

# Developing investigative leads from a probabilistic genotyping database

David W. Bauer<sup>1</sup>, PhD  
Nasir Butt<sup>2</sup>, PhD  
Jeffrey Oblock<sup>2</sup>  
Mark W. Perlin<sup>1</sup>, PhD, MD, PhD

<sup>1</sup>Cybergenetics  
Pittsburgh, PA

<sup>2</sup>Cuyahoga County  
Forensic Laboratory  
Cleveland, OH



Cybergenetics

Cybergenetics © 2003-2019

---

---

---

---

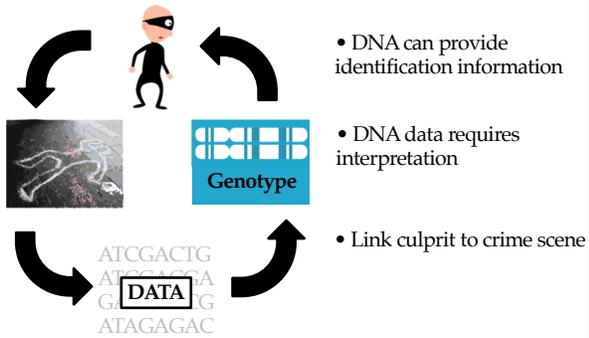
---

---

---

---

## DNA evidence chain <sup>2</sup>



---

---

---

---

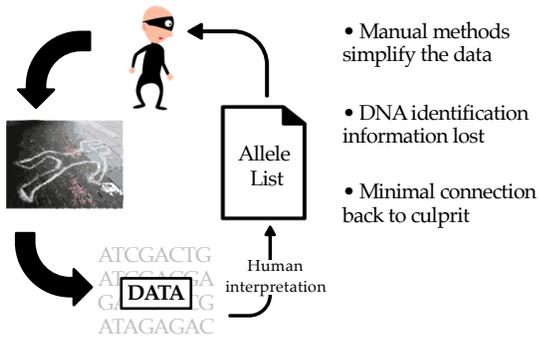
---

---

---

---

## A broken link <sup>3</sup>



---

---

---

---

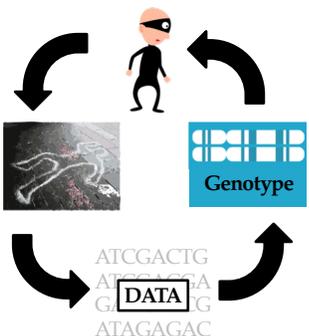
---

---

---

---

### Probabilistic Genotyping 4



The diagram illustrates the Probabilistic Genotyping process. It starts with a suspect (represented by a stick figure) and a crime scene (represented by a photograph of a fingerprint). Arrows indicate the flow of information: from the suspect to a 'Genotype' box, and from the crime scene to a 'DATA' box. The 'DATA' box contains a DNA sequence: ATCGACTG, ATAGAGAC, G, and A. The 'Genotype' box contains a grid of colored squares representing genetic markers. A circular arrow connects the 'Genotype' box back to the suspect, and another circular arrow connects the 'DATA' box back to the crime scene.

- TrueAllele™ probabilistic genotyping
- Uses all of the data
- Preserves ID information
- Connect culprit to crime scene

---

---

---

---

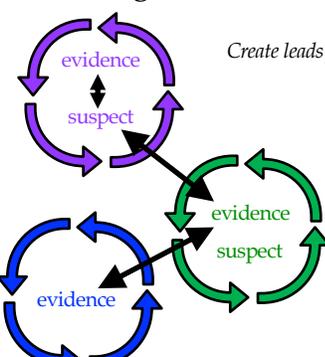
---

---

---

---

### Investigative TrueAllele Database 5



The diagram shows the Investigative TrueAllele Database. It features three interconnected circular loops: a purple loop for 'evidence' and 'suspect', a green loop for 'evidence' and 'suspect', and a blue loop for 'evidence'. Arrows indicate the flow of information between these loops. The text 'Create leads for criminal investigations' is written next to the loops.

- Link suspect to evidence within cases
- Link suspects to other cases
- Link evidence between crime scenes

---

---

---

---

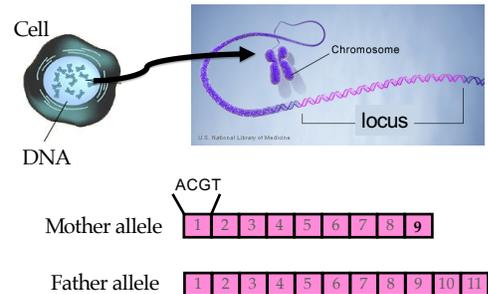
---

---

---

---

### STR genotype 6



The diagram illustrates the STR genotype. It shows a 'Cell' containing 'DNA', which is then shown as a 'Chromosome' with a 'locus' marked. Below this, the 'Mother allele' is shown as a sequence of 9 repeating units (1-9) and the 'Father allele' is shown as a sequence of 12 repeating units (1-12). The sequence is labeled 'ACGT'.

Mother allele: 1 2 3 4 5 6 7 8 9

Father allele: 1 2 3 4 5 6 7 8 9 10 11 12

---

---

---

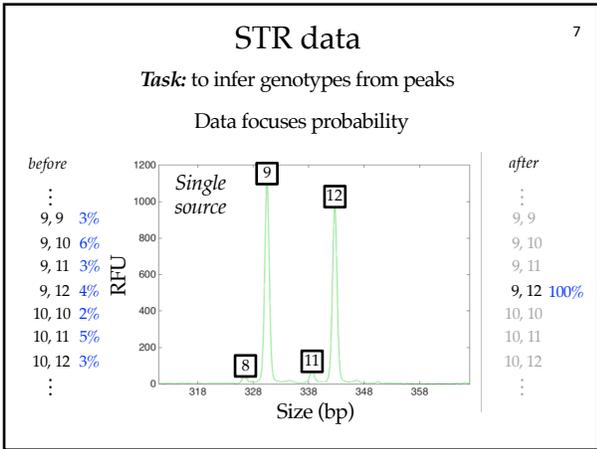
---

---

---

---

---




---

---

---

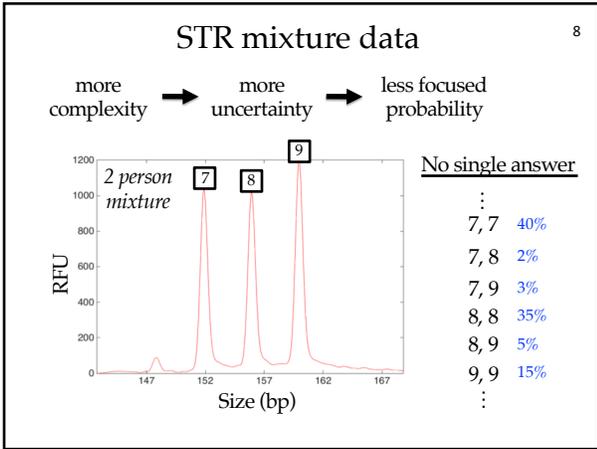
---

---

---

---

---




---

---

---

---

---

---

---

---

### Cleveland auto thefts 9

- Spike in auto thefts during summer / fall 2015
- Gang involvement suspected

37 cases

- 161 DNA evidence samples
- 72 references samples

---

---

---

---

---

---

---

---

## The challenges: Throughput

10

$$161 \text{ evidence samples} \times 72 \text{ references} = \sim 12,000 \text{ comparisons}$$

$$12,000 \text{ comparisons} \times 10 \text{ minutes per comparison} = > 2,000 \text{ human-hours}$$

- A lot of human time performing repetitive task

---

---

---

---

---

---

---

---

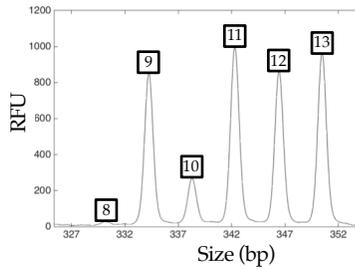
## The challenges: Data complexity

11



Reported conclusions: "...complexity of the data..."  
"...mixture of at least 3 individuals..."

"...therefore, no further conclusions can be drawn."



---

---

---

---

---

---

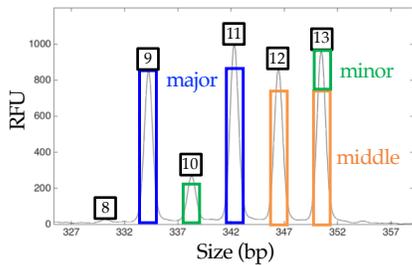
---

---

## Interpreting complex data

12

- TrueAllele computer proposes possibilities
- Compares with data to calculate probability
- Good explanations → higher probability



---

---

---

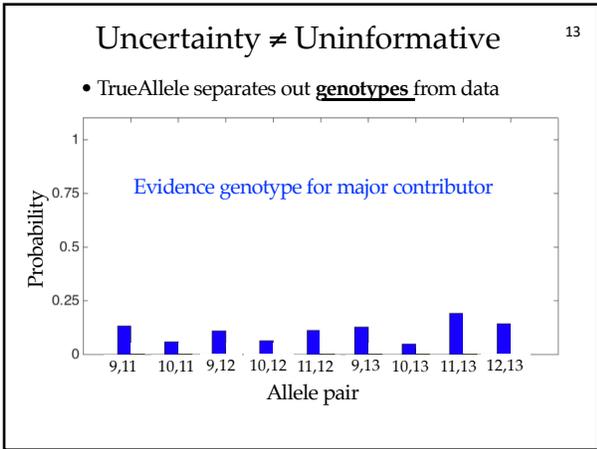
---

---

---

---

---




---

---

---

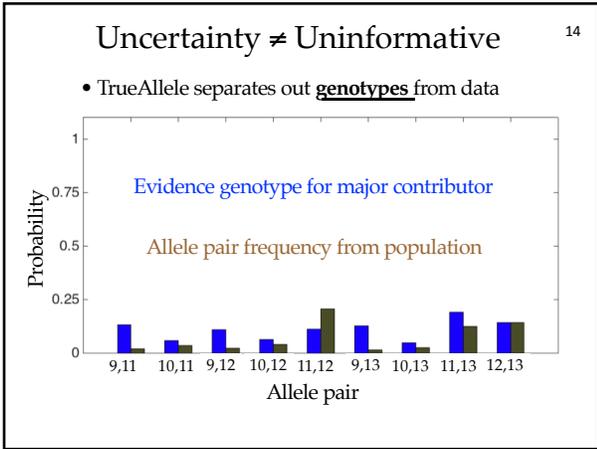
---

---

---

---

---




---

---

---

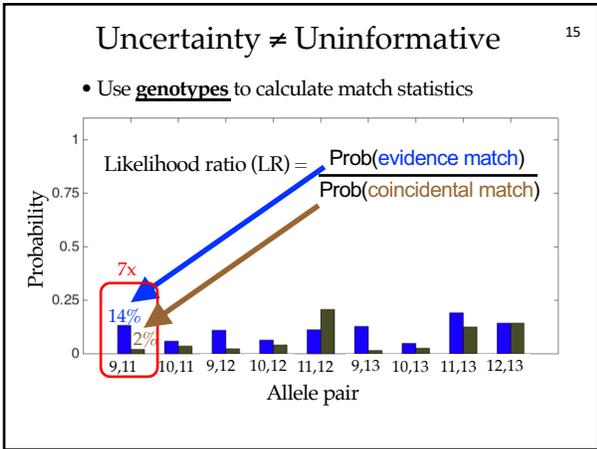
---

---

---

---

---




---

---

---

---

---

---

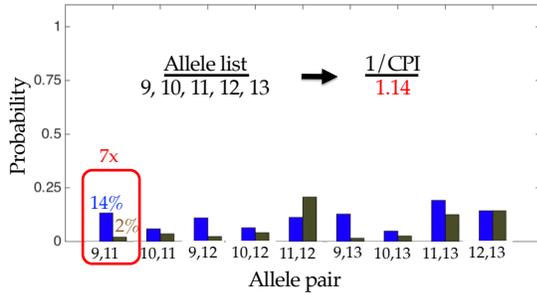
---

---

## Uncertainty ≠ Uninformative

16

- Allele lists loose identification information




---

---

---

---

---

---

---

---

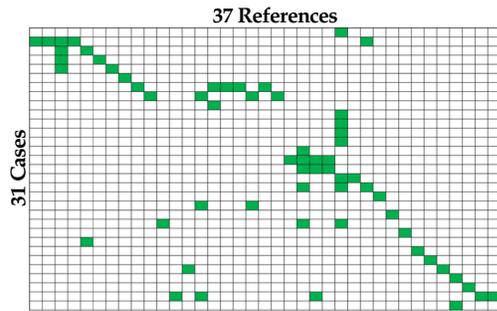
---

---

## Linking suspects to crimes

17

161 items → 1,200 comparisons → 127 ref-evi matches  
72 references




---

---

---

---

---

---

---

---

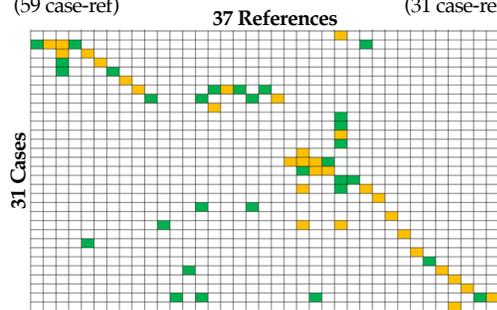
---

---

## The missing links

18

TrueAllele Database (59 case-ref) Manual & T.A. (31 case-ref)




---

---

---

---

---

---

---

---

---

---

## Method comparison

19

### Manual review

~2 months from data to interpretations

23 references

30 items 'inconclusive'

12 reported exclusions

### TrueAllele IPGD

Few hours of analyst time

37 references

At least one reference association for all 30

12 positive associations

---

---

---

---

---

---

---

---

---

---

## Match review

20

### Search

### Download

| Select                   | Case 1     | Case 2     | log(LR) |
|--------------------------|------------|------------|---------|
| <input type="checkbox"/> | 2015-4210  | 2015-4210  | 36.772  |
| <input type="checkbox"/> | 2015-4623  | 2015-4623  | 35.104  |
| <input type="checkbox"/> | 2015-4518  | 2015-4518  | 34.444  |
| <input type="checkbox"/> | 2015-4516  | 2015-4520  | 33.014  |
| <input type="checkbox"/> | 2015-4520  | 2015-4520  | 32.997  |
| <input type="checkbox"/> | 2015-4520  | 2015-4516  | 32.989  |
| <input type="checkbox"/> | 2015-4515  | 2015-4515  | 32.606  |
| <input type="checkbox"/> | 2015-5553  | 2015-5553  | 32.591  |
| <input type="checkbox"/> | 2015-4422  | 2015-4520  | 32.359  |
| <input type="checkbox"/> | 2015-4422  | 2015-4516  | 32.351  |
| <input type="checkbox"/> | 2015-4516  | 2015-4516  | 32.104  |
| <input type="checkbox"/> | 2015-4500  | 2015-4422  | 31.535  |
| <input type="checkbox"/> | 2015-4519  | 2015-4519  | 31.071  |
| <input type="checkbox"/> | 2015-4525  | 2015-4525  | 30.696  |
| <input type="checkbox"/> | 2018-12345 | 2018-54321 | 28.977  |
| <input type="checkbox"/> | 2018-12345 | 2018-12345 | 28.977  |
| <input type="checkbox"/> | 2018-54321 | 2018-12345 | 28.895  |
| <input type="checkbox"/> | 2015-4621  | 2015-4621  | 28.091  |
| <input type="checkbox"/> | 2015-4513  | 2015-4513  | 25.160  |
| <input type="checkbox"/> | 2015-4516  | 2015-4419  | 22.584  |

---

---

---

---

---

---

---

---

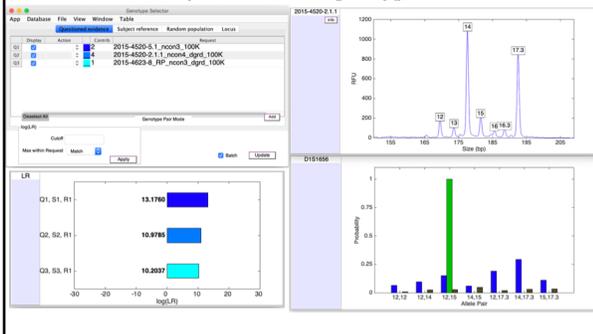
---

---

## Match review

21

- Quick and easy access to data, genotypes, match statistics




---

---

---

---

---

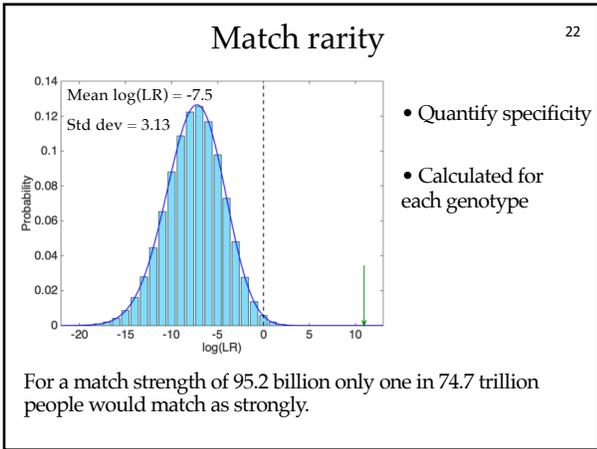
---

---

---

---

---




---

---

---

---

---

---

---

---

---

---

---

---

### Connecting crime scenes

23

Find Match

Genotype 1 **Match** Genotype 2

Filter CRFSL

Case Between

Rule EVI to EVI

Min log(LR)

After

Before

Recency newest

Limit 100

OK Cancel

- Use all the data
- Investigative leads between cases
- Database calculates match statistics between evidence items

| Select                   | Case 1     | Case 2     | log(LR) |
|--------------------------|------------|------------|---------|
| <input type="checkbox"/> | 2015-4516  | 2015-4520  | 33.014  |
| <input type="checkbox"/> | 2015-4520  | 2015-4516  | 32.989  |
| <input type="checkbox"/> | 2015-4422  | 2015-4520  | 32.359  |
| <input type="checkbox"/> | 2015-4422  | 2015-4516  | 32.351  |
| <input type="checkbox"/> | 2015-4520  | 2015-4422  | 31.535  |
| <input type="checkbox"/> | 2018-12345 | 2018-54321 | 28.977  |
| <input type="checkbox"/> | 2018-54321 | 2018-12345 | 28.895  |
| <input type="checkbox"/> | 2015-4516  | 2015-4419  | 22.584  |
| <input type="checkbox"/> | 2015-39065 | 2015-3906  | 21.980  |
| <input type="checkbox"/> | 2015-3906  | 2015-3905  | 19.756  |
| <input type="checkbox"/> | 2015-3981  | 2015-4319  | 18.838  |
| <input type="checkbox"/> | 2015-4354  | 2015-4515  | 18.351  |
| <input type="checkbox"/> | 2015-4420  | 2015-4419  | 17.871  |

$$\sum_{x \in G} \frac{\Pr(X = x | d_e) \Pr(Y = x | d_s)}{\Pr(X = x)}$$


---

---

---

---

---

---

---

---

---

---

---

---

### Evidence in action

24

- July & August 2013 in Bakersfield, CA
- Three separate break-ins and sexual assault cases

**The Bakersfield Californian** NEWS OBITS SPORTS EVENTS

Full Menu

HOME > NEWS

**Rapes have east Bakersfield residents on edge**

BY JASON KOTOWSKI Californian staff writer jkotoski@bakersfield.com

WEDNESDAY, AUG 21, 2013 6:43 PM

---

---

---

---

---

---

---

---

---

---

---

---

### Evidence in action

25

- 37 swabs and cuttings
- Low level DNA mixtures
- Evidence matches between cases
- CODIS hit on zip tie

---

---

---

---

---

---

---

---

### Evidence in action

26

- 37 swabs and cuttings
- Low level DNA mixtures
- Evidence matches between cases
- CODIS hit on zip tie
- Arrest (2013) and conviction (2015)

---

---

---

---

---

---

---

---

### World Trade Center

27

Powerful TrueAllele Database to identify victim remains

---

---

---

---

---

---

---

---

## Conclusions

28

- Automate DNA investigation
- Reduce interpretation bottleneck
- Extract more information from DNA data
- Match suspects to crimes
- Connect crime scenes together

---

---

---

---

---

---

---

## More information

[www.cybgen.com](http://www.cybgen.com)



- Courses
- Newsletters
- Newsroom
- Presentations
- Publications
- Webinars

**TrueAllele YouTube channel:**  
<http://www.youtube.com/user/TrueAllele>



Cybergenetics



dave@cybgen.com

---

---

---

---

---

---

---