

# A SOFTWARE PACKAGE FOR DESIGNING AND INTERPRETING FORENSIC DNA VALIDATION STUDIES

Sarah Riman<sup>1</sup>, Hari Iyer<sup>1</sup>, Edgar Robitaille<sup>2</sup>, Sicen Liu<sup>2</sup>, Asmitha Sathya<sup>2</sup>, Peter M Vallone<sup>1</sup>

<sup>1</sup>U.S. National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD

<sup>2</sup>U.S. Johns Hopkins University, 3400 N. Charles St, Baltimore, MD

## Background:

- Internal validation of measuring and interpreting single-source and mixture DNA profiles is essential in every forensic DNA laboratory.
- Validation is not a one-time process as laboratories continue to revalidate when changes or upgrades are introduced to their workflows.
- The field lacks much-needed open-source software that can assist in designing validation experiments and interpreting the resulting data.

## Motivation:

- Develop a standalone software (i.e., not running over a network) accessible to users through an easy-to-use graphical user interface (GUI) that can help practitioners in (1) designing validation studies to adequately cover a user selected factor space and (2) interpreting and visualizing the data from the validation studies.

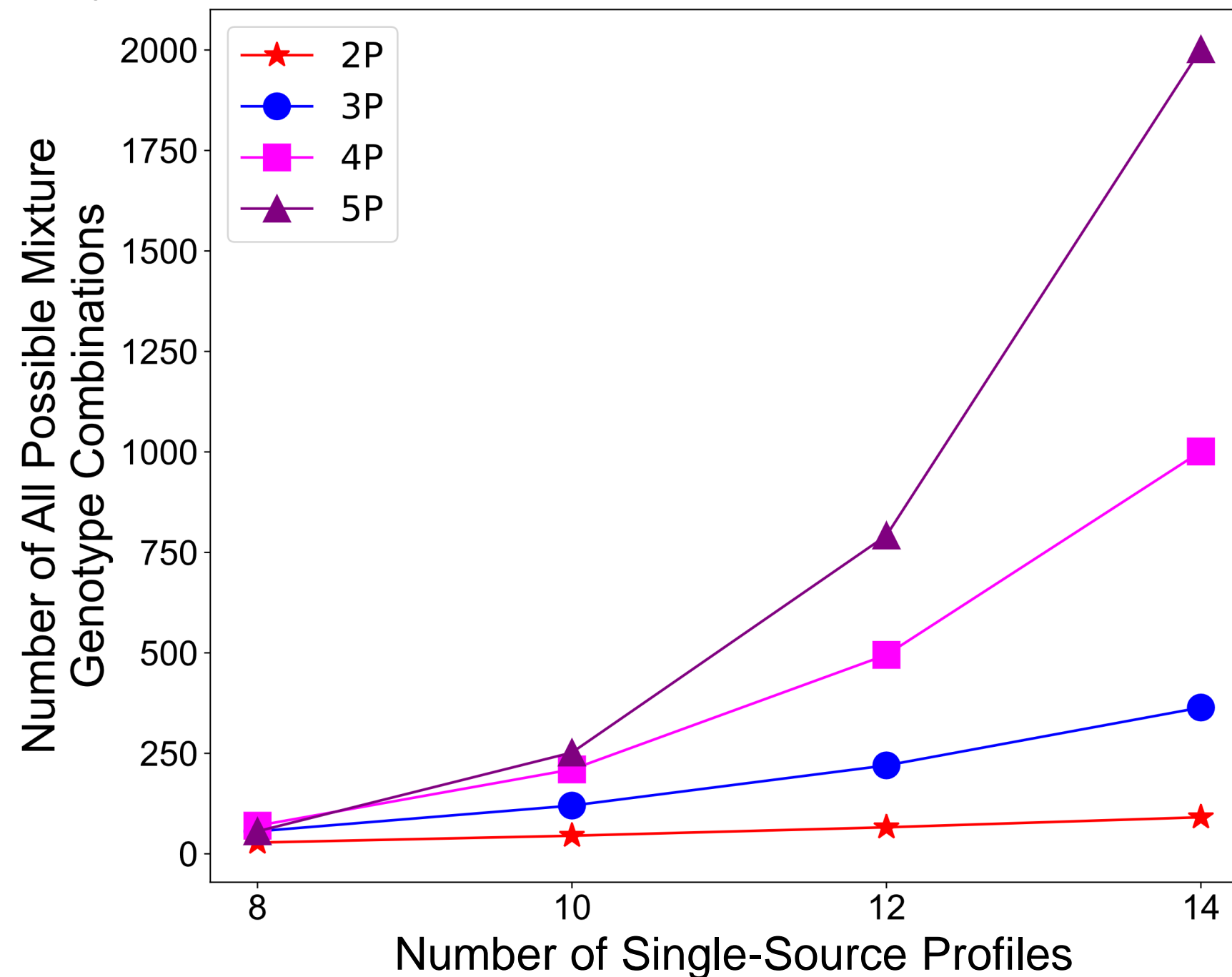
## Software Key Features

### 1 Construction of all possible mixture genotype combinations

Mixture Genotype Combination Formula:

$$\text{Combinations, } {}_nC_r = \frac{n!}{r!(n-r)!}$$

${}_nC_r$  = number of possible genotype combinations  
 $n$  = number of single-source profiles  
 $r$  = number of contributors

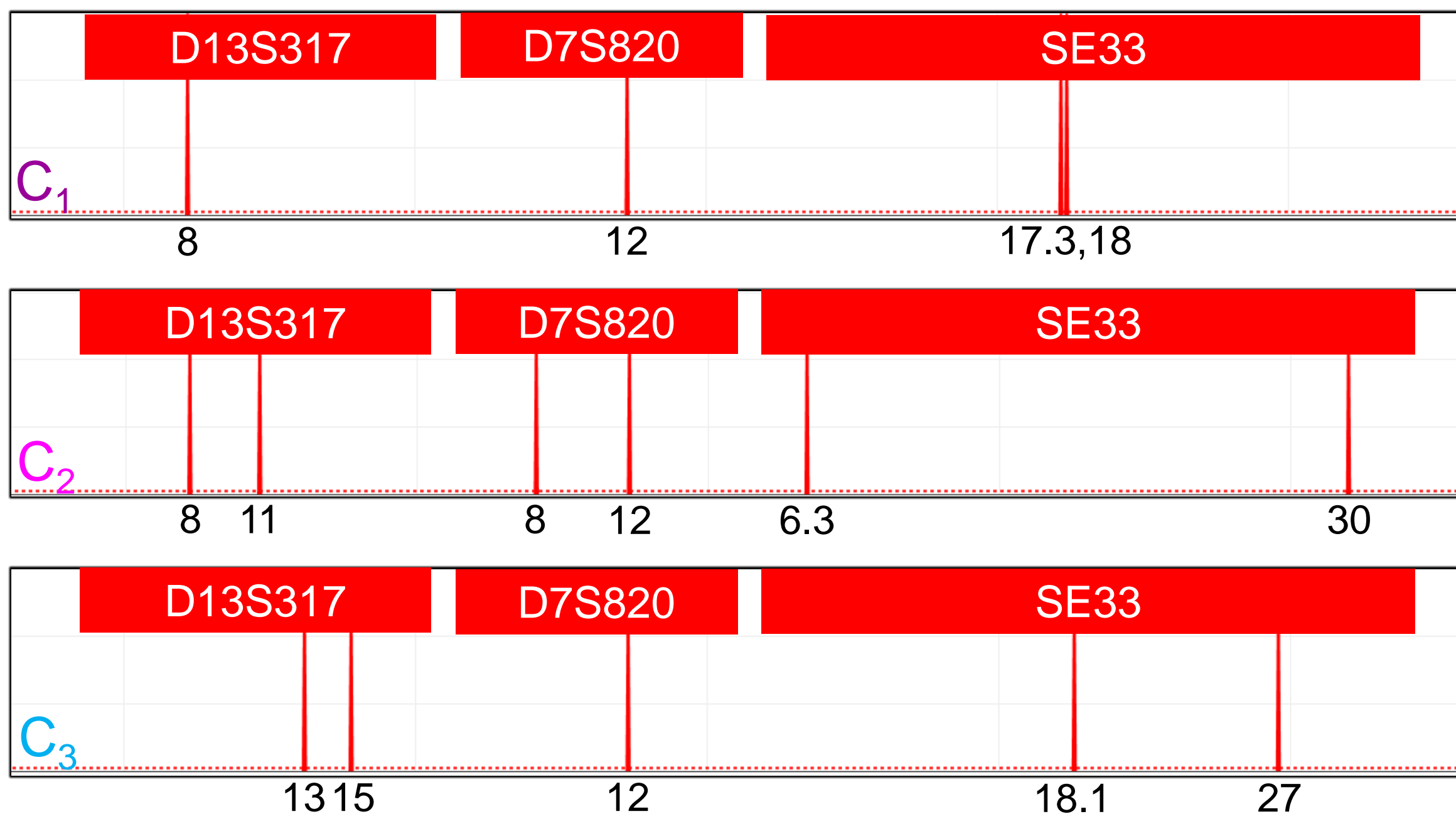


The tool will:

- Support loading user-provided ground truth single-source profiles genotyped by any STR multiplex kit in a txt file format.
- Generate all possible mixture genotype combinations depending on the number of single-source profiles and number of contributors (NoCs) chosen.
- Compute various metrics of interest for each generated mixture and select those mixtures that meet user specified requirements for developing an experimental design for the validation study.

### 2 Choosing a validation experimental design

A. As an illustration, the genotype for three loci from red dye channel are shown for three single-source samples ( $C_1$ ,  $C_2$ , &  $C_3$ )



This functionality is still under development and will aid the user with their choice of combinations

B. All Mixture genotype combinations with per locus statistic metrics simulated from  $C_1$ ,  $C_2$ , &  $C_3$  profiles to experimentally prepare 2P mixtures



C. Summary statistics across each simulated mixture combination (N=21 loci)

Mixture Combinations	$\Sigma$ Homozygote Counts	$\Sigma$ A-A 1 bp Difference	Min/Max # of Expected Alleles	$\Sigma$ Expected Alleles	ASR
$C_1, C_2$	5	1	1/4	57	0.58
$C_1, C_3$	14	2	2/4	64	0.2
$C_2, C_3$	9	0	3/4	72	0.09

Allele-allele sharing ratio (ASR) between the contributors that constitute each combination; ranges between 0 (no sharing) to 1 (maximum sharing).

$$ASR = 1 - \frac{(\text{actual number of observed peaks} - \text{minimum possible number of peaks})}{(\text{maximum possible number of peaks} - \text{minimum possible number of peaks})}$$

- Homozygosity:** counts of homozygote genotypes at a locus.
- A-A (1 bp):** instances of a single base-pair difference between two alleles at a given locus.
- Number of alleles:** counts of alleles expected to be observed.
- Expected alleles to be observed.

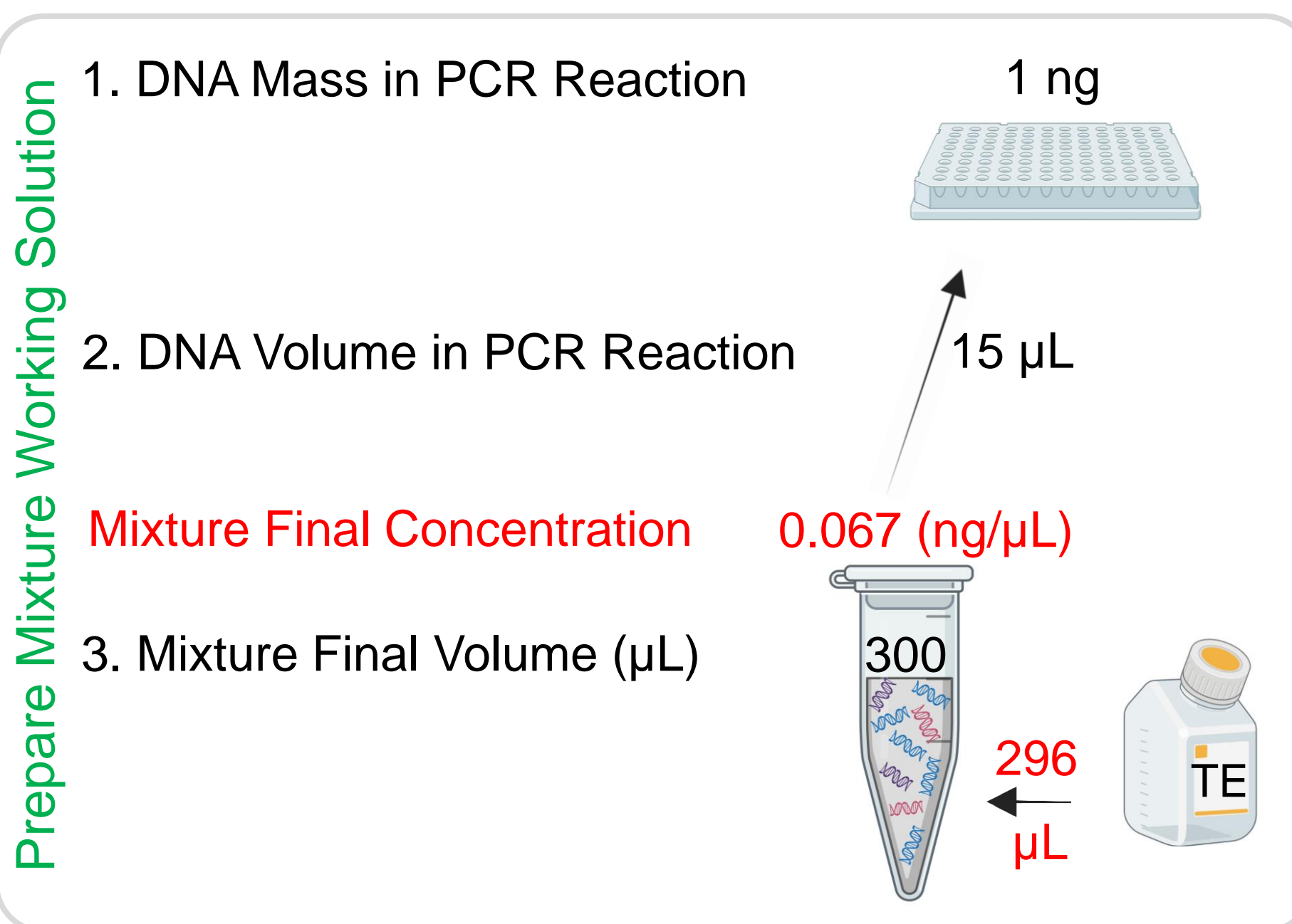
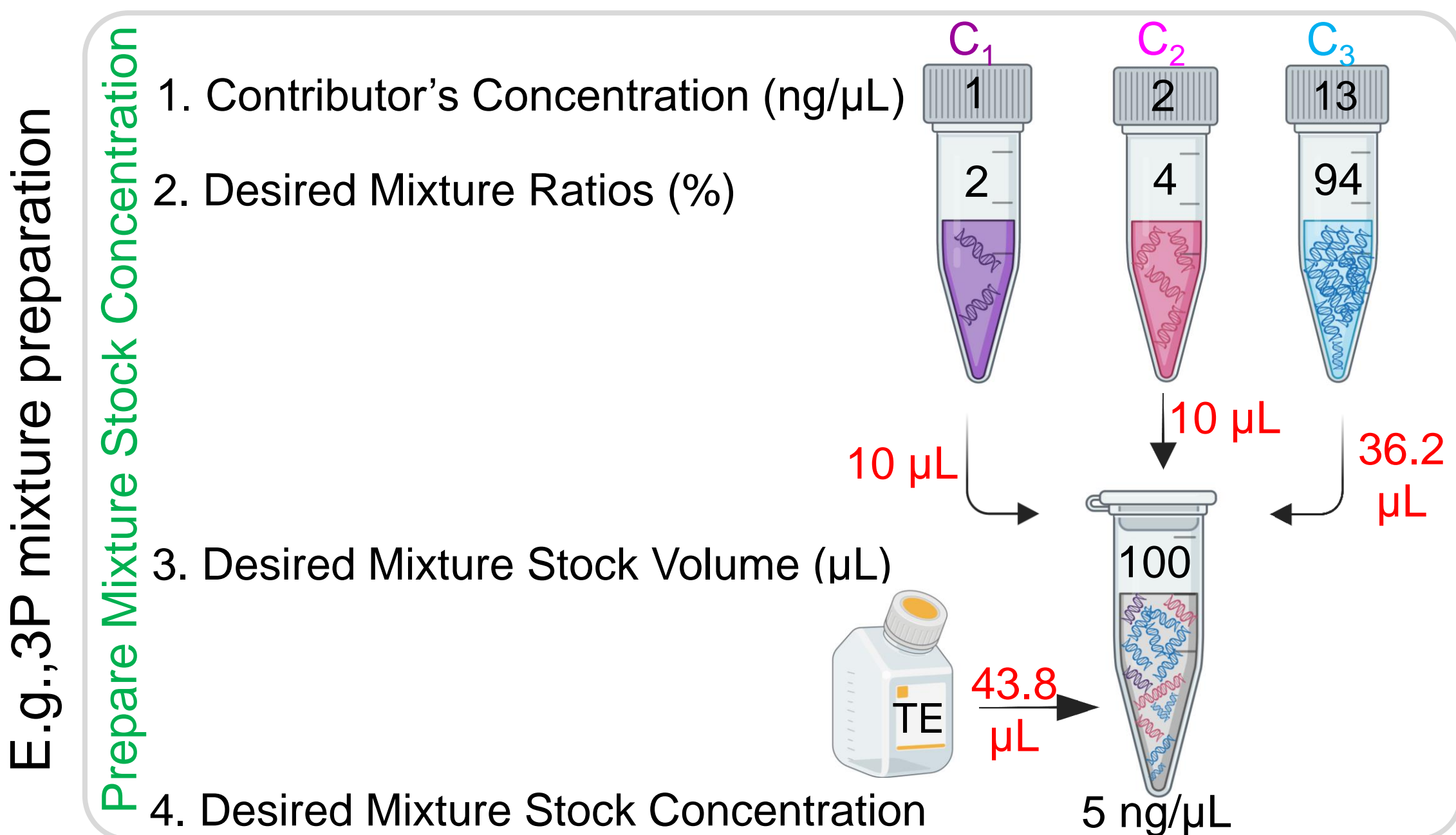
D. The software will ask the user to input the factor space desired to be covered by specifying:

- \* Experimental NoC
- \* Apparent NoC
- \* DNA quality (pristine or degraded)
- \* Total template amounts
- \* Level of A-A 1bp
- \* Level of allele sharing
- \* Contributor's template amounts or mixture ratios
- \* Number of runs per NoC value

Using statistical theory of factorial and fractional factorial experimental designs, the software will output candidate experimental plans to ensure reasonable coverage of the factor space based on user specifications.

### 3 Mixture Calculations

The software will take user's requirements (e.g., C's concentration, desired mixture ratios) and constraints (minimum pipetting amounts, DNA mass in PCR reaction, minimum mixture stock solution) and provide the optimal strategy for making the desired mixtures.



## Key Points

- We present the initial stages of the development of a standalone (i.e., not running over a network) and easy-to-use GUI.
- The GUI is designed to be integrated into the validation workflows.
- The tool will assist the practitioners in automating: (1) the design of validation samples, and (2) the interpretation and visualization of the resulting data.

**Funding:** This work was supported by the Accelerating Forensic Innovation for Impact through the NIST Special Programs Office: Forensic Genetics.  
**Disclaimer:** Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Commerce. Certain commercial software, instruments, and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by NIST, nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose. All work involving NIST samples has been reviewed and approved by the NIST Research Protections Office.  
**Acknowledgements:** We thank the U.S. forensic labs (CalDoJ, Jefferson County Regional Crime Laboratory, ATF, and Los Angeles County Sheriff's Department) for their valuable input and feedback on the software functionalities. We also thank Ashwatha Suresh from JHU Whiting School of Engineering.