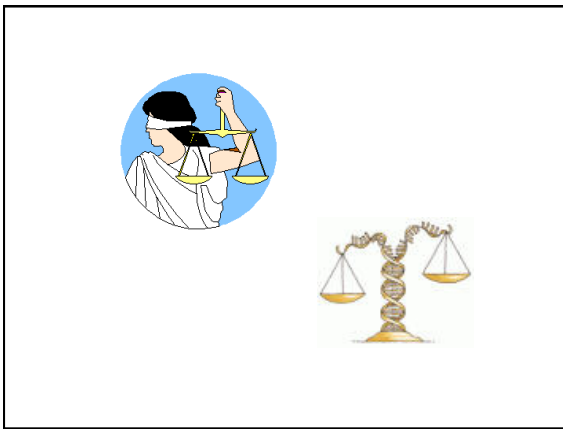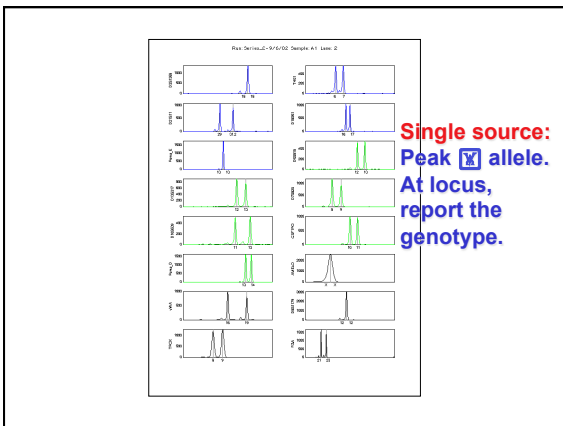**Simple Reporting of
Complex DNA Evidence:
Automated Computer Interpretation**

**Promega 14th International Symposium
on Human Identification
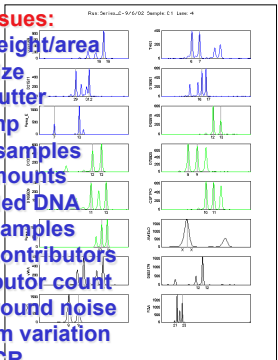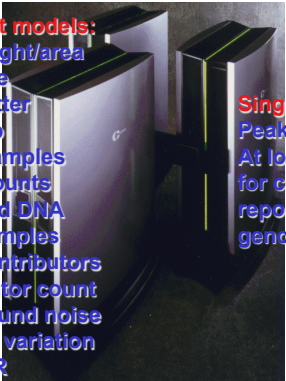Pointe Hilton Tapatio Cliffs Resort
Phoenix, AZ**

October, 2003

Cybergenetics

**Single source:
Peak = allele.
At locus,
report the
genotype.**

Data Issues:
Peak height/area
Peak size
PCR stutter
Pref amp
Mixed samples
Low amounts
Degraded DNA
Many samples
Many contributors
Contributor count
Background noise
Random variation
LCN PCR
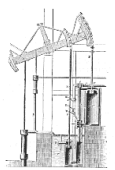
Math/stat models:
Peak height/area
Peak size
PCR stutter
Pref amp
Mixed samples
Low amounts
Degraded DNA
Many samples
Many contributors
Contributor count
Background noise
Random variation
LCN PCR

Single source:
Peak ⊠ allele.
At locus,
for contributor,
report the
genotype(s).

Recent Revolutions

- **science**
    - **PCR, sequencer**
- **technology**
    - **computer**
- **law**
    - **Daubert**
- **statistics**
    - **computational truth**

TrueAllele™

Expert system
  a computer program that replicates human expertise

History

System: 2 → 3
  Version: 14 → 15

Admissibility

Rule 702
1. Data
2. Method
3. Application

Daubert
1. Testable
2. Error rate
3. Peer review
4. General acceptance

Validation Design

Set 1	ng amplified			
Mixture ratio	1	0.5	0.25	0.125
10:0	A1	A2	A3	A4
9:1	B1	B2	B3	B4
7:3	C1	C2	C3	C4
5:5	D1	D2	D3	D4
3:7	E1	E2	E3	E4
1:9	F1	F2	F3	F4
0:10	G1	G2	G3	G4

10%
30%
50%
70%
90%

1, 1/2, 1/4, 1/8 ng

Set 2	ng amplified			
Mixture ratio	1	0.5	0.25	0.125
10:0	H1	H2	H3	H4
9:1	I1	I2	I3	I4
7:3	J1	J2	J3	J4
5:5	K1	K2	K3	K4
3:7	L1	L2	L3	L4
1:9	M1	M2	M3	M4
0:10	N1	N2	N3	N4

STR Data Generation

- premixed DNA templates: NIST
- lab protocols: Cybergenetics
- data generation (ten DNA labs)
  Florida, New York, Ohio, Pennsylvania,
  Virginia, Cellmark, UK FSS, Cybergenetics
- DNA sequencers:
  FMBio/II, 377, 310, 3100, 3700
- STR panels: PowerPlex (1, 2, 16),
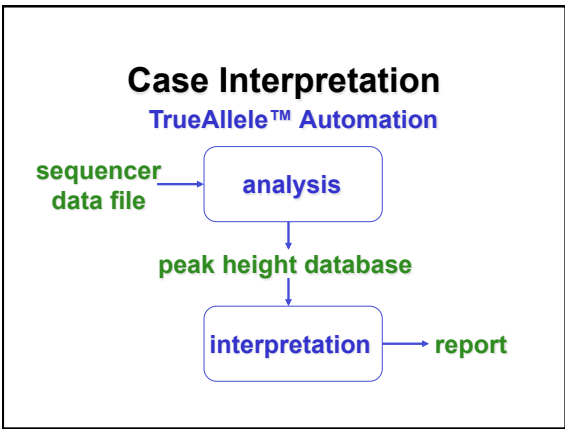 ProfilerPlus, Cofiler, SGMplus, Identifiler

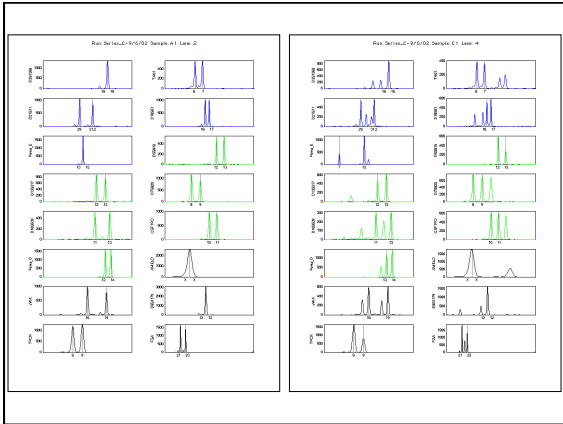## Data Analysis
TrueAllele™ Automation

sequencer
data file → analysis

peak height database

Case Interpretation
TrueAllele™ Automation

sequencer data file → [**analysis**]

↓

peak height database

↓

[**interpretation**] → **report**
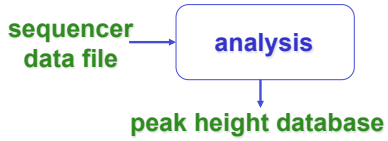
No Suspect Example
1 ng DNA, PowerPlex16, ABI/310

Contributors
A: Victim
G: Unknown Suspect

Samples
A1 (Victim)
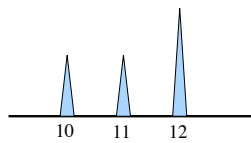C1 (Mixture) **70% A** + **30% G**

Simplicity Goals

- time < 1 minute
- information full power
- understandable yes
- admissible yes

1,440 minutes each day
Backlog?
(no suspect, convicted offender)

Discriminating Power

10   11   12

Conservative		Cautious		Exact	
10	10	10	12	10	12
10	11	11	12		
10	12	12	12		
11	11				
11	12				
12	12				

Concordance: DP

**Synthetic: NIJ study comparisons**
Cases: 1,000 no suspect rape kits

locus	TrueAllele™ Interpretation		Human Interpretation	
	allele 1	allele 2	allele 1	allele 2
CSF1PO	12	12	12	
D13S317	9	13	9	
D16S539	9	12	9	12
D18S51	13	15	13	15
D21S11	30	31	30	31
D3S1358	16	17	16	17
D5S818	12	12		
D7S820	10	10	10	
D8S1179	8	11	8	11
FGA	21	22		22
TH01	9	9.3	9	9.3
TPOX	8	8		
vWA	15	18	15	18

26/26   **18/26**

3.6×10^{16} 6.1×10^{11}

Lab Process Scenarios

• screen data, organize case
• second scorer review
• post-conviction DNA testing
• service for DNA analyst
• NDIS forensic upload
• NDIS profile download
• mass disasters
• serial crime
• advance troubleshooting
• …
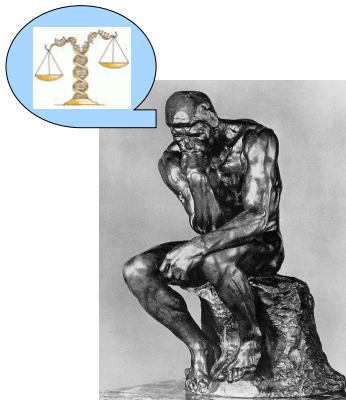
Conclusions

casework interpretation system:
   objective, unbiased, reproducible,
   reliable, admissible, available
reports: simple, understandable
studies: concordance, validation

   • time
   • information
   • understandable
   • admissible

Acknowledgements

National Institute of Justice:
   Award #2001-IJ-CX-K003

Collaborators at:
   Florida, New York, Ohio, Pennsylvania,
   Virginia, NIST, Cellmark, UK FSS

Cybergenetics:
   Developers, Quality Process, DNA Lab

Cybergenetics   perlin@cybgen.com