

# The Potential of New Technologies

Peter M. Vallone, PhD

Leader, Applied Genetics Group, NIST



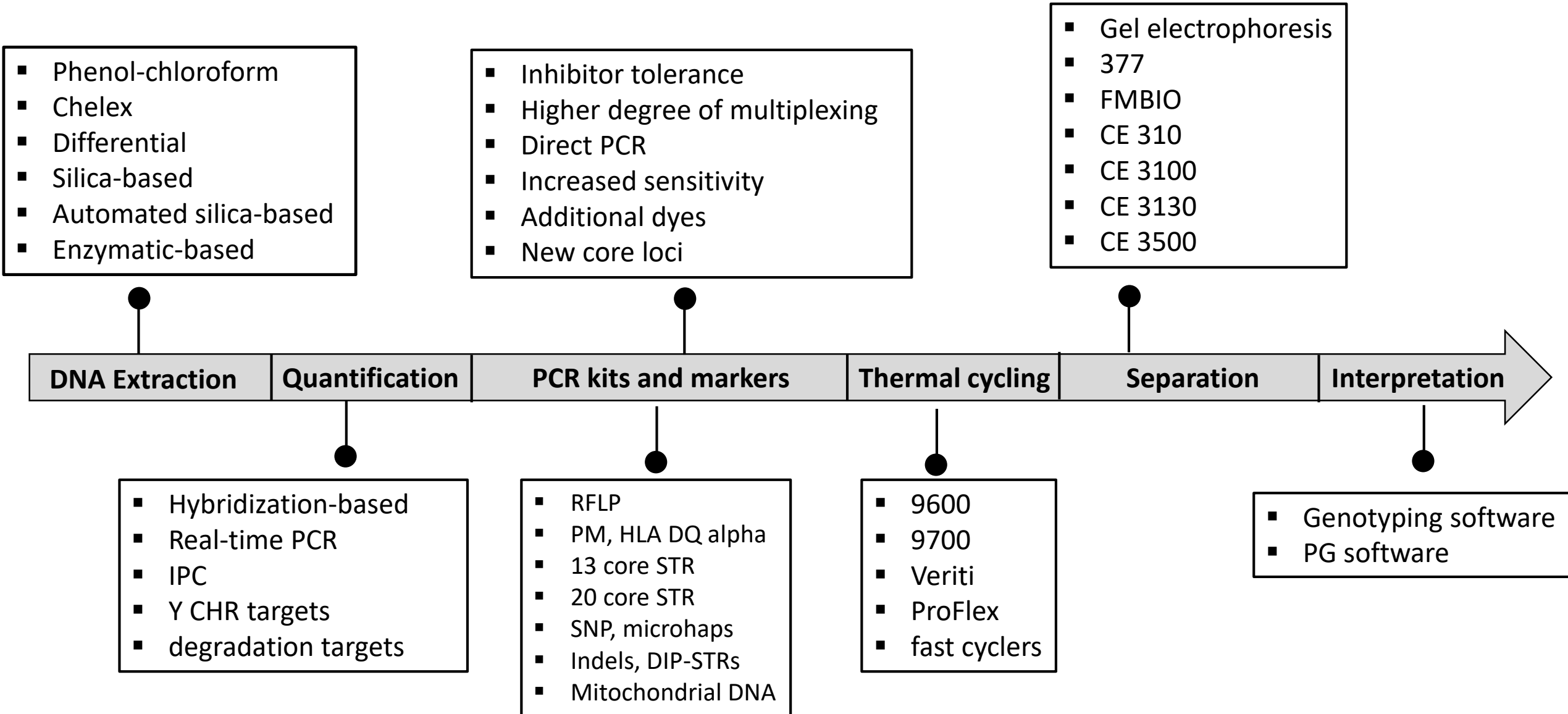
# Disclaimer

- *Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the National Institute of Standards and Technology*
- *Certain commercial equipment, 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.*

# Outline

- Adoption and implementation of new technologies
  - Advances in forensic DNA typing
  - Idealized process
- The mixture problem
  - General illustration
- Sequencing
  - STR and microhaplotype examples
- Cell separation techniques

# Advances in Forensic DNA Typing



# Advances in Forensic DNA Typing

New things continue to come our way...

Forensics

- Rapid DNA
- **NGS/MPS – Sequencing**
- Investigative genealogy
- **Cell separation techniques**

- Phenol-chloroform
- Chelex
- Differential
- Silica-based
- Automated silica-based
- Enzymatic-based



- Hybridization-based
- Real-time PCR
- IPC
- Y CHR targets
- degradation targets

- RFLP
- PM, HLA DQ alpha
- 13 core STR
- 20 core STR
- SNP, microhaps
- Indels, DIP-STRs
- Mitochondrial DNA

- 9600
- 9700
- Veriti
- ProFlex
- fast cyclers

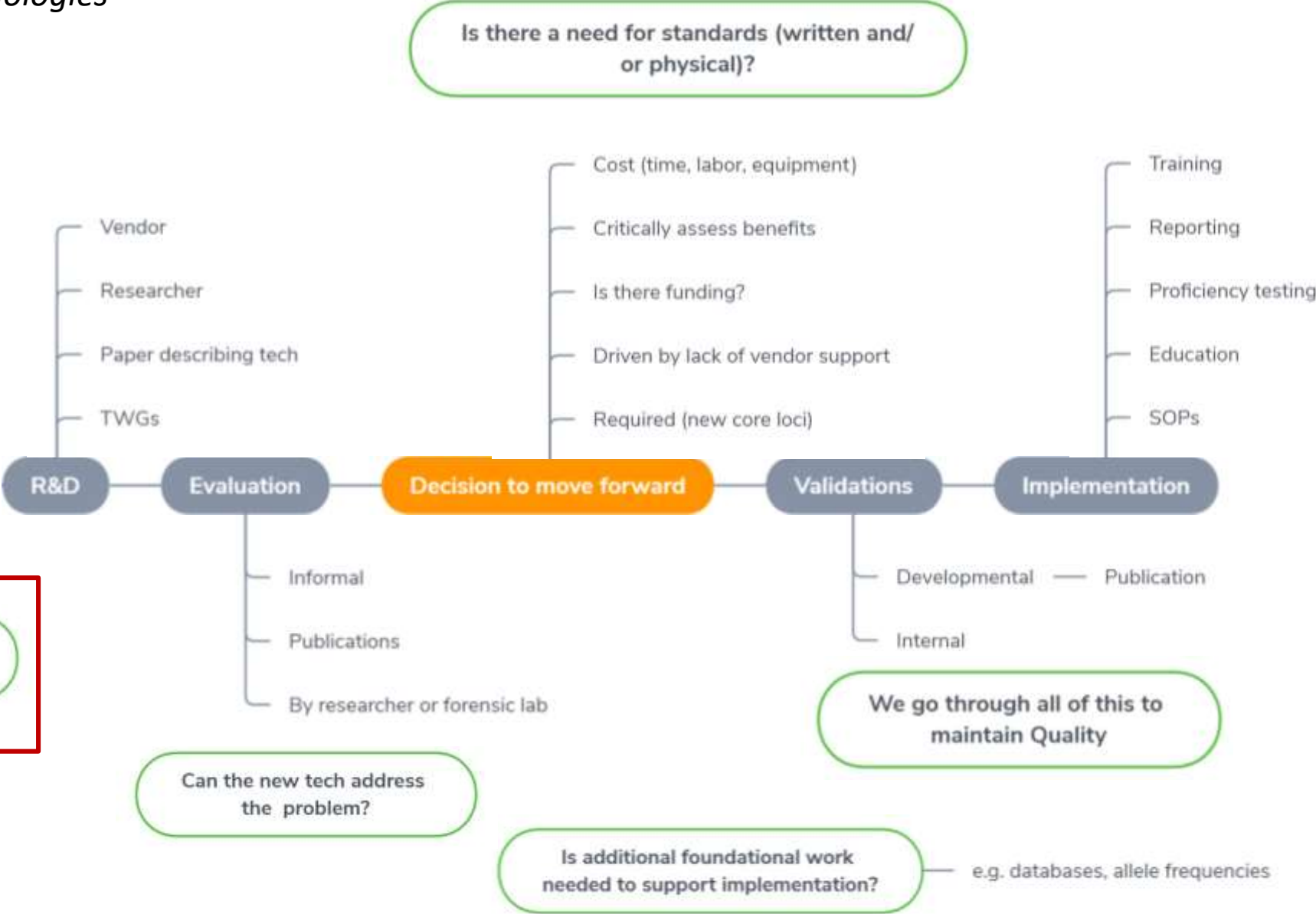
- Genotyping software
- PG software

# Analysis of a DNA Mixture

- DNA from two or more contributors are deposited
  - DNA may be in the cell or cell free
- Post DNA **extraction**, the alleles from all contributors are mixed together
  - DNA may be lost/reduced in the purification process
- **PCR** amplifies the alleles present post DNA extraction
  - Stochastic effects, degradation, inhibitors
- Currently PCR amplicons are **separated and detected** by CE methods
- **Interpretation** of the data (community is moving toward probabilistic genotyping)

*Challenges as we address more difficult cases: touch DNA, lower template amounts, more contributors...*

*Idealized process for the implementation of new technologies*







# Sequencing

Next-generation sequencing (NGS)

Massively parallel sequencing (MPS)

# Current NGS/MPS platforms and assays allow for the typing of forensically-relevant STR and SNP marker systems



*Verogen FGx platform  
ForenSeq panel (STRs, SNPs, mito)*



*Thermo Fisher S5 platform  
Precision ID panels (STRs, SNPs, mito)*



*PowerSeq™ 46GY System  
PowerSeq™ CRM Nested System*



*Qiagen GeneReader platform  
Large SNP panels and mito*

# Benefits of sequencing (general)

- Can provide further resolution of STR alleles

- Increased polymorphic content
- Length (CE) -> Sequence (NGS)

19 allele -> [GGAA]11 [GGCA]8

More markers  
Multiplexing of samples per run  
More information per run

- Technology can be applied to type additional markers systems

- Additional non-CODIS STR markers
- Insertion=deletion markers
- Mitochondrial (control region, whole mitochondrial genome)
- SNPs
  - Ancestry, Phenotype, ID, **Microhaplotypes**

# Sequencing *STRs for Mixtures*

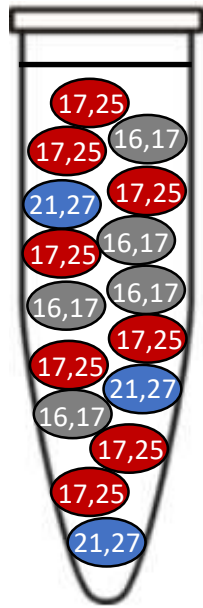
In Comparison to CE methods	Comment
<ul style="list-style-type: none"><li>• Additional alleles</li></ul>	<ul style="list-style-type: none"><li>• “Unmasking” of alleles identical by length</li><li>• Improve number of contributor estimates</li><li>• Sequence-based allele frequencies are applied</li><li>• Length-based alleles are back compatible to current databases</li></ul>
<ul style="list-style-type: none"><li>• Currently using targeted PCR</li></ul>	<ul style="list-style-type: none"><li>• Comparable sensitivity to CE</li><li>• May have an increased input range (&gt; 2 ng)</li><li>• Subject to stochastic effects</li></ul>
<ul style="list-style-type: none"><li>• Stutter products are sequenced</li></ul>	<ul style="list-style-type: none"><li>• <i>Potential</i> to <u>correlate</u> stutter product(s) to parent allele</li><li>• Allow for a more accurate modeling of stutter products</li></ul>
<ul style="list-style-type: none"><li>• Signal thresholds</li></ul>	<ul style="list-style-type: none"><li>• Discern noise (from instrument, PCR, seq, library error) from an allele; determine an AT</li></ul>
<ul style="list-style-type: none"><li>• Artifacts</li></ul>	<ul style="list-style-type: none"><li>• No dye artifacts; other concerns?</li></ul>
<ul style="list-style-type: none"><li>• Shorter PCR amplicons</li></ul>	<ul style="list-style-type: none"><li>• Improved performance with degraded samples</li></ul>
<ul style="list-style-type: none"><li>• Larger multiplexes</li></ul>	<ul style="list-style-type: none"><li>• More loci can be analyzed (autosomal, Y, X, mito)</li></ul>
<ul style="list-style-type: none"><li>• Interpretation</li></ul>	<ul style="list-style-type: none"><li>• A NGS-based probabilistic genotyping model for STRs?</li></ul>



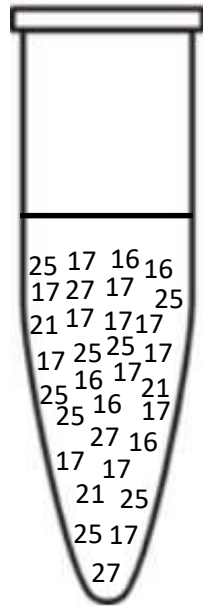
Collection of evidence



Evidence Processing



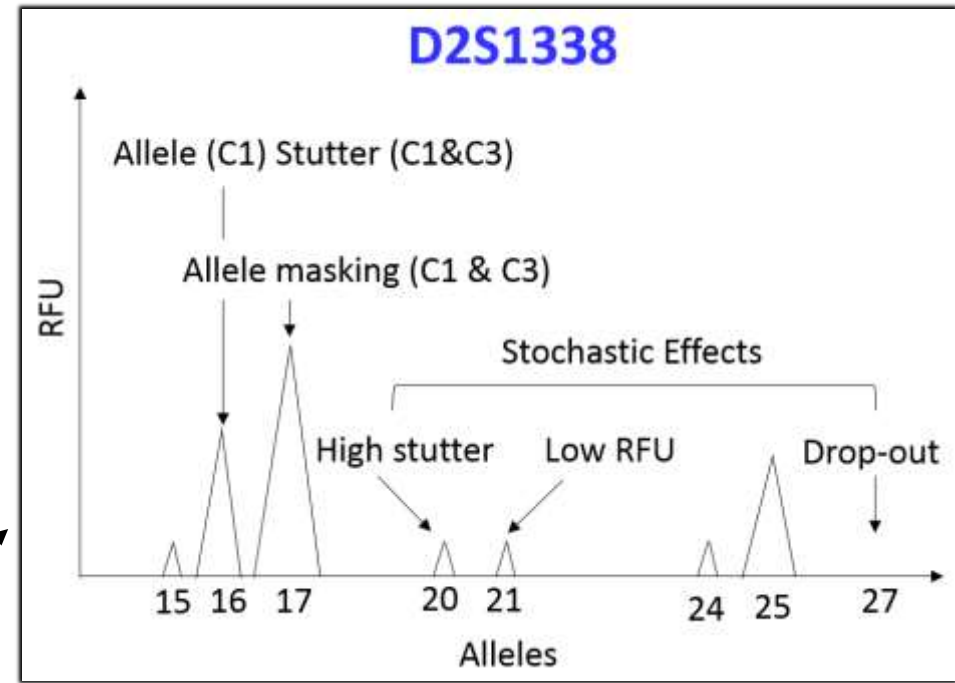
Extraction



Post extraction

PCR

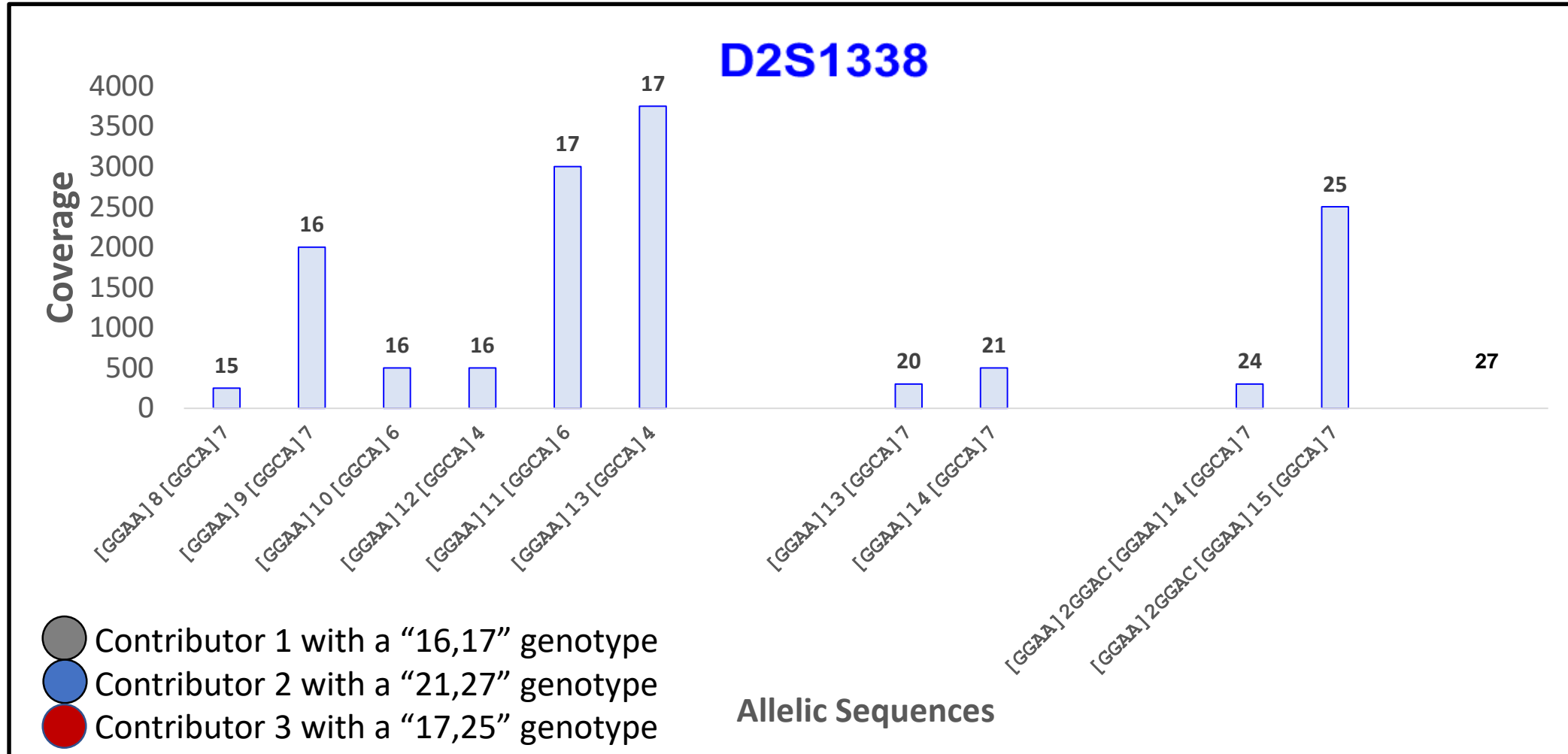
CE



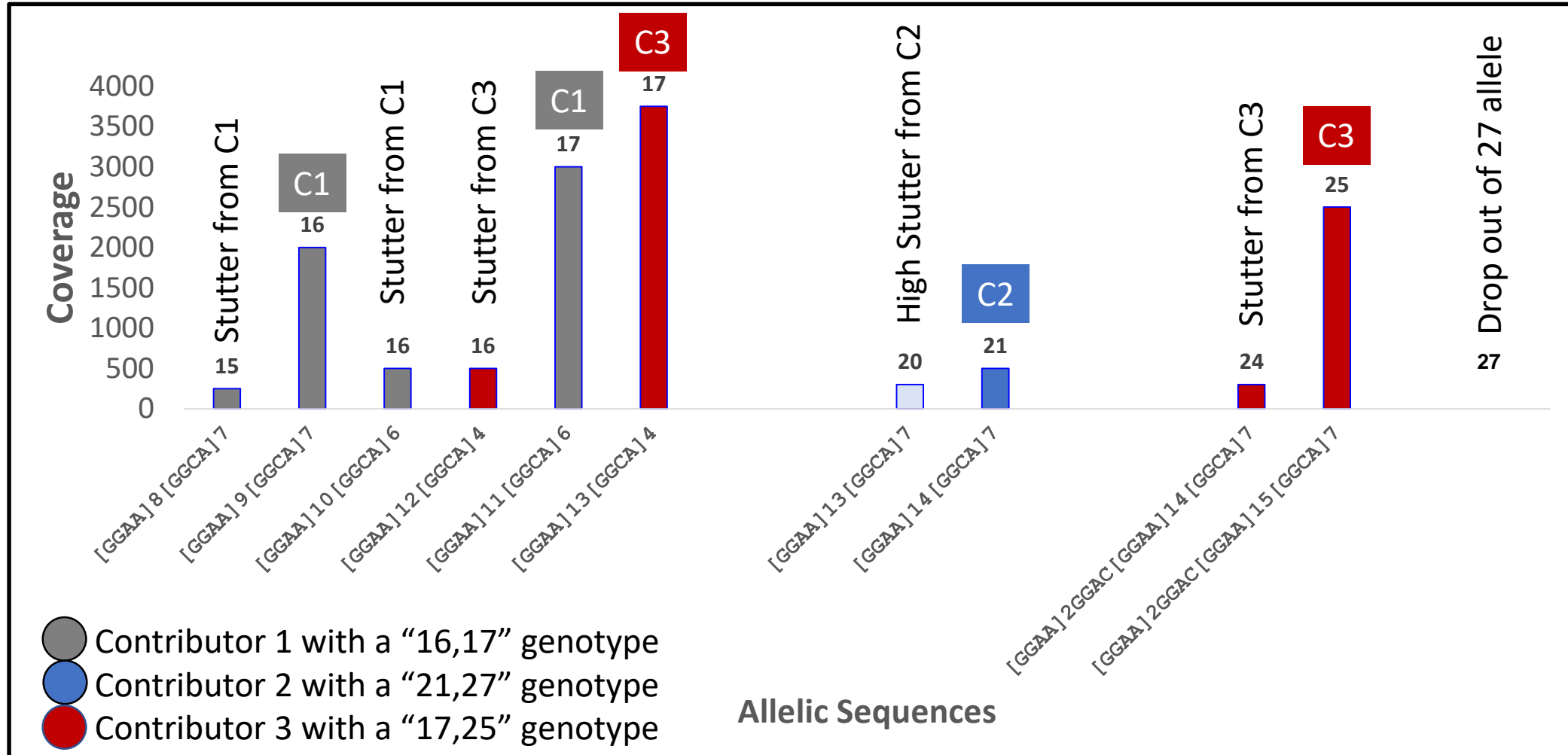
NGS data?

- Contributor 1 with a "16,17" genotype
- Contributor 2 with a "21,27" genotype
- Contributor 3 with a "17,25" genotype

# Looking at the data in 'Sequencing space'

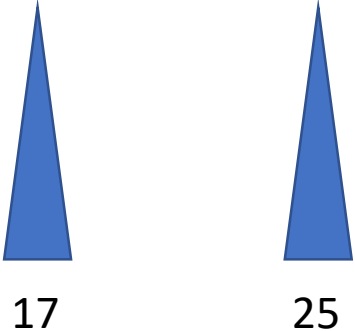


# Looking at the data in 'Sequencing space'



# Locus D2S1338 - Length-based allele frequencies

CE length-based information



Length	Freq	
15	0.0010	0.1%
16	0.0401	4.0%
<b>17</b>	<b>0.1419</b>	<b>14.2%</b>
18	0.0705	7.0%
19	0.1486	14.9%
20	0.1327	13.3%
21	0.0666	6.7%
22	0.0752	7.5%
23	0.1182	11.8%
24	0.0970	9.7%
<b>25</b>	<b>0.0835</b>	<b>8.3%</b>
26	0.0227	2.3%
27	0.0019	0.2%

13 alleles observed

Heterozygous genotype (17, 25)

$$2pq = 2(0.1419)(0.0835) = 0.0237$$

...1 in 42

Combined set of allele frequencies taken from Gettings et al. Sequence-based U.S. population data for 27 autosomal STR loci. Forensic Sci Int Genet. 2018 37:106-115



# Locus D2S1338 – Sequence-based allele frequencies

Sequence-based information

Heterozygous genotype

17 [GGAA]13 [GGCA]4  
25 [GGAA]2 GGAC [GGAA]15 [GGCA]7

$$2pq = 2(0.0029)(0.0734) = 0.0004$$

...1 in 2,349

15[GGAA]10 [GGCA]5	0.0010
16[GGAA]10 [GGCA]6	0.0217
16[GGAA]12 [GGCA]4	0.0145
16[GGAA]9 [GGCA]7	0.0019
16[GGAA]11 [GGCA]5	0.0019
17[GGAA]11 [GGCA]6	0.1366
<b>17[GGAA]13 [GGCA]4</b>	<b>0.0029</b>
17[GGAA]10 [GGCA]7	0.0014
17[GGAA]12 [GGCA]5	0.0010
18[GGAA]12 [GGCA]6	
18[GGAA]11 [GGCA]7	= 14.2%
18[GGAA]14 [GGCA]4	0.0024
18[GGAA]13 [GGCA]5	0.0019
18[GGAA]8 GAAA [GGAA]2 [GGCA]7	0.0010
18[GGAA]15 [GGCA]3	0.0005
19[GGAA]12 [GGCA]7	0.1076
19[GGAA]13 [GGCA]6	0.0333
19[GGAA]11 [GGCA]8	0.0024
19[GGAA]14 [GGCA]5	0.0024
19[GGAA]2 GGAC [GGAA]10 [GGCA]6	0.0014
19[GGAA]9 GAAA [GGAA]2 [GGCA]7	0.0005
19[GGAA]11 GGGA [GGCA]7	0.0005
19[GGAA]16 [GGCA]3	0.0005
20[GGAA]13 [GGCA]7	0.0893
20[GGAA]2 GGAC [GGAA]10 [GGCA]7	0.0121
20[GGAA]14 [GGCA]6	0.0092
20[GGAA]10 GAAA [GGAA]2 [GGCA]7	0.0087
20[GGAA]12 GGGA [GGCA]7	0.0068
20[GGAA]12 [GGCA]8	0.0043
20[GGAA]16 [GGCA]4	0.0010
20[GGAA]2 GGAC [GGAA]11 [GGCA]6	0.0005
20[GGAA]2 GGAC [GGAA]9 AGAA [GGCA]7	0.0005
20[GGAA]15 [GGCA]5	0.0005

21[GGAA]14 [GGCA]7	0.0256
21[GGAA]2 GGAC [GGAA]11 [GGCA]7	0.0232
21[GGAA]13 [GGCA]8	0.0082
21[GGAA]2 GGAC [GGAA]12 [GGCA]6	0.0068
21[GGAA]15 [GGCA]6	0.0014
21[GGAA]12 [GGCA]9	0.0005
21[GGAA]16 [GGCA]5	0.0005
21[GGAA]17 [GGCA]4	0.0005
22[GGAA]2 GGAC [GGAA]12 [GGCA]7	0.0410
22[GGAA]2 GGAC [GGAA]13 [GGCA]6	0.0145
22[GGAA]15 [GGCA]7	0.0101
22[GGAA]14 [GGCA]8	0.0043
22[GGAA]13 [GGCA]9	0.0039
22[GGAA]16 [GGCA]6	0.0010
22[GGAA]2 GGAC [GGAA]14 [GGCA]5	0.0005
23[GGAA]2 GGAC [GGAA]13 [GGCA]7	0.0960
23[GGAA]2 GGAC [GGAA]14 [GGCA]6	0.0130
23[GGAA]16 [GGCA]7	0.0029
23[GGAA]14 [GGCA]9	0.0024
23[GGAA]2 GGAC [GGAA]12 [GGCA]8	0.0019
23[GGAA]15 [GGCA]8	0.0019
24[GGAA]2 GGAC [GGAA]14 [GGCA]7	0.0835
24[GGAA]2 GGAC [GGAA]15 [GGCA]6	0.0106
24[GGAA]2 GGAC [GGAA]13 [GGCA]8	0.0024
24[GGAA]15 [GGCA]9	0.0005
<b>25[GGAA]2 GGAC [GGAA]15 [GGCA]7</b>	<b>0.0734</b>
25[GGAA]2 GGAC [GGAA]14 [GGCA]8	0.0072
25[GGAA]2 GGAC [GGAA]16 [GGCA]6	0.0029
26[GGAA]2 GGAC [GGAA]16 [GGCA]7	
26[GGAA]2 GGAC [GGAA]15 [GGCA]7	= 8.3%
26[GGAA]2 GGAC [GGAA]17 [GGCA]6	0.0014
26[GGAA]2 GGAC [GGAA]18 [GGCA]5	0.0005
27[GGAA]2 GGAC [GGAA]17 [GGCA]7	0.0014
27[GGAA]2 GGAC [GGAA]16 [GGCA]8	0.0005

67 alleles observed

# Microhaplotypes

Kidd and Speed Investigative Genetics (2015) 6:1  
DOI 10.1186/s13223-014-0018-3

**Investigative Genetics**

RESEARCH Open Access

## Criteria for selecting microhaplotypes: mixture detection and deconvolution

Kenneth K. Kidd<sup>a</sup> and William C. Speed

Forensic Science International: Genetics 12 (2014) 215–224

Contents lists available at ScienceDirect

**Forensic Science International: Genetics**

journal homepage: [www.elsevier.com/locate/fsig](http://www.elsevier.com/locate/fsig)

### Microhaplotypes in forensic genetics

Fabio Oldoni<sup>a</sup>, Kenneth K. Kidd<sup>a</sup>, Daniele Podini<sup>b</sup>

<sup>a</sup> Department of Forensic Science, The George Washington University, 2300 Mitchell Road NW, Washington, DC, 20007, United States  
<sup>b</sup> Yale University School of Medicine, Department of Genetics, 333 Cedar Street, New Haven, CT, 06510, United States

**ARTICLE INFO**

**Keywords:**  
Microhaplotypes  
Single nucleotide polymorphisms  
Massively parallel sequencing  
Geographic ancestry inference  
Human deconvolution  
Human identification  
Missing persons and relationship identification  
Probabilistic genotyping  
Clinical applications  
Non-human DNA

**ABSTRACT**

Microhaplotype loci (microhap, ML) are a novel type of molecular marker of less than 300 nucleotides, defined by two or more closely linked SNPs associated in multiple allelic combinations. The value of these markers is enhanced by massively parallel sequencing (MPS), which allows the sequencing of both parental haplotypes at each of the many multiplexed loci. This review describes the features of these multi SNP markers and discusses their value in forensic genetics, focusing on individualization, biogeographic ancestry inference, and mixture deconvolution. Forensic applications also include missing person identification, relationship testing, and medical diagnostic applications. The technique is not restricted to humans.

**1. Introduction: historical background of haplotype discovery**

**1.1. Discovery of haplotype blocks in the human genome**

The term 'haplotype' was first introduced by Ruggiero Cappellari in the late 180s to describe alleles within the human leukocyte antigen (HLA) region that are inherited together as a block [1]. Twenty years later the human genome project (HGP) launched an unprecedented international collaboration [2] foundational to the study of human genetics and biomedical research. The early work focused on mapping of human and mouse genes and sequencing the genomes of significantly smaller and easily studied organisms [3–6]. This was of paramount importance for the understanding of the hereditary architecture of disease [7–9] and provides an essential scaffold for the assembly and annotation of the human genome. The publication that precluded the completion of the sequencing of the first draft of the human genome in

**1.2. Large scale international haplotype map project**

With the goal of establishing a resource for the study of global variation in the human genome, the HGP-CEPH (Human Genome Diversity Panel - Centre d'Etude du Polymorphisme Humain) was established in 2002. At its core the HGP-CEPH [17,18] is a collection of 1064 cultured lymphoblastoid cell lines from 52 populations of different parts of the world deposited at the Fondation Jean Dausset in Paris.

Later studies, aimed at understanding human evolution and population structure [19,20] focused on linkage disequilibrium (LD). Comprehensive large-scale LD studies identified distinct LD blocks, defined as haplotype blocks or specific genomic regions with a restricted number of haplotypes occurring due to the limited number of ancestral recombination events within these regions [21,22]. With these considerations in mind, a large-scale Hap Map international project (i.e.,

Forensic Science International: Genetics 12 (2014) 215–224

Contents lists available at ScienceDirect

**Forensic Science International: Genetics**

journal homepage: [www.elsevier.com/locate/fsig](http://www.elsevier.com/locate/fsig)

## Current sequencing technology makes microhaplotypes a powerful new type of genetic marker for forensics

Kenneth K. Kidd<sup>a,\*</sup>, Andrew J. Pakstis<sup>a</sup>, William C. Speed<sup>a</sup>, Robert Lagacé<sup>b</sup>, Joseph Chang<sup>b</sup>, Sharon Wootton<sup>b</sup>, Eva Haigh<sup>a</sup>, Judith R. Kidd<sup>a</sup>

<sup>a</sup> Department of Genetics, Yale University School of Medicine, New Haven, CT 06520-8005, USA  
<sup>b</sup> Human Identification Group, Thermo Fisher Scientific, 180 Oyster Point Blvd., South San Francisco, CA 94080

**Abstract**

[3–5]. By the familial to identify or more individual loci of forensic use. The use of precisely defined sets oforphisms as an informative tool for forensic genetics is

International Journal of Legal Medicine  
<https://doi.org/10.1007/s00414-019-02010-7>

ORIGINAL ARTICLE

## Mixture deconvolution by massively parallel sequencing of microhaplotypes

Lindsay Bennett<sup>1</sup> · Fabio Oldoni<sup>2</sup> · Kelly Long<sup>2</sup> · Selena Cisana<sup>2</sup> · Katrina Madella<sup>2</sup> · Sharon Wootton<sup>3</sup>  
Joseph Chang<sup>3</sup> · Ryo Hasegawa<sup>3</sup> · Robert Lagacé<sup>3</sup> · Kenneth K. Kidd<sup>4</sup> · Daniele Podini<sup>2</sup> 

# Microhaplotypes

- Novel type of marker of < 300 base pairs
- Defined by two or more closely linked SNPs associated in multiple allelic combinations
  - Similar size amplicons can be used in a multiplex (not length-based alleles)
  - Each allele from a locus will be the same size
- Can also be used for ancestry prediction
- (Typically) fewer alleles than a STR locus

**\*Absence of stutter as the alleles are SNP-defined versus a repeating motif\***

- Still enriched by targeted PCR
- Allele frequencies are published (and still being generated)
- No core law enforcement database
- Will require a framework for interpretation (probabilistic genotyping?)

# Chromosome 1



mh01KK-001

## Genotypes

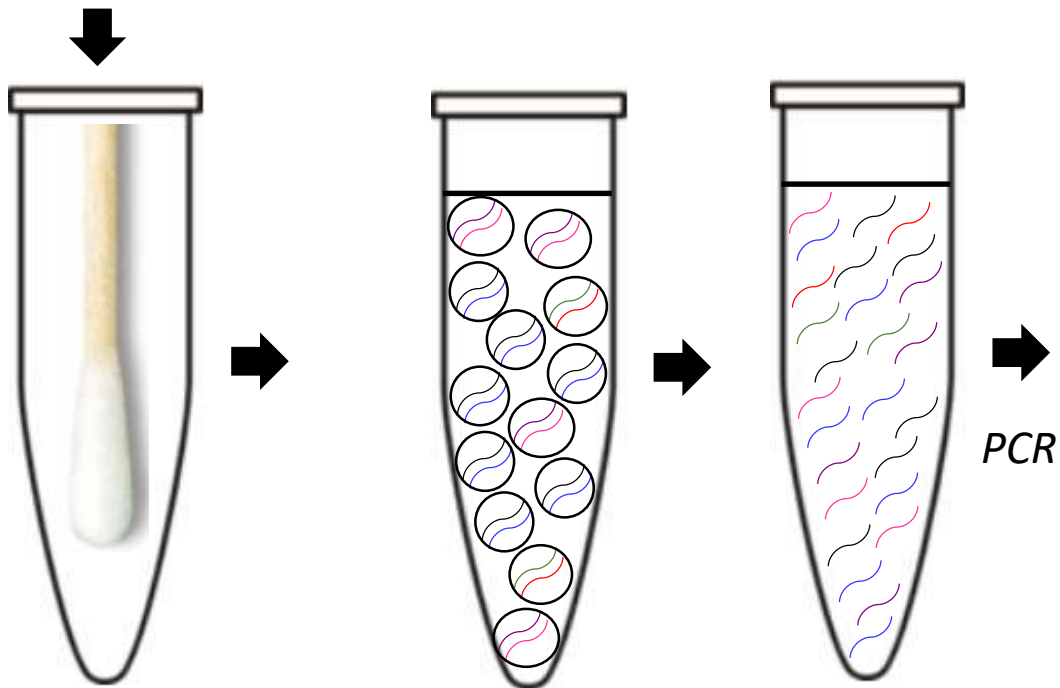
	SNP1	SNP2	SNP3	SNP4	MHs
Individual 1	C	G	G	C	CGGC, CGGC
Individual 2	C	G	G	T	CGGT, CAGC
Individual 3	T	A	G	C	TAGC, TAGT

Approximately 200-300 base pair amplicon



# Microhaplotypes

Collection of evidence

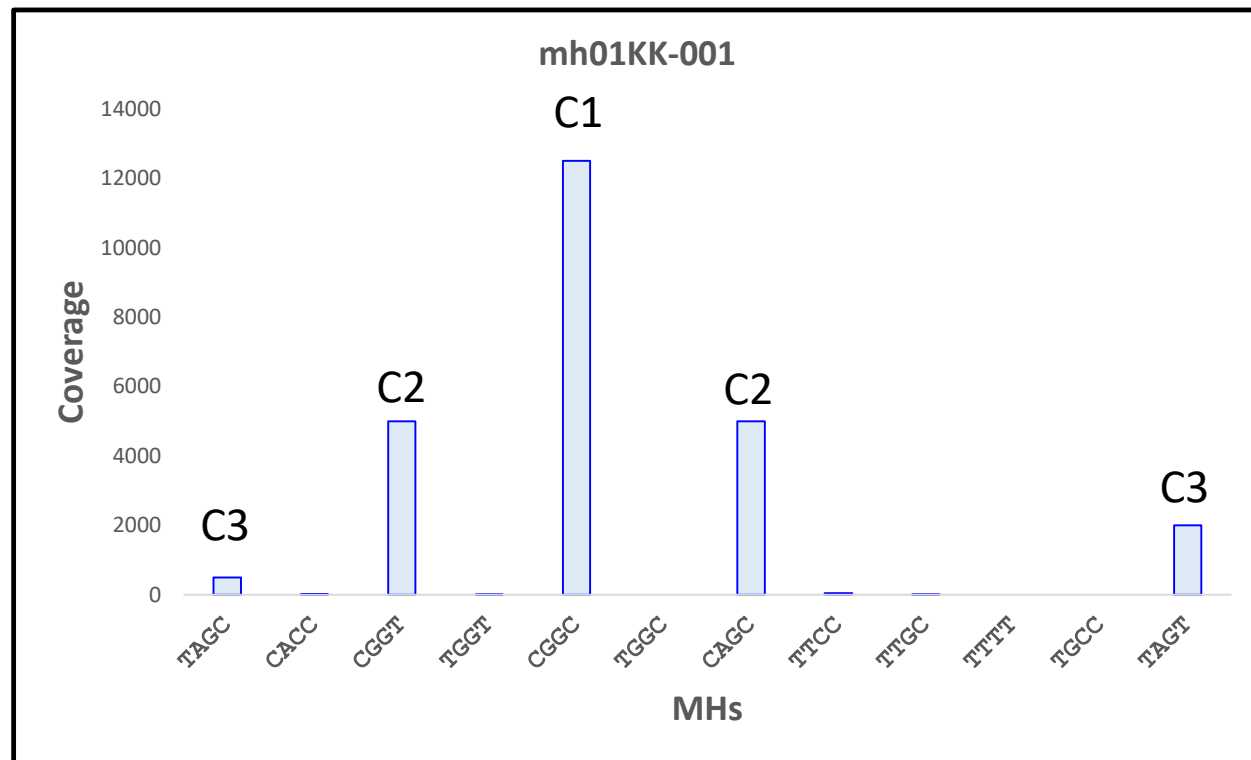





Evidence Processing

Extraction

Post extraction

PCR



-  Contributor 1 with a "CGGC, CGGC" microhaplotype
-  Contributor 2 with a "CGGT, CAGC" microhaplotype
-  Contributor 3 with a "TAGC, TAGT" microhaplotype

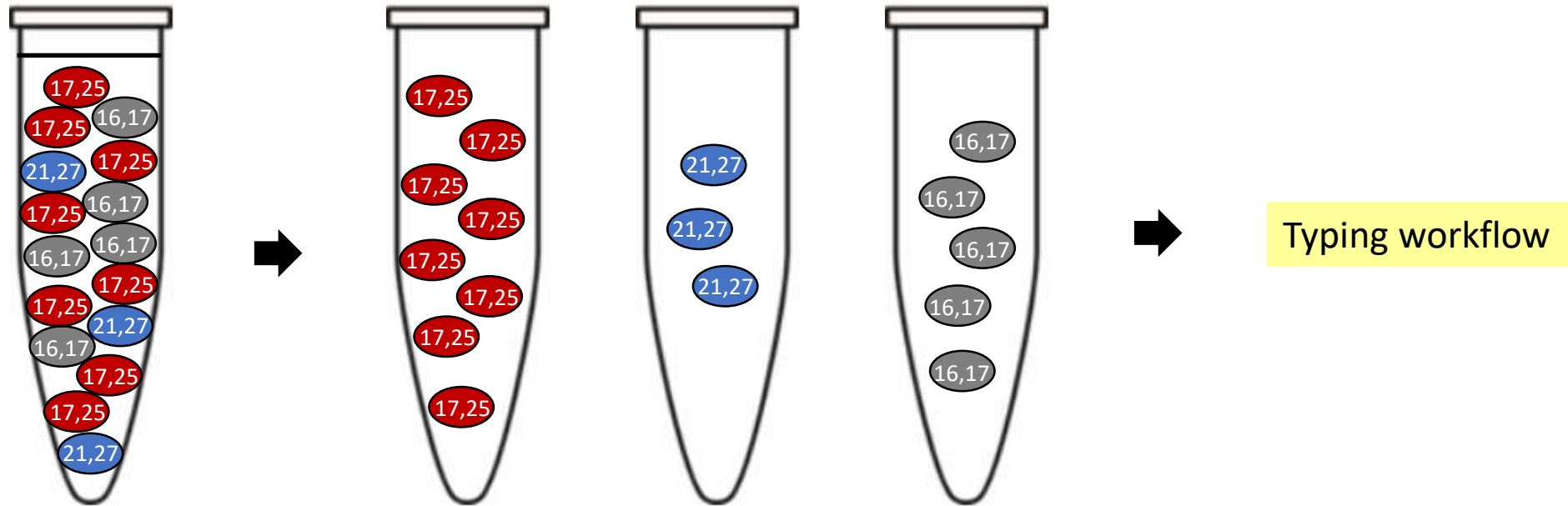
Will still observe some 'noise'  
 Hb imbalance at lower levels of DNA input (C3)  
**No stutter artifacts**

# Sequencing *Microhaplotypes for Mixtures*

In Comparison to CE/STR methods	Comment
<ul style="list-style-type: none"> <li><del>Additional Microhaplotype-based alleles</del></li> </ul>	<ul style="list-style-type: none"> <li><del>“Unmasking” of alleles identical by length</del></li> <li>Improve number of contributor estimates</li> <li><b>Microhaplotype</b>-based allele frequencies are applied</li> <li><del>Length-based alleles are back compatible to current databases</del></li> </ul>
<ul style="list-style-type: none"> <li>Currently using targeted PCR</li> </ul>	<ul style="list-style-type: none"> <li>Comparable sensitivity to CE</li> <li>May have an increased input range (&gt; 2 ng)</li> <li>Subject to stochastic effects</li> </ul>
<ul style="list-style-type: none"> <li><del>Stutter products are sequenced</del></li> <li><b>No stutter artifacts</b></li> </ul>	<ul style="list-style-type: none"> <li><del>Potential to correlate stutter product(s) to parent allele</del></li> <li><del>Allow for a more accurate modeling of stutter products.</del></li> </ul>
<ul style="list-style-type: none"> <li>Signal thresholds</li> </ul>	<ul style="list-style-type: none"> <li>Discern noise (from instrument, PCR, seq, library error) from an allele; determine an AT</li> </ul>
<ul style="list-style-type: none"> <li>Artifacts</li> </ul>	<ul style="list-style-type: none"> <li>No dye artifacts; other concerns?</li> </ul>
<ul style="list-style-type: none"> <li>Shorter PCR amplicons (compared to CE)</li> </ul>	<ul style="list-style-type: none"> <li>Improved performance with degraded samples</li> </ul>
<ul style="list-style-type: none"> <li>Larger multiplexes</li> </ul>	<ul style="list-style-type: none"> <li>A need for microhaplotype loci</li> </ul>
<ul style="list-style-type: none"> <li>Interpretation</li> </ul>	<ul style="list-style-type: none"> <li>A NGS-based probabilistic genotyping model for MHs?</li> </ul>

# Physical Separation of Cells

General concept – physical separation/sorting of cells before DNA typing workflow





# Physical Separation of Cells – How?!

- Some proposed methods
  - Partition into microreactors
  - Micro-manipulation (needle, laser)
  - Sort based on cell morphology or tagging

## Challenges

- Dried cells are more challenging than fresh solutions
- Single cell sampling methods are lower throughput, may require PCR optimization
- Consider the specificity/sensitivity of reagents that bind cells (antibodies)
- Is there DNA in the cell (and what about cell free DNA)?



## 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)

Lee Dean<sup>a</sup>, Ye Jin Kwon<sup>a</sup>, M. Katherine Philpott<sup>a</sup>, Cristina E. Stanciu<sup>a</sup>, Sarah J. Seashols-Williams<sup>a</sup>, Tracey Dawson Cruz<sup>a</sup>, Jamie Sturgill<sup>b</sup>, Christopher J. Ehrhardt<sup>a,\*</sup>

<sup>a</sup>Department of Forensic Science, Virginia Commonwealth University, 1015 Floyd Ave, Richmond, VA 23284, USA

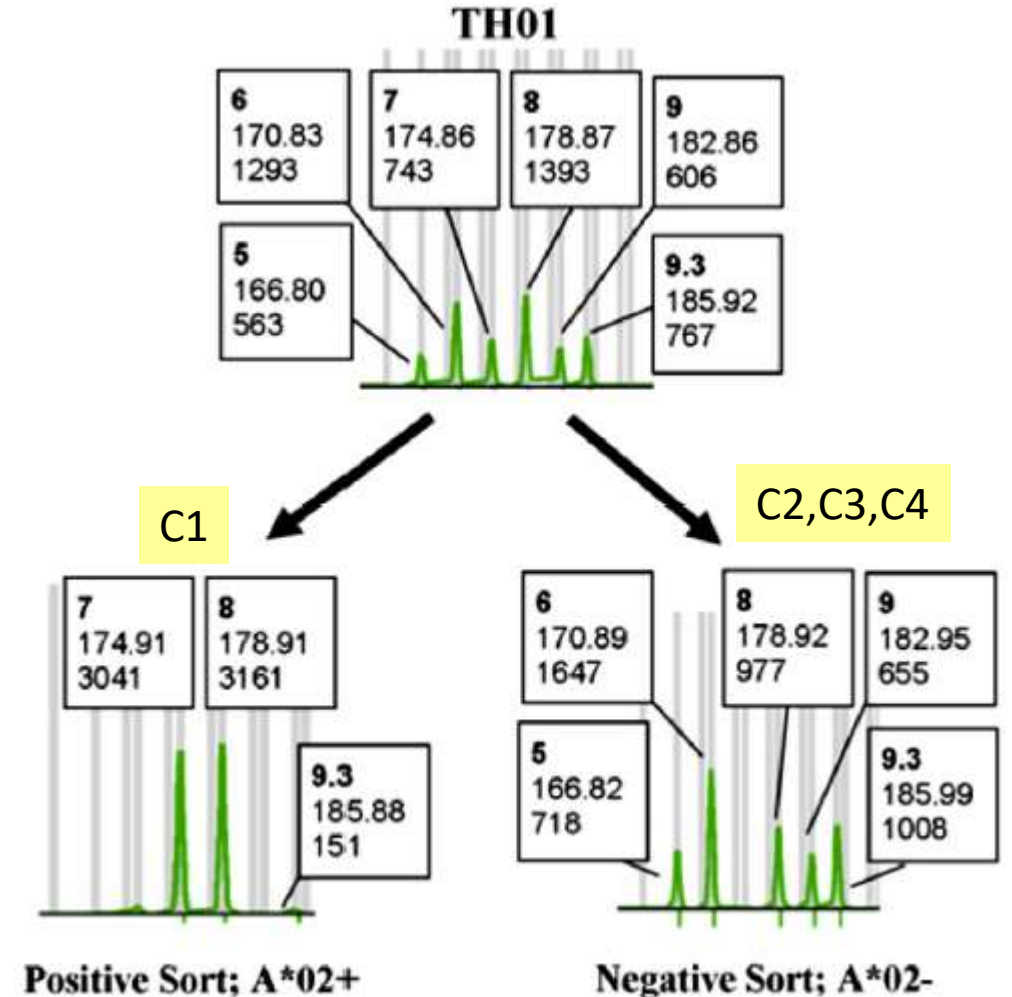
<sup>b</sup>School of Nursing, Virginia Commonwealth University, Medical College of Virginia, Richmond, VA 23284, USA

- Four person mixture (C1, C2, C3, C4)
- Binding cells with HLA allele (A\*02) antibody
  - C1 is HLA allele (A\*02) positive
- FACS separated
- Fractions typed

### Contributor genotypes

C1	7, 8
C2	6, 9
C3	5, 6
C4	8, 9.3

### Mixed DNA profile



# Single source DNA profile recovery from single cells isolated from skin and fabric from touch DNA mixtures in mock physical assaults

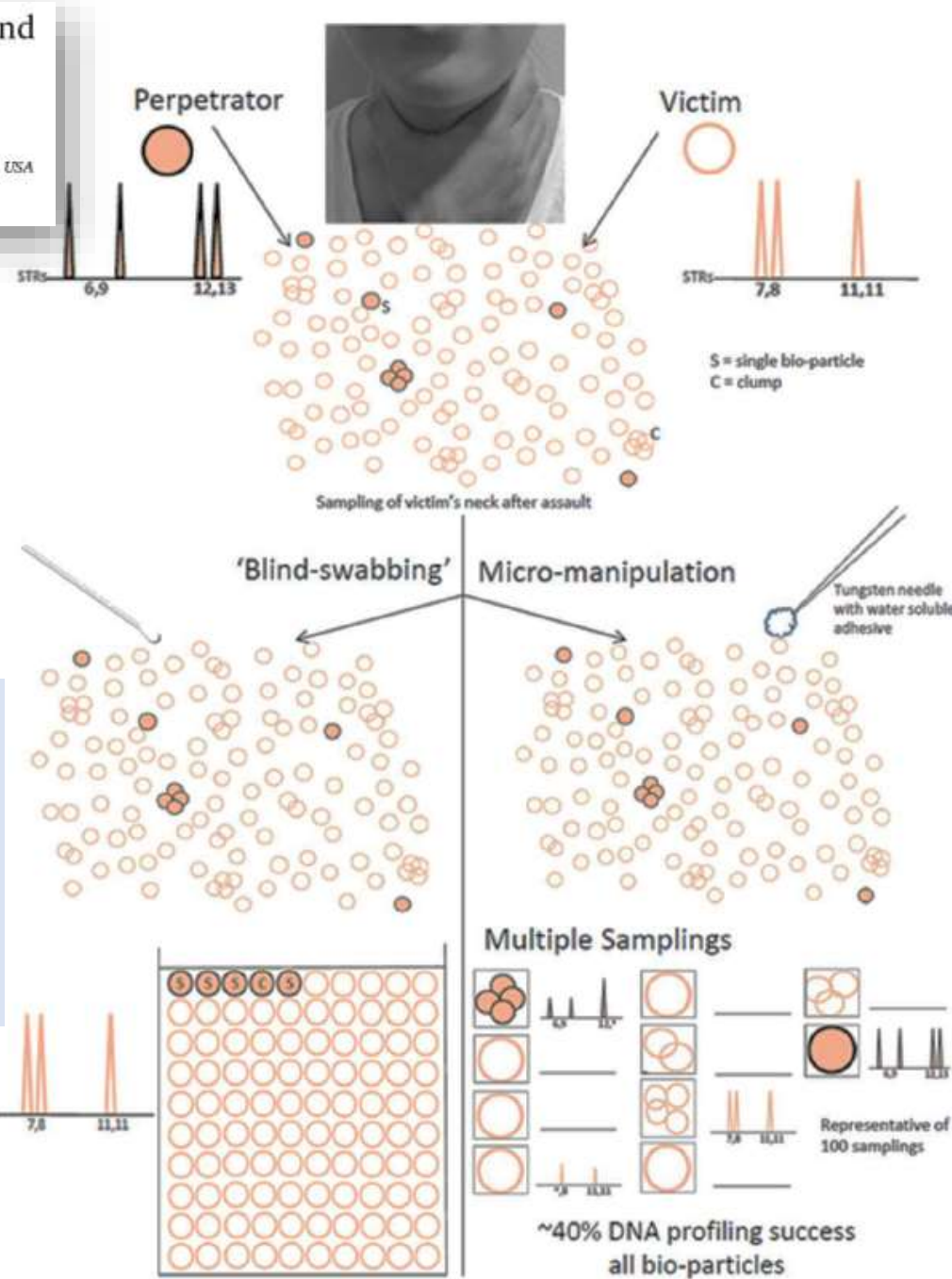
Katherine Farash<sup>a</sup>, Erin K. Hanson<sup>b</sup>, Jack Ballantyne<sup>a,b,c,\*</sup>

<sup>a</sup> Graduate Program in Forensic Science, Biochemistry Track, Department of Chemistry, University of Central Florida, PO Box 162366, Orlando, FL 32816-2366, USA

<sup>b</sup> National Center for Forensic Science, PO Box 162367, Orlando, FL 32816-2367, USA

<sup>c</sup> Department of Chemistry, University of Central Florida, PO Box 162366, Orlando, FL 32816-2366, USA

Science & Justice 58 (2018) 191–199

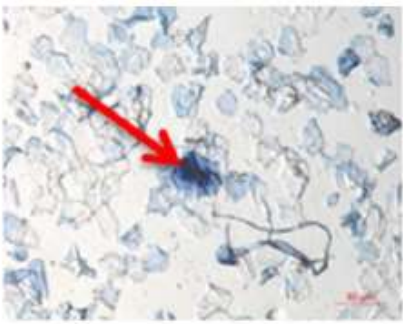


STANDARD  
ANALYSIS

“SMART”  
ENHANCED  
ANALYSIS

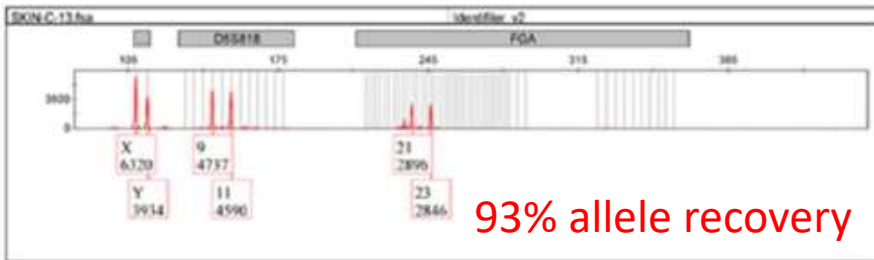
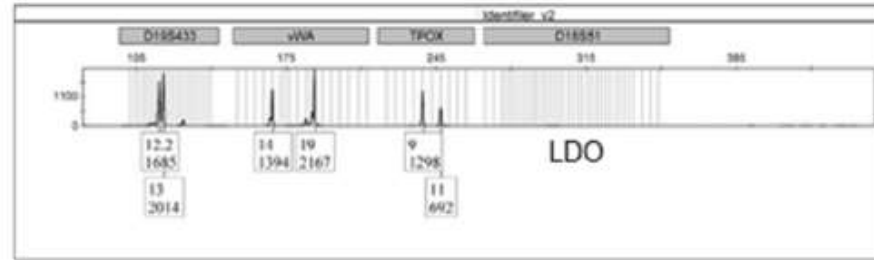
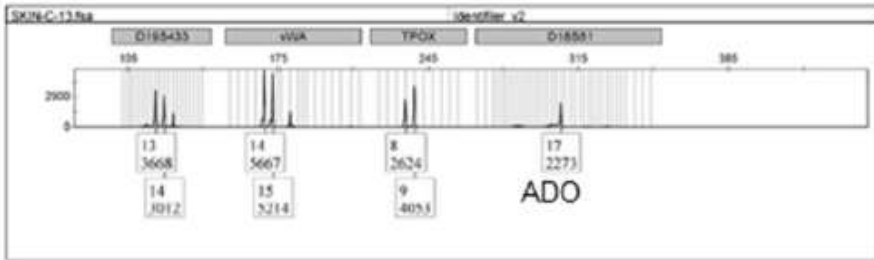
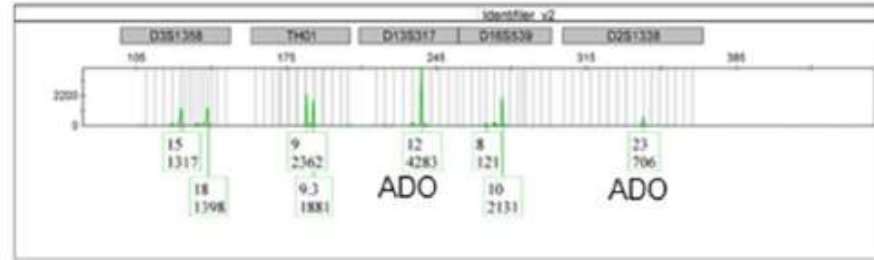
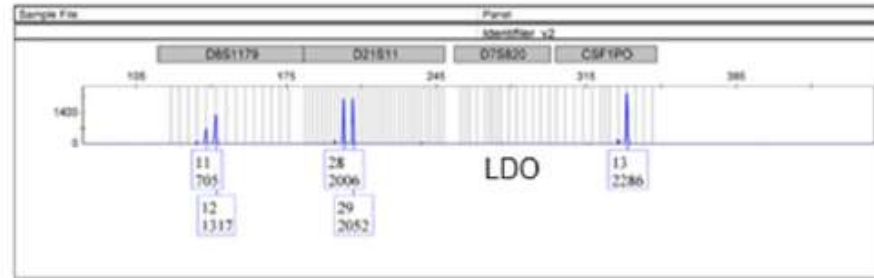
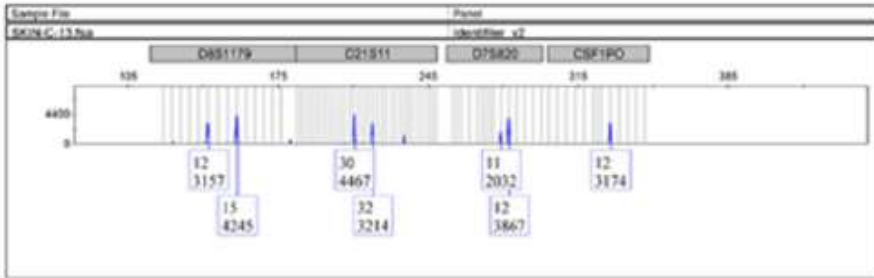
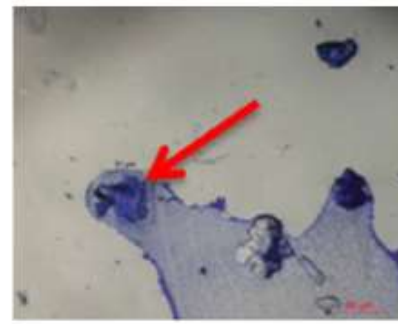
- Avoid ‘blind swabbing’ that might create mixtures
- A more directed sampling approach (“Smart”) **“sensitive, measurable, attainable, relevant, targeted”**
- Optimized direct PCR (lysis + PCR)

Fig. 1. Schema of standard versus “smart” enhanced analysis of touch DNA. Cells from the victim and perpetrator are indicated as unfilled and filled circles respectively.

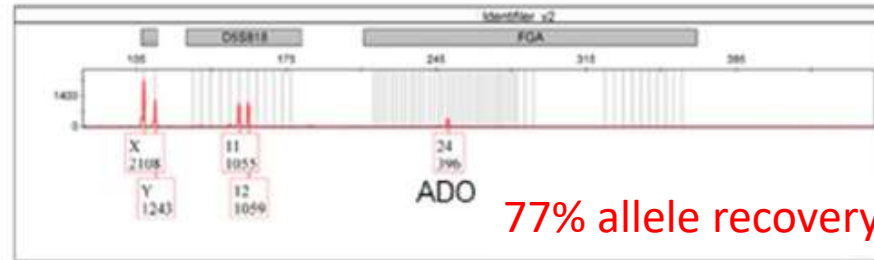


Wrist Grab  
"Clump"

Sleeve Grab  
"Single"



93% allele recovery



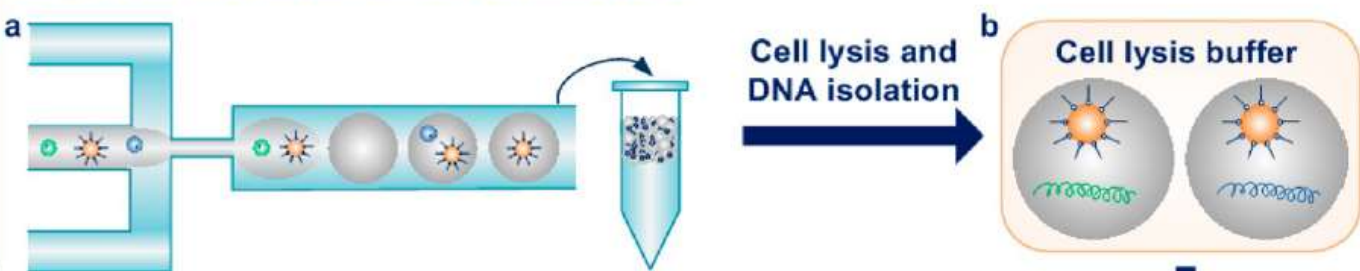
77% allele recovery



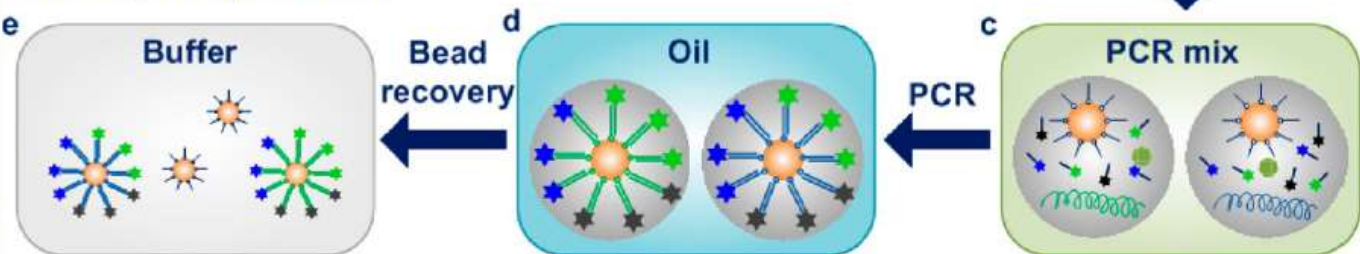
# Single-Cell Forensic Short Tandem Repeat Typing within Microfluidic Droplets

Tao Geng,<sup>†</sup> Richard Novak,<sup>‡</sup> and Richard A. Mathies<sup>\*,†,‡</sup><sup>†</sup>Department of Chemistry, University of California, Berkeley, California 94720, United States<sup>‡</sup>UCSF/UC Berkeley Graduate Program in Bioengineering, University of California, Berkeley, California 94720, United States

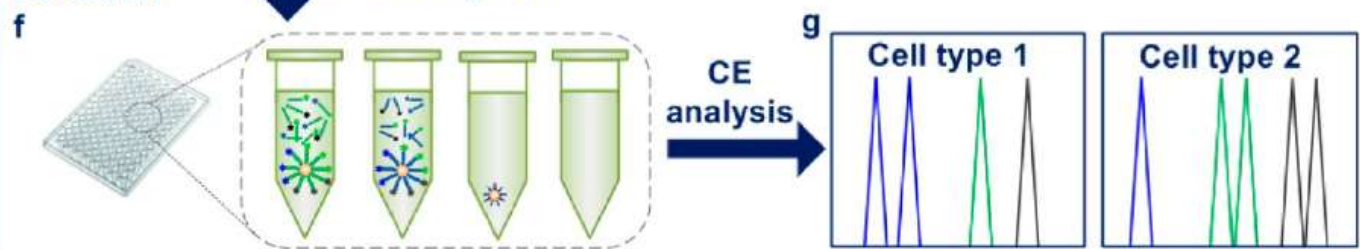
## Single-cell encapsulation and DNA isolation



## STR target amplification



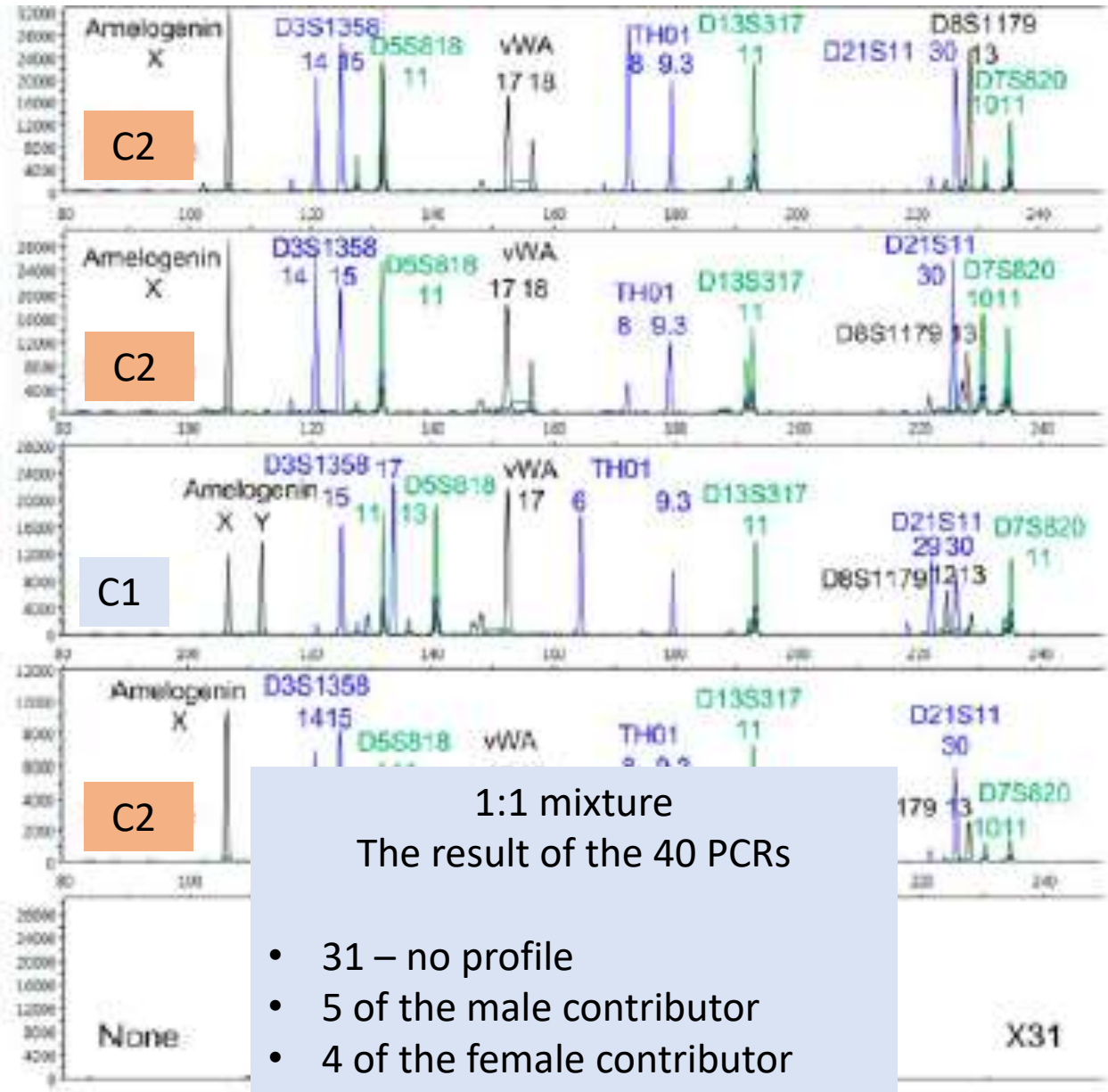
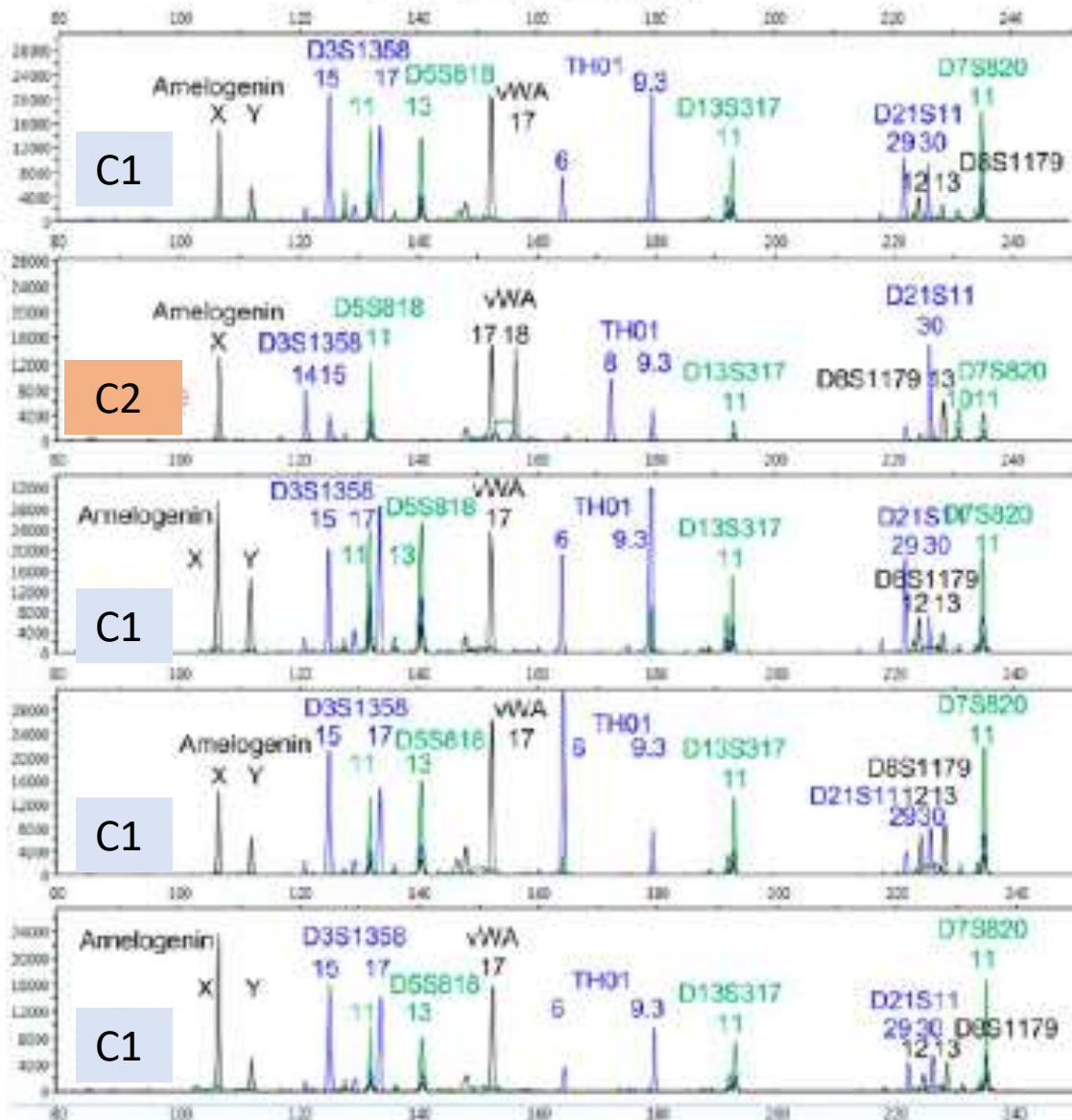
## Detection



- A cell and a primer-coated bead are encapsulated in agarose droplets
- Cell lysis and PCR take place in the droplet
- Droplets are dissolved and beads containing the PCR amplicons are diluted further for another round of PCR
- Products are separated and detected by CE
- CE runs will show: either no profile or a single source profile from the contributors

# Size (Base Pairs)

Relative fluorescence units



# Thoughts and considerations

- Define and understand the problem
- Understand the potential of the new technology to address the problem
- Consider the cost of implementation as a whole
- There is always something new on the horizon

# Thank you for your attention! Questions?



Contact: [Peter.Vallone@nist.gov](mailto:Peter.Vallone@nist.gov)

Thanks to

- **Dr. Sarah Riman (NIST)**
- Members of the NIST mixture review team (John, Hari, Rich, Melissa, and Sheila)
- Funding
  - NIST Special Programs Office: *Forensic DNA*