"Using Computer Technology to Overcome Bottlenecks in the Forensic DNA Testing Process and Improve Data Recovery from Complex Samples"

Goals of DNA Testing

1. Help identify perpetrators of crimes
2. Eliminate the wrongfully accused
3. Help prevent future crimes

Reducing DNA Testing Time

1. Hire more staff
2. Automate
   - Humans for tasks requiring intelligence
   - Machines for repetitive processes

Typical Laboratory Automation Setup

Identify Material	Human
DNA Extraction	Automated, ~15 years
DNA Quantification	Automated, ~15 years
DNA Amplification	Automated, ~10-15 years
Instrumental Analysis	Automated, ~20 years
DNA Profile Interpretation	Human

How Many DNA Profiles?

- 96-well plate
- 4 to 8 allelic ladders
- At least two PCR controls
- Several DNA extraction blanks

Typical plate could contain 80 to 90 DNA profiles

Interpretation Guidelines

- 2017 SWGDAM Guidelines for human interpretation

To avoid confirmation bias, evidence samples should be interpreted before comparison to known samples

To Avoid Confirmation Bias

1. Look at your evidence profile first
2. Infer DNA types from the evidence without knowledge of known types
3. Compare inferred DNA types to knowns
4. Determine if match exists

Human Interpretation

Eight hour workday

- 8 hours x 60 minutes = 480 minutes
- 480 minutes / 84 DNA profiles =

5 minutes, 42 seconds per DNA profile

Interpretation Bottleneck

- Volume of data
- Complexity
- Thresholds
- Data declared “Inconclusive” or “Too Complex for Interpretation”

So Throwing Out Data is the Only Way?

Alternate Approach - Automate

Automate DNA interpretation using computers

- Use humans for tasks requiring intelligence, use machines for repetitive processes
- DNA interpretation can be automated with the TrueAllele® technology

TrueAllele® Processing

ViewStation User Client

Database Server

Interpret/Match Expansion

Visual User Interface VUIer™ Software

Parallel Processing Computers

Automated Process

1. Upload entire plate to server
2. Computer interprets the data
3. Check results and compare:
   - To other data files on the same plate
   - To known profiles
   - To previous runs
4. Perform detailed processing on potential matches

Benefits of Automated DNA Interpretation

1. All data examined, nothing discarded
2. Speed
   - One plate in ~6 – 7 hours
3. No confirmation bias
   - Computer infers genotypes
   - No prior knowledge

Benefits of Automated DNA Interpretation

4. All data compared:
   - Evidence, references, lab staff, crime scene investigators, controls
   - Identify more case-to-case matches and potential contamination
5. CODIS specimen and candidate match assessment

Database Matching

- TrueAllele implemented in January 2016
- Server expansion modules installed May 2017
  - 20 processors (casework and database screening)
  - 8 (original) processors dedicated to database screening

TrueAllele Workflows

1. Casework
   - Traditional sample-by-sample analysis
2. Database
   - Process everything, look for matches
   - Confirm matches in Casework process

Database Matching

Uploaded all 7 years of data to BCSO TrueAllele database

- >7,500 DNA profiles
- >15,000 inferred genotypes
- Batched request run conditions:
  - 5K/5K
  - 3 unknown contributors

Database Matching

- ~30,000 potential matches returned
- Most were “within case”
- To date, ~80 previously unknown EVI-REF case-to-case matches

Match Evaluation

- Five samples from five different cases matched to each other
- EVI – EVI; no reference matches
- No matches to any other genotypes

Sample	Contributor	N Con	KL	2016 Case (log LR)
2014 Case	2	3	19.9574	12.5962
2015 Case	1	3	19.675	14.7183
2017 Case	3	3	14.6881	6.2815
2015 Case 2	1	3	17.401	8.0723

Case Evaluation

2014 Home Invasion (unknown suspect)
2015 Attempted Murder (known suspect)
2015 Burglary (two known suspects)
2016 Murder (known suspect)
2017 B&E Auto (three known suspects)

Case Evaluation

- No suspects in common
- 2015 Attempted Murder & 2016 Murder: identity was not in dispute
- 2017 B&E Auto: suspects were witnessed and caught soon after

Possible Causes

1. Wrongfully accused?
2. An accomplice that we don't know about?
3. Contamination?
4. Aliens?

Investigation

1. Worked in lab at different times
2. Worked in lab by three different examiners
3. Genotypes did not match lab staff
4. Genotypes did not match any case reference samples

Investigation

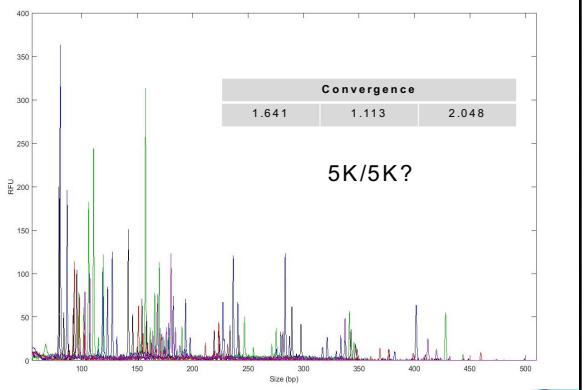
5. All cases worked by same agency
6. All items were touch DNA
7. All evidence was collected or handled by same investigator
8. Investigator was the only common link between all five cases

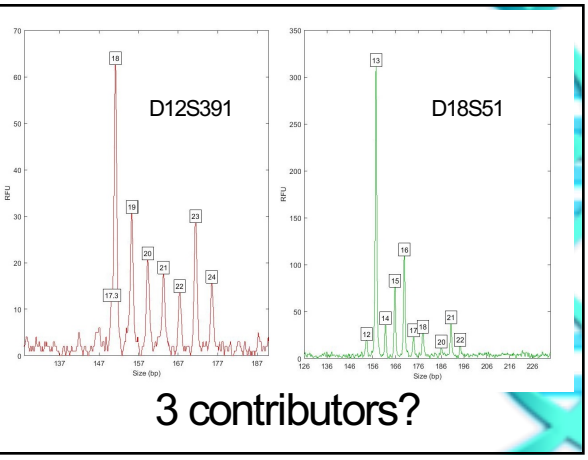
Sample	Contributor	N Con	KL	Reference (log LR)
2014 Home Invasion	2	3	19.9574	13.4844
2015 Burglary	1	3	19.675	15.4627
2016 Murder	1	3	23.3979	17.798
2017 B&E Auto	3	3	14.6881	11.0779
2015 Attempted Murder	1	3	17.401	4.0144

Run Conditions

Batched request:

- 5K/5K
- 3 unknown contributors





3 contributors?

Run Conditions

Batched request:

- 5K/5K
- 1, 2, or 3 unknown contributors

➢ Identify potential matches for detailed processing

Quality Control

- Two of the profiles had been entered into CODIS
  - LDIS match
- Other three were unsuitable for CODIS entry

CODIS Match Evaluation

CODIS Match Evaluation Example #1

Offender	Human Review, 30 minutes	TrueAllele, 5 minutes	LR	CPI (1 in)
#1	Uncertain	Eliminated	1.5	39,000
#2	Not eliminated	Eliminated	2.7	39,000
#3	Not eliminated	Match	73 billion	39,000

CODIS Match Evaluation Example #2

Offender	MME	CPI (1 in)	LR
#1	1.728×10^4	65	158 trillion

Previously Unidentified Matches

2014: Burglary case, uploaded to SDIS

- Offender hit
- Match confirmed in laboratory

2017: Process old data, upload to TrueAllele®

- 3 additional cases from 2012 - 2014
- Never entered into CODIS

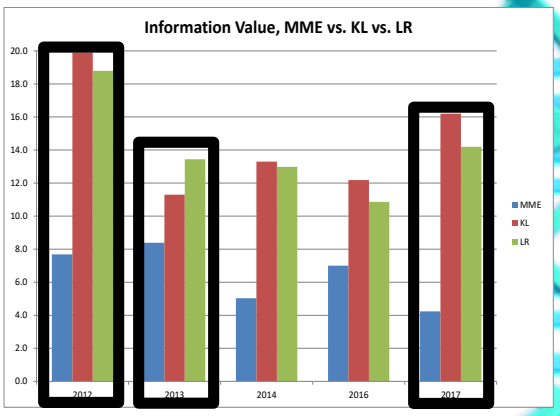
Sample	Minimum # Contributors	Major?	4x4 Rule?	CPI (1 in)	LR
2012	3	No	No	28,000	6 quintillion
2013	3	No	No	920	27 trillion
2014	3	No	No	760	9 trillion

Can we use the automated process for CODIS screening?

CODIS Screening

1. KL computed by TrueAllele®
   - Measures information value of inferred genotype
2. MME calculated by CODIS
   - Predicts matches at moderate stringency

Compare MME, KL, and LR for CODIS profile assessment



What We Are Implementing

1. Use KL to predict quality of match
2. Use MME to filter adventitious matches
   - High KL – build MME, search CODIS
   - Low KL – do not upload

Summary

- DNA interpretation is automatable
- Reduce/eliminate interpretation bottlenecks
- Output searched internally and screened for suitable CODIS profiles
- Improved quality control
- More information recovered from same amount of data

Questions?