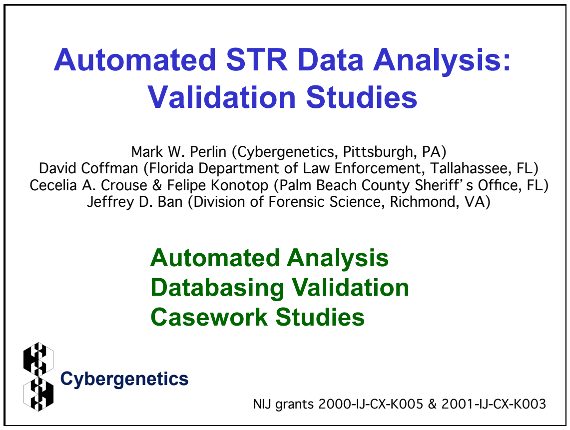


Automated STR Data Analysis: Validation Studies

Mark W. Perlin (Cybergenetics, Pittsburgh, PA)
David Coffman (Florida Department of Law Enforcement, Tallahassee, FL)
Cecelia A. Crouse & Felipe Konotop (Palm Beach County Sheriff's Office, FL)
Jeffrey D. Ban (Division of Forensic Science, Richmond, VA)

Automated Analysis
Databasing Validation
Casework Studies

Cybergenetics

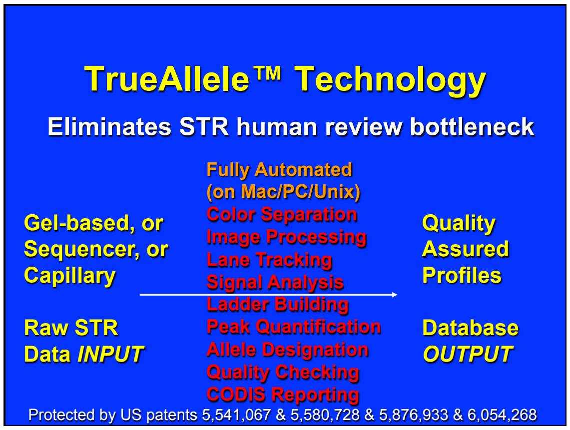
NIJ grants 2000-IJ-CX-K005 & 2001-IJ-CX-K003

Reviewing STR Data

Human data review bottleneck

Computer Automation
Quality Assurance
Database Integrity
Casework & Mixtures

Key goals:
- no error
- high throughput
- small staff



TrueAllele™ Technology

Eliminates STR human review bottleneck

Gel-based, or
Sequencer, or
Capillary

Raw STR
Data *INPUT*

Fully Automated
(on Mac/PC/Unix)
Color Separation
Image Processing
Lane Tracking
Signal Analysis
Ladder Building
Peak Quantification
Allele Designation
Quality Checking
CODIS Reporting

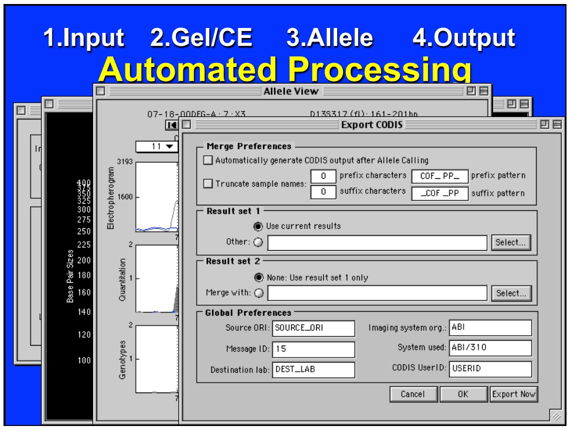
Quality
Assured
Profiles

Database
OUTPUT

Protected by US patents 5,541,067 & 5,580,728 & 5,876,933 & 6,054,268

1.Input 2.Gel/CE 3.Allele 4.Output

Automated Processing



Export CODIS

Merge Preferences

☐ Automatically generate CODIS output after Allele Calling

☐ Truncate sample names: 0 prefix characters COF_ PP_ prefix pattern; 0 suffix characters _COF _PP suffix pattern

Result set 1

● Use current results

Other: ○ [Select...]

Result set 2

● None: Use result set 1 only

Merge with: ○ [Select...]

Global Preferences

Source ORI: SOURCE_ORI

Imaging system org.: ABI

Message ID: 15

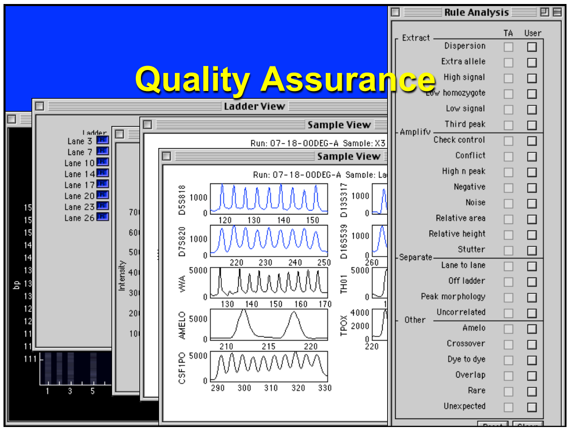
System used: ABI/310

Destination lab: DEST_LAB

CODIS UserID: USERID

[Cancel] [OK] [Export Now]

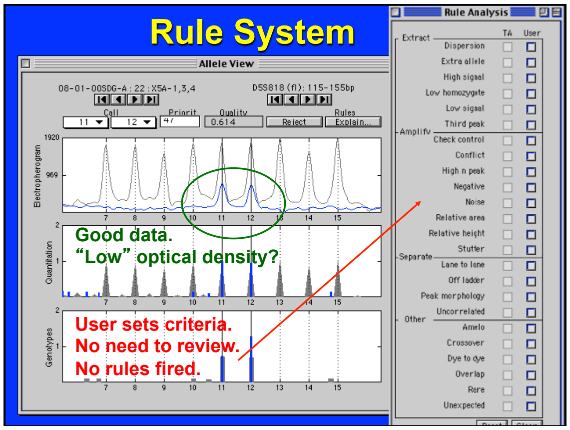
Quality Assurance



Rule Analysis

		TA	User
Extract	Dispersion	☐	☐
	Extra allele	☐	☐
	High signal	☐	☐
	Low homozygote	☐	☐
	Low signal	☐	☐
	Third peak	☐	☐
Amplify	Check control	☐	☐
	Conflict	☐	☐
	High n peak	☐	☐
	Negative	☐	☐
	Noise	☐	☐
	Relative area	☐	☐
	Relative height	☐	☐
	Stutter	☐	☐
Separate	Lane to lane	☐	☐
	Off ladder	☐	☐
	Peak morphology	☐	☐
	Uncorrelated	☐	☐
Other	Amelo	☐	☐
	Crossover	☐	☐
	Dye to dye	☐	☐
	Overlap	☐	☐
	Rare	☐	☐
	Unexpected	☐	☐

Rule System



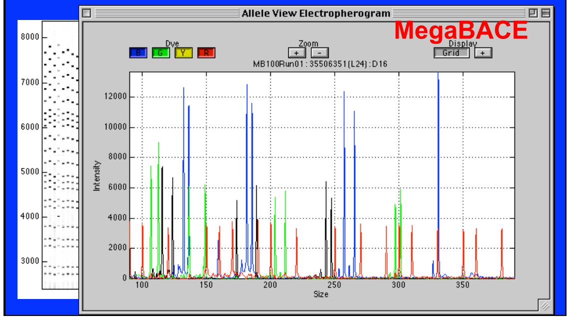
Good data.
“Low” optical density?

User sets criteria.
No need to review.
No rules fired.

Rule Analysis

		TA	User
Extract	Dispersion	☐	☐
	Extra allele	☐	☐
	High signal	☐	☐
	Low homozygote	☐	☐
	Low signal	☐	☐
	Third peak	☐	☐
Amplify	Check control	☐	☐
	Conflict	☐	☐
	High n peak	☐	☐
	Negative	☐	☐
	Noise	☐	☐
	Relative area	☐	☐
	Relative height	☐	☐
	Stutter	☐	☐
Separate	Lane to lane	☐	☐
	Off ladder	☐	☐
	Peak morphology	☐	☐
	Uncorrelated	☐	☐
Other	Amelo	☐	☐
	Crossover	☐	☐
	Dye to dye	☐	☐
	Overlap	☐	☐
	Rare	☐	☐
	Unexpected	☐	☐

Multi-Platform Engine

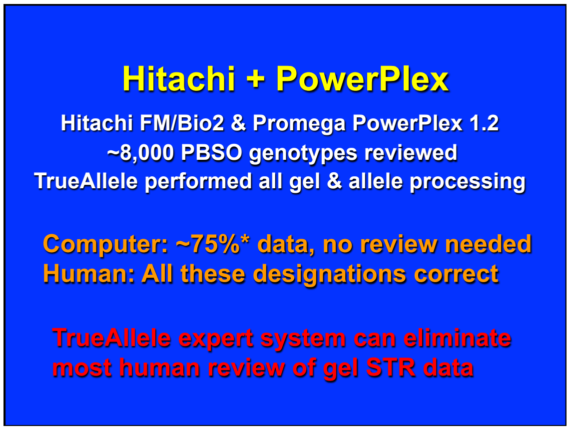


Validation Methods

1. Obtain original data
2. Process data in TrueAllele ES
   (auto-setup, process run, Q/A,
   call alleles, apply rules, check)
   computer: accept/reject/edit
3. Review all data
   one person, many computers
   human: accept/reject/edit
4. Generate results & stats

Rule Settings

		Gel		Capillary	
		On/Off	Value	On/Off	Value
Extract	Dispersion	On	60%	On	80%
	Extra Allele	On	25%	On	15%
	High Signal	On	9000	On	10000
	Low Signal	On	150	On	300
	Low Homozygote	On	250	On	500
	Third Peak	On	23%	On	15%
Amplify	Check Control	On	----	On	----
	Conflict	On	----	On	----
	High n Peak	On	30%	On	30%
	Negative	On	----	On	----
	Noise	On	----	On	----
	Relative Area	On	50%	On	50%
	Relative Height	On	50%	On	50%
	Stutter	On	20%	On	15%
Separate	Lane to Lane	Off	----	Off	----
	Off Ladder	On	0.4	On	0.4
	Peak Morphology	On	50%	On	20%
	Uncorrelated	On	0.4	On	0.4
Other	Amelo	On	----	On	----
	Crossover	On	----	On	----
	Dye to Dye	Off	----	Off	----
	Overlap	On	----	On	----
	Rare	On	----	On	----
	Unexpected	On	25%	On	15%

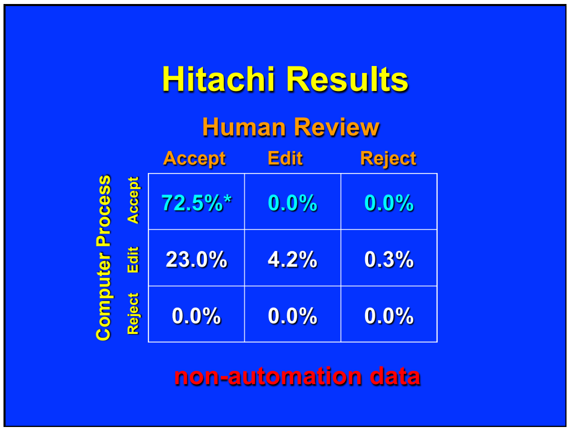


Hitachi + PowerPlex

Hitachi FM/Bio2 & Promega PowerPlex 1.2
~8,000 PBSO genotypes reviewed
TrueAllele performed all gel & allele processing

Computer: ~75%* data, no review needed
Human: All these designations correct

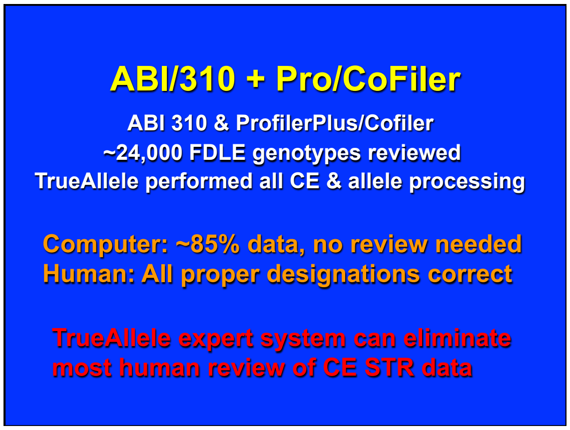
TrueAllele expert system can eliminate most human review of gel STR data



Hitachi Results

		Human Review		
		Accept	Edit	Reject
Computer Process	Accept	72.5%*	0.0%	0.0%
	Edit	23.0%	4.2%	0.3%
	Reject	0.0%	0.0%	0.0%

non-automation data

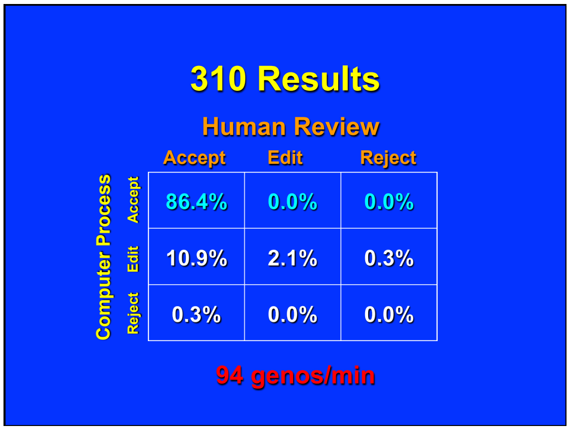


ABI/310 + Pro/CoFiler

ABI 310 & ProfilerPlus/Cofiler
~24,000 FDLE genotypes reviewed
TrueAllele performed all CE & allele processing

Computer: ~85% data, no review needed
Human: All proper designations correct

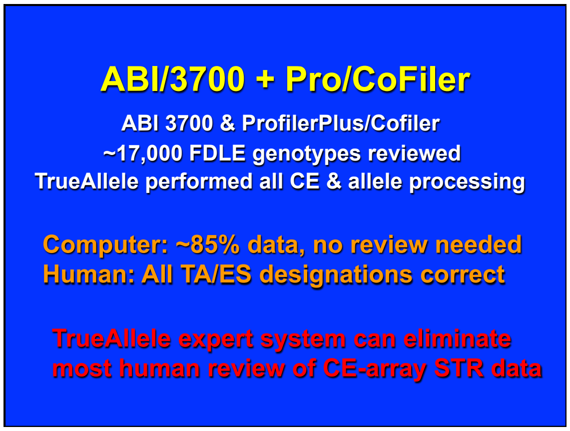
TrueAllele expert system can eliminate most human review of CE STR data



310 Results

		Human Review		
		Accept	Edit	Reject
Computer Process	Accept	86.4%	0.0%	0.0%
	Edit	10.9%	2.1%	0.3%
	Reject	0.3%	0.0%	0.0%

94 genos/min

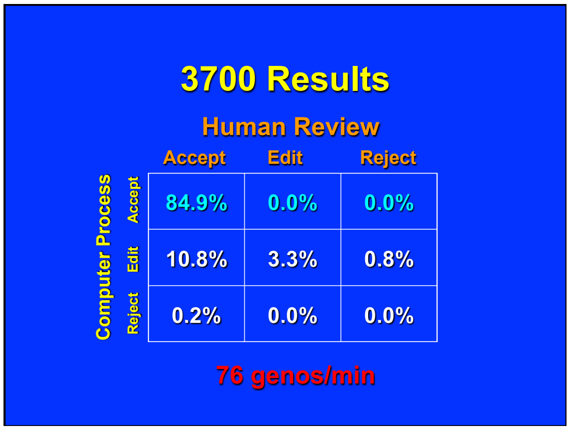


ABI/3700 + Pro/CoFiler

ABI 3700 & ProfilerPlus/Cofiler
~17,000 FDLE genotypes reviewed
TrueAllele performed all CE & allele processing

Computer: ~85% data, no review needed
Human: All TA/ES designations correct

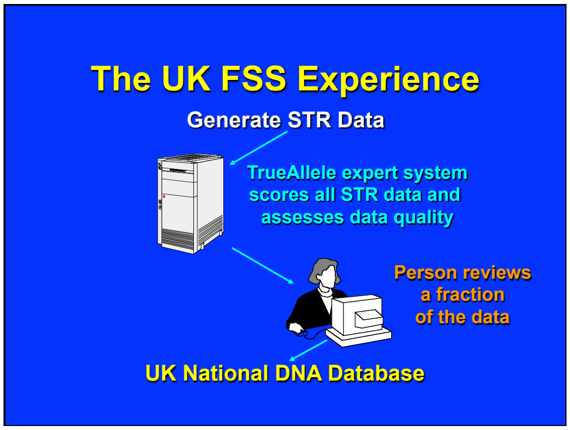
TrueAllele expert system can eliminate
most human review of CE-array STR data



3700 Results

		Human Review		
		Accept	Edit	Reject
Computer Process	Accept	84.9%	0.0%	0.0%
	Edit	10.8%	3.3%	0.8%
	Reject	0.2%	0.0%	0.0%

76 genos/min



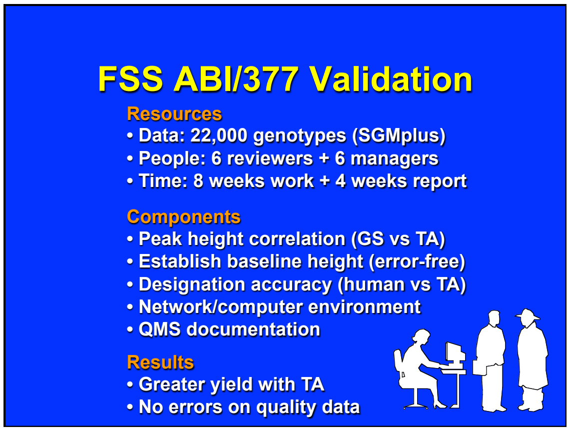
The UK FSS Experience

Generate STR Data

TrueAllele expert system scores all STR data and assesses data quality

Person reviews a fraction of the data

UK National DNA Database

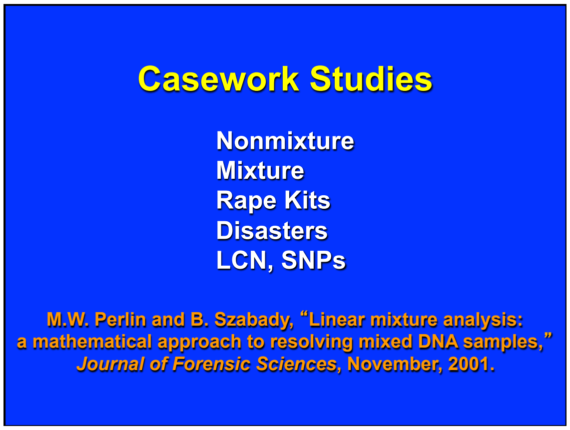


FSS ABI/377 Validation

Resources
- Data: 22,000 genotypes (SGMplus)
- People: 6 reviewers + 6 managers
- Time: 8 weeks work + 4 weeks report

Components
- Peak height correlation (GS vs TA)
- Establish baseline height (error-free)
- Designation accuracy (human vs TA)
- Network/computer environment
- QMS documentation

Results
- Greater yield with TA
- No errors on quality data



Casework Studies

Nonmixture
Mixture
Rape Kits
Disasters
LCN, SNPs

M.W. Perlin and B. Szabady, "Linear mixture analysis: a mathematical approach to resolving mixed DNA samples," *Journal of Forensic Sciences*, November, 2001.

Statistical Information

Prob{STR profile & peak quants}

Degenerate SGM+ alleles (6x6x6x10x...x6)
☒ 100,000,000 feasible profiles
Compute unknown minor contrib profiles
☒ @30% 1 feasible profile
☒ @10% 100 feasible profiles

LMA increases identification power
a million-fold; CODIS match

Validation Data Sets

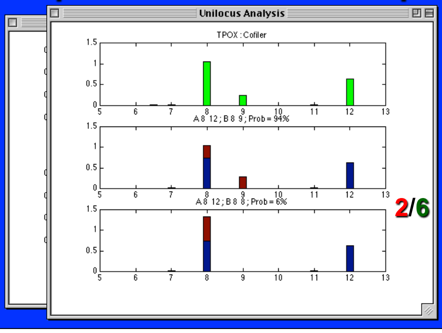
Collaborators: Florida, Virginia, New York, FBI, UK, Private Labs

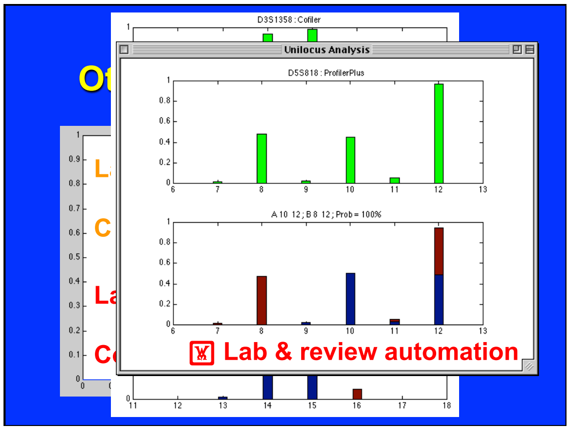
Data: synthetic lab mixtures, casework, rape kits, disasters

Input: TrueAllele quantitated peaks

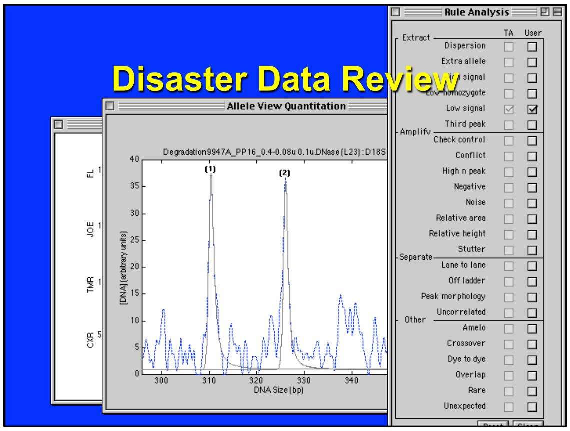
Studies: comparison, concordance, automated lab processes

Rape Kit: Unknown Suspect





Lab & review automation



Disaster Data Review

		TA	User
Extract	Dispersion	☐	☐
	Extra allele	☐	☐
	signal	☐	☐
	Low homozygote	☐	☐
	Low signal	☑	☑
	Third peak	☐	☐
Amplify	Check control	☐	☐
	Conflict	☐	☐
	High n peak	☐	☐
	Negative	☐	☐
	Noise	☐	☐
	Relative area	☐	☐
	Relative height	☐	☐
	Stutter	☐	☐
Separate	Lane to lane	☐	☐
	Off ladder	☐	☐
	Peak morphology	☐	☐
	Uncorrelated	☐	☐
Other	Amelo	☐	☐
	Crossover	☐	☐
	Dye to dye	☐	☐
	Overlap	☐	☐
	Rare	☐	☐
	Unexpected	☐	☐

Conclusions

- TrueAllele *databasing validation*
- Reduce time, error, staff & cost

- Ongoing *casework validation*
- Automate: data review & lab work
- Serve: police, courts, society
- Objective, comprehensive