An Automated Genotype Database for Associating Counterfeit Drug Evidence

American Academy of Forensic Sciences
Criminalistics Section

February 2023
Orlando, FL

Matthew Legler
William Allan, MS
Mark Perlin, PhD, MD, PhD
Cybergenetics © 2003-2023

Cybergenetics

Background
Counterfeit pharmaceuticals

- Counterfeit pharmaceuticals kill over 100,000 people each year
- In two decades, the number of incidents has grown from 200 to 4,400
- Pfizer Global Security tested 250 Xanax purchases online; 96% were counterfeit
- The annual $100 billion counterfeit drug market rivals the $250 billion illicit drug trade

---

Counterfeit drug project

- Private company has a collection of samples from known counterfeit drug packages
- DNA lab generated STR data over an 18-month period, November 2018 through May 2020
- Analyzed 133 samples, representing 103 package evidence items
- The lab analysis produced no information
- Can a TrueAllele Database find more?

Cybergenetics processing

- Process all of the data in TrueAllele® Casework
  - 103 items (swabs from vials, packaging, seals, etc.)
  - Data typically contained 2 to 5 contributors
- Task - Find connections between different collections of counterfeit drugs
- Data received: May 27, 2020
- First results returned: June 3, 2020

Manufacturer Case
Case 5

vial and label edges collected by Lab Tech AS February 2019

Laboratory: Item is "Not Searchable Mixture"
TrueAllele: Expected match statistic of a **nonillion** (10^{30})

Manufacturer Case
Case 5
Case 8

1 nonillion

metal seal, vial label, flip cap and bottom rim collected by Lab Tech JD May 2019

Laboratory: Both items are "Not Searchable Mixtures"
TrueAllele: Finds match statistic association of a **nonillion** (10^{30})

Slide 1

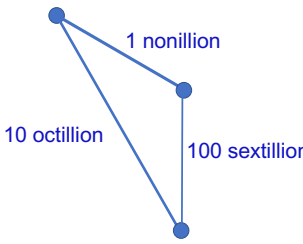Manufacturer Case

Case 5
Case 8
Case 19

Laboratory: All items are "Not Searchable Mixtures"
TrueAllele: Finds a cluster of matches

1 nonillion
10 octillion
100 sextillion

carton, vial, patient insert collected by Lab Tech PN January 2020

Slide 2
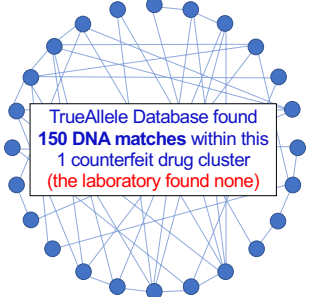
Nov 2018
Feb 2020

Manufacturer Case

Case 1	Case 12
Case 2	Case 13
Case 3	Case 14
Case 4	Case 15
Case 5	Case 16
Case 6	Case 17
Case 7	Case 18
Case 8	**Case 19**
Case 9	Case 20
Case 10	Case 21
Case 11	

TrueAllele Database found **150 DNA matches** within this 1 counterfeit drug cluster (the laboratory found none)

Slide 3

Manufacturer Case

Case 5

vial and label edges collected by Lab Tech AS February 2019

CODIS database searchable profile Generated by TrueAllele Database

Expected database match statistic of **10 billion** (10^{10})

Expected TrueAllele® match statistic of a **nonillion** (10^{30}) to the true perpetrator

Marker	Allele 1	Allele 2	Allele 3
AMELO	X	Y	
CSF1PO	11	14	
D10S1248	14	15	
D12S391	17.3	20	
D13S317	10	11	13
D16S539	11	12	
D18S51	14	16	
D19S433	12	14	
D1S1656	16	17.3	
D21S11	27	29	
D22S1045	11	15	
D2S1338	17	20+	
D2S441	14		
D3S1358	14	17	
D5S818	11	13+	
D7S820	11		
D8S1179	13	15	
FGA	22		
SE33	13.2	27.2	
TH01	7	9.3	
TPOX	8	11	
vWA	15	16	

Results – The Numbers

DNA data from 103 counterfeit packages

	TrueAllele Genotypes	Failed Allele Lists
Information	723 DNA connections	0
CODIS profiles	6	0
Interpretation time	1 week	1 year + 6 months

Probabilistic databases overcome allele database obstacles

Allele Databases
Limited uploads, limited information

Probabilistic databases
Everything can be uploaded, preserve information

Alleles
10, 11, 12

Probabilities
[10 11]   0.90
[10 12]   0.10

Probabilistic databases overcome allele database obstacles

MME vs KL

Moderate Match Estimate
must exceed a pre-set threshold
reduce CODIS hits

KL (Kullback–Leibler)
provides expected log(LR) information
not a limit on evidence upload

Information prediction: MME vs KL
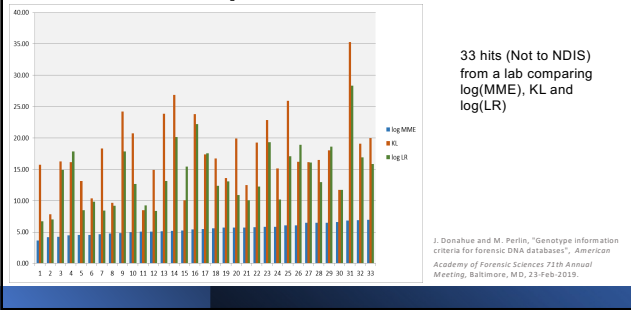
33 hits (Not to NDIS) from a lab comparing log(MME), KL and log(LR)

J. Donahue and M. Perlin, "Genotype information criteria for forensic DNA databases", *American Academy of Forensic Sciences 71th Annual Meeting, Baltimore, MD, 23-Feb-2019.*

Database as investigative tool
Drug distribution sites

- Fentanyl, Meth
  - Can you find the original producers and suppliers?
- Data – drug baggies, vials, scales, etc. collected at distribution sites
- Task – find shared contributors across multiple locations, produce profiles for database search

Database as investigative tool
Organized retail crime

- For organized retail crime and retail theft
- Is there a shared contributor on recovered items?
- Database to identify and link criminals and evidence
- Maps the crime scene connecting suspects and evidence items

Conclusions

- Able to interpret in 1 week, 1½ years worth of data
- Found 723 DNA matches and 6 CODIS-searchable profiles
- Did not require extensive manual data review
- Database automatically compared the DNA data
- An automated, informative DNA database can catch counterfeiters and save lives
- Can be extended to other big DNA data problems, such as illicit drug production and organized retail theft