) provide the ability to assess DNA mixture interpretation results from a simple mixture of semen and white blood cells. Summary reports of participant results are provided to the ANSI-ASQ National Accreditation Board (ANAB).

We did not find these PT results or reports to be publicly available for examination or review.

3.1.3.3. Forensic Assurance

In an effort to provide PT samples that are more like casework situations, Forensic Assurance has begun offering a PGS proficiency test. They supply data files for two evidentiary mixture samples (two-, three-, or four-person mixtures) and four known reference samples. Participants are required to estimate the number of contributors in the mixture profiles and compare the reference profiles to the mixture profiles using their laboratory's PGS and interpretation protocols. Participants return their likelihood ratio (LR) for each comparison along with the propositions used and a determination of which proposition is favored (i.e., H_1 versus H_2 or the numerator versus the denominator in their LR calculation).

We did not find these PT results or reports to be publicly available for examination or review.

3.1.3.4. GEDNAP Studies

The German DNA Profiling Group (GEDNAP) provides regular DNA PT exams for quality-assurance purposes (Rand et al. 2002, Rand et al. 2004). A GEDNAP "Stain Commission" designs the studies, which commonly contain challenging samples and mixtures. Each GEDNAP PT consists of three reference samples and four "stains" designed to mimic crime scene samples. Samples are prepared and sent out twice a year from a DNA laboratory in Münster, Germany. Each February, an annual Stain Workshop meeting is held ("Spurenworkshop" in the German language) to review the overall results obtained in the two studies from the prior year.

Typical errors are examined in an anonymous fashion to encourage quality improvements. Successful laboratories receive proficiency certificates. Over 200 laboratories from more than 40 different countries regularly participate in the GEDNAP PT DNA studies. Correct results are shared with each participating laboratory along with their score and a summary of any errors made.

We did not find these PT results or reports to be publicly available for examination or review.

3.1.4. Interlaboratory Studies on DNA Mixture Interpretation

Interlaboratory studies provide an opportunity to assess variations across laboratory protocols and can be useful barometers regarding the reproducibility of various approaches.

Twenty interlaboratory studies examining various aspects of DNA mixture interpretation and performance (see NISTIR 8351sup2, Table S2.12) have been conducted over the past 25 years. These studies have been conducted by researchers at the National Institute of Standards and Technology, the Spanish-Portuguese Working Group of the International Society for Forensic Genetics, the European Forensic Genetics Network of Excellence, the UK Forensic Science Regulator, the Defense Forensic Science Center, the Netherlands Forensic Institute, and developers of the STRmix PGS system. Most of these studies have been published (see citations in NISTIR 8351sup2). Some of the older studies are not terribly relevant to present-day work.

3.1.5. Available Research Data Sets

Research data sets have been produced to aid current and future DNA mixture studies. The largest and most widely used to date is the PROVEDIt (Project Research Openness for Validation with Empirical Data) data set maintained by Professor Catherine Grgicak at Rutgers University (and previously at Boston University), which contains almost 25,000 DNA profiles (Alfonse et al. 2018). Table 3.2 summarizes the PROVEDIt data set, which contains DNA profiles amplified with three STR kits (Identifiler Plus, PowerPlex 16HS, and GlobalFiler) and analyzed on two capillary electrophoresis (CE) platforms (ABI 3130 and ABI 3500). These data were generated under 144 laboratory conditions and are classified by total DNA amount, DNA treatment, contributor numbers, and mixture proportions.

Table 3.2. Summary of PROVEDIt data set collected by researchers at Boston University and Rutgers University. Available at https://lftdi.camden.rutgers.edu/provedit/files/ (accessed November 1, 2024).

Sample Preparation	Data Set	STR Kit	# PCR Cycles	CE	# Profiles	Single-Source	2p Mixture	3p Mixture	4p Mixture	5p Mixture
DNA extract mixtures	RD12	Identifiler Plus	29	ABI 3500	3212	2280	366	209	147	210
DNA extract mixtures	RD12	PowerPlex 16HS	32	ABI 3130	1024	795	57	52	60	60
Whole blood mixtures	RD14	Identifiler Plus	28	ABI 3130	10,261	8267	524	487	520	463
Whole blood mixtures	RD14	GlobalFiler	29	ABI 3500	10,195	8190	526	484	527	468
				TOTAL	24,692	19,532	1473	1232	1254	1201
				ABI 3500	13,407					
				ABI 3130	11,285					

The PROVEDIt data can be downloaded as raw data (.fsa and .hid files) or exported genotypes table (.csv files) from the Laboratory for Forensic Technology Development and Integration (LFTDI; https://lftdi.camden.rutgers.edu/provedit/files/). Among the 5160 mixture profiles, ranging from two-person (2p) up to five-person (5p) profiles, 76% contain a contribution of at least one individual of less than 20% of the total DNA content. Many of the samples, which were prepared with 37 different genotype combinations, were subjected to PCR inhibitors or purposely degraded to produce partial profiles (Alfonse et al. 2018).

The funding to generate this data set represents a substantial and important investment by the U.S. government over multiple years. In their article describing the PROVEDIt data set, the authors express their hope that “a large dataset would play a critical role in demonstrating the foundational validity and robustness of new or existing DNA identity testing technology” (Alfonse et al. 2018). Samples from the PROVEDIt data set have been used in PGS comparisons (e.g., Riman et al. 2019b, Cheng et al. 2021) and interlaboratory studies (e.g., Bright et al. 2019).

4. Chapter 4: DNA Interpretation at the Sub-Source Level

This chapter considers foundational issues related to the reliability of DNA mixture measurement and interpretation at the sub-source (DNA profile) or sub-sub-source (DNA component) level of the hierarchy of propositions (see Table 2.5). This interpretation is often performed using probabilistic genotyping software (PGS) systems to assign likelihood ratios (LRs) to reflect the strength of evidence based on competing propositions (H_1 and H_2). Reliability centers on trustworthiness of a system or its components. Validation data, interlaboratory studies, and proficiency tests help in the assessment of the degree of reliability of a system. To enable effective use of any information, responsibilities exist with both providers of information and stakeholders using it. The term "factor space" is used to describe the factors that may influence complexity, measurement, and interpretation variability – these factors include the number of contributors, the degree of allele sharing, the ratios of mixture components, and the amount and quality of the DNA tested. Available data from publicly accessible validation studies, proficiency tests, and interlaboratory studies are summarized in a supplemental document (NISTIR 8351sup2) where limitations of available information and aspects of factor space coverage are considered. This information includes data from 72 published articles and 20 internal validation summaries involving PGS systems, more than 10 years of proficiency test results involving more than 150,000 comparisons with DNA mixtures, and 20 interlaboratory studies over the past 25 years. The degree of reliability of a DNA mixture interpretation system, such as a DNA analyst using a probabilistic genotyping software program, depends on sample complexity. Results cannot be simply summarized into "reliable" or "unreliable" without considering the context of the factor space explored and supporting validation data using ground truth samples of similar complexity. This report emphasizes that proficiency tests need to continue to evolve and be representative of complex DNA mixtures seen in casework if these tests are intended to assess analysts' ability to conduct dependable DNA mixture interpretation. Finally, transparency of validation information is encouraged in the future to improve data sharing and support independent assessments.

4.1. Introduction

As discussed in Section 2.5, DNA interpretation at the sub-source level is increasingly performed using likelihood ratios (LRs) assigned by probabilistic genotyping software (PGS) systems[11] with various inputs (see Figure 2.3). A probabilistic genotyping framework offers the best available tool for DNA mixture interpretation at the sub-source or sub-sub-source levels within the hierarchy of propositions (see Table 2.5), but its fitness for purpose in any specific casework application must be supported by validation data from known samples that are similar in complexity to those seen in casework.

This chapter examines issues around assessing the reliability of assigned LR values. For example, given the growing use of PGS systems, there have been calls for software code review (e.g., Adams et al. 2018, Abebe et al. 2022). DNA interpretation is more than simply using a computer program to obtain an LR. The introduction of software into the DNA interpretation process has blurred the lines and added complexity in the way some DNA analysts see their role versus the PGS system they utilize. In an article titled "'Is your accuser me, or is it the software?' Ambiguity and contested expertise in probabilistic DNA profiling," researchers from Columbia University studying this issue found:

> "The algorithm gets constructed alternately either as merely a tool or as indispensable statistical backing; the analysts' authority as either independent of the algorithm or reliant upon it to resolve conflict and create a final decision; and forensic expertise as resting either with the analysts or with the software" (Pullen-Blasnik et al. 2024).

The current process is first discussed for performing and verifying validation studies involving DNA mixture interpretation for sub-source level determinations and then aspects of reliability and the need for empirical data in assessing reliability are examined.

4.1.1. Current Process for Performing and Verifying Validation Studies

In the process of implementing a PGS tool to assist with DNA mixture interpretation at the sub-source level, an individual laboratory performs some internal validation studies according to a set of requirements such as SWGDAM guidelines that outline the types of experiments that should be conducted (e.g., SWGDAM 2015). These experiments are documented in a validation summary that exists in the laboratory but is usually not made publicly accessible. Publicly accessible validation studies that were identified with this foundation study have been described in a supplementary document (Tables S2.3 and S2.4 in NISTIR 8351sup2).

There are no standardized procedures for conducting validation experiments or documenting the results so the details included and the amount of information described can vary among laboratories (see Table S2.4 in NISTIR 8351sup2). It is expected that a laboratory's internal validation summary and perhaps even the underlying data have been assessed by auditors for compliance to FBI Quality Assurance Standards requirements (see QAS Audit 2020, p. 25). An audit of a validation study under the FBI QAS requirements involves a "yes", "no", or "N/A (not applicable)" response to a series of questions, such as "have internal validation studies

[11] For a history of DNA mixture interpretation and PGS developments, see the accompanying supplemental document (NISTIR 8351sup1).

included, as applicable: precision and accuracy studies? sensitivity and stochastic studies? mixture studies?..." However, in an audit, there is no mechanism to assess performance reliability, only whether or not a specified type of study has been conducted and documented in records retained by the laboratory and made available to the auditor.

DNA mixture samples are interpreted using the assistance of a PGS system. Inputs to this PGS system include EPG peaks above the laboratory's analytical threshold (with their respective quantitative information including peak heights and sizes), an assigned NoC, specific propositions based on the NoC and case-specific details, STR allele frequencies from relevant populations, and population substructure adjustments (see Figure 2.3). Results reported that may include one or more assigned LRs along with verbal equivalents (e.g., strong support, moderate support) to assist users of that information.

According to the FBI QAS, protocols and threshold conditions that are established in a forensic laboratory (e.g., an analytical threshold of 50 RFU) "should be supported by the internal validations" (QAS 2020, Standard 9.1). Casework is conducted using these protocols and thresholds.

Casework results are typically trusted because they are analyzed under a "validated" method or protocol, which means (1) that experiments were conducted and analyzed by laboratory personnel (or the analysis might be outsourced to commercial providers), and (2) that a decision maker (in each U.S. forensic laboratory, this is the DNA Technical Leader) has declared testing was conducted according to a set of requirements (e.g., QAS 2020 and/or ISO/IEC 17025) informed by guidelines (e.g., SWGDAM 2015) or standards (e.g., ASB 018) and that the method is declared "fit-for-purpose" in their laboratory. Limitations can sometimes be placed on a DNA interpretation method in a forensic laboratory, such as use with up to four-person mixtures or down to 100 pg total DNA quantity, because of a decision maker's experience with their own internal validation experiments and/or interpretation of prior published studies regarding the ability to deconvolute mixture components.

4.1.2. Reliability

The "plain English" meaning of the word *reliability* is trustworthiness, which is determined by the degree to which a result is consistently accurate.[12] This is the sense in which the term reliability is used in this report. For a method to be reliable, it must produce results that are consistently accurate within a specified (and acceptable) uncertainty. An accurate result implies that a true value is known (ground truth) for the thing being measured. The word reliable is sometimes treated as though it has a binary meaning (i.e., something is reliable or not reliable). However, from a scientific perspective, it is more appropriate to speak in terms of a degree of reliability, reflecting the magnitude by which a result deviates from being accurate.[13]

[12] Oxford Dictionary (https://www.lexico.com/en/definition/reliability): a) The quality of being trustworthy or of performing consistently well. b) The degree to which the result of a measurement, calculation, or specification can be depended on to be accurate.
[13] We recognize that in legal settings, binary decisions (e.g., guilty or not guilty) need to be made. However, our focus is on the nonbinary scientific aspects of reliability rather than the binary legal ones.

An important hallmark of science is to develop reliable theories and methods *based on empirical data,* so that users of scientific knowledge or methods can have a high degree of trust in their claims, results, or predictions. Reliability is born out of demonstrations of accuracy, which is what allows for trust in both the initial assumptions as well as in the resulting claims. In this chapter, the basis for reliability in DNA mixture measurements and interpretation at the sub-source (DNA profile) level of the hierarchy of propositions (see Table 2.5) is explored.

It is generally accepted that measurement and interpretation of high-template, high-quality, single-source DNA samples have a high degree of reliability (NRC 2009, PCAST 2016). This reliability comes from testing and observing consistently accurate results when assigning allele pairs into genotypes. At the other extreme, measurement and interpretation of samples involving many contributors, and/or low DNA template amounts, present challenges in assigning allele pairs for specific contributors (e.g., Benschop et al. 2012, Benschop et al. 2015, Taylor & Buckleton 2015). This is likely the reason some laboratories adopt a policy of not interpreting highly complex mixtures (e.g., more than three contributors or not reporting an assigned LR below 1,000,000, see Hahn et al. 2024).

Using current laboratory methods, it is impossible to physically separate the DNA within a complex mixture into its constituent parts although research is progressing in this area (see Section 6.3). To interpret a DNA mixture, an analyst uses their best judgment or an algorithm[14] to estimate the number of contributors, which is a required input for most PGS systems, based on the observed DNA profile and then proceeds as described in Chapter 2 (see Figure 2.2).

The outcome of DNA mixture interpretation often includes a numeric output in the form of a likelihood ratio (LR). The numerical value could be expressed in the form of a specific LR (e.g., 2.3×10^{20}) or can be described in a log LR format (e.g., 20.4). Verbal categorical descriptions are sometimes used with LRs (SWGDAM 2018). As discussed in Chapter 2, an LR value, which is a ratio of two probabilities, cannot be compared to any reference standard to assess accuracy.

A 2016 virtual special issue of the journal *Science & Justice* discussed measuring and reporting the precision of forensic likelihood ratios (Morrison 2016). Some scientists hold the position that there is no true LR (e.g., Steele & Balding 2014, Gill et al. 2018), and some hold the position that there is no uncertainty associated with an LR assignment (Berger & Slooten 2016; see also Biedermann et al. 2016a, Taroni et al. 2016, Curran 2016, Morrison & Enzinger 2016, Sjerps et al. 2016, Taylor & Balding 2020).

DNA analysts have increasingly relied on one of several available PGS systems to deconvolute mixtures and to assign a strength of evidence numerical value to based on a pair of propositions selected by the analyst (see Chapter 2 and NISTIR 8351sup1). Interpretation of DNA profiles involves DNA analysts applying their judgment, training, and experience along with laboratory protocols, practices, and policies as they use information provided by a PGS system. Some PGS are proprietary, and others are open-source (see Table 2 in Butler & Willis 2020).

[14] Algorithms also incorporate judgments of the software developers when implementing the mathematical or statistical models. An important benefit of an algorithm is consistency, but its use does not guarantee accuracy.

4.1.3. Empirical Assessments of Reliability

Empirical assessments of reliability require that the process of interest be tested in ground-truth[15] known situations. For DNA mixture interpretation, this means that samples with known genotypes, a known number of contributors, known mixture ratios, known degrees of degradation, etc., have been tested (i.e., measured and interpreted), and results from such tests are publicly accessible to provide the basis for stakeholders to assess the degree of reliability of the process. The degree of reliability can be considered from developmental and internal validation experiments (method-focused), proficiency tests (analyst-focused), and interlaboratory studies (community-focused). Each type of assessment uses empirical data and addresses different questions.

Systematic approaches for analyzing the results of validation experiments and relevant concepts such as accuracy, bias, precision, and calibration are discussed in various textbooks (e.g., Vosk & Emery 2014, Zadora et al. 2014). Common statistical methods to summarize and visualize the data often include the average values and standard deviations of the results. At the other extreme, when the quantity of interest is binary (e.g., whether a proposition is true or false), differences from the expected value are summarized using error rates, which involve calculating a percentage of the times *true* is incorrectly classified as *false* (false negative errors) or *false* is incorrectly classified as *true* (false positive errors) (see also Swofford et al. 2024).

In their September 2016 report, the President's Council of Advisors on Science and Technology (PCAST) associated reliability with test results that have been demonstrated to be repeatable, reproducible, and accurate (PCAST 2016, p. 47). PCAST used the phrase "foundational validity" to reflect whether something was based on reliable principles and methods and "validity as applied" to reflect whether the individual performing the work was applying these principles and methods reliably in any particular case (PCAST 2016, pp. 42-66). The 2016 PCAST report emphasized that "the *only* way to establish scientifically that an examiner is capable of applying a foundationally valid method is through appropriate empirical testing to measure how often the examiner gets the correct answer" (PCAST 2016, p. 57, emphasis in the original). This point was reiterated in the January 2017 *An Addendum to the PCAST Report on Forensic Science in Criminal Courts*: "Forensic scientists rightly cite examiners' experience and judgment as important elements in their disciplines…However, experience and judgment alone – no matter how great – can *never* establish the validity or degree of reliability of any particular method. Only empirical testing of the method can do so" (PCAST 2017, p. 3, emphasis in the original).

KEY TAKEAWAY #4.1: To assess the degree of reliability of a component or a system for any forensic method or practice, detailed empirical data are needed, such as data from validation experiments, interlaboratory studies, and proficiency tests.

[15] Ground-truth requires knowing the correct answer before testing is performed and therefore is not possible with samples arising from crime-scene evidence.

4.1.4. Assessing Reliability of Likelihood Ratio Values from PGS Systems

The process of interpreting DNA mixtures is guided by principles of the underlying biology as well as statistical representations of the empirically observed relationship between genotypes and EPGs, all of which may be combined and codified in the form of models in PGS systems. These models assign probabilities (Figure 4.1). Their fitness for any given purpose is informed by results of validation studies involving experiments with ground-truth known data and covering the space of anticipated casework scenarios.

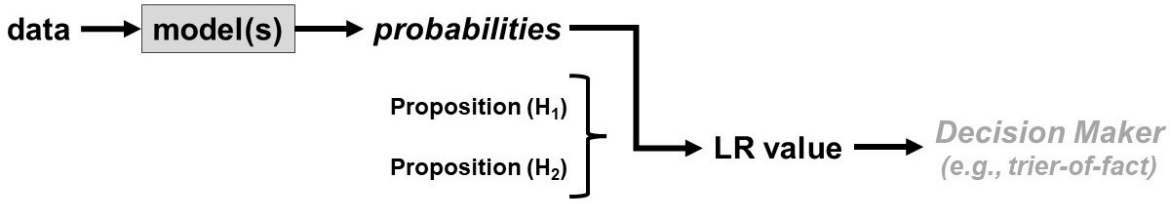

Figure 4.1. Relationship between data, models, and probabilities used to assign LR values given specific casework propositions (H_1 and H_2).

PGS systems used to perform DNA mixture interpretation at the sub-source level have multiple inputs that influence the LRs assigned, such as propositions used based on case-specific circumstances, the number of contributors, the STR allele frequencies for appropriate population group(s), and the input data from an evidentiary DNA STR profile EPG based on laboratory-established analytical thresholds (see Figure 2.4). The probabilities used in DNA mixture interpretation with PGS to assign a LR, such as weights for possible genotype combinations of contributors, result from biological models, computer algorithms, and statistical models. The resulting LR value assigned by a PGS system is dependent on the inputs provided.

Writing in 2018, the ISFG DNA Commission stated:

> "… there are no true likelihood ratios, just like there are no true models… Depending on our assumptions, our knowledge and the results we want to assess, different models will be adopted, hence *different values for the LR will be obtained*. It is therefore important to outline in our statements what factors impact evaluation (propositions, information, assumptions, data, and choice of model)" (Gill et al. 2018, emphasis added).

As discussed in Chapter 2, different experts using different assumptions, different statistical models, and different inference procedures often arrive at different LR values. Information regarding the extent to which their LR values agree or disagree is typically not available. There have been some comparisons conducted between various PGS systems (e.g., see Table 4.2 later in this chapter). There appears to be a general misconception that LR assessments made by different experts will be close enough to one another to not impact the final DNA mixture interpretation conclusions. Although they may be similar in many instances, this is not known for any particular case, and it is not advisable to take this for granted (e.g., Thompson 2023).

In addition, there are a number of different LR values that can be generated by a PGS system, such as a highest posterior density (HPD) LR to adjust for sampling uncertainty, a unified LR to account for both related and unrelated individuals under the alternative (H_2) proposition, a

population stratified LR to incorporate relative proportions of different subpopulations, a variable number of contributors (varNOC) LR estimation, or various combinations of these LR adjustments (Kelly et al. 2020). Appreciating the assumptions and information provided by each of these numbers is important to communicate what a specific LR reflects given sub-source propositions (see Table 2.4). Likewise, when considering other questions in the hierarchy of propositions (see Table 2.5), the underlying information supporting the LR assignments should be communicated to users of the findings that are reported.

The degree of reliability or trustworthiness of a given PGS system in a given case is dependent upon the number of instances in which that system has been tested with samples that are judged to be of similar complexity as the casework sample, the performance of the method among those instances, and how the characteristics (e.g., number of contributors, DNA amounts, level of degradation) of the ground-truth known samples compared to those of the sample in the case at hand. More validation samples and denser coverage of the space of application scenarios provide better estimates of casework-relevant reliability metrics.

The desired performance for a DNA mixture interpretation PGS model is often described in terms of trends. For example, authors of the STRmix developmental validation study wrote:

> "the log(LR) for known contributors (Hp true) should be high and should *trend* to 0 as less information is present within the profile. Information includes the amount of DNA from the contributor of interest, conditioning profiles (for example, the victim's profile on intimate samples), PCR replicates, and decreasing number of contributors… The LR should *trend* upwards to neutral [for known non-contributors] as less information is present within the profile" (Bright et al. 2016, emphasis added).

Beyond producing LRs that follow expected trends (which are an important starting point), it is valuable to also consider other questions. Since repeatability and reproducibility are components of reliability, it is fair to ask to what extent the LRs offered by different experts using different databases and different models differ from one another. Understanding what level of reproducibility there is among laboratories or among forensic scientists will help assess reliability.

Whereas each laboratory or expert may consider their assessments to be reliable, the recipients of such assessments in a given case need guidance on what to do in situations where variation among different LR assessments could potentially impact the outcome of a trial (e.g., Thompson 2023). In particular, because there are no standards to compare to and no traceability considerations as there are for conventional measurements, judgments of reliability by decision makers or triers of fact will be helped by comparing LRs from multiple systems and made by multiple experts (Gill et al. 2015). While individual forensic laboratories may not operate multiple PGS systems, some published studies have shared variation observed with different systems on the same samples as will be discussed later in this chapter (e.g., Greenspoon et al. 2024).

Likelihood ratios must satisfy an internal consistency requirement (called the property of being well-calibrated or "calibration accuracy," for short), which can be empirically tested (Ramos & Gonzalez-Rodriguez 2013, Meuwly et al. 2017, Hannig et al. 2019). The scientific validity of any

particular PGS system used in casework can be assessed, at least partly, by investigating (1) repeatability, (2) reproducibility, (3) calibration accuracy, and (4) efficiency or discriminating power. Such an exercise will help identify the better-performing PGS systems for consideration in casework applications. Some research has been conducted on reproducibility of a single PGS system in different laboratories (e.g., Riman et al. 2024a) and calibration and discrimination performance of two PGS systems (e.g., McCarthy-Allen et al. 2024).

A 2024 article that examined differences between two PGS systems discussed calibration:

> "The preferred way to examine the validity of a group of LRs is calibration…Calibration is an empirical test to determine whether, on average, a group of LRs is approximately correct…Calibration is not required in any of the guidance documents for [PGS] but the realization of the value of this postdates many of these documents. In the DNA context, calibration presents some challenges since, to calibrate, say, an LR of 10^9 requires over 10^9 [known non-contributor test] experiments or the use of importance sampling…The best path to finding reliable systems is calibration" (Buckleton et al. 2024a).

The accuracy of the reported LR value in any specific casework situation cannot be determined. However, results of LR assessments across a collection of casework-similar, ground-truth known, scenarios can assist in informing the receiver of the LR assessment in deciding how much weight they should give to the LR assessment in the case at hand.

The specific propositions selected impact the LRs obtained (see Table 2.4). This fact should encourage continued efforts to standardize development of propositions (e.g., Gittelson et al. 2016, Hicks et al. 2021, Duke et al. 2022). Ground-truth information in validation experiments can only inform us whether the LR assigned supports H_1 when this proposition is true or H_2 when this proposition is true, but it cannot tell us what the LR should be. Studies can, however, estimate the percentage of time the LRs are on the wrong side of 0 (when using $\log_{10}$(LR)) and providing adventitious exclusionary or inclusionary support (see Riman et al. 2021). Sometimes, data may be favorable to H_1 even when H_2 is true. This happens not just due to adventitious matches from a high degree of allele sharing among contributors in DNA mixtures, but also due to limitations of models, particularly with low LRs (see McCarthy-Allen et al. 2024). As discussed in the next section, there are a number of factors that influence DNA mixture interpretation.

4.1.5. Factors Influencing DNA Mixture Interpretation

A number of factors[16] influence the overall process of DNA mixture measurement and interpretation and eventually the obtained results. This includes: (a) DNA extraction and quantitation, STR kits, instruments, and PCR parameters used, (b) actual or apparent number of contributors, (c) degradation levels of DNA from contributors, (d) mixture ratios of DNA from contributors, (e) total DNA template amount, (f) relatedness of potential contributors and

[16] We use the term *factor space* to describe the totality of scenarios and associated variables (*factors*) that are considered likely to occur in actual casework. This may seem to some to be a new term, but it is not a new concept. For example, the FBI QAS (2020) Standard 8.3.2.1 states: "Mixture interpretation validation studies shall include samples with a range of the number of contributors, template amounts, and mixture ratios expected to be interpreted in casework." Likewise, guidance from the International Laboratory Accreditation Cooperation (ILAC) G19:06/2022, which informs accrediting bodies like ANAB and A2LA, states: "When developing their processes, forensic units shall show objective evidence that they have assessed the *factors* that can influence the results and have recorded these" (ILAC-G19:06/2022, Section 3.10; emphasis added).

degree of allele sharing, (g) statistical models used to perform interpretation, etc. Table 4.1 contains a more complete (but not exhaustive) list of factors when implementing PGS systems to assist DNA mixture interpretation at the sub-source level.

Table 4.1. Factors that influence DNA mixture measurements and interpretations with probabilistic genotyping software (PGS) systems. See also Table 2.1.

Areas/Topics	Influencing Factors
Measurement of STR Alleles and Genotypes in the DNA Profile EPG i.e., analyst filtered input file used by PGS system	• Peak position for short tandem repeat (STR) alleles • Peak morphology or resolution for STR alleles • Peak height for STR alleles • Relative peak heights for STR allele pairs (heterozygote balance) • Triallelic patterns • Presence of stutter products and their relative heights compared to associated STR alleles
Sample Complexity	• Number of contributors • Degree of allele sharing among contributors • Total DNA template and contributor template amounts • Mixture ratio of DNA from contributors • Sample quality including degree of degradation • Presence of stutter products and potential minor contributors in a DNA mixture • Allele drop-out or drop-in
Laboratory-/Analyst-Specific Decisions	• Sample processing methods (e.g., extraction, quantitation, target DNA template levels tested) • STR typing kit(s) used • Replicate testing • Number of PCR cycles • Capillary electrophoresis (CE) instrument used • Analytical threshold • Analyst training and experience (with lab protocols) • Assessments of suitability including limitations, if any • Population allele frequencies • Co-ancestry coefficient (i.e., theta value)
PGS Model Decisions	• PGS model used (i.e., discrete or continuous) • Laboratory-specific parameters for use in the PGS model (e.g., peak height variations, probability of allele drop-out, probability of allele drop-in) • Non-contributor data construction and testing

Areas/Topics	Influencing Factors
Software Implementing the PGS Model	• Choice of numerical methods for computing likelihood ratios (e.g., MCMC, numerical integration) • Choice of the number of iterations or numerical integration parameters (e.g., grid size) • Choice of diagnostic checks on the results
Case-Specific Considerations	• Propositions and assumptions

4.2. Review of Publicly Accessible Data and Details

As part of this assessment of the foundations of DNA mixture interpretation methods and practices, data and details were examined from publicly available sources. Information was gathered in five areas: (1) published developmental validation studies from STR typing kits, (2) published PGS studies, (3) publicly accessible PGS internal validation studies or summaries from forensic laboratories, (4) DNA mixture proficiency test results, and (5) interlaboratory studies assessing DNA mixture interpretation. Findings are summarized in a supplemental document (NISTIR 8351sup2). Interested readers are invited to examine the details there along with some observations on the collected information. In the rest of this chapter, some thoughts are provided on what can be learned from this publicly accessible information and how to strengthen DNA interpretation at the sub-source level.

4.3. Discussion

In Table 2.5 under factors considered at the sub-sub-source (DNA contributor) level in the hierarchy of propositions, two important aspects are included: (1) reliability of the DNA mixture interpretation protocol (as determined by validation studies with ground-truth information) used in similar situations to the case in question, and (2) an understanding of variability across the mixture factor space. PGS studies (e.g., Susik & Sbalzarini 2023a) have noted some general trends for the strength of evidence with sub-source level propositions from assigned LRs produced by PGS systems. These LRs typically decrease as

- the total amount of DNA material in the mixture decreases
- the amount of DNA material from the person of interest decreases
- the relative proportion of the contributor decreases (i.e., lower mixture ratio)
- the sample quality decreases (i.e., more DNA degradation or a higher presence of PCR inhibitors in the sample).

Demonstrating a trend though is not the same as showing the reliability of a specific LR in DNA mixtures with various levels of complexity. An assigned LR comes from the biological and statistical models and assumptions used when creating a PGS system. In a 2021 article, a researcher from the Netherlands Forensic Institute made the following observations:

> "The better the model, the more evidence it can find for actual contributors, but a worse model will never produce strong evidence for many non-contributors. The aim of the model is to reduce the uncertainty around the profiles of the contributors making use of the trace profile…Sometimes the mixture data may be poor…, but that does not necessarily imply the technology becomes unreliable, only that it can be expected to give less conclusive results…Therefore in mixture analysis, smaller likelihood ratios may be more prone to changes that are important for decision making when re-evaluated with a different model…" (Slooten 2021).

An article in *The New Yorker* in January 2024 discussed how mathematical models, like those used in PGS systems, have limits. The author shared some valuable insights that are worth repeating here:

> "Like model cars, model airplanes, and model trains, mathematical models aren't the real thing—they're simplified representations that get the salient parts right. Like fashion models, model citizens, and model children, they're also idealized versions of reality… [T]he point of modelling…is to see how far you can get by using only general scientific principles, translated into mathematics, to describe messy reality… Sometimes the phenomenon you want to model is simply unmodellable. All mathematical models neglect things; the question is whether what's being neglected matters… How much should we trust them, and why?... All models reflect choices about what to include and what to leave out…[Models] need to be measured against external data… If we don't understand how a model works, then we aren't in a good position to know its limitations until something goes wrong" (Rockmore 2024).

As discussed in the next section, comparison studies with different PGS systems and models can assist with understanding the level and magnitude of variations where models differ and where potential limitations may exist.

4.3.1. PGS Comparison Studies

To better understand how PGS systems using different algorithms and models may differ, assigned LRs from multiple PGS systems can be assessed using the same DNA samples or profiles. Some of the published PGS studies listed in the accompanying supplemental document examined multiple PGS systems (see Table S2.2 in NISTIR 8351sup2).

Table 4.2 distills information from 20 PGS comparison studies including the PGS systems compared, numbers and types of samples tested, and observations made for each cited reference. This information comes from 18 PGS studies[17] cited in Table S2.2 along with two others that involved casework mixtures without ground-truth knowledge of the number of contributors (Costa et al. 2022, Thompson 2023).

Table 4.2. NIST-extracted summary of published PGS comparison studies. For reviews on PGS systems, see Coble & Bright 2019, Butler & Willis 2020, and Gill et al. 2021a.

PGS Systems Compared Reference	Samples Tested	Observations Made
Lab Retriever (v.1.2.1) STRmix (assume v.2.0) Bille et al. 2014	Examined **a single Identifiler two-person mixture** with a low degree of allele sharing (10 of 15 loci displayed non-overlapping four alleles) at mixture ratios of 1:1, 1:2, 1:3, 1:4, and 1:5 with total template quantities of (100, 200, 300, 400, and 500) pg DNA amplified in duplicate (resulting in 50 mixture samples)	Plotted LR data points from the discrete (Lab Retriever) and continuous (STRmix) PGS systems along with random match probability (RMP) and combined probability of inclusion (CPI) mixture statistics (their Figure 1); reproducibility improves with higher qualities of total DNA; "information content associated with height is limited for the 1:1 mixtures but increases as we proceed toward the 2:1, 3:1, 4:1, and 5:1 mixtures"; the authors conclude: "It is noted that this trial was conducted on a relatively easy type of mixed DNA profile, two-person mixtures. Further comparison with three- and four-person mixtures and profiles where the person of interest is potentially masked is warranted."
LiRa likeLTD (v.4.4) LRmix (v.4.3) Puch-Solis & Clayton 2014	Examined **ten replicates of a single SGM Plus profile** (simulating a single-source, low-level DNA from Balding 2013 Table S1) where allele drop-out, drop-in, and uncertain designations are possibilities	Differences were observed with each PGS system (all are discrete models) even when only alleles (and no peak height differences) are considered; authors introduced concept of "ban evidential efficiency"; four experiments were conducted: (1) one-person profiles with no replicates, (2) one-person profiles consisting of two and three replicates, (3) two-person profiles, and (4) three- and four-person profiles; more variation was observed between PGS systems as profiles became more complex; with an example involving propositions of three people, results were (in *bans*): likeLTD (*9.3*), LiRa (*8.98*), and LRmix (*3.99*) – meaning that LiRa and LRmix were five *bans* or five orders of magnitude different in this example

[17] These 18 studies are listed in Table S2.2 as "*multiple" PGS systems – see rows #49 to #61, #64, #69, #70, #71, and #72 (NISTIR 8351sup2).

PGS Systems Compared Reference	Samples Tested	Observations Made
Lab Retriever (v.1.2.4) LRmix (v.4.3) STRmix (v.2.0) Bright et al. 2015	Used **two artificial Identifiler profiles** to create major/minor, balanced, and stochastic profiles (profiles are provided in supplementary material)	Performed four experiments: (1) comparison to the expected LR with no drop-out or drop-in, (2) the effect of drop-out, (3) the effect of drop-in, and (4) reproducibility; comparison of results identified a difference in how Lab Retriever calculated their population genetic model compared to the other two PGS systems; the authors suggest: "an essential feature of validation is the ability to specify exactly what the software is doing at least with regard to routine matters such as the population genetic model and the allele probabilities"
EuroForMix *DNAmixtures* Bleka et al. 2016a	Examined **a two-person mixture amplified with PowerPlex ESX 17**; also simulated three random DNA profiles where one, two, three, or four individuals contributed	Compared assigned LRs between EuroForMix and *DNAmixtures* by randomly generating single-source profiles and two- and three-person mixtures; the authors observed identical log LRs up to 11 decimal places for each considered proposition
EuroForMix LRmix Studio Bleka et al. 2016b	Examined **four two-person and 55 three-person mixtures amplified with NGM**; see Table 1 in their article; full dataset available at http://www.euroformix.com/data	Used receiver operating characteristic (ROC) plots to examine the rate of false positives versus true positives across different conditions; the authors reported: "LRmix still gave a high LR for true contributors up to four dropouts for a person of interest (POI) in a three-person mixture. However, the main benefit of EuroForMix was with the interpretation of major/minor mixtures where the minor was evidential. Here up to 11 allele dropouts for the POI in a three-person mixture could provide probative evidence, whilst LRmix may return a much lower LR or a false negative result. The two models are expected to return similar LR results when contributors have equal mixture proportions or for mixtures of higher order."
Kongoh LRmix Studio (v.2.1.3) EuroForMix (v.1.7) Manabe et al. 2017	Examined **18 mixtures** (6 two-person, 6 three-person, and 6 four-person) **amplified with Identifiler Plus**; see Tables S1 and S2 in their article	Used bar charts to compare LR values from a binary model, LRmix Studio, EuroForMix, and Kongoh for two-person (Figure 6), three-person (Figure 7), and four-person (Figure 8); the authors reported: "LR values of Kongoh tended to be similar to those of EuroForMix even in four-person mixtures…[except with a] minor POI of 7:1:1:1 mixtures with 0.25 ng DNA and with three drop-out alleles of the POI"

PGS Systems Compared Reference	Samples Tested	Observations Made
DNA•VIEW (v.37.17) EuroForMix (v.1.9.3) Lab Retriever (v.2.2.1) LRmix Studio (v.2.1.3) STRmix (v.2.3.06) Alladio et al. 2018	Examined **7 mixtures** (3 two-person and 4 three-person) plus a dilution series of a 1:1:1 mixture from 500 pg total down to 4 pg **amplified with seven STR kits** (GlobalFiler, NGM SElect, MiniFiler, PowerPlex Fusion, Fusion 6C, ESI 17 Fast, and ESX 17 Fast); mixtures were made with NIST SRM 2391c components A, B, and C	Plotted log(LR) data points from the five PGS systems by mixture ratio, NIST component, and STR kit; also plotted averaged log(LR) values from the two discrete PGS systems versus the three continuous PGS systems; created histograms to compare averaged discrete vs averaged continuous LR results for each NIST component against the overall DNA quantity in the dilution series; the authors reported: "[continuous PGS] results were always higher than the [discrete PGS] ones, regardless of the DNA amplification kit that was adopted" and "LR results provided by both [discrete PGS] models were very similar or identical" while "log(LR) results provided by [continuous PGS] models proved similar and convergent to one another, with slightly higher within-software differences (i.e., approximately 3-4 degrees [orders] of magnitude)"
EuroForMix (v.1.10.0 and v.1.11.4) Lab Retriever (v.2.2.1) LRmix Studio STRmix (v.2.5.11) Buckleton et al. 2018	Examined **one Identifiler and four Identifiler Plus profiles** and reference samples from five NIST MIX13 mock cases; data available at https://strbase.nist.gov/NIST_Resources/Interlabortory_Studies/Mix_13.zip	Provided LRs from each PGS system compared to 1/RMP for each reference sample in case 1 (Table 4), case 2 (Table 5), case 3 (Table 7), case 4 (Table 9), and case 5 (Table 11); the authors reported on the case 1 results: "All four [PGS] tested also included reference 1A with as much as four orders of magnitude difference between software systems (see Table 4). The continuous model software systems reported the larger LRs and the [discrete] software systems essentially reported the same LR"; these general trends were observed for cases 2, 3, and 4, namely (1) that the two discrete PGS systems yielded similar results (usually less than an order of magnitude apart) as did the two continuous PGS systems to one another and (2) continuous systems assigned higher LR values than discrete ones; the assigned LR results differed in case 5, which were discussed by the authors as an "over engineered" challenge involving a non-contributor reference profile possessing extensive allele overlap and that inclusion of this reference "should be termed an adventitious match not a false inclusion"
CEESIt (four models labeled A, B, C, D); see their Table 1 for model assumptions Swaminathan et al. 2018	Examined **101 Identifiler Plus profiles** (30 single-source, 41 two-person, 30 three-person samples) five times each; see Table S2 and Table S3 in their article	Provided summaries of minimum and maximum LR values for each model with some other statistics (Table 2); for each model 1010 LRs were produced (150 single-source, 410 two-person, and 450 three-person); the authors reported: "In all four models, intramodel variability in the LRs increased with an increase in the number of contributors and with a decrease in the contributor's template mass."

PGS Systems Compared Reference	Samples Tested	Observations Made
likeLTD (v.6.3.0) EuroForMix (v.1.11.4) You & Balding 2019	Examined **72 NGM SElect profiles** (36 single-source, 24 two-person, 12 three-person samples); see Table 1 in their article	Used ROC plots with different thresholds and an information gain ratio (IGR) compared to the inverse match probability (1/RMP) that serves as an upper bound; the authors reported: "Overall results from likeLTD and [EuroForMix] were similar, despite being based on different modelling assumptions."
LRmix Studio (v.2.1.3) STRmix (v.2.5.11) Rodriguez et al. 2019	Examined **102 two-person mixtures amplified with PowerPlex 21**; see Table 1 in their Supplemental file	Provided LRs for each sample and PGS system with H1 true LRs (Table 2 in Supplemental file) and H2 true test results (Table 3 in Supplemental file); also plotted log(LR) values against the number of drop-outs in the POI; the authors reported: "The capacity of the LR approach to discriminate between true and false propositions increases with the amount of correct information provided."
EuroForMix (v2.1.0) STRmix (v2.6) Riman et al. 2021 See also Buckleton et al. 2022 Riman et al. 2022	Examined **154 two-person, 147 three-person, and 127 four-person mixtures** from the PROVEDIt dataset; see Supplemental Table 4 in their article	Provided LR values for 1279 Hp-true tests (Supplemental Table 4) and 1279 Hd-true tests (Supplemental Table 5) for each software; explored LR distributions observed and used ROC plots, scatter plots, histograms with distribution of differences; evaluated apparent discrepancies between PGS models, adventitious exclusionary and inclusionary support, and verbal equivalent discordance; the authors reported: "in certain cases differences in numerical LR values from both software resulted in differences in one or more than one verbal categories (Table 8). These differences were substantially more with low template minor contributors and higher [number of contributors]…" Concerns were later raised (Buckleton et al. 2022) about stutter peak retention, input files used, universal versus dye-specific analytical thresholds, different rare allele models, and PGS diagnostics consideration. The authors' response noted: "The availability of the ground truth information for each mixture profile was beneficial in examining the possible reasons behind these differences", "different LR systems with comparable discrimination performance exhibited variability in the assigned LR values", and "we have fully disclosed our protocols so that anyone else can evaluate the performance metrics of those protocols" (Riman et al. 2022).

PGS Systems Compared Reference	Samples Tested	Observations Made
EuroForMix (v3.0.3 and 3.3.0) STRmix (v2.7.0) Cheng et al. 2021 Buckleton et al. 2024a	Examined **74 two-person, 30 three-person, and 25 four-person mixtures** from the PROVEDIt dataset; see Supplemental Materials in their article	Provided scatter plots of the $\log_{10}$LR versus the average peak height along with analysis of several divergent results; discovered a miscode in EuroForMix regarding the stutter models used; the authors concluded: "After taking into account the differences in allele probability models, the LRs from EuroForMix and STRmix for single-source profiles were the same to at least two significant figures" and "LRs for 84% of the comparisons for known contributors without rare alleles were within two orders of magnitude"; a follow-up article (Buckleton et al. 2024a) examines how the rare allele probability models differ and discusses the importance of and issues with calibration of assigned LR values; the authors conclude: "…all the comparison studies between [EuroForMix] and STRmix have demonstrated that the two different models can both be useful…"
EuroForMix (v3.3.1) MaSTR (v1.11) Adamowicz et al. 2022	Examined **8 two-person, 15 three-person, and 4 four-person mixture profiles**; observed up to 12 orders of magnitude difference between PGS systems (5.82×10^{28} with MaSTR vs 6.62×10^{16} with EuroForMix for a three-person mixture 1:1:2)	Provided LRs assigned by each PGS system on each of 27 samples examined (Supplemental Table S5); the authors stated: "the LRs generated by the two software packages using the same propositions were generally concordant. Some variation in the LR results between the two was expected…however, the variations tended to be relatively small and rarely changed the outcome on the verbal scale."
EuroForMix (v3.4.0) LRmix Studio (v2.1.3) STRmix (v2.7) Costa et al. 2022	Examined **156 casework mixture profiles** from the Portuguese Scientific Police Laboratory with estimated two or three contributors	Provided LRs assigned by each PGS system on each sample (Supplemental Tables S1 and S2); the authors stated: "Different software products are based on different approaches and mathematical or statistical models, which necessarily result in the computation of different LR values. The understanding by the forensic experts of the models and their differences among available software is therefore crucial. The better the expert understands the methodology, the better he/she will be able to support and/or explain the results in court or any other area of scrutiny."

PGS Systems Compared Reference	Samples Tested	Observations Made
STRmix (v2.6) EuroForMix (v3.4.0) Custom Hamiltonian Monte Carlo algorithms Susik et al. 2022 Susik & Sbalzarini 2023a, Susik & Sbalzarini 2023b	Examined **11 LRs from MIX13** and **428 mixtures** (**154 two-person, 147 three-person, and 127 four-person)** from the PROVEDIt dataset (same as Riman et al. 2021); see Supplemental Materials with their articles for plots visualizing precision of the methods	Provided average LRs over 10 runs with their Hamilton Monte Carlo (HMC) algorithm in comparison with 11 previous EuroForMix and STRmix results (Buckleton et al. 2018) from NIST MIX13 profiles (Table 4 in Susik et al. 2022); the authors stated: "Our algorithm reproduces most of the results of other solutions, suggesting its validity. Similar to other solutions, our algorithm provides more conservative LR values when a smaller number of contributors is chosen." Provided multiple plots in comparing their HMC algorithm with other models including histograms and ROC curves and developed a new metric, opposite of the neutral threshold (OotNT) (see Figures in Susik & Sbalzarini 2023a; Table 2 in Susik & Sbalzarini 2023b)
STRmix (v2.8) TrueAllele Casework VUIer (v3.3.8258.2R20b) Thompson 2023 See also Kalafut et al. 2024 Perlin et al. 2024	Explored variation in results from **a single low template DNA profile** through locus-by-locus comparison of the probabilities assigned to a defendant's genotype and the resulting likelihood ratios computed by both PGS systems	Provided detailed analysis and commentary on assumptions made and LRs assigned with both PGS sets of results examined; the author wrote: "This discussion highlights the importance of establishing the limits of validity for [probabilistic genotyping] programs…Finding the limits will be important to courts evaluating the admissibility of PG results in cases like this one where labs may be working near or even beyond those limits."
STRmix (v2.7) run in three laboratories (ESR, FBI, and NIST) Riman et al. 2024a	Examined **265 profiles (19 single-source, 59 two-person, 57 three-person, 55 four-person, 65 five-person, and 10 six-person mixtures)**; Supplemental Tables contain 913 rows of known contributor LRs (H_1-true) and 56,857 rows of known non-contributor LRs (H_2-true) calculations	Provided LRs assigned by each PGS system in the three laboratories along with their $\log_{10}$(LR) differences; over 92% of assigned LRs fell within the same order of magnitude (i.e., had a delta $\log_{10}$(LR) between 0 and 1) for the same input file across the three laboratories; Supplemental Table 3 has 10 replicate interpretations of the same profile from one laboratory; the authors noted: "The main objective was to highlight the types of profiles where the LR values were variable between repeat interpretations and discuss the causes of this variability."

PGS Systems Compared Reference	Samples Tested	Observations Made
STRmix (v2.6 & v2.6.3) TrueAllele Casework (v3.25.5840.1) and VUIer (v3.3.5743.1) Greenspoon et al. 2024	Examined **48 different mixture combinations (18 two-person, 18 three-person, and 12 four-person) mock casework samples** with 152 LR comparisons between TrueAllele and STRmix; "log(LR) values which were ≤2 log10 units apart (two ban) were considered reproducible"; "Log(LR) comparisons were further parsed into 'informative' and 'uninformative' differences. An uninformative difference, regardless of the magnitude, was a difference in log(LR) that was unlikely to affect the conclusion regarding a specific contributor's association with an item of evidence."	Provided LRs assigned by each PGS system on each of samples examined (Table 1); observed less than two orders of magnitude difference between PGS systems with 52% (79 of 152 comparisons), >2 and <4 orders of magnitude difference with 22% (33 of 152), and ≥4 orders of magnitude differences with 26% (40 of 152); the largest observed difference was 14 orders of magnitude (log(LR) -7.1 with TrueAllele vs log(LR) 6.1 with STRmix for a four-person mixture 5:2:2:1 with a degraded DNA minor contributor); the authors stated: "These systems converged on the same result >90% of the time and when moderate or substantial differences in log(LR) values were observed, most of these would not affect the conclusion of the reference profile association (or not) to a mixture as defined here." They further noted: "these systems are more likely to deviate with low-template contributors."
DNAStatistX (v2.1.0) EuroForMix (v4.0.8) McCarthy-Allen et al. 2024	Examined **428 mixtures** (**154 two-person, 147 three-person, and 127 four-person)** from the PROVEDIt dataset (following Riman et al. 2021; same input files as Susik & Sbalzarini 2023a so comparisons were also made with their HMC and STRmix results)	Provided scatter plots of the $\log_{10}$LRs with H_1-true and H_2-true tests comparing the various PGS systems (Figure 4) as well as analysis of false positives and false negatives for given LR ranges (Table 3), by software (Figure 5), by NOC using one software (Figure 6), and using ROC curves (Figure 7); the authors concluded: "The results confirm previous observations that high H_p-true LRs (LR>1000) were similar to other PG software when performed with the same input data and propositions."

Some of the early studies listed in Table 4.2 examined only one or two samples at various parts of the mixture factor space. Yet these studies are still helpful in understanding differences with various PGS models and verifying DNA principles described in Section 2.7 of this report. For example, a single two-person mixture with a low degree of allele sharing (e.g., 10 of 15 STR loci displayed non-overlapping heterozygous alleles) was examined at five mixture ratios and five different DNA template amounts with duplication PCR amplifications using a discrete and a continuous PGS system (Bille et al. 2014). This study reported that reproducibility improved between replicates with higher quantities of total DNA (see Principles 8, 9, 10, 11) and that the information content associated with peak height from a continuous system is limited for 1:1 balanced mixtures but increased with mixture ratio differences (i.e., 2:1 to 5:1) (see Principles 14 and 15).

Tools for comparing performance of the PGS systems and the assigned LR values include histograms with distribution differences, scatter plots, and receiver operating characteristic (ROC) curves to assess discrimination performance (i.e., the ability to differentiate) between

the H1-true tests and H2-true tests across different conditions (e.g., Bleka et al. 2016b, You & Balding 2019, Riman et al. 2021). The reported comparisons were described in various manners.

Some caveats in PGS comparisons included removing contributors with rare alleles as different models may be used when statistically accounting for rarely observed alleles. For example, a 2021 study comparing EuroForMix and STRmix reported: "LRs for 84% of the comparisons for known contributors without rare alleles were within two orders of magnitude" (Cheng et al. 2021).

A 2022 study in Table 4.2 reported after examining 27 samples that "the LRs generated by the two software packages [EuroForMix and MaSTR] using the same propositions were generally concordant… variations tended to be relatively small and rarely changed the outcome on the verbal scale" (Adamowicz et al. 2022). An increase in variation with assigned LRs between PGS systems was observed as profiles became more complex (e.g., Puch-Solis & Clayton 2014), and "these differences were substantially more with low template minor contributors and higher [number of contributors]" (Riman et al. 2021). A 2024 study using a different set of PGS systems concurred with this assessment of sample types that exhibit increased differences: "These systems are more likely to deviate with low-template contributors" (Greenspoon et al. 2024).

Publications sometimes mention use of a two-order-of-magnitude grouping with assigned LR values when considering whether these LRs are similar. Regarding their classification approach, a 2024 study stated:

> "Log(LR) values which were ≤ 2 $\log_{10}$ units apart (two ban) were considered reproducible. Log(LR) values that were >2, but <4 ban apart were classified to have a 'moderate' difference and those that were ≥ 4 apart to have a 'substantial' difference" (Greenspoon et al. 2024).

This 2024 study assessed 152 LR comparisons between STRmix and TrueAllele across 48 two-, three-, and four-person mixtures and observed that 52% (79 of 152 comparisons) were within two ban, 22% (33 of 152) were between two and four ban, and 26% (40 of 152) were more than four orders of magnitude different in their assigned LR values. However, only 9% (14 of 152) produced "informative differences" since many of these assigned LRs fell in the "inconclusive or uninformative range" and had an assigned log(LR) value between -3 and +3 (Greenspoon et al. 2024). One mixture in this study exhibited a 15 order of magnitude difference with an assigned log(LR) of -8 with TrueAllele and +7 with STRmix. The authors found examination of locus-specific LR values helpful in trying to understand algorithm differences between the two PGS systems (see Table 2 in Greenspoon et al. 2024), and they concluded:

> "This particular mixture exposed two of the greatest underlying differences in how these systems model mixtures: how missing allele information is incorporated into the process and the modeling of stutter" (Greenspoon et al. 2024).

These differences were influenced by the DNA template amounts and the potential for allele drop-out with low DNA quantities. The authors of this study state:

> "The [TrueAllele] system handles missing allele data or drop-in peaks profoundly differently from STRmix or EuroForMix in that it samples down to 10 RFU so every peak is modeled, and all possible genotype combinations are assessed" (Greenspoon et al. 2024).

4.3.2. Variability in Assigned LR Values from the Same Sample

Information in these publications summarized in Table 4.2 is often treated in aggregate and displayed as scatter plots. Without specific details about the samples, including the assigned LR values and metadata about the complexity of the mixture such as the degree of allele sharing, then reasons for differences cannot be independently assessed. While there have been many samples tested in the compiled information sources examined (Table 4.2 and NISTIR 8351sup2), it should be noted that important details are sometimes missing. For example, when differences in assigned LRs were observed in these publications, a reader typically cannot access the assigned LR values nor know anything about the degree of allele sharing in the mixture without the contributor genotypes.

Considering performance of the same PGS system in a single laboratory with different detection platforms, a 2022 study compared mixed DNA profiles produced from a rapid DNA testing platform[18] with those generated in a standard laboratory workflow. The authors noted that "classically expected trends in LR were seen for the RapidHIT ID data. Specifically, the LRs contracted towards one for both decreasing DNA amount and increasing profile complexity" (Ward et al. 2022). The study found that in general there were lower peak heights produced by the rapid DNA instrument compared to the standard laboratory workflow involving an ABI 3500xl capillary electrophoresis instrument and that the properties of peak height variability distributions differed (see Table 1 in Ward et al. 2022). Since assigned LR values from the same PGS system can be sensitive to peak height variation (see Bright et al. 2019), a comparison of LRs between the rapid DNA and standard workflow showed that almost a third (9 of 30) of the samples appeared to have assigned LRs that differed by at least 10 orders of magnitude on the same samples with one sample that differed by >20 orders of magnitude (see Figure 5 in Ward et al. 2022). Specific details about these differences were not described in the publication beyond several graphical plots that help understand trends but not details.

This type of limited information can be contrasted with another study that shared more details.

Considering performance of the same PGS system across three different laboratories, a 2024 study compared 265 DNA mixture profiles using identical input files, NOC settings, propositions, database of true and false donors, laboratory-specific PGS settings, allele frequencies, and subpopulation correction (Riman et al. 2024a). Over 92% of assigned LRs fell within the same order of magnitude for the same input file, i.e., had a delta $\log_{10}$(LR) between 0 and 1, across the three laboratories (see Figure 4 in Riman et al. 2024a). For the remaining 8%, the authors conducted further analysis and describe what they found:

[18] These results are from a research study. Including this information here is in no way meant to suggest that rapid DNA testing should be applied to DNA mixture interpretation or that LR values from rapid and conventional workflows should be compared. This comparison is discussed to point out that assigned LR values can differ for the same DNA sample depending on the inputs used, including the detection platform.

> "Investigation into the details that led to this higher than expected run-to-run LR variability [i.e., one order of magnitude difference across the three comparisons] was undertaken by interpreting these profiles a further ten times within the one laboratory and reviewing the summary statistics contained in the reports of the repeated interpretations…A review of the diagnostics identified five key reasons for poor precision" (Riman et al. 2024a).

The five key reasons are described in their publication, and examples provided "to highlight the types of profiles where the LR values were variable between repeat interpretations and discuss the causes of this variability" (Riman et al. 2024a). Importantly, the 913 known contributor LRs and 56,857 known non-contributor LRs for each laboratory and computed interlaboratory differences are all publicly accessible (see Supplemental Tables 1 and 2 in Riman et al. 2024a).

KEY TAKEAWAY #4.2: There is a growing body of scientific literature on DNA mixture interpretation. However, supporting data provided in the scientific literature is not always sufficiently detailed for an independent review of claims. Such data and details, if required as part of the journal publication acceptance process, will assist with independent review of published articles.

4.3.3. Reliability Assessment of LR Values

Many studies have demonstrated that when less information is available for a contributor in a DNA mixture profile, such as fewer peaks in an EPG due to low DNA quantities, then lower LR values are assigned. But showing trends (i.e., lower LR values when there is less information) may not be sufficient to stakeholders and users of information coming from assigned LRs. There can be an expectation among users of the information that a specific LR value has a particular meaning. What is fit-for-purpose with LR values? What influences reliability of assigned LRs?

Variability exists across forensic DNA laboratories as there are multiple parameters and settings in DNA measurement and interpretation processes. Some examples of variation include use of different STR typing kits or number of cycles when performing PCR amplification, different analytical thresholds with different capillary electrophoresis instruments, and different interpretation protocols. As noted in Table 4.1, there are many factors that influence DNA mixture interpretation.

In an article titled "The most consistent finding in forensic science is inconsistency," author Itiel Dror stated:

> "From a scientific point of view, it does not matter who or where an analysis is conducted, the same analysis must yield the same results. There can be no science if the same exact analysis gives different results" (Dror 2023).

In some situations, such as with a large number (e.g., LR $>10^{20}$), it may be acceptable to have an assigned LR value from one PGS system that differs by several orders of magnitude from an assigned LR value from a second PGS system. In other situations, such as with a small number

(e.g., LR <100), then variation of several orders of magnitude is more significant in how users of the information might view the assigned LR value.

A follow-up article titled "Extending the discussion on inconsistency in forensic decisions and results" addressed variation in assigned LR values using PGS systems (Buckleton et al. 2024b). These authors conclude:

> "There is a demonstrated variance to our results from inter-laboratory studies even in quite constrained situations. Some, but not all, of this variance arises from case-by-case human decision-making… One approach to improve consistency between laboratory's approaches to forensic casework is regular international inter-laboratory trials where feedback is given as quickly as possible by publication on a website. These trials should be complex but not unreasonably complex initially since they are aimed at motivating consistency rather than the oft-cited 'exploring the boundaries.' There will need to be a system for dealing with variability created by equipment or software differences…Trials such as we have described will identify the areas of decision-making where variability in decision-making exists and its magnitude…Once identified, the causes of variability can be addressed through changes in process and protocol…" (Buckleton et al. 2024b).

A framework for comparing PGS results amongst different laboratories has been proposed with 14 specific steps, including using only STR loci in common among the participating laboratories along with the same population allele frequencies, the same population genetic model, the same population substructure correction, and the same propositions (McNevin et al. 2021, McNevin & Barash 2024). Based on what has been reported in interlaboratory studies to-date (see Section 7 in NISTIR 8351sup2), there will still be variation in assigned LR values, likely due to PCR amplification variability particularly with low-level contributors experiencing stochastic variation (as noted with Principles 8, 9, and 10 in Chapter 2). Low template contributors add to potential uncertainty in possible genotype contributors.

Forensic DNA laboratories in the United States conduct internal validation studies based on the FBI Quality Assurance Standards (e.g., QAS 2020) and SWGDAM guidelines (e.g., SWGDAM 2015) "with the appropriate sample number and type to demonstrate the reliability and potential limitations of the method" (QAS 2020, Standard 8.3). Data from internal validation studies "may be shared by all locations in a multi-laboratory system" and "the summary of the shared validation data shall be available at each site" (QAS 2020, Standard 8.3.1.1). Results from the internal validation are intended to "define quality assurance parameters and interpretation guidelines, including, as applicable, guidelines for mixture interpretation and the application of appropriate statistical calculations" (QAS 2020, Standard 8.3.2). In particular for DNA mixture interpretation: "Mixture interpretation validation studies shall include samples with a range of the number of contributors, template amounts, and mixture ratios expected to be interpreted in casework" (QAS 2020, Standard 8.3.2.1) and be "documented and summarized" as well as "reviewed and approved by the technical leader prior to implementing a procedure for forensic applications" (QAS 2020, Standard 8.3.4). According to the QAS, an internal validation study must be assessed at least once at a minimum: "Each validation study shall be evaluated and approved during one external audit. Approved validation studies shall be documented in the Audit Document" (QAS 2020, Standard 15.2.2).

Information about performance and limitations of methods derived from validation studies should inform laboratory protocols. The FBI QAS emphasizes this aspect in two standards:

> "The laboratory shall have and follow analytical procedures *supported by the internal validations* and approved by the technical leader" (Standard 9.1, emphasis added).
>
> "The laboratory shall have and follow written guidelines for the interpretation of data that are *based on and supported by internal validation studies* (Standard 9.6, emphasis added).

Thus, internal validation studies are intended to inform laboratory protocols per the FBI QAS. Yet the authors of this report were unable to find publicly accessible information clearly linking protocols with validation data, largely because internal validation data is not publicly accessible.

Although a number of PGS studies have been published and internal validation studies conducted (see NISTIR 8351sup2, Tables S2.2 and S2.4), in their present form, publicly accessible internal validation summaries do not provide sufficient information to independently assess the degree of reliability of protocols in use by forensic laboratories today. Further, these summaries typically do not provide data points (e.g., LR values) and associated information and metadata (see Box 4.1) necessary to assess the degree of reliability and performance under potentially similar case scenarios.

Box 4.1. Desired Information for Reliability Assessments of LR Values in PGS Systems

Although not necessarily a comprehensive list, the following information should help an independent reviewer assess reliability of a DNA measurement and interpretation (end to end) system. With this information, reliability assessments could include (1) assessment of discrimination ability, (2) LR value calibration accuracy in PGS systems, and (3) some exploration of regions of the factor space where LR values assigned by a PGS system are more reliable versus less reliable. If such data are available for different PGS systems, then a performance comparison may be possible (e.g., You & Balding 2019, Riman et al. 2021).

1. Sample Number or Unique Identifier
2. Number of Contributors (NOC)
3. Target DNA Template Amounts
4. Degradation Status of DNA Template(s)
5. NOC Used for Analysis (Apparent NOC)
6. H_1 True? (Yes/No)
7. Person of Interest (POI) Position in the Mixture (if H_1 is true)
8. Reported Log_{10}(LR)
9. Mixture EPG Results*
10. POI Profile*
11. Known Contributor A Profile* and Any Additional Known Contributors
12. Noncontributor Profile (if H_1 is Not True): Is This Profile Simulated or Determined from an Actual Sample?
13. Analytical Threshold and Instrument Model Used for Analysis
14. PGS Parameters and Settings

* Profiles should only be collected and shared from individuals with explicit consent

In the supplemental document to this report summarizing publicly accessible information (NISTIR 8351sup2), it was noted that the FBI QAS and other accreditation requirements have not previously compelled forensic laboratories to make their validation data or summaries publicly accessible beyond "peer-reviewed publication of the underlying scientific principle(s) of a method" involved in developmental validation (see QAS 2020, Standard 8.2.2). Forensic laboratories conduct internal validation studies to demonstrate that specific methods perform as expected in their individual environments – and thus levels of performance seen as appropriate may vary across laboratories. And as noted above, auditors do assess the types of studies performed as part of QAS audit procedures. To date, public accessibility of this internal validation information has not been a primary goal for most forensic laboratories.

KEY TAKEAWAY #4.3: Currently, publicly accessible validation data does not have the detail (including metadata, protocols, conditions, etc.) to enable an external and independent assessment of the degree of reliability of DNA mixture interpretation practices, including the use of probabilistic genotyping software (PGS) systems.

4.3.4. Comments on Publicly Accessible Data

As mentioned earlier, an accompanying supplemental document provides details on publicly accessible information regarding DNA mixture interpretation located as part of this scientific foundation review (see NISTIR 8351sup2). Details are described there for the following:

- 3 STR kit developmental validation studies (Table S2.1),
- 72 published PGS studies (Table S2.2),
- 20 PGS internal validation summaries from forensic laboratories (Tables S2.3 and S2.4),
- Proficiency test results with DNA mixtures from
    - 109 datasets of biological samples, usually two-person mixtures (Table S2.5),
    - 22 datasets of electropherogram interpretation (Table S2.6), and
    - 4 datasets of biological samples interpreted with assigned LRs from PGS systems (Tables S2.7, S2.8, S2.9, S2.10, and S2.11), and
- 20 interlaboratory studies assessing DNA mixture interpretation (Table S2.12).

Developmental validation studies for STR typing kits[19] are focused on measurement aspects important to effective genotyping of single-source samples and parameters that can inform mixture interpretation guidelines. As discussed in Section 3 of the accompanying supplemental document (NISTIR 8351sup2), typically a single two-person mixture is examined with various mixture ratios being the primary variable explored. The overall success rate in detecting non-overlapping minor contributor STR alleles is a commonly used metric in these publications, yet the degree of allele overlap, which depends on the mixture components, is typically not described. While mixture studies in STR kit developmental validation efforts address a requirement in guidance documents (e.g., see Table S1.4 in NISTIR 8351sup1), they offer limited information on performance of any DNA mixture interpretation protocols.

Numerous PGS studies have been published in peer-reviewed journals. As discussed in Section 4 of the accompanying supplemental document (NISTIR 8351sup2), many of the publications did not contain information and details to aid independent review of the data in them. For example, assigned LR values and their associated propositions would be helpful to include in supplementary material accompanying peer-reviewed publications. A 2023 editorial in *Forensic Science International: Genetics* emphasized: "With their submission, authors must ensure that sufficient information is available for independent verification and replication of their findings" (Kayser et al. 2023).

Internal validation experiments are crucial to establish capabilities and limitations of methods and development of protocols used in a laboratory. As discussed in Section 5 of the accompanying supplemental document (NISTIR 8351sup2), some of the publicly accessible PGS internal validation summaries did not explicitly state critical information such as the number of samples tested, or the provided sample numbers do not agree is different parts of the

[19] The 2006 article on urban legends surrounding validation (Butler 2006) was written in the context of STR kit validation at that time and should not be extrapolated to efforts to assess DNA interpretation reliability with assigned LR values from modern PGS systems.

validation summary. The validation summaries examined often illustrated results in graphical form without details such as assigned LR values and propositions used. Sometimes it is not clear whether mixtures contain biological relatives and/or a high-degree of allele sharing, or whether the range of mixture ratios, DNA template quantities, and number of contributors mentioned in these summaries cover case scenarios encountered in that laboratory.

Participants in DNA mixture proficiency tests (PTs) generally do very well in terms of correctly including true contributors and correctly excluding non-contributors. As discussed in Section 6 of the accompanying supplemental document (NISTIR 8351sup2), these PTs are typically two-person mixtures prepared by mixing equal volumes of blood or semen containing high quantities of DNA template. For DNA interpretations from EPGs discussed in the accompanying supplemental document, all 15 false exclusions and 90% of inconclusive results came from three-person mixtures (see Section 6.4.1 in NISTIR 8351sup2). Assigned LR values are not required for most of the PTs. Since 2022, four publicly accessible PT datasets included PGS information. A four-person mixture with allele drop-out resulted in participants selecting numbers of contributors ranging from 3 to 5 and assigning LR values ranging from 10^2 to 10^{28} along with seven false exclusions (i.e., LR<1) using several PGS systems and different propositions (see Section 6.4.3 in NISTIR 8351sup2).

Over the past 25 years, there have been 20 published interlaboratory studies involving DNA mixtures. As discussed in Section 7 of the accompanying supplemental document (NISTIR 8351sup2), these studies reflect performance across a portion of the forensic DNA community at the time these studies were conducted. Most of the early studies focused on two-person mixtures. It has only been in the last decade or so that performance with low-level, high-contributor mixtures has been studied, in large measure due to PGS use expanding across the community. While several of the earlier studies provided biological samples or DNA extracts to participants, many interlaboratory studies have focused on interpretation variability and provided DNA profile EPGs, which means that laboratory-specific measurement variability including DNA extraction and PCR amplification efficiency cannot be assessed as part of the study. Some studies have noted that allowing participants to set their own propositions led to significant differences in the assigned LR values compared to requiring a common set of propositions across all participants. Variation with DNA mixture interpretation and assigned LR values exists among laboratories, even those utilizing PGS systems. Observations and limitations of interlaboratory studies need to be considered in the context of the specific mixtures examined given the experimental design and the participants involved.

Table 4.3 summarizes issues with available information and offers suggestions for future considerations for published PGS studies, internal validation data and summaries, proficiency tests, and interlaboratory studies.

KEY TAKEAWAY #4.4: Current proficiency tests are primarily focused on single-source samples and simple two-person mixtures containing large quantities of DNA. To appropriately assess the ability of analysts to interpret complex DNA mixtures, tests of analysts should include the types of samples often seen in forensic casework, such as mixtures with low-template components and more than two contributors.

Table 4.3. Issues with available information for the data sources examined in this study.

Data Sources	Issues with Available Information	Future Considerations
Published PGS Studies (see Table 4.2 and NISTIR 8351sup2, Table S2.2)	• a lack of uniformity and data details makes comparing information across studies difficult • the following are not consistently provided: contributor genotypes or degree of allele sharing, EPGs of mixtures, ground-truth information on the number of contributors (see Box 4.1)	Adopt a community-wide uniform approach to publishing information with details such as assigned LR datapoints (e.g., Riman et al. 2021) to enable independent assessment of PGS performance (see Box 4.1 and Key Takeaway #4.7)
Internal Validation Data and Summaries (see NISTIR 8351sup2, Table S2.4)	• few forensic laboratories currently provide publicly accessible internal validation data or summaries • contributor genotypes or degree of allele sharing is rarely provided	Adopt a community-wide uniform approach to sharing internal validation information and data to enable independent assessment of DNA mixture interpretation performance (see Box 4.1 and Key Takeaway #4.7)
Proficiency Tests (see NISTIR 8351sup2, Table S2.5 to S2.11)	• mixture PTs consist mainly of simple mixtures with high-quality and quantity DNA and some PTs only utilize single-source samples (e.g., Hundl et al. 2020) • since 2022, probabilistic genotyping PTs have been offered (see NISTIR 8351sup2, Table S2.7 to S2.11)	Require more challenging PT samples (e.g., UKFSR 2020) containing low-level, degraded DNA and mixtures with more than two contributors (see NISTIR 8351sup2)
Interlaboratory Studies (see NISTIR 8351sup2, Table S2.12)	• most previous studies are not relevant to today's PGS methods • laboratories use different methods (e.g., STR kits, PCR cycles, or population databases for allele frequencies) that influence LR assignment at the sub-source level	Conduct regular interlaboratory studies as suggested by others (e.g., Buckleton et al. 2024b) including studies independent of PGS developers (e.g., Hicklin et al. 2023)

4.3.5. Provider-User Responsibilities and Examples

A provider of information, including data, delivers that information to a potential user. The provider explains the intended use of the information. However, the user decides what to accept and how the information will be used. The user decides whether sufficient information exists to judge its reliability and validity relative to the intended application[20]. Users or their expert helpers should be informed and somewhat knowledgeable of the subject in question to assess what they receive.

In some settings, a DNA analyst may be the user of information and in other settings, they may be the provider of information. For example, when deciding on which method to utilize when performing an internal validation study, the DNA analyst may be the *user* of information provided by a developer of an instrument, commercial kit, or software program. As a user performing an internal validation study, the DNA Technical Leader in a forensic laboratory determines whether sufficient data have been collected to demonstrate that a method is fit for its intended purpose within their operational environment (QAS 2020, Standards 8.3.4 and 8.8.5).

On the other hand, when findings are communicated in a written report or through testifying as an expert witness in a court setting, a DNA analyst is the *provider* of information while a trier of fact (judge or jury) and lawyers asking questions in the admissibility hearing or trial are users of the provided testimony as are defendants in a trial (or persons of interest in an investigation). In a court situation, the judge, jury, and lawyers representing their clients determine whether sufficient information has been provided to determine reliability and validity.

When conducting an assessment of reported findings and considering whether results are fit-for-purpose, *whose purpose* matters? A forensic laboratory, an officer of the court, or a PGS provider may have a different perspective on whether a method is fit-for-purpose.

NIST has begun work on a separate scientific foundation review on the communication of forensic findings and how those findings are understood by recipients. This effort began with a workshop on communicating forensic findings that was held in June 2024 (CFF 2024) and is intended to go beyond communication of LRs to include verbal interpretation scales.

KEY TAKEAWAY #4.5: To enable effective use of any information, providers and users of the information both have responsibilities. While a provider explains the relevance and significance of the information and data, only an informed user can assess the degree of reliability and validity, and whether that information is fit-for-purpose for their application.

[20] With this scientific foundation review, the authors of this report serve as *both users and providers* in examining what data and information are publicly accessible (user role) and in describing our findings and their significance (provider role).

4.4. Looking to the Future

Publicly accessible data does not contain sufficient information to show how validation studies translate into protocols. In a mature discipline with an extensive body of literature in peer-reviewed journals that one would expect to find clear indications of how the DNA community measures and interprets mixtures. However, with limited information publicly available, the authors of this report could not assess the underlying data from which the community makes its decisions. While one cannot speak to reliability of assigned LR values from the multitude of PGS systems currently available, greater transparency in data from validation studies is encouraged (e.g., Riman et al. 2021) and clear linkages to protocols and their limitations to enable independent review of DNA mixture interpretation by anyone interested in doing so (e.g., Buckleton et al. 2022, Riman et al. 2022).

The discussion section of this chapter (Section 4.3) comments on limitations in currently available data from PGS systems used for DNA mixture interpretation. This section looks to the future in terms of desired data when conducting independent scientific assessments for LR values assigned by PGS systems and ways that these data might be evaluated to provide increased confidence in these results[21].

4.4.1. Performance Testing with Case-Similar Data

Generally speaking, models and interpretation methodologies developed using known DNA samples may be expected to perform satisfactorily (i.e., be fit-for-purpose) when applied in new but similar scenarios. PGS models may or may not work satisfactorily when applied to data that are unlike scenarios considered in the internal validation training set. Identification of those scenarios in which the performance of a specific method is judged to be inadequate will assist in establishing operational limits for the types of samples that may be reliably interpreted and also point to areas where the measurements or models require improvements. More research is needed in this area, particularly in understanding the impact of allele sharing on the limits of mixture deconvolution and LR assignment.

Alternatively, by demonstrating that a method performs well in scenarios more complex than the case at hand (e.g., test cases with more contributors, less DNA template, or more degradation), based on testing a large number of ground-truth known samples, can inspire confidence that the method may perform well in scenarios like the case at hand, even when there are few (or no) ground-truth known samples with closely matching characteristics. This is a primary purpose of internal validation studies.

Statistical tools are available for examining discrimination efficiency, especially for comparing two or more PGS systems. Receiver operating characteristics (ROC) plots are a tool for this purpose and have been used in evaluation of PGS systems (e.g., Bleka et al. 2016b, You & Balding 2019). Tools for examining calibration accuracy of LR assignments (e.g., Ramos et al. 2013, Hannig et al. 2019) are less widely used by forensic DNA analysts but have been explored

[21] Interested readers may also wish to consult slides from a September 2020 validation workshop covering discrimination power and LR accuracy calibration (ISHI 2020). This workshop covers use of receiver operating characteristics (ROC) plots and illustration of calibration.

by some groups (e.g., Bright et al. 2021, McCarthy-Allen et al. 2024). In addition, interlaboratory studies can assist with understanding areas of caution and consistency.

For example, a 2024 interlaboratory comparison of probabilistic genotyping parameters from eight laboratories concluded:

> "It is our opinion that the largest differences in the reported LR—when using a PGS which results in a likelihood ratio—comes from changes in the propositions that were used…We expect differences in all of these cases [from running different PGS systems] because the underlying interpretation models are different. Equally, *even when using the same software, such as STRmix, different laboratories may report very different LRs for the same sample*…Close inspection of the data presented here shows that by the time the template of a donor reaches about 300 rfu, the results are essentially identical, regardless of the stutter ratios, AT [analytical threshold], PCR cycles, or STR amplification kit. A template value of 300 (template is an approximation of rfu) corresponds to a rough maximum of 150 pg for 28 PCR cycle amplifications and about 100 pg DNA for 29 cycles—these values are the high end of the calculation for donor inputs that correspond to templates of about 300 [pg]" (Boodoosingh et al. 2024, p. 9, emphasis added).

This particular study appears to suggest that interlaboratory results are mostly in agreement above 100 pg to 150 pg per contributor, which is typically outside the stochastic range where significant allele dropout might be expected (see Walsh et al. 1992). What is left unsaid is whether, based on these findings, forensic laboratories should choose to limit their use of PGS for DNA mixture interpretation to samples with higher DNA quantities where allele dropout is unlikely. Protocols in forensic laboratories hopefully have a collection of ground-truth-known analyses underpinning them similar to those as outlined by the conclusions of this 2024 interlaboratory study (Boodoosingh et al. 2024).

As a forensic laboratory's decision maker, the DNA Technical Leader implements protocols that help determine at what point casework samples are considered outside the limits of the set of internal validation experiments performed. A single binary (i.e., yes/no) statement of reliability, based on aggregate performance across many types of samples and many different PGS systems, does not provide the information needed to judge the reliability of the measurement and interpretation in a particular case. Rather, what is needed in the context of a specific case is information concerning the performance of these methods when applied in casework-similar scenarios.

KEY TAKEAWAY #4.6: When assessing the degree of reliability of DNA mixture results for a specific case, the assessor (e.g., an expert user of the results) needs to have access to validation data from known samples that are similar in complexity to the sample in the case.

4.4.2. Data and Details for Independent Scientific Assessments

In a call for open science in forensics, two researchers discussed the importance of public release of a complete analyzable dataset for independent audit and verification of a federally funded study on identification decisions made by forensic firearm examiners (Albright & Scurich 2024). The benefit of data for independent review has been described in the following manner:

> "Data are the foundation on which everything else is built. A clear understanding of what data are is necessary to ensure their meaningful collection and recording. Careful consideration should be taken ahead of starting any project…and thus give the opportunity to adhere to the principles of FAIR (Findable, Accessible, Interoperable, and Reusable). This is applicable to all disciplines and research areas including forensic science… With this knowledge [about what data are needed], the forensic community can come together to decide how data should be organized and shared to strengthen the quality and integrity of research while providing greater transparency to published materials" (Hackman et al. 2024).

Science progresses best when data are accessible to be critically and independently evaluated by other scientists who are independent of the initial research, an aspect that is highlighted in the National Academy of Sciences' publication *On Being a Scientist: A Guide to Responsible Conduct in Research* (NAS 2009).

Available information on DNA mixture interpretation methods and practices may be limited as described in this report and its supplemental documents. Sometimes helpful, or even essential, information is missing. This makes it impossible to know what has actually been examined in a particular study. Note the "N.E.S." designations throughout Tables S2.2 and S2.4 in NISTIR 8351sup2 highlighting where important information is not explicitly stated in the referenced publication. The absence of this information does not necessarily mean that anything was done incorrectly in the study in question. Rather independent reviewers of the information may not be able to fully assess the work without additional efforts such as contacting the authors of a study and requesting further data, which may not be provided despite a request. *The community would benefit from a more uniform approach to both sharing information generally and sharing needed information to enable independent scientific assessments of PGS systems and other DNA mixture interpretation studies performed.*

Improvements in data and metadata transparency of internal validation experiments should assist future efforts with understanding reliability of assigned LR values in various casework scenarios.

The value of having a standard set of information to share when describing validation data can be seen with an approach taken by the digital PCR (dPCR) community, where "Minimum Information for Publication of Quantitative Digital PCR Experiments" (dMIQE) has been adopted and updated in 2020 (dMIQE Group 2020). This group noted:

> "To assist independent corroboration of conclusions, comprehensive disclosure of all relevant experimental details is required. To support the community and reflect the growing use of dPCR, we present an update to dMIQE, dMIQE2020, including a simplified dMIQE table format to assist researchers in providing key experimental information and

> understanding of the associated experimental process. Adoption of dMIQE2020 by the scientific community will assist in standardizing experimental protocols, maximize efficient utilization of resources, and further enhance the impact of this powerful technology" (dMIQE Group 2020).

The dPCR community has found it beneficial to supply a checklist of essential information that can be used by authors, reviewers, and editors when research articles are submitted for publication. This checklist includes details on specimens (types, numbers, sampling, storage), nucleic acid extraction (description of methods, volume used, number of replicates), dPCR protocol (instrument and model, primer and probe concentrations, template treatment, complete thermocycling parameters), assay validation (analytical specificity, analytical sensitivity, testing for inhibitors), and data analysis (description of dPCR experimental design, comprehensive details on negative and positive controls, repeatability, reproducibility, number of partitions measured, partition volume, statistical methods used for analysis, data transparency). For data transparency, raw data from dPCR experiments may be included as supplemental files.

In a spreadsheet that must be completed when a dPCR manuscript is submitted for publication, authors indicate "yes" or "no" for each item on the dMIQE2020 list. When "yes" is selected, a comment box in the spreadsheet can be used to describe the location of the required information (e.g., in a specific supplemental table to the manuscript). When "no" is selected, the comment box is used to outline rationale for the omission, such as why a particular item may not apply depending on the experiment(s) performed.

Adoption of a similar approach would benefit the forensic DNA community with future DNA mixture interpretation assessments to avoid omission of essential information in publications. Similar guidelines for minimum information on PGS validation experiments could be developed by SWGDAM[22], the Organization of Scientific Area Committee for Forensic Science (OSAC) Human Forensic Biology Subcommittee[23], or other similar organizations involving members of the forensic DNA community. In Box 4.1, some suggested information is provided that could be helpful in future reliability assessments of LR values assigned in PGS systems. Availability of these details should enable assessment of discrimination power and LR calibration accuracy for an associated method(s) (e.g., Zadora et al. 2014, Hannig & Iyer 2022).

Within the digital PCR community, dMIQE requirements have (1) enabled authors to design, perform, and report experiments with greater scientific integrity, (2) facilitated replication of experiments described in published studies in which these guidelines are followed, and (3) provided critical information that allows reviewers, editors, and the wider scientific community to evaluate the technical quality of submitted manuscripts against an established standard (dMIQE Group 2020).

A similar approach to the dMIQE data reporting requirements with studies involving PGS systems would benefit the forensic DNA community – both practitioners and users of their

[22] https://www.swgdam.org/ (accessed October 31, 2024)
[23] https://www.nist.gov/osac/human-forensic-biology-subcommittee (accessed October 31, 2024)

data. In addition, sharing more details on validation experiments could provide community-wide cost savings using a collaborative validation approach (Wickenheiser & Farrell 2020).

NIST does not have a regulatory role in forensic science, and therefore, cannot require that details or data be made publicly accessible. Nor do we, as authors of this report, necessarily have the expertise to offer input on all aspects of DNA mixture interpretation or other areas in forensic science. An examination of what guidance or requirements exist for DNA mixture interpretation when this report was written can be found in an accompanying supplemental document (see Table S1.4 in NISTIR 8351sup1).

Advances in and improvements to DNA interpretation at the sub-source and activity levels can take place as members of the forensic DNA community engage through development of specific guidance documents on these topics by SWGDAM, OSAC, or other similar organizations.

Over the past few years, specific LR data and details are being shared in some PGS publications. For example, researchers who conducted an interlaboratory comparison of probabilistic genotyping parameters from eight laboratories provided their assigned LR values in a supplemental data file in addition to a graphical representation of the data in their publication (Boodoosingh et al. 2024). Others have provided full tables of assigned LRs in their study as well as a table of anonymized genotypes of the mixture donors that can be used to assess allele overlap (e.g., Duke et al. 2022). In addition, studies have been conducted where contributor genotypes were shared as donor DNA came from deidentified extracts purchased from a local biobank following an Institutional Review Board approval (e.g., see Supplementary Table S1 in Adamowicz et al. 2022).

As future validation studies are designed and conducted using DNA samples collected under appropriate informed consent, contributor genotypes will be able to be shared according to relevant privacy laws. Availability of contributor genotypes will enable a better understanding of the impact of allele sharing on mixture deconvolution and LR assignment with PGS systems used for DNA interpretation given sub-source propositions.

KEY TAKEAWAY #4.7: To improve data sharing across laboratories and support independent assessments, the Scientific Working Group on DNA Analysis Methods (SWGDAM) and the Organization of Scientific Area Committees for Forensic Science (OSAC) are encouraged to develop minimum requirements and standard formats for data in validation studies and to recommend that validation data be made publicly accessible.

4.5. Summary

As forensic laboratories share their validation summaries and data used for making decisions to enable future independent review of their work, the field can be strengthened. Tables with sample details and LR values have been made available as supplemental files in some publications (e.g., Bright et al. 2019, Rodriguez et al. 2019, Riman et al. 2021, Duke et al. 2022). When aggregate graphs are provided in publications (e.g., Taylor 2014) or validation summaries do not include useful metadata for the data points displayed, an independent reviewer cannot assess or correlate the data and samples used to generate them.

There are many sources of uncertainty to consider when examining DNA mixture interpretation. Presence of multiple sources of uncertainty, by itself, does not decrease reliability of strength-of-evidence assessments. If the sources of uncertainty are acknowledged and appropriately modeled, the resulting LR statements are expected to be well-calibrated. If all (or almost all, in practice) of the discriminating (between H_1 and H_2) information present in the sample has been used in the LR assessment, then the PGS system is expected to have good discrimination power. Regardless of sources of uncertainty and complexity of the samples, reliability of a PGS system boils down to checking its calibration accuracy and discriminating power. A limitation to any reliability assessment is going to be the amount of casework-similar empirical data that is available for comparison in each specific case.

In the end, the reliability of LR values produced for DNA mixture interpretation purposes by a PGS system to address sub-source level questions may not be the primary concern of a user of this information. Rather how and when questions, those considered via activity-level propositions, may be of interest. This topic is discussed in the next chapter.

5. Chapter 5: DNA Interpretation at the Activity Level

This chapter considers foundational issues regarding the application of DNA test results in criminal investigations, particularly when small quantities of DNA are examined, and decisions are made about whether the DNA profile obtained is of value to an investigation. The literature is reviewed on mechanisms of DNA transfer, factors that affect the variability of transfer and persistence, and the potential transfer of contaminating DNA at any stage in an investigation. These published studies show it is possible to handle an item without transferring any detectable DNA to that item, that DNA may have been deposited before the crime and therefore may not be pertinent to the crime, and that DNA might be present due to indirect (secondary or tertiary) transfer. A common theme from the DNA transfer literature is that the association of a reference sample from a person of interest with a crime scene sample cannot automatically be used to infer involvement with the crime. The literature on case types dealing with transfer and methods of interpretation is reviewed, and the implications of the reviewed studies and outline strategies for dealing with questions about DNA transfer are considered. The suggested strategies are (1) to minimize contamination at all stages, not just in the laboratory; (2) to consider evidence in context, because the same findings will have different significance in different circumstances; (3) to ask and answer appropriate questions and work to ensure that stakeholders do not misunderstand that findings given source (or sub-source) propositions cannot address activity or offense propositions; (4) to use the Case Assessment and Interpretation model to identify the most probative samples and the hierarchy of propositions to identify the appropriate questions to be addressed; and (5) to separate investigation from evaluation, realizing that a likelihood ratio (LR) given sub-source propositions, which is very useful to identifying a suspect, will not address additional questions that may be of interest to the court. Thus, this chapter provides foundational information around the topic of interpreting DNA findings given activity level propositions, rather than DNA findings given sub-source level propositions as discussed in the previous chapter. Exploring findings given activity level propositions in essence seek to answer questions about how or when the DNA profile of interest ended up on a surface rather than who is the source of the DNA profile. The growing literature on DNA transfer, persistence, prevalence, and recovery (TPPR) has not been fully captured. DNA TPPR could be explored in greater detail as suggested in other recent reports (see EWG 2024, pp. 172-182; TFSC 2024, pp. 62-67).

5.1. Introduction

Every contact leaves a trace. This phrase, often associated with the pioneer French forensic scientist Edmond Locard, explains why investigators often seek support for two items having been in contact. However, one translation of what Locard said was:

> "The truth is that none can act with the intensity induced by criminal activities without leaving multiple traces of his path" (cited in Roux et al. 2015).

With this, one can see that the aphorism, *every contact leaves a trace*, is a simplification. Locard's statement implies at least two things. First, the trace is not only associated with the fact of contact, but also with an activity of greater or lesser intensity. Second, multiple traces of the activity can be expected, and therefore in most cases it would be inadequate for report users or the court to consider only a single trace in isolation.

Furthermore, to the extent that every contact does leave a trace, one needs a way to separate the relevant traces—those associated with the commission of the crime—from the irrelevant ones. Prior to the development of highly-sensitive methods, separating the relevant biological traces from the irrelevant presented less of a challenge because relatively large amounts of DNA were needed to produce a profile. For samples containing a large amount of DNA (e.g., a bloodstain the size of a coin), common sense was often sufficient for determining relevance. For example, with a visible blood or semen stain, the cell type could be determined, and the activity that caused a sample to be deposited could often be inferred, even by nonexperts.

That situation changed with the advent of methods that can detect very small quantities of DNA. The 1997 *Nature* publication "DNA Fingerprints from Fingerprints" (van Oorschot & Jones 1997) demonstrated that DNA could be recovered from touched samples, which typically do not leave visible residue and may not have an easily identifiable cell type. In addition, DNA can readily transfer under some circumstances (e.g., Szkuta et al. 2017b) and can persist for fairly long periods of time (e.g., van Oorschot et al. 2014a). A summary of the above papers indicates that the significance to the crime of samples that containing small quantities of DNA to the crime is often difficult to discern. Issues involving interpreting small quantities of DNA have been a subject of discussion in the scientific literature for at least the past two decades (e.g., Evett et al. 2002). When the DNA trace in question is of low level, poor quality, partial profile, or some combination of those attributes, a formal evaluation of such evidence should be done given appropriate activity level propositions (see Table 2.5).

Forensic science typically involves investigating multiple pieces of evidence to shed light on a past event that has taken place at a particular moment in time. Figure 5.1 illustrates the opportunities for the transfer of DNA at various stages before, during, and after a crime event. The potential for DNA transfers before the crime event mean that DNA found at a crime scene may be irrelevant to the crime, and, furthermore, that the DNA present may be in the form of a DNA mixture, which further complicates the process of interpretation.

To properly assess the relevance of a DNA sample to a crime event, it is necessary to understand the factors that affect the transfer of DNA and how long it persists in different circumstances. This chapter reviews the literature on this subject, which is commonly referred

to as DNA transfer, persistence, prevalence, and recovery (DNA-TPPR) (van Oorschot et al. 2019).

KEY TAKEAWAY #5.1: DNA can be transferred from one surface or person to another, and this can potentially occur multiple times. Therefore, the relevance of the DNA to the crime being investigated should be considered when evaluating the evidence.

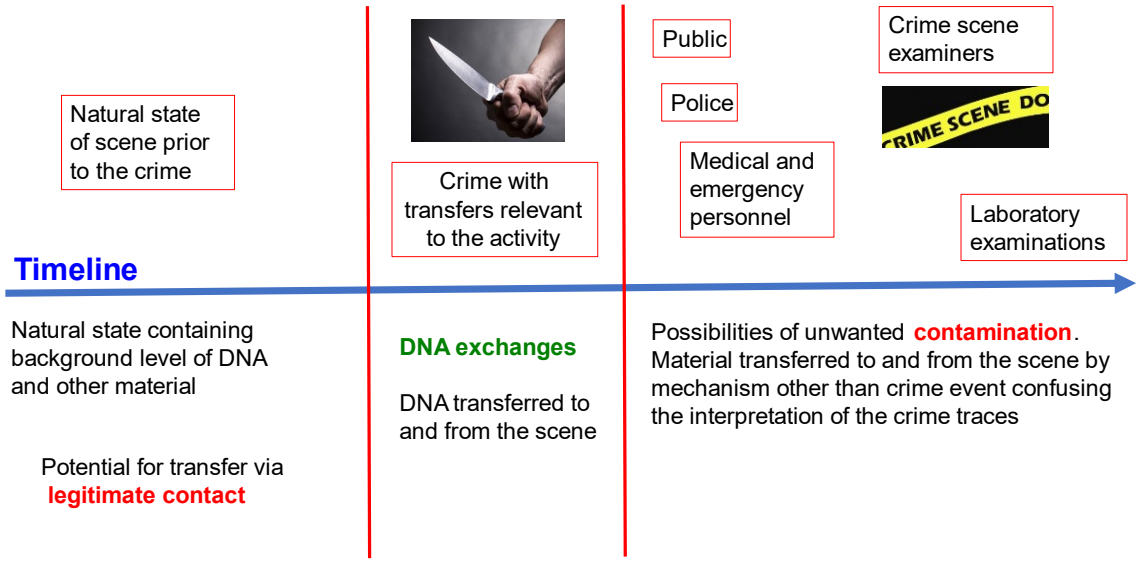

Figure 5.1. Timeline illustrating the potential for transfer via legitimate contact before the crime activity DNA exchange and the possibility of contamination after the crime event (adapted from Gill 2002).

5.1.1. An Example

Two hypothetical case scenarios are considered as an illustration of the importance of case context. Each scenario contains the same finding of a knife on which a three-person DNA mixture is detected. A reference sample from the person-of-interest (i.e., the individual suspected of stabbing) is also provided, and it is associated with the mixture profile found on the knife.

Case Scenario A

This case involves a stabbing in a private home. During a burglary attempt, the burglar is disturbed by the house home owner and grabs a knife from the kitchen, which he uses to stab the house home owner. In this case, the significance of DNA results on the knife handle may be obvious to a DNA analyst when considering what part of this evidence to sample.

Case Scenario B

This case involves a knife fight in a hotel kitchen. During this fight, a chef is thought to have been stabbed by a coworker. The knife is later recovered in a nearby alley. In this case, the relevance of DNA results on the knife handle may not be obvious.

Expectations and Risks: There is a lower risk of using an assigned LR value considering sub-source propositions in isolation with Case A than with Case B. In Case A, the burglar had no previous access to the house, and therefore finding an association with the knife would be probative. In Case B, the same finding needs more investigation before the relevance of a DNA result can be assessed. If a suspect in the hotel kitchen stabbing case had prior access to the knife as part of their job, then there is some expectation that a profile matching them would be detected on the knife handle before the stabbing occurred. A DNA mixture is a likely finding in either case scenario presented.

Considering Possible Contamination: To reduce the risk that a profile arose from contamination, additional scene samples could be taken, particularly from areas expected to be handled by the assailant. If that same DNA appears in multiple evidence items, contamination would be a less likely source (Jackson 2013, NRC 1996).

Ask and Answer the Right Questions: In both cases, the real question being sought from the DNA finding is whether the POI transferred their DNA to the knife handle while stabbing the victim. In other words, the important question is at the activity level. In Case A, if the POI's profile can be associated with the knife, then one might infer that the transfer happened during the stabbing (though the possibility of contamination must be considered). Therefore, the elevation of sub-source questions to activity questions (i.e., moving from assigned LR values with sub-source propositions to ones with activity propositions) is low risk. However, in Case B, the elevation of sub-source to activity level is of higher risk. Reporting an association between the POI and the knife, where there is the possibility of the POI's profile being present prior to the crime, cannot be taken to indicate that it was transferred there at the time of the stabbing. DNA transfer studies have demonstrated that the last person to handle an item may not be the major profile in a mixture (e.g., Taylor et al. 2016). In this example, there are at least three contributors to the mixture, so further investigation is necessary. During testimony, users of case reports and triers of fact should be made aware that an assigned LR value addressing a sub-source level question is not sufficient evidence that the POI transferred their DNA to the knife at the time of the stabbing because, at least in Case Scenario B, the POI had legitimate access to the knife and could have handled it prior to the crime (see EWG 2024, pp. 172-182).

Using a Case Assessment and Interpretation (CAI) Approach: For Case Scenario B, a preassessment of the case might prompt questions as to when the knife in question was last used and a decision on whether an assigned LR value with sub-source propositions would be helpful. Also, additional samples may be requested to get a fuller picture of the shedding characteristics of the POI to help assess whether DNA from regular use would be expected. This would depend on when the knife was last used and assumptions about how long it was handled during the knife fight stabbing incident. There may not be sufficient data available, in which case the findings would be neutral. At a minimum, the risk of misleading information based on LR values assigned with sub-source propositions alone must be emphasized to report users and the trier of fact when testifying.

5.2. Data Sources Used

The information in this chapter is based on peer-reviewed literature, most of which was found via multiple searches of the PubMed database. A search for "trace DNA" conducted on October 4, 2018, found 4085 papers. Most of the references from this search were not related to forensic DNA applications. Those relevant to small quantities of DNA for use in criminal investigations were retained. Further PubMed searches for "transfer, mixture DNA" in October 2018 located 270 articles, which were checked for relevance. Additional studies were found cited in the reference lists from three review articles that preceded this study (Wickenheiser 2002, Meakin & Jamieson 2013, Gill et al. 2015) and several additional reviews that were published during the course of this study (Taylor et al. 2018, Burrill et al. 2019, van Oorschot et al. 2019, Gosch & Courts 2019) and at least one other during the public comment period on our initial draft report (e.g., van Oorschot et al. 2021).

We divided the topics presented in the collected literature into several subject areas, as shown in Table 5.1.

Table 5.1. Subject areas examined as part of this review.

Subject Area	Number of Articles Reviewed [a]	Comments
Mechanisms of DNA transfer	16	Studies on how DNA transfers
Structured experiments to examine key variables affecting DNA transfer	40	Includes overlap with persistence
Studies on DNA transfer that mimic casework scenarios	19	Relevant to transfer and not covered in structured studies or casework section
Studies on contamination	26	Mainly studies to identify sources
Interpretation and evaluation	28	Papers particularly relevant to the issues outlined in this chapter
Casework reports	21	Useful information collated or taken from individual cases

[a] Articles were categorized according to the main message of the paper, but there is overlap, particularly among transfer, contamination, and casework. Additional sources, such as textbooks or classic references, are cited throughout the text and listed at the end of the chapter.

5.2.1. Obstacles to Comparing Data Across Studies

The existing studies are difficult to compare with each other for various reasons. For instance, different laboratories use different analytical parameters, which result in different strengths of evidence. In addition, over the past 20 years, the sensitivity of detection has increased, and aspects of how DNA profiles are produced have changed (see Chapter 2 and Appendix 1). These changes mean that a study from 2000 is not directly comparable with one from 2019.

For example, researchers may measure the efficiency of transfer based on the percentage of alleles detected, but there are subtle differences in this approach depending on how homozygous and shared alleles are counted. In other studies, only unique alleles are used to assess transfer efficiency. Sometimes this is not an issue because, unlike in a crime scene scenario, the contributors' profiles are known in a controlled research study. However, if a study records the criteria used in casework to assess their findings (e.g., Breathnach et al. 2016), a different set of criteria in another laboratory may make it difficult to compare results across studies.

DNA transfer studies have also increased in complexity. Many now consider multiple transfers and, as in real casework, consider profiles from very small quantities of DNA. This has prompted the use of probabilistic genotyping software (PGS) LR assignments rather than allele counting, which adds to the difficulty in comparing results across studies.

5.3. Reviewing the Data

The contents of the reviewed publications provided an overall view of the current state of knowledge. A great deal remains unknown about the mechanisms of primary DNA transfer, about the factors that affect secondary and higher-order transfers (transfer via one or more intermediaries, which can be animate or inanimate), and persistence. Strategies to improve research by gathering more systematic data have been suggested (Gosch & Courts 2019).

5.3.1. Mechanisms of DNA Transfer

Although there is widespread acceptance in the literature and in practice that DNA transfers, there is relatively little research on the actual mechanisms of transfer.

The term *touch DNA* is frequently used, but there is a lack of clarity about the underlying processes that allow the recovery of DNA when an item is handled. The most common view is that DNA originates from skin cells shed during the action of touching (Hanson et al. 2011). There is disagreement on this view, however, because the outer skin cells have no nuclei and therefore are not expected to contain nuclear DNA. There are alternative theories, but the number of studies as seen in the following paragraphs is limited.

Attempts to identify cell types via RNA analysis have been carried out in conjunction with nuclear DNA studies. A group of 22 collaborating laboratories carried out simultaneous extraction of RNA and DNA to identify the tissue source of the DNA and had some success with skin markers (Haas et al. 2015). Five messenger RNA (mRNA) markers were identified that demonstrated a high degree of specificity for skin. The use of these markers has enabled the detection and identification of skin using as little as approximately 5 pg to 25 pg of input total RNA from skin and, significantly, in swabs of human skin and various touched objects (Hanson et al. 2012). These researchers acknowledge that if touch DNA consisted of naked DNA in body secretions such as sweat or sebaceous fluid, skin-specific mRNA markers may be present at a concentration too low to be currently detected.

Several mRNA markers were used to determine whether different epidermal layers could help identify the type of activity, such as a firm grip or a casual touch, that gave rise to a transfer (Bhoelai et al. 2013). The study did not establish any relationship with the type of contact.

Because of the possibility that DNA may be transferred either in sweat or sebaceous fluid, there is a question as to whether touch-related DNA profiles come from extranuclear DNA rather than nuclear DNA in shed skin cells (Quinones & Daniel 2012, Zoppis et al. 2014). Testing of sweat collected from volunteers yielded an average of 11.5 ng of DNA from 1 mL cell-free sweat samples. This observation prompted the proposition that DNA transferred through the act of touching consists of cell-free nucleic acids of a length suited for STR analysis (Quinones & Daniel 2012). Another study suggested that DNA fragments on touched objects may originate from the epidermal cells of the cornified layer that are constantly sloughed off and are removed from the skin surface by sweat (Kita et al. 2008).

A morphological study using microscopy and immunology reported the following: “When swabs from touch samples were analyzed, using imaging and flow cytometry, 84–100% of DNA

detected was extracellular" (Stanciu et al. 2015). These experiments involved volunteers who held objects, with some having been asked to wash their hands prior to handling the objects. Hand washing resulted in a decrease in the amount of extracellular DNA but did not have a significant impact on the number of epidermal cells detected. The flow cytometry experiments showed two distinct fractions—fully differentiated keratinocytes (i.e., corneocytes) and cellular debris/fragments. Buccal cells were not observed, indicating saliva was not a significant source of the DNA found on subjects' hands (Stanciu et al. 2015).

It has been postulated that DNA in touch samples is transferred in the sebaceous fluid (Zoppis et al. 2014). These studies found that the ability to shed sebaceous fluid had a major influence on secondary transfer, which supports the view that dividing participants into "good" and "bad" shedders (see section 5.3.2.1) is too simplistic. Instead, the ability to shed sebaceous fluid will vary with age, hormonal condition, skin diseases, and the part of the skin that touched an object (e.g., Kamphausen et al. 2012). The relative tendency of fingertips or palms to produce DNA was examined with the view that the tips were the better source (Oleiwi et al. 2015). This study supports the claim that palms have relatively fewer sebaceous pores (Zoppis et al. 2014).

Some work has focused on the potential loss of DNA during extraction, with the possibility that touch samples may benefit from improved extraction methods (Vandewoestyne et al. 2013). It has been noted that a better understanding of the mechanism for DNA transfer will "increase our confidence in assigning a weight to DNA evidence obtained in such circumstances" (Quinones & Daniel 2012).

Researchers studying glass slides touched by donors have commented:

> "The underlying science of touch DNA recovered from criminal casework is directly related to the basic biology and genetics of normal skin regeneration and programmed cell death (apoptosis) and lends an understanding of the inherent variability in DNA recovery from handled items" (Hazell-Smithen et al. 2014).

This perspective is supported by an alternative method of sample collection involving searching surfaces for clumps of cells (Hanson & Ballantyne 2013, Farash et al. 2015, Farash et al. 2018). The approach of physically separating cells on a surface (see Chapter 6) has the advantage of being able to generate single-source DNA profiles and thus avoid the complex mixtures that arise when swabbing a surface containing cellular deposits from multiple individuals.

While the number of studies is low, it seems that the current sensitivity in DNA testing is sufficient to generate a profile from cornified-layer cells (Kita et al. 2008) that still contain DNA. The cornified layer and apoptosis may account for the possibility of additional alleles from degraded cells. Almost every transfer study discussed in this chapter has unexpected additional alleles that would support the possibility of cell-free DNA being present.

Figure 5.2 illustrates potential sources of DNA from touch evidence sample deposits taken from a 2019 comprehensive review on the topic, where the authors state: "Our current understanding of the cellular content of touch deposits and the origin of the potential trace DNA therein is extremely limited" (Burrill et al. 2019).

Additional work from these researchers "raise[s] questions about shed corneocyte DNA content previously assumed to be negligible" (Burrill et al. 2020).

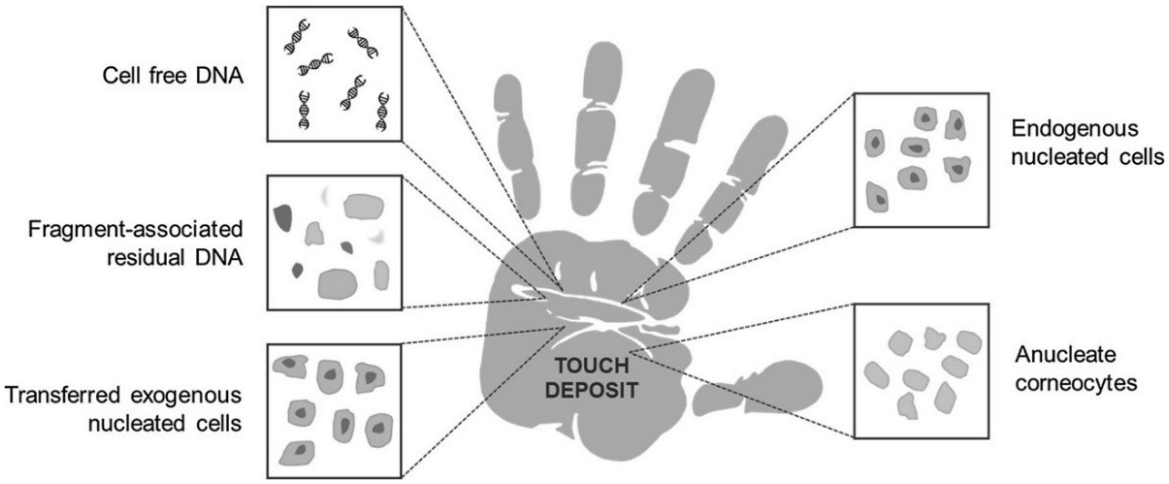

Figure 5.2. Concept map of potential sources of DNA deposited by touch/handling. It is currently well established that individuals may leave behind detectable DNA when they handle items, but the anatomical origin of that DNA remains unsolved. It is possible that the DNA typically recovered from handled items in forensic scenarios comes from nucleated cells from hands, anucleate cells from hands, nucleated cells transferred onto hands from elsewhere, residual cell fragments (including free nuclei) from hands, or from outside a cellular architecture in sweat on hands or residual transferred body fluids. Reproduced with permission from Burrill et al. (2019).

5.3.2. Structured Experiments to Examine Key Variables Affecting DNA Transfer

Several studies have been conducted to assess factors that affect the transfer and persistence of DNA. This transfer may occur via blood or saliva or small quantities of DNA of an unknown cell type. Available studies can be divided into two broad categories: (1) systematic studies that examine variables affecting the transfer and persistence of DNA, and (2) studies carried out to address specific case-like situations.

Table 5.2 provides details on structured experiments that examined key variables for the transfer and persistence of DNA. The purpose of each study and key findings have been summarized. Comparison of findings across these studies is difficult because the criteria and the methods used to measure transfer have evolved (e.g., different STR kits and PCR conditions).

A number of studies covered the following four topics, which are discussed in more detail below. The first topic involves **shedder status**, in which experiments are conducted to assess whether an individual sheds low or high amounts of DNA. The second topic involves **substrate effects**, in which experiments examine how DNA transfer is affected by the surface on which the sample is deposited. The third topic involves **persistence studies**, which examine the length of time DNA can be detected on a surface following deposition. The fourth topic involves studies concerning **non-self-DNA on individuals**, in which experiments are conducted looking for DNA not associated with the individual who touched an item.

Other variables that affected DNA transfer in these studies included moisture (Goray et al. 2010a, Lehmann et al. 2013, Verdon et al. 2013), pressure (Tobias et al. 2017), and friction (Verdon et al. 2013).

Table 5.2. Studies published from 1997 to 2018 involving structured experiments to examine key variables for the transfer and persistence of DNA.

No.	Reference and Title	Size of Study and Measurement Criteria	Purpose of Study	Key Results
1	van Oorschot and Jones (1997) DNA fingerprints from fingerprints	Various tests with 1 to 4 repeats Profiles: 2 ng to 150 ng DNA	Examine whether a DNA profile can be detected from items participants touch	Profiles generated 13/13; secondary transfer noted
2	Lowe et al. (2002) The propensity of individuals to deposit DNA and secondary transfer of low-level DNA from individuals to inert surfaces	8 participants, 3 time intervals repeated 5 times; 22 participants, one time interval repeated 3 times; 2 pairs, 3 time intervals, 5 replicates % profiles obtained	Study secondary transfer of DNA when body fluid is not known	Secondary transfer is possible; participants differ in their propensity to deposit DNA; time since handwashing is a key variable
3	Phipps and Petricevic (2007) The tendency of individuals to transfer DNA to handled items	60 participants with 5 volunteers chosen to test good shedder/ bad shedder theory; tested over four days Number of alleles obtained with relative proportion of full profiles, partial profiles, and no results	Check transfer of DNA and repeat Lowe et al. (2002) study	Handwashing is not a key factor as an individual cannot be relied upon to shed a consistent amount of DNA over time; results indicated that it may be more difficult than expected to classify individuals as *good* or *bad* shedders
4	Farmen et al. (2008) Assessment of individual shedder status and implications for secondary DNA transfer	9 participants tested with palms swabbed at 2 time intervals; handshakes followed more swabbing and holding a beaker Number of matching alleles	Assess shedder status and check effect on secondary transfer	Shedder categorization confirmed with a good shedder picked up on other participants' hands and objects; transfer noted on all occasions in this study
5	Goray et al. (2010a) Secondary DNA transfer of biological substances under varying test conditions	DNA 5 µL/mL, blood and, saliva on wool, cotton, and plastic using passive, pressure, and friction; each combination replicated four times % DNA transferred	Factors affecting secondary transfer; deposit including moisture level, the primary and secondary substrate, and type of contact	Initial deposit of DNA was 20 times greater when deposited onto porous cotton surface than onto a smooth and hard plastic surface, with less in reverse; nature of substrate and moisture were significant; other biological materials were the same

No.	Reference and Title	Size of Study and Measurement Criteria	Purpose of Study	Key Results
6	Goray et al. (2010b) Investigation of secondary DNA transfer of skin cells under controlled test conditions	One donor produced DNA skin cells; 6 times for each variable; 1 and 2 substrate, passive, pressure, and friction % DNA transferred; initial amounts of DNA needed to transfer to generate good profile (1 ng at that time) measured; results varied with conditions from 385 ng to 2 ng	Study of factors affecting secondary transfer of skin cells	Freshness of deposit not a factor; friction increased rate of transfer; skin cells deposited onto nonporous substrate transfer more readily but further transfers facilitated more by porous substrate. Nonporous to porous with friction most effective
7	Daly et al. (2012) The transfer of touch DNA from hands to glass, fabric, and wood	300 participants, 50/50 male /female held in their fist for 60 s; no distinction made between dominant or no-dominant hand Gene scanner and gene mapper 50 relative fluorescence units (RFU) for heterozygous and 200 for homozygous	Check the variation onto glass, wood, and cloth.	9% for glass samples, 23% for fabric, and 36% for wood; no difference between males and females; 22% classified as shedders; secondary transfer inferred by number of alleles
8	Lehmann et al. (2013) Following the transfer of DNA: How far can it go?	4 replicates of six transfers of wet and dry blood and touch DNA on cotton and glass Average % DNA transferred as Goray (2010a)	Measure the detectability of DNA following multiple transfers	Wet blood detected up to 4 transfers on cotton and 6 on glass; dry blood detected up to 2 on cotton and 6 on glass; DNA detected on first transfer on cotton and second on glass
9	Verdon et al. (2013) The influence of substrate on DNA transfer and extraction efficiency	6 fabrics as substrates, three nonporous substrates; wet and dry blood; passive and friction; 4 replicates % transfer DNA	Influence of nine substrate types on DNA transfer involving blood	High transfer when primary substrate nonporous and secondary porous; extraction most efficient from nonporous; friction and wet give best transfer
10	Poetsch et al. (2013) Influence of an individual's age on the amount and interpretability of DNA left on touched items	213 individuals at different stages of life Total DNA amount and allele counts	Effect of age on transfer	Amount of DNA of children and older participants could be distinguished
11	van Oorschot et al. (2014a) DNA transfer: The role of temperature and drying time	4 replicates of four temperatures in 13 time conditions % DNA transfer flaking blood from nonporous surfaces may affect yields	Time to dry biological fluids and effect on transfer	Exponential decay rates regardless of temperature; blood dries fairly quickly; transfer of DNA very dependent on dryness of sample, so timing since deposit needs to be considered

No.	Reference and Title	Size of Study and Measurement Criteria	Purpose of Study	Key Results
12	van Oorschot et al. (2014b) Persistence of DNA deposited by the original user on objects after subsequent use by a second person	54 pens and 88 nylon/polyester elastic bands "used" by one donor and given to second users; 46 items solely used by one individual given to a second user Relative % contribution of each participant using relative RFU contributions at each locus; where alleles were shared, RFU portion determined using RFU of other alleles at that locus	Check the persistence of DNA following prior use by an individual	% contribution of first user decreases in a linear manner with time; depends on substrate; hard porous surface loses first person's DNA quicker than soft porous item; unknown source alleles detected
13	Gršković et al. (2014) Impact of donor age, gender, and handling time on the DNA concentration left on different surfaces	60 participants touched 9 items; 540 samples Amounts only; no profiling carried out	Test correlation between donor age, gender, and handling time and trace DNA amount recovered on paper, plastic, and plastic-coated metal surfaces	Item texture, donor age, and gender influence trace DNA concentration; independent of handling time
14	Davies et al. (2015) Assessing primary, secondary, and tertiary DNA transfer using the Promega ESI-17 Fast PCR chemistry	Couples gripping plastic tubes, directly or following handshakes for 30 s % of unambiguous alleles actually detected compared to those available for detection; summing of the peak heights of all of the detected unambiguous alleles matching the expected donor, divided by the number of alleles expected	Measure the levels of DNA transfer from direct, secondary, and tertiary transfer	Variable nature of primary transfer; occasional secondary transfer greater than primary; even in primary transfer, nondonor alleles were detected; suggestion that there was a limit for template?

No.	Reference and Title	Size of Study and Measurement Criteria	Purpose of Study	Key Results
15	Lehmann et al. (2015) Following the transfer of DNA: How does the presence of background DNA affect the transfer and detection of a target source of DNA?	DNA, wet and dried blood used as substrate on glass and cotton; one donor as background on first set of six and different donors as background on second set of six; target DNA added to first substrate before transfers; replicated by 4 % DNA as noted by unique alleles	Transfer and detection in the presence of background	Presence of background DNA influenced the transfer of DNA differently depending on the combination of biological material and surface type; detection decreased after multiple contacts due to decreased DNA and complexity of mixtures
16	Fonneløp et al. (2015a) Secondary and subsequent DNA transfer during criminal investigation	3 donors deemed to be good shedders; 30 transfer chains; 11 repeats for wood, 9 for plastic, and 10 for metal Quantity of total DNA in ng and % DNA transferred as assessed by number of alleles above 200 RFUs	Primary transfer to wood, plastic, and metal and secondary transfer via nitrile gloves; onto fabric and paper	DNA can be transferred onto a third substrate via nitrile gloves in 5 out of 30 transfer chains
17	Fonneløp et al. (2015b) Persistence and secondary transfer of DNA from previous users of equipment	4 participants: 2 male, 2 female Alleles present to include; person could not be excluded or contributors cannot be detected	Study of persistence of DNA from previous user to new user's hand	Initial user alleles detectable up to 8 days after receiving the equipment
18	Goray and van Oorschot (2015) The complexities of DNA transfer during a social setting	3 participants repeated 5 times STRmix, to record exclusion, not excluded, and no. persons in the mixture	Study transfers with group having a drink together	DNA can be detected without actual contact between individuals; DNA of unknown source can be transferred from hands
19	Oldoni et al. (2015) Exploring the relative DNA contribution of first and second object's users on mock touch DNA mixtures	14 volunteers acting as first or second handlers of 5 plastic, 2 metal, 1 fabric and inside and outside of nitrile gloves giving 231 mixtures Relative peak height (50 RFU) of the two contributors with markers showing no allele sharing; average profile contribution was calculated over several samples paired	To gain knowledge on the relative contribution of DNA left behind by different users over time	Second handler contribution increased from 21% to 73% between 5 and 120 min; unexpected full profiles detected in 7 simulations suggesting indirect transfer

No.	Reference and Title	Size of Study and Measurement Criteria	Purpose of Study	Key Results
20	Meakin et al. (2015) The deposition and persistence of indirectly transferred DNA on regularly used knives	4 volunteers paired, experiment in triplicate repeated for 5 weeks at 1 hour, 1 day, and 1 week time intervals % profiles on the basis of unique alleles, RFU 100; total amount of DNA	Whether transferred DNA could be detected on regularly used items	DNA of person who shook hands with knife handler; regular user could be detected in 10:1 ratio, but alleles were detected for up to 1 week; unexpected alleles also detected, suggesting indirect transfer
21	Montpetit and O'Donnell (2015) An optimized procedure for obtaining DNA from fired and unfired ammunition	10 volunteers carried half their ammunition for 2 days before loading weapons, and the other half was loaded directly; each shooter loaded half of their cartridges into a magazine and tested unfired cartridges; the other half were fired and analyzed Quantities of DNA and reportable alleles recorded and interpretable profiles as judged by fixed criteria	Study to optimize collection and profiling of DNA from fired and unfired ammunition	Less than 50 pg on 78% (607 of 800), 27% (229 of 785); 40% had mixtures or indication that more than loader's genotype detected; available information is human handling at manufacture stage less than 1%
22	Oldoni et al. (2016) Shedding light on the relative DNA contribution of two persons handling the same object	Fourteen persons acting in pairs as first and second user handled a range of everyday items in three time simulations Alleles over 50 RFU counted so long as they appeared in 2 amps; % contribution calculated.	To understand the relative proportion of DNA deposited by different persons through time	Contribution from second user increased in time and became the major profile in many instances after 120 min; indirectly transferred DNA in 8/234 cases; a full profile in one case; evidence of shedder status; porous and nonporous effects
23	Samie et al. (2016) Stabbing simulations and DNA transfer	4 donors, 16 experiments, and 64 traces 30 RFU; allelic count and STRmix;70% more than 6 loci considered full profile	Study transfer of DNA from handler and check if handlers would transfer DNA from persons closely connected to them	DNA of person handling the knife in 83% of cases; person nearby not detected; 2, 3 and 4 person mixtures
24	Cale et al. (2016) Could secondary DNA transfer falsely place someone at the scene of the crime?	12 participants using 24 knives Quantity of DNA and allelic counts 50 RFU	Detection of interpretable secondary DNA profiles	After 2 min handshake, secondary DNA transfer was detected in 85% of the samples; in five samples, secondary contributor was major or only contributor

No.	Reference and Title	Size of Study and Measurement Criteria	Purpose of Study	Key Results
25	Goray et al. (2016) Shedder status—An analysis of self- and non-self-DNA in multiple handprints deposited by the same individuals over time	240 handprints from 10 individuals; self and nonself DNA determined Deposits varied 0.05 to 5 ng; total DNA; total alleles per locus; STRmix using depositor and staff elimination database; evaluation of mixture proportions	Determine if individuals deposit consistent quantities of their own DNA as well as variability	Some individuals shed more readily than others, but there is a lot of variation; nonself, usually as minor component in 79% of samples; depositor excluded from deposit in 7 samples; good shedders had less nonself DNA; total amount of DNA independent of ratio of self to nonself
26	Buckingham et al. (2016) The origin of unknown source DNA from touched objects	4 participants; seven tests % unique alleles and unique alleles of other participants; total adjusted peak height used to get % contribution DNA	Test whether the last person to handle an item can be detected in the DNA profile produced from that item	Nonself DNA common on a person's hands; material deposited and retrieved from an object is dependent on who touches what, how, and when; evidence of the prevalence and complexity of nonself DNA in its deposit and transfer
27	Helmus et al. (2016) DNA transfer—a never-ending story; a study on scenarios involving a second person as carrier	3 pairs, each participant acted as donor, giving 6 implementations per scenario of participants repeated twice Allele counting at each locus >50 RFU; classified as complete if each allele present without additional peaks or if 5 or more regardless of additional deemed to be partial; <5 alleles regarded as no profile	Study of second person as a carrier	DNA transfers from donor to cotton to plastic or cotton via second person 40% of 180 samples; cotton much more receptive than plastic; effect of gloves not as strong as expected
28	Manoli et al. (2016) Sex-specific age association with primary DNA transfer	128 individuals, experiment in triplicate, 768 swabs % alleles	Effect of age and sex on transfer of DNA; also test if shedder status remained constant in 1 and 2 transfers	DNA not always transferred; claim primary and secondary can be distinguished; 77% of participants changed shedder status through the trials; young males more likely to transfer than older males; actual results suggest females poorer shedders but not claimed by authors

No.	Reference and Title	Size of Study and Measurement Criteria	Purpose of Study	Key Results
29	Lacerenza et al. (2016) A molecular exploration of human DNA/RNA co-extracted from the palmar surface of the hands and fingers (PHF)	Samples collected from 30 males and 30 females Peak height, 50 RFU; 16 tissue markers for mucosa; saliva; semen; vaginal mucosa; menstrual secretions; and skin	Study to explore source of transferred DNA using DNA/RNA; levels of foreign material on hand surfaces of the general population	Nonskin cellular material observed in 15% of PHF; amount of DNA from these samples higher than skin cells only; donor alleles 75% in males and 60% in females; 30% females had mixtures with a component of 20% or more and 8% males had such mixtures
30	van den Berge et al. (2016) Prevalence of human cell material: DNA and RNA profiling of public and private objects and after-activity scenarios	549 samples, four categories: public (105); private; transfer-related; and washing machine samples RNA and DNA co-extracted; in-house multiplex used for RNA; known genotypes used with in-house software to assess contribution to mixtures; maximum allelic counts used to determine the minimum number of contributors	Gain understanding of cell material on surfaces contributing to background traces; DNA mRNA on various items	High DNA not related to increased number of contributors; major DNA on an individual may not be owner; in activity situations, perpetrator not always the major
31	Voskoboinik et al. (2017) Laundry in a washing machine as a mediator of secondary and tertiary DNA transfer	Eight new unworn socks - various cotton blends washed with typical laundry of four households - various washing conditions; six new unworn socks and a T-shirt laundered without additional items; 15 washing machine drums swabbed Amount of DNA and allele calls; 60 RFU detection threshold, 200 RFU stochastic threshold	Check the possibility of secondary and tertiary DNA transfer during laundry washing of worn and unworn garments in household and public washing machines	Secondary transfer detected in 22% of cases; tertiary transfer experiments indicated that the possibility of DNA transfer between separate washing cycles via the deposition of biological material in a washing or drying machine's drum is unlikely

No.	Reference and Title	Size of Study and Measurement Criteria	Purpose of Study	Key Results
32	Fonneløp et al. (2017) The implications of shedder status and background DNA on direct and secondary transfer in an attack scenario	20 participants, 60 experiments with test tubes; 17 simulated attacks with four samples from each Quantity of DNA; mixture interpretation according to International Society for Forensic Genetics (ISFG) guidelines; three-person mixture considered if major profile	Shedder status and effect of background DNA; simulated attacks	No aerosol transfer from talking; DNA transferred in attacks (16/17); background DNA from the environment can be confused with crime samples (1/148)
33	Szkuta et al. (2017a) Transfer and persistence of DNA on the hands and the influence of activities performed	Volunteers paired on 12 occasions; each of 24 participants acted as depositor or known contributor LR for POI using STRmix; up to 4 participants analyzed with caution LR of 100 billion reported rather than exact number	Whether nonself DNA transferred via handshake could be detected on surfaces and what effect activities had	Depositor of handprint main depositor; minor contributions from handshaker decreasing with the number of handshakes; main depositor excluded on several occasions; concept of “parking,” i.e., retransfer of DNA on used items
34	Meakin et al. (2017) Trace DNA evidence dynamics: An investigation into the deposition and persistence of directly and indirectly transferred DNA on regularly used knives	4 volunteers carrying out experiments on three separate weeks at 1 hour, 1 day; and 1-week intervals; 36 knives for examination in total Total DNA amount; peak heights and % unique alleles as well as RMP and LR using LRmix 2.0	To study directly and indirectly transferred DNA on regularly used knives; extension of 2015 study	When dealing with items already having a DNA load, it may be possible to use intrinsic qualities of profiles to distinguish between directly and indirectly transferred DNA
35	Tobias et al. (2017) The effect of pressure on DNA deposition by touch	2 participants, 36 samples Quantity of DNA and % profiles based on alleles	Test whether pressure affects the amount and the quality of DNA transferred by touch	Increase in pressure resulted in an increase in DNA from both donor and unknown sources; no difference between participants at 4 kPa but variation noted at 21 kPa and 37 kPa

No.	Reference and Title	Size of Study and Measurement Criteria	Purpose of Study	Key Results
36	Pfeifer and Wiegand (2017) Persistence of touch DNA on burglary-related tools	3 types of tools with and without gloves; 234 samples in total Completeness of profiles based on unique alleles; casework approach to reporting for German database entries, 4/8 of some markers together with 7/13 of another set; statistical comparisons conducted using GraphPad Prism	Explore the persistence of DNA on tool handles when more than one person touched them; different types of tools tested with and without gloves; experiments carried out to get data to address activity propositions in case of mixed profile on a screwdriver	Owner detected in 47% of cases before burglary and in 1/30 cases after mock burglary and never as major; more moderate action gives possible match to first or second user; 30% tools from households have reportable profile of owner; 57% have mixture that cannot be resolved; amounts varied in manner that did not help; one case of second user even though wearing gloves; nature of contact, substrate, and user characteristics variables
37	Bowman et al. (2018) Detection of offender DNA following skin-to-skin contact with a victim	Nine pairs tested three times; some changes resulting in total of 15 females and seven males; 266 samples collected; 72, 94, and 100 from time points 0 h, 3 h, and 24 h; skin and clothing sampled Unique alleles recorded and STRmix used for mixture interpretation	Test value of collecting DNA samples in mock assault situations	Support for H_p for 56% and 77% for medium and heavy pressure used in assault; amount of DNA falls off rapidly on skin but detectable on clothes up to 24 h; high amount of nonself alleles detected in control areas; information on shedder varying with time
38	Poetsch et al. (2018) Impact of several wearers on the persistence of DNA on clothes	4 females and 2 males wearing sweatbands for times from 10 min to days; each combination of times done with 6 different pair/trios of individuals, giving a total of 204 samples Amount of DNA and allele peaks interpreted when greater than or equal to 300 RFUs for single; allele counting at each locus >50 RFU; classified as complete if each allele present without additional peaks or if 5 or more regardless of additional deemed to be partial; <5 alleles regarded as no profile	Test how long DNA persists on an item used in daily routine and how long a piece of clothing must be worn to definitively leave detectable DNA behind	After 10 min, at least a partial profile of the second/third wearer of a piece of clothing could be demonstrated; even after the sweatband was worn for 3 days by the second wearer, the complete profile of the first wearer was still detectable in 42% of these samples

No.	Reference and Title	Size of Study and Measurement Criteria	Purpose of Study	Key Results
39	Helmus et al. (2018) Persistence of DNA on clothes after exposure to water for different time periods—a study on bathtub, pond, and river	5 participants; epithelial cells and separately a drop of blood added to clothes left in bathtub, pond, and small river for varying periods up to six months Allele peaks >50 RFU; complete profile if all alleles detected even if additional peaks; partial if more than half loci; and regarded as no profile if less than half of the evaluable loci in every allele of the individual in question was found	This study was conducted to attempt a general statement about the conditions under which sufficient DNA remains can be expected for molecular genetic analysis	Complete STR profiles could be detected even after immersion in water, dependent on conditions; longest time recorded was full profile after 2 weeks in a pond in winter
40	Ruan et al. (2018) Investigation of DNA transfer onto clothing during regular daily activities	50 participants supplied shirts, various areas sampled worn for 7–9 h and sampled again; 38 participants received 10 × 10 swatches to add to their laundry STRmix used to examine profiles produced Y-allele at the amelogenin locus in PowerPlex® 21 System	Check the transfer of DNA to clothing during regular activity; test the effect of laundering	The adventitious transfer of trace DNA means that the DNA recovered in forensic casework may not always have evidentiary relevance; freshly laundered clothes had interpretable mixtures from which uploadable foreign DNA profiles could be determined; in some cases, the donor of the clothing was not even the predominant DNA profile in the sample

5.3.2.1. Shedder Status

Shedder status refers to the greater or lesser tendency of an individual to shed DNA (Lowe et al. 2002). This is an important variable affecting transfer. There is a consensus that some people are better shedders than others (i.e., sharing a higher quantity of DNA), but there is less agreement about whether individual variation over time is of comparable magnitude. Different studies use different criteria to classify participants as good or bad shedders. Therefore, even though there is agreement that people vary, there is no universal scheme for classification.

The first article describing touch DNA results (van Oorschot & Jones 1997) noted variable amounts of DNA recovered from objects touched by different individuals, though these individuals were not formally classified as *good shedders* or *poor shedders* at that time. One of the first studies to examine shedder status specifically concluded:

> "While a good DNA shedder may leave behind a full DNA profile immediately after hand washing, poor DNA shedders may only do so when their hands have not been washed for a period of 6 hours" (Lowe et al. 2002).

Some studies have raised doubts about the ability to classify individuals as good or bad shedders (Phipps & Petricevic 2007), while other studies have confirmed that these categories can be useful (Farmen et al. 2008, Goray et al. 2016, Kanokwongnuwut et al. 2018). For example, some Australian researchers reported that,

> "while there is substantial variation in the quantities deposited by individuals on different occasions, some clear trends were evident with some individuals consistently depositing significantly more or less DNA than others" (Goray et al. 2016).

Another study carried out with 128 individuals found that shedder status varied with individuals over time in 77% of cases (Manoli et al. 2016). When age was studied, children and older participants could be distinguished (Poetsch et al. 2013, Grškovic et al. 2014). Younger males were more likely to shed than older males, though this effect was not noted in females (Manoli et al. 2016).

One study found that the amount of DNA transferred was not correlated with the length of time an item was handled (Gršković et al. 2014). Other studies investigating activities found time between activities had an impact. For example, when the deposition of a handprint was delayed, the activities performed by the individual had a substantial effect on the resultant detection of the contributing profile. In addition, multiple contacts with the same items increased the likelihood that the known contributor's DNA would be retained and subsequently detected due to the parking and retransfer of DNA on handled items (Szkuta et al. 2017b). Moisture was found to increase the amount of transfer (Goray et al. 2010a, Lehmann et al. 2013, Verdon et al. 2013).

Some studies examining secondary transfer have found that contributions from particular donors dominate, with this finding being explained by a shedder effect (Fonneløp et al. 2017, Buckingham et al. 2016, van Oorschot et al. 2014a). Other studies exploring the contribution of two and more people to the surface of an object proposed shedder status as a major factor explaining the variability in percentage contributions (Oldoni et al. 2015, Oldoni et al. 2016, Meakin et al. 2015, Goray et al. 2016).

It is more useful to think of shedder status as existing on a continuum—as opposed to there being good or bad shedders—as these studies do support the idea that some individuals routinely shed more DNA than others. The most convincing example in the literature thus far is a longitudinal study of contamination in an operational biology laboratory over a period of time (Taylor et al. 2016). In this study, DNA linked to one individual was greater in quantity and more widely distributed than DNA from a coworker with similar duties who was working nearby (Taylor et al. 2016).

It may be that a definite answer to the question of shedder status will not be possible until a better understanding of the mechanisms of DNA transfer is gained, as discussed earlier. However, the degree to which an individual sheds DNA is a variable that needs to be kept in

mind when considering the relevance of DNA in a mixture or in any situation where there is the question of how or when the DNA was deposited.

5.3.2.2. Substrate Effects

The material onto which DNA transfers (i.e., the substrate) affects how easily DNA will transfer or be retained. Researchers have examined the effect of moisture and substrate on the transfer of skin cells and noted that skin cells are deposited more readily onto porous substrates, such as cotton. However, secondary and higher-order transfers of skin cells are facilitated more by non-porous substrates, such as plastic. The most effective transfer chain was from non-porous to porous substrates with the use of friction (Goray et al. 2010b).

A study of 300 participants holding glass, cloth, and wood found the likelihood of obtaining a DNA profile was approximately 9% for glass samples (average recovery of ≈0.50 ng or ≈85 cells), 23% for fabric (average recovery of ≈1.2 ng or ≈200 cells), and 36% for wood (average recovery of ≈5.8 ng or ≈975 cells) (Daly et al. 2012). If this particular study, which was conducted with the STR kit SGM Plus using 28 cycles, was repeated with the higher-sensitivity DNA tests being routinely used today (e.g., the STR kit GlobalFiler with 29 cycles or PowerPlex Fusion with 30 cycles), then the DNA amounts detected via transfer would be expected to increase, while the relative behavior of surface types would probably remain the same.

Another study involving nine different substrates also found that the amount of DNA transfer was highly dependent on the porous or non-porous nature of a surface (Verdon et al. 2013). The finding that transfer was highest when the primary substrate was non-porous and the secondary substrate was porous is in keeping with an everyday experience of how materials behave. When transfers onto wood, plastic, and metal were considered in another study, nitrile gloves were found to be good vectors for additional transfers onto fabric and paper (Fonneløp et al. 2015a). More DNA was transferred onto the wood and plastic than onto the metal initially, but proportionally more was transferred from the metal onto the gloves. DNA was transferred in the highest concentration to plastic and plastic-coated metal, and least onto paper in a different study (Gršković et al. 2014).

Substrate effects were again noted when controlled experiments were carried out to check the persistence of DNA from a prior handler following handling by a second person:

> "The retrieval of the profile of the initial user of the object is dependent on the type of substrate and on how the object was used. When considering a hard, non-porous object, the first user's contribution to the profile drops approximately 50% immediately upon use by a second person and drops to approximately 15% after 90 minutes. When considering a worn object made of soft porous material, the first wearer's profile remained higher than that of a second wearer during the first 10 hours of wear by the second wearer, and still accounted for approximately 12% after 96 hours" (van Oorschot et al. 2014a).

Other researchers, when exploring the impact of a second user following a first user or habitual user, studied a range of materials, and though they reported that the second user became the major DNA contributor for all substrates after 120 min, they did note "extreme values" for both

non-porous plastic bracelets and porous nurse caps (Oldoni et al. 2015). Items of clothing, i.e., porous material, were used in two other studies that broadly sought to test whether wearer DNA could be identified (Breathnach et al. 2016, Magee et al. 2018).

A 2020 review explored the underlying mechanisms of metal-DNA interactions. It acknowledges how ionization and electron affinity of metals impact the degree of interaction with DNA, which is a negatively charged molecule. The proposal is that this bonding is responsible for the difficulty in recovering DNA from certain metal surfaces, and it shows that understanding these metal-DNA interactions are fundamental to improving the likelihood of getting interpretable profiles from trace samples (Bonsu et al. 2020).

5.3.2.3. Persistence Studies

For a DNA association to be relevant in a particular case, the DNA must have been deposited at the time the crime occurred (see Figure 5.1). If any cells or DNA molecules were left prior to the crime and they persist, then this *non-relevant DNA* could contribute to the crime scene evidence (e.g., possibly creating a mixture) and potentially influence the relevance of the final result. Therefore, it is important to understand the factors that affect the persistence of DNA.

DNA persistence has rarely been studied in isolation. One study, using the Profiler Plus kit with 28 cycles, detected DNA outdoors that had been deposited up to two weeks before (Raymond et al. 2009a). The sensitivity of the technology has increased since that time, so it is possible that today, profiles would be detectable for a longer period of time if they do persist. However, similar studies have not yet been undertaken with newer STR kits and CE instruments. In a study considering the persistence of primary and secondary transfer from previous users of equipment, alleles of the previous user were detected for up to eight days (Fonneløp et al. 2015b). In a study of the buildup of DNA contamination from staff members in a semi-controlled laboratory environment, DNA profiles were detected long after deposition, and in fact could be detected months later (Taylor et al. 2016).

A study on the detection of offender DNA following a simulated assault involving skin-to-skin contact showed a rapid decrease in the detection of the offender's DNA on the skin, though DNA profiles could still be detected up to 24 hours post-assault (Bowman et al. 2018). DNA could also be detected on clothing worn over the assaulted area up to 24 hours later, and the authors suggested that sampling from clothing worn over the assaulted area may be an additional or better avenue for the recovery of offender DNA post assault, when there has been a significant time lapse between assault and sampling (Bowman et al. 2018).

As will be discussed in a later section on digital penetration, there have been a number of persistence studies dealing with fingernails.

Information on persistence can also be gained from studies on the effect of a second user when the persistence of the first user is studied. The DNA of the initial user decreases with time, though in a study involving knives used by a person following a handshake, DNA from the handshaker was detectable on a knife handle for at least a week, albeit as a partial profile (Meakin et al. 2015).

5.3.2.4. Non-Self DNA on Individuals

Many of the studies summarized in Table 5.2 detected alleles or profiles that could not be accounted for by DNA from the individuals participating in the study. For example, foreign alleles were detected approximately 50% of the time, with 31% consisting of one to three alleles and 9% containing six or more (Manoli et al. 2016).

Such alleles from unknown sources have received more emphasis in various studies because of increases in DNA test sensitivity. The authors of one study, which sought to look at DNA transfers in a social setting rather than in structured experiments, reported that,

> "simple minor everyday interactions involving only a few items in some instances lead to detectable DNA being transferred among individuals and objects without them having contacted each other through secondary and further transfer. Transfer was also observed to be bi-directional. Furthermore, DNA of unknown source on hands or objects can be transferred and interfere with the interpretation of profiles generated from targeted touched surfaces" (Goray et al. 2015).

In another study, non-self DNA was detected on 79% of hands (Goray et al. 2016). Results from this study showed that in most situations, participants were majority contributors or the only source of the DNA deposited. An average of 74% of detected DNA derived from self, while the other 26% appeared to be non-self DNA. In instances involving participants that the researchers classified as *poor shedders*, non-self DNA rather than self-DNA was transferred. This was found to be the case in seven samples, 2.9% of the time (Goray et al. 2016).

A study about a new collection and extraction procedure for obtaining DNA from ammunition also provided an example of the detection of non-self-DNA (Montpetit & O'Donnell 2015). In this study, 10 volunteers handled various fired or unfired rounds of ammunition, which were then swabbed for DNA. With 97% of interpretable results, the volunteer who handled or loaded the ammunition was detected. However, non-self-DNA was detected unexpectedly: the DNA profile from a child of one of the volunteers was recovered from ammunition, but there was no opportunity for the child to touch the ammunition directly (Montpetit & O'Donnell 2015).

In a number of studies, the major profile was not always associated with the last person to handle an item (Cale et al. 2016, Buckingham et al. 2016, Goray et al. 2016). This may result from background DNA or from the handler depositing non-self DNA.

5.3.3. Studies on DNA Transfer that Mimic Casework Scenarios

5.3.3.1. Caution in Using DNA in Domestic Settings

Investigating crimes in domestic settings can be challenging. Numerous researchers have conducted experiments on transfer during clothes washing/laundering. This is important because moisture was noted as one of the factors affecting secondary transfer of biological materials and DNA (Goray et al. 2010a, Goray et al. 2010b). The potential for transfer of spermatozoa in washing machines has been accepted by forensic biologists for some time (Kafarowski et al. 1996). Later studies have also found transfer of DNA rather than spermatozoa

during washing (Brayley-Morris et al. 2015, Noël et al. 2016). Together, these washing studies suggest that finding DNA from one member of a household on another needs to be interpreted with caution. DNA from family members was detected on children's underwear even in instances where semen was not placed on the samples (Noël et al. 2016). In another study, DNA from blood of a household member was detected on laundered items, but DNA from saliva or epithelial abrasions was not detected (Kamphausen et al. 2015). A 2018 study reported that it is not uncommon for foreign DNA to transfer onto an individual's clothing during laundering and included a note of caution in relation to the investigation of crime in domestic situations (Ruan et al. 2018).

5.3.3.2. Mixtures in Sexual Assault Cases

In the early days of DNA profiling, most mixtures were from sexual assault cases where epithelial cells from the female victim were mixed with sperm and epithelial cells of the perpetrator. Although such samples can involve allele overlap and other complicating factors, sperm and epithelial cells are relatively easy to separate because sperm cells are more resistant to extraction, which allows the DNA from the two types of cells to be extracted without mixing. It is important to note that sexual assault samples may contain epithelial cells from the perpetrator (from seminal fluid, skin contact, saliva) which will be co-extracted with female epithelial cells; however, male epithelial cells are typically in the minority on swabs taken from the female victim and may not result in detectable alleles. Differential extraction (Gill et al. 1985) continues to be an important method in these types of cases.

5.3.3.3. Sexual Intercourse versus Social Contact

There are various other situations in sexual assaults where mixtures of unknown cell types are encountered. Researchers have tended to design specific experiments to address these issues, as seen below. Although the sample numbers in the experiments are limited, they do provide better information than uncalibrated experience in the absence of ground truth.

In some cases in which DNA is recovered, the trier of fact needs to assess whether the DNA transfer occurred during a sexual assault or through simple social contact. A series of experiments measured the amount of female DNA transferred to male undergarments and genitals following sexual intercourse and following non-intimate social contact that was designed to maximize transfer (Jones et al. 2016). In the experiments performed, it was not possible to replicate the high levels of DNA transferred from sexual intercourse by non-intimate contact (Jones et al. 2016). Although this study was confined to one couple carrying out the sexual intercourse experiments, the findings are in keeping with the effects of moisture on transfer seen in earlier transfer experiments (Lehmann et al. 2015).

A retrospective survey of sexual assault cases noted *positive findings* consisting of the victim's epithelial cells recovered from the attacker's penis highlighting the advantage of collecting such samples in sexual assault cases (Fonneløp et al. 2019). When such samples are examined and a female victim claims vaginal penetration, the defendant may offer an alternative explanation of secondary transfer of the victim's cells to his penis. Fourteen couples were recruited to test the

hypothesis that female DNA was more likely to be detected following intercourse than following social contact. The authors report the possibility of using their data to make a statistical model to distinguish "between samples taken after intercourse and samples taken after secondary transfer by skin contact" (Bouzga et al. 2020).

5.3.3.4. Digital Penetration

A 2015 study of digital penetration used information from Y-STR markers on vaginal swabs (McDonald et al. 2015). Conversely, earlier work focused on the possibility of obtaining DNA matching the female from under the penetrator's fingernails.

> "Full female profiles were obtained from all swabs collected at 0 and 6 hours after digital penetration, indicating that female DNA was always transferred and persisted in the short term. Furthermore, full female profiles were produced from three-quarters of samples collected after 12 hours whilst mixed profiles were produced in the majority of samples taken after 18 hours. The analysis of several variables indicated that hand washing had a significant effect on the persistence of female DNA profiles" (Flanagan & McAlister 2011).

An earlier study of fingernails at the autopsy stage did not record foreign profiles in the majority of cases (Cerri et al. 2009).

In a study involving

> "deliberate scratching of another individual (n = 30), 33% of individuals had a foreign DNA profile beneath their fingernails from which the person they scratched could not be excluded as a source; however, when sampling occurred 6 hours after the scratching event, only 7% retained the foreign DNA" (Matte et al. 2012).

In controlled experiments with females scratching males to simulate assaults, 95% (38 out of 40) of fingernail samples collected immediately and 60% (24 out of 40) of those collected five hours later were "suitable for comparison" (Iuvaro et al. 2018). Analyses of fingernail samples in criminal cases were also studied (Bozzo et al. 2015).

Clothing is also submitted in cases of alleged digital penetration. In an experiment designed to better target sampling, a mannequin was used to determine how much DNA was transferred by volunteers to parts of underwear (Ramos et al. 2020).

5.3.3.5. Wearer versus Toucher

In the past, it may have been common to use the DNA profile obtained on a garment as a proxy for the DNA profile of the person who wore the garment (e.g., Casey et al. 2016). However, the issue of increased sensitivity is again relevant. A 2018 study showed that the wearer profile was detected in all interpretable profiles, and it was the major profile 50% of the time (Magee et al. 2018). However, the definition of *interpretable* varies across laboratories (e.g., Benschop et al. 2017). Therefore, information obtained from many of these DNA transfer studies will only be

valuable in a particular case when carried out under similar conditions and interpretation criteria, which is why metadata about the samples and experimental conditions is important to collect.

An interlaboratory study considered upper garments after being worn by individuals who embraced (contact), went on an outing together (close proximity), or spent a day in another person's environment (physical absence). The wearer was typically but not always observed as the major contributor to the profiles obtained. The authors of the study noted: "DNA from the activity partner was observed on several areas of the garment following the embrace and after temporarily occupying another person's space. No DNA from the activity partner was acquired by the garments during the outing even though both participants were in close proximity" (Szkuta et al. 2020).

5.3.4. Studies on Contamination

Contamination is a type of DNA transfer. However, it is typically considered as a special case of transfer and is investigated separately from the types of DNA transfer studies discussed above. Many studies focus on contamination and on suitable methods to avoid it. A list of such studies is presented in Table 5.3.

Table 5.3. Studies published from 2003 to 2019 where measuring or investigating potential sources of contamination is the main focus.

No.	Reference and Title	Purpose of Study	Size of Study	Key Results	Implication
1	Rutty et al. (2003) The effectiveness of protective clothing in the reduction of potential DNA contamination of the scene of crime	Series of experiments were undertaken to determine the extent to which an investigator could contribute to any DNA contamination of a scene of crime under different simulated activities; effectiveness of protective clothing checked	18 experiments with one participant	In total, 413 alleles were identified in the 18 experiments, and 34 were not attributable to the subject and therefore considered to be contamination; vigorous activity, even when wearing protective garments, can cause contamination of a crime scene	Need for ongoing checks on the effectiveness of protective clothing
2	van Oorschot et al. (2005) Beware of the possibility of fingerprinting techniques transferring DNA	Check the potential of fingerprint brushes to transfer DNA	13 brushes used to powder surface containing saliva before powdering clean plates; DNA contaminated brushes used to powder 6 plastic sheets in another experiment	Transfer occurred when brushed over a biologically stained area or fresh print	Need to ensure fingerprint brushes are not transferring DNA

No.	Reference and Title	Purpose of Study	Size of Study	Key Results	Implication
3	Durdle et al. (2009) The transfer of human DNA by *Lucilia cuprina* (Meigen) (Diptera: Calliphoridae)	*Lucilia cuprina* were fed either human blood or human semen *ad libitum* and their artifacts were analyzed for human DNA content	Samples containing 1, 10, 30, and 50 artifacts	Blowfly artifacts can be a source of DNA at crime scenes, in addition to being a potential contaminant; data suggest the amount of DNA in artifacts can be dependent on the meal type	Depending on the environment, be conscious of ability of living things, other than humans, to transfer DNA
4	Preuße-Prange et al. (2009) The problem of DNA contamination in forensic case work—How to get rid of unwanted DNA?	Tested the efficiency of different (chemical and physical) procedures for DNA removal with focus on the commonly recommended ultraviolet (UV) irradiation	Saliva and pure DNA applied to glass slides for 9 time periods from 5 min to 24 h and exposed to UV sources at 8 and 48 cm	Pure DNA reduced more effectively than saliva	UV irradiation can only reduce the contamination but does not eliminate it completely;- importance of contamination avoidance prior to analysis
5	Daniel and van Oorschot (2011) An investigation of the presence of DNA on unused laboratory gloves	A preliminary investigation of three brands of laboratory gloves was undertaken to determine the levels of human DNA present on unused gloves from closed and open boxes	In total, 56 gloves were examined from six to seven closed boxes of three different brands	5 gloves from four of seven boxes of one brand had up to 20 alleles	Use certified DNA-free gloves
6	Digréus et al. (2011) Contamination monitoring in the forensic DNA laboratory and a simple graphical model for unbiased EPG classification	Devising a classification scheme for monitoring contamination events	25 EPGs compared with classification made by two reporting officers	Scheme operational	Potential for monitoring across laboratories

No.	Reference and Title	Purpose of Study	Size of Study	Key Results	Implication
7	Durdle et al. (2011) The change in human DNA content over time in the artefacts of the blowfly *Lucilia cuprina* (Meigen) (Diptera: Calliphoridae)	Check whether human DNA that can be profiled from blowfly changes with time	41, 43, and 22 samples tested for blood, semen, and saliva fed to blowflies	Blood and semen data showed that the amount of human DNA that could be extracted increased over the first 400 days but had decreased to one-month levels by 750 days; no changes in saliva over 60 days in the amount of human DNA that could be extracted	Issue for cases held in storage
8	Goray et al. (2012b) DNA transfer within forensic exhibit packaging: Potential for DNA loss and relocation	Investigation of how much DNA is "lost" from an exhibit due to its transfer to the inside of the packaging containing the exhibit, and transfer from one area of an exhibit to another	Multiple variables of substrate and packaging	Demonstrated that DNA could be transferred from the deposit area to either other parts of the item or to the bag itself and usually to both	% total DNA and number of alleles
9	Szkuta et al. (2013) The potential transfer of trace DNA via high-risk vectors during exhibit examination	Check level of DNA potentially transferred between high-risk vectors (scissors, forceps, gloves) and exhibits during the examination process in both light and heavy contamination/contact scenarios	24 swatches stained with 25 µL of blood used as source for multiple uses by three vectors	DNA transfer was observed for all vectors in both heavy- and light-contact scenarios; sufficient alleles to identify the origin except in case of forceps, where only a small number of alleles were transferred under light conditions	Tools and equipment should be cleaned or replaced immediately if they come into contact with substrate containing blood
10	Neuhuber et al. (2009) Female criminals—It's not always the offender!	Systematic search for errors in the investigative process following the contamination of multiple cases in 1993 and 2009 by female DNA	In 34 out of 191 swabs, peaks were found at 4 or more loci of the SGM+-kit; these 34 swabs corresponded to 2 manufacturers	Noted that cotton swabs that had been sterilized with radiation were often contaminated	Manufacturing process, as well as the products themselves used in collection of DNA trace evidence, should be reevaluated with the emphasis on preventing contamination

No.	Reference and Title	Purpose of Study	Size of Study	Key Results	Implication
11	Henry et al. (2015) A survey of environmental DNA in South Australia Police facilities	Survey of police areas where items are sometimes examined prior to submission to laboratories, 18 facilities across South Australia	20 various items sampled; number of times sampled varied from 1 to 29	50% had DNA, 4% originated from 1 person, 9% from 2 people, 19% from 3 people, and 18% from 4 or more people; 20% weak profile; 30% no profile	Need procedures to reduce environmental DNA in examination rooms
12	Kovács and Pádár (2015) Misinterpretation of sample contamination in a Hungarian case report	Case report of DNA from soft tissue from bone sent to two laboratories for identification with conflicting results, which were due to mix up	One bone sent to two laboratories	Results of a case study	The risk of contamination must never be ignored in forensic examination, and the evaluation of minor/major components of a mixed profile can lead to a wrong interpretation
13	Margiotta et al. (2015) Risk of DNA transfer by gloves in forensic casework	All the gloves used in one day by four operators were analyzed; For every glove evaluated, the presence of contamination DNA from the operator or from other samples was detected	16 pairs of gloves used by 4 operators; 5 negative controls from used and unused boxes	12.5% no alleles; 10% operator-related alleles; 12.5% alleles referable to the operator and to the test sample; 50% a mixture of alleles of the test sample and unknown subjects; 15% alleles of unknown subjects different from the operator	Operators must change gloves every time after touching items or surfaces, prior to touching the exhibit
14	van Oorschot et al. (2015) Considerations relating to the components of a laboratory DNA contamination minimisation monitoring (DCMM) program	Advice on what an environmental monitoring program should include	Discussion paper rather than experimental study	Discussion paper rather than experimental study	Information available on what needs to be considered for environmental monitoring

No.	Reference and Title	Purpose of Study	Size of Study	Key Results	Implication
15	Szkuta et al. (2015a) DNA transfer by examination tools—a risk for forensic casework?	Check if DNA and blood transferred to DNA-free surfaces via scissors, forceps, and gloves	Twenty sets of vectors, multiple donors, and four replicates per transfer set; transfer sets each contained blood and touch DNA	DNA-containing material can be transferred from exhibit to exhibit by scissors, forceps, and gloves	Encourage awareness amongst staff of the potential sources of contamination within the laboratory and during examination
16	Szkuta et al. (2015b) Residual DNA on examination tools following use	Check the proportion of DNA that remains on the high-risk vectors following contact with the substrate.	Transfer experiment as Szkuta et al. 2015a	While DNA-containing material is picked up by DNA-free vectors and transferred from exhibit to exhibit, sufficient DNA remains on these vectors, which can potentially result in further transfer and contamination through subsequent contact	See Szkuta et al. 2015a
17	Fonneløp et al. (2016) Contamination during criminal investigation: Detecting police contamination and secondary DNA transfer from evidence bags	Check level of contamination in police facilities and check scene-of-crime officers' profiles against casework from 2009 to 2015 A pilot study to assess whether DNA from the outside package of an exhibit could be transferred to a DNA sample was also carried out	Areas divided into high-, medium-, and low-risk areas and three gloves checked after checking case-created scenarios	Environmental DNA was detected in various samples from hot spots; furthermore, 16 incidences of previously undetected police-staff contamination were found; in 6 cases, the police officers with a matching DNA profile reported that they had not been involved with the case	Important to ensure that "best-practice" procedures are upgraded, and appropriate training is provided in order to ensure that police are aware of the increased contamination risks; specific recommendation s listed below

No.	Reference and Title	Purpose of Study	Size of Study	Key Results	Implication
18	Bolivar et al. (2016) Assessing the risk of secondary transfer via fingerprint brush contamination using enhanced-sensitivity DNA analysis methods	Check whether fingerprint brushes transfer DNA from fingerprint when using traditional profiling and low-template profiling methods	Six samples, six substrate controls, and six brush controls were collected from each of the three sets of latent and contaminant donors for a total of 18 samples, 18 substrate controls, and 18 brush controls	Although LCN improves the recovery of the DNA profile from the latent print evidence, it also increases the chance of detection of extraneous DNA, such as that transferred by fingerprint brush contamination	Improper procedures may lead to false exclusions or false associations between evidence and crime scene; therefore, procedures for examining latent print evidence should be carefully examined, especially when higher-sensitivity DNA analysis methods are utilized
19	Taylor et al. (2016) Observations of DNA transfer within an operational forensic biology laboratory	Investigation of the extent to which individuals at Forensic Science SA (FSSA) deposit their DNA on objects throughout the floor of the building where DNA examinations take place by examining monitoring and contamination events as well as specific sampling	138 samples were taken from areas across the floor	Evidence that some individuals shed DNA more readily than others over time; last person to handle an item not necessarily detected; primary transfer accounted for 9/14 contamination events	Questions of how and when did the DNA get there more challenging than statistical calculations; more studies needed to avoid more uninformative responses such as is possible
20	Neuhuber et al. (2017) Police officer's DNA on crime scene samples—Indirect transfer as a source of contamination and its database-assisted detection in Austria	Systematic investigation of contamination events	Between the years 2000 and 2016, 347 contamination incidents were detected in approximately 46,000 trace DNA samples (0.75%)	The DNA profiles were screened for contamination incidents by combining a manual check with database-assisted profile comparisons using the national Police Elimination Database (PED) as well as the profile comparison tool of the GeneMapperID-X software	The potential and importance of reference databases containing DNA profiles of police officers and examiners for the detection of contaminated crime scene samples is demonstrated

No.	Reference and Title	Purpose of Study	Size of Study	Key Results	Implication
21	Pickrahn et al. (2017) Contamination incidents in the pre-analytical phase of forensic DNA analysis in Austria—Statistics of 17 years	Continuation of work from Neuhuber et al. (2017)	347 contamination incidents in 17 years	The usefulness of reference profile databases that contain DNA profiles of police officers to detect contamination incidents of trace DNA material	With improved detection methods, it also becomes apparent that indirect transfer of biological material is a serious issue
22	Szkuta et al. (2017b) DNA decontamination of fingerprint brushes	Assessment of the contamination risk of reused fingerprint brushes through the transfer of dried saliva and skin deposits from and to glass plates; assessment of ability to eradicate DNA from brushes	7 new and used squirrel and fiberglass fingerprint brushes used in simulated casework scenarios using glass plates with saliva, single and multiple handprints as substrates; repeated 6–12 times on each substrate and 3 deposits on secondary surface following washings	No profiles observed on new fiberglass brushes, but yields of ≤1 ng on squirrel brushes containing alleles to imply 3 to 4 people; detectability dependent on secondary surface and on biological nature of material being transferred; squirrel brushes easy to clean effectively but fiberglass brushes became tangled and matted	A protocol needed to ensure brushes not used as vectors for transfer of DNA within and between crime scenes
23	Basset and Castella (2018) Lessons learned from a study of DNA contamination from police services and forensic laboratories in Switzerland	National inventory of contaminations to better understand their origin and to make recommendations in order to decrease their occurrence	Mean of 11.5 (9.6 to 13.4) contaminations per year per 1000 profiles sent to the Swiss DNA database	86% of these contaminations originated from police officers, whereas only 11% were from genetic laboratories; direct contact between the stain and the contaminant person occurred in only 51% of the laboratory contaminations, whereas this number increased to 91% for police collaborators	Improving sampling practices at the scene could be beneficial to reduce contaminations

No.	Reference and Title	Purpose of Study	Size of Study	Key Results	Implication
24	Helmus et al. (2019) Unintentional effects of cleaning a crime scene—When the sponge becomes an accomplice in DNA transfer	The aim of this study was to investigate whether DNA traces could be distributed by cleaning an object	Blood, saliva, and epithelial cells from 5 individuals; samples deposited onto two surface types and cleaned with wet sponge; 218 samples initially and 384 in a different experimental setup	It is not only possible but rather probable to distribute DNA from one place to another by cleaning the surface of an object as long as the DNA source is blood or saliva. Regarding DNA from epithelial cells, a transfer of enough DNA for a complete profile by wiping is unlikely	Disposable materials best for cleaning surfaces contaminated with biological fluids
25	Goray et al. (2019) DNA transfer: DNA acquired by gloves during casework examinations	The aim of this study was to investigate DNA transfer during actual casework examinations even when wearing gloves	96 gloves from the examination of 11 exhibits carried out by 5 examiners	Gloves used during examination can collect DNA from the exhibits; for instance, during trace sampling, such losses to the gloves can result in the reduction of DNA available, impacting the quality of the evidentiary profile; furthermore, DNA collected on the gloves could be redeposited on other parts of the exhibit	Profiles were interpreted and statistically evaluated using continuous probabilistic software STRmix (version 2.06) This software weights genotype combinations and allows comparison to persons of interest (POI) and the staff elimination database, expressed as likelihood ratios

The differences between DNA profiles produced by low-template and standard STR multiplex analysis were discussed when high-sensitivity approaches were introduced (Gill et al. 2000). At that time, 100 pg, or about 16 cells, was the minimum amount of DNA that would be analyzed. Duplicate analyses were recommended, and, interestingly, it was noted that laboratory contamination in the form of random alleles could not be eliminated.

The appearance of random additional alleles was previously encountered when profiling wildlife samples from bears (Taberlet et al. 1996). In that situation, the authors explained that the alleles must have arisen as an artifact during PCR because the alleles detected had not been encountered in that laboratory before and therefore could not have been the result of contamination.

Many of the studies on transfer and persistence in which ground truth is known note the presence of alleles not associated with subjects of the study. These alleles are generally

attributed to contamination. Such contamination could add to the difficulties of mixture deconvolution when dealing with casework.

The studies on contamination in Table 5.3 illustrate the various ways that contamination can occur during crime scene examination prior to receipt by the laboratory. The studies give information on possible vectors and other risks that could give rise to such false inclusions. The possibility of contamination from an innocent person's profile is discussed, and the value of elimination databases is supported (Pickrahn et al. 2017, Fonneløp et al. 2016). Miscarriages of justice have arisen because of contamination either before the laboratory or in the laboratory (e.g., Gill 2014, Gill 2016, Gill 2019a).

Contamination is often considered in the context of laboratory handling. The early application of low-template DNA outlined necessary precautions in the laboratory (Gill 2001). The main concern at the time was that contamination by stray alleles would cause false exclusions. These precautions included the need to carry out PCR amplification in a separate contained laboratory, that personnel wear disposable laboratory coats and face masks, that staff and police elimination databases be used, and that duplicate tests be performed when possible. A study was conducted on the risk of contamination via routine implements such as scissors and forceps (Szkuta et al. 2015a). Results obtained demonstrated not only that DNA transfers from exhibit to exhibit, but also that DNA persisted on the tools, making future transfers possible. This can give rise to the possibility of false inclusions as well as exclusions (Szkuta et al. 2015a).

Three studies examined the possibility of nitrile gloves acting as vectors (Fonneløp et al. 2015a, Szkuta et al. 2015a, Goray et al. 2019). The results illustrate the need for frequent and appropriate changing of gloves to avoid moving DNA from object to object. The finding of sufficient levels of DNA capable of providing STR alleles on unused gloves is an additional cause for concern (Daniel & van Oorschot 2011, Margiotta et al. 2015). A study on DNA acquired by gloves during casework found:

> "In many instances, the case-associated person of interest was observed within the profile generated. So too were profiles of the examiner or other staff members, predominantly from the first and last gloves used during the examination, which were associated with removing the exhibit from its packaging and repackaging it." (Goray et al. 2019)

Fonneløp et al. 2016 considered the possibility of contamination prior to receipt by a laboratory. It was demonstrated that DNA from the outside of bags could contaminate an exhibit during examination (Fonneløp et al. 2016). Fingerprint brushes also were the subject of a study as potential vectors for transfer of DNA. The additional concern in the case of brushes was that some new brushes had considerable detectable DNA (Szkuta et al. 2017b). The transfer of human DNA by blowfly *Lucilia cuprina* has also been reported (Durdle et al. 2009).

Contamination avoidance is a well-known concept in DNA laboratories (e.g., Butler 2011, p. 18). The UK Forensic Science Regulator (UKFSR) has issued guidance on avoiding contamination in the DNA laboratory (UKFSR 2015), during sexual assault forensic medical exams (UKFSR 2016a), and at the crime scene (UKFSR 2016b). The Scientific Working Group on DNA Analysis Methods

(SWGDAM) has also published guidelines on contamination prevention and detection (SWGDAM 2017b).

van Oorschot et al. (2015) discussed a program for monitoring and minimizing laboratory DNA contamination in the context of key performance indicators (KPIs) and the cost of such a program. Periodic sampling of work areas, blind proficiency testing of individuals, practitioner self-assessment of compliance, general compliance with audits, and practitioner observation and assessment were recommended. This approach called for root cause analysis when contamination was detected.

The Netherlands Forensic Institute (NFI) identified contamination as a particularly important quality concern. They published a study reporting on errors in casework during the period 2008 to 2012 (Kloosterman et al. 2014). NFI observed an increase in the number of cases of contamination over that time period. This increase was explained by an increase in the number of analyses, a more sensitive analytical system, an increase in the number of persons in the elimination databases (which allowed for more contamination to be recognized), and an increase in the requests for "touch DNA" evidence (Kloosterman et al. 2014). The NFI study distinguished between cases where there are multiple samples of DNA and those where the findings consist of a single low-level DNA sample. The authors noted that there are signals that would prompt a scientist to consider possible contamination, but only if the scientist is alerted to this possibility through tools like an elimination database (Kloosterman et al. 2014).

An article highlighted the possibility of DNA contamination in mortuaries and suggested that time and money may be wasted searching for profiles matching deceased individuals who may be already buried or cremated (Rutty 2000).

A 2017 report of contamination incidents in Austria over a 17-year period also highlighted the need for elimination databases (Pickrahn et al. 2017). The infamous Phantom of Heilbronn case involving contamination of swabs by the manufacturer (Neuhuber et al. 2009, Butler 2011, p. 79) may have prompted the study of potential contamination by police officers collecting evidence at crime scenes (Nuehuber et al. 2017). Such contamination, which causes false positive results and can potentially mislead investigations, is an ongoing challenge for forensic laboratories and a constant reminder of the ease with which DNA transfers. A later publication in this area presented lessons learned from a study of DNA contamination of police services and forensic laboratories in Switzerland (Basset & Castella 2018). An international documentary standard was published in 2016 to help address potential contamination in reagents and products used to collect and process DNA samples (ISO 18385:2016).

Given that DNA can transfer readily, precautions are needed both before and after evidence is submitted to a laboratory. Fonneløp et al. 2016 noted 16 instances of previously unknown police-staff contamination found later when reference samples became available and called for a national elimination database or elimination protocol in Norway. The difficulty of identifying contamination if elimination databases are not in place is implicit in the following statement:

> "… currently most morticians, pathologists, and even the police officers and their allied workers do not have their DNA profiles in the database for exclusion purposes" (Rutty 2000).

A Canadian forensic laboratory conducted 30 information sessions between 2010 and 2012 with multiple police agencies in their jurisdiction on the value of a DNA elimination database that resulted in 327 DNA profiles from crime scene personnel being added to their elimination database (Lapointe et al. 2015). Up to the time when their article was submitted for publication two years later, the authors noted that DNA profiles from 46 (14% of these 327) different crime scene workers matched to 58 criminal cases including 31 cases that were already in their national DNA database. The DNA profiles from crime scene workers were typically found as "touch DNA" on handled objects of evidence and usually contained less than 2.0 ng of DNA (Lapointe et al. 2015). The authors conclude: "Acquiring as many DNA profiles from crime scene workers for elimination purposes has now become an important, if not crucial, tool to help reduce erroneous investigative leads…, [and] crime scene workers may be lacking vital information to help them understand the necessity of contributing to such an important forensic tool" (Lapointe et al. 2015).

5.3.5. Studies Involving Casework Scenarios

As with any community of practice, some insight can be gained from a review of casework. Many groups have collated the type of samples from which DNA profiles have been successfully obtained (Castella & Mangin 2008, Dang et al. 2012, Djuric et al. 2008, Dziak et al. 2018, Mapes et al. 2016, van Oorschot 2012). Other groups have considered particular evidence or sample types, such as adult necks (Graham & Rutty 2008), sandals (Ferreira et al. 2013), zip-lock bags in drug cases (Hellerud et al. 2008), and ammunition (Montpetit & O'Donnell 2015).

Several studies investigated the use of low amounts of DNA in various property crimes and proposed considering factors in a wider context (Forr et al. 2018). Some transfer studies attempted to mirror casework (Raymond et al. 2008a, 2008b, 2009a, 2009b; Fonneløp et al. 2017), while others sought to assess outcomes in mock scenarios (Benschop et al. 2012, Goray et al. 2012a). Finally, case context and interpretation issues in specific case examples were explored by several authors (McKenna 2013, Jackson 2013, Jackson & Biedermann 2019).

5.3.6. Literature on How to Evaluate DNA Relevance in Context

The scientific publications examined in this chapter cover properties of low-template DNA and provide insights into how those properties affect transfer and persistence. In addition, several publications describe approaches to interpretation that explicitly consider relevance of the DNA to the crime. Publications that cover this last topic are listed in Table 5.4.

Table 5.4. Summary of topics and some associated references explored in the following section.

Topics	References
Insight on the impact of low-template DNA	Taberlet et al. 1996, Gill et al. 2000, Gill 2001, Gill 2002, Gill & Buckleton 2010, Gill et al. 2015, Benschop et al. 2015
Case assessment and interpretation model (CAI) and the hierarchy of propositions	Cook et al. 1998a, Cook et al. 1998b, Evett et al. 2000a, Evett et al. 2000b, Evett et al. 2002, Jackson et al. 2006
Theoretical frameworks for assessing transfer evidence	Biedermann & Taroni 2012, Champod 2013, Taylor et al. 2018, Taylor et al. 2017, Taroni et al. 2013, Taylor et al. 2019, Samie et al. 2020
Formulating propositions	Biedermann et al. 2016a, Hicks et al. 2015, Gittelson et al. 2016, Kokshoorn et al. 2017, Taylor et al. 2017
Distinction between investigation (police) and evaluation (court) uses of DNA	Jackson et al. 2006, ENFSI 2015, Gill et al. 2018, UKFSR 2018a, UKFSR 2021

The strategies and approaches presented in the publications listed in Table 5.4 are further discussed in the sections below.

5.4. Discussion

5.4.1. Implications from What Is Currently Known

The studies reviewed herein show that the amount and likelihood of DNA transfer vary widely under different conditions. However, the possibility of transfer cannot be ignored when interpreting DNA evidence. If it is ignored, DNA findings, when considered in isolation, have the potential to be misleading[24].

Based on review of the literature described above, it is possible to outline several ways in which DNA transfer might mislead an investigation. These include the following:

- **It is possible to handle an item without transferring any detectable DNA to it.**

  The absence of detectable DNA was noted in 11% of experiments by Manoli et al. (2016) and in 2.9% by Goray et al. (2016). In addition, Meakin et al. (2017) noted that full profiles were not always detected. The shedder status of the donor as well as effectiveness of extraction and analytical methods are all relevant here and have been extensively discussed in the literature (Lowe et al. 2002, Farmen et al. 2008, Taylor et al. 2016, Taylor et al. 2017, Taylor et al. 2018).

- **Genetic material may have been deposited before or after the crime and therefore may not be relevant to it.** This can happen because the person has legitimate access to the scene or item, or because the DNA was transferred in some other way (Raymond et al. 2009a, Goray & van Oorschot 2015). Studies examining persistence of original user or wearer following another user show that the substrate as well as the shedder status of the first and second user affect the findings (Fonneløp et al. 2015b, Oldoni et al. 2015, Oldoni et al. 2016, Meakin et al. 2015, Meakin et al. 2017, Pfeifer & Wiegand 2017).

- **Detected DNA might be present due to indirect (secondary or tertiary) transfer, whether from a person or an object.** These transfers can occur before or after the commission of a crime and be due to innocent activity in the area. They might also be the result of contamination during evidence collection, transport, and other stages of the investigation or during the laboratory submission, storage, and examination processes.

While the traditional view is to focus on the major contributor to a mixture based on the assumption that the profile belongs to the last person to handle an item, some studies have shown this is not always the case (e.g., Cale et al. 2016, Buckingham et al. 2016, Goray et al. 2016).

The LR given sub-source level propositions, no matter the magnitude, should never be used as the strength of the evidence with regard to activity level issues.

[24] A public comment on the June 2021 draft report from a forensic practitioner noted: "The value of forensic DNA testing as a tool in answering the question as to who is the/a source must also not be conflated with its far more limited value in answering the questions as to what bodily substance the DNA comes from, and when and how it came to be deposited. Developing a more disciplined mindset in relation to the concept of relevance as the driver of our examination strategies and interpreting and reporting results in that framework will help to ensure that we are always providing information of the highest value to the criminal justice system" (see PC24 in 8351-draft PCs).

The highly sensitive DNA methods that have become commonplace increase the likelihood of detecting irrelevant DNA. Peter Gill, in a review article covering the previous 20 years of development in the field, claims that all laboratories at the time are testing for low-template DNA (Gill et al. 2015). Although the definition of low-template DNA may be considered trivial, the impact of highly sensitive methods on interpretation is important: "the lower the amount of DNA present in a sample, the greater the chance that it may not be associated with a crime-event" (Gill et al. 2015).

Relevance was identified as an issue when low-template DNA work was first introduced: "Inevitably, there is a direct relationship between the quantity of DNA present and the relevance of the evidence" (Gill 2001). The authors of a study seeking to establish the limits for DNA mixtures using small amounts of DNA concluded:

> "The relevance of the evidence, rather than the DNA typing methodology or statistical model, may be the limiting factor for obtaining useful results for forensic casework and court going purposes." (Benschop et al. 2015)

Weight-of-evidence statistics (e.g., likelihood ratios given sub-source propositions; see Taylor et al. 2018) are often produced in forensic laboratories as stand-alone findings, perhaps with a brief disclaimer in the accompanying report that mentions the possibility of transfer but does not treat this issue sufficiently. The studies in this chapter suggest that this area would benefit from more attention during routine practice to avoid potentially misleading findings.

The following section discusses strategies to help ensure that LRs are considered in context and to mitigate the risk that DNA transfer might mislead an investigation.

KEY TAKEAWAY #5.2: Highly sensitive DNA methods increase the likelihood of detecting DNA mixtures that may or may not be related to the crime being investigated. In cases that involve very small quantities of DNA, it is especially important for the users of the information to consider context when determining the utility of the evidence.

5.4.2. Strategies for Mitigating the Risk of Misleading DNA Results

5.4.2.1. Minimize Contamination

Forensic scientists have known since the advent of forensic DNA methods that DNA can transfer readily (e.g., van Oorschot & Jones 1997). This is evidenced by the systems that laboratories have had in place since then to avoid contamination (e.g., Butler 2011, p. 18). However, the use of highly sensitive methods increases the probability of detecting small amounts of contaminating DNA. "Along with increased sensitivity comes the prospect of detecting contaminating DNA, complicating the interpretation of profiles" (Szkuta et al. 2013).

The contamination avoidance strategies in forensic laboratories that have long been in place are more important than ever. Furthermore, as evidenced by the studies outlined in Table 5.5, contamination can happen during a scene investigation. Therefore, contamination avoidance procedures must be in place during all stages of an investigation, from the crime scene through

the production of the profile. These studies also highlight the need for elimination databases (e.g., Basset & Castella 2018, Basset & Castella 2019) to avoid wasting resources following up on profiles that arise from the examination and also as a way of reducing complexity in mixtures.

Contamination can take various forms and consists of DNA from investigators and scientists and other personnel at the crime scene, the hospital, and/or laboratory from inappropriate handling of evidence items or transfer from one surface to another, which can be a particular risk when dealing with heavily blood-stained items.

Table 5.5. Examples of routes where contamination of DNA can occur as illustrated in the UK Regulator's guidance on DNA Anti-Contamination–Forensic Medical Examination in Sexual Assault Referral Centers and Custodial Facilities (UKFSR 2016a). Examiner refers to an individual conducting laboratory tests while practitioner is a forensic healthcare provider such as a nurse.

Direct transfer						
Sample	to	Environment/item				
Environment/item	to	Sample				
Consumable	to	Sample				
Person	to	Environment/item				
Indirect transfer—secondary transfer						
Environment/item	to	Examiner	to	Sample		
Environment/item	to	Consumable	to	Sample		
Environment/item	to	Practitioner	to	Sample		
Environment/item	to	Environment/item	to	Sample		
Person	to	Examiner	to	Sample		
Person	to	Environment/item	to	Sample		
Sample 1	to	Environment/item	to	Sample2		
Indirect transfer—tertiary transfer						
Person	to	Environment/item	to	Consumable	to	Sample
Person	to	Environment/item	to	Examiner	to	Sample
Environment/item	to	Environment/item	to	Examiner	to	Sample
Environment/item	to	Environment/item	to	Practitioner	to	Sample
Sample 1	to	Environment/item	to	Examiner	to	Sample 2

KEY TAKEAWAY #5.3: Contamination avoidance procedures should be robust both at the crime scene and in the laboratory. These procedures should include the maintenance of elimination databases containing samples from personnel who have access to crime scenes and evidence items.

5.4.2.2. Consider Evidence in Context

It is a principle of forensic science that results only have meaning in context (e.g., Evett & Weir 1998, Cook et al. 1998a, Cook et al. 1998b). The trend, however, is for the forensic scientist to have limited access to information about the case. This trend is driven in part by efforts to avoid confirmation bias. These efforts risk isolating the forensic scientist from contextual information that may be crucial when assessing relevance. It is possible to facilitate both approaches by sequential unmasking of information (Butler 2014, pp. 461–464).

One way of considering evidence in context is to view the case as a whole rather than simply evaluating a single sample in isolation. As noted at the beginning of the chapter, Locard spoke of a criminal "leaving *multiple* traces of his path…" (Roux et al. 2015, emphasis added). This observation should serve as a caution against expecting a single association to *solve* the crime. This is in keeping with views expressed by others (e.g., Gill 2014, Sense about Science 2017) that DNA should not be used as the sole evidence in a criminal case, and that it is inappropriate to assume that DNA always has greater value than other types of evidence. A 2020 publication outlines a method for combining different types of evidence (de Koeijer et al. 2020).

A miscarriage of justice that occurred in Australia demonstrated these points. In this case, DNA was the only evidence in an alleged rape, but that DNA was later shown to have resulted from cross-contamination in a sexual assault examination room. The judge who later reviewed the circumstances that led to the conviction stated:

> "In the present case, the obviously unreserved acceptance of the reliability of the DNA evidence appears to have so confined thought that it enabled all involved to leap over a veritable mountain of improbabilities and unexplained aspects that, objectively considered, could be seen to block the path to conviction" (Vincent 2010).

This review cautioned that DNA

> "must be carefully used and placed into proper perspective and understood that a calculation of statistical likelihood provides a dangerous basis for conviction, if it is upon that alone that proof beyond reasonable [doubt] rests" (Vincent 2010).

While the Australian case involved cross-contamination of evidence, the warning from the judge about misusing a statistical likelihood applies to any case that may involve DNA transfer. The LR, as typically used when interpreting DNA mixtures, is based only upon the analytical properties of the DNA. It does not provide information about other important aspects of the evidence, such as the quantity of DNA in the original sample or whether the cell type is known. Therefore, a large blood stain might produce a very similar LR to a swab from a light switch, yet the two have would very different meanings in the context of a case (e.g., Taroni et al. 2013). While an LR is an expression of the strength of evidence under a pair of propositions, the result should be considered in context (i.e., the result represents the evidence of what?).

5.4.2.3. Ask and Answer the Right Questions

Keith Inman and Norah Rudin have written: "One of the greatest unrecognized contributions that a criminalist can provide [to a case] is framing the correct question" (Inman & Rudin 2001).

The trier of fact needs to know the answers to multiple questions, many of which the forensic scientist cannot address. Who, what, when, where, how, and why all need to be answered at the criminal trial. The LR as typically used in DNA mixture interpretation addresses the *who* question, but it does not address the questions of *when* and *how* the DNA was deposited. This presents a risk that the trier of fact might use an answer to a relatively easy question to answer the more difficult questions. A 2019 review article, describing this phenomenon as an attribute substitution, stated: "If someone doesn't know the answer to a difficult question, they will substitute an easier question (even if subconsciously) and answer that instead" (Eldridge 2019). This tendency highlights the need to be clear about what questions are being addressed with any particular interpretive method.

KEY TAKEAWAY #5.4: DNA statistical results such as a likelihood ratio given sub-source propositions do not provide information about how or when DNA was transferred, or whether it is relevant to circumstances of a case. Therefore, reporting a likelihood ratio as a standalone number can be misleading without sharing the assumptions made in the LR assignments and the level in the hierarchy of propositions being addressed (i.e., considering what question is being answered).

5.4.2.4. Use Case Assessment and Interpretation

The references cited in Table 5.4 include a paper that introduces a framework for ensuring that case context is considered when evaluating evidence (Cook et al. 1998a). Case Assessment and Interpretation (CAI), which has come to be known as evaluative reporting, provides a systematic way to produce "an assessment of the strength to be attached to the findings in the context of alleged circumstances" (ENFSI 2015, see also UKFSR 2021).

CAI requires the forensic scientist to document their expectations in a given scenario before examining the evidence. For example, a violent assault involving significant bloodshed would typically be expected to yield multiple transfers to the assailant rather than trace amounts of DNA of no known cell type. Documenting expectations in this way can help avoid being "findings-led" (i.e., trying to make the findings fit the case). Without an assessment before examinations, the scientist can be accused of drawing the target after the shot is fired, also referred to as the Texas sharpshooter fallacy (Thompson 2009).

CAI serves as the basis of several guidelines developed over the last 10 years (AFSP 2009, ENFSI 2015, ANZPAA 2017). The principles of CAI include:

- The findings are assessed in the context of the case, because they have no intrinsic value in isolation.
- At least two propositions are considered when assigning an LR. The assessments are dependent on the propositions addressed.

- In order to avoid what is commonly referred to as "transposing the conditional," (Thompson & Schumann 1987), the scientist reports on the findings, not the propositions.

5.4.2.5. The Hierarchy of Propositions

The researchers who formulated the CAI framework outlined a hierarchy of propositions, with each level addressing different questions (Cook et al. 1998b), as discussed in Section 2.6. This helped to clarify the questions addressed during evidence evaluation (Cook et al. 1998b, Evett et al. 2000a). The questions addressed at the lower end of the hierarchy—source, sub-source, and sub-sub-source—are seen in Table 2.5. These levels only address questions about the source of the DNA profile. An example of a sub-source-level proposition might be that the DNA mixture contains DNA from the POI and the victim. These source- or sub-source-level propositions are based on the genotypes or alleles present in the evidence, but they do not address in any way *how* the DNA was deposited.

Above the source-level propositions are activity propositions, which address questions about *how* the DNA came to be present in a mixture. An activity proposition might be, for instance, that sexual activity occurred between the POI and the victim, or that the POI stabbed the victim with this knife. Activity-level propositions more directly address issues of interest to the court (Jackson 2013, Taylor et al. 2018), and they almost always involve greater uncertainty than source-level propositions.

Finally, offense-level propositions address questions of guilt or innocence. These questions are addressed by the courts rather than by forensic scientists (see Table 2.5).

It is vital that users of forensic science information understand the differences between levels in the hierarchy and that they do not use the LR for one level to address a question at a higher level. It has been noted:

> "Due attention must be paid to the position in the hierarchy of propositions that can be considered. This information must be effectively conveyed to the court to avoid the risk that an evaluation at one level is translated uncritically and without modification to evaluation at a higher level. We cannot over-emphasize the importance of this. A DNA match may inform decisions about the source of the DNA, but decisions about an activity, say sexual intercourse versus social contacts, involve additional considerations beyond the DNA profile." (Buckleton et al. 2014)

Peter Gill also discussed the risks of conflating source and activity propositions in his book *Misleading DNA Evidence: Reasons for Miscarriages of Justice* (Gill 2014). This book introduces the concept of an "association fallacy," where "a probability is transposed from one level of the framework of propositions to a higher level." Several miscarriages of justice have been shown to result from misleading DNA evidence due to this fallacy (Gill 2014, Gill 2016, Gill 2019a). This risk is increased by the fact that the vast majority of criminal cases in the United States are settled through plea bargaining (Gramlich 2019).

5.4.2.6. Activity-Level Propositions

Activity-level questions have received a growing attention in the published literature and in court (e.g., Taylor et al. 2018). The CAI approach involves formulating activity-level propositions to assign an LR with activity-level reporting (ALR). For example, in reference to scenario A given at the beginning of the chapter, the pair of activity-level propositions might be the POI stabbed the victim with this knife (H_1) and the POI did not touch this knife (H_2). There are many references in the literature to the suitability of this approach but little in the way of prescriptive assistance. A notable exception is a recent retrospective study of 74 cases from the Netherlands Forensic Institute that shared insights into assigning LRs given activity-level propositions and the sources of data used to assign probabilities to DNA TPPR events (Kokshoorn & Luijsterburg 2023). Bayesian networks have been suggested as a method with which to identify those variables that are most likely to impact the LRs given activity-level propositions (Taylor et al. 2017, Biederman & Taroni 2012, Taylor et al. 2019). Depending on the questions being addressed, an LR given sub-source propositions may not be appropriate. This is true when trying to differentiate the expected findings in light of the potential of primary or secondary transfer, for example.

Simulation and modeling are used to assess the impact of variables on LRs based on activity propositions (Samie et al. 2020). The results show that regardless of the DNA outcome, the most impactful variable is the "DNA match probability when the [defense] alleged that the person of interest (POI) had nothing to do with the incident" (Samie et al. 2020). When potential secondary transfer is under consideration, the DNA match probability is no longer the issue and variables associated with circumstances around DNA transfer are important. Extraction, sampling quantity of DNA on hands and background are the variables to be considered. The authors provide a tool to assess the impact of varying the latter two parameters (Samie et al. 2020).

LR values produced from activity-level propositions are generally much lower numbers than those produced from source-level propositions. An early paper illustrated this observation, showing an LR value on the order of 1000 given activity-level propositions, in contrast to what the authors describe as an infinite LR in favor of a sub-source level proposition (Evett et al. 2002). Some have argued that, given that activity propositions produce more conservative assessments of the weight of evidence and are more relevant to the issues of the court, their use is more appropriate (Biedermann et al. 2016b, Kokshoorn et al. 2017, Taylor et al. 2018, Szkuta et al. 2018, Gill et al. 2020a).

In addition, it is possible to obtain some value from the CAI approach after the production of a statistic by having another scientist carry out an assessment and assign probabilities for transfer, errors, contamination, etc., and then evaluate the findings in light of the previously analyzed results.

Some forensic laboratories in Europe have adopted evaluative reports utilizing LRs given activity-level propositions (e.g., Kokshoorn & Luijsterburg 2023), and guidance for doing so has been issued by the European Network of Forensic Science Institutes (ENFSI 2015, ENFSI 2022)

and DNA Commission of the International Society for Forensic Genetics (Gill et al. 2018, Gill et al. 2020a). The ENFSI Best Practice Manual for Human Biology and DNA Profiling states:

> "LRs given activity level propositions are typically many orders of magnitude lower than those calculated given sub-source level propositions. It is useful to demonstrate this even if there are limited data available" (ENFSI 2022, p. 29).

A 2021 survey of U.S. forensic DNA practitioners with 54 responses found "about half of the participants were uncomfortable with activity-level evaluations of DNA evidence, and raised various concerns" (Yang et al. 2022). Survey responses suggest six major concerns to be addressed before implementing activity-level reporting (ALR) in the United States: "(1) effect of [the] number of variables involved; (2) need for education for practitioners/legal system; (3) [an] inadequate number of activity studies with realistic scenarios; (4) difficulty of achieving admissibility in court; (5) need for standardized approaches/guidelines; and (6) requisite shift in perspective as to the validity of ALR" (Yang et al. 2022).

A follow-on survey with 21 questions received 162 responses from forensic science organizations across Europe, Australia, South America, Canada, Asia, and Africa (Prinz et al. 2024). Key concerns expressed by survey participants included filling the education gap for scientists and legal experts, having more DNA evidence-related data under realistic case scenarios, and formalizing an ALR approach (Prinz et al. 2024).

An expert working group on human factors in forensic DNA interpretation published a 436-page report in May 2024 that included an 11-page chapter titled "How and When Questions in DNA Analysis" (EWG 2024, pp. 172-182). This chapter discusses applying knowledge about DNA transfer in criminal cases and what is appropriate for DNA experts to say now with activity-level reporting along with the description of a path forward to improve current practice in the United States. This group recommended that "DNA analysts should not opine about the possibility or probability of direct or indirect transfer having occurred in a case" unless they are "appropriately trained to respond to such questions" (EWG 2024, p. 177). Some proposed responses are given for ways DNA analysts can respond when asked questions that exceed the boundaries of their methods and expertise (EWG 2024, pp. 175-176). A 2023 book *Forensic DNA Trace Evidence Interpretation: Activity Level Propositions and Likelihood Ratios* contains helpful training material with worked examples prepared by two forensic scientists, who are both researchers and practitioners, to assist with understanding and applying the concepts of activity-level reporting (Taylor & Kokshoorn 2023).

The Texas Forensic Science Commission issued a report in July 2024 that discussed 10 important and unresolved issues in performing evaluations given activity-level propositions in the U.S. legal system (TFSC 2024, pp. 30-59). Effective implementation of activity-level reporting requires structures in place to ensure that a DNA analyst is truly "expert" in this realm (see van Oorschot et al. 2019, p. 160) including (1) education and training for practitioners and criminal justice system stakeholders who are users of their findings, (2) documentary standards against which to assess the quality of work performed, (3) clear and detailed protocols guiding the application and calculations performed, and (4) robust validation studies and proficiency tests to establish reliability of reported results under known conditions before conducting routine casework with LRs given activity-level propositions (van Oorschot et al. 2017). While hundreds

of DNA-TPPR studies have been published (e.g., see reviews such as van Oorschot et al. 2019, van Oorschot et al. 2021), these efforts have not been systematic (Gosch & Courts 2019), which makes arriving at meaningful probabilities for a specific scenario challenging at best. Research efforts are underway to create LR frameworks and develop sensitivity analyses for activity-level evaluations (e.g., Gill et al. 2021b, Taylor et al. 2024).

5.4.2.7. The Value of CAI-Based Reasoning

In its fully realized form, CAI involves formulating activity-level propositions and assigning LRs given these activity-level propositions. Assigning those probabilities requires an understanding of DNA transfer and persistence. For instance, the probability that a person transferred DNA onto the handle of a knife during a stabbing would be affected by the material that the knife handle is made of (wood versus plastic), the shedder status of the person of interest, and the history of the knife. There may be insufficient empirical data to assign probabilities based on these factors. Some researchers have argued that, in that case, it would be appropriate to assign "subjective probabilities" as long these probability assessments are transparent and based on justifications that clarify the extent to which these assigned probabilities are informed by data (Biedermann et al. 2016a, ENFSI 2015). It is important to document expectations and identify propositions as required by CAI to consider context, avoid being findings-led, and ensure that the findings address appropriate questions.

Using a CAI approach requires additional time beyond reporting results using sub-source propositions and necessitates specific information for potential case scenarios that may not be readily available to the forensic laboratory or the DNA analyst conducting the casework. The efficiency and throughput of DNA laboratories may work against organizations taking on these issues and the fact that DNA analysts are not always aware of case context.

Confining the report to an assigned LR value given sub-source propositions and answering questions about relevance if and when they arise in court are not balanced efforts and are therefore likely to be biased to one side or the other depending on the circumstances (see EWG 2024, pp. 172-182). Discussion about the lack of suitability of this approach is well argued in Biedermann et al. (2016b) where the authors state:

> "As human beings, we refer to a lot of events as being 'possible' (i.e., the probability of the event is not 0), but forensic scientists should be more informative than this: they should assess *how probable* their results are given the propositions at hand, just like they do when they assess the probability of observing a given DNA profile if it came from some unknown person" (Biedermann et al. 2016b, p. 8, emphasis in the original).

Balance, transparency, logic, and robustness were suggested as four requirements for reporting of scientific findings (ENFSI 2015). Confining the requirements to robustness in isolation is not sufficient to ensure that the court is fully informed. Guidance on this topic for judges, lawyers, forensic scientists, and expert witnesses states:

> "The expert should lead the reader through the process of assigning probabilities for the scientific findings, describing and explaining whatever data have been relied on from (specified) other sources, as well as any information derived from the scientist's own

> personal experience. It should be possible for an informed reader to retrace the logic of the scientist's approach, recheck calculations or reconsider the scientist's findings in the light of changed assumptions. It should be crystal clear to anybody relying on the report how the expert arrived at her opinion, and what assumptions have been made" (Jackson et al. 2015, p. 72).

CAI was originally formulated to help assess which tests that would be most probative. With laboratories under ever-increasing pressure to conduct more tests, this type of analysis would help ensure that laboratory resources are used most effectively. The Resource Group (see Chapter 1) strongly supported the notion that decisions about what evidence items to test should be made by forensic experts rather than policy-makers. CAI provides a framework for making these types of decisions but requires that these experts be familiar with the transfer and persistence of DNA and their laboratory's ability to detect such transfers.

5.4.2.8. Separating Investigation from Evaluation

There are two phases in assessing evidence in a criminal case. During the investigative phase, the goal is to narrow the lines of inquiry and produce a suspect. During this phase, questions of relevance may be set aside while the police might identify other evidence that might provide context. During the subsequent evaluation phase, the scientist would evaluate the evidence by formulating competing propositions that are based on the surrounding case circumstances.

The DNA Commission of the International Society for Forensic Genetics (ISFG) distinguishes between investigative and evaluative modes when using LRs (Gill et al. 2018). The UK Forensic Science Regulator does as well (UKFSR 2018a, UKFSR 2021). Both sets of guidelines anticipate a scientist delivering results in an iterative manner. The challenges and advantages of this approach have been outlined previously (Buckleton et al. 2014). Separating the investigation and evaluation phases has a major impact on the propositions used in LR calculations. The investigator produces information or explanations for findings at a scene. The investigative mode is most appropriate when it is not possible to formulate a pair of propositions or when there is insufficient conditioning information (ENFSI 2015).

The ISFG DNA Commission states:

> "The scientist works in an investigative mode if there is no person of interest in the case. If a suspect is identified, then generally the scientist switches to evaluative mode with respect to this suspect and needs to assign the value of their results in the context of the case. If there is new information (in particular from the person of interest), the scientist will need to re-evaluate the results. It is thus important that reports contain a caveat relating to this aspect" (Gill et al. 2018).

At source level, an evaluation might consider including relatives in the propositions. It also might affect conditioning on particular genotypes if, for instance, the evidence includes the victim's DNA, as often happens in cases of sexual assault. At activity level, wider issues such as opportunities for transfer, persistence, and shedder status should also be considered.

These nuances in different uses of DNA and the effect of different propositions are well reflected in the literature. Nonetheless, in practice, the focus is on the number—that is, an assigned LR. Authors of a 2016 article on formulating propositions stated:

> In [their] experience, "this may be referred to as 'the number' by prosecutor and defense attorney. This practice breaks the connection between the LR and the propositions, and this is regrettable. Discussion in court very likely evolves to activity level, yet there is no direct relationship between the LR for sub-source level propositions and one for activity level propositions." (Gittelson et al. 2016)

The "number" (LR value) is like seeing the headline of an advertisement without reading the small print and considering the propositions behind the number, which are typically included in the report and case files. Kwong recognized this in a *Harvard Law Review* article:

> "Yet despite the perception of DNA evidence as definitive proof, when DNA evidence involves complex mixtures of multiple individuals' DNA, science is not as simple as it appears on television." (Kwong 2017).

The evaluation stage is an opportunity to use the risk-mitigating strategies outlined previously, to review the findings in light of the case context, to assess the possibility of contamination or error, and to formulate activity propositions. Some researchers have proposed additional sampling with an effort to seek information about other genotypes in the mixture, or conduct *ad hoc* transfer experiments that apply to the particulars of the case. This has been referred to as "sense making" by Paul Roberts (Roberts & Stockdale 2018).

The Deputy Commissioner for Crime of the Victoria Police in Australia has commented:

> "DNA matching [is] very valuable to police for intelligence and evidentiary purposes, but, when used as evidence, *[has] to be seen as one part of a circumstantial case and not as the entirety of it*" (Vincent 2010, emphasis added).

KEY TAKEAWAY #5.5: The fact that DNA can transfer between objects does not negate the value of DNA evidence. However, the value of DNA evidence depends on the circumstances of the case.

5.4.3. Growing Awareness of DNA Transfer and Persistence

Interest in DNA transfer and persistence studies has grown over the last 20 years (e.g., van Oorschot et al. 2021, Sessa et al. 2023). A 2018 review noted a growth from five papers published in 2000 to 35 articles on the topic in 2015 (Kokshoorn et al. 2018). In spite of an increase in the number of published studies on DNA transfer, the results of these studies have not been combined to deal with broad questions about transfer mechanisms (Taylor et al. 2017, Gosch & Courts 2019). Rather, information from published studies can be seen as a way of gaining sufficient knowledge to address the questions being raised in court about how DNA is deposited. For example, a 2020 publication examined two types of firearms handled in four realistic, casework-relevant handling scenarios to explore levels of DNA from non-handlers that could inform alternative activity-level propositions in gun-related crimes (Gosch et al. 2020). A

logical framework in which questions of transfer mechanism can be approached probabilistically has been published, together with identification of the gaps that need to be addressed (Taylor et al. 2017).

One of the reasons there is so much variation in the results of the transfer studies is that results can vary across laboratories, as interlaboratory studies show (Steensma et al. 2017, Szkuta et al. 2020). Therefore, any laboratory planning to assist the court by offering probabilities based on these studies will need to adjust for their own level of sensitivity. For example, if the laboratory has a higher level of sensitivity than a particular study, their likelihood of detecting transfer may be higher than the study would suggest.

A 2017 publication by prominent Australian researchers in this area stated:

> "The forensic community needs to acknowledge that the expertise required to perform activity level assessments in relation to DNA-TPPR is distinct from that required for sub-source-level evaluations, and that expertise does not necessarily transfer between the two tasks. Furthermore, the relevant governance standards, accreditation, competency testing, and ongoing proficiency testing applicable within each jurisdiction in relation to DNA-TPPR associated activity-level evaluations should be equivalent in scope and depth to those related to sub-source-level evaluations" (van Oorschot et al. 2017).

KEY TAKEAWAY #5.6: There is a growing body of knowledge about DNA transfer, persistence, prevalence, and recovery (TPPR). However, significant knowledge gaps exist, including: (1) appropriate TPPR data for casework-like scenarios, and (2) education of and standardized approaches for users on how to apply the LR framework to activity-level questions in a specific case.

5.5. Summary

One of the foundational principles of forensic DNA analysis is that DNA transfers and persists (see Principle 2 in Chapter 2). This is what makes it possible to investigate crimes using DNA in the first place. However, this also means that the association of DNA to a crime cannot be taken for granted and needs to be assessed. Whether DNA is transferred directly or indirectly may affect its application to a criminal investigation. This is the obvious overall implication from the studies presented in the earlier part of this chapter.

Furthermore, an LR (or other statistic) produced by mixture interpretation methods given sub-source propositions considers only the rarity of the profiles. It does not say anything about whether the DNA is associated with the crime, and mixture profiles may well contain genotypes of individuals not connected to the crime. Therefore, it is important that an LR not be used in isolation. Instead, one must consider an assigned LR within the larger context of the case and ensure that stakeholders do not use the sub-source "number" alone as an indication of the contribution of DNA to the case (see Gill 2014, pp, 154-158; EWG 2024, pp. 172-182). In an article discussing a 2012 California case where Lukas Anderson's DNA was recovered from a murder victim's fingernail clipping and later discovered to have likely been transferred by paramedics to the victim with whom Anderson had no known contact, a prominent researcher of DNA transfer is quoted: "No one should ever rely solely on DNA evidence to judge what's going on" (Worth 2018).

6. Chapter 6: New Technologies: Potential and Limitations

New technologies are often investigated to assess whether they can provide solutions to existing problems in the forensic community. The adoption and implementation of these technologies depends upon a cost/benefit analysis within forensic laboratories. An appreciation of fundamental challenges in DNA mixture interpretation can provide an impetus to consider whether new approaches can bring desired improvements. The ability to analyze short tandem repeat alleles by sequence, in addition to length, promises to bring some new capabilities to forensic DNA laboratories. Next-generation sequencing platforms also enable additional genetic markers to be examined. Microhaplotypes have been pursued for their potential to improve DNA mixture interpretation. Additionally, cell separation techniques offer the potential to separate contributors prior to DNA extraction.

6.1. Technology Development and Drivers

Previous chapters have examined measurement and interpretation issues (Chapter 4) and case context and relevance for DNA mixtures (Chapter 5). This chapter explores the potential and limitations of new technologies to assist with DNA mixture interpretation.

As described in a supplemental document (NISTIR 8351sup1), DNA technologies (and interpretation approaches) have advanced over the past three decades. These advancements have been fueled largely due to ongoing efforts in biotechnology, specifically the commercialization of new instruments and techniques for clinical analysis and large-scale DNA sequencing efforts. Having multiple uses for a single technology allows commercial manufacturers to develop application-specific products with minimal risk. Thus, “piggy-backing” onto these broader advances provides capabilities to the forensic DNA community that would not be available otherwise. A prime example is the capillary electrophoresis (CE) technology that was developed for chemists to separate molecules according to size and charge, but also enabled the sequencing of billions of nucleotides for the Human Genome Project (Lander et al. 2001).

Over the past 25 years, CE technology has been the mainstay in forensic DNA laboratories around the world for separation and detection of short tandem repeat (STR) markers, starting with the ABI 310 Genetic Analyzer and then multi-capillary ABI 3100, 3130, and 3500 systems (Butler 2011, pp. 141-165). Some high-throughput forensic laboratories have also implemented the 3700 or 3730 Genetic Analyzers with 48 or 96 capillaries.

The polymerase chain reaction (PCR) is also used broadly in molecular biology, and forensic applications combine this method with fluorescently labeled primers to enable various configurations of STR typing kits. These kits have evolved both in terms of sensitivity and the number of targeted STR markers – the latter in keeping with increases to DNA database core sets (Gill et al. 2006a, Hares 2012, Hares 2015). Modern CE-based STR kits examine over 20 locations in the human genome from only a few cells (Butler 2011, Butler 2014). An increase in STR typing kit sensitivity improves detection of proportionally lower-level contributors in DNA mixtures, potentially resulting in a greater number of alleles in a mixed DNA sample. Although collecting more information is generally viewed as positive, examining additional data can add

to the complexity of interpretation and communication of results obtained from an electropherogram (EPG).

Millions of STR profiles, primarily single-source reference samples from convicted offenders or arrestees, now exist in national DNA databases around the world, with substantial resources invested to create these law enforcement databases. With increasing knowledge of the human genome, new genetic markers are being proposed for forensic identification purposes. This is described later in this chapter. However, adoption is challenging due to the existence of large STR profile databases (see Butler 2015). Before implementing a new technology, the degree of potential improvement needs to be considered in terms of the amount of information gained along with the cost and effort of changing.

The marketplace has played an important role in developing forensic DNA typing technology. The forensic DNA community uses commercial DNA extraction and quantification kits, STR typing kits, CE instruments for detection, and software for analysis and data interpretation (Figure 6.1). The adoption of commercially available options has led to more uniformity of methods employed in laboratories and consistent quality control. However, these same benefits can result in an increased reliance on ready-made solutions. This can result in lost opportunities for innovation.

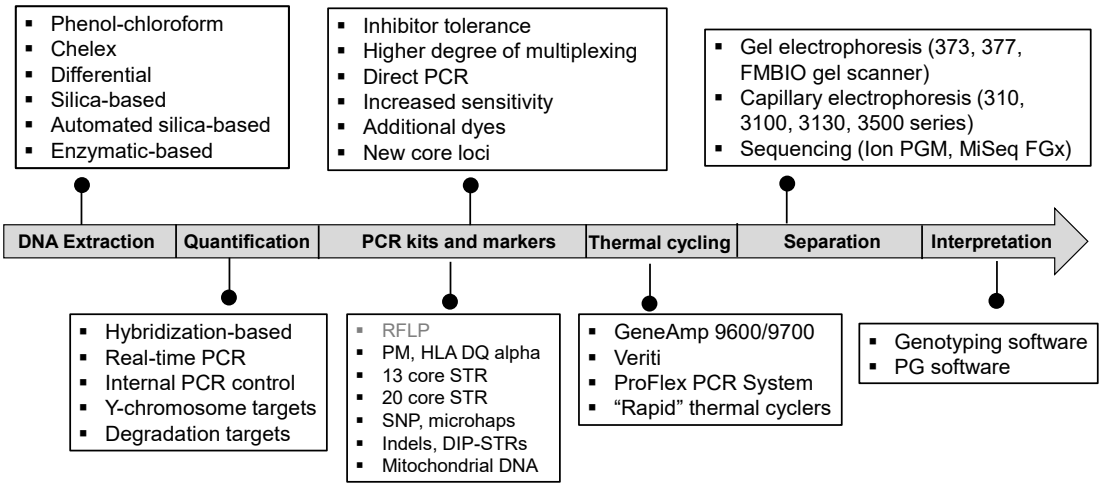

Figure 6.1. Advances and introduction of new technology to support the STR typing workflow.

Commercial suppliers must consider production and sales volume in deciding which products to develop and maintain in the marketplace. Thus, even if new technologies are developed, they may not be implemented in the forensic arena for reasons that can be either technology-based or market-driven. A proposed solution with a new technology may not sufficiently address the problem it is trying to solve to warrant change. A forensic laboratory determines whether the cost (including time and labor) of purchasing, training, performing internal validation experiments, implementing, and maintaining new procedures or equipment is expected to provide a satisfactory solution to an existing problem. While forensic laboratories can perform developmental validations for methods established in-house, most methods originate in the commercial sector where the vendor performs the developmental validation. Vendors often

collaborate with a forensic laboratory on the developmental validation, but most forensic laboratories are solely performing internal validation studies (see QAS 2020).

Adopting a new method or technology is not necessarily a linear process. Therefore, understanding the complexity of DNA mixture analysis and the way a new technology may or may not overcome known difficulties is important. Although a formal process for adoption and implementation does not exist, general steps can be considered. Table 6.1 lists considerations in deciding whether to adopt a new technology.

Table 6.1. Steps and considerations for implementing a new technology or method into practice.

Steps	Considerations
Research and Development	• Review work performed by commercial vendors or researchers • Seek input from technical working groups or previous adopters
Evaluation	• Perform informal studies (e.g., beta tests) • Examine early-stage publications describing the potential of the new technology performed by researchers or other practitioners
Decision to Move Forward	• Assess the "cost" (e.g., personnel time, new equipment) • Consider available funding for adoption • Weigh the changes and potential impact (e.g., adding new core loci, potential changes in vendor support) • Critically assess the benefits of the new technology to address issues and consider potential limitations
Internal Validation	• Examine published developmental validation studies (typically performed by the vendor) • Conduct internal validation studies • Perform additional supporting experiments as needed
Implementation	• Prepare standard operating procedures (SOPs), conduct training and competency testing, and establish proficiency testing and reporting/testimony guides
Other	• Evaluate whether additional documentary or physical standards are needed

6.2. Fundamental Mixture Challenges

In this section, the challenges are examined that are fundamental to DNA mixtures and areas of possible improvement via new technologies.

Sample collection, extraction, and quantitation are the first steps in the DNA measurement and interpretation workflow (see Figure 2.1 in Chapter 2). Improvements in DNA extraction efficiencies can help ensure maximal recovery of the evidence and, in theory, reduce the potential for stochastic variation observed with lower amounts of DNA (e.g., minor components in a mixture).

A DNA mixture arises when cells from multiple contributors are present in a sample. These cells are physically distinct prior to DNA extraction, but the DNA from those cells commingles and mixes during and after the extraction process (Figure 6.2). Thus, if cells from different contributors to a sample could be physically separated prior to extraction, then cells from each contributor could potentially be analyzed separately as a single-source sample. For example, chemical differences of the cell walls of sperm enable differential extraction to partition a sexual assault victim's epithelial cells from a perpetrator's sperm cells (Gill et al. 1985). However, when cells from multiple contributors are co-extracted, DNA mixtures result.

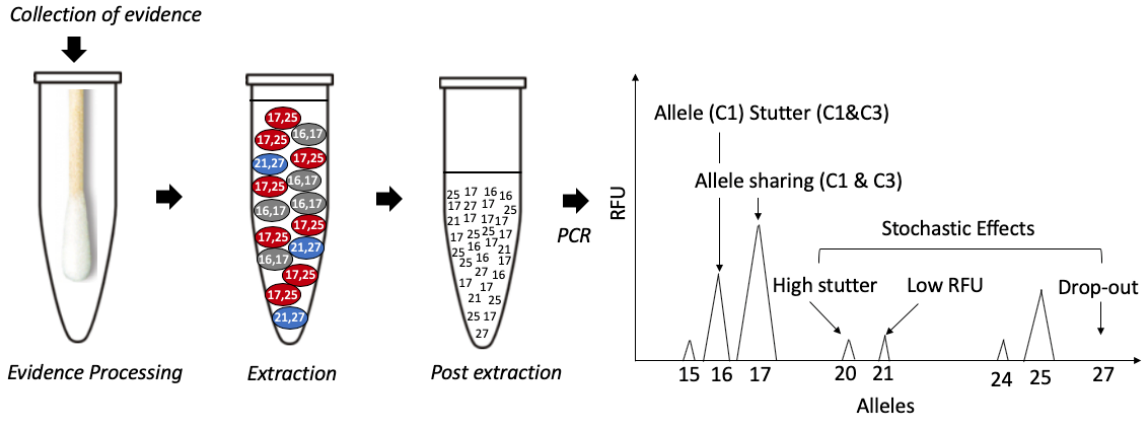

Figure 6.2. Illustration of steps involved in generating a DNA mixture profile and some of the possible factors in interpretation. If an evidentiary swab contains a mixture of cells from three contributors (Contributor 1 (C1) [grey], Contributor 2 (C2) [blue], Contributor 3 (C3) [red]) and the corresponding genotypes at one STR locus as an illustration are (Contributor 1 [16,17], Contributor 2 [21,27], Contributor 3 [17,25]), then allele sharing occurs with the "17" allele. If only a few cells are recovered for one or more of the mixture contributors, then stochastic effects, such as high stutter, heterozygote peak imbalance, and allele drop-out may occur.

From a measurement and interpretation standpoint, several challenges are fundamental to DNA mixture interpretation (see Chapter 2). Briefly, with any PCR system, there will be ***stochastic variation*** when small amounts of DNA are analyzed. Stochastic effects impact the recovery of alleles and genotypes from mixture samples and lead to uncertainty in assigning alleles to genotypes and genotypes to contributor profiles. When STR markers are examined, ***stutter products*** add noise to the system. Stutter products impact uncertainty when alleles from minor contributor(s) overlap with stutter peaks of alleles from major contributor(s). Use of non-repetitive genetic markers (described further in section 6.4.2) can avoid stutter products

but may not possess the genetic variation of STRs, which are needed to improve detection of genotypes from multiple contributors. Finally, ***sharing of common alleles*** can mask the presence of contributor alleles and affect the ability to estimate the number of contributors. When combined with stochastic variation and the existence of stutter products, allele sharing increases the complexity of a DNA mixture.

Allele sharing is illustrated in Figure 6.2 with allele 17 of Contributor 1 and Contributor 3. Stutter products (of allele 17) can also overlap an allele of the same length (allele 16). Stochastic effects can lead to high stutter (what appears to be an allele 20) and missing information (drop-out of allele 27). The illustration in Figure 6.2 does not account for further complications in the data caused by DNA degradation, PCR inhibitors, contamination (see Chapter 5), or cell-free DNA that may also be present in collected forensic evidence. STR allele sequencing technologies that rely on PCR amplification will still be subject to these fundamental mixture issues.

6.3. Possible Improvements: Physical Separation of Cells

Physically separating cells from different contributors prior to DNA extraction and STR typing can reduce the need for DNA mixture interpretation (Figure 6.3). This separation is an attractive concept but presents new challenges of working directly with cells prior to DNA extraction.

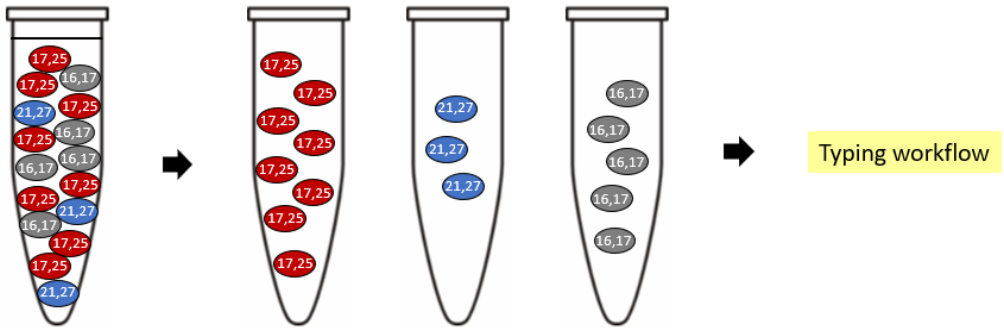

Figure 6.3. Illustration of physical separation and sorting of cells based on properties unique to a contributor's cell-type.

Separating cells from multiple contributors can sometimes be performed with laser-capture microdissection (Ballantyne et al. 2013) or micromanipulation (Farash et al. 2015). Cell separation can also be based on a unique property, such as the binding of a specific antibody to a unique feature on the cell surface (Verdon et al. 2015, Fontana et al. 2017). This type of work has included fluorescence-assisted cell sorting (FACS) methods and fluorescently labeled antibodies (Verdon et al. 2015, Dean et al. 2015, Stokes et al. 2018). Proof-of-concept research has been conducted, but the work is laborious and usually demonstrated on fresh samples.

In one micro-manipulation approach, 40 discrete "bio-particles" (20 single and 20 clumped cells) were collected under a microscope and subjected to PCR conditions optimized for low-level DNA detection, resulting in recovery of single-source STR profiles in 41% of the 479 tested samples (Farash et al. 2018). Another approach for recovering individual cells is the DEPArray system, which is an image-based, microfluidic digital sorter that can isolate pure cells (Fontana et al. 2017, Williamson et al. 2018). DNA profile recovery can also be improved through separating PCR inhibitors and DNA templates using a digital agarose droplet microfluidic approach (Geng & Mathies 2015). Similarly, agarose reactors can also allow for single-cell PCR within an encapsulated droplet (Geng et al. 2014).

One of the challenges of the FACS and microreactor methods is that crime scene evidence is typically composed of dried cells and may also contain cell-free DNA adhering to the outside of cells (Wang et al. 2017). The reconstitution of cells is not always straightforward, and it is important to maintain the integrity of the cell membrane to avoid mixing DNA from multiple cells. Dried cell membranes are more permeable and fragile, which may lead to cell breakage and DNA loss during sample preparation (Verdon et al. 2015). In addition to demonstrating success with samples subjected to real-world conditions, cell separation workflows would need to be streamlined prior to widespread adoption in the forensic laboratory.

Some examples of using single-cell analysis to reduce mixture sample complexities have been published since the draft of this report was released (Duffy et al. 2023, Grgicak et al. 2024, Huffman & Ballantyne 2023a, Huffman & Ballantyne 2023b, Huffman et al. 2023, Kulhankova et al. 2023, Kulhankova et al. 2024, Schulte et al. 2023, Schulte et al. 2024).

6.4. Possible Improvements: Sequencing

Next-generation sequencing (NGS), also known as massively parallel sequencing (MPS) in the forensic community, has been used for more than a decade to perform high-throughput DNA sequencing for biotechnology discovery purposes (Hert et al. 2008). NGS is widely described as important to the future of forensic DNA testing (Børsting & Morling 2015, Alonso et al. 2017, Alonso et al. 2018). Table 6.2 summarizes potential benefits and issues with the use of new sequencing technologies for DNA mixture interpretation. Compared to existing CE-based methods, NGS provides an additional dimension and more detailed resolution of genetic information, which includes the sequence of targeted PCR amplicons and accompanying stutter products with STR alleles.

In a 2015 review article, the authors stated:

> "Sequencing of complex and compound STRs with many alleles of the same size may simplify mixture interpretation, if the contributors have alleles of the same size with different sequence compositions or if the true allele of the minor contributor has a different sequence than the stutter artifact of the major contributor" (Børsting & Morling 2015).

Furthermore, the authors noted the difference between detecting alleles and distinguishing alleles from artifacts and noise:

> "It was recently demonstrated that sequences from the minor contributor in 1:100 or 1:50 mixtures were detectable by NGS – something that is not possible with the current PCR-CE technology. In these types of mixtures, the reads from the minor contributor will be difficult to separate from stutters and noise sequences, however, the mere fact that they could be identified opens up for new possibilities in mixture interpretation and it is certainly something that should be explored further" (Børsting & Morling 2015).

Table 6.2. Summary of the application of STR sequencing technologies to DNA mixtures.

Topics	Comments on Capabilities, Limitations, and Unknowns in Comparison to CE Methods
Smaller PCR Amplicons than CE	• Smaller and more consistently sized PCR products across STR loci (without the need to separate by size on an EPG) improve performance, particularly with degraded samples
Larger Multiplexes than CE, Potential Additional Markers	• Additional markers can be analyzed simultaneously to include more autosomal STRs, X and Y chromosome STRs, mitochondrial genome, single nucleotide polymorphisms (SNPs), and microhaplotypes • Additional information could potentially improve estimates for the number of contributors in a DNA mixture • Need to assess whether the observed mixture ratios of contributors are maintained across the examined loci
Targeted PCR Similar to CE	• Sensitivity similar to CE methods • Sequencers may tolerate a higher PCR DNA input than CE • Stochastic effects still present with low amounts of DNA
Different Artifacts from CE	• Fluorescent dye artifacts are not present (e.g., spurious EPG noise peaks, spectral “pull up”, or dye blobs) • Sequence-based artifacts may arise (e.g., homopolymers, phasing)
Different Determination of Thresholds	• Analytical thresholds, which discern noise sequences from biological sequences of STR alleles, are based on sequence data rather than CE molecule fluorescence
Sequenced Stutter Products	• Potential exists to discern a stutter product from a minor contributor allele if the allele sequence differs • Examination of the sequence context can allow a more accurate modeling of stutter product amounts • STR markers consisting of multiple repetitive regions may produce multiple stutter products per allele
Additional STR Alleles	• STR sequences may differentiate some identical-by-length STR alleles, separating some mixture components possessing shared alleles, which in turn may assist in an improved estimate of the number of contributors to the mixture • Not all STR loci experience significant gains from sequencing (e.g., TPOX, TH01) • Additional STR alleles requires sequence-based allele frequencies for statistical calculations • Sequenced STR alleles are compatible with current DNA databases using length-based STR information
Interpretation	• To take full advantage of sequencing capabilities for mixtures, an NGS-based probabilistic genotyping model will be required

Figure 6.4 illustrates the expected results from sequencing of the STR locus that was typed with CE methods and shown in Figure 6.2. Stochastic sampling effects similar to those encountered with CE data will continue to exist with amplified and sequenced low-template samples. For example, high stutter (from C2) and allele drop-out (27 allele of C2) are not addressed through sequencing, and allelic imbalances (not shown) could still impact the genotype determination of a contributor.

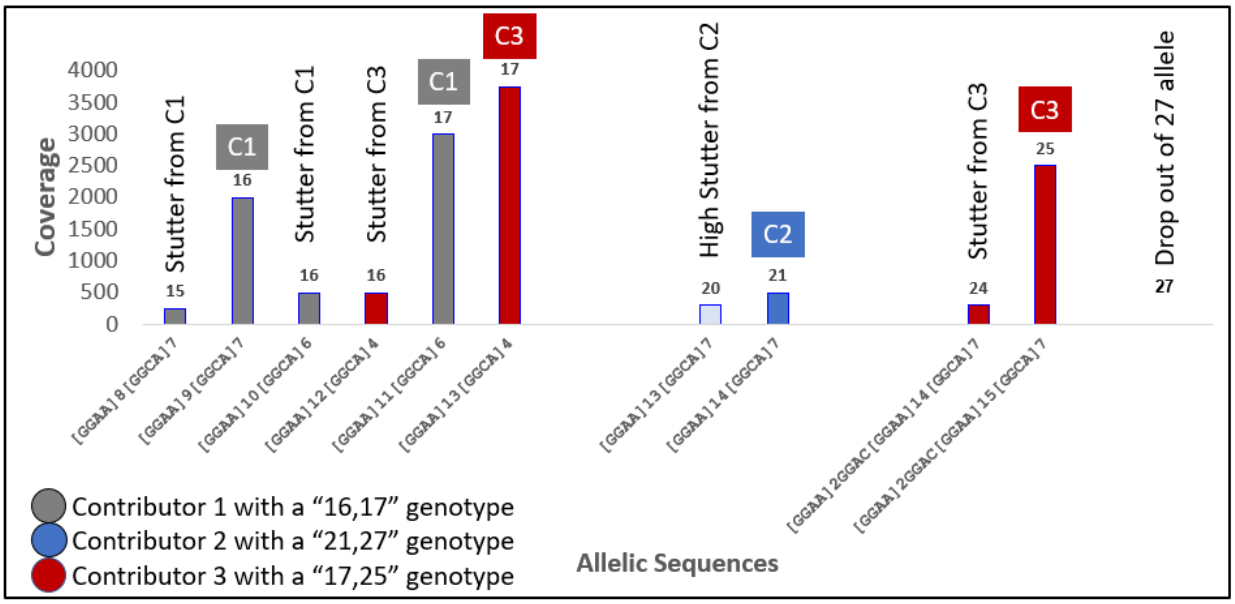

Figure 6.4. Illustration of results in "sequencing space" for the mixture example in Figure 6.2. The allele length and sequence are represented along the horizontal axis while relative sequence abundance (coverage) for the various alleles and stutter products is shown on the vertical axis. The same length "17" alleles from contributor 1 (C1) and contributor 3 (C3) can be resolved from one another. In addition, the stutter products from C1 and C3 can be separated by sequence from the "16" allele of C1.

Note that in Figure 6.4, the "17" allele (from C1 and C3) are distinguishable from one another through sequencing as are their corresponding "N-1" stutter products. In general, the degree of allele sharing is expected to decrease corresponding to an increase of observed alleles by sequencing, along with improved resolution and characterization of stutter artifacts. Each of these sequenced "17" alleles will have an associated sequenced-based allele frequency that would be applied in a statistical calculation, strengthening "matches" compared to a length-based STR analysis. The magnitude of the improvement will depend on the exact scenario and allele combinations, with gains expected primarily from the more complex STR markers, such as D12S391, D2S1338, and D21S11 (e.g., Gettings et al. 2018), as shown in sequenced-based allele frequency publications (summarized in Table 1 of Gettings et al. 2019). STR sequence-based nomenclature formats are under discussion (Parson et al. 2016, Phillips et al. 2018, Gettings et al. 2019), and will need to be determined to facilitate data exchange across laboratories (Gettings et al. 2024).

6.4.1. NGS Studies of STR Markers with DNA Mixtures

Because sequencing forensic STR markers is relatively new, much of the initial NGS mixture-related work in the literature consists of straightforward mixture *detection* experiments, rather than deconvolution with an associated statistical weight. These experiments can be thought of as "proof-of-concept" detection of the minor allele in a mixture to determine whether it is comparable to CE-based methods. This is not dissimilar to DNA mixture experiments designed for and performed in a developmental validation for CE-based methods (see NISTIR 8351sup2).

Table 6.3. Examples of factor space covered in two STR sequencing assay evaluations using two-person mixtures of various mixture ratios and input DNA amounts (Fordyce et al. 2015, van der Gaag et al. 2016).

Fordyce et al. 2015			van der Gaag et al. 2016	
Mixture Ratio	**Input DNA**		**Mixture Ratio**	**Input DNA**
1000 : 1	10 ng : 10 pg			
100 : 1	5 ng : 50 pg		99 : 1	5.94 ng : 60 pg
50 : 1	5 ng : 100 pg			
20 : 1	2 ng : 100 pg		19 : 1	1.14 ng : 60 pg
10 : 1	1 ng : 100 pg		9 : 1	540 pg : 60 pg
5 : 1	1 ng : 200 pg		4 : 1	400 pg : 100 pg
2 : 1	1 ng : 500 pg			
1 : 1	500 pg : 500 pg		1 : 1	250 pg : 250 pg

Using the Ion Torrent NGS platform (the Ion PGM) and a 10-plex assay consisting of amelogenin and mostly simple STR loci (CSF1PO, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, and D16S539), a 2015 Danish study examined two-person mixtures with eight mixture ratios (Fordyce et al. 2015), as shown in Table 6.3. Mixtures were easily deconvoluted down to 20:1 for the vWA and D3S1358 STR markers, although some minor contributor alleles were not identified by the associated software and required manual allele calling. This work also described stutter artifacts as a challenge:

> "The main factor hindering mixture deconvolution down to 100:1 was the stutters corresponding to the major contributor alleles. Hence, if stutters could be reduced, perhaps with an optimized PCR and possibly improved software, then it should be possible to deconvolute mixtures down to 100:1" (Fordyce et al. 2015).

In 2016, a group from The Netherlands used the PowerSeq assay to examine 45 mixtures, which consisted of five, two-person mixtures at ratios shown in Table 6.3 (van der Gaag et al. 2016). The amount of input DNA was inferred in this analysis and not explicit in the original text. For all the mixtures at all 22 PowerSeq STR markers, the authors state each allele for both contributors was detected in the expected ratio. Alleles in overlapping or stutter positions were not included in this analysis. The authors conclude:

> "When analysing alleles with abundance below 5% of the highest allele of the locus, additional PCR/sequence error variants were observed for several loci which can complicate the interpretation of a DNA sample. Therefore, the analysis of minor contributions of 5% or less in a mixture without prior knowledge of the ratio between the different donors, remains difficult for some, but not all loci, using the current experimental and analysis setup for this assay. Increasing the sequencing coverage increases the read counts of these artefacts as well and will not help to distinguish them from genuine alleles" (van der Gaag et al. 2016).

Published NGS studies have focused on simple two-person mixture examples in an effort to count the number of minor alleles detected in the mixture (e.g., Jäger et al. 2017). This is often reported for non-overlapping alleles between samples in the mixture and provides a general indicator of the minor allele detection capability. Full minor profiles are commonly detected at about 9:1 ratio range with allele drop-out starting to occur at the 19:1 level and greater (e.g., Alonso et al. 2018), which is essentially equivalent to CE-based methods used currently.

The need for robust thresholds to enable confident allele calling (e.g., Riman et al. 2020) and a systematic framework to account for sequenced stutter artifacts is often recommended. Research in these areas is underway in the community (Zeng et al. 2017, Alonso et al. 2018, Vilsen et al. 2018a, Vilsen et al. 2018b, Riman et al. 2019a) and should enable progress toward the goal of sequence-based interpretation. To date, the research has been largely proof-of-concept, and less effort has been spent on assigning a likelihood ratio or conducting a statistical analysis of results (e.g., Chan Mun Wei et al. 2018). As an understanding of sequence noise and sequence-specific stutter are developed (e.g., Just & Irwin 2018), this information can assist future NGS-specific models for probabilistic genotyping. The ability to *detect* alleles in a mixture is not the same as exploring the *interpretation capabilities* of NGS. These types of studies are still needed to understand the levels of measurement and interpretation errors that might occur.

Additional autosomal STR markers have been evaluated to ascertain their value in mixture detection based on sequence variation. Dozens of new highly polymorphic STRs have been identified (Tan et al. 2017, Novroski et al. 2018). In addition, *in-silico* analysis of two-, three-four-, and five-person mixtures was performed to rank the best STR markers for distinguishing alleles, which improved the estimates of the number of contributors in a mixture (Young et al. 2019).

Probabilistic genotyping software systems have added modules to assist with NGS data of DNA mixtures (e.g., Bleka et al. 2022, Cheng et al. 2023).

6.4.2. Alternate Markers

As described previously, the PCR amplification process for detecting STR alleles creates stutter product artifacts that interfere with unambiguous identification of minor contributors in an unbalanced mixture. Single nucleotide polymorphisms (SNPs) have been characterized for forensic use and explored to extend the capabilities of mixture interpretation. An important advantage of STR markers with mixture interpretation is the existence of many possible alleles within a population. This provides a greater chance of distinguishing multiple contributors from one another because of non-overlapping alleles compared to bi-allelic SNPs (Butler et al. 2007). SNPs and other alternative marker systems will be incompatible with existing STR databases. However, SNPs are amenable to array-based detection methods, which may be less expensive than STR sequencing for databasing single-source samples. While generally not used for samples containing DNA mixtures, array-based SNP genotyping data can also be used for genetic genealogy searches (Greytak et al. 2019).

The ability to examine many more markers in parallel has become possible because of the new sequencing technologies described previously. For example, a Danish research group using the Ion Torrent NGS platform examined a 169plex SNP typing assay with 11 two-person mixtures with ratios ranging from 1:1 to 1:1000 (Børsting et al. 2014). They were able to observe all minor contributor SNP types in a 1:100 mixture when the overall number of reads was sufficiently high to cross a detection threshold for the minor allele. Maintaining a signal balance across all of the tested markers becomes challenging when more markers are examined. In a proof-of-concept study of a probe capture method for 451 target SNPs, the authors indicate an expected ability to detect 85% to 100% of alleles unique to the minor contributor with two-person male-male mixtures from 10 ng of total DNA template (Bose et al. 2018). This study observed allele drop-out when the minor contributor was approximately 10% or less (Bose et al. 2018).

The multi-allelic possibilities of microhaplotype (MH) markers, which are defined by two or more closely linked SNPs within a single PCR product (Figure 6.5), extend the possibilities for DNA mixture interpretation (Kidd et al. 2014). MH markers tend to be less polymorphic than STRs, so a greater number may be needed for identification purposes. An attractive aspect of using microhaplotype markers with DNA mixtures is the lack of stutter artifacts during PCR amplification. Although the absence of stutter artifacts should reduce the complexity of the interpretation, PCR-related issues still occur and need to be addressed as part of interpretation. This includes measuring rates of allele drop-out and defining minimum signal thresholds.

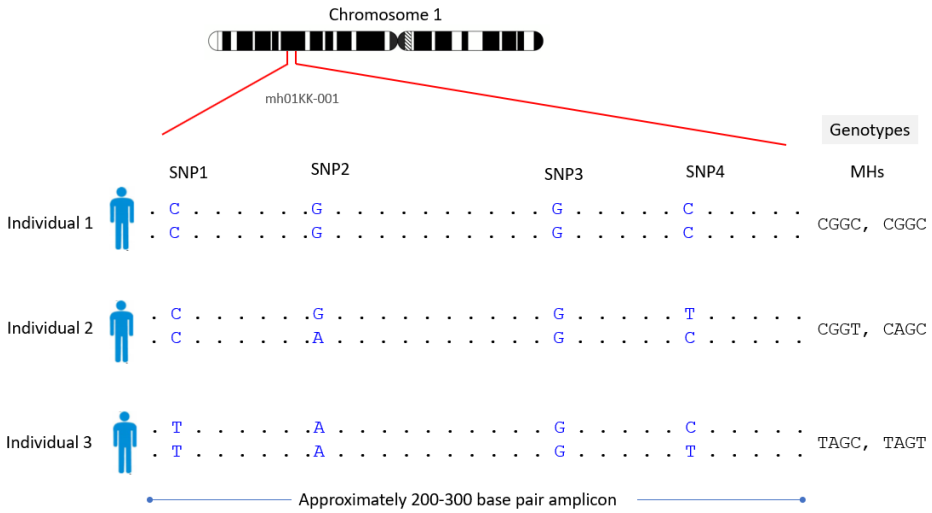

Figure 6.5. Schematic illustrating microhaplotypes in three individuals.

Different panels of microhaplotype markers have been developed by various research groups (e.g., van der Gaag et al. 2018, Chen et al. 2018, Voskoboinik et al. 2018, Bennett et al. 2019). Selection of standard MH markers and panels will be important as will more testing to explore the ability of these new markers to improve mixture interpretation in the future.

These research studies demonstrate the possibilities for new DNA markers to assist in mixture interpretation but will require much more extensive study before they can be incorporated into laboratory workflows. Although these new markers may be free of PCR stutter artifacts, stochastic effects will still exist with PCR-based approaches. These stochastic effects, combined with overlapping alleles when there are multiple contributors, will continue to make DNA mixture interpretation challenging when small amounts of DNA are examined.

6.5. Summary and Key Takeaways

The ultimate decision to implement new technologies in forensic laboratories should be driven by a real-use case and by those responsible for producing and reporting the information. A vendor or members of the general public may encourage forensic DNA laboratories to adopt a new approach or technology without appreciating the investments required to make a change.

KEY TAKEAWAY #6.1: Fundamental measurement and interpretation issues surrounding DNA mixtures, as described in Chapter 2, should be understood before attempting to apply a new technology.

Consideration needs to be given to whether supporting factors and resources will be available upon implementation. This includes allele frequencies, analysis software, interpretation methods, training, and support for potential admissibility hearings.

KEY TAKEAWAY #6.2: Implementation of new technologies requires a thorough understanding of the socio-technical benefits and limitations in addition to the required investment of time and effort put forth for its adoption by the laboratory.

An overall assessment is important and should include 1) how a new technology works, 2) what its limits are, 3) how it might specifically help improve a process or address a problem to be solved, and 4) whether this new technology can be justly and equitably implemented. This assessment is critical in evaluating whether implementation will be worthwhile.

References

[NISTIR 8351-draft PCs] Public comments received on NISTIR 8351-DRAFT DNA Mixture Interpretation: A NIST Scientific Foundation Review. Published December 3, 2021. https://www.nist.gov/system/files/documents/2021/12/03/Public-Comments-NISTIR8351draft_0.pdf (accessed August 23, 2024).

[AAFS 2019] DNA Mixture Interpretation Principles: Observations from a NIST Scientific Foundation Review. Workshop from 17 presenters at the American Academy of Forensic Scientists (Baltimore, MD) on February 18, 2019. Available at https://strbase-archive.nist.gov/AAFS2019-W10.htm (accessed October 8, 2024).

[AAFS 2021] MVPs of Forensic DNA: Examining the Most Valuable Publications in the Field. Virtual workshop presented by John Butler, Robin Cotton, Mechthild Prinz, and Charlotte Word at the American Academy of Forensic Sciences on February 16, 2021. Available at https://strbase.nist.gov/NIST_Resources/Presentations/2021/AAFS2021-W19-Handouts.pdf (accessed October 8, 2024).

[AAFS 2022] The National Institute of Standards and Technology (NIST) Forensic DNA Activities: Foundations, Research, and Standards. Hybrid workshop presented by John Butler, Peter Vallone, John Paul Jones, Katherine Gettings, Melissa Taylor, Sarah Riman, and Carolyn Steffen at the American Academy of Forensic Sciences on February 21, 2022. Available at https://strbase.nist.gov/NIST_Resources/Presentations/2022/AAFS2022-W2-NIST-Forensic-DNA-Activities-FINAL.pdf (accessed October 8, 2024).

[AAFS 2024] DNA Mixture Interpretation: History of the Field and NIST Foundation Study Findings. Presentation by John Butler at the American Academy of Forensic Sciences on February 23, 2024. Available at https://strbase.nist.gov/NIST_Resources/Presentations/2024/AAFS2024_Butler_DNA-Mixture-History-and-Foundation-Study.pdf (accessed October 8, 2024).

Abebe R, Hardt M, Jin A, Miller J, Schmidt L, Wexler R (2022) Adversarial scrutiny of evidentiary statistical software. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, pp. 1733-1746. https://doi.org/10.1145/3531146.3533228

Adamowicz MS, Rambo TN, Clarke JL (2022) Internal validation of MaSTR™ probabilistic genotyping software for the interpretation of 2-5 person mixed DNA profiles. *Genes* 13(8): 1429. https://doi.org/10.3390/genes13081429

Adams N, Koppl R, Krane D, Thompson W, Zabell S (2018) Letter to the Editor-Appropriate standards for verification and validation of probabilistic genotyping systems. *Journal of Forensic Sciences* 63(1): 339-340. https://doi.org/10.1111/1556-4029.13687

[AFSP 2009] Association of Forensic Science Providers (2009) Standards for the formulation of evaluative forensic science expert opinion. *Science & Justice* 49:161-164. https://doi.org/10.1016/j.scijus.2009.07.004

Aitken C, Nordgaard A, Taroni F, Biedermann A (2018) Commentary: Likelihood Ratio as Weight of Forensic Evidence: A Closer Look. *Frontiers in Genetics* 9:224. https://doi.org/10.3389/fgene.2018.00224

Aiken C & Nordgaard A (2018) Letter to the Editor-The roles of participants' differing background information in the evaluation of evidence. *Journal of Forensic Sciences* 63(2):648-649. https://doi.org/10.1111/1556-4029.13712

Alfonse LE, Tejada G, Swaminathan H, Lun DS, Grgicak CM (2017) Inferring the number of contributors to complex DNA mixtures using three methods: Exploring the limits of low-template DNA interpretation. *Journal of Forensic Sciences* 62(2):308-316. https://doi.org/10.1111/1556-4029.13284

Alfonse LE, Garrett AD, Lun DS, Duffy KR, Grgicak CM (2018) A large-scale dataset of single and mixed-source short tandem repeat profiles to inform human identification strategies: PROVEDIt. *Forensic Science International: Genetics* 32:62-70. https://doi.org/10.1016/j.fsigen.2017.10.006

Alladio E, Omedei M, Cisana S, D'Amico G, Caneparo D, Vincenti M, Garofano P (2018) DNA mixtures interpretation - A proof-of-concept multi-software comparison highlighting different probabilistic methods' performances on challenging samples. *Forensic Science International: Genetics* 37:143-150. https://doi.org/10.1016/j.fsigen.2018.08.002

Albright TD & Scurich N (2024) A call for open science in forensics. *Proceedings of the National Academy of Sciences of the United States of America* 121(24): e2321809121. https://doi.org/10.1073/pnas.2321809121

Almirall J, Arkes H, Lentini J, Mowrer F, Pawliszyn J (2017) *Forensic Science Assessments: A Quality and Gap Analysis – Fire Investigation.* American Association for the Advancement of Science. Available at https://www.aaas.org/resources/fire-investigation (accessed July 5, 2024).

Alonso A, Müller P, Roewer L, Willuweit S, Budowle B, Parson W (2017) European survey on forensic applications of massively parallel sequencing. *Forensic Science International: Genetics* 29:e23-e25. https://doi.org/10.1016/j.fsigen.2017.04.017

Alonso A, Barrio PA, Müller P, Köcher S, Berger B, Martin P, Bodner M, Willuweit S, Parson W, Roewer L, Budowle B (2018) Current state-of-art of STR sequencing in forensic genetics. *Electrophoresis* 39:2655-2668. https://doi.org/10.1002/elps.201800030

ANSI/ASB Standard 018 (2020) Standard for Validation of Probabilistic Genotyping Systems. Available at https://www.aafs.org/asb-standard/standard-validation-probabilistic-genotyping-systems (accessed July 1, 2024).

ANZPAA (2017) An introductory guide to evaluative reporting. Available at https://www.anzpaa.org.au/nifs/publications/general (accessed July 5, 2024).

Balding DJ & Nichols RA (1994) DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. *Forensic Science International* 64(2-3):125-140. https://doi.org/10.1016/0379-0738(94)90222-4

Balding DJ & Buckleton J (2009) Interpreting low template DNA profiles. *Forensic Science International: Genetics* 4(1):1-10. https://doi.org/10.1016/j.fsigen.2009.03.003

Balding DJ (2013) Evaluation of mixed-source, low-template DNA profiles in forensic science. *Proceedings of the National Academy of Sciences of the United States of America* 110(30):12241-12246. https://doi.org/10.1073/pnas.1219739110

Ballantyne J, Hanson EK, Perlin MW (2013) DNA mixture genotyping by probabilistic computer interpretation of binomially-sampled laser capture cell populations: combining quantitative data for greater identification information. *Science & Justice* 53:103-114. https://doi.org/10.1016/j.scijus.2012.04.004

Barrio PA, Crespillo M, Luque JA, Aler M, Baeza-Richer C, Baldassarri L, Carnevali E, Coufalova P, Flores I, García O, García MA, González R, Hernández A, Inglés V, Luque GM, Mosquera-Miguel A, Pedrosa S, Pontes ML, Porto MJ, Posada Y, Ramella MI, Ribeiro T, Riego E, Sala A, Saragoni VG, Serrano A, Vannelli S (2018) GHEP-ISFG collaborative exercise on mixture profiles (GHEP-MIX06). Reporting conclusions: results and evaluation. *Forensic Science International: Genetics* 35: 156-163. https://doi.org/10.1016/j.fsigen.2018.05.005

Basset P & Castella V (2018) Lessons from a study of DNA contaminations from police services and forensic laboratories in Switzerland. *Forensic Science International: Genetics* 33:147-154. https://doi.org/10.1016/j.fsigen.2017.12.012

Basset P & Castella V (2019) Positive impact of DNA contamination minimization procedures taken within the laboratory. *Forensic Science International: Genetics* 38:232-235. https://doi.org/10.1016/j.fsigen.2018.11.013

Bayes T (1763) An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions Royal Society of London* 53: 370-418.

Bennett L, Oldoni F, Long K, Cisana S, Maddela K, Wootton S, Chang J, Hasegawa R, Lagace R, Kidd KK, Podini D (2019) Mixture deconvolution by massively parallel sequencing of microhaplotypes. *International Journal of Legal Medicine* 133(3):719-729. https://doi.org/10.1007/s00414-019-02010-7

Benschop CC, van der Beek CP, Meiland HC, van Gorp AG, Westen AA, and Sijen, T. (2011) Low template STR typing: effect of replicate number and consensus method on genotyping reliability and DNA database search results. *Forensic Science International: Genetics* 5(4): 316-328. https://doi.org/10.1016/j.fsigen.2010.06.006

Benschop CC, Haned H, de Blaeij TJ, Meulenbroek AJ, Sijen T (2012) Assessment of mock cases involving complex low template DNA mixtures: A descriptive study. *Forensic Science International: Genetics* 6(6):697-707. https://doi.org/10.1016/j.fsigen.2012.04.007

Benschop C, Haned H, Sijen T (2013) Consensus and pool profiles to assist in the analysis and interpretation of complex low template DNA mixtures. *International Journal of Legal Medicine* 127(1):11-23. https://doi.org/10.1007/s00414-011-0647-5

Benschop CC, Haned H, Yoo SY, Sijen T (2015) Evaluation of samples comprising minute amounts of DNA. *Science & Justice* 55(5): 316-322. https://doi.org/10.1016/j.scijus.2015.05.002

Benschop CC, Connolly E, Ansell R, Kokshoorn B (2017) Results of an inter and intra laboratory exercise on the assessment of complex autosomal DNA profiles. *Science & Justice* 57(1): 21-27. https://doi.org/10.1016/j.scijus.2016.10.001

Berger CEH & Slooten K (2016) The LR does not exist. *Science & Justice* 56: 388-391. https://doi.org/10.1016/j.scijus.2016.06.005

Bhoelai B, de Jong B, Sijen T (2013) Can mRNA markers distinguish traces generated by different types of contact? *Forensic Science International: Genetics Supplement Series* 4(1):e3–e4. https://doi.org/10.1016/j.fsigss.2013.10.001

Bieber FR, Buckleton JS, Budowle B, Butler JM, Coble MD (2016) Evaluation of forensic DNA mixture evidence: protocol for evaluation, interpretation, and statistical calculations using the combined probability of inclusion. *BMC Genetics* 17(1):125. https://doi.org/10.1186/s12863-016-0429-7

Biedermann A & Taroni F (2012) Bayesian networks for evaluating forensic DNA profiling evidence: A review and guide to literature. *Forensic Science International: Genetics* 6(2):147-157. https://doi.org/10.1016/j.fsigen.2011.06.009

Biedermann A, Bozza S, Taroni F, Aitken C (2016a) Reframing the debate: a question of probability, not of likelihood ratio. *Science & Justice* 56:392-396. https://doi.org/10.1016/j.scijus.2016.05.008

Biedermann A, Champod C, Jackson G, Gill P, Taylor D, Butler J, Morling N, Hicks T, Vuille J, Taroni F (2016b) Evaluation of forensic DNA traces when propositions of interest relate to activities: Analysis and discussion of recurrent concerns. *Frontiers in Genetics* 7:215. https://doi.org/10.3389/fgene.2016.00215.

Bill M, Gill P, Curran J, Clayton T, Pinchin R, Healy M, Buckleton J (2005) PENDULUM – a guideline-based approach to the interpretation of STR mixtures. *Forensic Science International* 148:181-189. https://doi.org/10.1016/j.forsciint.2004.06.037

Bille T, Bright JA, Buckleton J (2013) Application of random match probability calculations to mixed STR profiles. *Journal of Forensic Sciences* 58(2):474-485. https://doi.org/10.1111/1556-4029.12067

Bille TW, Weitz SM, Coble MD, Buckleton J, Bright JA (2014) Comparison of the performance of different models for the interpretation of low level mixed DNA profiles. *Electrophoresis* 35:3125-3133. https://doi.org/10.1002/elps.201400110

Bille T, Coble MD, Bright JA (2022) Exploring the advantages of amplifying the entire extract versus splitting the extract and interpreting replicates using a continuous model of interpretation. *Australian Journal of Forensic Sciences* 54(5): 584-595. https://doi.org/10.1080/00450618.2021.1882568

Blake E, Mihalovich J, Higuchi R, Walsh PS, Erlich H (1992) Polymerase chain reaction (PCR) amplification and human leukocyte antigen (HLA)-DQ alpha oligonucleotide typing on biological evidence samples: casework experience. *Journal of Forensic Sciences* 37(3):700-726.

Bleka Ø, Storvik G, Gill P (2016a) EuroForMix: An open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts. *Forensic Science International: Genetics* 21:35-44. https://doi.org/10.1016/j.fsigen.2015.11.008

Bleka Ø, Benschop CC, Storvik G, Gill P (2016b) A comparative study of qualitative and quantitative models used to interpret complex STR DNA profiles. *Forensic Science International: Genetics* 25:85-96. https://doi.org/10.1016/j.fsigen.2016.07.016

Bleka Ø, Just R, Agudo MM, Gill P (2022) MPSproto: An extension of EuroForMix to evaluate MPS-STR mixtures. *Forensic Science International: Genetics* 61:102781. https://doi.org/10.1016/j.fsigen.2022.102781

Bolivar PA, Tracey M, McCord B (2016) Assessing the risk of secondary transfer via fingerprint brush contamination using enhanced sensitivity DNA analysis methods. *Journal of Forensic Sciences* 61(1):204–211. https://doi.org/10.1111/1556-4029.12911

Bonsu DOM, Higgins D, Austin JJ (2020) Forensic touch DNA recovery from metal surfaces - A review. *Science & Justice* 60(3):206-215. https://doi.org/10.1016/j.scijus.2020.01.002

Børsting C, Fordyce SL, Olofsson J, Mogensen HS, Morling N (2014) Evaluation of the Ion Torrent™ HID SNP 169-plex: A SNP typing assay developed for human identification by second generation sequencing. *Forensic Science International: Genetics* 12:144-154. https://doi.org/10.1016/j.fsigen.2014.06.004

Børsting C & Morling N (2015) Next generation sequencing and its applications in forensic genetics. *Forensic Science International: Genetics* 18:78-89. https://doi.org/10.1016/j.fsigen.2015.02.002

Boodoosingh S, Kelly H, Curran JM, Kalafut T (2024) An inter-laboratory comparison of probabilistic genotyping parameters and evaluation of performance on DNA mixtures from different laboratories. *Forensic Science International: Genetics* 71:103046. https://doi.org/10.1016/j.fsigen.2024.103046

Bose N, Carlberg K, Sensabaugh G, Erlich H, Calloway C (2018) Target capture enrichment of nuclear SNP markers for massively parallel sequencing of degraded and mixed samples. *Forensic Science International: Genetics* 34:186-196. https://doi.org/10.1016/j.fsigen.2018.01.010

Bouzga MM, Dorum G, Gundersen K, Kohler P, Hoff-Olsen P, Fonnelop AE (2020) Is it possible to predict the origin of epithelial cells? - A comparison of secondary transfer of skin epithelial cells versus vaginal mucous membrane cells by direct contact. *Science & Justice* 60(30):234-242. https://doi.org/10.1016/j.scijus.2020.02.003

Bowman ZE, Mosse KSA, Sungaila AM, van Oorschot RAH, Hartman D (2018) Detection of offender DNA following skin-to-skin contact with a victim. *Forensic Science International: Genetics* 37:252-259. https://doi.org/10.1016/j.fsigen.2018.09.005

Bozzo WR, Colussi AG, Ortíz MI, Laborde L, Pilili JP, Carini G, Lojo MM (2015) Analysis of DNA from fingernail samples in criminal cases. *Forensic Science International: Genetics Supplement Series* 5:e601-e602. https://doi.org/10.1016/j.fsigss.2015.09.237

Brayley-Morris H, Sorrell A, Revoir AP, Meakin GE, Syndercombe Court D, Morgan RM (2015) Persistence of DNA from laundered semen stains: Implications for child sex trafficking cases. *Forensic Science International: Genetics* 19:165-171. https://doi.org/10.1016/j.fsigen.2015.07.016

Breathnach M, Williams L, McKenna L, Moore E (2016) Probability of detection of DNA deposited by habitual wearer and/or the second individual who touched the garment. *Forensic Science International: Genetics* 20:53-60. https://doi.org/10.1016/j.fsigen.2015.10.001

Bregu J, Conklin D, Coronado E, Terrill M, Cotton RW, Grgicak CM (2013) Analytical thresholds and sensitivity: establishing RFU thresholds for forensic DNA analysis. *Journal of Forensic Sciences* 58(1):120-129. https://doi.org/10.1111/1556-4029.12008

Bright JA & Curran JM (2014) Investigation into stutter ratio variability between different laboratories. *Forensic Science International: Genetics* 13:79-81. https://doi.org/10.1016/j.fsigen.2014.07.003

Bright JA, Evett IW, Taylor D, Curran JM, Buckleton J (2015) A series of recommended tests when validating probabilistic DNA profile interpretation software. *Forensic Science International: Genetics* 14:125-131. https://doi.org/10.1016/j.fsigen.2014.09.019

Bright JA, Taylor D, McGovern C, Cooper S, Russell L, Abarno D, Buckleton J (2016) Developmental validation of STRmix, expert software for the interpretation of forensic DNA profiles. *Forensic Science International: Genetics* 23:226-239. https://doi.org/10.1016/j.fsigen.2016.05.007

Bright JA, Cheng K, Kerr Z, McGovern C, Kelly H, Moretti TR, Smith MA, Bieber FR, Budowle B, Coble MD, Alghafri R, Allen PS, Barber A, Beamer V, Buettner C, Russell M, Gehrig C, Hicks T, Charak J, Cheong-Wing K, Ciecko A, Davis CT, Donley M, Pedersen N, Gartside B, Granger D, Greer-Ritzheimer M, Reisinger E, Kennedy J, Grammer E, Kaplan M, Hansen D, Larsen HJ, Laureano A, Li C, Lien E, Lindberg E, Kelly C, Mallinder B, Malsom S, Yacovone-Margetts A, McWhorter A, Prajapati SM, Powell T, Shutler G, Stevenson K, Stonehouse AR, Smith L, Murakami J, Halsing E, Wright D, Clark L, Taylor DA, Buckleton J (2019) STRmix™ collaborative exercise on DNA mixture interpretation. *Forensic Science International: Genetics* 40:1-8. https://doi.org/10.1016/j.fsigen.2019.01.006

Bright J-A & Coble M (2020) *Forensic DNA Profiling: A Practical Guide to Assigning Likelihood Ratios*. CRC Press: Boca Raton.

Bright JA, Jones Dukes M, Pugh SN, Evett IW, Buckleton JS (2021) Applying calibration to *LR*s produced by a DNA interpretation software. *Australian Journal of Forensic Sciences* 53(2):147-153. https://doi.org/10.1080/00450618.2019.1682668

Brinkac LM, Richetelli N, Davoren JM, Bever RA, Hicklin RA (2023) DNAmix 2021: Laboratory policies, procedures, and casework scenarios summary and dataset. *Data in Brief* 48:109150. https://doi.org/10.1016/j.dib.2023.109150

Brookes C, Bright JA, Harbison S, Buckleton J (2012) Characterising stutter in forensic STR multiplexes. *Forensic Science International: Genetics* 6(1):58-63. https://doi.org/10.1016/j.fsigen.2011.02.001

Buckingham AK, Harvey ML, van Oorschot RAH (2016) The origin of unknown source DNA from touched objects. *Forensic Science International: Genetics* 25:26-33. https://doi.org/10.1016/j.fsigen.2016.07.015

Buckleton J, Triggs CM, Walsh SJ (Editors) (2005) *Forensic DNA Evidence Interpretation*. CRC Press: Boca Raton.

Buckleton JS, Curran JM, Gill P (2007) Towards understanding the effect of uncertainty in the number of contributors to DNA stains. *Forensic Science International: Genetics* 1(1):20-28. https://doi.org/10.1016/j.fsigen.2006.09.002

Buckleton J & Curran J (2008) A discussion of the merits of random man not excluded and likelihood ratios. *Forensic Science International: Genetics* 2(4):343-348. https://doi.org/10.1016/j.fsigen.2008.05.005

Buckleton J, Bright JA, Taylor D, Evett I, Hicks T, Jackson G, Curran JM (2014) Helping formulate propositions in forensic DNA analysis. *Science & Justice* 54(4):258-261. https://doi.org/10.1016/j.scijus.2014.02.007

Buckleton JS, Bright J-A, Taylor D (Editors) (2016) *Forensic DNA Evidence Interpretation, Second Edition*. CRC Press: Boca Raton.

Buckleton JS, Bright J-A, Cheng K, Budowle B, Coble MD (2018) NIST Interlaboratory studies involving DNA mixtures (MIX13): a modern analysis. *Forensic Science International: Genetics* 37:172-179. https://doi.org/10.1016/j.fsigen.2018.08.014

Buckleton J, Robertson B, Curran J, Berger C, Taylor D, Bright JA, Hicks T, Gittelson S, Evett I, Pugh S, Jackson G, Kelly H, Kalafut T, Bieber FR (2020) A review of likelihood ratios in forensic science based on a critique of Stiffelman "No longer the Gold standard: Probabilistic genotyping is changing the nature of DNA evidence in criminal trials". *Forensic Science International* 310: 110251. https://doi.org/10.1016/j.forsciint.2020.110251

Buckleton J, Bright JA, Taylor D, Wivell R, Bleka Ø, Gill P, Benschop C, Budowle B, Coble M (2022) Re: Riman et al. Examining performance and likelihood ratios for two likelihood ratio systems using the PROVEDIt dataset. *Forensic Science International: Genetics* 59: 102709. https://doi.org/10.1016/j.fsigen.2022.102709

Buckleton J, Susik M, Curran JM, Cheng K, Taylor D, Bright JA, Kelly H, Wivell R (2024a) A diagnosis of the primary difference between EuroForMix and STRmix™. *Journal of Forensic Sciences* 69(1):40–51. https://doi.org/10.1111/1556-4029.15387

Buckleton J, Bright JA, Taylor D, Curran J, Kalafut T (2024b) Extending the discussion on inconsistency in forensic decisions and results. *Journal of Forensic Sciences* 69(4):1125-1137. https://doi.org/10.1111/1556-4029.15558

Budowle B, Moretti TR, Niezgoda SJ, Brown BL (1998). CODIS and PCR-based short tandem repeat loci: law enforcement tools. *Proceedings of the Second European Symposium on Human Identification* (pp. 73–88). Madison, Wisconsin: Promega Corporation. Available at https://www.promega.com/-/media/files/resources/conference-proceedings/ishi-02/oral-presentations/17.pdf?la=en (accessed July 1, 2024).

Budowle B, Onorato AJ, Callaghan TF, Della Manna A, Gross AM, Guerrieri RA, Luttman JC, McClure DL (2009) Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework. *Journal of Forensic Sciences* 54(4):810-821. https://doi.org/10.1111/j.1556-4029.2009.01046.x

Burrill J, Daniel B, Frascione N (2019) A review of trace "touch DNA" deposits: Variability factors and an exploration of cellular composition. *Forensic Science International: Genetics* 39:8-18. https://doi.org/10.1016/j.fsigen.2018.11.019

Burrill J, Daniel B, Frascione N (2020) Illuminating touch deposits through cellular characterization of hand rinses and body fluids with nucleic acid fluorescence. *Forensic Science International: Genetics* 46:102269. https://doi.org/10.1016/j.fsigen.2020.102269

Butler JM, Buel E, Crivellente F, McCord BR (2004). Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. *Electrophoresis* 25:1397-1412. https://doi.org/10.1002/elps.200305822

Butler JM (2006) Debunking some urban legends surrounding validation within the forensic DNA community. *Profiles in DNA* 9(2):3-6. Available at https://www.promega.com/resources/profiles-in-dna/2006/debunking-some-urban-legends-surrounding-validation-within-the-forensic-dna-community/ (accessed November 4, 2024).

Butler JM (2007) Short tandem repeat typing technologies used in human identity testing. *Biotechniques* 43(4):Sii-Sv. https://doi.org/10.2144/000112582

Butler JM, Coble MD, Vallone PM (2007) STRs vs. SNPs: thoughts on the future of forensic DNA testing. *Forensic Science, Medicine and Pathology* 3(3):200-205. https://doi.org/10.1007/s12024-007-0018-1

Butler JM (2009) *Fundamentals of Forensic DNA Typing*. Elsevier Academic Press: San Diego.

Butler JM & Hill CR (2010) Scientific issues with analysis of low amounts of DNA. *Profiles in DNA,* 13(1). Available at https://www.promega.com/resources/profiles-in-dna/2010/scientific-issues-with-analysis-of-low-amounts-of-dna/ (accessed July 1, 2024).

Butler JM (2011) *Advanced Topics in Forensic DNA Typing: Methodology*. Elsevier Academic Press: San Diego.

Butler JM (2014) *Advanced Topics in Forensic DNA Typing: Interpretation*. Elsevier Academic Press: San Diego.

Butler JM (2015) The future of forensic DNA analysis. *Philosophical Transactions of the Royal Society B* 370:20140252. https://doi.org/10.1098/rstb.2014.0252

Butler JM, Kline MC, Coble MD (2018a) NIST interlaboratory studies involving DNA mixtures (MIX05 and MIX13): variation observed and lessons learned. *Forensic Science International: Genetics* 37: 81-94. https://doi.org/10.1016/j.fsigen.2018.07.024

Butler JM, Iyer H, Press R, Taylor M, Vallone PM, Willis S (2018b) DNA mixture interpretation principles: insights from the NIST scientific foundation review. *Proceedings of the 29th International Symposium on Human Identification.* Available at https://www.promega.com/products/pm/genetic-identity/ishi-conference-proceedings/29th-ishi-oral-presentations/ (accessed July 5, 2024).

Butler JM, Iyer H, Press R, Taylor M, Vallone PM, Willis S (2019) NIST scientific foundation review on DNA mixture interpretation. Poster presented at the International Society for Forensic Genetics. Available at https://strbase-archive.nist.gov/pub_pres/ButlerISFG2019poster.pdf (accessed July 5, 2024).

Butler JM & Willis S (2020) INTERPOL review of forensic biology and forensic DNA typing 2016-2019. *Forensic Science International: Synergy* 2:352-367. https://doi.org/10.1016/j.fsisyn.2019.12.002. Presentation available at https://strbase-archive.nist.gov/pub_pres/Butler-INTERPOL-DNAreview-Oct2019.pdf (accessed July 1, 2024).

Butler JM (2023) Recent advances in forensic biology and forensic DNA typing: INTERPOL review 2019-2022. *Forensic Science International: Synergy* 6:100311. https://doi.org/10.1016/j.fsisyn.2022.100311

Cale CM, Earll ME, Latham KE, Bush GL (2016) Could secondary DNA transfer falsely place someone at the scene of a crime? *Journal of Forensic Sciences* 61(1):196-203. https://doi.org/10.1111/1556-4029.12894

Casey DG, Clayson N, Jones S, Lewis J, Boyce M, Fraser I, Kennedy F, Alexender K (2016) Correspondence: A response to Meakin and Jamieson DNA transfer: Review and implications for casework. *Forensic Science International: Genetics* 21:117-118. https://doi.org/10.1016/j.fsigen.2015.12.013

Castella V& Mangin P (2008) DNA profiling success and relevance of 1739 contact stains from caseworks. *Forensic Science International: Genetics Supplement Series* 1:405-407. https://doi.org/10.1016/j.fsigss.2007.10.071

Cerri N, Verzeletti A, Cortellini V, Cincotta A, De Ferrari F (2009) Prevalence of mixed DNA profiles in fingernail swabs from autoptic cases. *Forensic Science International: Genetics Supplement Series* 2(1):163-164. https://doi.org/10.1016/j.fsigss.2009.08.015

[CFF 2024] Workshop on Communicating Forensic Findings: Current Practices and Future Directions held June 25-26, 2024. Available at https://www.nist.gov/news-events/events/2024/06/communicating-forensic-findings-current-practices-and-future-directions (accessed October 11, 2024).

Chambers GK, Cordiner SJ, Buckleton JS, Robertson B, Vignaux GA (1997) Forensic DNA profiling: the importance of giving accurate answers to the right questions. *Criminal Law Forum* 8(3): 445-459.

Champod C (2013) DNA transfer: Informed judgment or mere guesswork? *Frontiers in Genetics* 4:300. https://doi.org/10.3389/fgene.2013.00300.

Chan Mun Wei J, Zhao Z, Li SC, Ng YK (2018) NGS-based likelihood ratio for identifying contributors in two- and three-person DNA mixtures. *Computational Biology and Chemistry* 74:428-433. https://doi.org/10.1016/j.compbiolchem.2018.03.010

Chen P, Yin C, Li Z, Pu Y, Yu Y, Zhao P, Chen D, Liang W, Zhang L, Chen F (2018) Evaluation of the microhaplotypes panel for DNA mixture analyses. *Forensic Science International: Genetics* 35:149-155. https://doi.org/10.1016/j.fsigen.2018.05.003

Cheng K, Bleka Ø, Gill P, Curran J, Bright J-A, Taylor D, Buckleton J (2021) A comparison of likelihood ratios obtained from EuroForMix and STRmix. *Journal of Forensic Sciences* 66(6):2138-2155. https://doi.org/10.1111/1556-4029.14886

Cheng K, Bright JA, Kelly H, Liu YY, Lin MH, Kruijver M, Taylor D, Buckleton J (2023) Developmental validation of STRmix™ NGS, a probabilistic genotyping tool for the interpretation of autosomal STRs from forensic profiles generated using NGS. *Forensic Science International: Genetics* 62:102804. https://doi.org/10.1016/j.fsigen.2022.102804

Clayton TM, Whitaker JP, Sparkes R, Gill P (1998) Analysis and interpretation of mixed forensic stains using DNA STR profiling. *Forensic Science International* 91(1):55-70. https://doi.org/10.1016/s0379-0738(97)00175-8

Coble MD, Bright JA, Buckleton JS, Curran JM (2015) Uncertainty in the number of contributors in the proposed new CODIS set. *Forensic Science International: Genetics* 19:207-211. https://doi.org/10.1016/j.fsigen.2015.07.005

Coble MD & Bright J-A (2019) Probabilistic genotyping software: an overview. *Forensic Science International: Genetics* 38:219-224. https://doi.org/10.1016/j.fsigen.2018.11.009

Cook R, Evett IW, Jackson G, Jones PJ, Lambert JA (1998a) A model for case assessment and interpretation. *Science & Justice* 38(3):151-156. https://doi.org/10.1016/S1355-0306(98)72099-4

Cook R, Evett IW, Jackson G, Jones PJ, Lambert JA (1998b) A hierarchy of propositions: Deciding which level to address in casework. *Science & Justice* 38(4):231-239. https://doi.org/10.1016/S1355-0306(98)72117-3

Costa C, Figueiredo C, Amorim A, Costa S, Miguel Ferreira P, Pinto N (2022) Quantification of forensic genetic evidence: Comparison of results obtained by qualitative and quantitative software for real casework samples. *Forensic Science International: Genetics* 59: 102715. https://doi.org/10.1016/j.fsigen.2022.102715

Cotton EA, Allsop RF, Guest JL, Frazier RR, Koumi P, Callow IP, Seager A, Sparkes RL (2000) Validation of the AMPFlSTR SGM plus system for use in forensic casework. *Forensic Science International* 112(2-3):151-161. https://doi.org/10.1016/s0379-0738(00)00182-1

Cowell RG (2009) Validation of an STR peak area model. *Forensic Science International: Genetics* 3(3):193-199. https://doi.org/10.1016/j.fsigen.2009.01.006

Cowell RG, Graversen T, Lauritzen SL, Mortera J (2015) Analysis of forensic DNA mixtures with artefacts. *Journal of the Royal Statistical Society Applied Statistics Series C* 64(1):1-48. Available at https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/rssc.12071 (accessed July 5, 2024).

Crespillo M, Barrio PA, Luque JA, Alves C, Aler M, Alessandrini F, Andrade L, Barretto RM, Bofarull A, Costa S, García MA, García O, Gaviria A, Gladys A, Gorostiza A, Hernández A, Herrera M, Hombreiro L, Ibarra AA, Jiménez MJ, Luque GM, Madero P, Martínez-Jarreta B, Masciovecchio MV, Modesti NM, Moreno F, Pagano S, Pedrosa S, Plaza G, Prat E, Puente J, Rendo F, Ribeiro T, Sala A, Santamaría E, Saragoni VG, Whittle MR (2014) GHEP-ISFG collaborative exercise on mixture profiles of autosomal STRs (GHEP-MIX01, GHEP-MIX02 and GHEP-MIX03): results and evaluation. *Forensic Science International: Genetics* 10:64-72. https://doi.org/10.1016/j.fsigen.2014.01.009

Crow JF (1999) Hardy, Weinberg and language impediments. *Genetics* 152:821-825. https://doi.org/10.1093/genetics/152.3.821

[CSAFE 2019] DNA Mixture Interpretation: Thoughts and Lessons Learned from a NIST Scientific Foundation Review. Webinar presented by John Butler and Hari Iyer and organized by the Center for Statistics and Applications in Forensic Evidence. Available at https://forensicstats.org/blog/portfolio/dna-mixture-interpretation-thoughts-and-lessons-learned-from-a-nist-scientific-foundation-review/ (accessed October 8, 2024).

Curran JM (2008) A MCMC method for resolving two person mixtures. *Science & Justice* 48(4):168-177. https://doi.org/10.1016/j.scijus.2007.09.014

Curran JM (2016) Admitting to uncertainty in the LR. *Science & Justice* 56:380-382. https://doi.org/10.1016/j.scijus.2016.05.005

[DAB 2000] DNA Advisory Board (2000) Statistical and population genetics issues affecting the evaluation of the frequency of occurrence of DNA profiles calculated from pertinent population database(s). *Forensic Sci. Comm.* 2(3). Available at https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/july2000/dnastat.htm (accessed July 1, 2024).

Daly DJ, Murphy C, McDermott SD (2012) The transfer of touch DNA from hands to glass, fabric and wood. *Forensic Science International: Genetics* 6:41-46. https://doi.org/10.1016/j.fsigen.2010.12.016

Dang HW, Mao J, Wang H, Huang JP, Bai XG (2012) Research status and prospects of DNA test on difficult specimens. *Fa Yi Xue Za Zhi* 28:52-54.

Daniel R & van Oorschot RAH (2011) An investigation of the presence of DNA on unused laboratory gloves. *Forensic Science International: Genetics Supplement Series* 3:e45-e46. https://doi.org/10.1016/j.fsigss.2011.08.022

Davies C, Thomson J, Kennedy F (2015) Assessing primary, secondary and tertiary DNA transfer using the Promega ESI-17 Fast PCR chemistry. *Forensic Science International: Genetics Supplement Series* 5:e55-e57. https://doi.org/10.1016/j.fsigss.2015.09.022

Dean L, Kwon YJ, Philpott MK, Stanciu CE, Seashols-Williams SJ, Cruz TD, Sturgill J, Ehrhardt CJ (2015) Separation of uncompromised whole blood mixtures for single source STR profiling using fluorescently-labeled human leukocyte antigen (HLA) probes and fluorescence activated cell sorting (FACS). *Forensic Science International: Genetics* 17:8-16. https://doi.org/10.1016/j.fsigen.2015.03.003

Debernardi A, Suzanne E, Formant A, Pène L, Dufour AB, Lobry JR (2011) One year variability of peak heights, heterozygous balance and inter-locus balance for the DNA positive control of AmpFℓSTR© Identifiler© STR kit. *Forensic Science International: Genetics* 5(1):43-49. https://doi.org/10.1016/j.fsigen.2010.01.020

de Koeijer JA, Sierps MJ, Vergeer P, Berger CEH (2020) Combining evidence in complex cases – a practical approach to interdisciplinary casework. *Science & Justice* 60(1):20-29. https://doi.org/10.1016/j.scijus.2019.09.001

Dembinski GM, Sobieralski C, Picard CJ (2018) Estimation of the number of contributors of theoretical mixture profiles based on allele counting: Does increasing the number of loci increase success rate of estimates? *Forensic Science International: Genetics* 33:24-32. https://doi.org/10.1016/j.fsigen.2017.11.007

Digréus P, Andersson AC, Nordgaard A, Ansell R (2011) Contamination monitoring in the forensic DNA laboratory and a simple model for unbiased EPG classification. *Forensic Science International: Genetics Supplement Series* 3:e299-e300. https://doi.org/10.1016/j.fsigss.2011.09.012

Djuric M, Varljen T, Stanojevic A, Stojkovic O (2008) DNA typing from handled items. *Forensic Science International: Genetics Supplement Series* 1:411-412. https://doi.org/10.1016/j.fsigss.2007.10.161

dMIQE Group (2020) The digital MIQE guidelines update: Minimum information for publication of quantitative digital PCR experiments for 2020. *Clinical Chemistry* 66(8):1012-1029. https://doi.org/10.1093/clinchem/hvaa125

DNA Identification Act of 1994 (1994) Public Law No. 103-322. Available at https://www.govinfo.gov/content/pkg/STATUTE-108/pdf/STATUTE-108-Pg1796.pdf (accessed August 5, 2024).

Dror IE (2023) The most consistent finding in forensic science is inconsistency. *Journal of Forensic Sciences* 68(6):1851-1855. https://doi.org/10.1111/1556-4029.15369

Duewer DL, Kline MC, Redman JW, Newall PJ, Reeder DJ (2001) NIST Mixed Stain Studies #1 and #2: interlaboratory comparison of DNA quantification practice and short tandem repeat multiplex performance with multiple-source samples. *Journal of Forensic Sciences* 46:1199-1210.

Duffy KR, Lun DS, Mulcahy MM, O'Donnell L, Sheth N, Grgicak CM (2023) Evidentiary evaluation of single cells renders highly informative forensic comparisons across multifarious admixtures. *Forensic Science International: Genetics* 64:102852. https://doi.org/10.1016/j.fsigen.2023.102852

Duke K, Cuenca D, Myers S, Wallin J (2022) Compound and conditioned likelihood ratio behavior within a probabilistic genotyping context. *Genes* 13(11): 2031. https://doi.org/10.3390/genes13112031

Durdle A, van Oorschot RAH, Mitchell RJ (2009) The transfer of human DNA by *Lucilia cuprina* (Meigen) (Diptera: Calliphoridae). *Forensic Science International: Genetics Supplement Series* 2(1):180–182. https://doi.org/10.1016/j.fsigss.2009.08.095

Durdle A, Mitchell RJ, van Oorschot RAH (2011) The change in human DNA content over time in the artefacts of the blowfly *Lucilia cuprina* (Meigen) (Diptera: Calliphoridae). *Forensic Science International: Genetics Supplement Series* 3(1):e289–e290. https://doi.org/10.1016/j.fsigss.2011.09.007

Dziak R, Peneder A, Buetter A, Hageman C (2018) Trace DNA sampling success from evidence items commonly encountered in forensic casework. *Journal of Forensic Sciences* 63(3):835–841. https://doi.org/10.1111/1556-4029.13622

Eldridge H (2019) Juror comprehension of forensic expert testimony: A literature review and gap analysis. *Forensic Science International: Synergy* 1:24-34. https://doi.org/10.1016/j.fsisyn.2019.03.001

[ENFSI 2015] Willis S, McKenna L, McDermott S, O'Donell G, Barrett A, Rasmusson B, Nordgaard A, Berger CEH, Sjerps MJ, Lucena-Molina JJ, Zadora G, Aitken C, Lovelock T, Lunt L, Champod C, Biedermann A, Hicks TN, Taroni F (2015) *ENFSI Guideline for Evaluative Reporting in Forensic Science*. Available at http://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf (accessed August 5, 2024).

[ENFSI 2022] ENFSI Best Practice Manual for Human Forensic Biology and DNA Profiling. Available at https://enfsi.eu/wp-content/uploads/2022/12/ENFSI-DNA-BPM-03.pdf (accessed August 8, 2024).

Ensenberger MG, Lenz KA, Matthies LK, Hadinoto GM, Schienman JE, Przech AJ, Morganti MW, Renstrom DT, Baker VM, Gawrys KM, Hoogendoorn M, Steffen CR, Martín P, Alonso A, Olson HR, Sprecher CJ, Storts DR (2016) Developmental validation of the PowerPlex(®) Fusion 6C System. *Forensic Science International: Genetics* 21:134-144. https://doi.org/10.1016/j.fsigen.2015.12.011

Evett IW, Buffery C, Willott G, Stoney D (1991) A guide to interpreting single locus profiles of DNA mixtures in forensic cases. *Journal of Forensic Science Society* 31(1):41-47. https://doi.org/10.1016/s0015-7368(91)73116-2

Evett IW (1995) Avoiding the transposed conditional. *Science & Justice* 35(2):127-131. https://doi.org/10.1016/S1355-0306(95)72645-4

Evett IW & Weir BS (1998) *Interpreting DNA Evidence: Statistical Genetics for Forensic Scientists*. Sinauer Associates: Sunderland, Massachusetts.

Evett IW, Jackson G, Lambert JA (2000a) More on the hierarchy of propositions: Exploring the distinction between explanations and propositions. *Science & Justice* 40(1):3-10. https://doi.org/10.1016/S1355-0306(00)71926-5

Evett IW, Jackson G, Lambert JA, McCrossan S (2000b) The impact of the principles of evidence interpretation on the structure and content of statements. *Science & Justice* 40(4):233-239. https://doi.org/10.1016/S1355-0306(00)71993-9

Evett IW, Gill PD, Jackson G, Whitaker J, Champod C (2002) Interpreting small quantities of DNA: The hierarchy of propositions and the use of Bayesian networks. *Journal of Forensic Sciences* 47(3):520-530.

[EWG 2024] Taylor M, Romsos E, Ballantyne K, Moore Boswell D, Busey T. (2024) *Forensic DNA Interpretation and Human Factors: Improving the Practice Through a Systems Approach*. NIST Interagency Report (NISTIR) 8503, National Institute of Standards and Technology, Gaithersburg, MD. Available at https://doi.org/10.6028/NIST.IR.8503 (accessed May 10, 2024).

Farash K, Hanson EK, Ballantyne J (2015) Enhanced genetic analysis of single human bioparticles recovered by simplified micromanipulation from forensic 'touch DNA' evidence. *Journal of Visualized Experiments* 97:e52612. Video available at https://www.jove.com/v/52612 (accessed August 5, 2024).

Farash K, Hanson EK, Ballantyne J (2018) Single source DNA profile recovery from single cells isolated from skin and fabric from touch DNA mixtures in mock physical assaults. *Science & Justice* 58(3):191-199. https://doi.org/10.1016/j.scijus.2017.12.006

Farmen RK, Jaghø R, Cortez P, Frøyland ES (2008) Assessment of individual shedder status and implication for secondary DNA transfer. *Forensic Science International: Genetics Supplement Series* 1:415-417. https://doi.org/10.1016/j.fsigss.2007.08.015

Ferreira STG, Paula KA, Moraes AV (2013) DNA typing of trace DNA recovered from different areas of sandals found at a homicide crime scene investigation: A comparative study. *Forensic Science International: Genetics Supplement Series* 4(1):e372-e373. https://doi.org/10.1016/j.fsigss.2013.10.189

Findlay I, Taylor A, Quirke P, Frazier R, Urquhart A (1997) DNA fingerprinting from single cells. *Nature* 389:555-556. https://doi.org/10.1038/39225

Flanagan N & McAlister C (2011) The transfer and persistence of DNA under the fingernails following digital penetration of the vagina. *Forensic Science International: Genetics* 5(5):479-483. https://doi.org/10.1016/j.fsigen.2010.10.008

Fonneløp AE, Egeland T, Gill P (2015a) Secondary and subsequent DNA transfer during criminal investigation. *Forensic Science International: Genetics* 17:155-162. https://doi.org/10.1016/j.fsigen.2015.05.009

Fonneløp AE, Johannessen H, Gill P (2015b) Persistence and secondary transfer of DNA from previous users of equipment. *Forensic Science International: Genetics Supplement Series* 5:e191-e192. https://doi.org/10.1016/j.fsigss.2015.09.077

Fonneløp AE, Johannessen H, Egeland T, Gill P (2016) Contamination during criminal investigation: Detecting police contamination and secondary DNA transfer from evidence bags. *Forensic Science International: Genetics* 23:121-129. https://doi.org/10.1016/j.fsigen.2016.04.003

Fonneløp AE, Ramse M, Egeland T, Gill P (2017) The implications of shedder status and background DNA on direct and secondary transfer in an attack scenario. *Forensic Science International: Genetics* 29:48-60. https://doi.org/10.1016/j.fsigen.2017.03.019

Fonneløp AE, Johannessen, H, Heen, G, Molland, K, Gill, P. (2019) A retrospective study on the transfer, persistence and recovery of sperm and epithelial cells in samples collected in sexual assault casework. *Forensic Science International: Genetics* 43:102153. https://doi.org/10.1016/j.fsigen.2019.102153

Fontana F, Rapone C, Bregola G, Aversa R, de Meo A, Signorini G, Sergio M, Ferrarini A, Lanzellotto R, Medoro G, Giorgini G, Manaresi N, Berti A (2017) Isolation and genetic analysis of pure cells from forensic biological mixtures: The precision of a digital approach. *Forensic Science International: Genetics* 29:225-241. https://doi.org/10.1016/j.fsigen.2017.04.023

Fordyce SL, Mogensen HS, Børsting C, Lagacé RE, Chang CW, Rajagopalan N, Morling N (2015) Second-generation sequencing of forensic STRs using the Ion Torrent™ HID STR 10-plex and the Ion PGM™. *Forensic Science International: Genetics* 14:132-140. https://doi.org/10.1016/j.fsigen.2014.09.020

Forr C, Schei B, Stene LE, Ormstad K, Hagemann CT (2018) Factors associated with trace evidence analyses and DNA findings among police reported cases of rape. *Forensic Science International* 283:136-143. https://doi.org/10.1016/j.forsciint.2017.12.025

Geng T, Novak R, Mathies RA (2014) Single-cell forensic short tandem repeat typing within microfluidic droplets. *Analytical Chemistry* 86:703-712. https://doi.org/10.1021/ac403137h

Geng T & Mathies RA (2015) Minimizing inhibition of PCR-STR typing using digital agarose droplet microfluidics. *Forensic Science International: Genetics* 14:203-209. https://doi.org/10.1016/j.fsigen.2014.10.007

Gettings KB, Borsuk LA, Steffen CR, Kiesler KM, Vallone PM (2018) Sequence-based U.S. population data for 27 autosomal STR loci. *Forensic Science International: Genetics* 37:106-115. https://doi.org/10.1016/j.fsigen.2018.07.013

Gettings KB, Ballard D, Bodner M, Borsuk LA, King JL, Parson W, Phillips C (2019) Report from the STRAND Working Group on the 2019 STR sequence nomenclature meeting. *Forensic Science International: Genetics* 43:102165. https://doi.org/10.1016/j.fsigen.2019.102165

Gettings KB, Bodner M, Borsuk LA, King JL, Ballard D, Parson W, Benschop CCG, Børsting C, Budowle B, Butler JM, van der Gaag KJ, Gill P, Gusmão L, Hares DR, Hoogenboom J, Irwin J, Prieto L, Schneider PM, Vennemann M, Phillips C (2024) Recommendations of the DNA Commission of the International Society for Forensic Genetics (ISFG) on short tandem repeat sequence nomenclature. *Forensic Science International: Genetics* 68:102946. https://doi.org/10.1016/j.fsigen.2023.102946

Gill P, Jeffreys AJ, Werrett DJ (1985) Forensic application of DNA 'fingerprints.' *Nature* 318:577-579. https://doi.org/10.1038/318577a0

Gill P, Sparkes R, Kimpton C (1997) Development of guidelines to designate alleles using an STR multiplex system. *Forensic Science International* 89(3):185-197. https://doi.org/10.1016/s0379-0738(97)00131-x

Gill P, Whitaker J, Flaxman C, Brown N, Buckleton J (2000) An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. *Forensic Science International* 112(1):17-40. https://doi.org/10.1016/s0379-0738(00)00158-4

Gill P (2001) Application of low copy number DNA profiling. *Croatian Medical Journal* 42(3):229-232.

Gill P (2002) Role of short tandem repeat DNA in forensic casework in the UK—Past, present, and future perspectives. *Biotechniques* 32(2):366-372. https://doi.org/10.2144/02322rv01

Gill P, Curran J, Elliot K (2005) A graphical simulation model of the entire DNA process associated with the analysis of short tandem repeat loci. *Nucleic Acids Research* 33(2):632-643. https://doi.org/10.1093/nar/gki205

Gill P, Fereday L, Morling N, Schneider PM (2006a) The evolution of DNA databases—recommendations for new European STR loci. *Forensic Science International* 156(2-3):242-244. https://doi.org/10.1016/j.forsciint.2005.05.036

Gill P, Brenner CH, Buckleton JS, Carracedo A, Krawczak M, Mayr WR, Morling N, Prinz M, Schneider PM, Weir BS (2006b) DNA Commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Science International* 160:90-101. https://doi.org/10.1016/j.forsciint.2006.04.009

Gill P & Buckleton J (2010) A universal strategy to interpret DNA profiles that does not require a definition of low-copy-number. *Forensic Science International: Genetics* 4(4):221-227. https://doi.org/10.1016/j.fsigen.2009.09.008

Gill P, Gusmao L, Haned H, Mayr WR, Morling N, Parson W, Prieto L, Prinz M, Schneider H, Schneider PM, Weir BS (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. https://doi.org/10.1016/j.fsigen.2012.06.002

Gill P & Haned H (2013) A new methodological framework to interpret complex DNA profiles using likelihood ratios. *Forensic Science International: Genetics* 7(2):251-263. https://doi.org/10.1016/j.fsigen.2012.11.002

Gill P (2014) *Misleading DNA Evidence: Reasons for Miscarriages of Justice.* Elsevier Academic Press: San Diego.

Gill P, Haned H, Bleka Ø, Hansson O, Dørum G, Egeland T (2015) Genotyping and interpretation of STR-DNA: Low-template, mixtures and database matches—Twenty years of research and development. *Forensic Science International: Genetics* 18:100-117. https://doi.org/10.1016/j.fsigen.2015.03.014

Gill P (2016) Analysis and implications of the miscarriages of justice of Amanda Knox and Raffaele Sollecito. *Forensic Science International: Genetics* 23:9-18. https://doi.org/10.1016/j.fsigen.2016.02.015

Gill P, Hicks T, Butler JM, Connolly E, Gusmão L, Kokshoorn B, Morling N, van Oorschot RAH, Parson W, Prinz M, Schneider PM, Sijen T, Taylor D (2018) DNA Commission of the International Society for Forensic Genetics: Assessing the value of forensic biological evidence—Guidelines highlighting the importance of propositions. Part I: Evaluations of DNA profiling comparisons given (sub-) source

propositions. *Forensic Science International: Genetics* 36:189-202. https://doi.org/10.1016/j.fsigen.2018.07.003

Gill P (2019a) DNA evidence and miscarriages of justice. *Forensic Science International* 294:e1-e3. https://doi.org/10.1016/j.forsciint.2018.12.003

Gill, P. (2019b) Interpretation continues to be the main weakness in criminal justice systems: developing roles of the expert witness and court. *WIREs Forensic Sci.* 1:e1321. Available at https://doi.org/10.1002/wfs2.1321.

Gill P, Hicks T, Butler JM, Connolly E, Gusmão L, Kokshoorn B, Morling N, van Oorschot RAH, Parson W, Prinz M, Schneider PM, Sijen T, Taylor D (2020a) DNA Commission of the International Society for Forensic Genetics: Assessing the value of forensic biological evidence - guidelines highlighting the importance of propositions. Part II: Evaluation of biological traces considering activity level propositions. *Forensic Science International: Genetics* 44:102186. https://doi.org/10.1016/j.fsigen.2019.102186

Gill P, Bleka Ø, Hansson O, Benschop C, Haned H (2020b) *Forensic Practitioner's Guide to the Interpretation of Complex DNA Profiles.* Elsevier Academic Press: San Diego.

Gill P, Benschop C, Buckleton J, Bleka Ø, Taylor D (2021a) A review of probabilistic genotyping systems: *EuroForMix*, *DNAStatistX* and *STRmix™*. *Genes* 12(10): 1559. https://doi.org/10.3390/genes12101559

Gill P, Bleka Ø, Roseth A, Fonneløp AE (2021b) An LR framework incorporating sensitivity analysis to model multiple direct and secondary transfer events on skin surface. *Forensic Science International: Genetics* 53:102509. https://doi.org/10.1016/j.fsigen.2021.102509

Gittelson S, Kalafut T, Myers S, Taylor D, Hicks T, Taroni F, Evett IW, Bright JA, Buckleton J (2016) A practical guide for the formulation of propositions in the Bayesian approach to DNA evidence interpretation in an adversarial environment. *Journal of Forensic Sciences* 61(1):186-195. https://doi.org/10.1111/1556-4029.12907

Gittelson S, Berger CEH, Jackson G, Evett IW, Champod C, Robertson B, Curran JM, Taylor D, Weir BS, Coble MD, Buckleton JS (2018) A response to "Likelihood ratio as weight of evidence: A closer look" by Lund and Iyer. *Forensic Science International* 288:e15-e19. https://doi.org/10.1016/j.forsciint.2018.05.025

Goodwin W, Linacre A, Hadi S (2010) *An Introduction to Forensic Genetics, Second Edition*. Wiley: Hoboken, NJ.

Goray M, Eken E, Mitchell RJ, van Oorschot RAH (2010a) Secondary DNA transfer of biological substances under varying test conditions. *Forensic Science International: Genetics* 4:62-67. https://doi.org/10.1016/j.fsigen.2009.05.001

Goray M, Mitchell RJ, van Oorschot RAH (2010b) Investigation of secondary DNA transfer of skin cells under controlled test conditions. *Legal Medicine* 12:117-120. https://doi.org/10.1016/j.legalmed.2010.01.003

Goray M, Mitchell JR, van Oorschot RAH (2012a) Evaluation of multiple transfer of DNA using mock case scenarios. *Legal Medicine* 14:40-46. https://doi.org/10.1016/j.legalmed.2011.09.006

Goray M, van Oorschot RAH, Mitchell JR (2012b) DNA transfer within forensic exhibit packaging: Potential for DNA loss and relocation. *Forensic Science International: Genetics* 6(2):158-166. https://doi.org/10.1016/j.fsigen.2011.03.013

Goray M & van Oorschot RAH (2015) The complexities of DNA transfer during a social setting. *Legal Medicine* 17:82-91. https://doi.org/10.1016/j.legalmed.2014.10.003

Goray M, Fowler S, Szkuta B, van Oorschot RAH (2016) Shedder status—An analysis of self- and non-self-DNA in multiple handprints deposited by the same individuals over time. *Forensic Science International: Genetics* 23:190-196. https://doi.org/10.1016/j.fsigen.2016.05.005

Goray M, Pirie E, van Oorschot RAH (2019) DNA transfer: DNA acquired by gloves during casework examinations. *Forensic Science International: Genetics* 38:167-174. https://doi.org/10.1016/j.fsigen.2018.10.018

Gosch A & Courts C (2019) On DNA transfer: The lack and difficulty of systematic research and how to do it better. *Forensic Science International: Genetics* 40:24-36. https://doi.org/10.1016/j.fsigen.2019.01.012

Gosch A, Euteneuer J, Preuß-Wössner J, Courts C (2020) DNA transfer to firearms in alternative realistic handling scenarios. *Forensic Science International: Genetics* 48:102355. https://doi.org/10.1016/j.fsigen.2020.102355

Graham EA & Rutty GN (2008) Investigation into “normal” background DNA on adult necks: Implications for DNA profiling of manual strangulation victims. *Journal of Forensic Sciences* 53(5):1074-1082. https://doi.org/10.1111/j.1556-4029.2008.00800.x

Gramlich J (2019) Only 2% of federal criminal defendants go to trial, and most who do are found guilty. Pew Research Center. Available at https://www.pewresearch.org/fact-tank/2019/06/11/only-2-of-federal-criminal-defendants-go-to-trial-and-most-who-do-are-found-guilty/ (accessed on August 6, 2024).

Grasso JA & Woodard JW (1967) DNA synthesis and mitosis in erythropoietic cells. *Journal of Cell Biology* 33(3):645-655. https://doi.org/10.1083/jcb.33.3.645

Graversen T & Lauritzen S (2015) Computational aspects of DNA mixture analysis. *Statistics & Computing* 25(3): 527-541. https://doi.org/10.1007/s11222-014-9451-7

Green DM & Swets JA (1966) *Signal Detection Theory and Psychophysics.* Wiley: New York.

Greenspoon SA, Schiermeier-Wood L, Jenkins BC (2024) A tale of two PG systems: A comparison of the two most widely used continuous probabilistic genotyping systems in the United States. *Journal of Forensic Sciences* 69(5):1840-1860. https://doi.org/10.1111/1556-4029.15571

Grgicak CM, Bhembe Q, Slooten K, Sheth NC, Duffy KR, Lun DS (2024) Single-cell investigative genetics: Single-cell data produces genotype distributions concentrated at the true genotype across all mixture complexities. *Forensic Science International: Genetics* 69:103000. https://doi.org/10.1016/j.fsigen.2023.103000

Greytak EM, Moore C, Armentrout SL (2019) Genetic genealogy for cold case and active investigations. *Forensic Science International* 299:103-113. https://doi.org/10.1016/j.forsciint.2019.03.039

Grisedale KS & van Daal A (2012) Comparison of STR profiling from low template DNA extracts with and without the consensus profiling method. *Investigative Genetics* 3(1): 14. https://doi.org/10.1186/2041-2223-3-14

Gršković B, Markulin D, Zidkova A, Cmjac J, Anđelinović Š, Marijanović I, Tomašević L, Popović M, Primorac D, Mršić G (2014) Impact of donor age, gender and handling time on the DNA concentration left on different surfaces. *International Journal of Biomedicine* 4(3):169–174. Available at http://www.ijbm.org/articles/Article4_3_FR.pdf (accessed August 6, 2024).

Haas C, Hanson E, Banemann R, Bento AM, Berti A, Carracedo A, Courts C, Cock G, Drobnic K, Fleming R, Franchi C, Gomes I, Hadzic G, Harbison SA, Hjort B, Hollard C, Hoff-Olsen P, Keyser C, Kondili A, Maronas O, McCallum N, Miniati P, Morling N, Niederstatter H, Noel F, Parson W, Porto MJ, Roeder AD, Sauer E, Schneider PM, Shanthan G, Sijen T, Syndercombe Court D, Turanska M, van den Berge M, Vennemann M, Vidaki A, Zatkalikova L, Ballantyne J (2015) RNA/DNA co-analysis from human skin and contact traces—Results of a sixth collaborative EDNAP exercise. *Forensic Science International: Genetics* 16:139–147. https://doi.org/10.1016/j.fsigen.2015.01.002

Hackman L, Mack P, Ménard H (2024) Behind every good research there are data. What are they and their importance to forensic science. *Forensic Science International: Synergy* 8:100456. https://doi.org/10.1016/j.fsisyn.2024.100456

Hahn M, Courts C, Eckert M, Fimmers R, Grethe S, Kranz S, Leuker C, Oppelt C, Razbin S, Templin M, Vennemann M, Zimmermann P, Anslinger K (2024) Authors' response. *Journal of Forensic Sciences* 69(2):736-738. https://doi.org/10.1111/1556-4029.15426

Hannig J, Riman S, Iyer H, Vallone PM (2019) Are reported likelihood ratios well calibrated? *Forensic Science International: Genetics Supplement Series* 7:572-574. https://doi.org/10.1016/j.fsigss.2019.10.094

Hannig J & Iyer H (2022) Testing for calibration discrepancy of reported likelihood ratios in forensic science. *Journal of the Royal Statistical Society Series A: Statistics in Society* 185(1):267-301. https://doi.org/10.1111/rssa.12747

Hanson E, Haas C, Jucker R, Ballantyne J (2011) Identification of skin in touch/contact forensic samples by messenger RNA profiling. *Forensic Science International: Genetics Supplement Series* 3:e305-306. https://doi.org/10.1016/j.fsigss.2011.09.015

Hanson E, Haas C, Jucker R, Ballantyne J (2012) Specific and sensitive mRNA biomarkers for the identification of skin in 'touch DNA' evidence. *Forensic Science International: Genetics* 6(5):548-558. https://doi.org/10.1016/j.fsigen.2012.01.004

Hanson EK & Ballantyne J (2013) "Getting blood from a stone" Ultrasensitive forensic DNA profiling of microscopic bio-particles recovered from "touch DNA." *Methods in Molecular Biology* 1039:3-17. https://doi.org/10.1007/978-1-62703-535-4_1

Hares DR (2012) Expanding the CODIS core loci in the United States. *Forensic Science International: Genetics* 6(1):e52-e54. https://doi.org/10.1016/j.fsigen.2011.04.012

Hares DR (2015) Selection and implementation of expanded CODIS core loci in the United States. *Forensic Science International: Genetics* 17:33-34. https://doi.org/10.1016/j.fsigen.2015.03.006

Hardy GH (1908) Mendelian proportions in a mixed population. *Science* 17:49-50. https://doi.org/10.1126/science.28.706.49

Hazell-Smithen M, Callahan T, Miller Coyle H (2014) Touch DNA and the ability to detect the correct source. *International Journal of Advanced Science and Technology* 1(1):45-54. Available at http://www.identacode.org/Touch-DNA-and-the-Ability-to-Detect-the-Correct-Source.pdf (accessed August 6, 2024).

Hellerud B, Johannessen H, Haltbakk H, Hoff-Olsen P (2008) Zip lock poly bags in drug cases—A valuable source for obtaining identifiable DNA results? *Forensic Science International: Genetics Supplement Series* 1(1):433-434. https://doi.org/10.1016/j.fsigss.2007.10.013

Helmus J, Bajanowski T, Poetsch M (2016) DNA transfer—A never ending story. A study on scenarios involving a second person as carrier. *International Journal of Legal Medicine* 130(1):121-125. https://doi.org/10.1007/s00414-015-1284-1

Helmus J, Zorell S, Bajanowski T, Poetsch M (2018) Persistence of DNA on clothes after exposure to water for different time periods—A study on bathtub, pond, and river. *International Journal of Legal Medicine* 132(1):99-106. https://doi.org/10.1007/s00414-017-1695-2

Helmus J, Pfeifer M, Feiner LK, Krause LJ, Bajanowski T, Poetsch M (2019) Unintentional effects of cleaning a crime scene—When the sponge becomes an accomplice in DNA transfer. *International Journal of Legal Medicine* 133(3):759-765. https://doi.org/10.1007/s00414-018-1983-5

Henry J, McGowan P, Brown C (2015) A survey of environmental DNA in South Australia Police facilities. *Forensic Science International: Genetics Supplement Series* 5:e465-e466. https://doi.org/10.1016/j.fsigss.2015.09.184

Hert DG, Fredlake CP, Barron AE (2008) Advantages and limitations of next-generation sequencing technologies: a comparison of electrophoresis and non-electrophoresis methods. *Electrophoresis* 29(23): 4618-4626. https://doi.org/10.1002/elps.200800456

Hicklin RA, Richetelli N, Emerick BL, Bever RA, Davoren JM (2023) Variation in assessments of suitability and number of contributors for DNA mixtures. *Forensic Science International: Genetics* 65: 102892. https://doi.org/10.1016/j.fsigen.2023.102892

Hicks T, Kerr Z, Pugh S, Bright JA, Curran J, Taylor D, Buckleton J (2021) Comparing multiple POI to DNA mixtures. *Forensic Science International: Genetics* 52:102481. https://doi.org/10.1016/j.fsigen.2021.102481

Hicks T, Buckleton J, Castella V, Evett I, Jackson G (2022) A logical framework for forensic DNA interpretation. *Genes* 13(6):957. https://doi.org/10.3390/genes13060957

Holt CL, Buoncristiani M, Wallin JM, Nguyen T, Lazaruk KD, Walsh PS (2002) TWGDAM validation of AmpFlSTR PCR amplification kits for forensic DNA casework. *Journal of Forensic Sciences* 47(1):66-96.

Huffman K & Ballantyne J (2023a) Validation of probabilistic genotyping software for single cell STR analysis. *Genes* 14(3): 674. https://doi.org/10.3390/genes14030674

Huffman K & Ballantyne J (2023b) Single cell genomics applications in forensic science: Current state and future directions. *iScience* 26(11): 107961. https://doi.org/10.1016/j.isci.2023.107961

Huffman K, Kruijver M, Ballantyne J, Taylor D (2023) Carrying out common DNA donor analysis using DBLR™ on two or five-cell mini-mixture subsamples for improved discrimination power in complex DNA mixtures. *Forensic Science International: Genetics* 66:102908. https://doi.org/10.1016/j.fsigen.2023.102908

Hundl C, Neuman M, Rairden A, Rearden P, Stout P (2020) Implementation of a blind quality control program in a forensic laboratory. *Journal of Forensic Sciences* 65(3):815-822. https://doi.org/10.1111/1556-4029.14259

[ILAC G19] International Laboratory Accreditation Cooperation (ILAC) (2022) Modules in a Forensic Science Process. G19:06/2022. Available at https://ilac.org/publications-and-resources/ilac-guidance-series/ (accessed April 20, 2023).

Inman K & Rudin N (2001) *Principles and Practice of Criminalistics: The Profession of Forensic Science*. CRC Press: Boca Raton.

Inman K, Rudin N, Cheng K, Robinson C, Kirschner A, Inman-Semerau L, Lohmueller KE (2015) Lab Retriever: a software tool for calculating likelihood ratios incorporating a probability of drop-out for forensic DNA profiles. *BMC Bioinformatics* 16:298. https://doi.org/10.1186/s12859-015-0740-8

[ISHI 2019] DNA Mixture Interpretation Principles and Best Practices. Workshop presented by John Butler, Hari Iyer, and Sheila Willis at the International Symposium on Human Identification (Palm Springs, CA) on September 26, 2019. Available at https://strbase-archive.nist.gov/pub_pres/ISHI2019-MixtureWorkshop.pdf (accessed October 8, 2024).

[ISHI 2020] Validation Principles, Practices, Parameters, Performance Evaluations, and Protocols. Virtual workshop presented by John Butler and Hari Iyer at the International Symposium on Human Identification on September 18, 2020. Available at https://strbase.nist.gov/NIST_Resources/Presentations/2020/ISHI2020-ValidationWorkshop-Butler_Iyer-Slides.pdf (accessed October 8, 2024).

[ISHI 2021] NIST Scientific Foundation Review on DNA Mixture Interpretation. Presentation by John Butler at the International Symposium on Human Identification (Orlando, FL) on September 16, 2021. Available at https://strbase.nist.gov/NIST_Resources/Presentations/2021/Butler-ISHI2021-NIST-DNA-Mixture-Report.pdf (accessed October 8, 2024).

[ISO 18385:2016] International Organization for Standardization (2016) Minimizing the risk of human DNA contamination in products used to collect, store and analyze biological material for forensic purposes – Requirements. Available at https://www.iso.org/standard/62341.html (accessed August 23, 2024).

[ISO/IEC 17025:2017] International Organization for Standardization (2017) General requirements for the competence of testing and calibration laboratories. Available at https://www.iso.org/ISO-IEC-17025-testing-and-calibration-laboratories.html (accessed August 23, 2024). *Section 7.2.2.1(e) mentions interlaboratory comparisons as a technique used for method validation and Section 7.7.2(b) encourages laboratories to monitor their performance by participation in interlaboratory comparisons.*

Iuvaro A, Bini C, Dilloo S, Salmo S, Pelotti S (2018) Male DNA under female fingernails after scratching: Transfer and persistence evaluation by RT-PCR analysis and Y-STR typing. *International Journal of Legal Medicine* 132(6):1603-1609. https://doi.org/10.1007/s00414-018-1839-z

Jackson G, Jones S, Booth G, Champod C, Evett IW (2006) The nature of forensic science opinion – a possible framework to guide thinking and practice in investigation and in court proceedings. *Science & Justice* 46(1):33-44. https://doi.org/10.1016/s1355-0306(06)71565-9

Jackson G (2013) The impact of commercialization on the evaluation of DNA evidence. *Frontiers in Genetics* 4:227. https://doi.org/10.3389/fgene.2013.00227.

Jackson G, Aitken C, Roberts P (2015) Practitioner Guide No, 4: *Case Assessment and Interpretation of Expert Evidence: Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses*, Royal Statistical Society (145 pages). Available at https://rss.org.uk/news-publication/publications/law-guides/ (accessed on October 4, 2024).

Jackson G & Biedermann A (2019) "Source" or "activity": What is the level of issue in a criminal trial? *Significance* 16(2):36-39. https://doi.org/10.1111/j.1740-9713.2019.01253.x

Jäger AC, Alvarez ML, Davis CP, Guzmán E, Han Y, Way L, Walichiewicz P, Silva D, Pham N, Caves G, Bruand J, Schlesinger F, Pond SJK, Varlaro J, Stephens KM, Holt CL (2017) Developmental validation of the MiSeq FGx Forensic Genomics System for Targeted Next Generation Sequencing in Forensic DNA Casework and Database Laboratories. *Forensic Science International: Genetics* 28:52-70. https://doi.org/10.1016/j.fsigen.2017.01.011

Jeffreys AJ (1987) Highly variable minisatellites and DNA fingerprints. *Biochemical Society Transactions* 15(3):309-317. https://doi.org/10.1042/bst0150309

Jones S, Scott K, Lewis J, Davidson G, Allard JE, Lowrie C, McBride BM, McKenna L, Teppett G, Rogers C, Clayson N, Baird A (2016) DNA transfer through nonintimate social contact. *Science & Justice* 56(2):90-95. https://doi.org/10.1016/j.scijus.2015.10.004

Just RS & Irwin JA (2018) Use of the LUS in sequence allele designations to facilitate probabilistic genotyping of NGS-based STR typing results. *Forensic Science International: Genetics* 34:197-205. https://doi.org/10.1016/j.fsigen.2018.02.016

Kadane JB (2011) *Principles of Uncertainty*. CRC Press: Boca Raton.

Kafarowski E, Lyon AM, Sloan MM (1996) The retention and transfer of spermatozoa in clothing by machine washing. *Canadian Society of Forensic Science Journal* 29(1):7-11. https://doi.org/10.1080/00085030.1996.10757042

Kalafut T, Schuerman C, Sutton J, Faris T, Armogida L, Bright JA, Buckleton J, Taylor D (2018) Implementation and validation of an improved allele specific stutter filtering method for electropherogram interpretation. *Forensic Science International: Genetics* 35:50-56. https://doi.org/10.1016/j.fsigen.2018.03.016

Kalafut T, Curran JM, Coble MD, Buckleton J (2024) Commentary on: Thompson WC. Uncertainty in probabilistic genotyping of low template DNA: a case study comparing STRmix™ and TrueAllele™. J Forensic Sci. 2023;68 (3):1049-63. doi: 10.1111/1556-4029.15225. *Journal of Forensic Sciences* 69(1): 371-377. https://doi.org/10.1111/1556-4029.15405

Kamphausen T, Schadendorf D, von Wurmb-Schwark N, Bajanowski T, Poetsch M (2012) Good shedder or bad shedder—The influence of skin diseases on forensic DNA analysis from epithelial abrasions. *International Journal of Legal Medicine* 126:179-183. https://doi.org/10.1007/s00414-011-0579-0

Kamphausen T, Fandel SB, Gutmann JS, Bajanowski T, Poetsch M (2015) Everything clean? Transfer of DNA traces between textiles in the washtub. *International Journal of Legal Medicine* 129:709-714. https://doi.org/10.1007/s00414-015-1203-5

Kanokwongnuwut P, Martin B, Kirkbride KP, Linacre A (2018) Shedding light on shedders. *Forensic Science International: Genetics* 36:20-25. https://doi.org/10.1016/j.fsigen.2018.06.004

Katsanis SH & Wagner JK (2013) Characterization of the standard and recommended CODIS markers. *Journal of Forensic Sciences* 58(Suppl 1):S169-S172. https://doi.org/10.1111/j.1556-4029.2012.02253.x

Kayser M (2007) Uni-parental markers in human identity testing including forensic DNA analysis. *Biotechniques* 43(6):S16-S21.

Kayser M, Gill P, Parson W, Gusmão L, Linacre A, Vallone P, Carracedo A (2023) Editorial considerations for publication in Forensic Science International: Genetics. *Forensic Science International: Genetics* 65:102877. https://doi.org/10.1016/j.fsigen.2023.102877

Kelly H, Bright JA, Buckleton JS, Curran JM (2014) A comparison of statistical models for the analysis of complex forensic DNA profiles. *Science & Justice* 54(1):66-70. https://doi.org/10.1016/j.scijus.2013.07.003

Kelly H, Bright JA, Kruijver M, Cooper S, Taylor D, Duke K, Strong M, Beamer V, Buettner C, Buckleton J (2018) A sensitivity analysis to determine the robustness of STRmix™ with respect to laboratory

calibration. *Forensic Science International: Genetics* 35:113-122. https://doi.org/10.1016/j.fsigen.2018.04.009

Kelly H, Bright JA, Coble MD, Buckleton J (2020) A description of the likelihood ratios in the probabilistic genotyping software STRmix™. *WIREs Forensic Sciences* 2(6):e1377. https://doi.org/10.1002/wfs2.1377

Kidd KK, Pakstis AJ, Speed WC, Lagace R, Chang J, Wootton S, Haigh E, Kidd JK (2014) Current sequencing technology makes microhaplotypes a powerful new type of genetic marker for forensics. *Forensic Science International: Genetics* 12:215-224. https://doi.org/10.1016/j.fsigen.2014.06.014

Kirchner L (2017) Putting crime scene DNA analysis on trial. ProPublica (October 11, 2017). Available at https://www.propublica.org/article/putting-crime-scene-dna-analysis-on-trial (accessed October 8, 2024).

Kita T, Yamaguchi H, Yokoyama M, Tanaka T, Tanaka N (2008) Morphological study of fragmented DNA on touched objects. *Forensic Science International: Genetics* 3:32-36. https://doi.org/10.1016/j.fsigen.2008.09.002

Kloosterman A, Sjerps M, Quak A (2014) Error rates in forensic DNA analysis: Definition, numbers, impact and communication. *Forensic Science International: Genetics* 12:77-85. https://doi.org/10.1016/j.fsigen.2014.04.014

Kokshoorn B, Blankers BJ, de Zoete J, Berger CEH (2017) Activity level DNA evidence evaluation: On propositions addressing the actor or the activity. *Forensic Science International* 278:115-124. https://doi.org/10.1016/j.forsciint.2017.06.029

Kokshoorn B, Aarts LHJ, Ansell R, Connolly E, Drotz W, Kloosterman AD, McKenna LG, Szkuta B, van Oorschot RAH (2018) Sharing data on DNA transfer, persistence, prevalence and recovery: Arguments for harmonization and standardization. *Forensic Science International: Genetics* 37:260-269. https://doi.org/10.1016/j.fsigen.2018.09.006

Kokshoorn B, Luijsterburg M (2023) Reporting on forensic biology findings given activity level issues in the Netherlands. *Forensic Science International* 343: 111545. https://doi.org/10.1016/j.forsciint.2022.111545

Kolowski J, Barden C, Brown C, Leete J, Czyryca C (2016) Proficiency testing trends following the 2009 National Academy of Sciences report, “Strengthening Forensic Science in the United States: A Path Forward”. *Forensic Science Policy and Management* 7: 98-105. https://doi.org/10.1080/19409044.2016.1221485

Kovács G & Pádár Z (2015) Misinterpretation of sample contamination in a Hungarian casework. *Forensic Science International: Genetics Supplement Series* 5:e425-e427. https://doi.org/10.1016/j.fsigss.2015.09.169

Kulhankova L, Montiel González D, Bindels E, Kling D, Kayser M, Mulugeta E (2023) Single-cell transcriptome sequencing allows genetic separation, characterization and identification of individuals in multi-person biological mixtures. *Communications Biology* 6(1):201. https://doi.org/10.1038/s42003-023-04557-z

Kulhankova L, Bindels E, Kayser M, Mulugeta E (2024) Deconvoluting multi-person biological mixtures and accurate characterization and identification of separated contributors using non-targeted single-cell DNA sequencing. *Forensic Science International: Genetics* 71:103030. https://doi.org/10.1016/j.fsigen.2024.103030

Kwong K (2017) The algorithm says you did it: The use of black box algorithms to analyze complex DNA evidence. *Harvard Journal of Law & Technology* 31(1):275-301.

Lacerenza D, Aneli S, Omedei M, Gino S, Pasino S, Berchialla P, Robino C (2016) A molecular exploration of human DNA/RNA co-extracted from the palmar surface of the hands and fingers. *Forensic Science International: Genetics* 22:44-53. https://doi.org/10.1016/j.fsigen.2016.01.012

Lander ES, et al. (International Human Genome Sequencing Consortium) (2001) Initial sequencing and analysis of the human genome. *Nature* 409(6822):860-921. https://doi.org/10.1038/35057062. Erratum in: Nature 2001 Aug 2;412(6846):565. Erratum in: Nature 2001 Jun 7;411(6838):720.

Lapointe M, Rogic A, Bourgoin S, Jolicoeur C, Séguin D (2015) Leading-edge forensic DNA analyses and the necessity of including crime scene investigators, police officers and technicians in a DNA elimination database. *Forensic Science International: Genetics* 19: 50–55. https://doi.org/10.1016/j.fsigen.2015.06.002

Lazaruk K, Walsh PS, Oaks F, Gilbert D, Rosenblum BB, Menchen S, Scheibler D, Wenz HM, Holt C, Wallin J (1998) Genotyping of forensic short tandem repeat (STR) systems based on sizing precision in a capillary electrophoresis instrument. *Electrophoresis* 19(1):86-93. https://doi.org/10.1002/elps.1150190116

Leclair B, Sgueglia JB, Wojtowicz PC, Juston AC, Frégeau CJ, Fourney RM (2003) STR DNA typing: increased sensitivity and efficient sample consumption using reduced PCR reaction volumes. *Journal of Forensic Sciences* 48(5):1001-1013.

Leclair B, Frégeau CJ, Bowen KL, Fourney RM (2004) Systematic analysis of stutter percentages and allele peak height and peak area ratios at heterozygous STR loci for forensic casework and database samples. *Journal of Forensic Sciences* 49(5):968-980.

LehmannVJ, Mitchell RJ, Ballantyne KN, van Oorschot RAH (2013) Following the transfer of DNA: How far can it go? *Forensic Science International: Genetics Supplement Series* 4:e53-e54. https://doi.org/10.1016/j.fsigss.2013.10.027

Lehmann VJ, Mitchell RJ, Ballantyne KN, van Oorschot RAH (2015) Following the transfer of DNA: How does the presence of background DNA affect the transfer and detection of a target source of DNA? *Forensic Science International: Genetics* 19:68-75. https://doi.org/10.1016/j.fsigen.2015.05.002

Li CC, Weeks DE, Chakravarti A. (1993) Similarity of DNA fingerprints due to chance and relatedness. *Human Heredity* 43:45-52. https://doi.org/10.1159/000154113

Lindley DV (1977) A problem in forensic science. *Biometrika* 64:207-213. https://doi.org/10.2307/2335686

Lindley DV (2014) *Understanding Uncertainty*, Revised Edition. John Wiley and Sons: Hoboken.

Lohmueller KE & Rudin N (2013) Calculating the weight of evidence in low-template forensic DNA casework. *Journal of Forensic Sciences* 58 Suppl 1:S243-S249. https://doi.org/10.1111/1556-4029.12017

Lowe A, Murray C, Whitaker J, Tully G, Gill P (2002) The propensity of individuals to deposit DNA and secondary transfer of low level DNA from individuals to inert surfaces. *Forensic Science International* 129:25–34. https://doi.org/10.1016/s0379-0738(02)00207-4

Ludeman MJ, Zhong C, Mulero JJ, Lagacé RE, Hennessy LK, Short ML, Wang DY (2018) Developmental validation of GlobalFiler™ PCR amplification kit: a 6-dye multiplex assay designed for amplification of casework samples. *International Journal of Legal Medicine* 132(6):1555-1573. https://doi.org/10.1007/s00414-018-1817-5

Lund SP & Iyer H (2017) Likelihood ratio as weight of forensic evidence: a closer look. *Journal of Research of National Institute of Standards and Technology* 122:27. Available at https://nvlpubs.nist.gov/nistpubs/jres/122/jres.122.027.pdf (accessed August 23, 2024).

Lund S & Iyer H (2024) The influence of validation data on logical and scientific interpretations of forensic expert opinions. arXiv:2403.02663v1. https://doi.org/10.48550/arXiv.2403.02663

Lynch PC & Cotton RW (2018) Determination of the possible number of genotypes which can contribute to DNA mixtures: non-computer assisted deconvolution should not be attempted for greater than two person mixtures. *Forensic Science International: Genetics* 37:235-240. https://doi.org/10.1016/j.fsigen.2018.09.002

Magee AM, Breathnach M, Doak S, Thornton F, Noone C, McKenna LG (2018) Wearer and non-wearer DNA on the collars and cuffs of upper garments of worn clothing. *Forensic Science International: Genetics* 34:152-161. https://doi.org/10.1016/j.fsigen.2018.02.011

Manabe S, Morimoto C, Hamano Y, Fujimoto S, Tamaki K (2017) Development and validation of open-source software for DNA mixture interpretation based on a quantitative continuous model. *PLoS ONE* 12(11):e0188183. https://doi.org/10.1371/journal.pone.0188183

Manoli P, Antoniou A, Bashiardes E, Xenophontos S, Photiades M, Stribley V, Mylona M, Demetriou C, Cariolou MA (2016) Sex-specific age association with primary DNA transfer. *International Journal of Legal Medicine* 130(1):103-112. https://doi.org/10.1007/s00414-015-1291-2

Mapes AA, Kloosterman AD, van Marion V, de Poot CJ (2016) Knowledge on DNA success rates to optimize the DNA analysis process: From crime scene to laboratory. *Journal of Forensic Sciences* 61(4):1055-1061. https://doi.org/10.1111/1556-4029.13102

Margiotta G, Tasselli G, Tommolini F, Lancia M, Massetti S, Carnevali E (2015) Risk of DNA transfer by gloves in forensic casework. *Forensic Science International: Genetics Supplement Series* 5:e527-e529. https://doi.org/10.1016/j.fsigss.2015.09.208

Marquis R, Biedermann A, Cadola L, Champod C, Gueissaz L, Massonnet G, Mazzella WD, Taroni F, Hicks T (2016) Discussion on how to implement a verbal scale in a forensic laboratory: Benefits, pitfalls and suggestions to avoid misunderstandings. *Science & Justice* 56:364-370. https://doi.org/10.1016/j.scijus.2016.05.009

Martire KA & Kemp RI (2018) Considerations when designing human performance tests in the forensic sciences. *Australian Journal of Forensic Sciences* 50(2):166-182. http://dx.doi.org/10.1080/00450618.2016.1229815

Matte M, Williams L, Frappier R, Newman J (2012) Prevalence and persistence of foreign DNA beneath fingernails. *Forensic Science International: Genetics* 6(2):236-243. https://doi.org/10.1016/j.fsigen.2011.05.008

McCarthy-Allen M, Bleka Ø, Ypma R, Gill P, Benschop C (2024) 'Low' LRs obtained from DNA mixtures: On calibration and discrimination performance of probabilistic genotyping software. *Forensic Science International: Genetics* 73:103099. https://doi.org/10.1016/j.fsigen.2024.103099

McDonald A, Jones E, Lewis J, O'Rourke P (2015) Y-STR analysis of digital and/or penile penetration cases with no detected spermatozoa. *Forensic Science International: Genetics* 15:84-89. https://doi.org/10.1016/j.fsigen.2014.10.015

McKenna L (2013) Understanding DNA results within the case context: Importance of the alternative proposition. *Frontiers in Genetics* 4:242. https://doi.org/10.3389/fgene.2013.00242

McNevin D, Wright K, Barash M, Gomes S, Jamieson A, Chaseling J (2021) Proposed framework for comparison of continuous probabilistic genotyping systems amongst different laboratories. *Forensic Sciences* 1(1):33-45. https://doi.org/10.3390/forensicsci1010006

McNevin D & Barash M (2024) Comparison of likelihood ratios from probabilistic genotyping for two-person mixtures across different assays and instruments. *Forensic Sciences* 4:441-452. https://doi.org/10.3390/forensicsci4030028

Meakin G & Jamieson A (2013) DNA transfer: Review and implications for casework. *Forensic Science International: Genetics* 7:434-443. https://doi.org/10.1016/j.fsigen.2013.03.013

Meakin G., Butcher EV, van Oorschot RAH, Morgan RM (2015) The deposition and persistence of indirectly-transferred DNA on regularly-used knives. *Forensic Science International: Genetics Supplement Series* 5:e498-e500. https://doi.org/10.1016/j.fsigss.2015.09.197

Meakin GE, Butcher EV, van Oorschot RAH, Morgan RM (2017) Trace DNA evidence dynamics: An investigation into the deposition and persistence of directly- and indirectly-transferred DNA on regularly-used knives. *Forensic Science International: Genetics* 29:38-47. https://doi.org/10.1016/j.fsigen.2017.03.016

Meuwly D, Ramos D, Haraksim R (2017) A guideline for the validation of likelihood ratio methods used for forensic evidence evaluation. *Forensic Science International* 276:142-153. https://doi.org/10.1016/j.forsciint.2016.03.048

Montpetit S & O'Donnell P (2015) An optimized procedure for obtaining DNA from fired and unfired ammunition. *Forensic Science International: Genetics* 17:70-74. https://doi.org/10.1016/j.fsigen.2015.03.012

Moore D, Clayton T, Thomson J (2020) A comprehensive study of allele drop-in over an extended period of time. *Forensic Science International: Genetics* 48:102332. https://doi.org/10.1016/j.fsigen.2020.102332

Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM, Smerick JB, Budowle B (2001a) Validation of short tandem repeats (STRs) for forensic usage: performance testing of fluorescent multiplex STR systems and analysis of authentic and simulated forensic samples. *Journal of Forensic Sciences* 46(3):647-660.

Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM, Brown AL, Budowle B (2001b) Validation of STR typing by capillary electrophoresis. *Journal of Forensic Sciences* 46(3):661-676.

Moretti TR, Moreno LI, Smerick JB, Pignone ML, Hizon R, Buckleton JS, Bright J-A, Onorato AJ (2016) Population data on the expanded CODIS core STR loci for eleven populations of significance for forensic DNA analyses in the United States. *Forensic Science International: Genetics* 25:175-181. https://doi.org/10.1016/j.fsigen.2016.07.022

Morling N, Bastisch I, Gill P, Schneider PM (2007) Interpretation of DNA mixtures--European consensus on principles. *Forensic Science International: Genetics* 1(3-4):291-292. https://doi.org/10.1016/j.fsigen.2007.06.007

Morrison GS (2016) Special issue on measuring and reporting the precision of forensic likelihood ratios: Introduction to the debate. *Science & Justice* 56(5):371-373. https://doi.org/10.1016/j.scijus.2016.05.002

Morrison GS & Enzinger E (2016) What should a forensic practitioner's likelihood ratio be? *Science & Justice* 56:374-379. https://doi.org/10.1016/j.scijus.2016.05.007

[NAS 2009] National Academy of Sciences (2009) *On Being a Scientist: A Guide to Responsible Conduct in Research* (Third Edition). The National Academies Press: Washington, DC. https://doi.org/10.17226/12192

[NRC I] National Research Council (1992) *DNA Technology in Forensic Science.* The National Academy Press: Washington DC. Available at https://nap.nationalacademies.org/catalog/1866/dna-technology-in-forensic-science (accessed August 24, 2024).

[NRC II] National Research Council (1996) *The Evaluation of Forensic DNA Evidence*. The National Academy Press: Washington DC. Available at https://nap.nationalacademies.org/catalog/5141/the-evaluation-of-forensic-dna-evidence (accessed August 24, 2024).

[NRC 2009] National Research Council, Committee on Identifying the Needs of the Forensic Science Community (2009) *Strengthening Forensic Science in the United States: A Path Forward*. The National Academies Press: Washington DC. Available at https://nap.nationalacademies.org/catalog/12589/strengthening-forensic-science-in-the-united-states-a-path-forward (accessed August 24, 2024).

Neuhuber F, Dunkelmann B, Höckner G, Kiesslich J, Klausriegler E, Radacher M (2009) Female criminals: It's not always the offender! *Forensic Science International: Genetics Supplement Series* 2(1):145-146. https://doi.org/10.1016/j.fsigss.2009.08.026

Neuhuber F, Kreindl G, Kastinger T, Dunkelmann B, Zahrer W, Cemper-Kiesslich J, Grießner I (2017) Police officer's DNA on crime scene samples—Indirect transfer as a source of contamination and its database-assisted detection in Austria. *Forensic Science International: Genetics Supplement Series* 6:e608-e609. https://doi.org/10.1016/j.fsigss.2017.11.001

[NISTIR 8225] Butler, J.M., Iyer, H., Press, R., Taylor, M.K., Vallone, P.M. and Willis, S. (2020) NIST scientific foundation reviews. (National Institute of Standards and Technology, Gaithersburg, MD), NISTIR 8225. https://doi.org/10.6028/NIST.IR.8225

[NISTIR 8351sup1] Butler, J.M. (2024) History of DNA Mixture Interpretation. *Supplemental Document to DNA Mixture Interpretation: A NIST Scientific Foundation Review.* (National Institute of Standards and Technology, Gaithersburg, MD), NIST IR 8351sup1. https://doi.org/10.6028/NIST.IR.8351sup1

[NISTIR 8351sup2] Butler, J.M. (2024) Summarized Information from Publicly Accessible Validation Data and Proficiency Test Results Covering DNA Mixture Interpretation. *Supplemental Document to DNA Mixture Interpretation: A NIST Scientific Foundation Review.* (National Institute of Standards and Technology, Gaithersburg, MD), NIST IR 8351sup2. https://doi.org/10.6028/NIST.IR.8351sup2

Noël S, Lagacé K, Rogic A, Granger D, Bourgoin S, Jolicoeur C, Séguin D (2016) DNA transfer during laundering may yield complete genetic profiles. *Forensic Science International: Genetics* 23:240-247. https://doi.org/10.1016/j.fsigen.2016.05.004

Novroski NMM, Woerner AE, Budowle B (2018) Potential highly polymorphic short tandem repeat markers for enhanced forensic identity testing. *Forensic Science International: Genetics* 37:162-171. https://doi.org/10.1016/j.fsigen.2018.08.011

Oldoni F, Castella V, Hall D (2015) Exploring the relative DNA contribution of first and second object's users on mock touch DNA mixtures. *Forensic Science International: Genetics Supplement Series* 5:e300-e301. https://doi.org/10.1016/j.fsigss.2015.09.119

Oldoni F, Castella V, Hall D (2016) Shedding light on the relative DNA contribution of two persons handling the same object. *Forensic Science International: Genetics* 24:148-157. https://doi.org/10.1016/j.fsigen.2016.07.002

Oleiwi AA, Morris MR, Schmerer WM, Sutton R (2015) The relative DNA-shedding propensity of the palm and finger surfaces. *Science & Justice* 55(5):329-334. https://doi.org/10.1016/j.scijus.2015.04.003

Paoletti DR, Doom TE, Krane CM, Raymer ML, Krane DE (2005) Empirical analysis of the STR profiles resulting from conceptual mixtures. *Journal of Forensic Sciences* 50(6):1361-1366.

Parson W, Ballard D, Budowle B, Butler JM, Gettings KB, Gill P, Gusmao L, Hares DR, Irwin JA, King JL, de Knijff P, Morling N, Prinz M, Schneider PM, Neste CV, Willuweit S, Phillips C (2016) Massively parallel sequencing of forensic STRs: Considerations of the DNA commission of the International Society for

Forensic Genetics (ISFG) on minimal nomenclature requirements. *Forensic Science International: Genetics* 22:54-63. https://doi.org/10.1016/j.fsigen.2016.01.009

Perlin MW, Legler MM, Spencer CE, Smith JL, Allan WP, Belrose JL, Duceman BW (2011) Validating TrueAllele® DNA mixture interpretation. *Journal of Forensic Sciences* 56(6):1430-1447. https://doi.org/10.1111/j.1556-4029.2011.01859.x

Perlin MW, Butt N, Wilson MR (2024) Commentary on: Thompson WC. Uncertainty in probabilistic genotyping of low template DNA: A case study comparing STRmix™ and TrueAllele®. J Forensic Sci. 2023;68(3):1049-63. *Journal of Forensic Sciences* 69(4):1516-1518. https://doi.org/10.1111/1556-4029.15518

Pfeifer CM & Wiegand P (2017) Persistence of touch DNA on burglary-related tools. *International Journal of Legal Medicine* 131(4):941-953. https://doi.org/10.1007/s00414-017-1551-4

Phillips C, Gettings KB, King JL, Ballard D, Bodner M, Borsuk L, Parson W (2018) "The devil's in the detail": Release of an expanded, enhanced and dynamically revised forensic STR Sequence Guide. *Forensic Science International: Genetics* 34:162-169. https://doi.org/10.1016/j.fsigen.2018.02.017

Phipps M & Petricevic S (2007) The tendency of individuals to transfer DNA to handled items. *Forensic Science International* 168(2–3):162-168. https://doi.org/10.1016/j.forsciint.2006.07.010

Pickrahn I, Kreindl G, Muller E, Dunkelmann B, Zahrer W, Cemper-Kiesslich J, Neuhuber F (2017) Contamination incidents in the pre-analytical phase of forensic DNA analysis in Austria—Statistics of 17 years. *Forensic Science International: Genetics* 31:12-18. https://doi.org/10.1016/j.fsigen.2017.07.012

Poetsch M, Bajanowski T, Kamphausen T (2013) Influence of an individual's age on the amount and interpretability of DNA left on touched items. *International Journal of Legal Medicine* 127:1093-1096. https://doi.org/10.1007/s00414-013-0916-6

Poetsch M, Pfeifer M, Konrad H, Bajanowski T, Helmus J (2018) Impact of several wearers on the persistence of DNA on clothes—A study with experimental scenarios. *International Journal of Legal Medicine* 132(1):117-123. https://doi.org/10.1007/s00414-017-1742-z

Possolo A (2015) *Simple Guide for Evaluating and Expressing the Uncertainty of NIST Measurement Results.* NIST Technical Note 1900. Available at https://nvlpubs.nist.gov/nistpubs/TechnicalNotes/NIST.TN.1900.pdf (accessed August 24, 2024).

[PCAST 2016] President's Council of Advisors on Science and Technology. Report to the President - Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods. September 2016. Available at https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf (accessed August 24, 2024).

[PCAST 2017] President's Council of Advisors on Science and Technology. *An Addendum to the PCAST Report on Forensic Science in Criminal Courts*. Available at https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensics_addendum_finalv2.pdf (accessed August 24, 2024).

Press R (2017) NIST to assess the reliability of forensic methods for analyzing DNA mixtures. Available at https://www.nist.gov/news-events/news/2017/10/nist-assess-reliability-forensic-methods-analyzing-dna-mixtures (accessed October 8, 2024).

Press R (2019) DNA Mixtures: A Forensic Science Explainer (April 3, 2019). Available at https://www.nist.gov/featured-stories/dna-mixtures-forensic-science-explainer (accessed October 8, 2024).

Preuße-Prange A, Renneberg R, Schwark T, Poetsch M, Simeoni E, von Wurmb-Schwark N (2009) The problem of DNA contamination in forensic case work: How to get rid of unwanted DNA? *Forensic Science International: Genetics Supplement Series* 2(1):185-186. https://doi.org/10.1016/j.fsigss.2009.08.164

Prinz M, Pirtle D, Oldoni F (2024) Global survey on evaluative reporting on DNA evidence with regard to activity-level propositions. *Journal of Forensic Sciences* 69(3):798-813. https://doi.org/10.1111/1556-4029.15488

Puch-Solis R & Clayton T (2014) Evidential evaluation of DNA profiles using a discrete statistical model implemented in the DNA LiRa software. *Forensic Science International: Genetics* 11:220-228. https://doi.org/10.1016/j.fsigen.2014.04.005

Pullen-Blasnik H, Eyal G, Weissenbach A (2024) 'Is your accuser me, or is it the software?' Ambiguity and contested expertise in probabilistic DNA profiling. *Social Studies of Science* 54(1):30-58. https://doi.org/10.1177/03063127231186646

[QAS 2020] FBI Quality Assurance Standards (QAS) for Forensic DNA Testing Laboratories. (2020). Available at https://www.swgdam.org/publications (accessed August 24, 2024).

[QAS 2020 Audit] FBI Quality Assurance Standards Audit for Forensic DNA Testing Laboratories in Accordance with the Quality Assurance Standards for Forensic DNA Testing Laboratories Effective July 1, 2020. (2020). Available at https://www.swgdam.org/publications (accessed August 24, 2024).

Quinones I & Daniel B (2012) Cell free DNA as a component of forensic evidence recovered from touched surfaces. *Forensic Science International: Genetics* 6:26-30. https://doi.org/10.1016/j.fsigen.2011.01.004

Ramos D & Gonzalez-Rodriguez J (2013) Reliable support: Measuring calibration of likelihood ratios. *Forensic Science International* 230:156-169. https://doi.org/10.1016/j.forsciint.2013.04.014

Ramos D, Gonzalez-Rodriguez J, Zadora G, Aitken C (2013) Information-theoretical assessment of the performance of likelihood ratio computation methods. *Journal of Forensic Sciences* 58(6):1503-1518. https://doi.org/10.1111/1556-4029.12233

Ramos P, Handt O, Taylor D. (2020) Investigating the position and level of DNA transfer to undergarments during digital sexual assault. *Forensic Science International: Genetics* 47:102316. https://doi.org/10.1016/j.fsigen.2020.102316

Rand S, Schürenkamp M, Brinkmann B (2002) The GEDNAP (German DNA profiling group) blind trial concept. *International Journal of Legal Medicine* 116(4):199-206. https://doi.org/10.1007/s00414-002-0285-z

Rand S, Schürenkamp M, Hohoff C, Brinkmann B (2004) The GEDNAP blind trial concept part II. Trends and developments. *International Journal of Legal Medicine* 118(2):83-89. https://doi.org/10.1007/s00414-003-0421-4

Raymond JJ, van Oorschot RAH, Walsh SJ, Roux C (2008a) Trace DNA analysis: Do you know what your neighbour is doing? A multi-jurisdictional survey. *Forensic Science International: Genetics* 2(1):19-28. https://doi.org/10.1016/j.fsigen.2007.07.001

Raymond JJ, Walsh SJ, van Oorschot RAH, Gunn PR, Evans L, Roux C (2008b) Assessing trace DNA evidence from a residential burglary: Abundance, transfer and persistence. *Forensic Science International: Genetics Supplement Series* 1:442-443. https://doi.org/10.1016/j.fsigss.2007.10.040

Raymond JJ, van Oorschot RAH, Gunn PR, Walsh SJ, Roux C (2009a) Trace evidence characteristics of DNA: A preliminary investigation of the persistence of DNA at crime scenes. *Forensic Science International: Genetics* 4(1):26-33. https://doi.org/10.1016/j.fsigen.2009.04.002

Raymond JJ, van Oorschot RAH, Gunn PR, Walsh SJ, Roux C (2009b) Trace DNA success rates relating to volume crime offences. *Forensic Science International: Genetics Supplement Series* 2(1):136-137. https://doi.org/10.1016/j.fsigss.2009.08.072

Riman S, Iyer H, Borsuk LA, Vallone PM (2019a) Understanding the behavior of stutter through the sequencing of STR alleles. *Forensic Science International: Genetics Supplement Series* 7:115-116. https://doi.org/10.1016/j.fsigss.2019.09.045

Riman S, Iyer H, Vallone PM (2019b) Exploring DNA interpretation software using the PROVEDIt dataset. *Forensic Science International: Genetics Supplement Series* 7:724-726. https://doi.org/10.1016/j.fsigss.2019.10.152

Riman S, Iyer H, Borsuk LA, Vallone PM (2020) Understanding the characteristics of sequence-based single-source DNA profiles. *Forensic Science International: Genetics* 44:102192. https://doi.org/10.1016/j.fsigen.2019.102192

Riman S, Iyer H, Vallone PM (2021) Examining performance and likelihood ratios for two likelihood ratio systems using the PROVEDIt dataset. *PLoS ONE* 16(9): e0256714. https://doi.org/10.1371/journal.pone.0256714

Riman S, Iyer H, Vallone PM (2022) A response to a correspondence letter by Buckleton et al. on: Riman et al. (2021) Examining performance and likelihood ratios for two likelihood ratio systems using the PROVEDIt dataset, PLoS One 16(9):e0256714. *Forensic Science International: Genetics* 59: 102710. https://doi.org/10.1016/j.fsigen.2022.102710

Riman S, Bright JA, Huffman K, Moreno LI, Liu S, Sathya A, Vallone PM (2024a) A collaborative study on the precision of the Markov chain Monte Carlo algorithms used for DNA profile interpretation. *Forensic Science International: Genetics* 72:103088. https://doi.org/10.1016/j.fsigen.2024.103088

Riman S, Iyer H, Robitaille E, Liu S, Sathya A, Vallone PM (2024b) A Software Package for Designing and Interpreting Forensic DNA Validation Studies. Poster by Sarah Riman at the International Society for Forensic Genetics on September 11, 2024. Available at https://strbase.nist.gov/NIST_Resources/Presentations/2024/ISFG2024_Riman_SoftwareTool.pdf (accessed October 8, 2024).

Roach JC, Glusman G, Smit AF, Huff CD, Hubley R, Shannon PT, Rowen L, Pant KP, Goodman N, Bamshad M, Shendure J, Drmanac R, Jorde LB, Hood L, Galas DJ (2010) Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* 328(5978):636-639. https://doi.org/10.1126/science.1186802

Roberts P & Stockdale M (eds.) (2018) *Forensic Science Evidence and Expert Witness Testimony: Reliability Through Reform?* Edward Elgar Publishing: Northampton, MA.

Rockmore, D (2024) How much of the world is it possible to model? *The New Yorker* (January 15, 2024). Available at https://www.newyorker.com/culture/annals-of-inquiry/how-much-of-the-world-is-it-possible-to-model (accessed October 31, 2024).

Rodriguez JJRB, Bright J-A, Salvador JM, Laude RP, de Ungria MCA (2019) Probabilistic approaches to interpreting two-person DNA mixtures from post-coital specimens. *Forensic Science International:* 300:157-163. https://doi.org/10.1016/j.forsciint.2019.04.037

Romsos EL, Borsuk LA, Steffen CR, Riman S, Kiesler KM, Vallone PM (2024) Development of a forensic DNA research grade test material. *Journal of Forensic Sciences* (in press). https://doi.org/10.1111/1556-4029.15639

Roux C, Talbot-Wright B, Robertson J, Crispino F, Ribaux O (2015) The end of the (forensic science) world as we know it? The example of trace evidence. *Philosophical Transactions of the Royal Society B* 370:20140260. https://doi.org/10.1098/rstb.2014.0260

Rowan KE, Wellner GA, Grgicak CM (2016) Exploring the impacts of ordinary laboratory alterations during forensic DNA processing on peak height variation, thresholds, and probability of dropout. *Journal of Forensic Sciences* 61(1):177-185. https://doi.org/10.1111/1556-4029.12899

Ruan T, Barash M, Gunn P, Bruce D (2018) Investigation of DNA transfer onto clothing during regular daily activities. *International Journal of Legal Medicine* 132:1035-1042. https://doi.org/10.1007/s00414-017-1736-x

Rutty GN (2000) Human DNA contamination of mortuaries: Does it matter? *Journal of Pathology* 190(4):410-411. https://doi.org/10.1002/(sici)1096-9896(200003)190:4%3C410::aid-path532%3E3.0.co;2-u

Rutty GN, Hopwood A, Tucker V (2003) The effectiveness of protective clothing in the reduction of potential DNA contamination of the scene of crime. *International Journal of Legal Medicine* 117(3):170-174. https://doi.org/10.1007/s00414-002-0348-1

Saiki RK, Walsh PS, Levenson CH, Erlich HA (1989) Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. *Proceedings of the National Academy of Sciences of the United States of America* 86(16):6230-6234. https://doi.org/10.1073/pnas.86.16.6230

Samie L, Hicks T, Castella V, Taroni F (2016) Stabbing simulations and DNA transfer. *Forensic Science International: Genetics* 22:73-80. https://doi.org/10.1016/j.fsigen.2016.02.001

Samie L, Champod C, Taylor D, Taroni F (2020) The use of Bayesian Networks and simulation methods to identify the variables impacting the value of evidence assessed under activity level propositions in stabbing cases. *Forensic Science International: Genetics* 48:102334. https://doi.org/10.1016/j.fsigen.2020.102334

Schneider PM, Gill P, Carracedo A (2006a) Editorial on the recommendations of the DNA commission of the ISFG on the interpretation of mixtures. *Forensic Science International* 160:89. https://doi.org/10.1016/j.forsciint.2006.05.036

Schneider PM, Fimmers R, Keil W, Molsberger G, Patzelt D, Pflug W, Rothämel T, Schmitter H, Schneider H, Brinkmann B (2006b) Allgemeine Empfehlungen der Spurenkommission zur Bewertung von DNA-Mischspuren. *Rechtsmedizin* 16: 401-404. [German version was published in 2009 in English] http://doi.org/10.1007/s00194-006-0411-1

Schneider PM, Fimmers R, Keil W, Molsberger G, Patzelt D, Pflug W, Rothämel T, Schmitter H, Schneider H, Brinkmann B (2009) The German Stain Commission: recommendations for the interpretation of mixed stains. *International Journal of Legal Medicine* 123(1):1-5. [originally published in German in 2006 – *Rechtsmedizin* 16: 401-404] https://doi.org/10.1007/s00414-008-0244-4

Schulte J, Marciano MA, Scheurer E, Schulz I (2023) A systematic approach to improve downstream single-cell analysis for the DEPArray™ technology. *Journal of Forensic Sciences* 68(6):1875-1893. https://doi.org/10.1111/1556-4029.15344

Schulte J, Caliebe A, Marciano M, Neuschwander P, Seiberle I, Scheurer E, Schulz I (2024) DEPArray™ single-cell technology: A validation study for forensic applications. *Forensic Science International: Genetics* 70:103026. https://doi.org/10.1016/j.fsigen.2024.103026

Sense about Science (2017) *Making Sense of Forensic Genetics*. Available at http://senseaboutscience.org/activities/making-sense-of-forensic-genetics/ (accessed August 24, 2024).

Sessa F, Pomara C, Esposito M, Grassi P, Cocimano G, Salerno M (2023) Indirect DNA transfer and forensic implications: a literature review. *Genes* 14(12):2153. https://doi.org/10.3390/genes14122153

Sjerps MJ, Alberink I, Bolck A, Stoel RD, Vergeer P, van Zanten JH (2016) Uncertainty and LR: to integrate or not to integrate, that's the question. *Law, Probability and Risk* 15(1): 23-29. https://doi.org/10.1093/lpr/mgv005

Slooten K (2018) The information gain from peak height data in DNA mixtures. *Forensic Science International: Genetics* 36:119-123. https://doi.org/10.1016/j.fsigen.2018.06.009

Slooten K (2021) The analogy between DNA kinship and DNA mixture evaluation, with applications for the interpretation of likelihood ratios produced by possibly imperfect models. *Forensic Science International: Genetics* 52:102449. https://doi.org/10.1016/j.fsigen.2020.102449

Smith PJ & Ballantyne J (2007) Simplified low-copy-number DNA analysis by post-PCR purification. *Journal of Forensic Sciences* 52(4):820-829. https://doi.org/10.1111/j.1556-4029.2007.00470.x

Stanciu CE, Philpott MK, Kwon YJ, Bustamante EE, Ehrhardt CJ (2015) Optical characterization of epidermal cells and their relationship to DNA recovery from touch samples. *F1000Research* 4:1360. https://doi.org/10.12688/f1000research.7385.1

Steele CD & Balding DJ (2014) Statistical evaluation of forensic DNA profile evidence. *Annual Review of Statistics and Its Application* 1:361-384. https://doi.org/10.1146/annurev-statistics-022513-115602

Steensma K, Ansell R, Clarisse L, Connolly E, Kloosterman AD, McKenna LG, van Oorschot RAH, Szkuta B, Kokshoorn B (2017) An inter-laboratory comparison study on transfer, persistence and recovery of DNA from cable ties. *Forensic Science International: Genetics* 31:95-104. https://doi.org/10.1016/j.fsigen.2017.08.015

Stiffelman B (2019) No longer the gold standard: probabilistic genotyping is changing the nature of DNA evidence in criminal trials. *Berkeley Journal of Criminal Law* 24: 110. https://doi.org/10.15779/Z384Q7QQ6X

Stiffelman B (2021) Juror understanding of the weight of evidence presented as a likelihood ratio and the impact on the deliberative process: a response to Buckleton, et al. *Forensic Science International: Genetics* 51:102454. https://doi.org/10.1016/j.fsigen.2020.102454

Stokes NA, Stanciu CE, Brocato ER, Ehrhardt CJ, Greenspoon SA (2018) Simplification of complex DNA profiles using front end cell separation and probabilistic modeling. *Forensic Science International: Genetics* 36:205-212. https://doi.org/10.1016/j.fsigen.2018.07.004

Stringer P, Scheffer JW, Scott P, Lee J, Goetz R, Ientile V, Eckhoff C, Turbett G, Carroll D, Harbison S (2009) Interpretation of DNA mixtures—Australian and New Zealand consensus on principles. *Forensic Science International: Genetics* 3(2):144-145. https://doi.org/10.1016/j.fsigen.2008.09.003

Susik M, Schönborn H, Sbalzarini IF (2022) Hamiltonian Monte Carlo with strict convergence criteria reduces run-to-run variability in forensic DNA mixture deconvolution. *Forensic Science International: Genetics* 60:102744. https://doi.org/10.1016/j.fsigen.2022.102744

Susik M & Sbalzarini IF (2023a) Analysis of the Hamiltonian Monte Carlo genotyping algorithm on PROVEDIt mixtures including a novel precision benchmark. *Forensic Science International: Genetics* 64:102840. https://doi.org/10.1016/j.fsigen.2023.102840

Susik M & Sbalzarini IF (2023b) Variational inference accelerates accurate DNA mixture deconvolution. *Forensic Science International: Genetics* 65:102890. https://doi.org/10.1016/j.fsigen.2023.102890

Swaminathan H, Qureshi MO, Grgicak CM, Duffy K, Lun DS (2018) Four model variants within a continuous forensic DNA mixture interpretation framework: effects on evidential inference and reporting. *PLOS ONE* 13(11): e0207599. https://doi.org/10.1371/journal.pone.0207599

SWGDAM (2015) SWGDAM Guidelines for the Validation of Probabilistic Genotyping Systems. Available at https://www.swgdam.org/publications (accessed August 24, 2024).

SWGDAM (2016) SWGDAM Validation Guidelines for DNA Analysis Methods. Available at https://www.swgdam.org/publications (accessed August 24, 2024).

SWGDAM (2017a) SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories. Available at https://www.swgdam.org/publications (accessed August 24, 2024).

SWGDAM (2017b) Contamination Prevention and Detection Guidelines for Forensic DNA Laboratories. Available at https://www.swgdam.org/publications (accessed August 24, 2024).

SWGDAM (2018) Recommendations of the SWGDAM Ad Hoc Working Group on Genotyping Results Reported as Likelihood Ratios. Available at https://www.swgdam.org/publications (accessed August 24, 2024).

Swofford H, Lund S, Iyer H, Butler J, Soons J, Thompson R, Desiderio V, Jones JP, Ramotowski R (2024) Inconclusive decisions and error rates in forensic science. *Forensic Science International: Synergy* 8:100472. https://doi.org/10.1016/j.fsisyn.2024.100472

Szkuta B, Harvey ML, Ballantyne KN, van Oorschot RAH (2013) The potential transfer of trace DNA via high risk vectors during exhibit examination. *Forensic Science International: Genetics Supplement Series* 4(1):e55-e56. https://doi.org/10.1016/j.fsigss.2013.10.028

Szkuta B, Harvey ML, Ballantyne KN, van Oorschot RAH (2015a) DNA transfer by examination tools—A risk for forensic casework? *Forensic Science International: Genetics* 16:246-254. https://doi.org/10.1016/j.fsigen.2015.02.004

Szkuta B, Harvey ML, Ballantyne KN, van Oorschot RAH (2015b) Residual DNA on examination tools following use. *Forensic Science International: Genetics Supplement Series* 5:e495-e497. https://doi.org/10.1016/j.fsigss.2015.09.196

Szkuta B, Ballantyne KN, van Oorschot RAH (2017a) Transfer and persistence of DNA on the hands and the influence of activities performed. *Forensic Science International: Genetics* 28:10-20. https://doi.org/10.1016/j.fsigen.2017.01.006

Szkuta B, van Oorschot RAH, Ballantyne KN (2017b) DNA decontamination of fingerprint brushes. *Forensic Science International* 277: 41-50. https://doi.org/10.1016/j.forsciint.2017.05.009

Szkuta B, Ballantyne KN, Kokshoorn B, van Oorschot RAH (2018) Transfer and persistence of non-self-DNA on hands over time: Using empirical data to evaluate DNA evidence given activity level propositions. *Forensic Science International: Genetics* 33:84-97. https://doi.org/10.1016/j.fsigen.2017.11.017

Szkuta B, Ansell R, Boiso L, Connolly E, Kloosterman AD, Kokshoorn B, McKenna LG, Steensma K, van Oorschot RAH (2020) DNA transfer to worn upper garments during different activities and contacts: An inter-laboratory study. *Forensic Science International: Genetics* 46:102268. https://doi.org/10.1016/j.fsigen.2020.102268

Taberlet P, Griffin S, Goossens B, Questiau S, Manceau V, Escaravage N, Waits LP, Bouvet J (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research* 24(16):3189-3194. https://doi.org/10.1093/nar/24.16.3189

Tan B, Zhao Z, Zhang Z, Li S, Li SC (2017) Search for more effective microsatellite markers for forensics with next-generation sequencing. *IEEE Trans Nanobioscience* 16(5):375-381. https://doi.org/10.1109/tnb.2017.2712795

Taroni F, Biedermann A, Vuille J, Morling N (2013) Whose DNA is this? How relevant a question? (a note for forensic scientists). *Forensic Science International: Genetics* 7:467-470. https://doi.org/10.1016/j.fsigen.2013.03.012

Taroni F, Bozza S, Biedermann A, Aitken C (2016) Dismissal of the illusion of uncertainty in the assessment of a likelihood ratio. *Law, Probability and Risk* 15(1):1-16. https://doi.org/10.1093/lpr/mgv008

Taylor D, Bright JA, Buckleton J (2013) The interpretation of single source and mixed DNA profiles. *Forensic Science International: Genetics* 7(5):516-528. https://doi.org/10.1016/j.fsigen.2013.05.011

Taylor D (2014) Using continuous DNA interpretation methods to revisit likelihood ratio behaviour. *Forensic Science International: Genetics* 11:144-153. https://doi.org/10.1016/j.fsigen.2014.03.008

Taylor D, Bright JA, Buckleton J (2014) The 'factor of two' issue in mixed DNA profiles. *Journal of Theoretical Biology* 363:300-306. https://doi.org/10.1016/j.jtbi.2014.08.021

Taylor D & Buckleton J (2015) Do low template DNA profiles have useful quantitative data? *Forensic Science International: Genetics* 16:13-16. https://doi.org/10.1016/j.fsigen.2014.11.001

Taylor D, Abarno D, Rowe E, Rask-Nielsen L (2016) Observations of DNA transfer within an operational forensic biology laboratory. *Forensic Science International: Genetics* 23:33-49. https://doi.org/10.1016/j.fsigen.2016.02.011

Taylor D, Biedermann A, Samie L, Pun KM, Hicks T, Champod C (2017) Helping to distinguish primary from secondary transfer events for trace DNA. *Forensic Science International: Genetics* 28:155-177. https://doi.org/10.1016/j.fsigen.2017.02.008

Taylor D, Kokshoorn B, Biedermann A (2018) Evaluation of forensic genetics findings given activity level propositions: A review. *Forensic Science International: Genetics* 36:34-49. https://doi.org/10.1016/j.fsigen.2018.06.001

Taylor D, Samie S, Champod C (2019) Using Bayesian networks to track DNA movement through complex transfer scenarios. *Forensic Science International: Genetics* 42:69-80. https://doi.org/10.1016/j.fsigen.2019.06.006

Taylor D & Balding D (2020) How can courts take into account the uncertainty in a likelihood ratio? *Forensic Science International: Genetics* 48:102361. https://doi.org/10.1016/j.fsigen.2020.102361

Taylor D & Kokshoorn B (2023) *Forensic DNA Trace Evidence Interpretation: Activity Level Propositions and Likelihood Ratios*. CRC Press: Boca Raton.

Taylor D, Kokshoorn B, Champod C (2024) A practical treatment of sensitivity analyses in activity level evaluations. *Forensic Science International* 355:111944. https://doi.org/10.1016/j.forsciint.2024.111944

[TFSC 2024] Texas Forensic Science Commission (July 26, 2024) Final Report on Complaint No. 23.67; Tiffany Roy; (Timothy Kalafut, Ph.D.; Evaluation of Biological/DNA Results Given Activity Level Propositions). Available at https://www.txcourts.gov/media/1458950/final-report-complaint-2367-roy-tiffany-073024_redacted.pdf (accessed September 5, 2024).

Thompson WC & Schumann EL (1987) Interpretation of statistical evidence in criminal trials: The prosecutor's fallacy and the defense attorney's fallacy. *Law and Human Behavior 11*(3):167-187. https://doi.org/10.1007/BF01044641

Thompson WC (2009) Painting the target around the matching profile: The Texas sharpshooter fallacy in forensic DNA interpretation. *Law, Probability and Risk* 8(3):257-276. https://doi.org/10.1093/lpr/mgp013

Thompson WC & Newman EJ (2015) Lay understanding of forensic statistics: evaluation of random match probabilities, likelihood ratios, and verbal equivalents. *Law and Human Behavior* 39(4):332-349. https://doi.org/10.1037/lhb0000134

Thompson W, Black J, Jain A, Kadane J (2017) *Forensic Science Assessments: A Quality and Gap Analysis – Latent Fingerprint Examination.* American Association for the Advancement of Science. Available at https://www.aaas.org/resources/latent-fingerprint-examination (accessed August 24, 2024).

Thompson WC (2023) Uncertainty in probabilistic genotyping of low template DNA: A case study comparing STRMix™ and TrueAllele™. *Journal of Forensic Sciences* 68(3):1049-1063. https://doi.org/10.1111/1556-4029.15225

Tobias SHA, Jacques GS, Morgan RM, Meakin GE (2017) The effect of pressure on DNA deposition by touch. *Forensic Science International: Genetics Supplement Series* 6:e12-e14. https://doi.org/10.1016/j.fsigss.2017.09.020

Torres Y, Flores I, Prieto V, López-Soto M, Farfán MJ, Carracedo A, Sanz P (2003) DNA mixtures in forensic casework: a 4-year retrospective study. *Forensic Science International* 134(2-3):180-186. https://doi.org/10.1016/s0379-0738(03)00161-0

[UKFSR 2015] UK Forensic Science Regulator (2015) *Guidance: The Control and Avoidance of Contamination in Laboratory Activities Involving DNA Evidence Recovery and Analysis*. FSR-G-208. Changed to *Guidance: DNA Contamination Controls: Laboratory* (FSR-GUI-0018). Available at

https://www.gov.uk/government/publications/laboratory-dna-anti-contamination-guidance (accessed October 31, 2024).

[UKFSR 2016a] UK Forensic Science Regulator (2016a) *Guidance: DNA Anti-Contamination—Forensic Medical Examination in Sexual Assault Referral Centres and Custodial Facilities*. FRS-G-207. Changed to *Guidance: DNA Contamination Controls: Forensic Medical Examinations* (FSR-GUI-0017). Available at https://www.gov.uk/government/publications/sexual-assault-referral-centres-and-custodial-facilities-dna-anti-contamination (accessed October 31, 2024).

[UKFSR 2016b] UK Forensic Science Regulator (2016b) *Guidance: The Control and Avoidance of Contamination in Crime Scene Examination Involving DNA Evidence Recovery*. FSR-G-206. Changed to *Guidance: Crime Scene DNA: Anti-contamination* (FSR-GUI-0016). Available at https://www.gov.uk/government/publications/crime-scene-dna-anti-contamination-guidance (accessed October 31, 2024).

[UKFSR 2018a] UK Forensic Science Regulator (2018a) DNA mixture interpretation, FSR-G-222. Available at https://www.gov.uk/government/publications/dna-mixture-interpretation-fsr-g-222 (accessed October 31, 2024).

[UKFSR 2018b] UK Forensic Science Regulator (2018b) Software validation for DNA mixture interpretation, FSR-G-223. Available at https://www.gov.uk/government/publications/software-validation-for-dna-mixture-interpretation-fsr-g-223 (accessed October 31, 2024).

[UKFSR 2020] UK Forensic Science Regulator (2020) Proficiency Testing Guidance for DNA Mixture Analysis and Interpretation. FSR-G-224. Available at https://www.gov.uk/government/publications/proficiency-testing-guidance-dna-mixture-analysis-and-interpretation (accessed October 31, 2024).

[UKFSR 2021] UK Forensic Science Regulator (2021) Development of Evaluative Opinions (FSR-C-118). Available at https://www.gov.uk/government/publications/development-of-evaluative-opinions (accessed October 31, 2024).

Vallone PM, Romsos EL, Borsuk LA, Steffen CR, Riman S, Kiesler KM (2024) NIST Research Grade Test Materials: A New Collaborative Effort to Address Measurement Challenges. Presentation by Peter Vallone at the International Society for Forensic Genetics on September 11, 2024. Available at https://strbase.nist.gov/NIST_Resources/Presentations/2024/ISFG2024_Vallone_RGTM.pdf (accessed October 8, 2024).

van den Berge M, Ozcanhan G, Zijlstra S, Lindenbergh A, Sijen T (2016) Prevalence of human cell material: DNA and RNA profiling of public and private objects and after activity scenarios. *Forensic Science International: Genetics* 21:81-89. https://doi.org/10.1016/j.fsigen.2015.12.012

van der Gaag KJ, de Leeuw RH, Hoogenboom J, Patel J, Storts DR, Laros JFJ, de Knijff P (2016) Massively parallel sequencing of short tandem repeats-Population data and mixture analysis results for the PowerSeq™ system. *Forensic Science International: Genetics* 24:86-96. https://doi.org/10.1016/j.fsigen.2016.05.016

van der Gaag KJ, de Leeuw RH, Laros JFJ, den Dunnen JT, de Knijff P (2018) Short hypervariable microhaplotypes: A novel set of very short high discriminating power loci without stutter artefacts. *Forensic Science International: Genetics* 35:169-175. https://doi.org/10.1016/j.fsigen.2018.05.008

Vandewoestyne M, Van Hoofstat D, Franssen A, Van Nieuwerburgh F, Deforce D (2013) Presence and potential of cell free DNA in different types of forensic samples. *Forensic Science International: Genetics* 7(2):316-320. https://doi.org/10.1016/j.fsigen.2012.12.005

van Oorschot RA & Jones MK (1997) DNA fingerprints from fingerprints. *Nature* 387:767. https://doi.org/10.1038/42838

van Oorschot RA, Treadwell S, Beaurepaire J, Holding NL, Mitchell RJ (2005) Beware of the possibility of fingerprinting techniques transferring DNA. *Journal of Forensic Sciences* 50(6):1417-1422.

van Oorschot RAH (2012) Assessing DNA profiling success rates: Need for more and better collection of relevant data. *Forensic Science Policy and Management* 3(3):37-41. https://doi.org/10.1080/19409044.2012.719581

van Oorschot RA, Glavich G, Mitchell RJ (2014a) Persistence of DNA deposited by the original user on objects after subsequent use by a second person. *Forensic Science International: Genetics* 8(1):219-225. https://doi.org/10.1016/j.fsigen.2013.10.005

van Oorschot RA, McArdle R, Goodwin WH, Ballantyne KN (2014b) DNA transfer: The role of temperature and drying time. *Legal Medicine* 16(3):161-163. https://doi.org/10.1016/j.legalmed.2014.01.005

van Oorschot RAH, Found B, Ballantyne KN (2015) Considerations relating to the components of a laboratory DNA contamination minimisation monitoring (DCMM) program. *Forensic Science Policy and Management* 6(3-4):91-105. https://doi.org/10.1080/19409044.2015.1085926

van Oorschot RAH, Szkuta B, Ballantyne KN, Goray M (2017) Need for dedicated training, competency assessment, authorisations and ongoing proficiency testing for those addressing DNA transfer issues. *Forensic Science International: Genetics Supplement Series* 6:e32-e34. https://doi.org/10.1016/j.fsigss.2017.09.013

van Oorschot RAH, Szkuta B, Meakin GE, Kokshoorn B, Goray M (2019) DNA transfer in forensic science: A review. *Forensic Science International: Genetics* 38:140-166. https://doi.org/10.1016/j.fsigen.2018.10.014

van Oorschot RAH, Meakin GE, Kokshoorn B, Goray M, Szkuta B (2021) DNA transfer in forensic science: recent progress towards meeting challenges. *Genes* 12(11): 1766. https://doi.org/10.3390/genes12111766

Verdon TJ, Mitchell RJ, van Oorschot RAH (2013) The influence of substrate on DNA transfer and extraction efficiency. *Forensic Science International: Genetics* 7(1):167-175. https://doi.org/10.1016/j.fsigen.2012.09.004

Verdon TJ, Mitchell RJ, Chen W, Xiao K, van Oorschot RA (2015) FACS separation of non-compromised forensically relevant biological mixtures. *Forensic Science International: Genetics* 14:194-200. https://doi.org/10.1016/j.fsigen.2014.10.019

Vilsen SB, Tvedebrink T, Eriksen PS, Bøsting C, Hussing C, Mogensen HS, Morling N (2018a) Stutter analysis of complex STR MPS data. *Forensic Science International: Genetics* 35:107-112. https://doi.org/10.1016/j.fsigen.2018.04.003

Vilsen SB, Tvedebrink T, Eriksen PS, Hussing C, Børsting C, Morling N (2018b) Modelling allelic drop-outs in STR sequencing data generated by MPS. *Forensic Science International: Genetics* 37:6-12. https://doi.org/10.1016/j.fsigen.2018.07.017

Vincent FHR (2010) Report: Inquiry into the circumstances that led to the conviction of Mr. Farah Abdulkadir Jama. Available at http://netk.net.au/DNA/Jama.pdf (accessed August 26, 2024).

Vosk T & Emery AF (2014) *Forensic Metrology: Scientific Measurement and Inference for Lawyers, Judges, and Criminalists*. CRC Press: Boca Raton.

Voskoboinik L, Amiel M, Reshef A, Gafny R, Barash M (2017) Laundry in a washing machine as a mediator of secondary and tertiary transfer. *International Journal of Legal Medicine* 123(2):373-378. https://doi.org/10.1007/s00414-017-1617-3

Voskoboinik L, Motro U, Darvasi A (2018) Facilitating complex DNA mixture interpretation by sequencing highly polymorphic haplotypes. *Forensic Science International: Genetics* 35:136-140. https://doi.org/10.1016/j.fsigen.2018.05.001

Walsh PS, Erlich HA, Higuchi R (1992) Preferential PCR amplification of alleles: mechanisms and solutions. *PCR Methods & Applications* 1(4):241-250. https://doi.org/10.1101/gr.1.4.241

Walsh PS, Fildes NJ, Reynolds R (1996) Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. *Nucleic Acids Research* 24(14):2807-2812. https://doi.org/10.1093/nar/24.14.2807

Wang C, Stanciu CE, Ehrhardt CJ, Yadavalli VK (2017) Nanoscale characterization of forensically relevant epithelial cells and surface associated extracellular DNA. *Forensic Science International* 277:252-258. https://doi.org/10.1016/j.forsciint.2017.06.019

Ward D, Henry J, Taylor D (2022) Analysis of mixed DNA profiles from the RapidHIT ID platform using probabilistic genotyping software STRmix. *Forensic Science International: Genetics* 58: 102664. https://doi.org/10.1016/j.fsigen.2022.102664

Weinberg W (1908) Uber den Nachwels der Vererbung beim Menschen. *Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg* 64:369-382 [English translation: “On the demonstration of heredity in man,” Papers on Human Genetics, Prentice-Hall (1963) translation by S.H. Boyer].

Weir BS, Triggs CM, Starling L, Stowell LI, Walsh KA, Buckleton J (1997) Interpreting DNA mixtures. *Journal of Forensic Sciences* 42(2):213-222.

Weir BS (2001) DNA match and profile probabilities: comment on Budowle et al. (2000) and Fung and Hu (2000). *Forensic Science Communications* 3(1). Available at https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/jan2001/weir.htm (accessed August 26, 2024).

Westen AA, Nagel JH, Benschop CC, Weiler NE, de Jong BJ, Sijen T (2009) Higher capillary electrophoresis injection settings as an efficient approach to increase the sensitivity of STR typing. *Journal of Forensic Sciences* 54(3):591-598. https://doi.org/10.1111/j.1556-4029.2009.01022.x

Whitaker JP, Cotton EA, Gill P (2001) A comparison of the characteristics of profiles produced with the AMPFlSTR SGM Plus multiplex system for both standard and low copy number (LCN) STR DNA analysis. *Forensic Science International* 123(2-3):215-223. https://doi.org/10.1016/s0379-0738(01)00557-6

Wickenheiser RA (2002) Trace DNA: A review, discussion of theory, and application of the transfer of trace quantities of DNA through skin contact. *Journal of Forensic Sciences* 47(3):442-450.

Wickenheiser RA (2006) General guidelines for categorization and interpretation of mixed STR DNA profiles. *Canadian Society of Forensic Sciences Journal* 39:179-216. https://doi.org/10.1080/00085030.2006.10757144

Wickenheiser R & Farrell L (2020) Collaborative versus traditional method validation approach: discussion and business case. *Forensic Science International: Synergy* 2:230-237. https://doi.org/10.1016/j.fsisyn.2020.08.003

Williamson VR, Laris TM, Romano R, Marciano MA (2018) Enhanced DNA mixture deconvolution of sexual offense samples using the DEPArray™ system. *Forensic Science International: Genetics* 34:265-276. https://doi.org/10.1016/j.fsigen.2018.03.001

Word CJ (2011) Mixture interpretation: why is it sometimes so hard? *Profiles in DNA (Promega)*. Available at https://www.promega.com/resources/profiles-in-dna/2011/mixture-interpretation-why-is-it-sometimes-so-hard/ (accessed August 26, 2024).

Worth K (2018) Framed for murder by his own DNA. PBS Frontline. Available at https://www.pbs.org/wgbh/frontline/article/framed-for-murder-by-his-own-dna/ (accessed October 7, 2024).

Yang YJ, Prinz M, McKiernan H, Oldoni F (2022) American forensic DNA practitioners' opinion on activity level evaluative reporting. *Journal of Forensic Sciences* 67: 1357-1369. https://doi.org/10.1111/1556-4029.15063

You Y & Balding D (2019) A comparison of software for the evaluation of complex DNA profiles. *Forensic Science International: Genetics* 40:114-119. https://doi.org/10.1016/j.fsigen.2019.02.014

Young BA, Gettings KB, McCord B, Vallone PM (2019) Estimating number of contributors in massively parallel sequencing data of STR loci. *Forensic Science International: Genetics* 38:15-22. https://doi.org/10.1016/j.fsigen.2018.09.007

Zadora G, Martyna A, Ramos D, Aitken C (2014) *Statistical Analysis in Forensic Science: Evidential Value of Multivariate Physicochemical Data.* Wiley: Hoboken, NJ.

Zeng X, King JL, Budowle B (2017) Investigation of the STR loci noise distributions of PowerSeq™ Auto System. *Croatian Medical Journal* 58(3):214-221. https://doi.org/10.3325/cmj.2017.58.214

Zoppis S, Muciaccia B, D'Alessio A, Ziparo E, Vecchiotti C, Filippini A (2014) DNA fingerprinting secondary transfer from different skin areas: Morphological and genetic studies. *Forensic Science International: Genetics* 11:137-143. https://doi.org/10.1016/j.fsigen.2014.03.005

Appendix A. Glossary and Acronyms

Allele
one of two or more versions of a genetic sequence at a particular location (a locus) in the genome; alleles targeted in STR analysis can vary by sequence in addition to length

Allele drop-in
allele peak(s) in an electropherogram (EPG) that are not reproducible across multiple independent amplification events; also, a hypothesis/postulate for the observation of one or more allelic peaks in an electropherogram that are inconsistent with the assumed/known contributor(s) to a sample

Allele (or locus) drop-out
loss of allele (or both alleles) information from a DNA profile; failure of an otherwise amplifiable allele to produce a signal above the analytical threshold because the allele was not present, or was not present in sufficient quantity, in the aliquot that underwent polymerase chain reaction (PCR) amplification

Amplification
an increase in the number of copies of a specific DNA fragment; in forensic DNA testing laboratories, this refers to the use of the PCR technique to produce copies of DNA alleles at specific genetic loci

Artifact
any non-allelic product of the amplification process (e.g., a stutter product), an anomaly of the detection process, such as spectral pull-up, or a dye blob, which is by-product of primer synthesis, that may be observed in an electropherogram; may complicate interpretation of a DNA profile when they cannot be distinguished from actual allele(s) data

Bracketing approach
considers results from samples that are more complex or less complex than the casework sample of interest as a pragmatic way of understanding case-specific reliability of an interpretation system

Binary method
an interpretation approach in which there are only two values (possible or not possible) for each decision

CE
capillary electrophoresis; an electrophoretic technique for separating DNA or other molecules by their size or charge based on migration through a narrow glass tube filled with a liquid polymer

Complex mixture
a DNA profile resulting from comingled DNA of two or more contributors that is difficult to interpret due to uncertainty in the determination of contributor genotypes; factors complicating mixture interpretation include, but are not limited to, low quantity DNA, low quality (degraded) DNA, the number of contributors, and the amount of allele sharing

Contamination
inadvertent introduction of biological material including DNA alleles into a sample at any stage from collection to testing

Continuous approach
a statistical model and accompanying probabilistic genotyping method that evaluates DNA profiles using peak height information to assign weights to the observed peak heights for different combinations of contributor genotypes at all tested loci

CPI
combined probability of inclusion; the product of the probabilities of inclusion calculated for each locus; the probability of inclusion at each locus estimates the probability that a randomly selected, unrelated individual is not excluded from being one of the sources of DNA present in a mixture profile and is calculated as the square of the sum of the relative frequencies of the observed alleles at the locus; sometimes referred to as Random Man Not

Excluded (RMNE); can only be appropriately used when all alleles from all contributors are present in the DNA profile

Deconvolution
separation of component DNA genotypes of contributors to a mixed DNA profile based on quantitative peak height information and any underlying assumptions (e.g., the number of contributors to the mixture, mixture ratios, or known contributors)

Discrete approach
a statistical model and accompanying probabilistic genotyping method that evaluates DNA profiles solely on the presence or absence of alleles without considering peak height information and utilizes probabilities of allele drop-out and drop-in

DNA
deoxyribonucleic acid

DNA mixture
sample that contains DNA from more than one individual

DNA mixture interpretation
an effort to (1) infer possible genotypes for detectable sample contributors (a process sometimes referred to as *deconvolution* of the mixture components) and (2) provide the strength of evidence for a person of interest being part of an evidentiary DNA profile

DNA profile
a string of values (numbers or letters) compiled from the results of DNA testing at one or more genetic markers (loci); can be single-source or a mixture from multiple contributors

EPG
electropherogram; graphic representation of the separation of molecules by electrophoresis in which data appear as “peaks” along a line; the format in which DNA typing results are presented with the horizontal axis displaying the observed peaks (which could be STR alleles or artifacts such as stutter products) in order of increasing size and the vertical axis recording the relative amount of DNA detected based on the fluorescent signal collected

Empirical (assessments/data/methods)
information gathered by direct observation

Factor space and **factor space coverage**
the totality of scenarios and associated variables (factors) that are considered likely to occur in actual casework; with DNA mixture interpretation, factors include the number of contributors, the degree of allele sharing, the ratios of mixture components, and the amount and quality of the DNA tested

Genotype
the variation in a DNA sequence that distinguishes one individual of a species, also described as the genetic constitution of an individual organism; the pair of alleles present at a tested STR locus

Ground truth
a situation where the correct answer is known by design

Interpretation
the process of giving meaning to findings; includes data and statistical analysis and usually produces an opinion on evidence examined

Known samples
biological material whose identity or DNA type is established

Locus (pl. Loci)
a unique physical location of a gene (or a specific sequence of DNA in the case of STRs) on a chromosome; the plural form of locus is pronounced /LOW-sigh/

LR
likelihood ratio; the probability of the evidence under one proposition divided by the probability of the evidence under an alternative, mutually exclusive proposition; the magnitude of its value is commonly used to express a strength of the evidence based on the propositions proposed

Measurand
property intended to be measured

Measurement
an experimental or computational process that, by comparison with a standard, produces an estimate of the true value of a property of a material or virtual object or collection of objects, or of a process, event, or series of events, together with an evaluation of the uncertainty associated with that estimate and intended for use in support of decision-making

Microhaplotypes
regions of DNA containing two or more closely linked single nucleotide polymorphisms (SNPs) associated with multiple allelic combinations (haplotypes)

Next generation sequencing
a high-throughput DNA sequencing technology where millions or billions of DNA strands can be sequenced in parallel; also called massively parallel sequencing

ng
nanogram; a billionth of a gram (10^{-9} g); there is 1 ng of DNA in ≈150 human cells

NIST
National Institute of Standards and Technology

NRC
National Research Council

PCR
polymerase chain reaction; an *in vitro* process that yields millions of copies of targeted DNA regions through repeated cycling of a biochemical reaction involving a DNA polymerase enzyme

pg
picogram; a trillionth of a gram (10^{-12} g); there are ≈6 pg of DNA in a single diploid human cell

PGS
probabilistic genotyping software; a computer program that utilizes statistical genetics, biological models, computer algorithms, and probability distributions to infer genotypes and assign likelihood ratios using either discrete or continuous approaches

Principles
fundamental, primary, or general scientific laws or truths from which others are derived

Proficiency test
a quality assurance measure used to monitor performance of a scientist and identify areas in which improvement may be needed; can be internal (produced by the agency undergoing the test) or external (produced by an outside test provider); proficiency tests can be either open (where the scientist is aware the samples being tested are a proficiency test) or blind (where the scientist is unaware the samples being tested are a proficiency test)

Reliability
providing consistently accurate results

Relevance
the quality or state of being closely connected or appropriate

RFLP
restriction fragment length polymorphism; an analysis method used in early DNA testing

RFU
relative fluorescence unit; an arbitrary measure of the heights of peaks in an electropherogram

ROC curve
receiver operating characteristic curve; a graphical plot that examines the relationship between sensitivity (fraction of true positives) and specificity (fraction of false positives)

SRM
Standard Reference Material; a certified reference material supplied by NIST

Stochastic effects or variation
the observation of intra-locus peak imbalance and/or allele drop-out resulting from random, disproportionate amplification of alleles in low-quantity DNA samples; allele drop-in and elevated stutter product levels may also result

STR
short tandem repeat; an identical (or similar) DNA sequence arranged in direct succession where the repeat sequence unit is 2 base pairs (bp) to 6 bp in length; the number of repeat units varies among individuals

SWGDAM
Scientific Working Group on DNA Analysis Methods; formerly known as TWGDAM, Technical Working Group on DNA Analysis Methods; an FBI-sponsored group that develops quality assurance standards and guidelines for forensic DNA and DNA databasing laboratories in the United States and Canada

Uncertainty
the lack of certainty or sureness of an event; measurement uncertainty is the doubt about the true value of the measurand [property intended to be measured] that remains after making a measurement (see Possolo 2015)