STR DNA PROFILING – THE STANDARD FOR CELL LINE AUTHENTICATION

Yvonne A. Reid, Ph.D.

Manager/Scientist, Cell Biology Program, ATCC

Doug Storts, Ph.D.

Head of Research – Nucleic Acid Technologies, Promega Corporation

June 5th, 2014

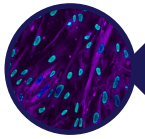
ATCC™ | THE ESSENTIALS OF LIFE SCIENCE RESEARCH GLOBALLY DELIVERED™

ATCC

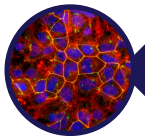
- Founded in 1925, ATCC is a non-profit organization with headquarters in Manassas, VA
- ATCC serves and supports the scientific community with industry-standard products and innovative solutions
- World's leading biological resource center and provider of biological standards
- Broad range of biological materials
  - Cell lines
  - Microorganisms
  - Native & synthetic nucleic acids
  - Reagents

ATCC®

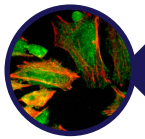
Outline



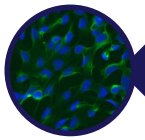
History of misidentified cell lines



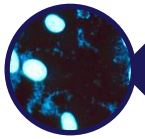
Receiving a cell line into the laboratory



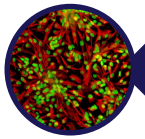
Creating a master cell bank – The seed stock concept



Steps for reducing cellular and microbial contamination



Tests for microbial infection



Species confirmation

ATCC®

Definition of “Cell Line Authentication”

Process by which cell line identity is confirmed

ATCC®

1952: HeLa – first immortalized human cell line

Henrietta Lacks

- Died at 31 years of age
- Aggressive grandular cancer (adenocarcinoma of the cervix)

George Gey

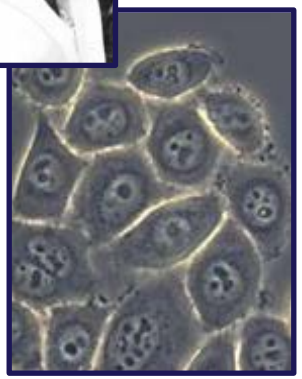
- 1952 HeLa cell line established by Mary Kubrick in George Gey's Laboratory

HeLa cells

- First human immortalized, continuous cell line to be developed
- HeLa (ATCC® CCL-2™) – 2nd cell line added to the ATCC Cell Biology Collection

Henrietta Lacks

George Gey



HeLa cells

ATCC

1960s: Poor culture conditions led to contamination

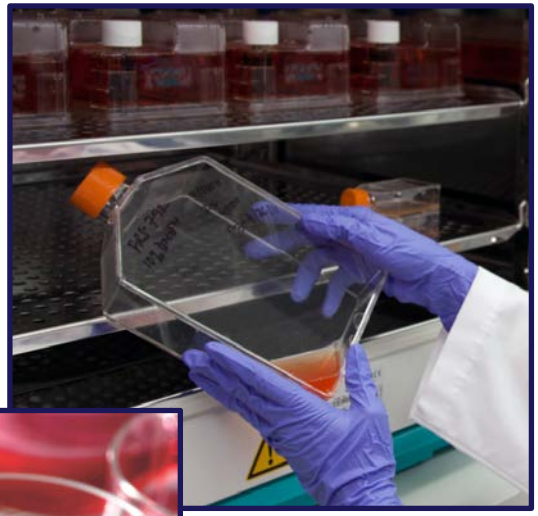
Poor tissue culture environment

- No disposable, plastic culture dishes
- No commercial media
  - Beef embryo extracts
  - Human cord blood
  - Chick plasma
- Cells grown on bench-top
- Bunsen burners and steam used for sterilization
- Technicians wore surgical masks, coats, gloves, booties, hair covers

Preparation of media (ATCC c.1978)

Better tissue culture conditions – contamination persists…

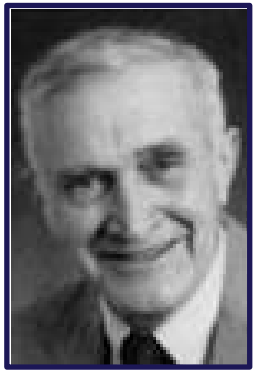
- The widespread used of HeLa resulted in cross-contamination and mix-up of cell lines.
- Recent data suggest that cross-contamination is still a major ongoing problem with modern cell cultures.



ATCC®

1962: Stanley Gartler describes HeLa contamination of cell lines

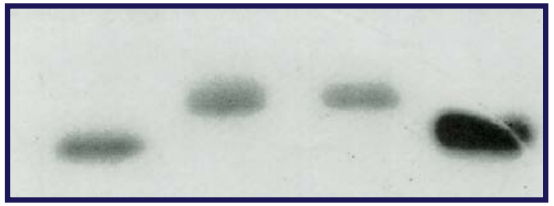
Name	Description	ATCC catalog no.	Origin	G6PD variant
HeLa	Cervical adenocarcinoma, human	ATCC® CCL-2™	African	Type A (fast)
KB	Oral epidermoid carcinoma, human	ATCC® CCL-17™	Caucasian	Type A (fast)
HEp-2	Larynx epidermoid carcinoma, human	ATCC® CCL-23™	Caucasian	Type A (fast)
Chang liver	Liver, human	ATCC® CCL-13™	Caucasian	Type A (fast)
Int-407	Embryonic intestine, human	ATCC® CCL-6™	Caucasian	Type A (fast)



Dr. Stanley Gartler

Isoenzyme Analysis

Glucose-6-phosphate dehydrogenase (G6PD)



Conclusion: 90% (18/20) human cell lines are 'HeLa'

ATCC®

Gartler SM, NCI Monogr. 26:176, 1967; Gartler, SM, Nature 217:750, 1968

1970s: Walter Nelson-Rees describes interspecies cross-contamination

Actual (43/466 (9.2%))	Purported (62 Laboratories)
Dog	Horse, Human, Mink, Mouse
Hamster	Mouse, Human, Marmoset, Rat
Mongoose	Human
Human	Gibbon
Mink	Human
Monkey	Horse, Human
Mouse	Human
Rabbit	Dog
Rat	Chicken, Human, Mink, Monkey

Dr. Walter Nelson-Rees

Cytogenetic Analysis – Karyotyping found HeLa markers in many human cell lines

ATCC

Nelson-Rees, WA, et al. Science 212,446, 1981

1984-2003: interspecies and intraspecies misidentification of cell lines

Cellular cross-contamination					
Year	**No.**	**%**	**Type of contamination**	**Technology**	**Reference**
1984	275	35%	Interspecies	Karyotyping	Hukku, B. et al. Eukaryotic cell culture. Plenum Press, 1984
1999	252	18%	Intraspecies	STR profiling	Drexler, HG et al. Leukemia 13:1999.
2003	550	15%	Intraspecies	STR profiling	Drexler, HG et al. Leukemia 17:2003

"Less than 50% of researchers regularly verify the identities of their cell lines using standard methods such as DNA fingerprinting by STR analysis"

ATCC®

Buehring, G.C., et al. (2004) In Vitro Cell Dev Biol 40:211

2002-2013: Misidentification of cell lines persists…

Year	Title of article	Reference
2004	LCC15-MB cells are MDA-MB-435: a review of misidentified **breast and prostate** cell lines.	Clin Exp Metastasis. 21(6):535, 2004.
2007	MDA-MB-435: The Questionable Use of a Melanoma Cell Line as a Model for **Human Breast** Cancer is Ongoing	Cancer Biology & Therapy 6:9, 1355, 2007.
2008	Deoxyribonucleic Acid Profiling Analysis of 40 Human **Thyroid Cancer** Cell Lines Reveals Cross-Contamination Resulting in Cell Line Redundancy and Misidentification.	J Clin Endocrinol Metab. 93(11):4331, 2008.
2009	Genetic Profiling Reveals Cross-Contamination and Misidentification of **6 Adenoid Cystic Carcinoma** Cell Lines: ACC2, ACC3, ACCM, ACCNS, ACCS and CAC2.	PLoS one. 4(6):e6040, 2009
2010	Verification and Unmasking of Widely Used Human **Esophageal Adenocarcinoma** Cell Lines.	JNCI. 102(4):271, 2010
2013	Misidentification of Putative Medullary **Thyroid Cancer** Cell Lines RO-H85-1 and RO-D81-1	J Clin Endocrinol Metab, 98(3):954, 2013

ATCC®

Impact of misidentified cell lines on applied research

Misidentification of frequently used esophageal adenocarcinoma cell lines			
Cell Line Name	**Purported**	**STR confirmed (ATCC STRProfile database)**	
SEG-1	Esophageal adenocarcinoma cell line	H460 (ATCC® HTB-177™)	Lung carcinoma (large cell lung cancer)
BIC-1	Esophageal adenocarcinoma cell line	SW620 (ATCC® CCL-227™)	Colorectal adenocarcinoma
SK-GT-5	Esophageal adenocarcinoma cell line	SK-GT-2	Gastric fundus carcinoma

Experimental results based on contaminated cell lines …

- Clinical trail recruiting EAC patients
- 100 scientific publications
- At least 3 NIH cancer research grants
- 11 US patents

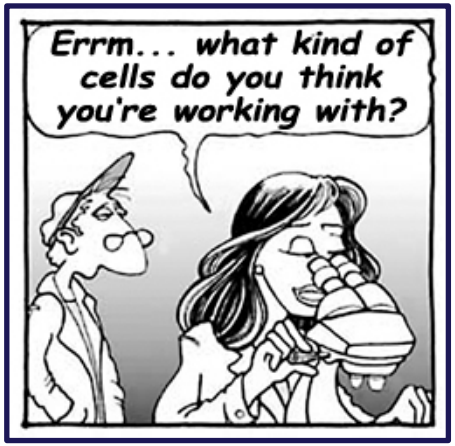
ATCC®

Boonstra, J.J., et al. (2010) JNCI.102(4):271

Are your cells REALLY what you think they are?

Common sources of cellular contamination

- Getting cell lines from a colleague down the hall
- Continuous culturing of working cell banks
- Use of feeder cells
- Mislabeling of culture flasks
- Working with multiple cell lines, concurrently
- Using one reservoir of growth medium for multiple cell lines
- CO_2 Incubator



ATCC®

Consequences of using misidentified cell lines

- Loss of cell line
- Loss of time and money
- Misinformation in the public domain
- Discordant or irreproducible results
- Private embarrassment /public humiliation

"For decades, biologists working with contaminated or misidentified cell lines have wasted time and money and produced spurious results; journals and funding agencies say it's not their job to solve this problem"

ATCC®

Rhitu Chatterjee. Cases of Mistaken Identity (2007) Science 15:928

Receiving a new cell line into the laboratory

Record history of cell line

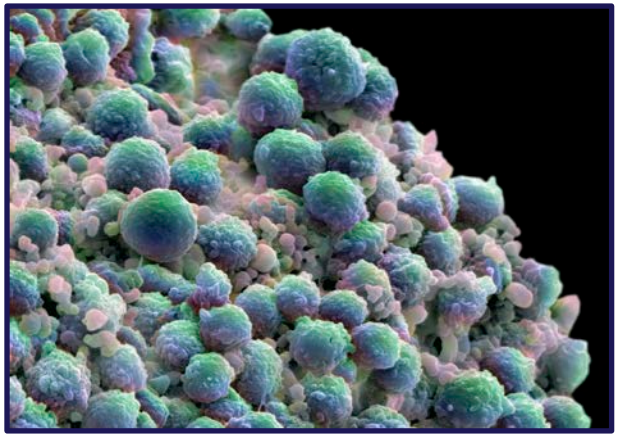
- Originator
- Institution/laboratory
- Date of origin
- Publication on deriving the cell line

ATCC®

Receiving a new cell line into the laboratory

Record all background information pertaining to cell line

- Cell line name
- Tissue of origin
- Species, strain
- Passage number
- Population doubling level
- Unique characteristics
- Unique function

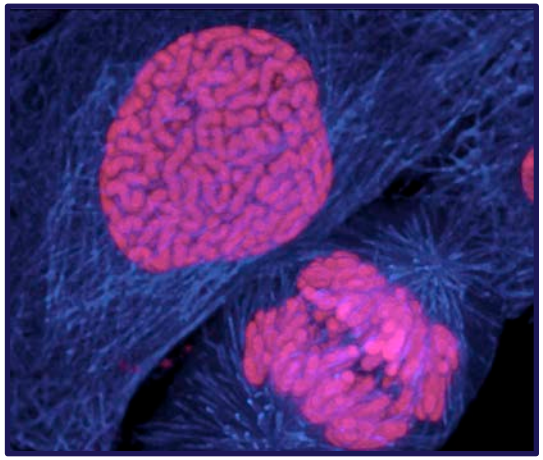


Prostate cancer cells

Receiving a new cell line into the laboratory

Record information on how to grow cell line

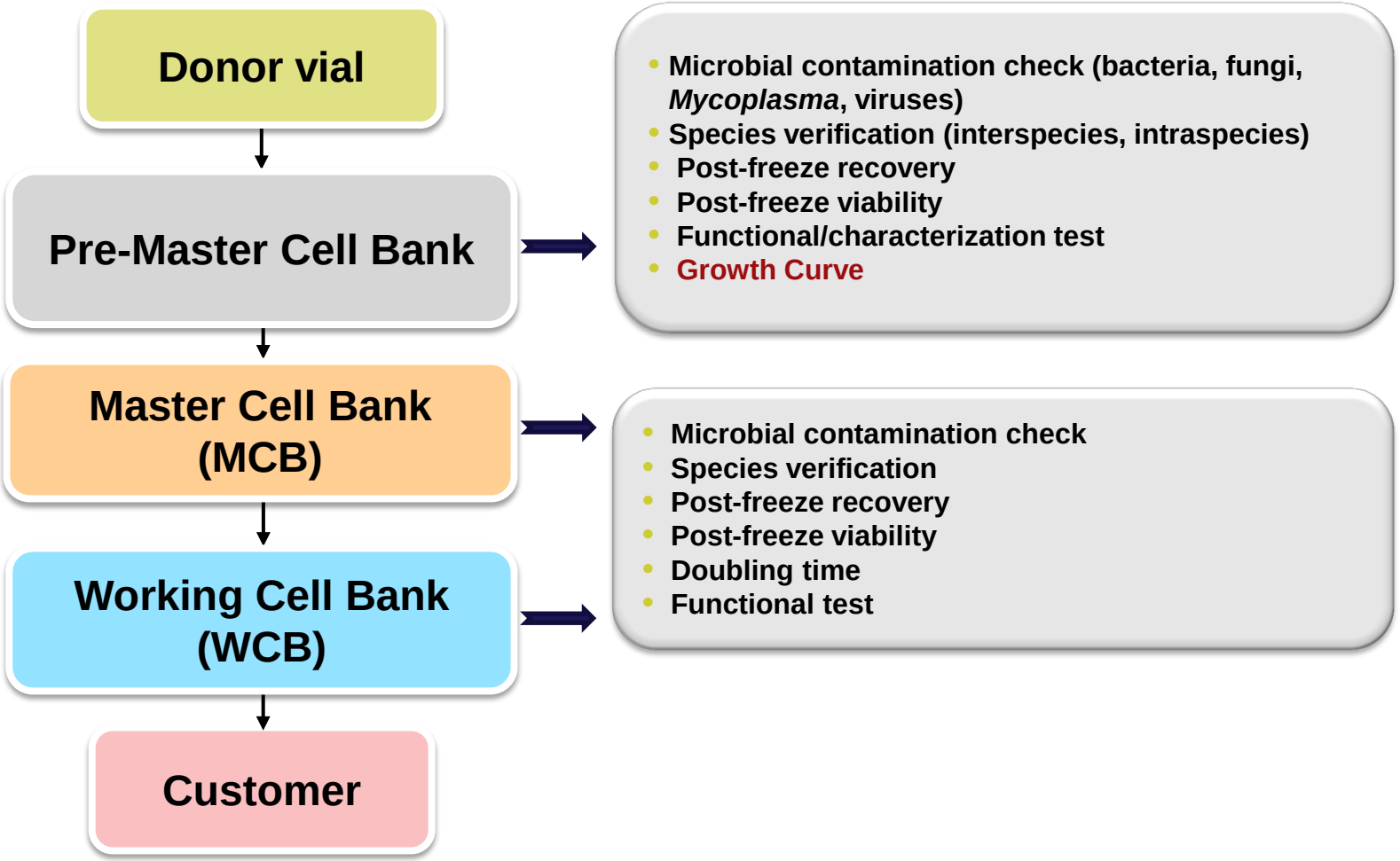
- Complete growth medium (include additives)
- Type of serum (to include source)
- Source of additives, concentration
- Procedure for thawing cells
- Procedure for subculturing
- Cryopreservation medium and procedure
- Doubling time
- Expected pre-freeze, post-freeze viability
- Storage temperature



Human breast cancer cells

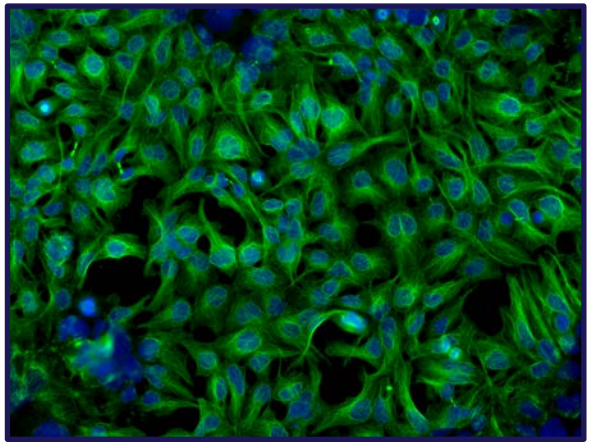
ATCC®

Accessioning scheme: ‘Seed Stock’ concept



Advantages of the ‘Seed Stock’ concept

- Ensures safe stocks of early material (Master Cell Bank)
- Reduces passage number (prevents genotypic and phenotypic drift)
- Prevents lot-to-lot variability
- Reduces the chance of cellular and microbial contamination
- Ensures reproducibility of experimental data

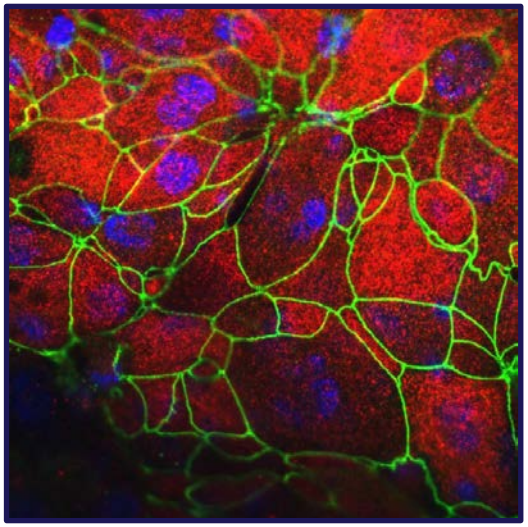


HEK-293 cells stained with beta tubulin

ATCC®

Steps for reducing cellular/microbial contamination

- Obtain cell line from a reputable source
- Good documentation
- Highly trained technicians
- Good aseptic techniques
- Use one reservoir of medium per cell line
- Aliquot stock solutions/reagents
- Monitor for cell line identity and characteristics contamination, routinely
- Use seed stock (create master stocks)
- Keep cell line passage number low

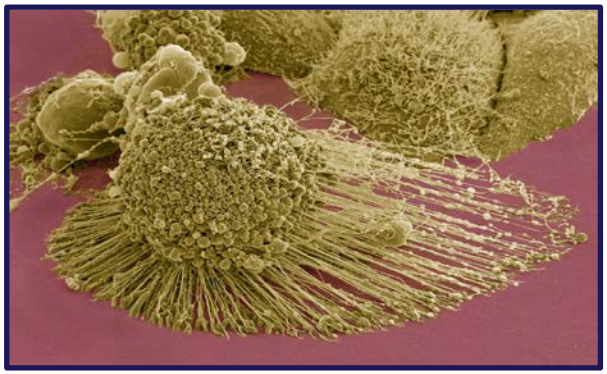


ATCC® HTB-174™, NCI-H441, human papillary adenocarcinoma

ATCC®

Steps for reducing cellular/microbial contamination

- Label flasks - name of cell line, passage number, date of transfer
- Use barcoded flasks when available
- Work with one cell line at a time in biological safety cabinet
- Clean biological safety cabinet between each cell line
- Allow a minimum of 5 minutes between each cell line
- Quarantine “dirty” cell lines from “clean” cell lines
- Manageable work load (reduce accidents)
- Clean laboratory (reduce bioburden)
- Legible handwriting (printed labels)



ATCC® CCL-2™; HeLa, cervical carcinoma

ATCC®

Test for microbial contamination

Bacteria and Fungi

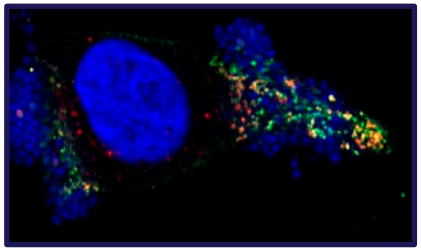
- Microbiological culture (Aerobic, Anaerobic)
- PCR

Mycoplasma

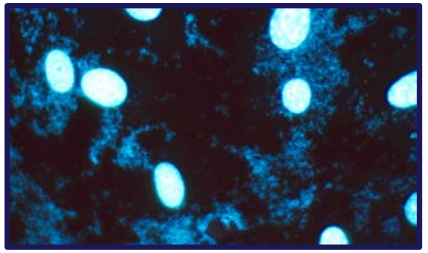
- Direct agar culture
- Indirect Hoechst stain
- PCR

Viruses

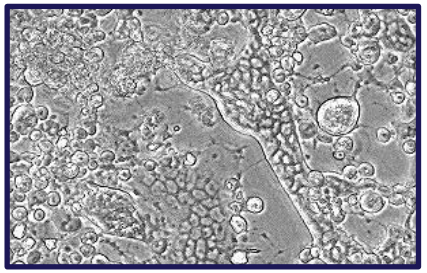
- Cytopathic effect (CPE)
- Indirect immunofluorescent antibody (IFA)
- Enzyme immunoassay (EIA)
- PCR



HeLa contaminated with *E. coli*



M. hyorhinis infection



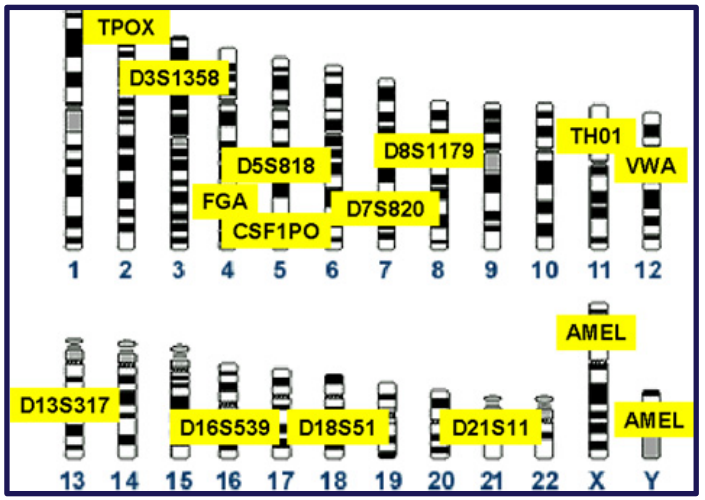
CPE of human herpes virus

ATCC®

Confirm species and identity of cell line

- Cytochrome C Oxidase subunit 1 (COI) for interspecies identification
- STR analysis (DNA profiling) for intraspecies identification and authentication of human cell lines

Barcode of Life
Barcoding for species Identification

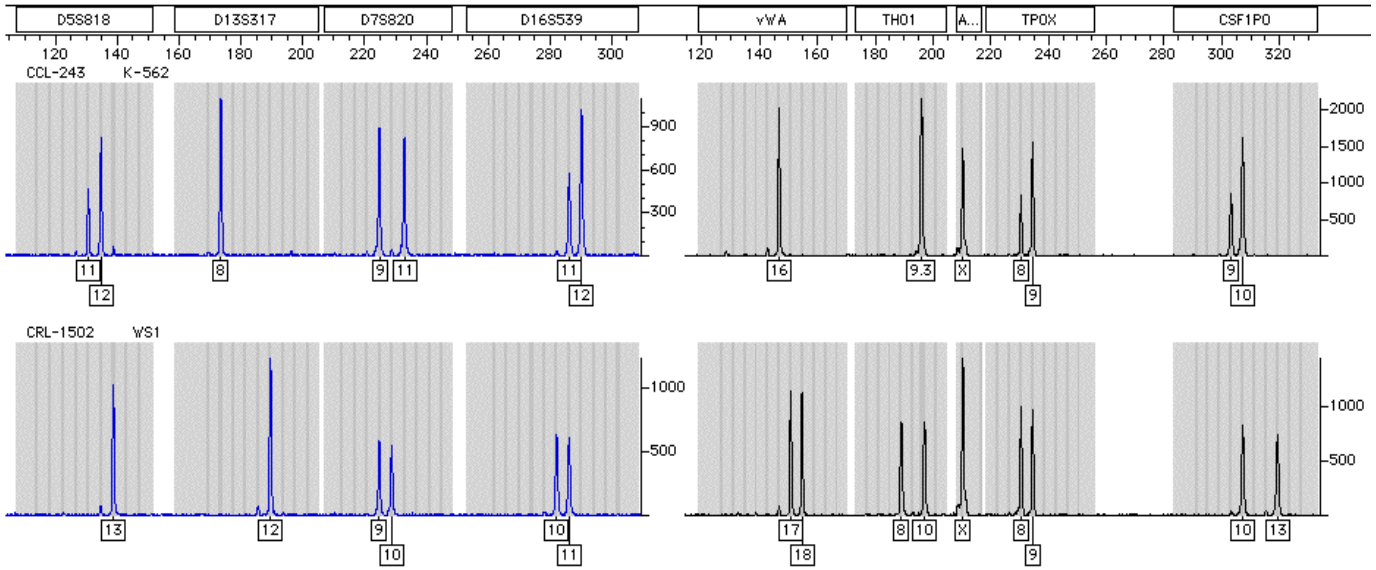


ATCC

Authenticate your human cell line by STR Analysis!!

K562

WS1



Electropherogram of 2 unrelated (unique) cell lines

THANK YOU

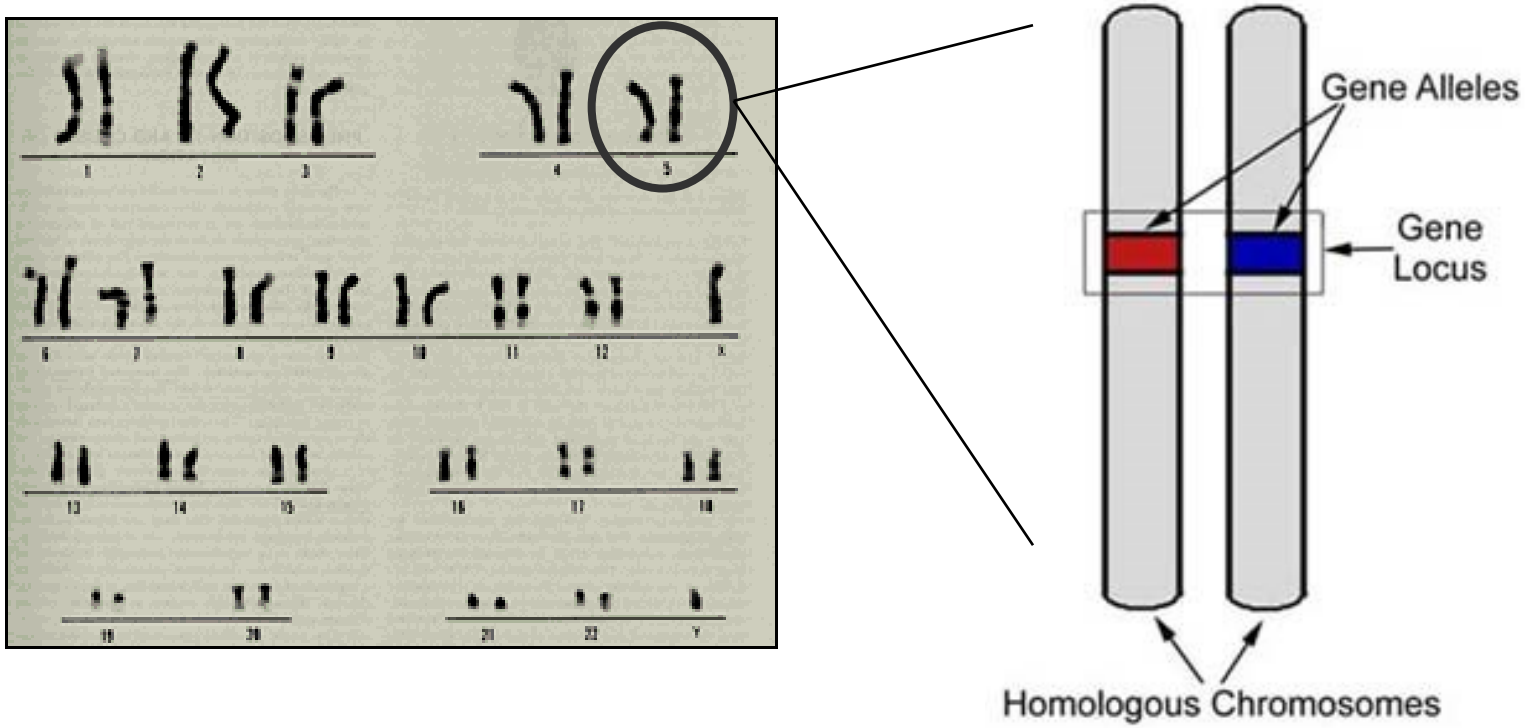
ATCC™ | THE ESSENTIALS OF LIFE SCIENCE RESEARCH GLOBALLY DELIVERED™

Overview

Short Tandem Repeat Technology

Promega Corporation

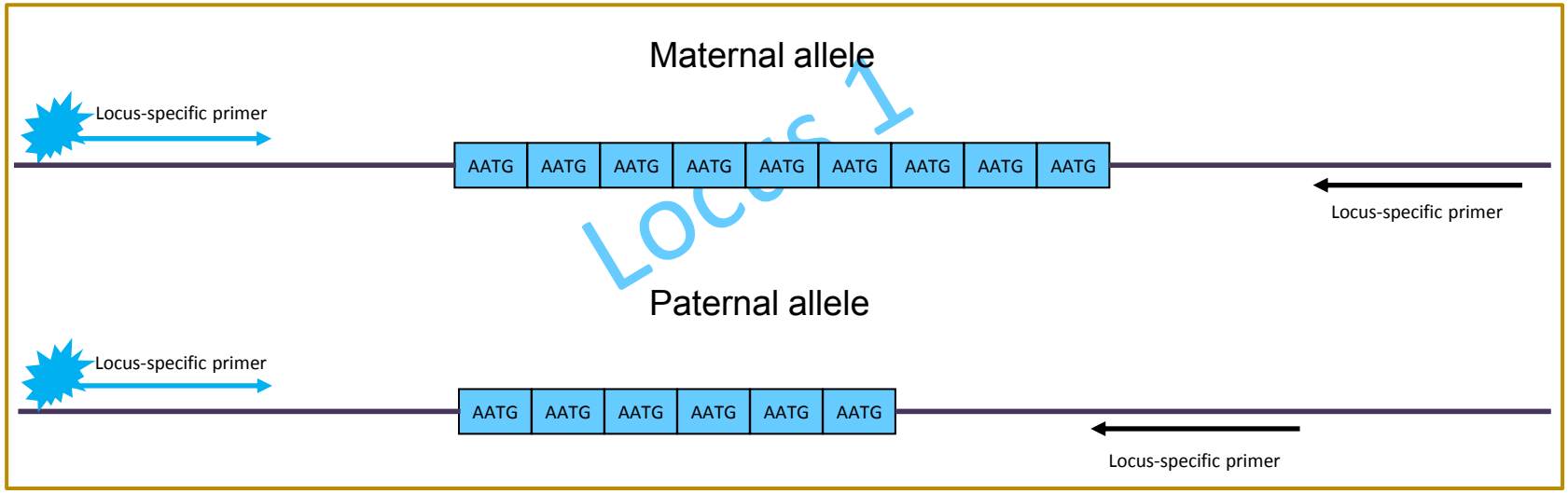
Gene Locus vs. Gene Allele



- **LOCUS** identifies the specific physical location of a gene or STR on a chromosome. Both chromosomes of a homologous pair contain this locus.
- **ALLELE** for the gene or STR contained at that locus may be the same on both chromosomes (**homozygous**), or different on each chromosome (**heterozygous**)

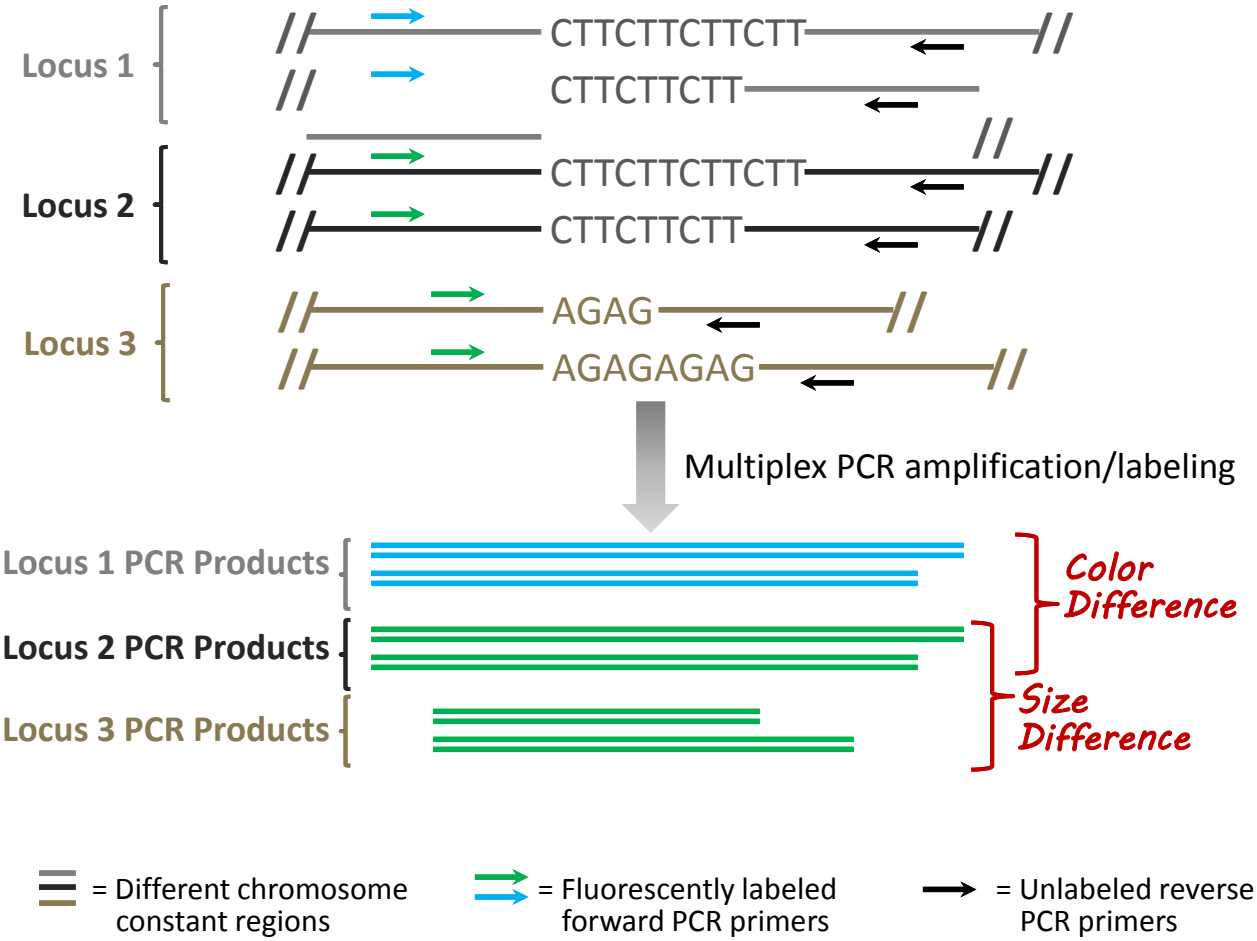
What are STRs?

- Short Tandem Repeats (STR) commonly used for identity testing
  - Trinucleotide: AAT AAT AAT AAT
  - Tetranucleotide: AATG AATG AATG AATG
  - Pentanucleotide: AAAGA AAAGA AAAGA AAAGA
- Multiplex of loci discriminated by fluorescent label and size range



Promega STR kits target polymorphic loci in non-coding regions

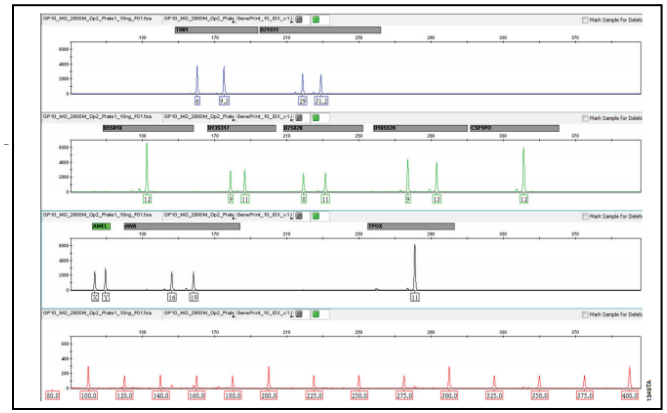
Multiplexing by Size and Color

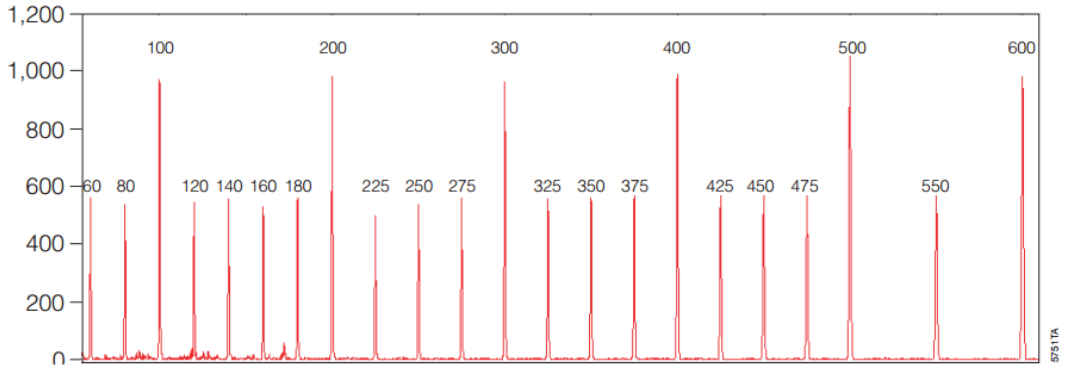


Detection of STRs

- Amplicons separated and detected on capillary electrophoresis (CE) instruments
  - Applied Biosystems Genetic Analyzers (3130, 3500, etc.)

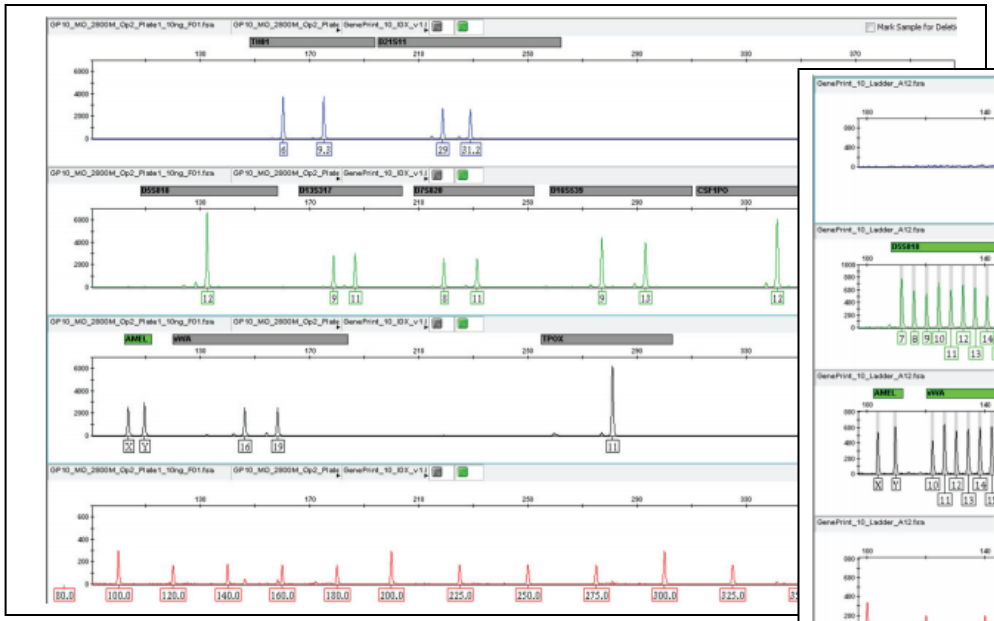
- Fragments are “sized” to allow comparison between samples
  - Use an Internal Lane Standard (ILS) to



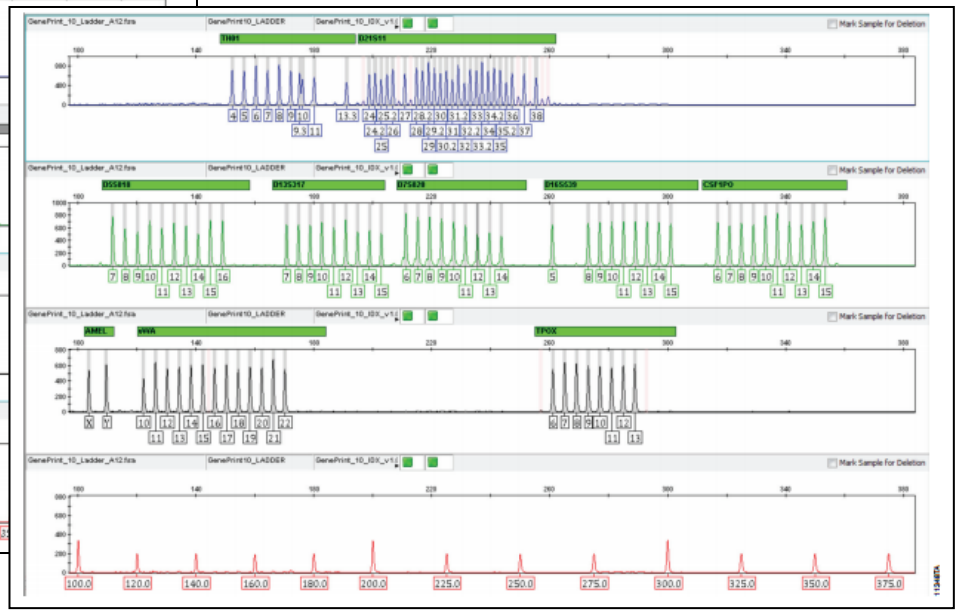


Assigning Allele Types to Peaks

- Samples are compared to an allelic ladder to determine actual alleles (everything compared to Internal Lane Standard)
  - Nomenclature is based on number of repeats

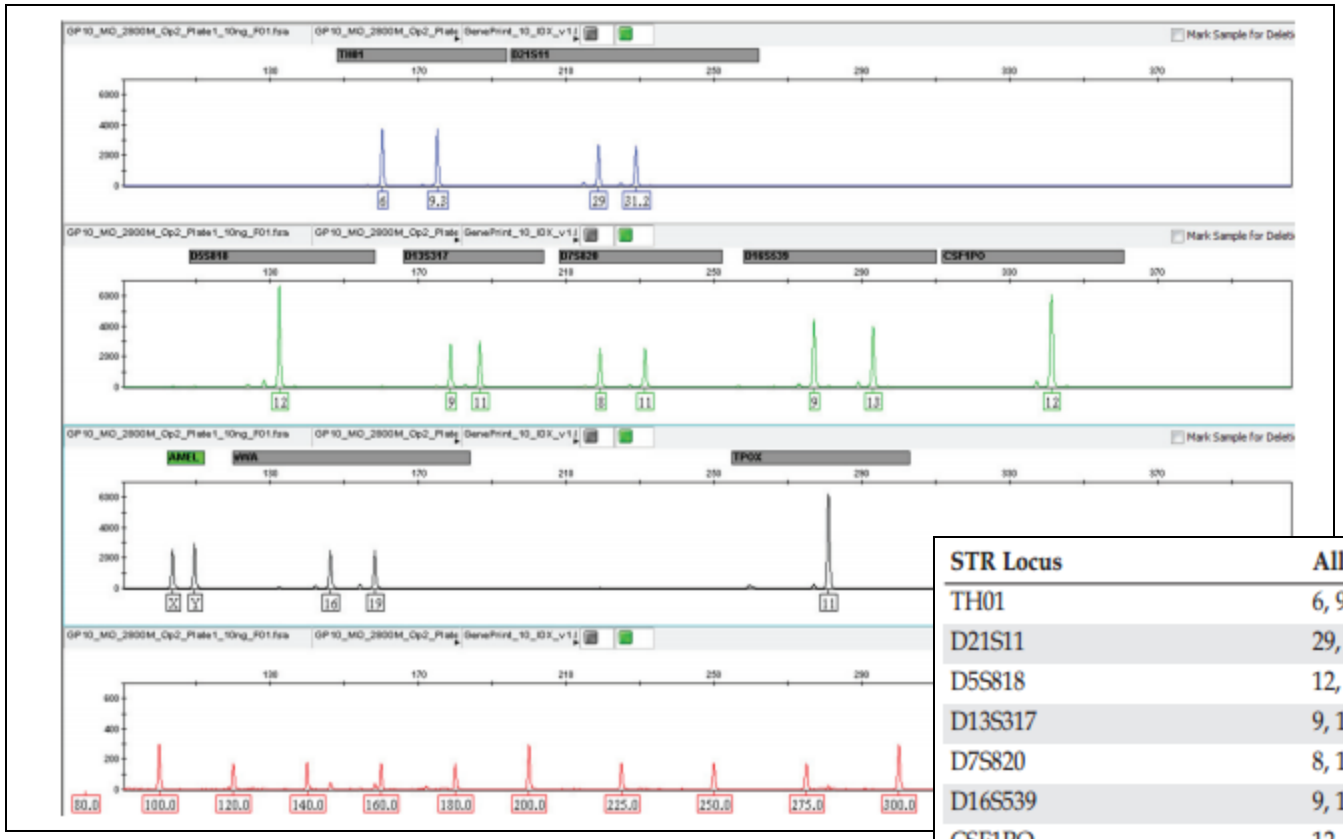


Sample



Allelic Ladder (run in separate capillary)

Output



Simple text-based table

STR Locus	Alleles
TH01	6, 9.3
D21S11	29, 31.2
D5S818	12, 12
D13S317	9, 11
D7S820	8, 11
D16S539	9, 13
CSF1PO	12, 12
Amelogenin	X, Y
vWA	16, 19
TPOX	11, 11

PowerPlex® 18D System

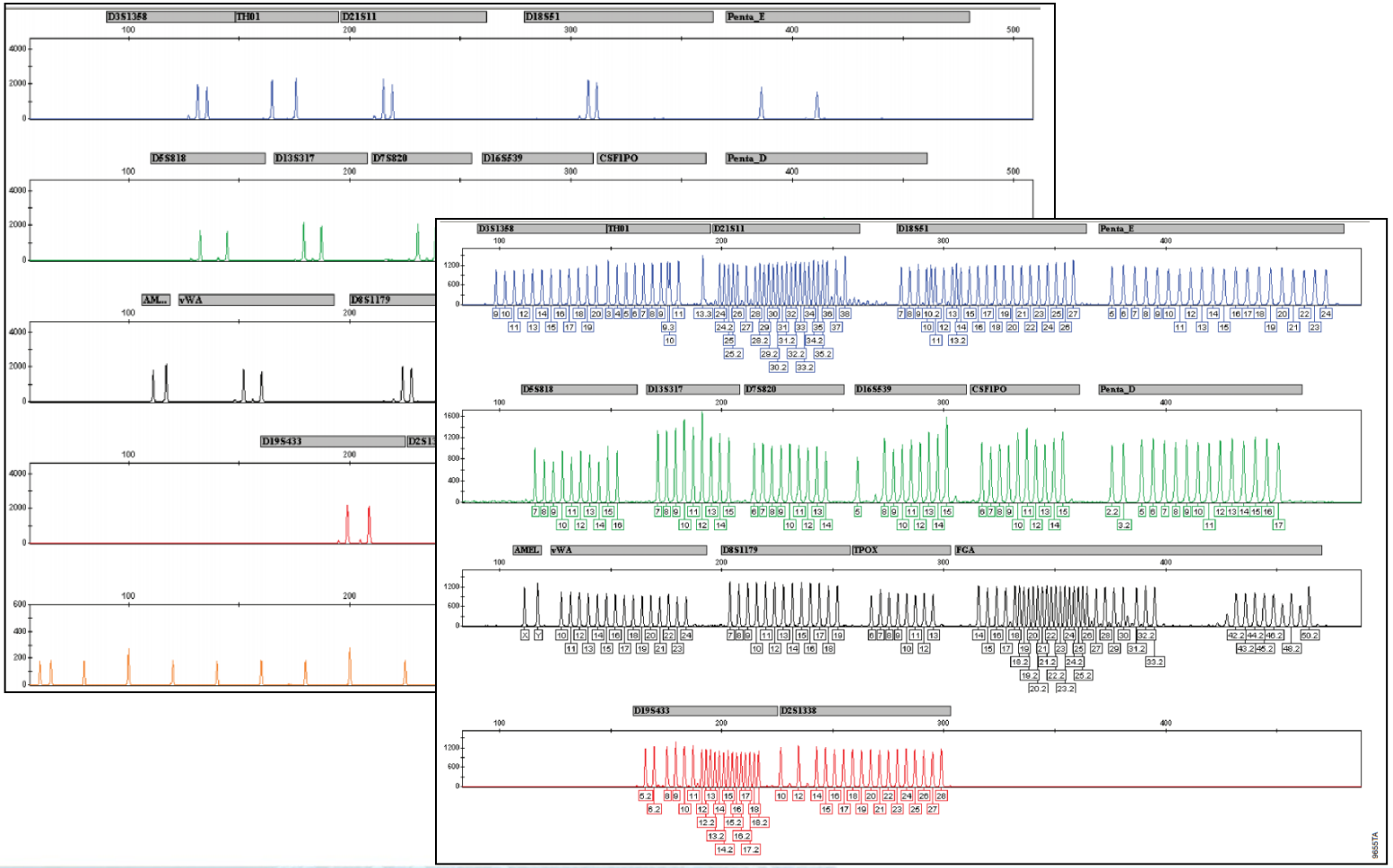
FAM

JOE

TMR-ET

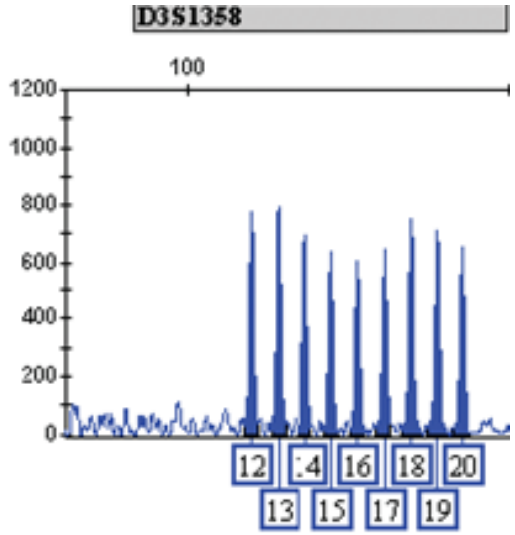
CXR-ET

ILS500
(CC5)



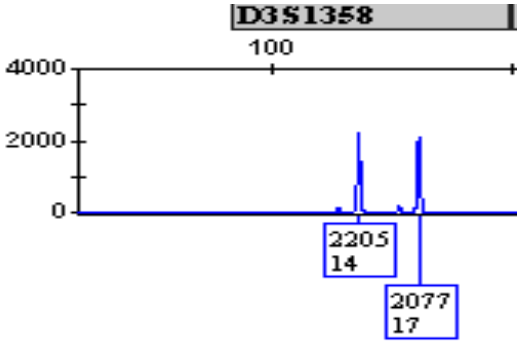
Potential Versus Actual Alleles at First Locus

Alleles commonly encountered in the human population at this location

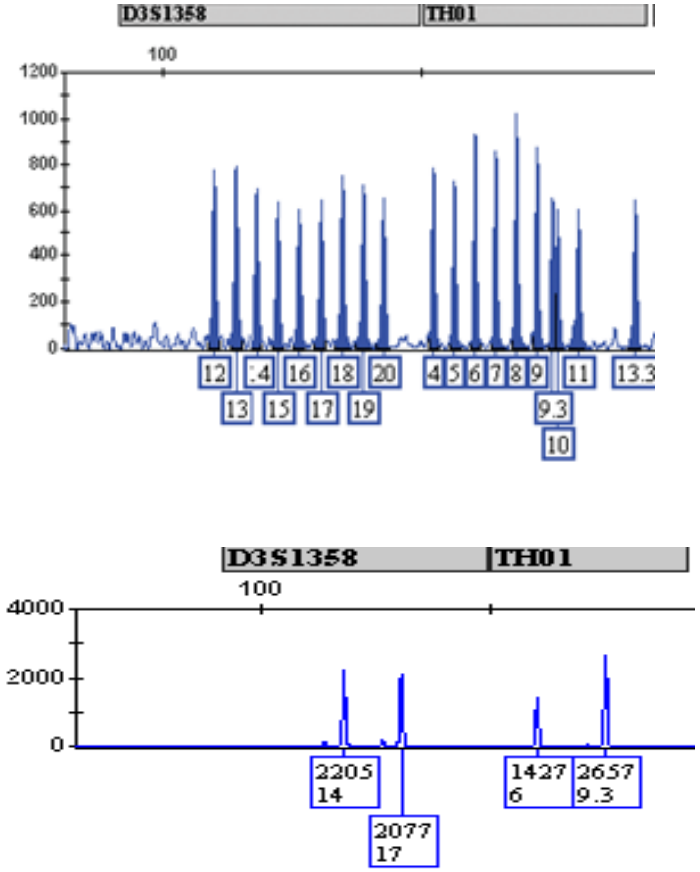


1 of 9 from mom
+ 1 of 9 from dad
1 of 18 possibilities

Profile generated from an individual



…using Two Loci



1 of 9 from mom

+ 1 of 9 from dad

1 of 18 possibilities

at first location (D3S1358)

x

1 of 10 from mom

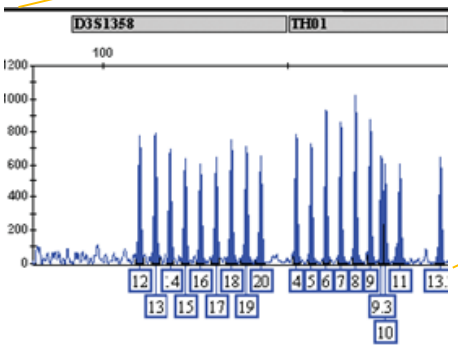
+ 1 of 10 from dad

1 of 20 possibilities

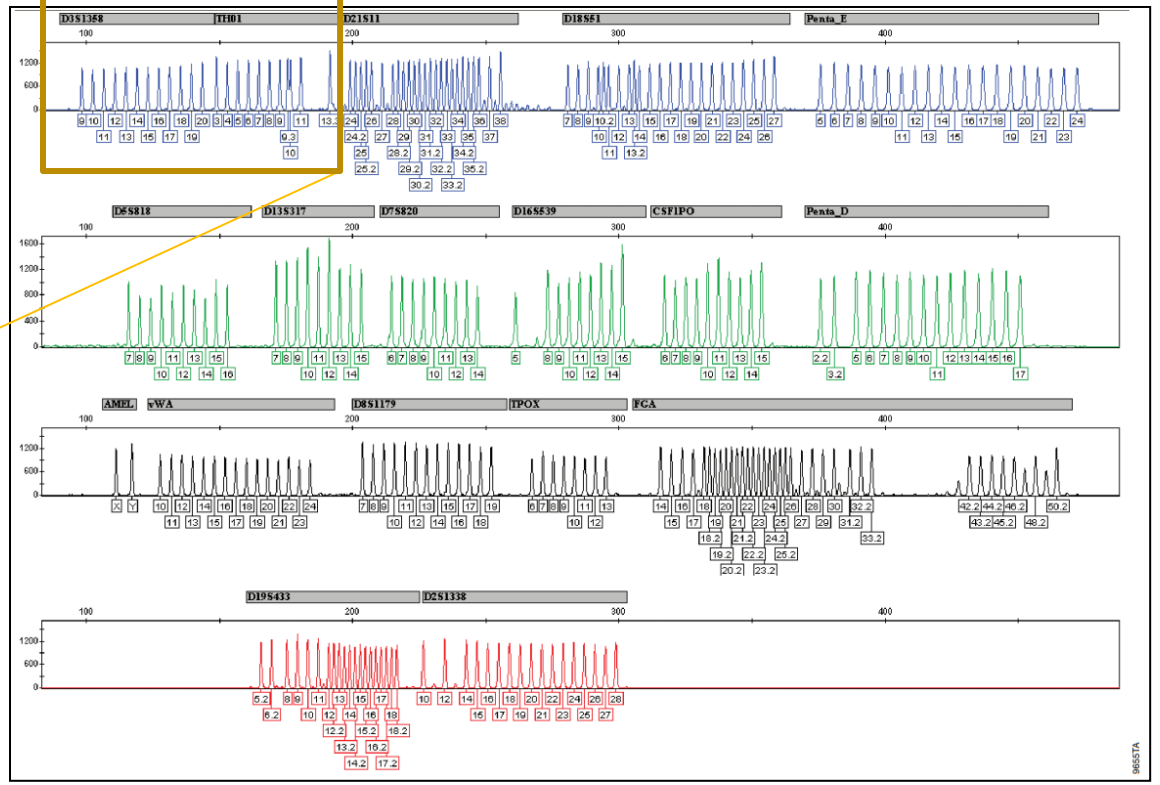
at second location (TH01)

=1:18 x 1:20 = 1:360

…using 18 Loci



1:18 x 1:20



Matching probability: 1 in 18 x 1 in 20 x 1 in 50… ≅ 1 in 10^{22}

10^{22} = 10,000,000,000,000,000,000,000

STR Analysis Workflow

DNA Sample

Multiplex PCR

- Amplification of select STR loci
- Simultaneous fluorescent labeling

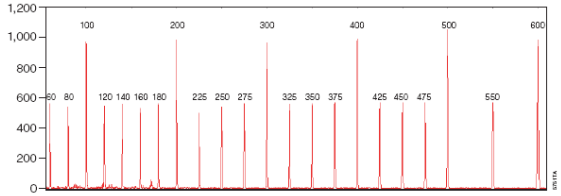
Capillary Electrophoresis (CE)

- Add Internal Lane Standard
- CE to size separate
- Fluorescent detection
- Also run Allelic Ladder in parallel

Data Analysis

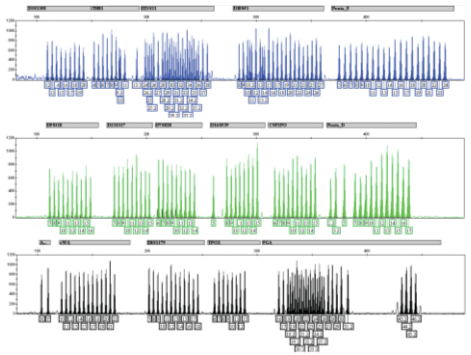
- Calculate sizes based on ILS
- Compare fragment sizes to allelic ladder to determine STR alleles present in sample
- Compare to databases

Internal Lane Standard



Labeled size standards in a different color than STR fragments

Allelic Ladder



Labeled fragments of all possible alleles for each STR locus

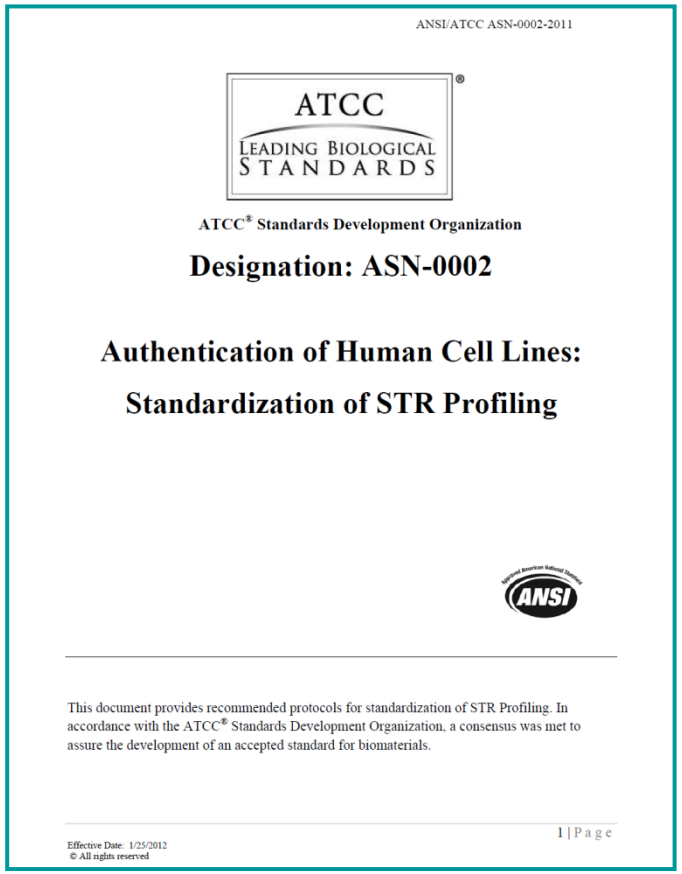
Strengths/Limitations

- STR systems can determine…
  - Relatedness of a cell line to a reference standard
  - Cell line cross contamination
  - Gender (sex)
- STR systems cannot distinguish…
  - Different cell lines created from the same individual (donor)
  - Cell lines created from identical twins
- Promega STR systems are designed to be human-specific
  - Cannot be used to genotype non-human species

Standardization

ANSI Standard

- ASN-0002 consensus standard describing the use of short tandem repeat (STR) analysis for cell line authentication
  - 8 STR loci plus gender marker
- National Center for Biotechnology Information (NCBI) database containing genotypes of validated cell lines
- Many journals changing editorial policy to require cell line authentication



ANSI/ATCC ASN-0002-2011

ATCC
LEADING BIOLOGICAL STANDARDS

ATCC® Standards Development Organization

Designation: ASN-0002

Authentication of Human Cell Lines:

Standardization of STR Profiling

ANSI

This document provides recommended protocols for standardization of STR Profiling. In accordance with the ATCC® Standards Development Organization, a consensus was met to assure the development of an accepted standard for biomaterials.

1 | Page

Effective Date: 1/25/2012

Standard Loci

- STR Loci (minimum set)
  - TH01, TPOX, vWA, CSF1PO, D16S539, D7S820, D13S317, and D5S818
- Gender marker
  - Amelogenin

Matching Algorithm

- First, combine total number of alleles observed in the Test Sample and Reference (TOTAL ALLELES)
- Second, count the number of alleles shared by the Test Sample and the Reference Sample (SHARED ALLELES)
- Third, use the Match Algorithm to calculate a percent match result for the two samples:

$$\text{Percent Match} = \frac{\text{SHARED ALLELES} \times 2}{\text{TOTAL ALLELES}}$$

- Related samples generally yield a result in the 80-100% match range
  - A percent match <80% should be investigated

International Cell Line Authentication Committee
http://standards.atcc.org/kwspub/home/the_international_cell_line_authentication_committee-iclac_/

GenePrint® and PowerPlex® STR Systems

Partial listing of Promega STR systems

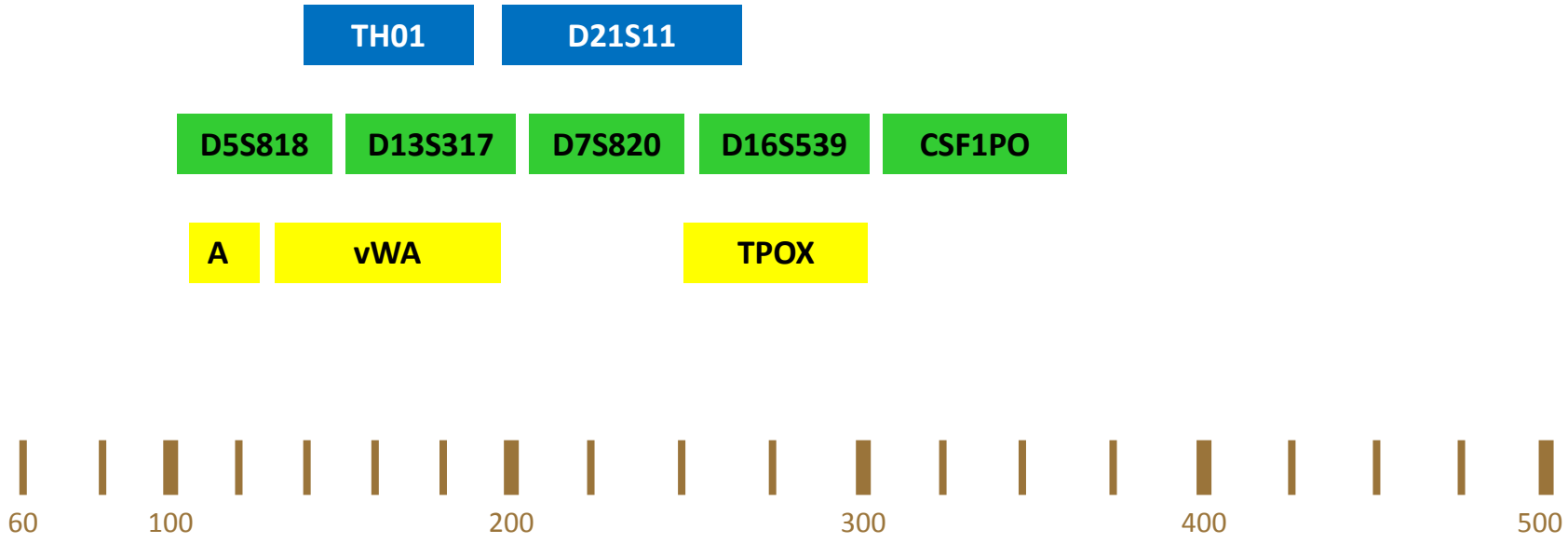
System	Loci	Matching Probability	Meets or Exceeds ANSI Standard
GenePrint® 10	10	1 in 3 x 10^{9}	√
PowerPlex® 16 HS	16	1 in 2.8 x 10^{19}	√
PowerPlex® 18D	18	1 in 3.5 x 10^{22}	√
PowerPlex® 21	21	1 in 6.7 x 10^{27}	√
PowerPlex® Fusion	24	1 in 1.4 x 10^{28}	√

Most systems compatible with direct amplification protocols

GenePrint® 10 System

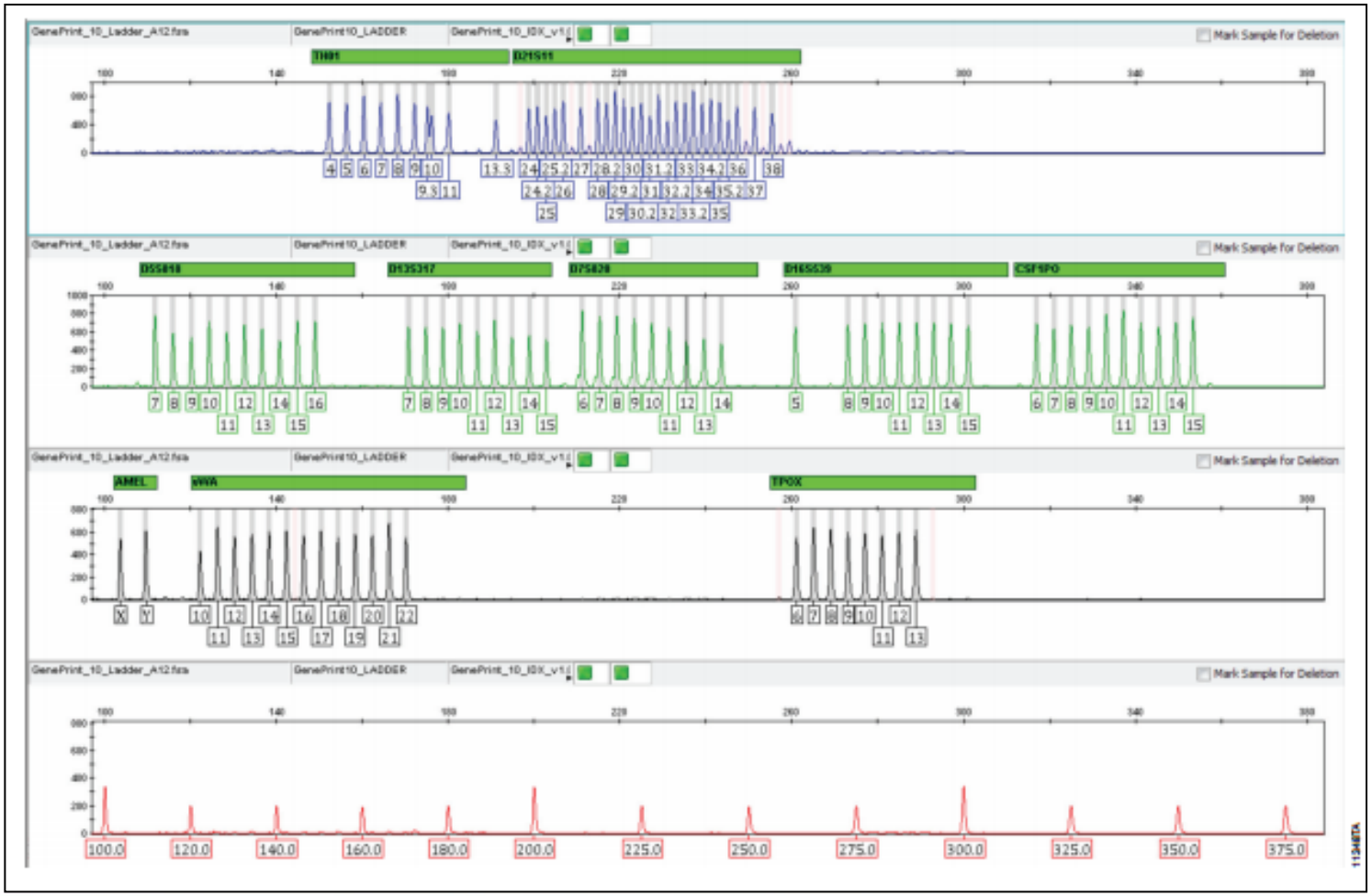
Promega Corporation

GenePrint® 10 System

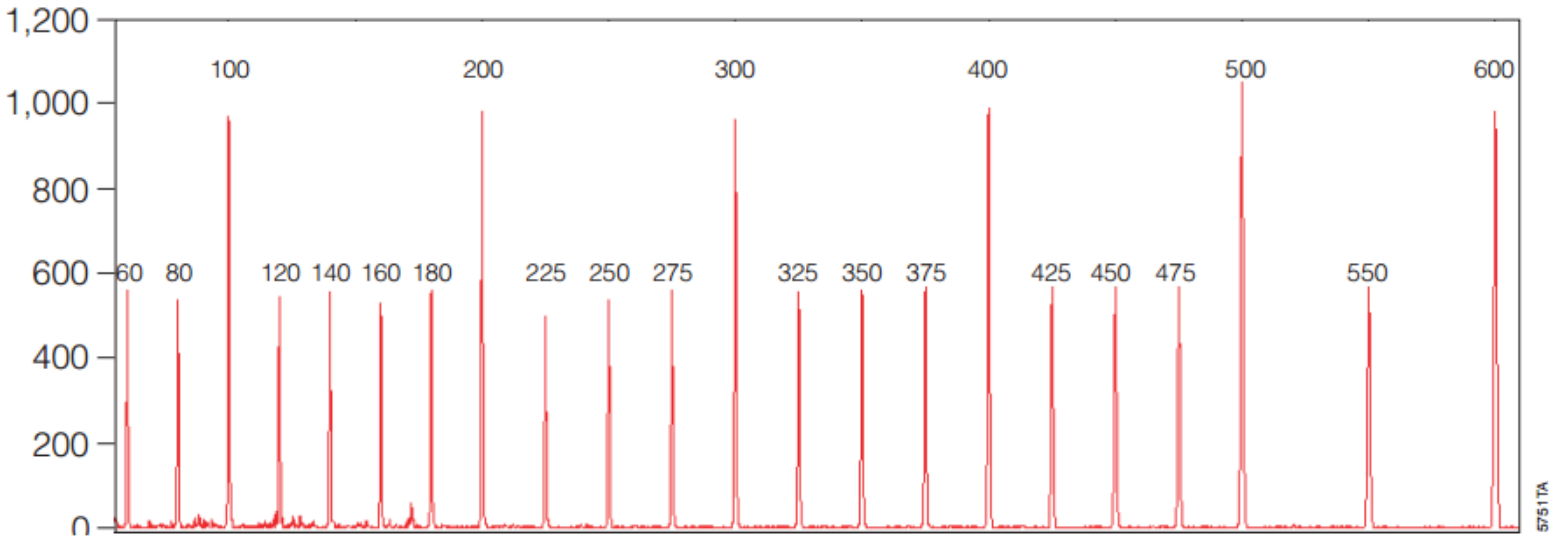


- Designed for cell line authentication
- Contains the loci recommended by the ASN-0002 consensus standard, plus D21S11

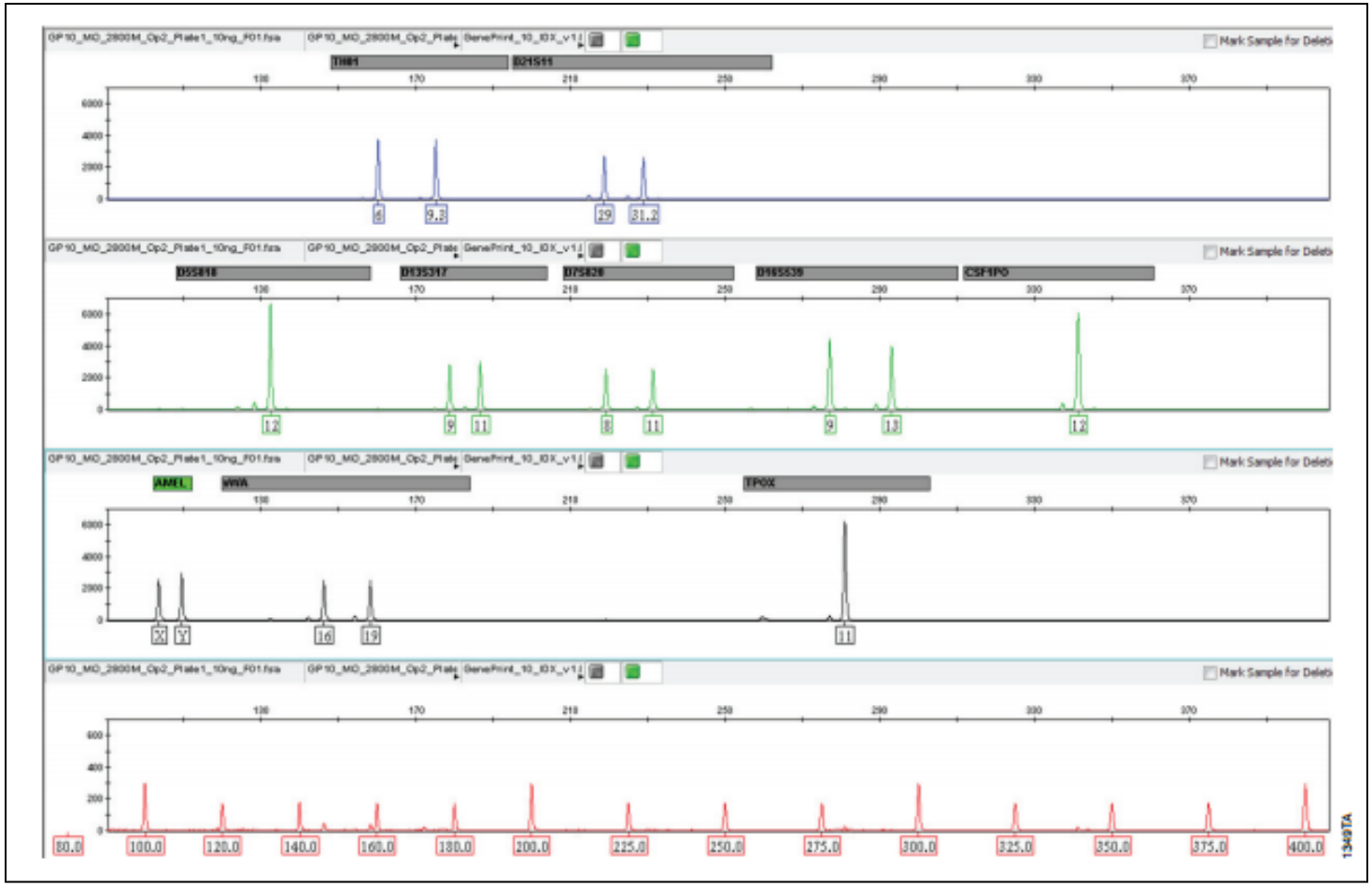
GenePrint® 10 System – Allelic Ladder



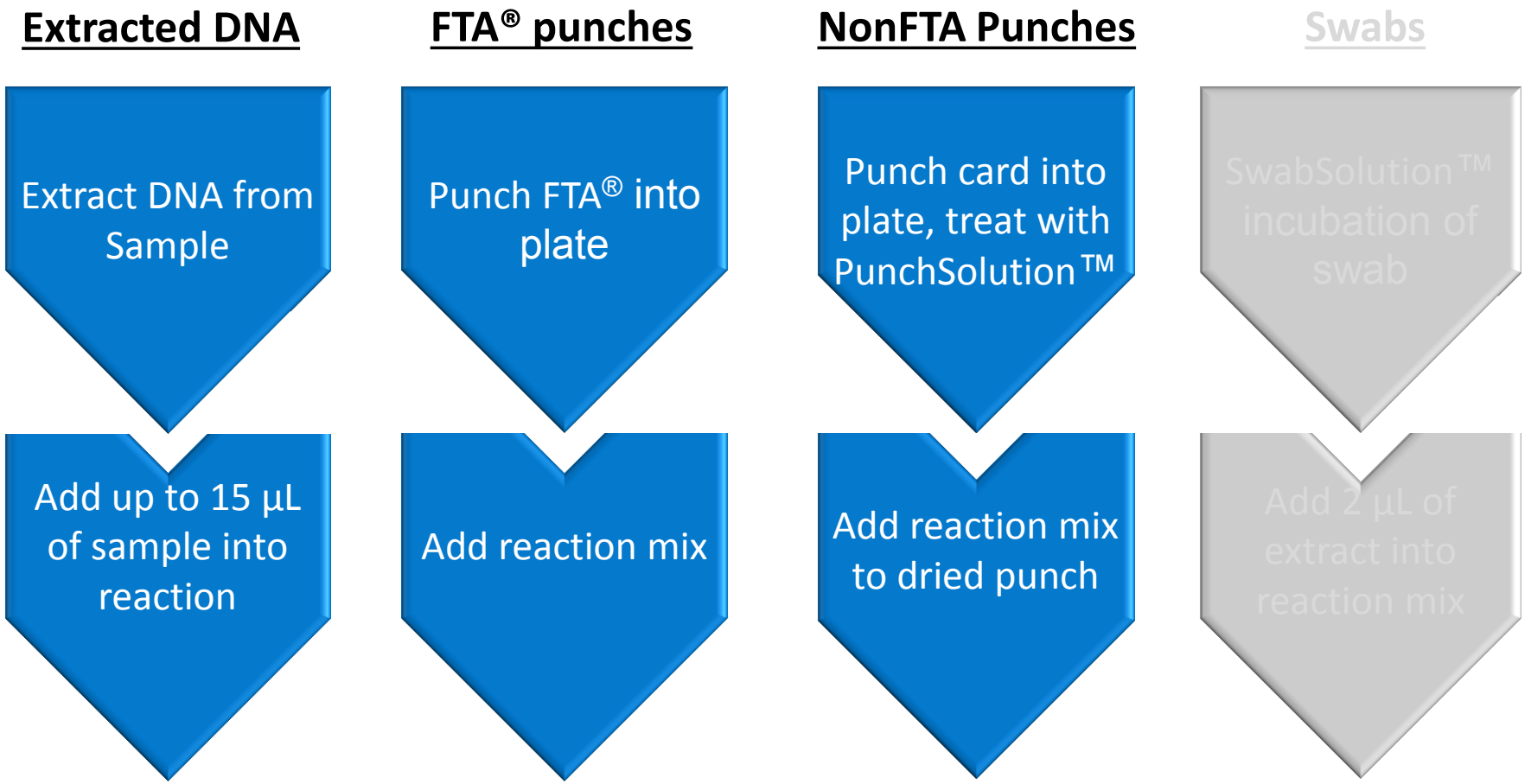
Internal Lane Standard 600



GenePrint® 10 Control DNA

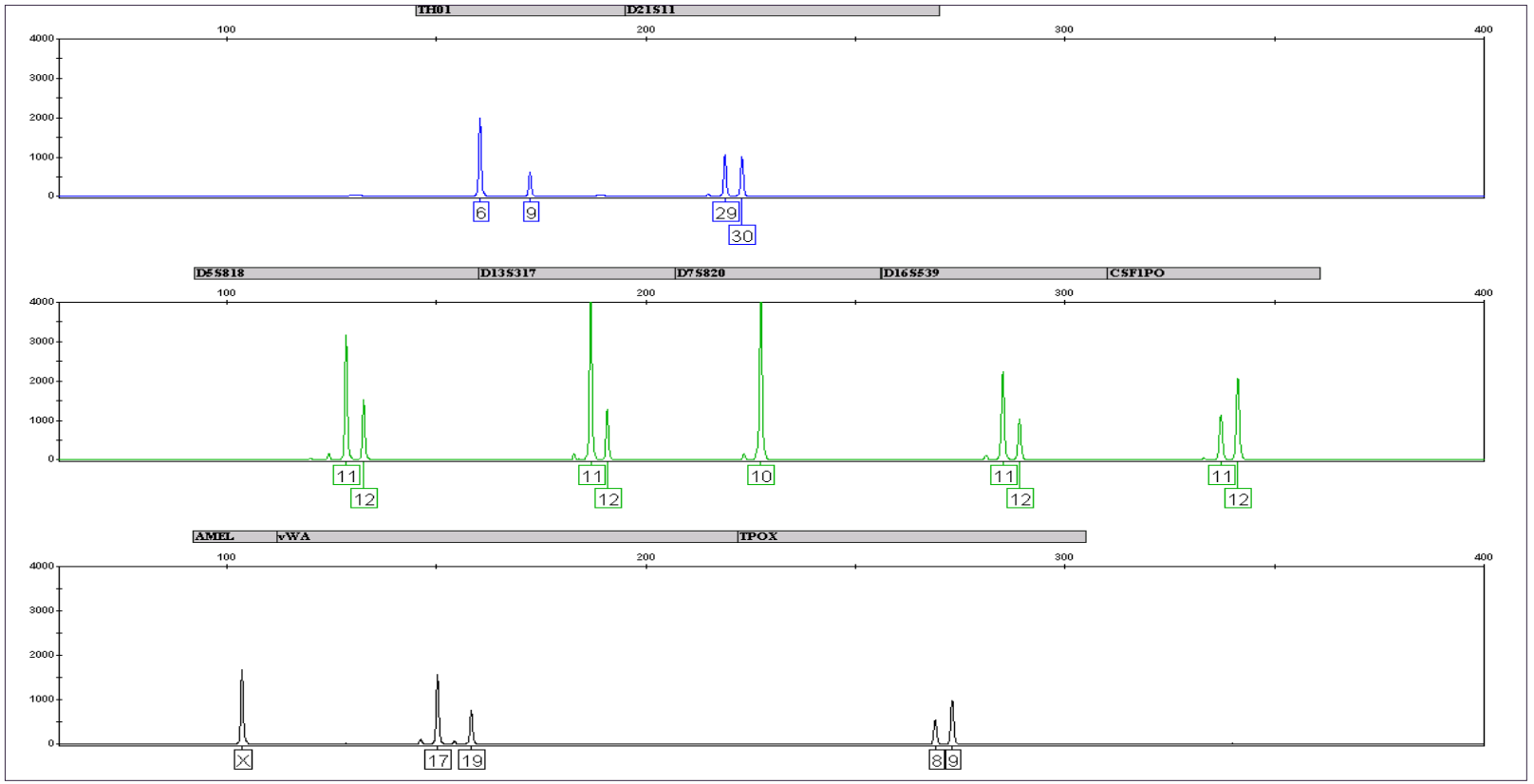


Variety of Sample Types
Simple Protocols



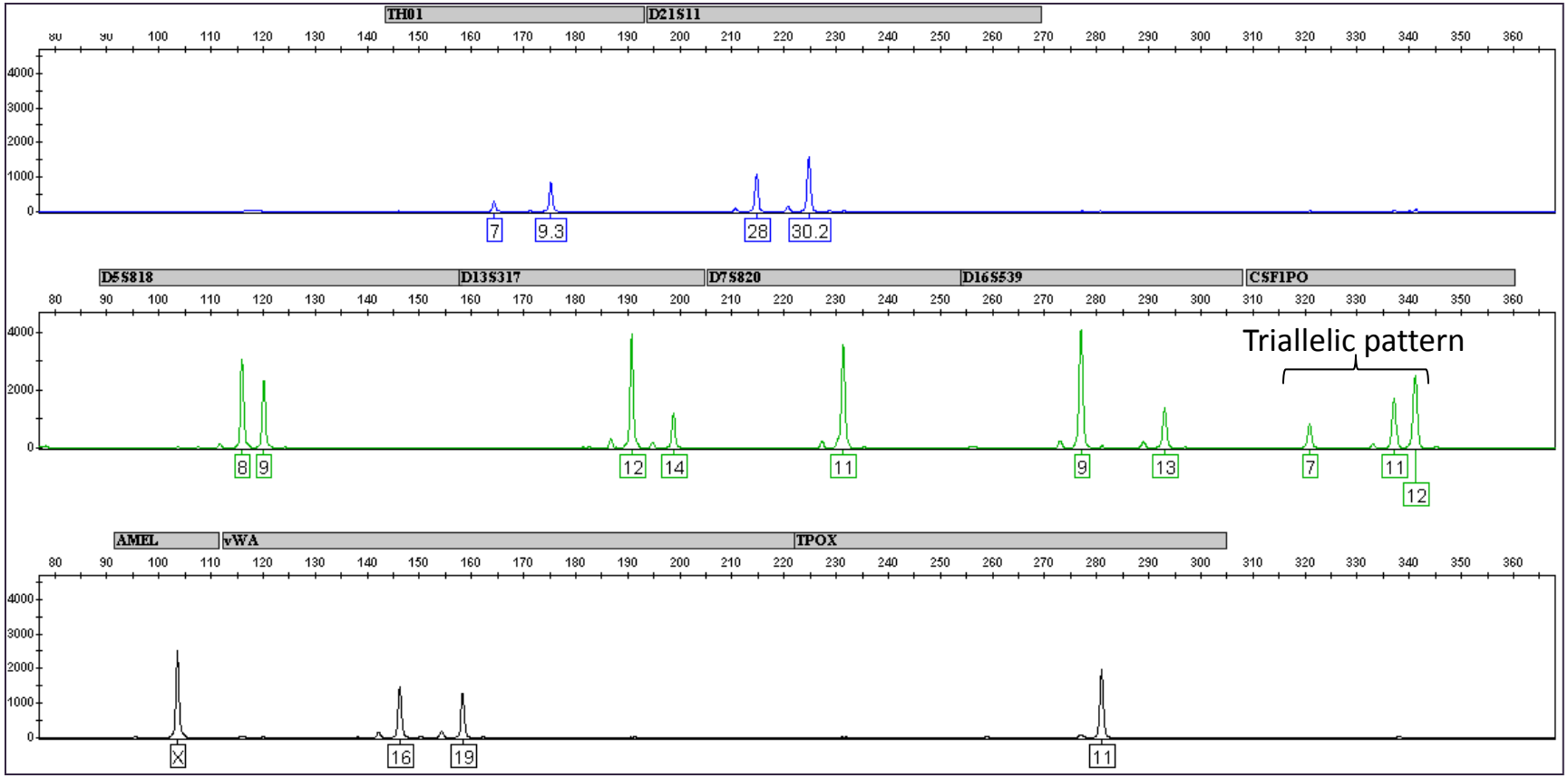
Cycling time <90 minutes for all applications

Direct Amplification of HT-29



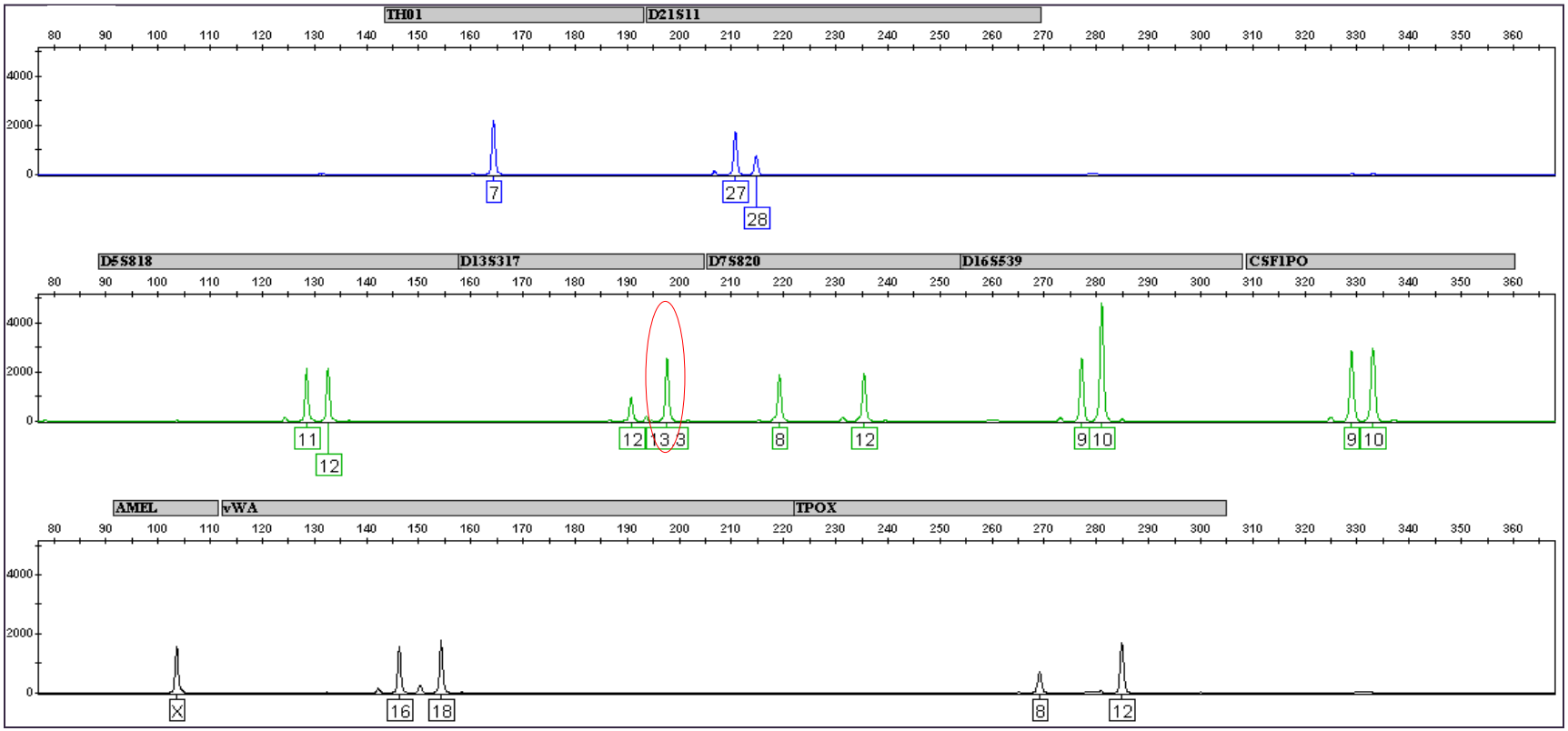
Direct amplification of a 1.2mm punch from an FTA® card using the GenePrint®10 System

HEK293 Profile – Extracted DNA



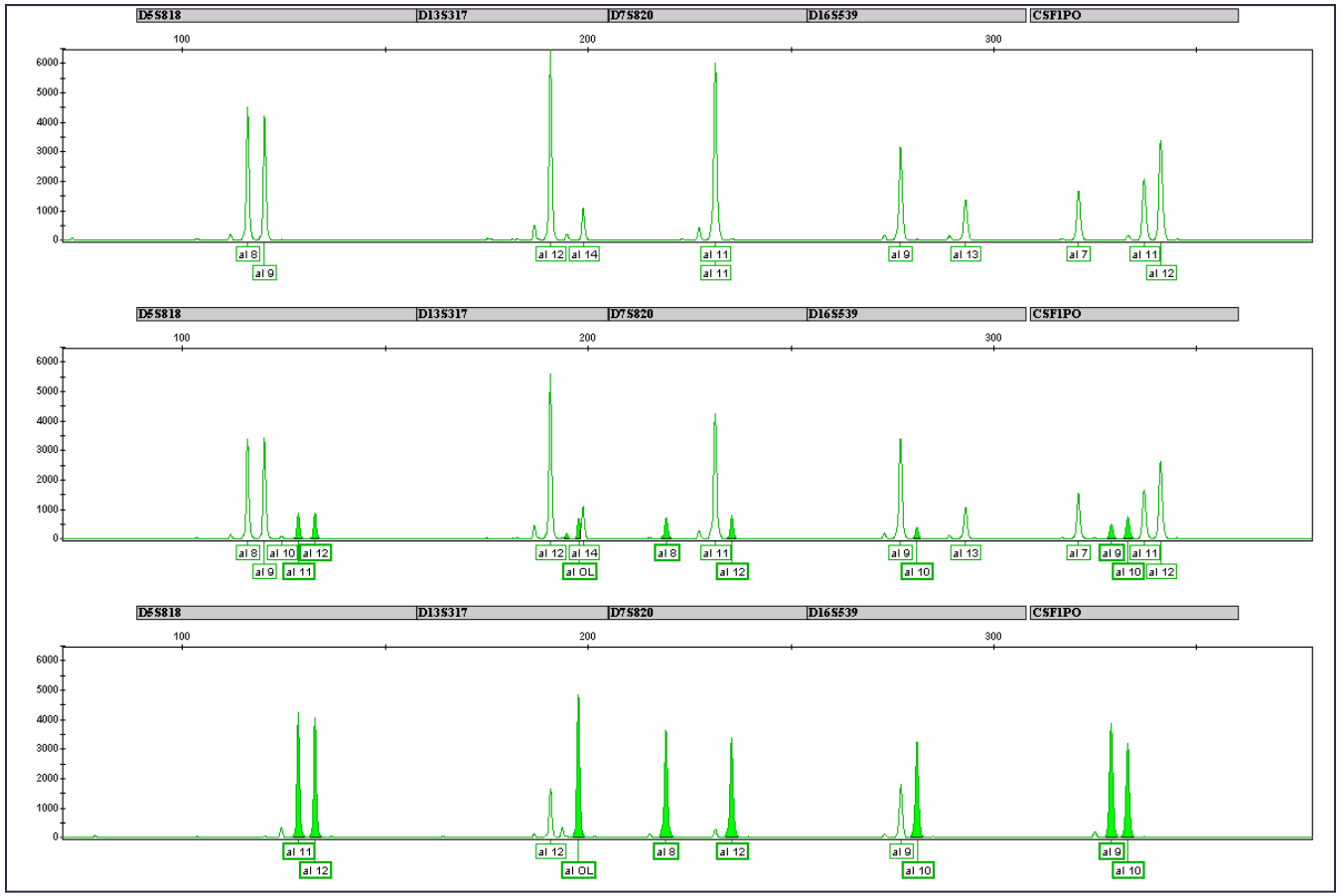
Often cell lines have triallelic patterns and other chromosomal instabilities

HeLa Profile – Extracted DNA



The 13.3 allele at D13S317 is characteristic of HeLa

Cell Line Mixtures



HEK293

HEK293 & HeLa

HeLa

Confidence in the Identity of Your Cells

Promega STR technologies and the experts at ATCC bring you confidence in the identity of your cell lines

Results and Interpretation in 3 Easy Steps

1. Place your Order for the service
2. Spot cells onto Supplied Sample Collection Card
3. Mail card to ATCC in the pre-addressed envelope

- STR analysis of 17 STR loci plus Amelogenin for gender identification
- Allele table and electropherograms supporting allele calls at each locus
- Known reference profiling against the ATCC STR database
- Comprehensive interpretation of the results
- Results emailed back in 5 business days
- Currently only available in the US, Canada, and Puerto Rico www.atcc.org/STR

Everything was going along fine until they discovered their HeLa cell line expressed Y chromosome markers.

Dunham JH and Guthmiller P. Doing Good Science: Authenticating Cell Line Identity. [Internet] 2012. [cited: year, month, date]. Available from: http://www.promega.com/resources/articles/pubhub/cell-line-authentication-with-strs-2012-update/

Additional information: http://www.promega.com/products/str-analysis/cell-line-authentication/