12-1-2014

Authentication of Human Cell Lines and Detection of Human Cell Line Contamination by Cost-Effective Direct Amplification of DNA

Vikram Reddy Vemireddy
University of North Texas Health Science Center at Fort Worth, vemireddy598@yahoo.com

AUTHENTICATION OF HUMAN CELL LINES AND DETECTION OF HUMAN CELL LINE CONTAMINATION BY COST-EFFECTIVE DIRECT AMPLIFICATION OF DNA

THESIS

Presented to the Graduate Council of the

Graduate School of Biomedical Sciences

University of North Texas

Health Science Center at Fort Worth

Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

By

Vikram Vemireddy

Fort Worth, Texas

November 2014

ACKNOWLEDGEMENTS

I would first like to thank my family for their support, encouragement, and faith in me during my academic endeavors. I thank Dr. Rhonda Roby and my committee members, Dr. Patricia Gwirtz, Dr. Alakananda Basu and Dr. Raghu Krishnamoorthy at the University of North Texas Health Science Center for their guidance throughout this project. To Peter Mosley, Vijay Nagala, Smirthi Rajendaran, Mark Sprouse, Nicole Philips and Alessandra Alicea-Centeno, I give my sincerest thanks for your support and encouragement. I would also like to show my deepest gratitude to my classmates for the friendships we have made throughout the last two and half years. I could not have picked a better group with which to share this experience and I wish you all the best.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

A-T – Adenine-Thymine

AT - Analytical thresholds

ATCC – American Tissue Culture Collection

BCA – Bicinchoninic acid assay

DNA – Deoxyribonucleic acid

EDTA – Ethylenediaminetetraacetic acid

FTA – Fast Technology Analysis

GATA – Guanine- Adenine-Thymine-Adenine

HLA – Human leukocyte antigen

ICLAC – International Cell Line Authentication Committee

ID – Identification

PBS – Phosphate buffered saline

PCR – Polymerase chain reaction

RFU – Relative Fluorescence Unit

RPMI – Roswell Park Memorial Institute

RT-PCR – Reverse transcriptase polymerase chain reaction

SSM – Slipped strand mispairing

ST – Stochastic threshold

STR – Short tandem repeat

TBST – Tris-Buffered Saline and Tween 20

UNTHSC – University of North Texas Health Science Center

CHAPTER 1

INTRODUCTION

Human cell lines are important tools in scientific research. The American Tissue Culture Collection (ATCC, Manassas, VA) database has genetic profiles of a few thousand human cell lines that are used in research (1). Each cell line was collected from a single human source, labeled, and supplied for research use. Because these cells are from a single source, each cell line contains a unique genetic profile that can be used to differentiate it from other cells. The cells have been collected from various organs of the body, such as breast, prostate, liver and others. The use of multiple cell lines tends to lead to contamination between cell lines, and this contamination can happen anywhere from the start to the end of research work. The contamination of cell lines may result in a major impact on an investigator's credibility and financial resources, because the data published by the researchers must be retracted. This is why the International Cell Line Authentication Committee (ICLAC) recommends cell line authentication at several key intervals during the conduct of a research project, such as when the cell line stock has arrived, when the new experimentation is started, during research passage, and before publication of data (2). The recommendation for cell line authentication by the ICLAC has resulted in a major cost to investigators and has created a need for a simple and cost-effective method of verifying cell lines.

Cell line authentication is the process of addressing a possible contamination problem by confirming the identity of a cell line. Until now, investigators have used various methods, such as karyotyping, isoenzyme analysis and immunotyping and human leukocyte antigen (HLA) typing, to authenticate their cell lines, but without success (3). Recently, short tandem repeat

(STR) analysis has shown promise in identifying each cell line because each cell line has a unique genetic profile. In this analysis, STR loci are amplified and analyzed to discriminate between two cell lines. The STR loci sequences are four base repeat units that can be highly variable among individuals, which makes these genetic markers effective for human identification and cell line authentication purposes (4). The STR amplification process is complex and time-consuming, because it involves multiple steps, such as deoxyribonucleic acid (DNA) extraction, purification, and amplification. In order to make this STR amplification method simple and quick, manufactures have been making kits for direct STR amplification that do not require DNA extraction and purification. In the direct amplification method, DNA can be amplified from cells spotted on storage cards, such as the Human ID Bloodstain Card. The DNA can be stored, transported, and processed at room temperature (5, 6). These advantages make the direct amplification method an ideal choice for cell line authentication. Many laboratories (such as Promega Corp., Madison, WI) are performing direct STR amplification for cell line authentication at a high price. This high cost is mainly due to the use of expensive storage media, such as indicating FTA™ micro cards (GE Healthcare Life Sciences, Piscataway, NJ), and expensive kits, such as PowerPlex 18D System (Promega Corp., Madison, WI). This creates a huge cost burden on researchers, as they may have to allocate a large portion of their research funds to cell line authentication.

An ideal cell line authentication method should be simple, quick, and cost-effective. The forensic team at the University of North Texas Health Science Center in Fort Worth (UNTHSC) is a leader in forensic research and has experience in using other storage media, such as the Human ID Bloodstain Cards. These cards are less expensive and can be used in for cell line authentication. Less expensive kits, such as the *GenePrint*® 10 System (Promega Corp.,

Madison, WI) can be used to amplify DNA on the Human ID Bloodstain Card. The combination of Human ID Bloodstain Card and *GenePrint*® 10 System would make cell line authentication a more cost-effective method for investigators compared to the method used by Promega Corp.

The ATCC Standards Development Organization standard "ASN-0002" recommends the use of Amelogenin and the following eight STR loci for cell line authentication: CSF1PO, D5S818, D7S820, D13S317, D16S539, TH01, TPOX, and vWA (7). The discriminatory power is approximately 1.2×10^8 when using these markers, which is sufficient to discriminate between cell lines because there are only a few thousand human cell lines in use. The *GenePrint*®10 System amplifies nine STR loci, including the ASN-0002 loci (TH01, TPOX, vWA, Amelogenin, CSF1PO, D16S539, D7S820, D13S317 and D5S818) as well as D21S11. The cost for the *GenePrint*®10 System kit is currently $400. This kit can be used for 50 amplifications. The cost of the Huamn ID Bloodstain Card is $2.01. This card can be used for 4 samples. Thus, the direct amplification method can be offered to investigators at an affordable price by using the combination of the Human ID Bloodstain Card and the *GenePrint*® 10 System kit.

After the Polymerase Chain Reaction (PCR) amplification, the resulting amplified products are separated by capillary electrophoresis. Data are collected using the Data Collection Software v3.0 (Life Technologies, Carlsbad, CA). The data are imported to GeneMapper® *ID* software version 3.2 (Life Technologies, Carlsbad, CA) for analysis. The STR data are compared to the ATCC database, which was established by the ATCC Standard Development Organization. The database reports results as a percentage by counting the number of shared alleles between the STR profile generated from the samples and the STR profile in its database and then divide it by total number alleles in the STR profile generated from the samples (7). A cell line can be considered to be authenticated when greater than 79.5% of the alleles in its STR

profile match profiles in the database. In a small number of cell lines, such as cancer cells, STR profiles show a greater degree of instability, resulting in a percent match less than 80%. The unrelated cell lines generally show percent match figures of 55% or less. Based on this match threshold, cell lines that are below 56% match are considered unrelated (8).

Table 1: Comparison of Current Method vs. the Proposed Direct STR Amplification Method

Methods & Materials	**Current Method**	**Proposed Direct STR Amplification Method**
Storage Cards & Cost	Indicating FTA™ Card $6 per Sample	Human ID Bloodstain Card $0.50 per Sample
Kits & Cost	Powerplex®18D System $22 per Sample	Geneprint® 10 System $8 per Sample
STR Profiles Generated Meets ATCC ASN-0002 Standards	Yes	Yes
STR Profiles Generated Compatible with ATCC Database	Yes`	Yes

Abbreviations: STR = short tandem repeat; ATCC = American Tissue Culture Collection; ID = Identification; Current Method = Promega Corp Laboratories Method.

Table 1 compares the use of storage cards and kits commonly used, and the STR profiles generated by the current method using the Powerplex®18D System and the proposed direct STR amplification method developed at UNTHSC. The cost of cell line authentication can be reduced by $19.50 per sample by using the Human ID Bloodstain Card and the *GenePrint*® 10 System Kit. The STR profile generated by the proposed methods meets the ASN-0002 standard and is compatible with the ATCC Database. Thus, the proposed method using the combination of Human ID Bloodstain Card and the *GenePrint*® 10 System Kit for cell line authentication would be a cost-effective direct amplification method that can be used by investigators multiple times with ease.

CHAPTER 2

BACKGROUND

Human Cell Lines

Human cell lines are important raw materials used in scientific research and are initially collected from donors are called primary cells (3). Primary cells are cultured on growth media in a dish until they occupy all the space on the dish. The primary cells then have to be transferred to a fresh plate containing additional media in order to provide more room for them to divide. These transferred cells are called sub cultured cells. After the sub cultured cells occupy all the space on the dish, the cells have to be transferred again to another dish. After the first subculture, the cells that are grown are called cell lines. These cell lines are labelled, genetically profiled and supplied for research use.

Contamination of Human Cell Lines

An ever growing scientific community requires a supply of many human cell lines for their work. The human cell lines cannot be differentiated from one cell line from another line with the naked eye. A consequence of increased use of multiple cell lines and inability to differentiate through the naked eye are the risks of contamination of one human cell line with another human cell line. The investigators do not know whether a cell line is an actual cell line or not until they verify the cell lines using a validated method. The true identity of human cell lines is crucial in scientific research as many medical products are specifically targeted towards a particular cell (7). The contamination of human cell lines during research may eventually result in a risk of exposing human to unnecessary medical products. Although misidentification of cell lines can be avoided by using proper techniques in the laboratory, this avoidance may not necessarily assure a true

identity of a cell line. The fact that each human cell line is collected from an individual donor ensures that each human cell line has a unique genetic profile. This characteristic profile is used in the human identification process (7). Consequently, the unique genetic profile in each human cell line alerted researchers to the value and usefulness of the STR analysis method to authenticate cell lines. Thus, the STR analysis method is now used in human identification can be used for assureing the true identity of a cell line during research. Because cell line contamination can happen anywhere during the conduct of research, investigators are required to verify their cell line identity multiple times during research.

Short Tandem Repeats (STR) Profiling

The STR sequences in the DNA are the sequences consisting of a repeated nucleotide base pair sequence, and the repeat sequence used in the cell line authentication is four base pair repeat (e.g., GATA-GATA-GATA) (7). The regions containing the repeat sequences present throughout the DNA, and of which a few are selected for to use in cell line authentication purpose. The number of repetitions in each repeat sequence varies from individual to individual, thus, makes the basis for discriminating cell lines of different human origin. For cell line authentication, the ANS-0002 standard recommends the use of nine STR loci as there are a few hundred human cell lines currently in use (7). The STR regions are amplified and analyzed for the number of repetitions at each selected STR loci.

Cell Line Authentication Standards

The ATCC established the ANSI/ATCC ASN-0002-2011 standard in 2011 for cell line authentication (7). The main objective of this standard was to simplify the authentication process. This was achieved by establishing a searchable database containing all the STR profiles

of cell lines it supplies and recommend the minimum number of STR loci required to discriminate the human cell lines (7). The ANSI/ATCC ASN-0002-2011 standard recommends the use of eight autosomal STR loci and one sex determining loci for cell line authentication.

Current Methods of Authentication

The STR profiling is a complex process involves DNA extraction, purification, and quantification before PCR amplification. To simplify this process, many manufacturers have been manufacturing kits that can amplify DNA directly without prior DNA extraction, purification, and quantification steps. Many laboratories offer the direct amplification of DNA for cell line authentication service to investigators, but the cost of this service is very high. This high cost is due to these laboratories using expensive storage media, such as indicating micro FTA™ cards, and expensive kits, such as a Powerplex® 18D system for cell line authentication. Investigators need a more cost-effective method to authenticate their cell lines, because they may have limited funding resources available for their research, and only a universal practice of cell line authentication can reduce the frequency of use of misidentified cell lines (9). The current internship practicum project was designed to identify a cell line authentication method that is simple, quick, and cost-effective. For this purpose the Human ID Bloodstain Cards and the *GenePrint*® 10 System Kit were used in the direct STR amplification method. The STR profile generated by the direct STR amplification method was a complete STR profile and can be compared with ATCC database.

Purpose of Cell Line Authentication

The main purpose of cell line authentication is to preserve the value of scientific research by identifying false cell lines. The therapeutic options for some diseases require drugs that target a

particular receptor on the cells of diseases (7). This type of therapy requires true cell lines of that particular disease in the research study. If cell contamination occurs during the study, the drugs that are discovered will not work against the disease and in addition the patient will be exposed to unnecessary side effects of the drugs.

Practicum Project Problem/Specific Aims

This practicum project addresses the need for simple, quick, and cost-effective method for cell line authentication that can be used by investigators. It is proposed that such a cost-effective direct STR amplification can become the method of choice used to authenticate cell lines and identify human cell line contamination. The specific aims of this practicum project are:

Specific Aim 1- Develop and optimize a method for spotting cultured cell lines onto the storage card.

Specific Aim 2- Determine that the Human ID Bloodstain Cards are a useful storage medium for storing DNA at room temperature.

Specific Aim 3- Establish a direct amplification method to detect contamination of human cell lines.

Significance

Cell line authentication is an important aspect of scientific research because the time and money spent by researchers is large. Also, many journals and funding agencies demand that investigators authenticate their cell lines multiple times during the progress of their research (10, 11). The proposed method amplifies Amelogenin and eight STR loci for human cell line authentication. This method might be the ideal choice for cell line authentication as it is simple,

quick, and possesses a minimum risk of contamination. The ultimate remedy for the cell line contamination problem is checking cell line identity multiple times during a study by all investigators. The cost-effective direct STR amplification method using Human ID Bloodstain Cards and *GenePrint*® 10 system will be a useful tool for investigators during their research.

CHAPTER 3

MATERIALS AND METHODS

Sample Preparation

Cell lines are grown in suspension or adherent cultures. The culture solution containing cell lines in which cells grow without requiring an anchoring medium are called a suspension culture. Some cell lines that require an anchoring medium are grown on a solid medium, and the anchoring media containing the cell lines are called adherent cultures. Cell lines used for this practicum project were obtained from a suspension or an adherent culture currently used by UNTHSC laboratories. The cell lines from a suspension culture can be used directly after determining the concentration of the cells/μL. But, the cell lines from an adherent culture must be trypsinized into a suspension solution before they are used. In order to transform an adherent culture into a suspension solution, the surface of adherent cell culture was first rinsed with 5 mL of Phosphate Buffer Saline (PBS). After this rinsing, 0.05% trypsin/EDTA (ethylenediaminetetraacetic acid) solution (Gibco®, Life TechnologiesTM) was added to detach the cells. The RPMI-1640 Medium (ATCC, Manassas, VA) was then used to disperse the cells into a single cell suspension, and, following this dispersion, cell lines were counted using a hemocytometer (Hausser Scientific, Horsham, PA) and centrifuged at 500g for 5 min. After centrifugation, the cell pellet was resuspended in 50μL of RPMI-1640 Medium, and 20μL of

each suspension (consisting of either 5000 cells/µL, 12,500/µL, or 25,000 cells/µL) was applied onto a Human ID Bloodstain Card to create samples for the direct STR amplification.

Processing of Human ID Bloodstain Card Samples

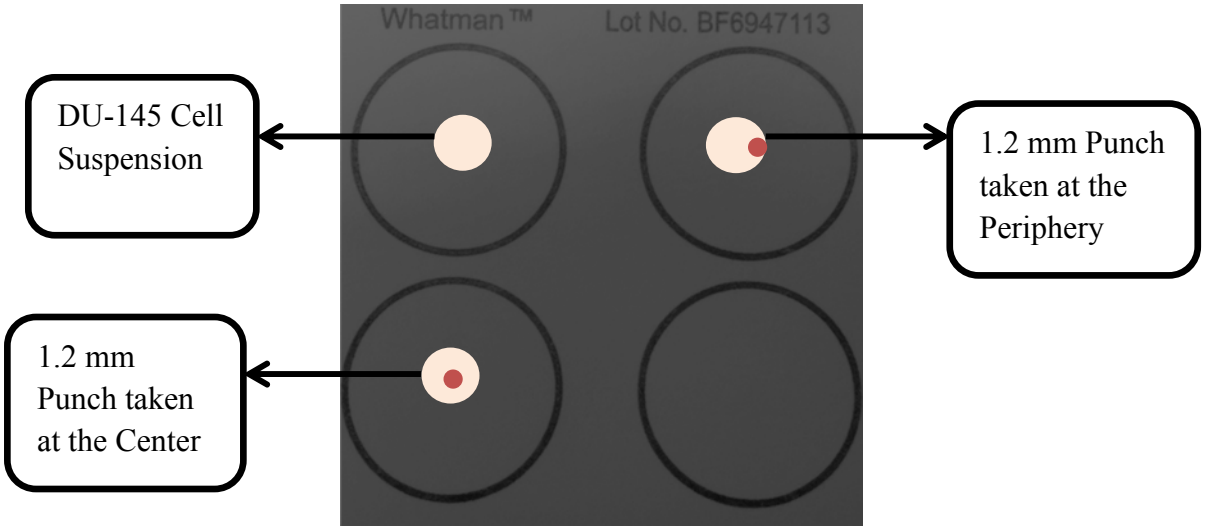
In the template pre-PCR hood, the cell suspension solution spot on the Human ID Bloodstain Cards was visualized with the Crime-Lite® 2 (Foster + Freeman Ltd, Sterling, VA). Two punches of 1.2 mm diameter were taken from the cards using a 1.2 mm Harris Micro-Punch™ (Ted Pella, Inc., Redding, CA). Initially, the punches were taken at the center of the circle on spot on each card. Samples from each concentration were amplified to determine the optimal concentration of cells to be spotted on the cards. Later, the 1.2 mm punches were taken in the periphery of the circle on the spot to determine the ideal card punch location.

Table 2: Human ID Bloodstain Card Samples

Sample #DU-145	**Cells/µL**	**Card Punch Location**
1 and 4	5000	Center & Periphery
2 and 5	12,500	Center & Periphery
3 and 6	25,000	Center & Periphery

Table 2 and Figure 1 shows the Human ID Bloodstain Card samples made from each concentration. Initially, samples 1, 2, and 3 consisted of punches taken at the center of the circle. To determine the best card punch location, samples 4, 5, and 6 consisted of punches taken in the periphery of the circle of the cell suspension spot. The positive and negative controls and the allelic ladder were included for each set of amplifications.

Figure 1. Human ID Bloodstain Card with the Spotted DU-145 Cell Suspension Solution and the Card Punch Location.



Each 1.2 mm punch was placed in 0.5 mL tubes containing 10µL of UNTHSC Incubation Buffer (UNTHSC, Fort Worth, TX: due to proprietary reasons, the composition of this buffer cannot be divulged). The samples were then incubated on a dry bath (AccuBlock™ Digital Dry Bath) at 70°C for 30 minutes or until dry. In the no template pre-PCR hood, the GenePrint® 10 PCR Amplification Mix was prepared by mixing 15 µL of double distilled water (ddH_2O), and 5 µL of *GenePrint*® 10 5x Master Mix and 5µL of *GenePrint*® 10 5x Primer Pair Mix (Promega, Madison, WI) for each Human ID Bloodstain Card and the negative control samples. For the positive control, the *GenePrint*® 10 PCR Amplification Mix was prepared by mixing 10 µL of ddH_2O, and 5 µL of *GenePrint*® 10 5x Master Mix and 5µL of *GenePrint*® 5x Primer Pair Mix.

Polymerase Chain Reaction Amplification (PCR)

In order to analyze the DNA by the 3130*xl* Genetic Analyzer (Life Technologies, Carlsbad, CA), it must be in sufficient quantity. In the PCR, the required segment of the DNA consisting of a

few nucleotide base pairs is amplified into multiple copies. Following the incubation of 1.2 mm sample punches, 25 µL of the *GenePrint*® 10 PCR Amplification Mix was added to the test tube containing the samples. Two control samples, positive and negative, were amplified along with Human ID Bloodstain Card samples. For the positive control test tube, 5 µL of positive control DNA (2800M) supplied with the *GenePrint*® 10 System Kit and the 20 µL of *GenePrint*® 10 PCR Amplification Mix prepared for positive control were added. To the negative control tube, only 25 µL of *GenePrint*® 10 Amplification Master Mix was added. The samples were briefly centrifuged and loaded onto a GeneAmp® PCR System 9700 (Life Technology, Carlsbad, CA). A PCR amplification of 26 cycles was performed using the thermal cycling parameters listed in Table 3.

Table 3: Amplification Parameters for the Direct Amplification Method

Number of Cycles	**Temperature**	**Time (minutes:seconds)**
HOLD	96°C	1:00
26	94°C 59°C 72°C	0:10 1:00 0:30
HOLD	60°C	10:00
HOLD	4°C	Infinity (∞)

Table 3 shows the thermal cycling parameters as outlined in the *GenePrint*® 10 System user manual for a PCR amplification of 26 cycles. Initially, the temperature was raised to 96°C and held for 1 min to dissociate double stranded DNA into two single strands. This was followed by 26 cycles of denaturing at 94°C for 10 s, annealing at 59°C for 1 min, and extension at 72°C for

30 s was performed. During the annealing step, the fluorescently labelled primers bind to the dissociated single-stranded DNA template. In the extension step, the Taq DNA polymearase synthesizes a new strand complementary to the single strand DNA by extending the primer attached to the single stranded DNA. After the last PCR cycle, the reaction temperature was held at 60°C for 10 min to finish any remaining single-stranded DNA fully extended. Finally, the temperature was lowered to 4°C. The amplified products were either stored at 4°C for a short time or immediately processed in a post-PCR hood for STR analysis.

Post-PCR Sample Preparation

After the PCR amplification, the samples were transferred to post-PCR hood. In the post-PCR hood, an aliquot was prepared by mixing 9 µL Hi-Di™ Formamide (Life Technologies, Carlsbad, CA) and 1µL of Internal Lane Standard (ILS) 600 supplied with the *GenePrint*® 10 System Kit for each sample. In the corresponding wells of each MicroAmp® Optical 96-Well Reaction Plate (Life Technology, Carlsbad, CA), 1 µL of *GenePrint*® 10 System allelic ladder or1 µL of amplified products were added. To each well of the plate containing an allelic ladder or sample, 10 µL of aliquot was added. The microamp plate was sealed and briefly centrifuged before loading onto the 3130*xl* Genetic Analyzer for capillary electrophoresis and data collection.

Capillary Electrophoresis and Data Collection

The *GenePrint*® 10 5x Primer Mix used to amplify the DNA contains primers labelled with fluorescent dyes. The primers are short strands consisting of a few nucleotide base pairs and complimentary to the initial starting base pairs at both ends of the DNA. These primers are required for synthesis of a new strand complementary to the DNA because the enzyme, DNA

Polymerase, that synthesizes the new strand can only synthesize by adding new nucleotides to the primers. The primers were labelled with either Blue (Fluorescein) or Green (JOE) or Yellow (TMR) dyes. The fluorescently labelled primers were extended by Taq DNA polymerase during the extension step, thus became part of the amplified product. In the capillary electrophoresis process, the amplified products were separated by their molecular weight. The capillary electrophoresis method consists of 16 submillileter capillaries filled with POP-6™ polymer, a 1X buffer with EDTA jar (Life Technology, Carlsbad, CA), a 1X buffer with EDTA reservoir, a POP-6™ polymer vial (Life Technology, Carlsbad, CA) and two ddH_2O reservoirs. The DNA fragments in the sample were injected into capillaries while a high voltage current is applied. The 3130*xl* Genetic Analyzer injection time and voltage were set at 5 Sec and 3 kV, respectively. The migration of the DNA fragments into the capillaries is called eletrokinetic migration because the negatively charged DNA fragments are injected into capillaries and start moving towards the positive electrode. The fluorescently labelled DNA fragments and the negative charge they carry forms the underlying mechanism involved in capillary electrophoresis. The polymer in the capillaries separates the DNA fragments by molecular weight. The smaller DNA fragments move faster than the larger DNA fragments through the polymer. A laser beam is used to fluoresce the dyes on the DNA fragments before the fragments reach the positive electrode. The fluorescent signals emitted from the DNA fragments are captured by a camera positioned near the positive electrode, and the fluorescent signals are converted into a digital data by 3130*xl* Genetic Analyzer Data Collection Software v3.0 (Life Technology, Carlsbad, CA).

STR Data Analysis

Short tandem repeats (STR) are repeated sequences of DNA (e.g., AATT, AATT, AATT). The number of repeats varies from individual to individual in several places of the DNA. The

variation in the number of repeats is useful in discriminating cell lines. The data were collected by Data Collection Software v 3.0 were imported to the GeneMapper *ID* software v 3.2 (Life Technologies, Carlsbad, CA) for analysis. The analytical threshold, stochastic threshold, peak height ratios, stutter percentages, and precision were calculated. The suggested criteria for determining the STR profile obtained from the samples was from a single cell line or from multiple cell lines are the number of STR peaks at a locus, the heterozygous peak height ratios and the stutter peak heights (12). After the data analysis, the STR profile generated was compared with the STR profile of the cell line in the ATCC STR database. Figure 2 shows STR profile of the positive control (2800M).

Figure 2. STR Profile of the Positive Control (2800M)

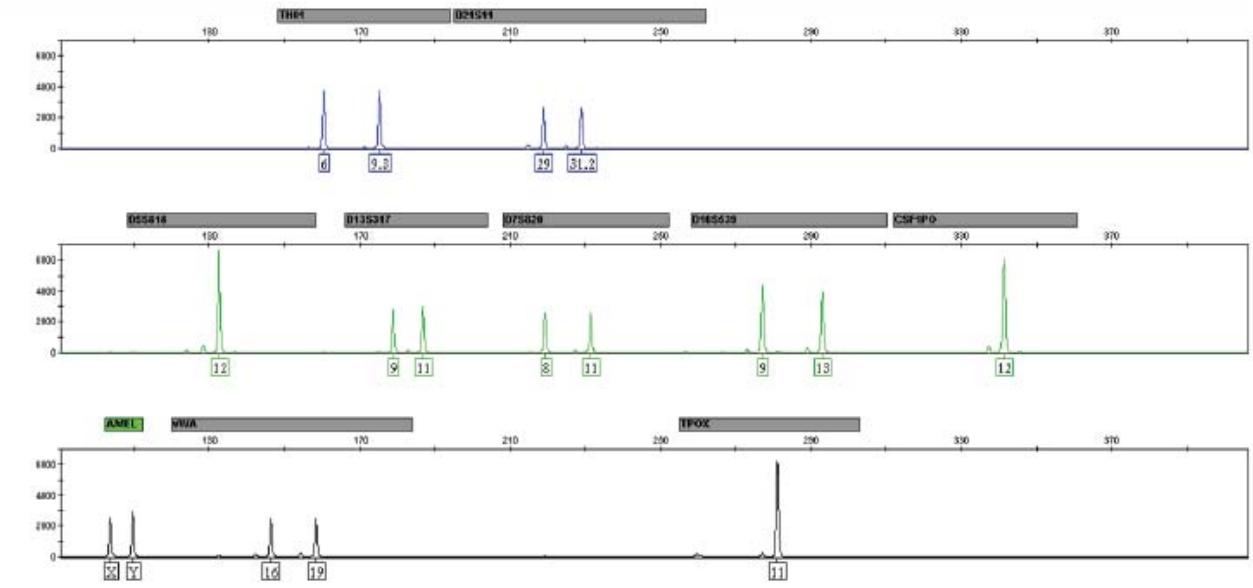


Figure 2 shows the STR profile of the positive control with all alleles of 10 loci. The expected alleles, TH01-6, TH01-9.3, D21S11-29, D21S11-31.2, D5S818-12, D13S317-9, D13S317-11, D7S820-8, D7S820-11, D16S630-9, D16S630-13, CSF1PO-12, AMEL-X, AMEL-Y, vWA-16,

vWA-19, and TPOX-11, were observed. Of these 10 loci, 3 were homozygous and 7 were heterozygous. The homozygous loci are CSF1PO, D5S818, and TPOX.

CHAPTER 3

RESULTS AND DISCUSSION

The Human ID Bloodstain Card samples were processed as described above. The punches taken at the center of the circle/spot on the card were amplified to determine the optimal concentration of cells to be spotted on the cards. The punches taken in the periphery of the circle/spot on the card were amplified to determine the best card punch location. An allelic ladder and a positive and a negative control were included with the amplification of card samples. Initially, the electrophreogram obtained from the allelic ladder and the positive and the negative control were analyzed to see the expected results. After confirming the expected results in the controls and allelic ladder, the STR profiles generated from the samples were analyzed. The STR peak heights from the Human ID Bloodstain Card samples were analyzed to determine optimal concentration of cells to be spotted and the best card punch location.

Analytical Threshold

Analytical thresholds (AT) are the minimum relative fluorescent unit (RFU) value, that all STR peaks must be greater than in order to be confirmed as true products of amplification. The analytical threshold should be determined for each post-PCR method used in the laboratory (13). The analytical thresholds can be calculated from the background noise peaks seen in electrophreogram of positive or negative control samples. Whenever the total quantity of DNA

greater than 1 ng is amplified, the background noise tends to increase. In these circumstances, the background noise peaks observed in positive controls are recommended to calculate analytical thresholds. When smaller quantities of DNA are amplified, the background noise peaks of negative controls can be used to calculate analytical threshold (13). In this study, the background noise of positive controls was used to calculate analytical threshold, because the quantity of input DNA was 50 ng. The mean peak height and standard deviation of background noise peaks were measured. The analytical threshold was calculated for each dye of the *GenePrint*® 10 System by adding 10 times the standard deviation to the mean (14).

Table 4: Mean, Standard Deviation and Analytical Threshold of Each Dye of the *GenePrint*® 10 System.

Dye	Average RFU	Standard Deviation	Analytical Threshold
Blue (Fluorescein)	10	5	60
Green (JOE)	12	6	72
Yellow (TMR)	8	6	68

Table 4 shows three dye colors (Blue, Green, and Yellow) used in the *GenePrint® 10* System for detection of nine STR loci. The analytical threshold was calculated for each dye. The highest analytical threshold (72 RFU) obtained of the three dyes was set as the analytical threshold for this practicum study (14).

Stochastic Threshold

The stochastic threshold (ST) is the minimum relative fluorescence units (RFU) value above which there is a high degree of confidence that if only one peak is observed, that peak is a true homozygous. Stochastic thresholds are calculated from the STR data of the heterozygous loci

with one allele being above the analytical threshold (i.e., False Homozygote) and the sister alleles below the minimum threshold. The mean and standard deviation of the false homozygous loci were calculated. The stochastic threshold was calculated by adding three times standard deviation to the mean (16). The mean and standard deviation of false homozygous loci observed in the samples were 123 RFU and 14 RFU, respectively. Therefore, the calculated stochastic threshold was 165 RFU.

Peak Height Ratios

The peak height ratios were calculated for all heterozygous loci (AMELEOGENIN, CSF1PO, D5S818, and D16S539) of DU-145 cell line. Normally, both alleles of a heterozygous loci are amplified equally, resulting in 100% peak height ratio (17). In some instances, such as low template DNA, peak height ratios fall below 60%. To analyze the STR data properly, the optimal heterozygous peak height ratio should be above 60% (18).

Stutter Peaks

Stutter products are most prevalent artifacts seen in an electrophreogram. The mechanism of stutter products is slipped strand mispairing (SSM) during PCR, resulting in either the insertion or deletion of one repeat unit on the new strand (12). In the STR analysis, stutter peaks that are greater in height than the set analytical threshold and stutter peaks that are exceeding the expected stutter percentage may mimic contamination of cell lines (7, 17). Several factors contribute to the occurrence of stutter products, such as number of STR repeats, the Adeninin – Thymine (A-T) content of the STR repeat, and interruption in STR repeats (19). The Promega Corp., (Madison, WI) provided expected stutter percentages for each loci amplified by

GenePrint® 10 System kit. Table 5 shows the average stutter percentages calculated by amplifying several DNA samples from different individuals of the general population.

Table 5: Expected Stutter Percentages of the STR Loci Amplified by *GenePrint*® 10 System Kit

Loci	**Stutter Percent (%)**
D21S11	22
TH01	6
CSF1P0	10
D5S818	11
D7S820	10
D13S317	12
D16S539	13
TPOX	6
vWA	14

Table 5 shows the expected stutter percentages provided by the Promega Corp (Madison, WI) for *GenePrint*® 10 System Kit. The percentages were calculated by amplifying several DNA samples from a population. The mean of the observed stutter peak heights of each allele obtained from the population samples were calculated. The stutter percentages were determined from the observed stutter peak height data (17). The expected stutter percentages were used during the analysis of the STR profile of Human ID Bloodstain Card samples. If the observed stutter percentage of the card sample was greater than expected stutter percentages of an allele, then the stutter peak was included and name and number were given to the stutter peak similar to true allele. If the observed stutter percentage of the card samples was less than the expected stutter percentage of an allele, then the stutter peak was excluded from the STR profile.

Precision

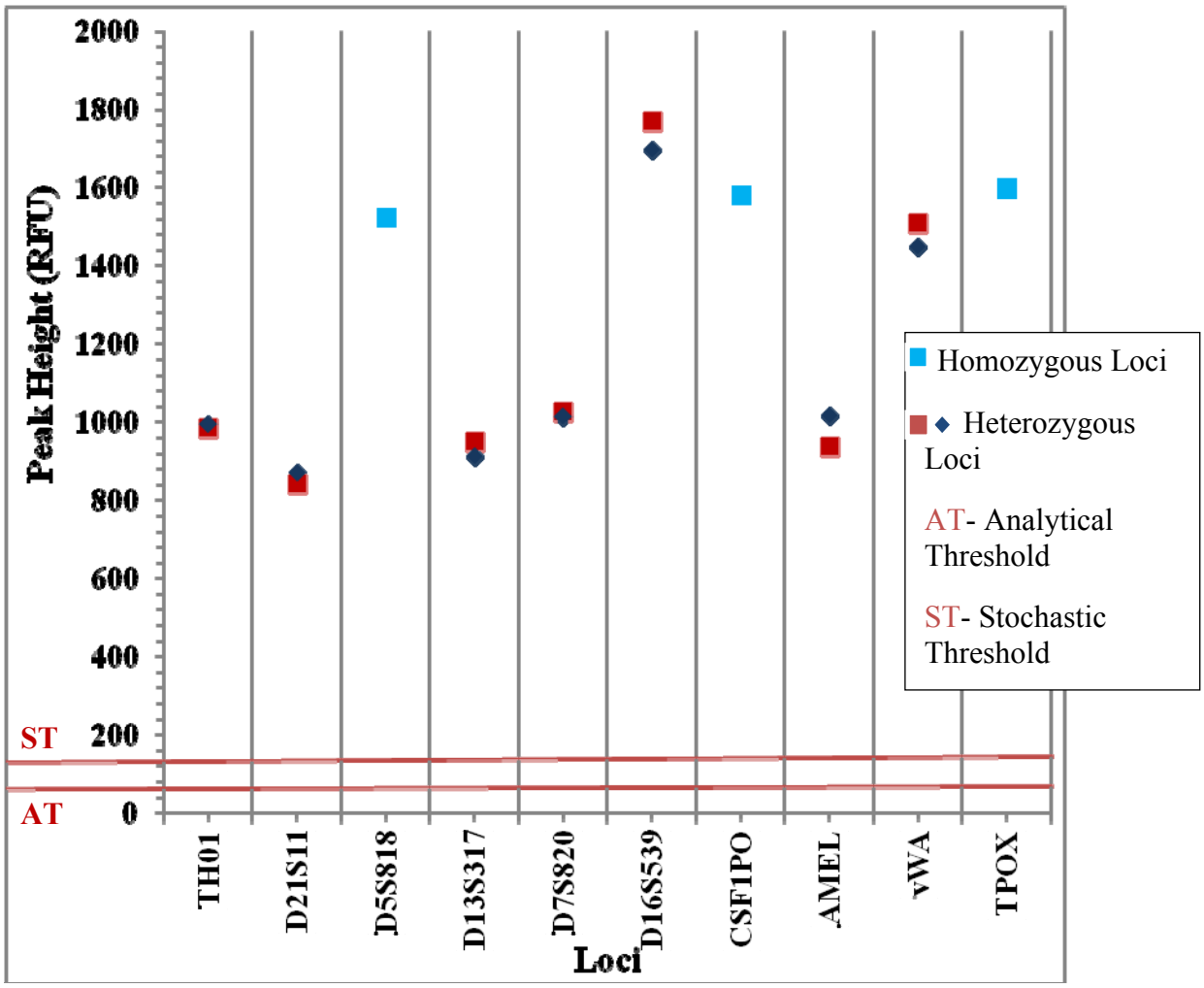
Each time a sample was subjected to capillary electrophoresis, the positive control and the allelic ladder were also included. The number of base pairs in each allele of all the allelic ladders and the positive controls that were run with samples were measured. The mean and standard deviation of the base pairs of each allele was calculated. The standard deviation should be less than 0.16 base pairs to have a high degree of precision (14). The *GenePrint*® 10 System Kit showed a high degree of precision with a standard deviation less than 0.15 base pairs for both the positive and the allelic ladder samples.

Analysis of STR Peak Heights Obtained from Positive Control Sample

In the positive control sample, 50 ng of extracted DNA (2800M) was amplified as described above. The STR peak height data in relative fluorescence units (RFU) obtained from the positive control sample is shown in Figure 3.

Figure 3 shows the STR peak height data of each allele of homozygous or heterozygous loci obtained from the positive control. The expected alleles of 10 STR loci (Ameleogenin, CSF1PO, D5S818, D7S820, D13S317, D16S539, D21S11, THO1, TPOX, and vWA) were observed. The allele peak heights were greater than the analytical threshold (72 RFU) and the stochastic threshold (165 RFU). The homozygous loci (D5S818, CSF1PO and TPOX) and the heterozygous loci (Ameleogenin, D7S820, D13S317, D21S11, THO1, and vWA) peak heights were between 750 RFU-2000 RFU. The peak height ratios of the heterozygous loci (Ameleogenin, D7S820, D13S317, D16S539, D21S11, THO1 and vWA) were above 60%. The peak heights were below the maximum threshold 8000 RFU of the 3130*xl* Genetic Analyzer (17).

Figure 3: The STR Peak Heights Measured from Positive Control



The STR profile generated from the positive control was a complete profile containing all alleles of 10 STR loci. The profile meets the acceptable criteria of ANS-0002 standards, such as the peak heights being 2 to 3 times above the analytical threshold, did not saturate the 3130*xl* Genetic Analyzer, and the allele STR peaks were given the expected name and number (7). Also, the STR profile meets the suggested criteria of the single source profile because only expected alleles were observed, the heterozygous peak height ratios were above 90% and there were no stutter peaks (12).

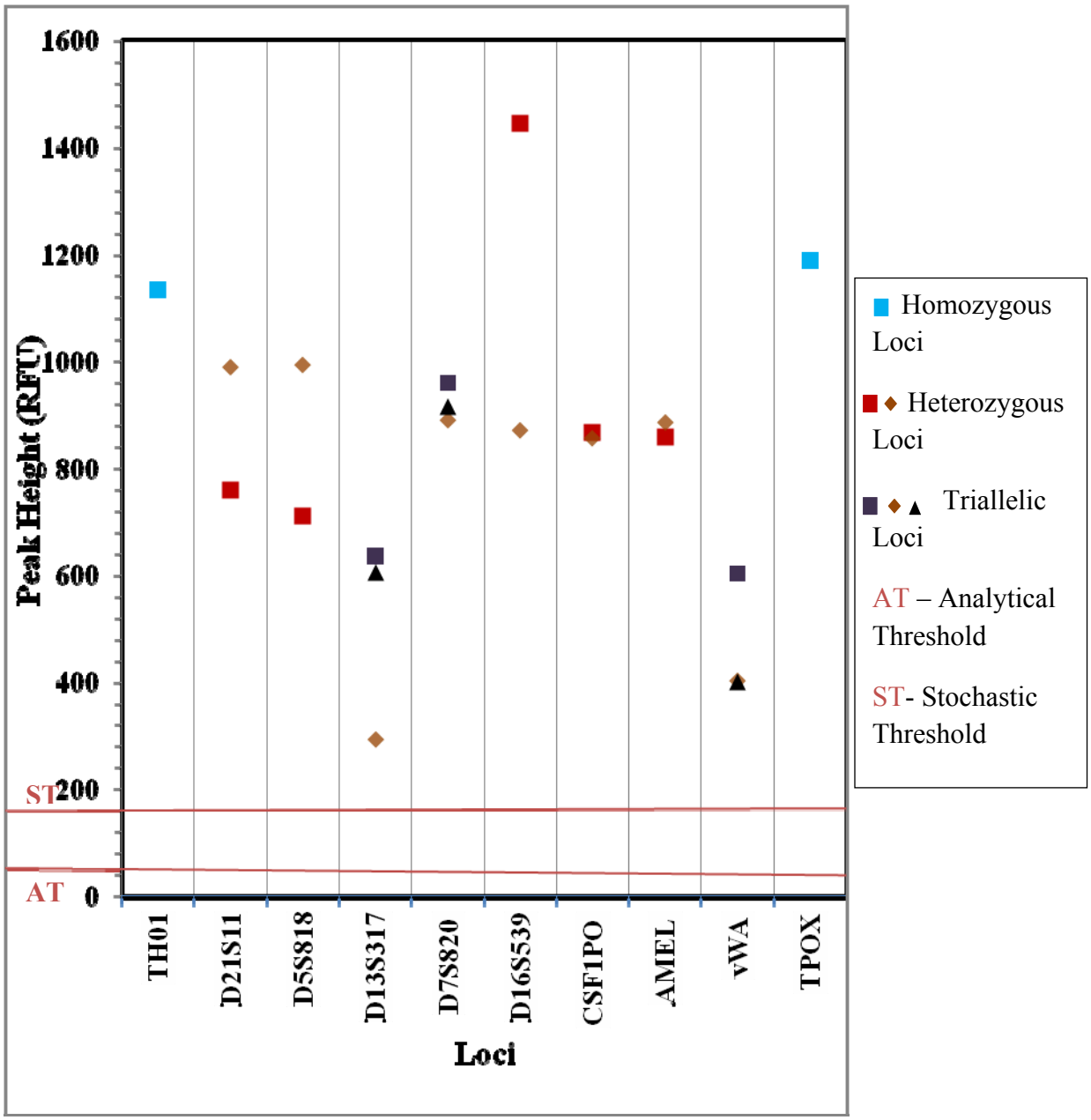
Analysis of STR Peak Heights Obtained from the Human ID Bloodstain Card Samples

STR Peak Heights from the Concentration of 5000 cell/µL

Two punches of 1.2 mm diameter were taken from the cards and processed as described above. The punch taken at the center of the circle was labelled as sample 1, and the punch taken at the periphery of the circle was labelled as sample 4. The STR peak height data in obtained from samples 1 and 4 are shown in Figures 4 and 5.

Figure 4 shows the STR peak height data of each allele obtained from sample 1. The expected 10 STR loci (Ameleogenin, CSF1PO, D5S818, D7S820, D13S317, D16S539, D21S11, THO1, TPOX, and vWA) were observed. The allele peak heights were greater than the analytical threshold (72 RFU), and the stochastic threshold (165 RFU). The homozygous loci (THO1 and TPOX) and the heterozygous loci (CSF1PO, Ameleogenin, D5S818, and D21S11) peak heights were between 750 RFU-1600 RFU. The peak height ratios of heterozygous loci (Ameleogenin, CSF1PO, D5S818, and D21S11) were above 60%, except for D16S539 (48%). The stutter products were observed for the alleles D5S818-13, D13S317-12, D16S539-11, and D16S539-13 were n-4 stutter products with stutter percentages were 10, 10, 6, and 12, respectively. The stutter percentages were less than the expected stutter percentages for the STR loci and the stutter peak heights were below the analytical threshold.

Figure 4: The STR Peak Heights Measured from Sample 1



The STR profile generated from sample 1 was a complete profile with all alleles of 10 STR loci, and the STR profile can be compared with the ATCC search database to verify the true identity of the cell line. The profile meets the acceptable criteria, such as the peak heights 2 to 3 times

above the analytical threshold, did not saturate the 3130xl Genetic Analyzer, and the allele STR peaks were given the expected name and number. The profile also meets the suggested criteria of single source profile such as, the number of STR peaks at each loci were as expected, the majority heterozygous peak height ratios were above 60 %, and the stutter peak heights were below the analytical threshold and the stutter percentages were below the expected stutter percentages.

Figure 5: The STR Peak Heights Calculated in Sample 4

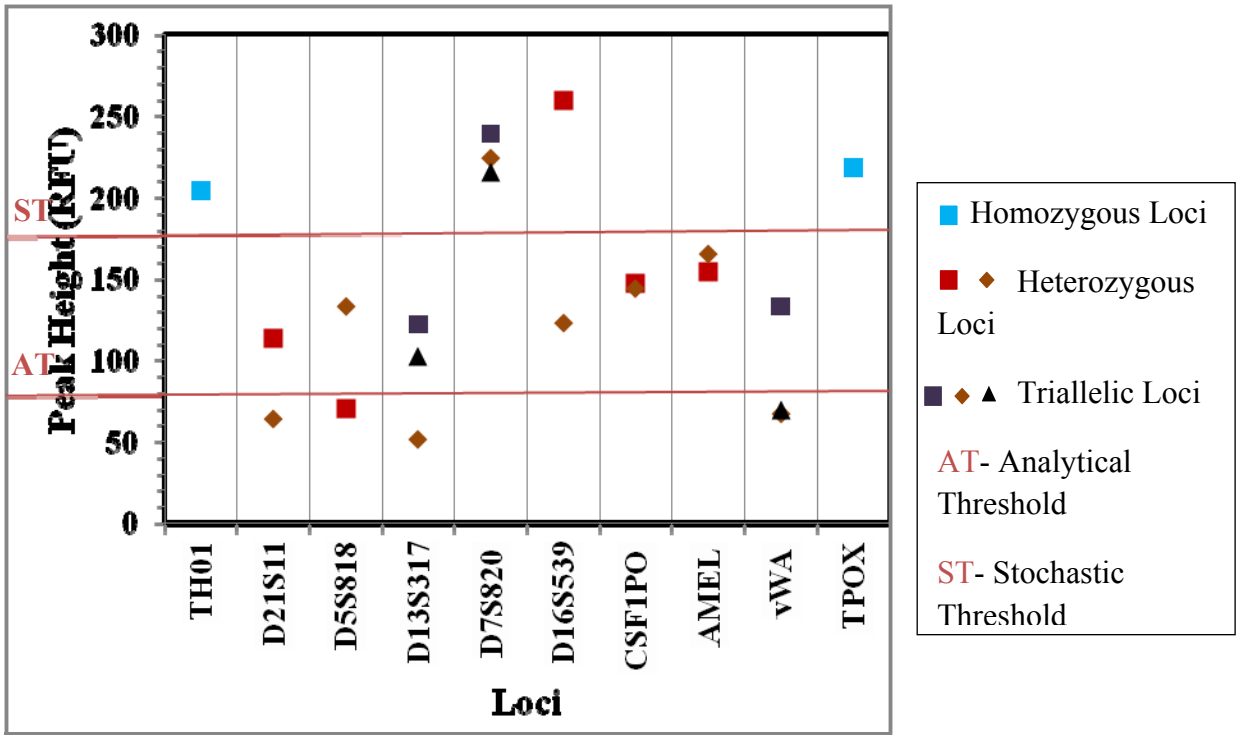


Figure 5 shows the STR peak height data of each allele obtained from sample 4. The alleles with peak heights less than the analytical threshold (72) were D21S11-33, D5S818-13, D13S317-13, vWA-18 and vWA-19. The heterozygous loci peak height ratios were above 60%, except for D16S539 (46%). The lower peak heights of sample 4 compared to sample 1 indicates that the

fewer cells were concentrated in the periphery of the circle than the center of the circle. The stutter products were not observed for the alleles.

The STR profile data obtained and shown in Figure 5 did not meet the acceptable criteria of the ASN-0002 standard because a few of the peak heights were below the analytical threshold, and the peak heights below the analytical threshold were not given a name and number. Thus, the profile obtained was a partial profile. The STR data indicate that the card punch location in the periphery of the cell suspension solution spot on the Human ID Bloodstain Card was not the ideal place to obtain a sample.

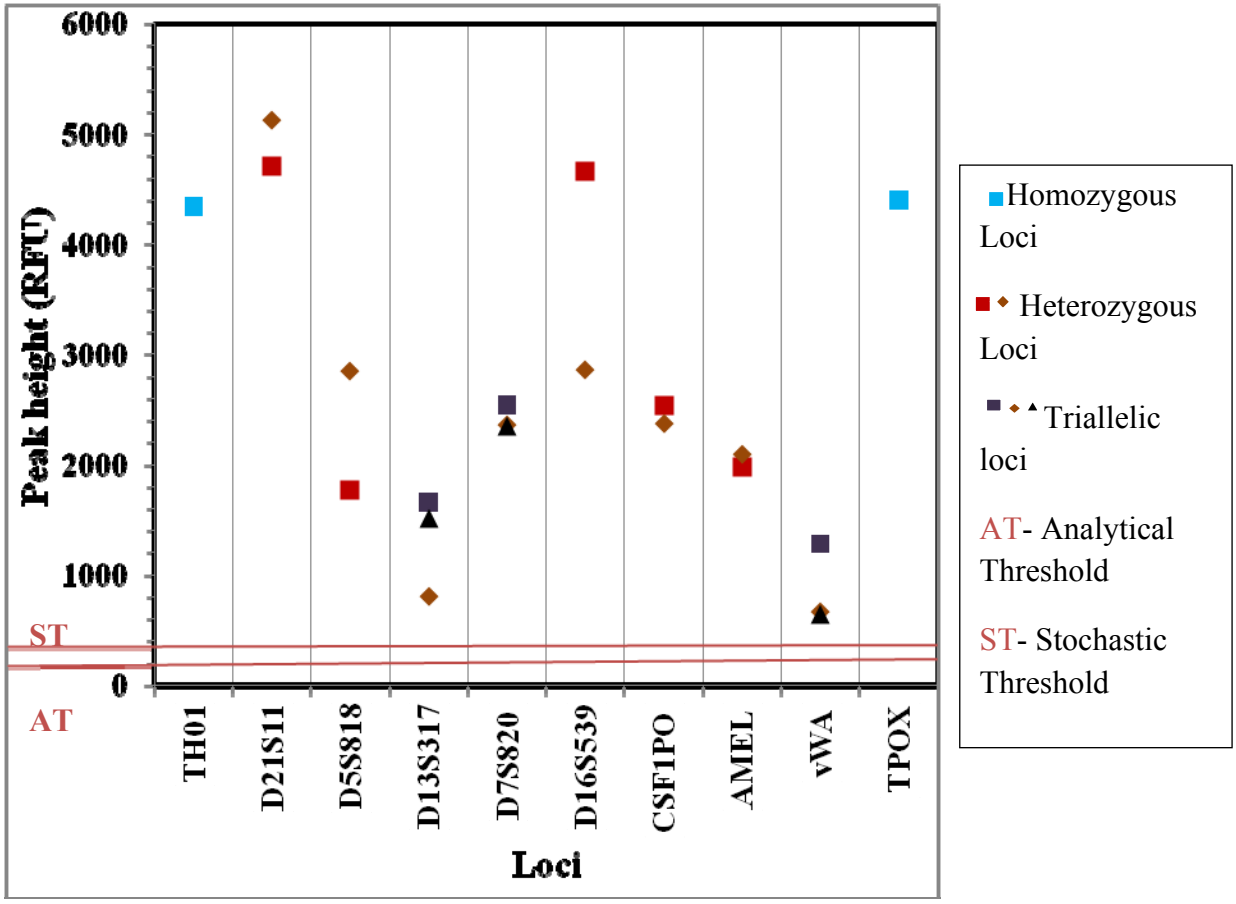
STR Peak Heighst from the Concentration of 12,500 cell/µL

Two punches of 1.2 mm diameter were taken from the card and processed as described above. The punch taken at the center of the circle was labelled as sample 2 and the punch taken at periphery was labelled as sample 5. The STR peak heights obtained from both samples are shown in Figures 6 and 7.

Figure 6 shows the STR peak height data of each allele obtained from sample 2. The expected 10 STR loci (Ameleogenin, CSF1PO, D5S818, D7S820, D13S317, D16S539, D21S11, THO1, TPOX, and vWA) were observed. The allele peak heights were greater than the analytical threshold (72 RFU), and the stochastic threshold (165 RFU). The homozygous loci (THO1 and TPOX) and heterozygous loci peak were between 1800-5500 RFU. The peak height ratios of heterozygous loci were above 60%, except for D16S539 (49%). The higher peak heights observed in sample 2 when compared sample 1 were due to a greater concentration of cells used in sample 2 compared to sample 1. The stutter products were observed for the alleles D5S818-10, D5S818-13, D7S820-10, D13S317-12, D16S539-11, D16S539-13 and THO1 were n-4

stutter products with stutter percentages were 14, 12, 12, 16, 15, 14, and 8, respectively. The stutter percentages were greater than expected stutter percentages for the loci. The stutter peak heights were greater than the analytical threshold.

Figure 6: The STR Peak Heights of Sample 2.



The STR profile shows expected alleles of all 10 STR loci plus a few additional allele peaks. The additional peaks were stutter peaks that were greater than analytical threshold. The additional peaks were given an allele name and number. The STR profile did not meet the suggested criteria of the single source profile (19) because additional peaks were observed than the expected peaks indicating possible contamination. The STR data indicate that concentration of

cells used in sample 2 was not the optimal concentration of cells to use for authentication of cell lines.

Figure 7: The STR Peak Heights of Sample 5

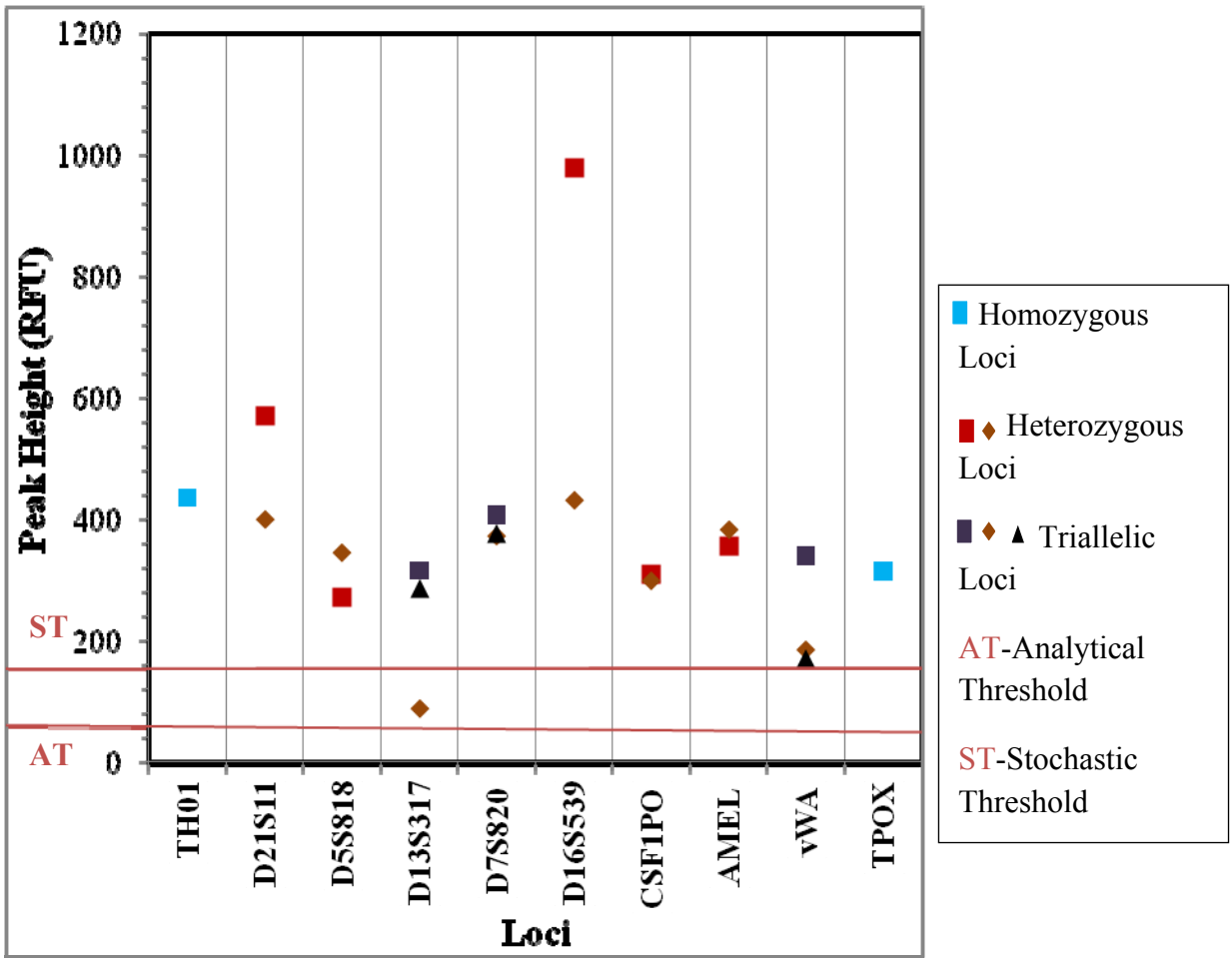


Figure 7 shows the STR peak height data of each allele obtained from sample 5. The expected 10 STR loci (Ameleogenin, CSF1PO, D5S818, D7S820, D13S317, D16S539, D21S11, THO1, TPOX, and vWA) were observed. The allele peak heights were greater than the analytical threshold (72 RFU), and the allele peak heights were greater than the stochastic threshold (165 RFU) except D13S317-13 allele. The homozygous and heterozygous peak heights in relative fluorescence units were between 200-900 RFU. The peak height ratios of heterozygous loci

(CSF1PO, Ameleogenin, D5S818, D16S539, D21S11) were above 60%, except for D16S539 (45%). The lower peak heights of sample 5 compared to sample 2 indicates that the fewer cells were concentrated in the periphery of the circle. The stutter peaks observed for the alleles D5S818-13, D13S317-12, D16S539-11, and D16S539-13 were n-4 stutter products with stutter percentages 11, 12, 6 and 13, respectively. The stutter percentages were not greater than the expected stutter percentages.

The STR profile was a complete profile showing all 10 STR loci. The STR profile data did not meet the acceptable criteria as some of the peak heights were not greater than 2 to 3 times of analytical threshold. The STR data indicate that fewer cells were concentrated in the periphery of the cell suspension spot on the Human ID Bloodstain Card and was not the ideal card punch location.

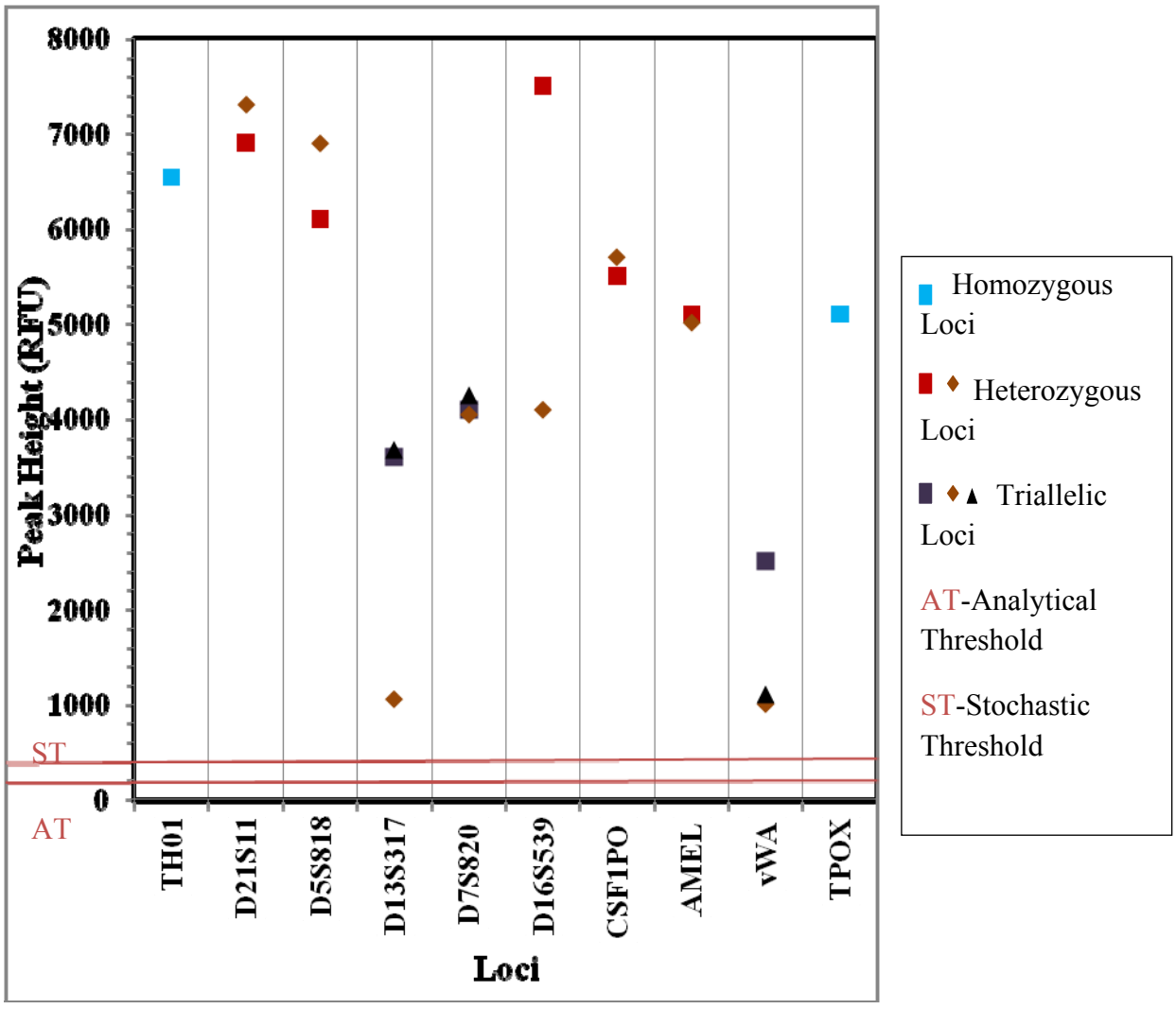
STR Peak Height from the concentration of 25,000 cell/µL

Two punches of 1.2 mm diameter were taken and processed as described above. The punch taken at the center of the circle was labelled as sample 3, and the punch taken in the periphery was labelled as sample 6. The STR peak heights obtained from both samples shown in Figures 8 and 9.

Figure 8 shows the STR peak height data of each allele obtained from sample 3. The expected 10 STR loci (Ameleogenin, CSF1PO, D5S818, D7S820, D13S317, D16S539, D21S11, THO1, TPOX, and vWA) were observed. The allele peak heights were greater than the analytical threshold (72 RFU), stochastic threshold (165 RFU). The homozygote and heterozygote loci peak heights were between 4500-8000 RFU. The peak height ratios of heterozygous loci were above 60%, except for D16S539 (49%). The stutter products were observed for the alleles

CSF1PO, D5S818-10, D5S818-13, D7S820-10, D13S317-12, D16S539-11, D16S539-13 and THO1 were n-4 stutter products with stutter percentages were 14, 16, 15, 14, 15, 15, 14 and 10, respectively. The stutter percentages were greater than expected stutter percentages for the loci. The stutter peak heights were greater than the analytical threshold.

Figure 8: The STR Peak Heights of Sample 3



The STR profile shows expected alleles of all10 STR loci plus a few additional allele peaks. The additional peaks were stutter peaks that were greater than analytical threshold. The additional peaks were given an allele name and number. The STR profile did not meet did not meet

suggested criteria of the single source profile because additional peaks were observed than the expected peaks indicating possible contamination. The STR profile did not meet acceptable criteria because some of the STR peaks reached the saturation point of the 3130*xl* Genetic Analyzer (8000 RFU). The STR data indicate that concentration of cells used in sample 3 was not the optimal concentration of cells required for accurate cell line authentication.

Figure 9: The STR Peak Heights of Sample 6

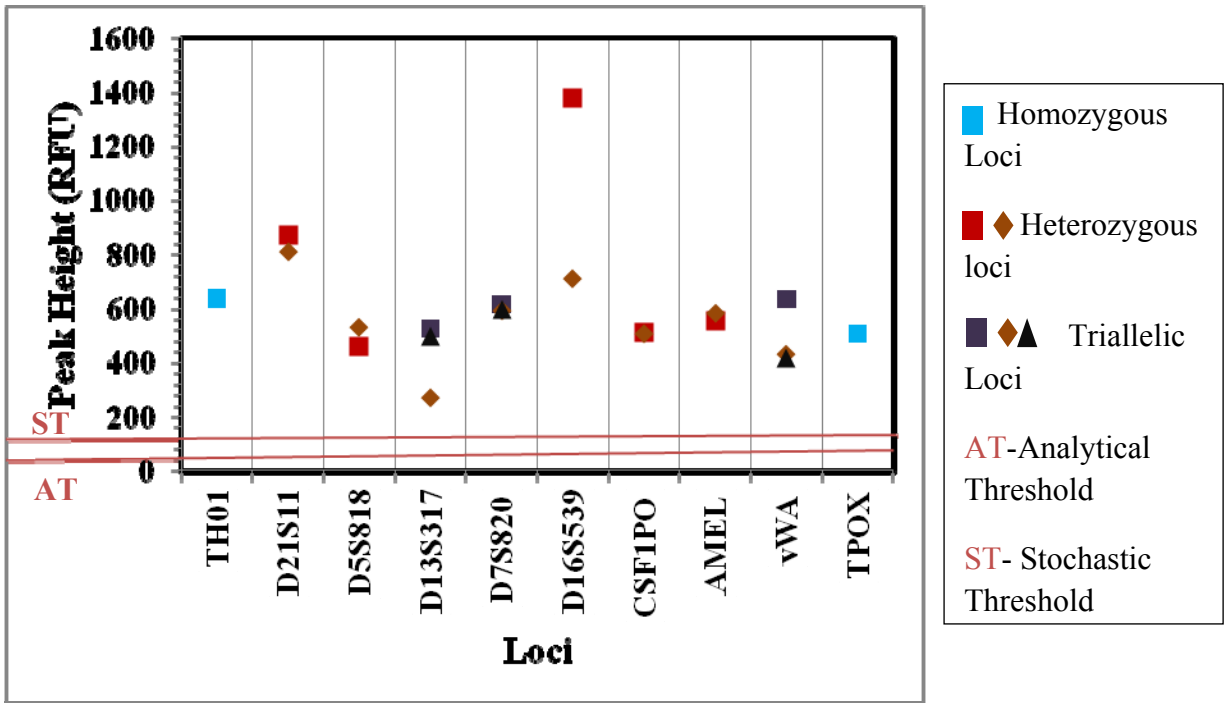


Figure 9 shows the STR peak height data of each allele obtained from sample 6. The expected 10 STR loci (Ameleogenin, CSF1PO, D5S818, D7S820, D13S317, D16S539, D21S11, THO1, TPOX, and vWA) were observed. The allele peak heights were greater than the analytical threshold (72 RFU) and greater than the stochastic threshold (165 RFU). The homozygote and heterozygote loci peak heights were between 450-1400 RFU. The peak height ratios of heterozygous loci (CSF1PO, Ameleogenin, D5S818, D21S11) were above 60%, except for

D16S539 (45%). The lower peak heights of sample 6 compared to sample 3 indicates that the fewer cells were concentrated in the periphery of the circle. The stutter products were observed for the alleles CSF1PO, D5S818-10, D5S818-13, D7S820-10, D13S317-12, D16S539-11, D16S539-13 and THO1 were n-4 stutter products with stutter percentages were 10, 9, 12, 10, 11, 12, and 6, respectively. The stutter percentages were not greater than expected percentages.

The STR profile was complete profile with all 10 STR loci. Even though, the STR profile meets the acceptable criteria of ASN-0002 and the suggested criteria of a single source profile, the STR peak height data indicate that fewer cells were concentrated in the periphery of the cell suspension spot on the Human ID Bloodstain Card and was not the best card punch location.

CONCLUSIONS

This practicum project used the Human ID Bloodstaind and the *GenePrint*® 10 System kit to establish the use of a cost-effective direct amplification method. The results showed that the direct STR amplification method using the *GenePrint*® 10 System and the Human ID Bloodstain Card produced a quality STR profile. The STR peak heights in relative fluorescence units obtained from each sample were analyzed and compared. The STR profile obtained from sample 1 showed all 10 STR loci. The homozygous peak heights were above 1000 RFU and heterozygous peak heights were between 750-1000 RFU. The peak height ratios of heterozygous loci were above 60%, except D16S539 (48%). The stutter percentages observed for each allele were below than the expected stutter percentage. This study showed that the parameters spotting 5000cells/μL in the center of the card and taking a center punch location as used in sample 1 are the optimal concentration and the best card punch location for amplification of samples using the proposed cost-effective method because the STR profile obtained in the sample met acceptable

criteria of ASN-0002 and the suggested criteria of a single source profile. Cell line authentication using the *GenePrint*® 10 System and the Human ID Bloodstain Card demonstrates that the authentication can be performed at a greatly reduced cost and will reduce the financial burden of conducting cell line authentication by investigators.

Limitations

STR profiling can discriminate human cell lines that are collected from different donors. However, this technique will not discriminate between cells that are collected from the different tissues of one donor. Also, the analysis of STR profiles from tumor tissue is sometimes complicated by their intrinsic genetic instability. The common genetic instability occurrences in human cancer cell lines are loss of heterozygosity, peak imbalance, and multiple peaks at several loci (4). Normally, a maximum of two alleles is observed at each locus, one from the mother and one from the father. Due to the intrinsic genetic instability characteristics of some tumor cell lines, it is observed that different possibilities of allele numbers, such as 1, 2, or 3, are at each locus. These characteristics require careful attention when analyzing the data, which is why the ATCC search database has three fields for STR loci instead of the two fields normally present for human reference samples. STR profiles that contain more than three peaks at more than three loci may be due to somatic mutation, trisomy, gene duplication or cellular contamination (7). When there is a situation where multiple peaks are observed, data must be analyzed with caution to identify any possible cell line contamination.

Future Studies

In addition to human cells, the ATCC supplies nonhuman cells from 150 different species for research use. During an investigation, these nonhuman cell lines can be contaminated by other

human or nonhuman cell lines. The contamination between any two different species, such as mice and rats, can be detected by amplifying a species specific marker. But this species specific marker will not determine to which individual mouse or rat, the cell line belongs to (20). It would be beneficial for researchers if a species specific STR analysis method and an effective STR database were established to identify contamination between two nonhuman cell lines belonging to the same species, such as between two mouse cell lines.

REFERENCES

1. Dirks WG, Drexler HG. Online verification of human cell line identity by STR DNA typing. *Methods Mol Biol*. 2011; 731:45–55.
2. Masters JR. Cell-line authentication: End the scandal of false cell lines. *Nature*. 2012; 492(7428):186.
3. Kerrigan L and Nims RW. Authentication of human cell-based products: The role of a new consensus standard. *Regen Med*. 2011; 6(2):255-260.
4. Reid Y, Storts D, Riss T, Minor L. Authentication of human cell lines by STR DNA profiling analysis. *Assay Guidance Manual*. Bethesda (MD). 2004.
5. Dobbs L. J, Madigan M. N, Carter A. B., Earls L. Use of FTA Gene Guard Filter Paper for the Storage and Transportation of Tumor Cells for Molecular Testing. *Arch. Pathol. Lab. Med*. 2002; 126:56–63.
6. Milne E, van Bockxmeer F. M, Robertson L, Brisbane J. M, Ashton L. J, Scott R. J, Armstrong B. K. Buccal DNA Collection: Comparison of Buccal Swabs with FTA Cards. *Cancer Epidemiol. Biomarkers*. 2006; 15:816–819.
7. ATCC SDO, Authentication of Human Cell Lines: Standardization of STR Profiling ATCC SDO document ASN-0002. Manassas, VA: ATCC Standards Development Organization. 2011.
8. Capes-Davis A, Reid YA, Kline MC, Storts DR, Strauss E, Dirks WG. Match criteria for human cell line authentication: Where do we draw the line? *Int J Cancer*. 2013; 132(11):2510-9.

9. Van Bokhoven, A., Varella-Garcia, M., Korch, C., and Miller, G. J. TSU-Pr1 and JCA-1 cells are derivatives of T24 bladder carcinoma cells and are not of prostatic origin. *Cancer Res*. 2001; 61:6340-6344.

10. Reid Y. Characterization and authentication of cancer cell lines: An overview. *Methods Mol Biol*. 2011; 731:35–43.

11. Perkel J. Curing cell lines. *Biotechniques.* 2011; 51:85–90.

12. Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM, Smerick JB, Budowle B. Validation of short tandem repeats (STRs) for forensic usage: Performance testing of fluorescent multiplex STR systems and analysis of authentic and simulated forensic samples. *J Forensic Sci*. 2001; 46(3):647-60.

13. J. Bregu, D. Conklin, E. Coronado, M. Terrill, R.W. Cotton, C.M. Grgicak, Analytical thresholds and sensitivity establishing RFU thresholds for forensic DNA analysis, *J. Forensic Sci*. 2013; 58(1):120–129.

14. Flores S, Sun J, King J, Budowle B. Internal validation of the GlobalFiler express PCR amplification kit for the direct amplification of reference DNA samples on a high-throughput automated workflow. *Forensic Sci Int Genet*. 2014; 10:33-9.

15. P. Gill, R. Puch-Solis, J. Curran, The low-template-DNA (stochastic) threshold – its determination relative to risk analysis for national DNA databases, *Forensic Sci.Int. Genet.* 2009; 104–111.

16. Gill P, Puch-Solis R, Curran J. The low-template-DNA (stochastic) threshold--its determination relative to risk analysis for national DNA databases. *Forensic Sci Int Genet*. 2009; 3(2):104-11.

17. Leclair B, Fregeau CJ, Bowen KL, Fourney RM. Systematic analysis of stutter percentages and allele peak height and peak area ratios at heterozygous STR loci for forensic casework and database samples. *J Forensic Sci*. 2004; 49(5):968-80.

18. Timken MD, Klein SB, Buoncristiani MR. Stochastic sampling effects in STR typing: Implications for analysis and interpretation. *Forensic Sci Int Genet*. 2014;11:195-204.

19. Brookes C, Bright JA, Harbison S, Buckleton J. Characterizing stutter in forensic STR multiplexes. *Forensic Sci Int Genet*. 2012; 6(1):58-63.

20. Holder M, Cooper P. Species identification and authentication of human and rodent cell cultures using polymerase chain reaction analysis of vomeronasal receptor genes. *Cytotechnology*. 2011; 63:553–558.