Cell Line Authentication Using the *GenePrint*® 10 and *GenePrint*® 24 Systems

A Spectrum Compact CE System Application Note

Promega Corporation

Sample Types:
Human cultured cell lines
Human cultured cell lines spotted on lytic storage cards

Instrument Requirements:

- Spectrum Compact CE System (Cat.# CE1304)

Promega Reagents:

- *GenePrint*® 10 System (Cat.# B9510)
- *GenePrint*® 24 System (Cat.# B1870, B1874)

Other Reagents and Consumables Required:

- Spectrum Compact CE System consumables (cartridge, polymer, septa, retainer, buffer; see Ordering Information at end of document for full list)
- PowerPlex® 4C Matrix Standard (Cat.# DG4800)
- *GenePrint*® 5C Matrix Standard (Cat.# B1930)
- Hi-Di™ Formamide (Thermo Fisher Scientific Cat.# 4401457)
- FTA® cards (Whatman)

Optional Instruments and Reagents:

- Maxwell® RSC Instrument (Cat.# AS4500)
- Maxwell® RSC Cultured Cells DNA Kit (Cat.# AS1620)

DNA Analysis Software:

- GeneMapper® Software Version 6 (Applied Biosystems) or similar

Introduction

Imagine you're studying colon cancer using a colon cell line model. After three painstaking years of research, your first major publication describes several new findings on the molecular pathways in this model. Or perhaps you identify sensitivity to a particular drug treatment. All of this work—your findings, its effect on the medical community, your reputation and grant access—hinge on your understanding of this cell line as a colon cancer model. But what happens if that cell line, gifted from the lab across the hall, has been mislabeled during routine passage? Or was overtaken by the faster growing HeLa cells also cultured in your lab?

Cell mislabeling or contamination is such a rare occurrence, you don't have to worry about that, right? Unfortunately, misidentified cells are all too common. In 1999, an estimated 18% of cell lines submitted to a German cell bank were misidentified (1). The Web of Science database uncovered 32,755 research articles that use cell lines known to be misidentified (2). And the TRIP Lab, a core facility providing cell line authentication, estimated that 28% of cell lines tested in 2017 were either contaminated or misidentified (3). These numbers are startling, and certainly not something you want to bet your career or the lives of patients on.

But there is good news, too. Major granting institutions and many publications are now requiring cell line authentication. Awareness of this issue is improving in the research community. And this shift in awareness and funding is having an effect. After outreach to researchers in its institution, the TRIP Lab showed that misidentified and contaminated cell lines dropped from 28% to 3.8% in just 2 years (3).

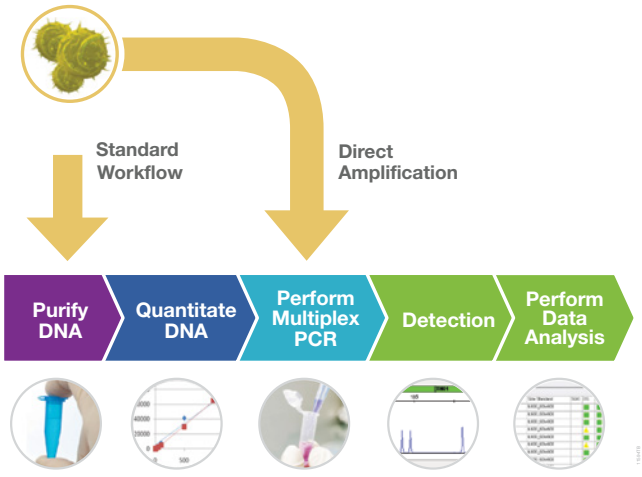

Figure 1. Overview of the process for cell line authentication.

Best Practices in Cell Line Authentication

To deal with the problem of misidentified or contaminated cells, a committee of experts published the ANSI/ATCC ASN-0002-2011 consensus guidelines for best practices in cell line authentication based on STR genotyping (4). Short tandem repeats (STRs) are 2–7bp repeating DNA sequences in the genome that are typically highly polymorphic between individuals in a population. These regions can be amplified by PCR using primers outside of the repeat sequence, and the resulting amplicons sized using capillary electrophoresis to determine the number of repeats for each individual. This technology is commonly used in the forensic community for identification of individuals, and can also be used to distinguish cell lines from different human donors. Most STR chemistries currently available can be performed as multiplexes and use multiple dye channels for electrophoresis for concurrent analysis of many STR loci at the same time.

Following the publication of the ANSI/ATCC guidelines, the International Cell Line Authentication Committee (ICLAC) was established to maintain a register of cross-contaminated and misidentified cell lines. Their website is also an excellent collection of resources for cell line authentication (CLA), including links to several current STR databases, guides and advice for individual scientists, and policy recommendations for institutions seeking to establish rigorous science with authenticated cell lines as the norm (5). We discuss the basic tenets below to familiarize you with the cell line authentication steps that can be performed in your lab using the Spectrum Compact CE System.

First, verify that any cell lines that you intend to use are not listed in the ICLAC Register of Misidentified Cell Lines (5). This register currently contains 552 cell lines (Version 10, released 25 March 2020) that are known to be misidentified or contaminated. Checking this database is an easy and free way to avoid loss of research time and funds spent pursuing experiments with erroneously labeled cell lines.

Second, perform STR genotyping on DNA purified from your cell line of interest, or from cells conserved on sample storage cards (e.g., Whatman® FTA® cards). If you have access to instrumentation like the Spectrum Compact CE System, you can genotype the cells in your own lab. Alternatively, samples can be sent to a high-quality core facility that follows the ANSI/ATCC guidelines. Resources available on the ICLAC website recommend performing such profiling at the beginning and end of each project, whenever observing a new phenotypic behavior or if performing phenotypic selection (e.g., drug selection), before submitting any grants or manuscripts, when freezing cell stocks, or when initiating new cell lines. While there is expense involved with these tests, the cost relative to a typical cell line project are minimal and help safeguard the results of your research project and the reputation of your lab and institution.

Third, compare the obtained genotype to a reference database, like the ATCC STR Database (**www.atcc.org/en/STR_Database.aspx**), the DSMZ STR Profile Database (**www.dsmz.de/services/human-and-animal-cell-lines/online-str-analysis**), or Cellosaurus (**https://web.expasy.org/cellosaurus-str-search/**). Each of these databases include a search option to compare your obtained test sample genotype to the cell lines included in the database and will return a percent match. Cell lines are typically aneuploid and may show genetic drift relative to the reference (4). While this should be minimized with good cell culture practices (e.g., using cell lines only for a limited number of passages) some variability is expected. Therefore a threshold of 80% genotype match has been established to claim cell line authentication. Note that cell lines established from the same donor or from identical twins may not be distinguished using STR genotyping.

Three different formulas are commonly used to determine the percent match (6). The formula used may vary between search engines and can drastically affect the percent match. This percent match is key to determining if your cell line is correctly identified, so make sure you understand how it is being calculated in whatever database you query.

Match Formula	Percent Match Equation
Masters	$100 \times \frac{\text{(Number shared alleles)}}{\text{(Total number alleles in Test Sample)}}$
Alternative Masters	$100 \times \frac{\text{(Number shared alleles)}}{\text{(Total number alleles in Reference Sample)}}$
Tanabe	$100 \times \frac{\text{(Number shared alleles} \times 2\text{)}}{\text{(Total number alleles in Test Sample + total number alleles in Reference Sample)}}$

Fourth, examine any cell lines for cross-contamination based on the indicated genotype. If using the Masters or Tanabe formulas, extra alleles present due to cross-contamination would lower the percent match. In contrast, extra alleles in the test sample do not affect the percent match calculated using the Alternative Masters formula (used by the ATCC STR Database). In this case, percent match can be calculated manually using either the Masters or Tanabe formula and should be >80% for uncontaminated cell lines. The

ANSI/ATCC Guidelines indicate that it is unlikely that any single cell line would have more than two alleles in three or more loci. Multiple alleles at multiple loci are another indication of cross-contamination.

And finally, testing alone cannot substitute for good cell culture practices, including a written lab policy and training protocol. In addition to performing STR genotyping for cell line authentication, consider additional measures to avoid other types of contamination, such as regular mycoplasma testing and testing for mouse cell contamination if mouse cell lines are used in the lab. Several companies and core facilities offer reagents and fee-for-service mycoplasma testing. Promega offers custom primers for identifying mouse cell line contamination that can be combined with STR genotyping.

Methods

DNA was purified from 7×10^5–2×10^6 cells of the indicated cell lines using the Maxwell® RSC Cultured Cells DNA Kit (Cat.# AS1620) on a Maxwell® RSC Instrument (Cat.# AS4500). DNA was quantified using absorbance on a NanoDrop® 8000 Spectrophotometer (Thermo Fisher Scientific) and diluted to 2ng/µl or 1ng/µl in 10mM Tris-HCl (pH 8.0), 100µM EDTA and 20µg/ml glycogen. For detecting cell mixtures, gDNA from HEK293 was mixed with HeLa gDNA at 20%, 10%, 5%, 4%, 3%, 2% or 1%, keeping the total DNA concentration constant. DNA from single-source cell lines or mixtures were amplified in triplicate with the *GenePrint*® 10 System described in the GenePrint® *10 System Technical Manual* #TM589, along with three no-template-control (NTC) reactions and one positive control. The same samples were also amplified in triplicate with the *GenePrint*® 24 System as described in the GenePrint® *24 System Technical Manual* #TM525, using either 2.5ng DNA input and 26 cycles or 5.0ng DNA input and 25 cycles.

For direct amplification, HEK293 and HeLa cell lines were trypsinized and counted using a Countess Cell Counter. Cells were diluted in PBS and mixed to contain 50%, 30%, 25%, 20%, 15%, 10%, 5% or 2.5% HeLa cells in a background of HEK293 cells at a total concentration of 1×10^6 cells/ml. Forty microliters of single-source cells, cell mixtures or PBS was spotted in triplicate on Whatman® FTA® Classic Cards and allowed to dry for at least 15 minutes in a tissue-culture hood, then stored at room temperature for 1 week before testing. Single 1.2mm punches from each card were amplified the *GenePrint*® 10 System following the protocol in Technical Manual #TM589, Section 11.3, including three PBS control punches, two NTC reactions and two positive control reactions.

GenePrint® 10 amplification products were denatured in Hi-Di™ formamide with Internal Lane Standard 600 (ILS600), and samples electrophoresed on a Spectrum Compact CE System with Polymer4 and the preloaded assay Promega_4Dye_ILS600_36_P4. *GenePrint*® 24 amplification products were denatured in Hi-Di™ formamide with WEN Internal Lane Standard 500 (ILS500) and samples electrophoresed on a Spectrum Compact CE System with Polymer7 using the preloaded assay settings for Promega_5Dye_WENILS_36_P7. All datasets included at least one allelic ladder per 16 samples (four injections). Injection and run conditions are given in Table 1.

Table 1. Default Fragment Analysis Parameters on the Spectrum Compact CE System. Reactions amplified with the *GenePrint*® 10 System were injected using Polymer4 and the preloaded assay Promega_4Dye_ILS600_36_P4. Reactions amplified with the *GenePrint*® 24 System were injected using Polymer7 and the preloaded assay Promega_5Dye_WENILS_36_P7.

	GenePrint® 10 System	*GenePrint*® 24 System
Assay Name	Promega_4Dye_ ILS600_36_P4	Promega_5Dye_ WENILS_36_P7
Injection Voltage	1.6kV	1.5kV
Injection Time	9 seconds	9 seconds
Run Voltage	13kV	13kV
Run Time	1,930 seconds	1,290 seconds
Oven Temperature	60°C	60°C

Data exported as .fsa files were analyzed using GeneMapper® Software, Version 6. For the *GenePrint*® 10 System, data were analyzed with the default peak detection threshold of 50RFU and with cutoff values reduced to 0.01 for improved mixture detection. For the *GenePrint*® 24 System, data were analyzed with a 75RFU peak detection threshold (based on analytical threshold testing in our laboratory) and the default cutoff value of 0. Artifacts, bleedthrough and n+1 repeat stutter (n + 3bp, 4bp or 5bp) above the peak detection threshold were omitted manually. Cell line identity was confirmed by querying the ATCC STR Database. Peak-height ratio was calculated for single-source samples as the shorter peak height divided by the taller peak height × 100 for heterozygous markers. Homozygous markers, trialleleic markers (e.g., HEK293 CSF1PO) and Y-chromosome markers (e.g., DYS391) were omitted from peak height analysis. All data are reported as themean ± standard deviation across triplicate reactions.

Results

Genomic DNA was purified from A549, HEK293, HeLa, Raji, Ramos and SKBR3 cell lines obtained locally and amplified with both the *GenePrint*® 10 System and *GenePrint*® 24 System for cell line authentication. Example electropherograms for the allelic ladders and Raji Burkitt's lymphoma cell line are shown in Figures 2 and 3. Data were analyzed in GeneMapper™ Software 6 and genotypes determined for each sample. In all cases, genotypes were concordant across replicates and regardless of system (*GenePrint*® 10 or *GenePrint*® 24 System) or DNA input.

To confirm the identity of each cell line, the ATCC STR Database was queried with the resultant genotypes. This database currently only lists genotype for the 9 core loci recommended by the ANSI/ATCC guidelines, which are included in both *GenePrint*® Systems, and uses the Alternative Masters formula to determine percent match. Four of the six cell lines—HeLa, Raji, Ramos and SK-BR-3—were identical to their reference genotypes (data not shown). A549, however, showed loss of one allele at CSF1PO relative to the canonical A549 ATCC reference sample. An example report for this query is shown in Tables 2 and 3, including the percent match

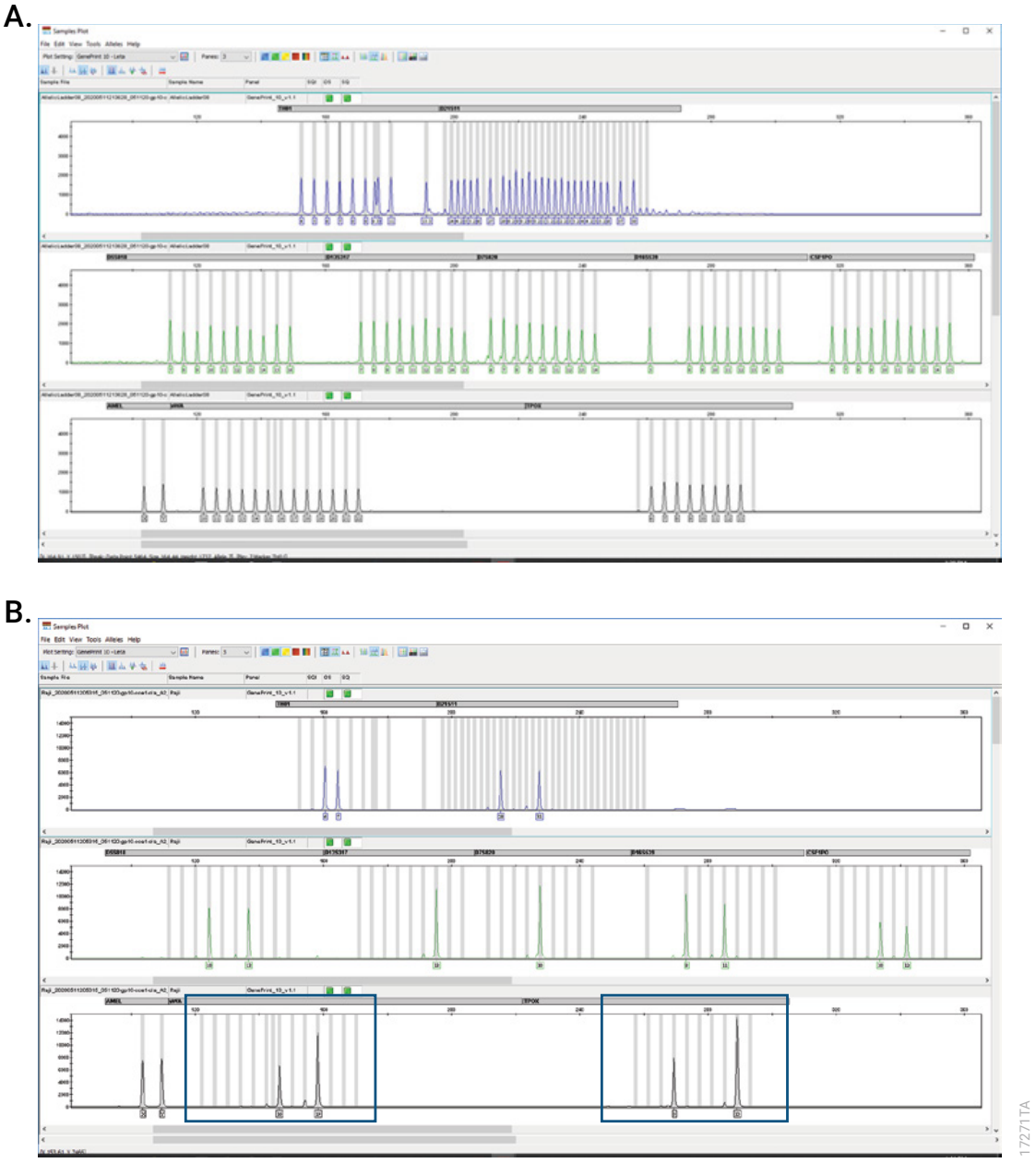

Figure 2. The *GenePrint*® 10 System allelic ladder (Panel A) and example Raji cell line electropherogram (Panel B) analyzed on a Spectrum Compact CE System. Outlined sections indicate the vWA and TPOX loci, which show allelic imbalance in the Raji cell line.

as calculated with each of the common formulas. Regardless of formula, our A549 test sample is considered correctly identified with >80% match to the expected reference genome.

Similarly, the HEK293 test sample shows loss of one allele (at D7S820) and gain of one allele (at CSF1PO) relative to the canonical ATCC reference genotype (Table 2). The percent match, regardless of formula, is 94%, indicating a match to the reference HEK293 CRL-1573 sample. However, there are many variations on HEK293 cells, including a large number of cell lines that have been derived from the original HEK293 founders and transformed with other genetic elements. Using STR genotyping, the specific HEK293 derivative cannot be determined.

Human DNA is typically expected to show allelic balance at STR loci, with either a single peak for a homozygous genotype, or two peaks of near equal peak height for a heterozygous genotype (i.e., a peak height ratio of 70–100%). However, cell lines are prone to allelic imbalance and instability, resulting in electropherograms that can vary from the typical peak height ratios. This can be seen for Raji cells in loci vWA and TPOX

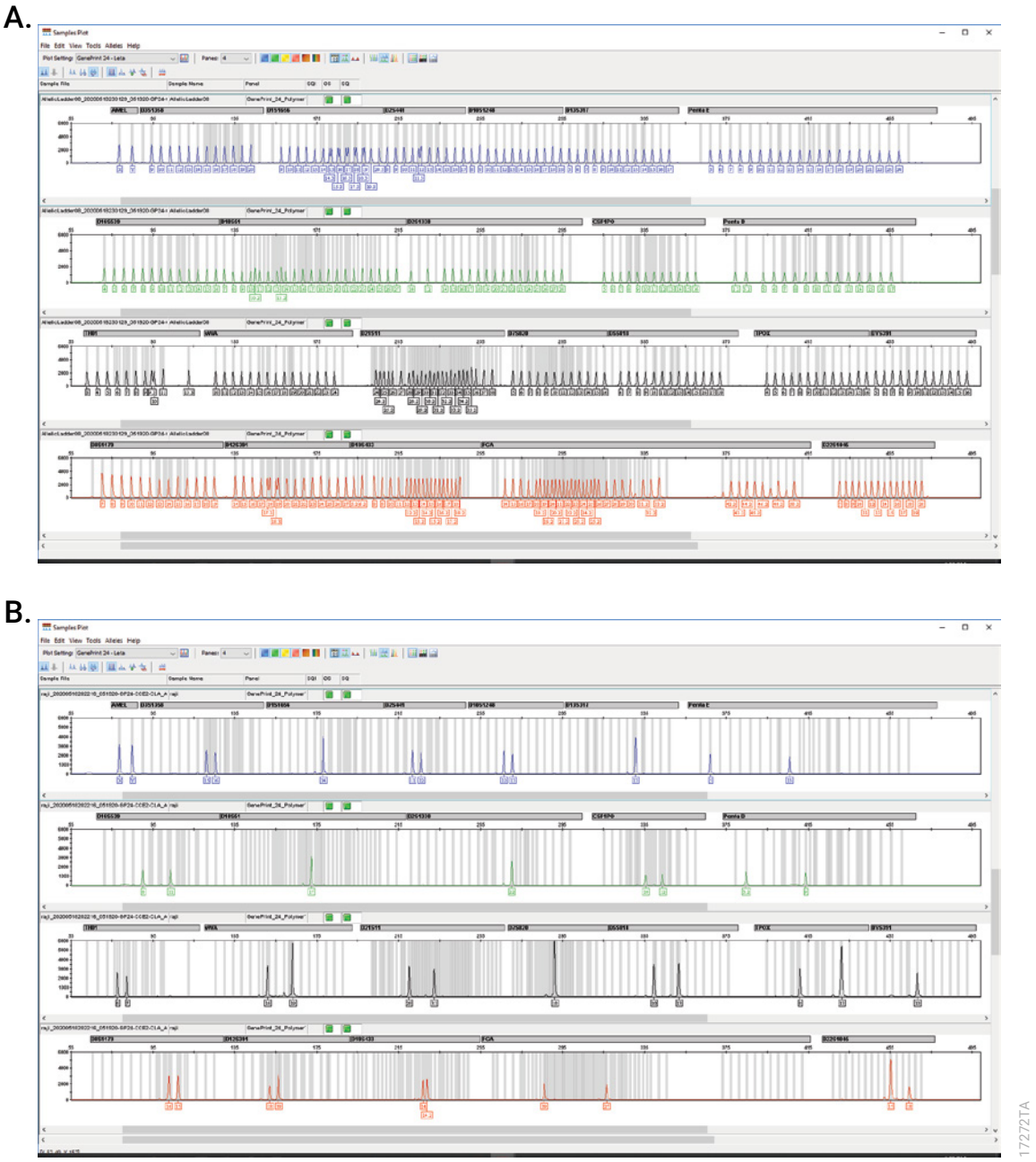

Figure 3. The *GenePrint*® 24 System allelic ladder (Panel A) and example Raji cell line electropherogram (Panel B) analyzed on a Spectrum Compact CE System. The example cell line data was amplified from 5ng input of Raji cell line gDNA.

Table 2. Cell Identification Query Results from Cell Line Genotypes Amplified Using the *GenePrint*® 10 and *GenePrint*® 24 Systems. The nine core loci included in the ATCC STR Database (gray) were used for the query.

Loci	Test Sample: A549				Test Sample: HEK293					
	Query Profile		ATCC Reference Profile: A549 (CCL-185)		Query Profile			ATCC Reference Profile: HEK293 (CRL-1573)		
AMEL	X	Y	X	Y	X			X		
D3S1358	16				15	17				
D1S1656	17	18.3			15	17.3				
D2S441	10	13			11	15				
D10S1248	13	16			14					
D13S317	11		11		12	14		12	14	
Penta E	7	11			7	15				
D16S539	11	12	11		9	13		9	13	
D18S51	14	17			17	18				
D2S1338	24				19					
CSF1PO	10		10	12	7	11	12	11	12	
Penta D	9				9	10				
TH01	8	9.3	8	9.3	7	9.3		7	9.3	
vWA	14		14		16	19		16	19	
D21S11	29				28	30.2				
D7S820	8	11	8	11	11			11	12	
D5S818	11		11		8	9		8	9	
TPOX	8	11	8	11	11			11		
DYS391	10									
D8S1179	13					12	14			
D12S391	18					19	21			
D19S433	13					15	18			
FGA	23					23				
D22S1045	15					16				

Table 3. Calculating the Percent Match. The percent match between the test sample and the ATCC reference sample was calculated based on the results in Table 2 according to the ANSI Guidelines (Alternative Masters formula) as well as the Masters and Tanabe formulas (4, 6–7).

	A549	HEK293
Number of Shared Alleles Between the Test and Reference Samples	14	15
Total Number of Alleles in the Test Sample	14	16
Total Number of Alleles in the Reference Sample	15	16
Masters Formula Match	100%	94%
Alternative Masters Formula Match	93%	94%
Tanabe Formula Match	97%	94%

(see blue box in Figure 2) and should not be automatically interpreted as cross-contamination. Peak height ratios for all heterozygous loci are shown in Figure 4, demonstrating the degree of allelic imbalance across these cell lines. For all cell lines tested, the degree of allelic imbalance is consistent between the *GenePrint*® 10 and *GenePrint*® 24 Systems for the 10 common loci. Note that all of these loci are balanced in single-source human gDNA with these same kits (data not shown).

STR analysis can help identify cross-contamination of cell lines with other human cells. To demonstrate sensitive contamination detection, HeLa gDNA was mixed with HEK293 gDNA down to 1%, amplified with the *GenePrint*® Systems, and analyzed with the Spectrum Compact CE System. Alleles were automatically determined with GeneMapper™ Software 6. Example electropherograms are shown in Figure 5 while Figure 6 shows the detection of alleles unique to the contaminating HeLa cell line. As shown in the electropherograms, single-source HeLa and HEK293 cell lines have different genotypes for locus D5S818. One of these unique HeLa alleles is automatically detected above background in this example of a 5% HeLa contamination mixture, whereas both are detected in a 20% mixture. Using the *GenePrint*® 10 System, as little as 2% HeLa gDNA contamination results in

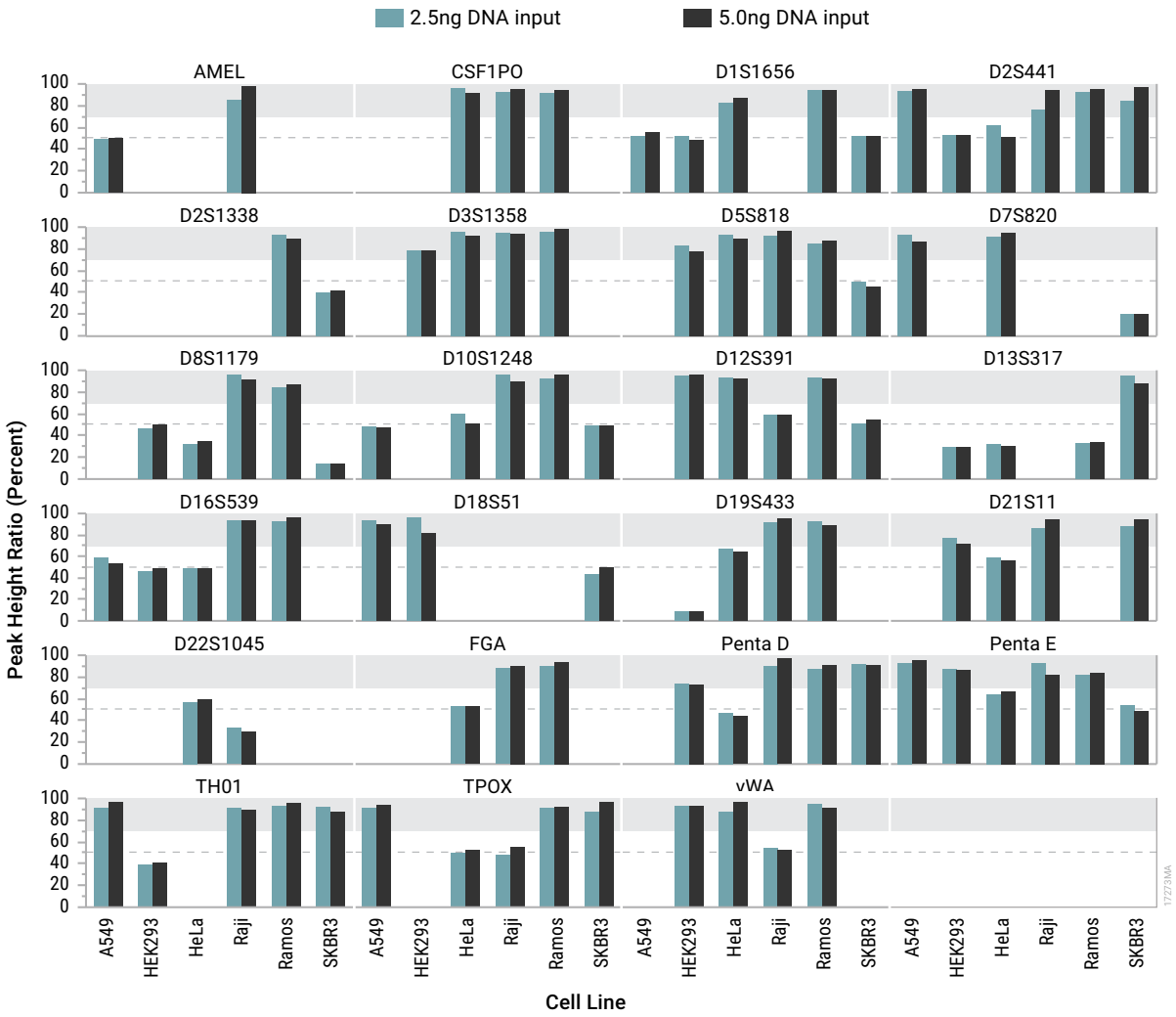

Figure 4. Allelic imbalance in cell line DNA amplified with the *GenePrint*® 24 System and injected on the Spectrum Compact CE System. Peak-height ratios for all heterozygous loci were calculated as the shorter peak height divided by the taller peak height and expressed as a percent. All data shown are mean of triplicate amplification reactions. Each locus is shown in an individual panel. Data was omitted for the Y chromosome marker DYS391 and cell lines that were homozygous or triallelic (HEK293 locus CSF1PO). Data in the shaded region (70–100%) show balanced alleles. The dashed line indicates a peak height ratio of 50% for reference.

five unique alleles detected (out of 12, relative to HEK293). The *GenePrint*® 24 System shows similar performance, and further expands the number of loci amplified to better discriminate between similar cell lines and to improve the ability to detect cross-contamination. Many of these loci are now included in other CLA databases, like Cellosaurus.

Using the core nine loci recommended by the ANSI/ATCC guidelines, we then calculated the percent match of these gDNA mixtures using the Masters formula. The Masters formula overlooks alleles missing in the test sample relative to the reference genotype, but is sensitive to extra alleles in the test sample, and therefore contamination. Using this formula (Figure 7), even 2% contamination of HEK293 with HeLa reduced the percent match below 80%, flagging the sample for manual review. On review, contamination would be obvious, detecting 4–6 novel alleles in 3–4 loci.

Many core facilities provide cell line authentication via mail service where laboratories can submit cell line samples spotted on sample storage cards. Using this same approach, HeLa cells were mixed with HEK293 cells down to 2.5% and spotted on Whatman® FTA® Classic Cards. Cell line DNA was then directly amplified from 1.2mm punches of these storage cards using the *GenePrint*® 10 System. As shown in Figure 8, the Spectrum Compact CE System could sensitively detect HeLa cell contamination. On average, four unique HeLa alleles were detected with as low as 2.5% cross-contamination, and cross-contamination could be identified by a low-percent match with as little as 2.5–5.0% HeLa cells (Figure 8).

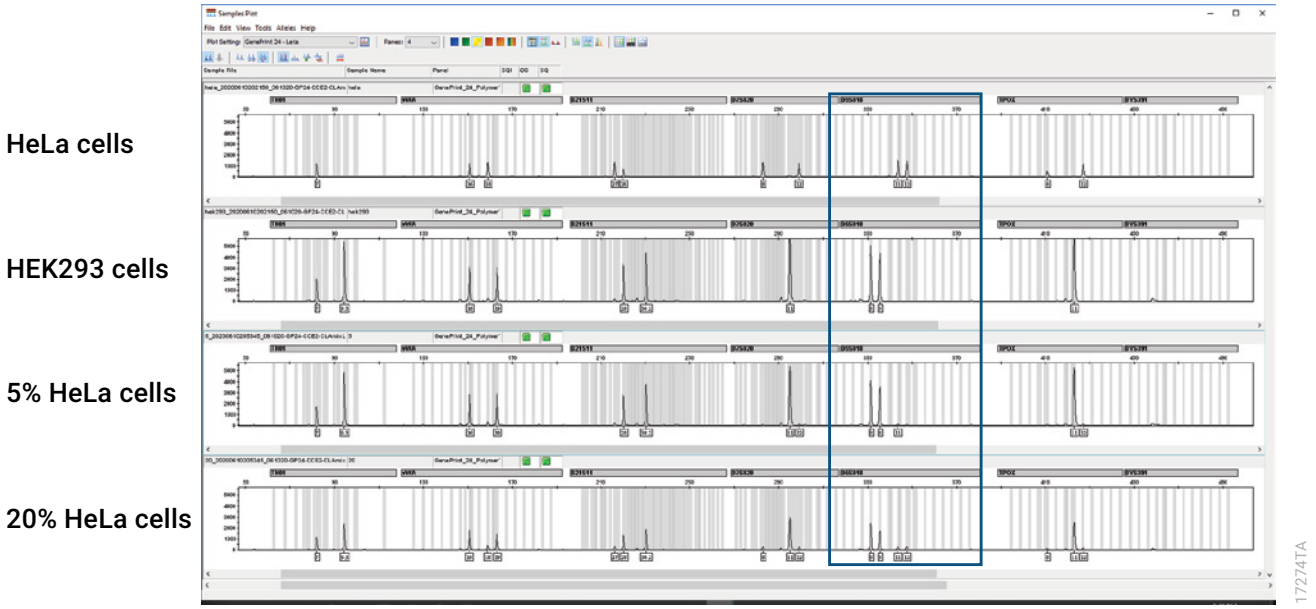

Figure 5. Example TMR (yellow) dye channel electropherograms showing the unique HeLa cell alleles detected in a cell mixture, amplified with the *GenePrint*® 24 System and analyzed on the Spectrum Compact CE System. Outlined in blue is the locus D5S818.

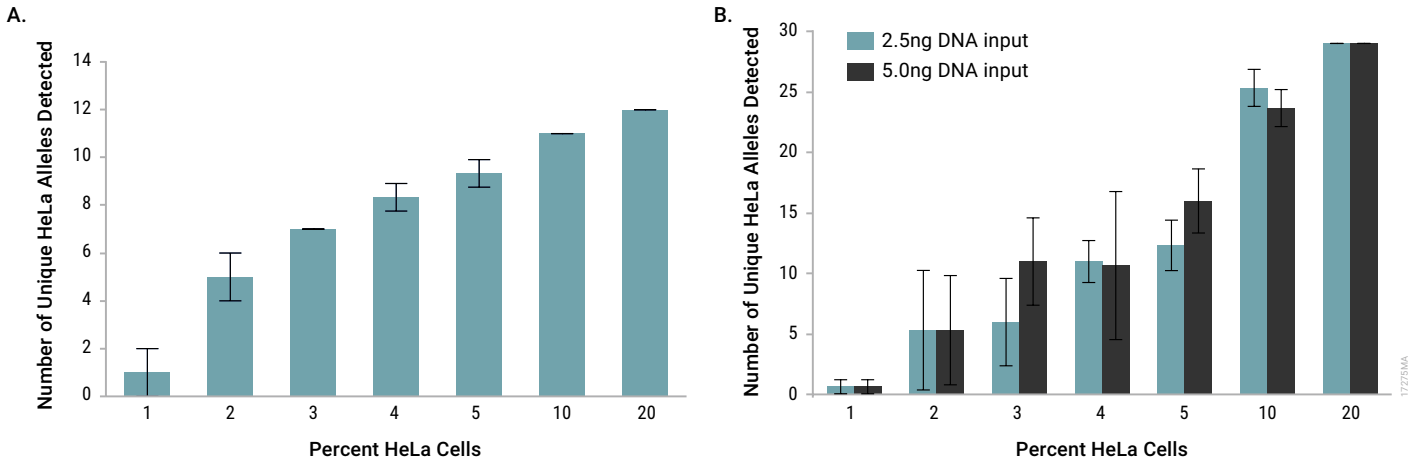

Figure 6. Detecting unique HeLa alleles in gDNA mixtures amplified with the *GenePrint*® 10 System (Panel A) or the *GenePrint*® 24 System (Panel B) and analyzed on the Spectrum Compact CE System. Compared to the HEK293 cell line, HeLa cells have 12 unique alleles when amplifying 10ng input gDNA with the *GenePrint*® 10 System and 30 unique alleles when amplified with the *GenePrint*® 24 System. Mean ± standard deviation shown, n = 3.

Conclusion

Cell line authentication is expected for grant applications submitted to the NIH and required for publication in many scientific journals. STR analysis is a simple, definitive process used for authenticating cell lines that assures researchers about the integrity and reproducibility of their data. The Spectrum Compact CE System, paired with our STR-based *GenePrint*® cell line authentication reagents, provides in-lab cell line authentication. The low- to medium-throughput instrument offers small-batch processing and minimizes reagent waste to accommodate individual labs or small lab clusters. And the sensitivity achieved with the Spectrum Compact CE System can help identify cell line cross-contamination before starting your experiments, preventing any negative consequences for your publications and reputation.

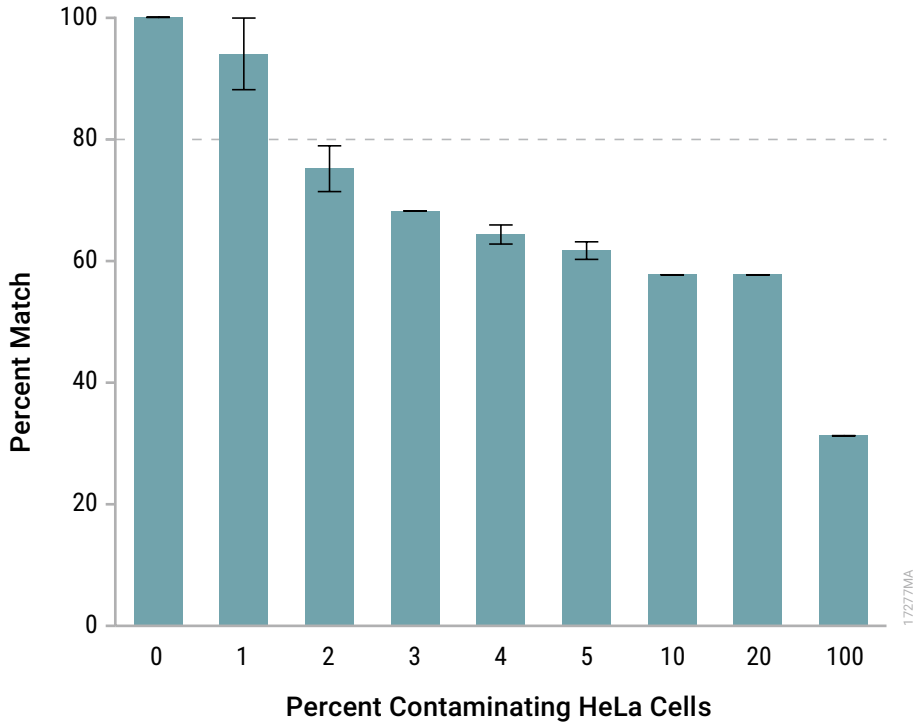

Figure 7. Percent match of mock-contaminated HEK293 cell line compared with the single-source genotype for HEK293, amplified using the *GenePrint*® 10 System and analyzed on a Spectrum Compact CE System. The alleles identified in the core eight STR loci and Amelogenin were compared to the genotype for single-source HEK293 cells used in this study. The percent match was calculated according to the Masters formula. Mean of n = 3 ± standard deviation shown. Matches >80% would be deemed related or matched, while matches <80% would be flagged as unrelated or contaminated samples.

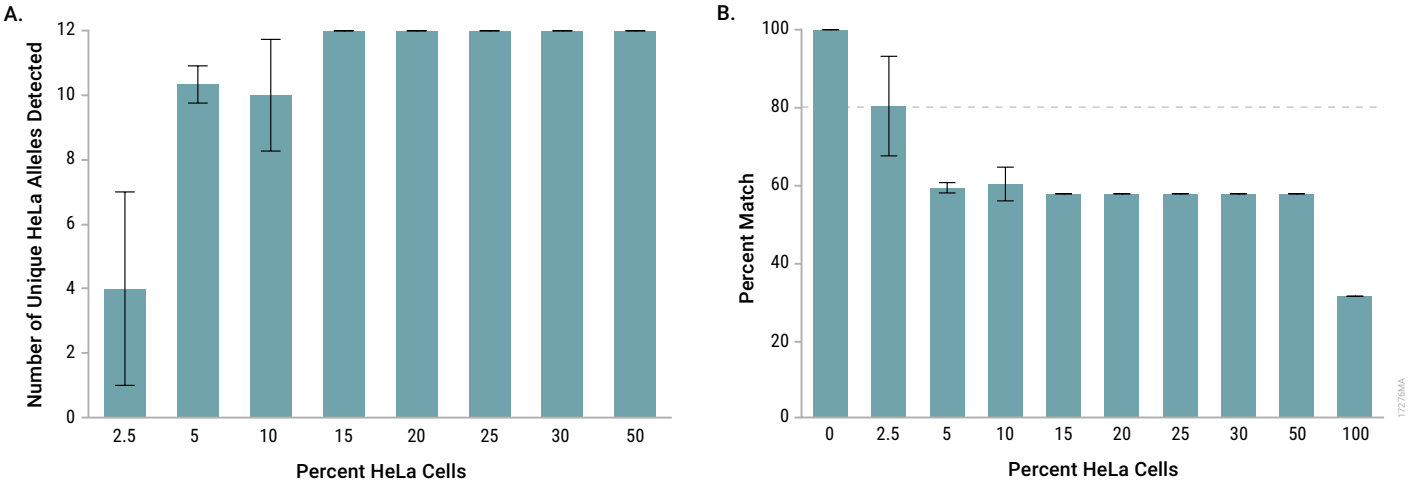

Figure 8. Detecting unique HeLa alleles in cell mixtures spotted on FTA® cards, directly amplified with the *GenePrint*® 10 System and analyzed on a Spectrum Compact CE System. Panel A. Number of unique HeLa alleles detected relative to the HEK293 cell line. **Panel B.** Percent Match calculated using the Masters formula. Mean ± standard deviation shown for n = 3 independent FTA® cards per mixture.

References

1. American Type Culture Collection Standards Development Organization Workgroup ASN-0002 (2010) Cell line misidentification: The beginning of the end. *Nat. Rev. Cancer* **10**, 441–8.
2. Horbach, S.P.J.M. and Halffman, W. (2017) The ghosts of HeLa: How cell line misidentification contaminates the scientific literature. *PLOS ONE* **12**, e0186281.
3. TRIP Services (n.d.) *Cell Line Authentication*.Retrieved August 12, 2020, from **https://research.pathology.wisc.edu/cell-line-authentication/**
4. American Type Culture Collection (2011) *Authentication of Human Cell Lines: Standardization of STR Profiling*. ANSI/ATCC ASN-0002-2011.
5. International Cell Line Authentication Committee (2020, March 25) *Register of Misidentified Cell Lines*. **https://iclac.org/databases/cross-contaminations/**
6. Capes-Davis, A. *et al.* (2013) Match criteria for human cell line authentication: Where do we draw the line? *Int. J. Cancer* **132**, 2510–9.
7. American Type Culture Collection (n.d.) *STR Database*. Retrieved August 11, 2020, from **www.atcc.org/STR_Database.aspx**

Ordering Information

Product	Cat.#
Spectrum Compact CE System*	CE1304
Spectrum Compact Capillary Cartridge, 4-Capillary, 36cm*	CE2340
Spectrum Compact Polymer4*	CE2304
Spectrum Compact Polymer7*	CE2307
Spectrum Compact Buffer*	CE2300
Spectrum Compact Cathode Septa Mat*	CE2301
Spectrum Compact Cathode Retainer*	CE2302
Spectrum Compact Strip Base & Retainer, 32-Well*	CE2332
Strip Septa Mat, 8-Well*	CE2308
GenePrint® 10 System*	B9510
PowerPlex® 4C Matrix Standard*	DG4800
GenePrint® 24 System*	B1870, B1874
GenePrint® 5C Matrix Standard**	B1930

*Not for Medical Diagnostic Use.

**For In Vitro Research Use Only. Not for Diagnostic Procedures.