STR Profiling for Human Cell Line Authentication

Brian Shapiro, Ph.D.
Technical Writer, ATCC
About ATCC

- Founded in 1925, ATCC is a non-profit organization with headquarters in Manassas, VA.

- World’s premiere biological materials resource and standards development organization.

- ATCC collaborates with and supports the scientific community with industry-standard biological products and innovative solutions.

- Strong team of 400+ employees; over one-third with advanced degrees.
Outline

- History of misidentified cell lines
- Confirm species
  - STR analysis for human cell line authentication
  - Case studies
- Steps for reducing cellular and microbial contamination
- New regulations requiring cell line authentication
1952: HeLa - first immortalized human cell line

**Henrietta Lacks**
- Died at 31 years of age
- Aggressive glandular 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 Cell Biology Collection
1960s: Poor culture conditions lead 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
Better tissue culture conditions - contamination persists...

- The widespread use of HeLa resulted in cross-contamination and mix-up of cell lines.

- Recent data suggest that cross-contaminations are still a major ongoing problem with modern cell cultures.
1962: Gartler describes HeLa contamination of cell lines

2nd Biennial Review Conference on Cell Tissue and Organ Culture, Bedford, Pa, 1962

Type A (fast)
Type B (slow)
Origin

Isoenzyme analysis
Glucose-6-phosphate dehydrogenase (G6PD)

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>ATCC® No.</th>
<th>Origin</th>
<th>G6PD variant</th>
</tr>
</thead>
<tbody>
<tr>
<td>HeLa</td>
<td>Cervical adenocarcinoma; human</td>
<td>CCL-2™</td>
<td>African</td>
<td>Type A (fast)</td>
</tr>
<tr>
<td>KB</td>
<td>Oral epidermoid carcinoma, human</td>
<td>CCL-17™</td>
<td>Caucasian</td>
<td>Type A (fast)</td>
</tr>
<tr>
<td>HEP-2</td>
<td>Larynx epidermoid carcinoma, human</td>
<td>CCL-23™</td>
<td>Caucasian</td>
<td>Type A (fast)</td>
</tr>
<tr>
<td>Chang liver</td>
<td>Liver, human</td>
<td>CCL-13™</td>
<td>Caucasian</td>
<td>Type A (fast)</td>
</tr>
<tr>
<td>Int-407</td>
<td>Embryonic intestine; human</td>
<td>CCL-6™</td>
<td>Caucasian</td>
<td>Type A (fast)</td>
</tr>
</tbody>
</table>

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

1970s: Nelson-Rees describes interspecies cross-contamination of cell lines

<table>
<thead>
<tr>
<th>Actual (43/466 (9.2%))</th>
<th>Purported (62 Laboratories)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dog</td>
<td>Horse, Human, Mink, Mouse</td>
</tr>
<tr>
<td>Hamster</td>
<td>Mouse, Human, Marmoset, Rat</td>
</tr>
<tr>
<td>Mongoose</td>
<td>Human</td>
</tr>
<tr>
<td>Human</td>
<td>Gibbon</td>
</tr>
<tr>
<td>Mink</td>
<td>Human</td>
</tr>
<tr>
<td>Monkey</td>
<td>Horse, Human</td>
</tr>
<tr>
<td>Mouse</td>
<td>Human</td>
</tr>
<tr>
<td>Rabbit</td>
<td>Dog</td>
</tr>
<tr>
<td>Rat</td>
<td>Chicken, Human, Mink, Monkey</td>
</tr>
</tbody>
</table>

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

1984-2003: Interspecies and intraspecies misidentification of cell lines

<table>
<thead>
<tr>
<th>Year</th>
<th>No.</th>
<th>%</th>
<th>Type of contamination</th>
<th>Technology</th>
<th>Reference</th>
</tr>
</thead>
</table>

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

2002-2013: Misidentification of cell lines persists...

<table>
<thead>
<tr>
<th>Year</th>
<th>Title of article</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td>Beware of imposters: MA-1, a novel MALT Lymphoma cell line is misidentified and corresponds to Pfeiffer, a diffuse large B-cell lymphoma cell line</td>
<td>Genes, Chromosomes and Cancer 52 (10): 986, 2013.</td>
</tr>
</tbody>
</table>
Impact of misidentified cell lines on applied research

<table>
<thead>
<tr>
<th>Misidentification of frequently used esophageal adenocarcinoma cell lines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell Line</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>SEG-1</td>
</tr>
<tr>
<td>BIC-1</td>
</tr>
<tr>
<td>SK-GT-5</td>
</tr>
</tbody>
</table>

Experimental results based on contaminated cell lines...

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

The use of misidentification of cell lines is widespread!

- It is estimated that 1/3 of all cell lines used in the life sciences are misidentified.
- The majority of cell lines used in pre-clinical cancer research are human in origin.

Over 96% of misidentified cell lines in the ICLAC database are human in origin.

ICLAC database of misidentified cell lines:

- Total: 438
- Human: 423
- Non-Human: 15

ICLAC data base of misidentified cell lines: (http://iclac.org/databases/cross-contaminations/)
Consequences of using misidentified cell lines

- Loss of cell line
- Loss of time and money
- Misinformation in the public domain
- Discordant or irreproducible results
- Tarnished reputation

“If we’re not using what we think we’re using, we’re not testing our hypotheses. We’re just gumming up the literature. I’m not sure what we’re doing, but that’s not science.”

Jeffrey Boatright, Emory University, The Big Clean Up, The Scientist Magazine®, September 1, 2015
Short Tandem Repeat (STR) analysis for intraspecies identification of human cell line

<table>
<thead>
<tr>
<th>DNA location</th>
<th>Degree of repetition</th>
<th>Number of loci</th>
<th>Repeat unit length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satellite DNA (centromere)</td>
<td>$10^3$ to $10^7$</td>
<td>1 to 2</td>
<td>2 to several thousand bp</td>
</tr>
<tr>
<td>Minisatellite DNA (telomere)</td>
<td>2 to several hundred</td>
<td>Many thousands</td>
<td>9 to 100 bp</td>
</tr>
<tr>
<td>Microsatellite DNA (STRs); randomly scattered</td>
<td>5 to about a hundred</td>
<td>$10^4$ to $10^5$</td>
<td>1 to 6 bp</td>
</tr>
</tbody>
</table>

**STR profiling a method for human cell line authentication!**
STR analysis for human cell line identity

- Target sequence consists of microsatellite DNA
- Typically use 1-2 ng DNA
- 1 to 2 fragments
- Discrete alleles allow digital record of data
- Markers distributed throughout the genome
- Highly variable within populations; highly informative
### Properties of STRs for DNA profiling

<table>
<thead>
<tr>
<th>Locus name</th>
<th>Chromosome location</th>
<th>Repeat motif</th>
<th>No. repeating units</th>
<th>No. alleles observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>D16S539</td>
<td>16q24-qtr</td>
<td>GATA</td>
<td>5-15</td>
<td>10</td>
</tr>
<tr>
<td>D7S820</td>
<td>7q11.21-22</td>
<td>GATA</td>
<td>6-15</td>
<td>22</td>
</tr>
<tr>
<td>D13S317</td>
<td>13q22-q31</td>
<td>TATC</td>
<td>5-15</td>
<td>14</td>
</tr>
<tr>
<td>D5S818</td>
<td>5p21-q31</td>
<td>AGAT</td>
<td>7-16</td>
<td>10</td>
</tr>
<tr>
<td>CSF1PO</td>
<td>5q33.3-34</td>
<td>TAGA</td>
<td>6-16</td>
<td>15</td>
</tr>
<tr>
<td>TPOX</td>
<td>2p23-pter</td>
<td>GAAT</td>
<td>6-13</td>
<td>20</td>
</tr>
<tr>
<td>vWA</td>
<td>12p23-pter</td>
<td>[TCTA] [TCTG]</td>
<td>10-24</td>
<td>28</td>
</tr>
<tr>
<td>THO1</td>
<td>11p15.5</td>
<td>TCAT</td>
<td>3-14</td>
<td>20</td>
</tr>
<tr>
<td>Amelogenin</td>
<td>Gender determination (not STR marker)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Power of discrimination** 1:1.2 x 10E8

**Retrospective studies on 500 human cell lines – minimum of 8 STR markers required to uniquely identify a human cell line**
Characteristics of STR markers

- High discriminating power
- High observed heterozygosity >70%
- Robust and reproducible results
- Low stutter characteristics
- Low mutation rate
Advantages of STR analysis

- Banding pattern is reproducible
- PCR amplifiable, high throughput
- Small size range allows multiplexing
- Allelic ladders simplify interpretation
- Small product size compatible with degraded DNA
- Rapid processing is attainable
- Probes are human-specific
- High stringency conditions
Outline of STR profiling procedure

**DNA sample**

- Amplification of STR loci
- Simultaneous fluorescent labeling

**Multiplex PCR**

- Addition of Internal Lane Standard
- CE to separate fragments
- Fluorescent detection
- Run allelic ladder in parallel

**Capillary Electrophoresis (CE)**

- Calculate size based on Internal Lane Standard
- Compare fragment sizes to allelic ladders to determine STR alleles
- Compare to databases

**Data Analysis**

**Internal Lane Standards**

**Allelic ladder (3 loci)**
Requirements for performing STR analysis

- Gene sequencer
- Thermocycler
- Primer kits from manufacturers
- STR database of human cell lines
- Experienced technicians
STR DNA polymorphism

**A: Homozygous at locus D16S539**

- GATA
- 8 repeating units

**B: Heterozygous at locus D16S539**

- GATA
- 10 repeating units
- 9 repeating units

**Unique STR DNA profile for each cell line derived from unrelated individuals**
Informativeness of STR markers

Example: 6 STR makers

<table>
<thead>
<tr>
<th></th>
<th>D5S818</th>
<th>D13S317</th>
<th>D7S820</th>
<th>D16S539</th>
<th>CSF1PO</th>
<th>Penta D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency from maternal chromosome</td>
<td>1:10</td>
<td>1:9</td>
<td>1:9</td>
<td>1:9</td>
<td>1:10</td>
<td>1:15</td>
</tr>
<tr>
<td>Frequency from paternal chromosome</td>
<td>1:10</td>
<td>1:9</td>
<td>1:9</td>
<td>1:9</td>
<td>1:10</td>
<td>1:15</td>
</tr>
<tr>
<td>Probability</td>
<td>1:20 (0.042)</td>
<td>1:18 (0.056)</td>
<td>1:18 (0.056)</td>
<td>1:18 (0.056)</td>
<td>1:20 (0.042)</td>
<td>1:30 (0.033)</td>
</tr>
</tbody>
</table>

Total probability = 1:93000000 (1: 9.3 x 10^7)
Gender is important for cell line identification

Amelogenin gene: AMELX (female), AMELY (male)

AMELX gene contains a 6 bp deletion in the intron 1
Unrelated human cell lines: STR analysis

<table>
<thead>
<tr>
<th></th>
<th>D5S818</th>
<th>D13S317</th>
<th>D7S820</th>
<th>D16S539</th>
<th>vWA</th>
<th>THO1</th>
<th>Amel.</th>
<th>TPOX</th>
<th>CSF1PO</th>
</tr>
</thead>
<tbody>
<tr>
<td>K562</td>
<td>11, 12</td>
<td>8</td>
<td>9, 11</td>
<td>11, 12</td>
<td>16</td>
<td>9.3</td>
<td>X</td>
<td>8, 9</td>
<td>9, 10</td>
</tr>
<tr>
<td>WS1</td>
<td>13</td>
<td>12</td>
<td>9, 10</td>
<td>10, 11</td>
<td>17, 18</td>
<td>8, 10</td>
<td>X</td>
<td>8, 9</td>
<td>10, 13</td>
</tr>
</tbody>
</table>

2 unrelated cell lines, separate individuals, unique STR DNA profiles
Related human cell line identification: STR analysis

Two related cell lines, same individual, identical DNA STR profile

<table>
<thead>
<tr>
<th>D5S818</th>
<th>D13S317</th>
<th>D7S820</th>
<th>D16S539</th>
<th>vWA</th>
<th>THO1</th>
<th>Amel.</th>
<th>TPOX</th>
<th>CSF1PO</th>
</tr>
</thead>
<tbody>
<tr>
<td>HAAE-2</td>
<td>12,13</td>
<td>11,12</td>
<td>8,10</td>
<td>12,13</td>
<td>14,18</td>
<td>7,9</td>
<td>X,Y</td>
<td>10,11</td>
</tr>
<tr>
<td>HFAE-2</td>
<td>12,13</td>
<td>11,12</td>
<td>8,10</td>
<td>12,13</td>
<td>14,18</td>
<td>7,9</td>
<td>X,Y</td>
<td>10,11</td>
</tr>
</tbody>
</table>
STR DNA profile links clone to parent

Clonal derivative has identical DNA profile to parental cell line

BG01 HuES (parent)

BGO1V HuES (clone)

<table>
<thead>
<tr>
<th>STR</th>
<th>BG01</th>
<th>BG01V</th>
</tr>
</thead>
<tbody>
<tr>
<td>D5S818</td>
<td>10, 12</td>
<td>10, 12</td>
</tr>
<tr>
<td>D13S317</td>
<td>11, 12</td>
<td>11, 12</td>
</tr>
<tr>
<td>D7S820</td>
<td>10, 11</td>
<td>10, 11</td>
</tr>
<tr>
<td>D16S539</td>
<td>9, 11</td>
<td>9, 11</td>
</tr>
<tr>
<td>vWA</td>
<td>16, 17</td>
<td>16, 17</td>
</tr>
<tr>
<td>TH01</td>
<td>7, 9.3</td>
<td>7, 9.3</td>
</tr>
<tr>
<td>TPOX</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>CSF1PO</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Amel.</td>
<td>X, Y</td>
<td>X, Y</td>
</tr>
</tbody>
</table>

Diploid; normal chromosomes

Aneuploid; +1 copy chrs. 12, 17, +1 copy chr. X

Karyotype (g-band)
STR analysis for monitoring genomic stability of cell banks

Donor

Token (Pre-MCB)

Seed (MCB)

Distribution (WCB)
Common sources of misidentification of cells

Common sources of cellular contamination
- Getting cell lines from a colleague down the hall
- Continuous culturing of working cell banks
- Use of feeder cells
- Mislabling of culture flasks
- Working with multiple cell lines, concurrently
- Using one reservoir of growth medium for multiple cell lines

Are your cells REALLY what you think they are?
Case study 1: Cellular cross-contamination

SK-OV-3
Ovary

SK-OV-3 + cell line X
Case study 2: Gender misidentification

Human cell line purported to be of female origin

Y material found by STR profiling and confirmed by Y chromosome paint
Case study 3: Misidentified cell line

Cell line X

<table>
<thead>
<tr>
<th>Loci</th>
<th>Cell line X</th>
<th>HT-29 (ATCC®HTB-38™)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D5S818</td>
<td>11.12</td>
<td>11.12</td>
</tr>
<tr>
<td>D13S317</td>
<td>11.12</td>
<td>11.12</td>
</tr>
<tr>
<td>D7S820</td>
<td>10</td>
<td>11.12</td>
</tr>
<tr>
<td>D16S539</td>
<td>11.12</td>
<td>11.12</td>
</tr>
<tr>
<td>CSF1PO</td>
<td>6,9</td>
<td>X</td>
</tr>
<tr>
<td>vWA</td>
<td>17,19</td>
<td>17,19</td>
</tr>
<tr>
<td>TH01</td>
<td>X</td>
<td>8,9</td>
</tr>
<tr>
<td>Amel.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPOX</td>
<td>8,9</td>
<td></td>
</tr>
</tbody>
</table>

HT-29: Human colon adenocarcinoma cell line
Case study 3 (cont’d): Misidentified cell line

Morphology, karyotyping, STR analysis of cell line similar to HT-29 cell line
Case study 4: Microsatellite instability

Instability of D13S317 – allele 9

Seed (MCB) 1

Distribution (WCB) 1

Distribution (WCB) 2

Distribution (WCB) 3
STR - a standard for the authentication of human cell lines

ASN-0002 - Authentication of Human Cell Lines: Standardization of STR Profiling

- The standard describes a consistent, inexpensive, and universally applicable method for authenticating new and established cell lines and their criteria for use

- **Chair:** John R.W. Masters, University College of London
- **Co-Chair:** Yvonne A. Reid, ATCC
- **Final action by ANSI:** January 25, 2012
- **Published date:** February 2, 2012

Steps for reducing cellular contamination

- Good documentation
- Highly trained technicians
- Good aseptic techniques
- Use one reservoir of medium per cell line
- Aliquot stock solutions/reagents

NCI-H441 (ATCC® HTB-174™) human papillary adenocarcinoma differentiated under air-liquid interface conditions
Steps for reducing cellular contamination

- Label flasks with name of cell line, passage number, date of transfer (use barcoded flasks when available)
- Work with one cell line at a time in a biological safety cabinet
- Clean the biological safety cabinet between each cell line
- Allow a minimum of 15 minutes between each cell line
Steps for reducing cellular contamination

- Quarantine “dirty” cell lines from “clean” cell lines
- Manageable work load (reduce accidents)
- Clean laboratory (reduce bioburden)
- Legible handwriting (printed labels)
Steps for reducing cellular contamination

- Monitor for cell line identity routinely
- Create a “good” working environment
- Review and approve laboratory notebooks
- Obtain each cell line from a reputable source

HUVEC (ATCC® CRL-1730™) expressing CD34
Many journals now make cell line authentication a pre-requisite for publication and funding agencies are requiring cell authentication in their decisions. Cell line authentication utilizing Short Tandem Repeat (STR) profiling aids in the detection of misidentified, cross-contaminated, or genetically drifted cells. STR profiling prevents serious errors leading to wasted time money and retracted research. Source: http://highwire.stanford.edu/
# Publications requiring/recommending cell line authentication

<table>
<thead>
<tr>
<th>BioTechniques</th>
<th>International Journal of Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancer Discovery (AACR)</td>
<td>Journal of Clinical Endocrinology &amp; Metabolism</td>
</tr>
<tr>
<td>Cancer Epidemiology, Biomarkers and Prevention (AACR)</td>
<td>Journal of Endocrinology</td>
</tr>
<tr>
<td>Cancer Immunology Research (AACR)</td>
<td>Journal of Molecular Biology</td>
</tr>
<tr>
<td>Cancer Prevention Research (AACR)</td>
<td>Journal of Molecular Endocrinology</td>
</tr>
<tr>
<td>Cancer Research (AACR)</td>
<td>Journal of the National Cancer Institute</td>
</tr>
<tr>
<td>Carcinogenesis</td>
<td>Molecular Cancer Research (AACR)</td>
</tr>
<tr>
<td>Cell Biochemistry and Biophysics</td>
<td>Molecular Cancer Therapeutics (AACR)</td>
</tr>
<tr>
<td>Cell Biology International</td>
<td>Molecular Endocrinology</td>
</tr>
<tr>
<td>Clinical Cancer Research (AACR)</td>
<td>Molecular Vision</td>
</tr>
<tr>
<td>Clinical Orthopaedics and Related Research</td>
<td>Nature</td>
</tr>
<tr>
<td>Endocrine Reviews</td>
<td>Nature Cell Biology</td>
</tr>
<tr>
<td>Endocrine-Related Cancer</td>
<td>Nature Methods</td>
</tr>
<tr>
<td>Endocrinology</td>
<td>Neuro-Oncology</td>
</tr>
<tr>
<td>In Vitro Cellular &amp; Developmental Biology – Animal</td>
<td>PLoS ONE</td>
</tr>
</tbody>
</table>

The number of publications requiring/recommending cell line authentication as re-requisite for publication is growing...
Cell line authentication to improve reproducibility in cancer research

NIH revised guidelines to applications for funding - Enhancing Reproducibility through Rigor and Transparency (effective Jan. 25, 2016)

- Authentication of Key Biological and/or Chemical Resources –
  - NIH expects that key biological and/or chemical resources will be regularly authenticated to ensure their identity and validity for use in the proposed studies
  - Do not limit authentication to cell lines; include specialty chemicals, antibodies, and biologics

Cell line authentication to improve reproducibility in cancer research

NIH Principles and Guidelines for Reporting Preclinical Research

5-points guidelines – No. 5: Consider establishing best practices guidelines for:

- Cell lines: to report source, authentication, and Mycoplasma contamination status
- Over 130 signatories of journals, scientific associations, and societies
Cell line authentication to improve reproducibility in cancer research

Prostate Cancer Foundation – Cell Line Authentication Initiative

- Perform genetic and pathogen tests for cell lines for future funding

Nature Journals

- Authors are asked to report the source of the cell line, if the cell line is authenticated, and if it has been checked against a database of misidentified cell lines
Economic impact of using misidentified cell lines

NIH Reporter for projects using “cell line” or “cell culture” – US $3.7B

The Economics of Reproducibility in Preclinical Research; PLoS Biology June 9, 2015
When to perform STR profiling

- When you first receive a cell line into the laboratory from a unreliable source
- After 10 passages
- After preparing a cell bank
- When in doubt
Testing services for STR profiling of human cell lines

Testing Services

- Cell Banks
- Paternity testing labs
- Universities
- Core labs

Important consideration when choosing a Testing Service

- Highly trained, experienced technicians
- Access to database of STR profiles
Disclaimers

© American Type Culture Collection. The ATCC trademark and trade name, and any other trademarks listed in this publication are trademarks owned by the American Type Culture Collection unless indicated otherwise.
Thank you for joining today!

Register for more ATCC webinars at www.atcc.org/webinars

Learn more about our cell authentication resources at www.atcc.org/str

Please email additional questions to tech@atcc.org