University of North Texas Health Science Center
UNTHSC Scholarly Repository

Theses and Dissertations

Summer 8-2009

Bone Marrow Engraftment Monitoring Using Mixture Deconvolution Software Designed for Forensic Casework.

Alexandra Newman
University of North Texas Health Science Center at Fort Worth, allynewman@gmail.com

Follow this and additional works at: http://digitalcommons.hsc.unt.edu/theses

Recommended Citation
Newman, A. , "Bone Marrow Engraftment Monitoring Using Mixture Deconvolution Software Designed for Forensic Casework." Fort Worth, Tx: University of North Texas Health Science Center; (2009).
http://digitalcommons.hsc.unt.edu/theses/43

This Internship Practicum Report is brought to you for free and open access by UNTHSC Scholarly Repository. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of UNTHSC Scholarly Repository. For more information, please contact Tom.Lyons@unthsc.edu.

BONE MARROW ENGRAFTMENT MONITORING USING MIXTURE DECONVOLUTION SOFTWARE DESIGNED FOR FORENSIC CASEWORK

INTERNSHIP PRACTICUM REPORT

Presented to the Graduate Council of the Graduate School of Biomedical Sciences

University of North Texas Health Science Center at Fort Worth

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

By

Alexandra Newman, B.S.

Fort Worth, Texas

August 2009

Newman, Alexandra, Bone Marrow Engraftment Monitoring Using Mixture Deconvolution Software Designed for Forensic Casework. Master of Science (Forensic Genetics), August 2009, 44 pages, 8 tables, 14 figures, 22 references.

Following a bone marrow transplant, patients are monitored closely for evidence of graft rejection or recurrence of the original disease. Bone marrow transplantation creates a donor-recipient cellular chimerism in the patient, which can be quantitively measured through short tandem repeat (STR) analysis of peripheral whole blood to determine the percent chimerism of the sample. Increasing recipient chimerism is an indication of graft rejection or relapse. Software programs designed to analyze forensic mixture samples have the potential to be useful in analyzing post-transplant mixed chimeric samples. Post-transplant samples were analyzed using three mixture deconvolution software programs. The programs were fast, accurate and consistent in determining the mixing proportions of the samples and the three programs gave concordant results.

ACKNOWLEDGEMENTS

I would like to give great appreciation to my major professor Dr. Joseph Warren and to my committee members Dr. John Planz and Dr. Arthur Eisenberg for their guidance and support throughout my studies at the University of North Texas Health Science Center. I couldn't have done this project without the data I received from Dr. Senitzer, Dr. Pereira, and Dr. Susskind. I would also like to thank Valerie Bostwick of the National Institute of Justice Expert Systems Testbed Project at Marshall University for sharing all of her technical knowledge with me. Finally, I would like to thank my project mentor Dr. Rhonda Roby for all of her help, encouragement, and support.

TABLE OF CONTENTS

page

LIST OF TABLES…….…….iv

LIST OF FIGURES…….v

CHAPTER

I. INTRODUCTION……...1

II. MATERIALS AND METHODS……...……10

III. RESULTS AND DISCUSSION…….....21

IV. CONCLUSIONS……...39

REFERENCES……..…42

LIST OF TABLES

page

1. Mixing proportions of the NEST Project mock mixture samples…..…….....…11

2. GeneMapper® *ID-X* calculation of observed mixing proportions……14

3. FSS-i^3 calculation of observed mixing proportions…….……..16

4. Average mixing proportion and mixing proportion per locus as reported by GeneMapper® *ID-X*, FSS-i^3, and DNA_DataAnalysis for the NEST Project samples…28

5. Comparison of average mixing proportions across 16 loci for the proficiency test samples as calculated by GeneMapper® *ID-X*, FSS-i^3, DNA_DataAnalysis, ASHI, and SCCA……..32

6. Case #1, proportion of recipient cells present in post-transplant whole blood and CD3 T-cell subpopulation……..…35

7. Case #2, proportion of recipient cells present in peripheral whole blood at varying times post-transplant……36

8. Comparison of the average mixing proportions at 16 loci before and after the removal of stutter…….…..37

LIST OF FIGURES

page

1. Diagram of hematopoiesis from a multipotent hematopoietic stem cell…….……2

2. Possible clinical responses to increasing mixed chimerism……4

3. Schematic representation of an STR locus for a heterozygous individual……..5

4. Schematic representation of possible donor and recipient allele patterns with corresponding formulas for mixing proportion calculations..……...…….....7

5. Flowchart of functions performed by DNA_DataAnalysis…….……..20

6. Electropherogram of ASHI sample 2295-EMO-68 which was processed with three mixture deconvolution programs……..…22

7. Sample output of selected and unselected genotype combinations tables from GeneMapper® *ID-X*…….....23

8. Example of a GeneMapper® *ID-X* uninformative locus…….…….….24

9. Sample output of possible genotype combinations and mixing proportions from FSS-i^3 ……....25

10. Sample output from DNA_DataAnalysis of the mixing proportions at four loci..……26

11. Sample output from GeneMapper® *ID-X* for ASHI sample 2241-EMO-5 at the D3S1358 locus……...33

12. Sample output from FSS-i^3 for ASHI sample 2241-EMO-5 at the D3S1358 locus……..34

13. Sample output from DNA_DataAnalysis for ASHI sample 2241-EMO-5 at the D3S1358 locus……...34

14. The effect of stutter on a minor allele for ASHI sample 1875-EMO-15……...……38

CHAPTER I

INTRODUCTION

Bone Marrow Transplantation

Bone marrow transplantation is the transplant of hematopoietic stem cells and is used to treat many diseases including leukemia, severe combined immunodeficiency disease (SCID), severe aplastic anemia, and lymphoma [1, 2]. The procedure involves ablation of the recipient's bone marrow using chemotherapy or full body irradiation. The complete ablation of the recipient's bone marrow is called myeloablative conditioning, while the partial ablation is called non-myeloablative conditioning. Non-myeloablative conditioning uses a lower dose of chemotherapy or irradiation and is primarily used for those who could not survive the myeloablative routine, such as the elderly. The recipient bone marrow ablation routine is followed by transplantation of the donor's hematopoietic stem cells, which are most commonly removed from the hip bone using a needle. Alternatively, hematopoietic stem cells can be collected from peripheral whole blood by continuous-flow apheresis after administering recombinant stimulating factor, although transplant of peripheral blood stem cells has been shown to increase the risk of graft-versus-host-disease (GVHD) [3]. The donor's hematopoietic stem cells will then repopulate the blood's cellular components in approximately 9-14 days, as seen in Figure 1. In addition, the donor cytotoxic T-cell and natural killer cell's "graft-versus-tumor" effect will eradicate any remaining diseased recipient cells.

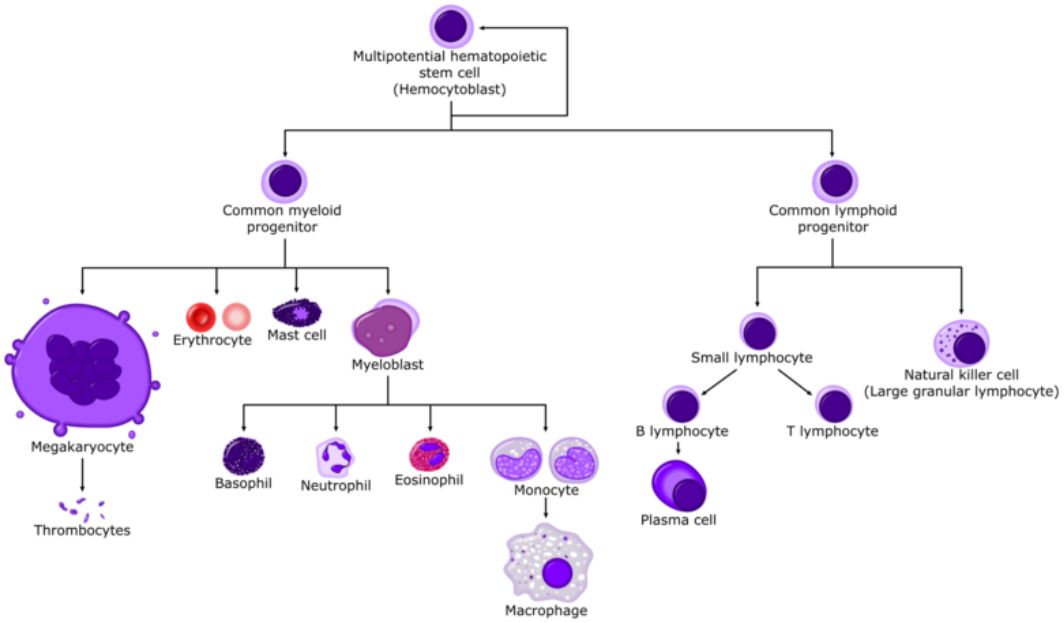

Figure 1: Diagram of hematopoiesis from a multipotent hematopoietic stem cell [4].

To decrease the risk of graft rejection and GVHD, the donor and recipient should have identical alleles in the human leukocyte antigens (HLA) genes [2]. The HLA genes fall into the Type I and Type II categories. In general, Type I mismatches result in an increased risk of graft rejection, whereas a Type II mismatch results in an increased risk of GVHD [3]. Patients that have undergone bone marrow transplantation are carefully monitored for evidence of graft rejection, GVHD, and recurrence of the original disease. Engraftment monitoring allows a physician to predict these negative events so that preventative action can be taken [1, 2, 4-6].

The goal of bone marrow engraftment monitoring is to determine what percentage of the hematopoietic cells in the patient are derived from the healthy donor bone marrow and what percentage of the hematopoietic cells are derived from the diseased host bone marrow. In the early post-transplant period, a dynamic relationship exists between the engrafted donor cells and the patient cells, resulting in mixed chimerism. The term “chimera” describes an individual with cells which derived from two distinct zygotic lineages. A complete changeover of the bone

marrow to the donor's hematopoietic and lymphoid cells is called complete chimerism. A partial changeover, called mixed chimerism, occurs when there are both donor and recipient hematopoietic and lymphoid cells present in varying ratios [1, 2, 4-6]. In addition, it is possible for patients to have complete chimerism in one sub-population of cells, such as natural killer cells, while having mixed chimerism in another cell sub-population. This is known as split chimerism [6].

Patients are tested at specific intervals to determine if the mixed chimerism is increasing, decreasing, or stable. Increasing mixed chimerism is defined as "increasing recipient DNA compared to the foregoing sample by at least 5%". Patients with complete chimerism or low-level stable mixed chimerism have a better survival rate and lowered occurrence of relapse of the original disease. Patients with increasing mixed chimerism may be at an increased risk of graft rejection, recurrence of the original disease, or GVHD. In addition, the quantity of recipient's cells in the blood can impact the course of therapy, such as cessation of immunosuppressive drugs or repeat donor lymphocyte infusion [6].

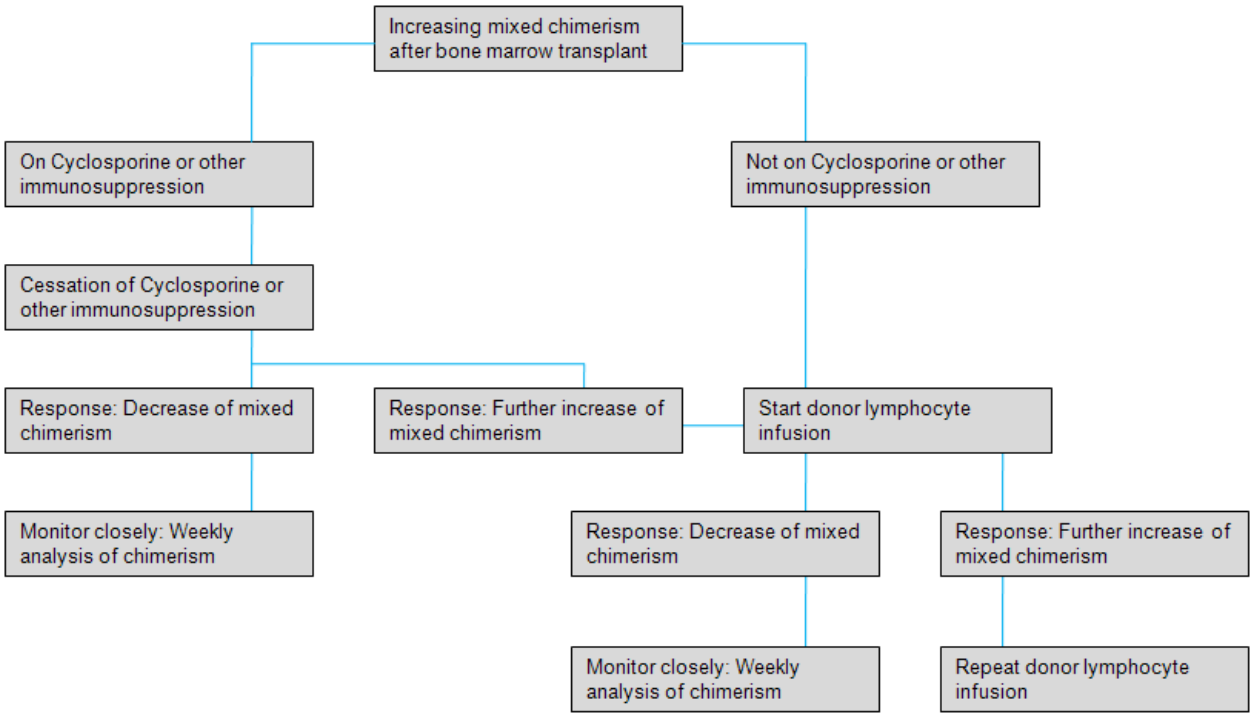

Figure 2: Possible clinical responses to increasing mixed chimerism [6].

History of Engraftment Monitoring

Since the first successful bone marrow transplant, there have been several methods for chimerism monitoring. The earliest method involved the detection of protein-based differences between the donor and the recipient, such as red blood cell antigen typing of the ABO, Rh, Duffy, or Kidd alleles and HLA disparities. However, it is rare that a transplant is performed between an ABO or HLA mismatched pair, so these methods have limited use. Other protein-based methods include detection of leukocyte enzyme polymorphisms and immunoglobulin allotype differences. The main drawback of these methods is that the number of protein polymorphisms is small; therefore, it is unlikely that an informative marker will be found between many donor-recipient pairs. To overcome this limitation, assays which detect genetic disparities between the donor and recipient were developed. Initial methods detected disease-

specific nuclear translocations, microscopic identification of the Y-chromosome in sex-mismatched transplants, and restriction fragment length polymorphism sites. The major limitations of these assays were low sensitivity or only being applicable to sex-mismatched pair transplants [7].

Since the early 1990's, short tandem repeat (STR) analysis has become the most widely accepted method to evaluate chimerism. STRs, or microsatellites, are tandemly repeated units that are 2-6 basepairs in length which are found throughout the genome. The number of times a particular STR marker repeats can be highly variable between individuals. STR analysis involves PCR-based amplification of specific loci on the genome using fluorescently labeled primers [7]. The primers are designed to bind in the region flanking the STR repeat. The amplified fragments are then separated by size using a slab gel or capillary electrophoresis. During electrophoresis, a high-energy laser excites the fluorophore, which then emits light at a lower wavelength. A charge-couple device is used to collect and amplify the signal from the flurophore and convert it to an electronic signal, which is displayed visually as an electropherogram. The advantages of STR analysis over previous chimerism analysis methods are the high polymorphism which allows for greater discrimination, high sensitivity, and wide availability of commercial kits [8].

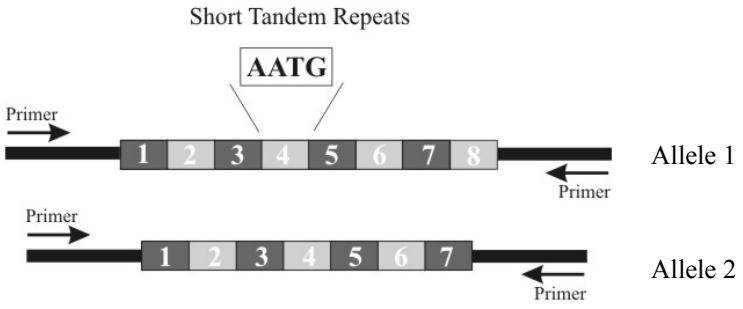

Figure 3: Schematic representation of an STR locus for a heterozygous individual. Allele 1 represents an 8 repeat unit. Allele 2 represents a 7 repeat unit. The simple STR repeat is represented by AATG.

STRs and Engraftment Monitoring

In STR analysis, peripheral whole blood is drawn weekly, bi-weekly or monthly post-transplant to evaluate chimerism. Once complete chimerism has been established, chimerism analysis can be performed at greater intervals, such as four times per year [6]. Depending on the disease of the patient, the laboratory may choose to analyze the peripheral whole blood or to first separate out specific cell sub-types, such as CD3 or CD19 T-cells, before analysis. The DNA is then extracted, quantified, and amplified using either a commercial kit or using in-house primers. Quantification of the percentage of donor and host DNA in the post-transplant sample is calculated manually using peak area comparisons as seen in Figure 4. If an allele is shared, only the allele which distinguishes the donor from the recipient is used [1]. Alternatively, peak height can be used to determine the percentage of donor and recipient DNA in the sample. A 2008 College of American Pathologists Monitoring Engraftment survey of 79 laboratories found that 53.2% of laboratories are using peak height, while 46.8% of laboratories are using peak area to calculate the mixing proportions of donor and recipient cells [9].

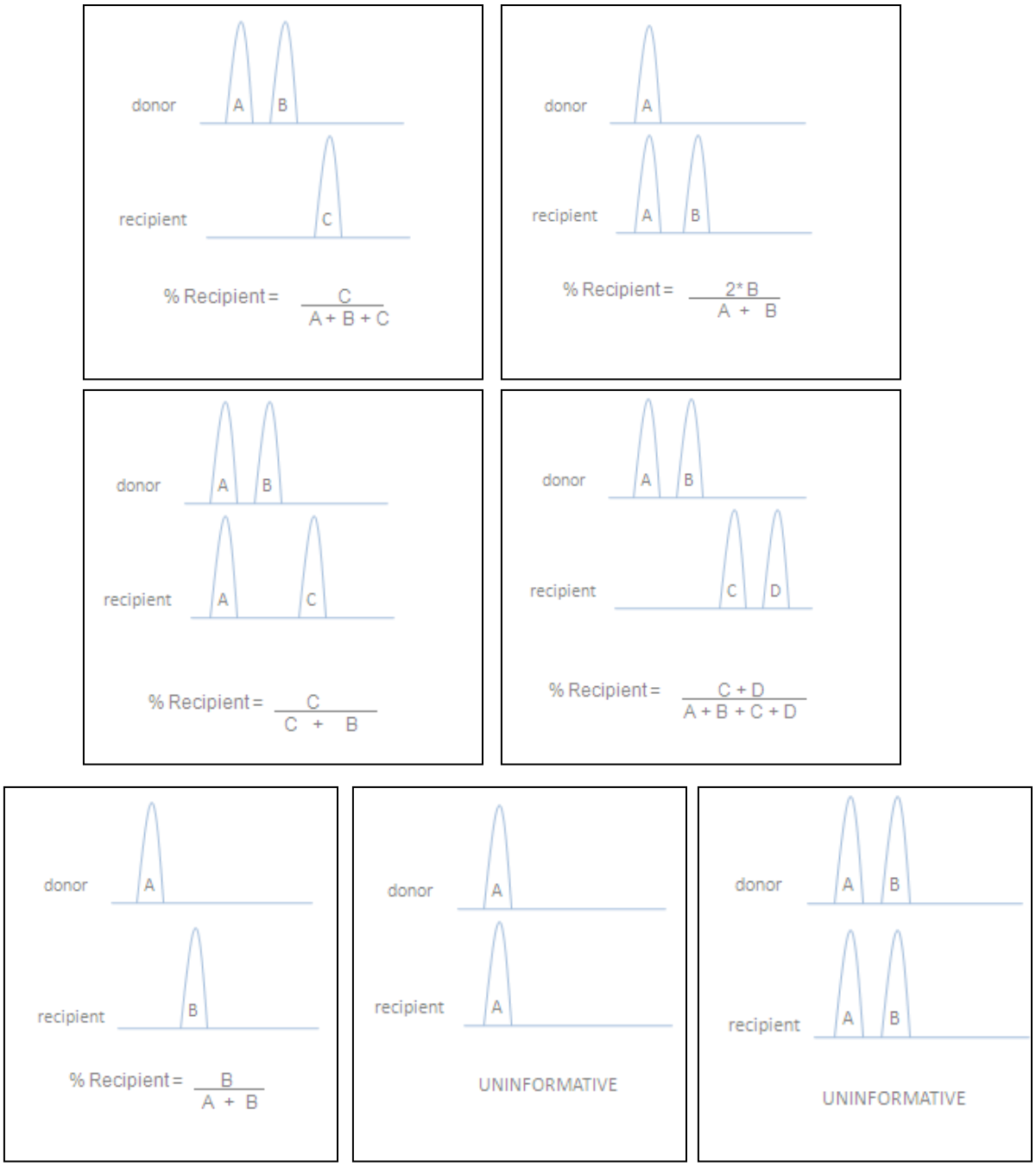

Figure 4: Schematic representation of possible donor and recipient allele patterns with corresponding formulas for mixing proportion calculations.

Fewer STR markers are needed for chimerism analysis than traditional forensic analysis; most literature suggests that recipient-donor alleles can be distinguished using between two and six loci [7, 10]. Loci which contain more polymorphisms are more "informative," or able to distinguish the two contributors to the mixture, than those with fewer polymorphisms. However, it is common that the bone marrow recipient and donor are biologically related; therefore, it is more likely that there will be shared alleles than if the patient and donor are unrelated [11]. For example, in a biallelic system, Mendelian genetics dictates that a sibling pair will have a 25% chance of sharing both alleles and a 50% chance of sharing one allele. If both alleles are shared, that locus is considered "uninformative" and cannot be used to calculate mixing proportions. Shared alleles and stutter are two phenomena that must be considered when interpreting data. Stutter peaks are produced when the template DNA strand slips during replication resulting in a peak which is one repeat unit shorter than the original allele. The amount of stutter is dependent on the repeat motif of the locus and the allele size. When either a donor or recipient allele is in a stutter position to another peak, this can alter the mixing proportion calculation because it is not know how much of the peak height is due to the true allele and how much is due to stutter. Some laboratories will completely exclude a locus if a donor or recipient allele is in a stutter position [11].

Currently, hospitals performing bone marrow transplant monitoring must perform the STR analysis, interpret the data, and calculate the mixture proportions within 24 hours of drawing the patient's blood. In bone marrow transplant monitoring, as in forensic casework, data analysis and calculations of mixture proportions are a bottleneck which must be addressed. Recently, various expert systems and mixture deconvolution software programs have been developed to aid analysts in this time-consuming step. These mixture deconvolution software

programs were originally designed for mixture analysis in forensic casework, such as a sexual assault sample or a mixed blood stain. The programs have the ability to quickly identify potential genotype combinations at each locus, calculate mixing proportions, subtract stutter, and identify uninformative loci. All of these functions also have the ability to be useful in mixture deconvolution of bone marrow engraftment samples [13].

In addition to being time-consuming, manual calculation of mixing proportions is prone to error and inconsistencies between analysts and laboratories. The 2008 College of American Pathologists report on proficiency testing from 58 laboratories performing bone marrow transplants found wide variability in the mixing proportion values obtained between the laboratories. For example, sample ME-08 was prepared as a 17% mixture of "A" cells, however, the results manually calculated by the participating laboratories had a range between 7-23% [9]. It is possible that the use of mixture deconvolution programs could help to standardize how laboratories calculate mixing proportions and lower inconsistencies between laboratories.

CHAPTER II

MATERIALS AND METHODS

Samples

National Institute of Justice Expert System Testbed (NEST) Project mock mixture samples were extracted using QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany) and the DNA was quantified using PicoGreen® dsDNA Quantitation Reagent (Molecular Probes, Eugene, Oregon). The samples were then amplified using the AmpFLSTR® Identifiler™ PCR Amplification Kit (Applied Biosystems) for 28 cycles; this kit contains the tetranucleotide STR markers CSF1PO, D2S1338, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D19S433, D21S11, FGA, TH01, TPOX, vWA, and the sex marker Amelogenin. Samples were also amplified using the PowerPlex® 16 System (Promega Corporation, Madison, Wisconsin) for 22 cycles; this kit contains the tetranucleotide STR markers CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, FGA, TH01, TPOX, and vWA, as well as the pentanucleotide STR markers Penta E and Penta D and the sex marker Amelogenin. All samples were amplified on the ABI PRISM® GeneAmp® PCR System 9700 (Applied Biosystems). Amplified product underwent capillary electrophoresis on an ABI PRISM® 3100 Genetic Analyzer (Applied Biosystems). The NEST Project samples are a series of mixtures with known mixing proportions, as seen in Table 1. NEST A1-G1 are mixtures of varying proportions between individual "A" and individual "X". NEST H1-N1 are mixtures of

varying proportions between individual “B” and individual “Y”.

Sample name	Concentration
NEST A1	A30:01X_1.50ng
NEST B1	A10:01X_1.50ng
NEST C1	A03:01X_1.50ng
NEST D1	A01:01X_1.50ng
NEST E1	A01:03X_1.50ng
NEST F1	A01:10X_1.50ng
NEST G1	A01:30X_1.50ng
NEST H1	B30:01Y_1.50ng
NEST I1	B10:01Y_1.50ng
NEST J1	B03:01Y_1.50ng
NEST K1	B01:01Y_1.50ng
NEST L1	B01:03Y_1.50ng
NEST M1	B01:10Y_1.50ng
NEST N1	B01:30Y_1.50ng

Table 1: Mixing proportions of the NEST Project mock mixture samples.

Samples processed at the City of Hope National Medical Center were previously extracted using QIAamp DNA Blood Mini Kit and the DNA was quantified using the NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts). The samples were then amplified using the AmpFLSTR® Identifiler™ PCR Amplification Kit (Applied Biosystems) for 28 cycles on an ABI PRISM® GeneAmp® PCR System 9700 (Applied Biosystems). Amplified product underwent capillary electrophoresis on an ABI PRISM® 3130*xl* Genetic Analyzer (Applied Biosystems).

Proficiency test samples were previously processed at the Seattle Cancer Care Alliance. These proficiency tests were administered by the American Society for Histocompatibility and Immunogenentics (ASHI) as part of the College of American Pathologists (CAP) accreditation process for laboratories performing engraftment monitoring. The samples were amplified using

the PowerPlex® 16 System. The proficiency tests assess the laboratory's ability to correctly perform the analyses, interpret the samples, and calculate the correct mixing proportions. Each proficiency test contains two whole blood samples from individuals and three or five admixtures of the individuals at varying concentrations [14].

First, the NEST Project data of known mixing proportions will be analyzed using four mixture deconvolution software programs: GeneMapper® *ID-X* Software version 1.1 (Applied Biosystems), FSS-i^3 Expert Systems Software version 4.2.2 (Promega Corporation), TrueAllele® Casework version 2.9 (Cybergenetics, Pittsburg, Pennsylvania), and DNA_Data Analysis Software (U.S. Army Criminal Investigation Laboratory, Fort Gillem, Georgia). The NEST Project samples are defined mixture set that have been amplified at varying concentrations. Next, transplant data from City of Hope National Medical Center and the ASHI proficiency tests will be analyzed. In addition, a time study will be conducted to compare the time needed to compute manual calculations to that needed to process the data through the deconvolution software.

Use of these mixture deconvolution software programs has the potential to drastically reduce the amount of time required to calculate the mixing proportions, and is likely to be performed more consistently between analysts than manual calculations. These mixture deconvolution programs have the ability to rapidly determine the number of contributors, the possible genetic profiles of each contributor at each locus, and the mixing weight of the contributors [13].

GeneMapper ID-X

GeneMapper® *ID-X* version 1.1 is a software program developed by Applied Biosystems. GeneMapper *ID-X* is an expert system, which is software program that can identify peaks, assign alleles, ensure that the data meets the laboratory's defined criteria, describe the rationale behind the decisions, and makes no incorrect calls, all without human intervention [13]. In addition, GeneMapper® *ID-X* contains a mixture analysis tool, which acts as an expert assistant. Mixture analysis can be a difficult and time-consuming step; however, many of the steps can be automated based on a series of rules which guide the analyst to make an informed and more consistent interpretation of the data. The GeneMapper® *ID-X* Mixture Analysis tool can quickly identify the number of contributors to a sample; samples containing two contributors are evaluated based on mixture proportions and peak height ratios [15]. The mixing proportion (Mx) is the measure of relative proportion of the minor contributor in a two-person mixture at all loci. The observed mixing proportion (Mx_{obs}) is the relative proportion of the minor contributor to a two-person mixture at a single locus. The GeneMapper® *ID-X* Mixture Analysis Tool calculates the Mx_{obs} using the formulas seen in Table 2 [13]: The Mx_{obs} is used to calculate the Mx_{avg} by averaging across all 3-peak and 4-peak loci.

Locus	Genotype Combinations	Mx_{obs}
4 peaks	AB, CD	$\frac{C+D}{A+B+C+D}$
3 peaks no shared peaks	AA, BC	$\frac{B+C}{A+B+C}$
3 peaks One shared peak	AB, BC	$\frac{C}{A+C}$
2 peaks No shared peaks	AA, BB	$\frac{B}{A+B}$
2 peaks One shared peak	AA, AB AB, BB	*N/A*
2 peaks Two shared peaks	AB, AB	N/A
1 peak	AA, AA	N/A

Table 2: GeneMapper® *ID-X* calculation of the observed mixing proportions (Mx_{obs}).

The software program generates a list of possible genotype combinations from the detected peaks at each locus, then uses the heterozygote peak height ratio and mixing proportion to evaluate whether the genotype combinations should be placed into the likely or "selected" combinations table or into the unlikely or "unselected" table. GeneMapper® *ID-X* uses a residual calculation, $R = \sum \left(PP_{\exp} - PP_{obs} \right)^2$,which is the difference between the observed and expected peak proportions at a locus, to determine if the proposed genotype combination is placed into the "selected" table. The residual threshold is set at 0.04 by the manufacturer; all genotype combinations greater than 0.04 are placed into the "unselected" genotypes table. The program also calculates the peak height ratios of each proposed heterozygote genotype; the peak height ratios must be greater than 0.5 to place the genotype combination into the "selected" table.

The list of possible genotype combinations of the individuals can be further narrowed by applying a known reference profile, such the victim from an intimate sample or the bone marrow

transplant recipient. GeneMapper® *ID-X* allows the user to apply one known reference sample. Where a known reference is specified, the software indicates where a match occurs and further filters the display options to only those possible genotype combinations where the known is included [15]. When only two or three alleles are detected, the software will indicate that allele dropout is possible. This is designated by an "F" allele in the possible genotype combinations; the F allele is used when a peak falls below the mixture analysis threshold but above the analysis threshold. Once all the possible genotypes for the mixture are included or excluded, GeneMapper® *ID-X* contains statistical features which are able to calculate the Random Match Probability, Combined Probability of Inclusion/Exclusion, and Likelihood Ratios [13].

FSS-i^3

FSS-i^3 Expert System Software version 4.2.2 is a software program developed by the Forensic Science Service software business unit and distributed in the United States by Promega Corporation. FSS-i^3 Expert System Software is comprised of three sub-programs: i-STRess, i-STReam, and i-ntegrity [14]. The data are first processed in GeneMapper® *ID*, GeneMapper® *ID-X*, or GeneScan®/Genotyper® to determine the scan data points, peak height, peak area, and basepair size. The data are then exported and uploaded to the i-STRess function of FSS-i^3. i-STRess is a data interpretation tool that produces DNA profiles by applying configurable rules and filters. The data are then analyzed in i-STReam. The i-STReam software uses the least-squares residual algorithm to produce the best-fit analysis of a two-person mixture using peak height or area and the allele designations determined by i-STRess. The Mx is the mixing proportion for an allele pair at one locus, and is calculated by the formulas seen in Table 3. The least-squares method approximates the mixing proportions (MxP) of the two DNA donors present in the mixture, and a list of all possible genotypes is generated in the i-STReam summary

sheet.

Locus	Genotype Combinations	Mx_{obs}
4 peaks	AB, CD	$\frac{C+D}{A+B+C+D}$
3 peaks no shared peaks	AA, BC	$\frac{B+C}{A+B+C}$
3 peaks One shared peak	AB, BC	$\frac{3C-A+B}{2(A+B+C)}$
2 peaks No shared peaks	AA, BB	$\frac{B}{A+B}$
2 peaks One shared peak	AA, AB	$\frac{2B}{A+B}$
2 peaks One shared peak	AB, BB	$\frac{B-A}{A+B}$
2 peaks Two shared peaks	AB, AB	*N/A*
1 peak	AA, AA	*N/A*

Table 3: FSS-i^3 calculation of the observed mixing proportions (Mx_{obs}).

Each candidate genotype is evaluated using the determined mixture ratio and user-defined settings, which can be specified to fit a laboratory's interpretation protocols. The user-defined settings include more than 20 rules which the user can optimize, such as relative fluorescence unit (RFU) threshold cutoff values, flags for imbalanced sister peaks, and flags for low heterozygote or homozygote RFU levels. The software then identifies the genotype combinations of two DNA contributors to the mixture that best fits the data at each locus. FSS-i^3 does not allow the user to apply known reference profiles to the mixture analysis. In cases where more than one profile combination fits the data, these alternatives are also included by the software. The i-ntegrety module allows a laboratory to check for potential sample-to-sample contamination within a batch [16]. Like GeneMapper® *ID-X*, FSS-i^3 allows the application of

the F-allele rule for potential allele dropout. This is applied when a homozygote peak falls below the set threshold of 150 RFU. FSS-i^3 does not allow the user to apply known reference samples; however, in the analysis of bone marrow engraftment data, where both of the contributor's profiles are known, the user can eliminate all incorrect genotype combinations which the program generates in the summary sheet [13].

TrueAllele® Casework

TrueAllele® Casework is a mixture deconvolution program developed by Cybergenetics Corporation [17]. The data are first processed in TrueAllele® Database; then, in TrueAllele® Casework, four windows are displayed: profile, match, mixture, and data. The profile window displays the contributors' possible genotypes at each locus; for each genotype a probability is assigned. The match window gives a likelihood ratio that the contributor in the mixture is a match to a single source sample which is present in the TrueAllele® Casework database. The match window also displays the strength, or overall locus contribution, to the likelihood ratio [13].

TrueAllele® Casework functions by repeatedly forming a hypothesis about one of the statistical modeling parameters, such as mixing proportions, genotypes, PCR artifacts, DNA amount, background noise, and data uncertainty. Once the parameters are combined, the system generates a hypothetical data pattern which is compared to the short tandem repeat (STR) data to form a probability value [18]. This process is repeated approximately 10,000 times, using a shuffling method [14]. Ultimately, the software will assign a higher probability to those genotypes which best fit the observed data. The software will converge on an answer which contains random variation, represented by a histogram. The center value of the histogram represents the mean value of the unknown parameter, such as mixing proportion, and the

width of the histogram represents the standard deviation [18]. TrueAllele® Casework allows the user to apply one known reference sample to the mixture [12].

DNA_DataAnalysis

DNA_Data Analysis is a mixture deconvolution software program which was written in Visual Basic Application with Microsoft® Excel as the interface. The program was developed by the U.S. Army Criminal Investigation Laboratory (USACIL). DNA_DataAnalysis is not intended to be an expert system, but to aid analysts in mixture interpretation. The raw data are first inspected and analyzed in GeneMapper® *ID* v.3.2 to determine the appropriate allele calls and to remove artifacts. The resulting peak height, base pair size, and allele calls are then exported into a genotypes table which can be imported into the DNA_DataAnalysis software. DNA_DataAnalysis is capable of interpreting samples processed using the PowerPlex® 16 System and the AmpFLSTR® Identifiler™, COfiler™, Profiler Plus™, and Yfiler™ PCR Amplification Kits.

When a locus is clicked, a box will be displayed which lists all possible genotype combinations, the peak height ratio for each combination, and the mixing proportions. At this step, the user can designate the genotype of one or both contributors, thus reducing the list of possible genotype combinations to only those which are consistent with the contributors. DNA_Data Analysis allows the user to match the mixture to a maximum of six single-source reference profiles. DNA_Data Analysis has a feature which allows the user to subtract the effect of stutter if one contributors' allele is in a stutter positions. The default program of the software is programmed with the manufacturer's defined stutter percentages for each locus.

Alternatively, the user can define his or her own stutter percentages.

DNA_Data Analysis also includes statistical features which are able to calculate the Random Match Probability, Combined Probability of Inclusion/Exclusion, and Likelihood Ratios using the allele frequencies from the Federal Bureau of Investigation database [13]. DNA_DataAnalysis is the only program being evaluated which has the capability of determining mixing proportions of three-person mixtures. The most commonly used functions performed by DNA_DataAnalysis can be seen in the flowchart below.

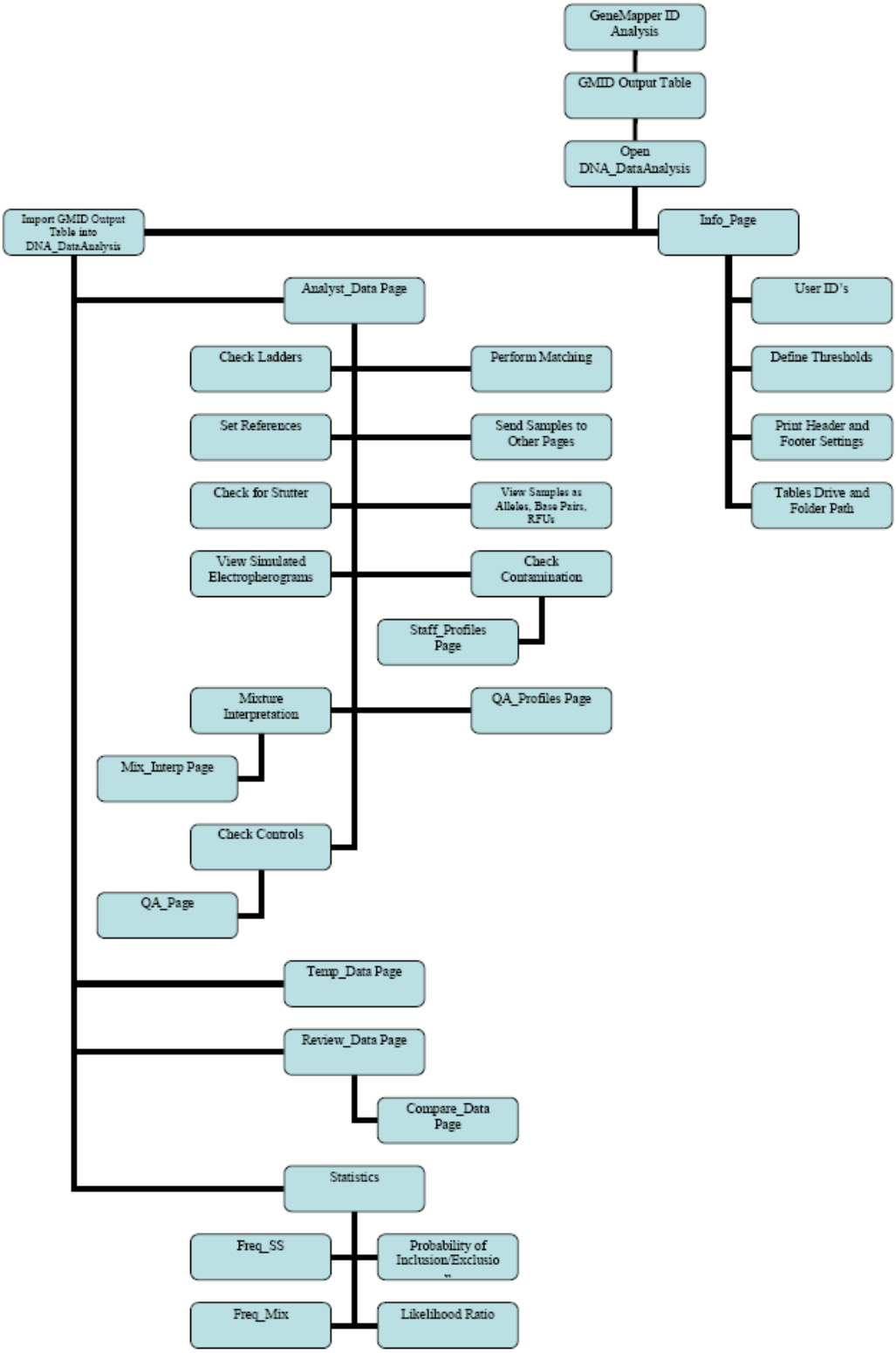

Figure 5: Flowchart of functions performed by DNA_DataAnalysis [19].

CHAPTER III

RESULTS AND DISCUSSION

Evaluation of the Various Software Programs

Each of the available mixture deconvolution software programs include different features which make them more suitable or less suitable for bone marrow engraftment monitoring. Because the software programs were originally designed for forensic casework mixtures, not all the features translate well to processing bone marrow engraftment mixture samples.

GeneMapper® *ID-X* may be the most logical choice to introduce to laboratories performing engraftment monitoring because it is an easy transition to make from GeneMapper® *ID*, which almost all laboratories that perform short tandem repeat (STR) analysis use for fragment analysis. A sample electropherogram from GeneMapper® *ID-X* can be seen in Figure 6. One limitation of GeneMapper® *ID-X* is that it only allows the user to input one reference sample; while this narrows the number of possible genotype combinations, most loci still contain multiple possible genotype combinations, as seen in Figure 7. This causes the user to have to manually place all incorrect genotype combinations into the "unselected" table, which is a moderately time-consuming step. In the example seen in Figure 7, there are 23 possible genotype combinations for the 16 loci, therefore, 7 genotype combinations which do not correspond to the donor-recipient profiles must be removed by the user. The average time needed

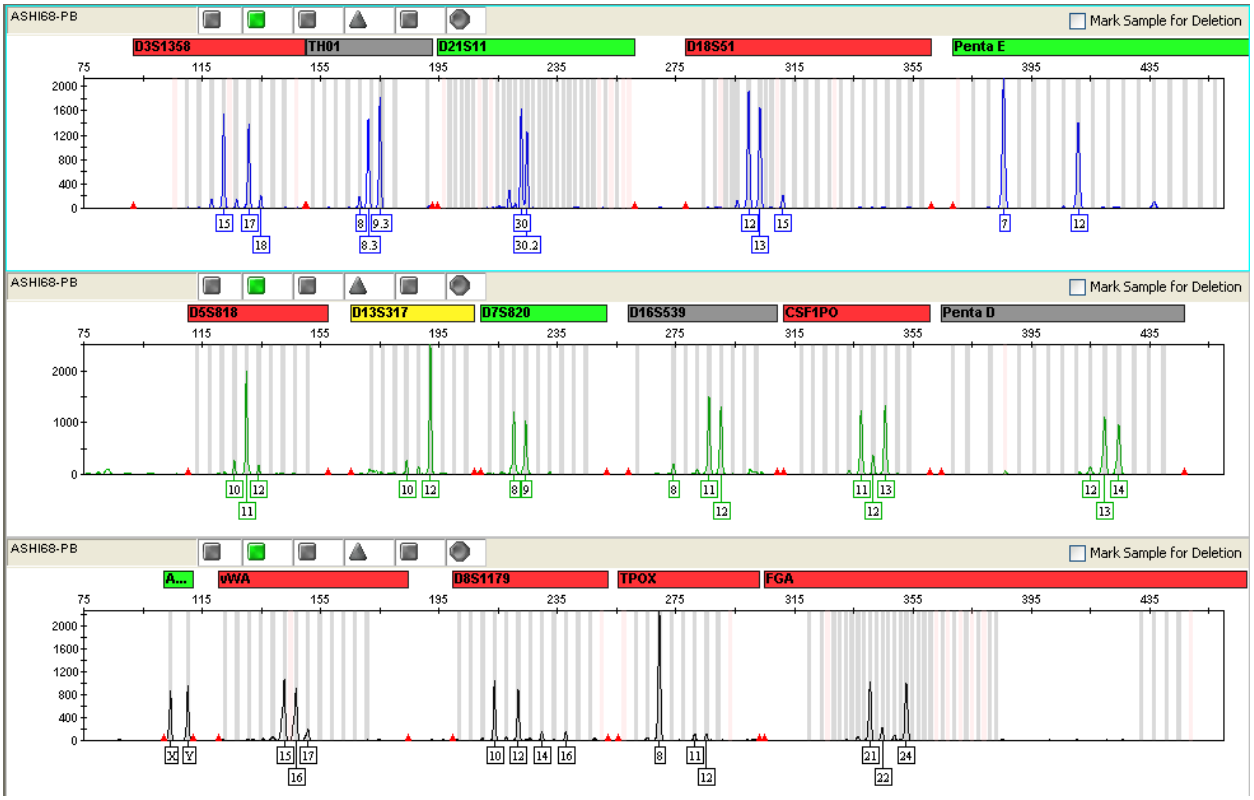

Figure 6: Electropherogram of ASHI sample 2295-EMO-68 which was processed with three mixture deconvolution programs.

to determine the correct mixing proportions of a two-person mixture using GeneMapper® *ID-X* was 2 minutes, 30 seconds. Another limitation of GeneMapper® *ID-X* is that the program will not generate the mixing proportion for a two-peak locus with one shared allele in which one contributor is homozygous and one contributor is heterozygous, as seen in Figure 8.

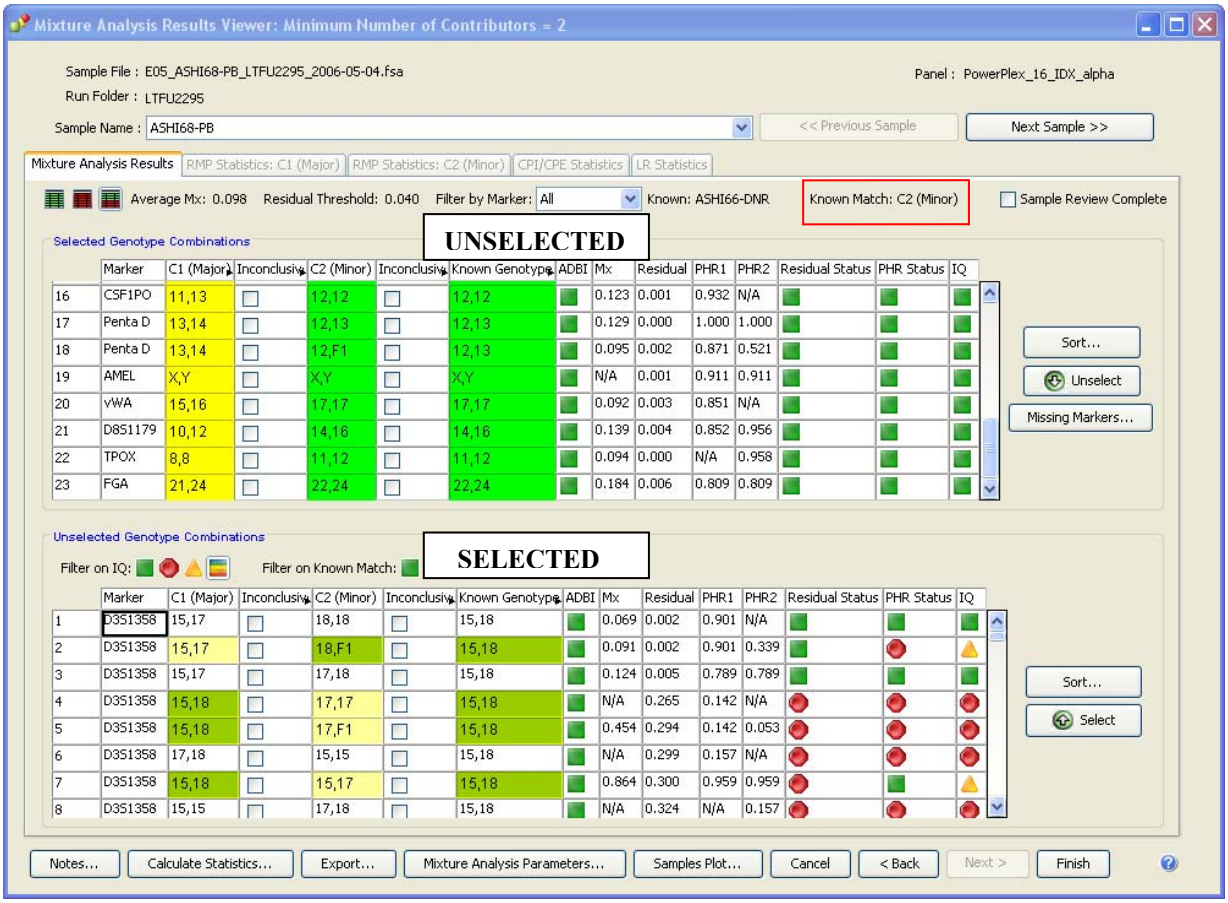

Figure 7: Sample output of selected and unselected genotype combinations tables from GeneMapper® *ID-X*. A sample 2295-68 from the ASHI proficiency test set with one reference profile applied. The results of the sample produced 23 selected genotype combinations and 258 unselected genotype combinations for the mixture

Other programs, such as DNA_DataAnalysis and FSS-i[3], as well as engraftment monitoring laboratories, consider this type of locus to be informative. However, GeneMapper® *ID-X* considers this an uninformative locus, and as such, will not calculate the mixing proportion. This will result in a further reduction of the number of loci used in the average mixing proportion calculation. This may be problematic for engraftment monitoring, especially in transplants between relatives, which are more likely to have such an allele pattern.

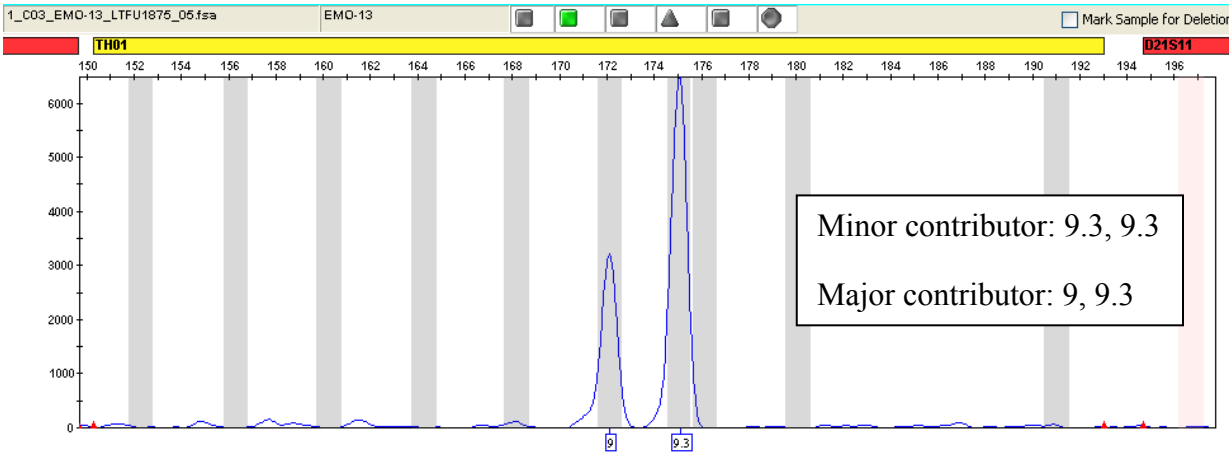

Figure 8: Example of GeneMapper® ID-X uninformative locus.

FSS-i^3, while well-suited for forensic casework, is the least suitable program for bone marrow engraftment monitoring. FSS-i^3 does not allow the user to apply either the donor or recipient reference profile to the mixture analysis; therefore, a long list of possible genotype combinations is generated, as seen in Figure 9. This list must be visually searched by the user for a match to the correct genotypes of the donor and recipient to find the mixing proportions at each locus. This process is time-consuming and prone to human-error. The time needed to determine the mixing proportions at 16 loci using FSS-i^3 was approximately 11 minutes, 20 seconds per sample.

FSS-i^3 does not allow the user to apply references because the program was designed by forensic scientists for forensic samples which should be analzyed with the utmost impartiality. The DNA commission of the International Society of Forensic Genetics recommendations for interpretation of mixtures state that an analyst should determine the number of contributors to a mixture, estimate the mixing proportion, and determine all possible genotype combinations *before* comparing the sample to a reference [20]. However, engraftment monitoring is quite different from forensic casework in the need for impartiality because the contributors to the

mixture are already known. Therefore, a mixture deconvolution program which allows the user to apply two known reference profiles, such as that of the donor and the recipient, would be best suited for engraftment monitoring.

-			Possible Contributors				Pref Amp Rule				Mix Prop Rule		-							
Locus	Allele	Height/Area	Contributor 1		Contributor 2		Contributor 1		Contributor 2		Mix Est		RC	Contributor 1		Contributor 2		Den	Num	Comments
D3S1358	15	1879	17	18	15	15	2394%	N	-	Y	49% 1:1	N	-	-	-	-	-	-	-	-
	17	1867	17	18	15	17	95%	Y	95%	Y	97% <1:10	N	-	-	-	-	-	-	-	-
	18	78	15	18	15	17	97%	Y	97%	Y	97% <1:10	N	-	-	-	-	-	-	-	-
	-	-	18	18	15	17	-	Y	101%	Y	98% <1:10	N	-	-	-	-	-	-	-	-
	-	-	15	17	15	18	97%	Y	97%	Y	3% >10:1	Y	Include	~~15~~	~~17~~	~~15~~	~~18~~	-	-	-
	-	-	17	18	15	18	2%	N	2%	N	50% 1:1	N	-	-	-	-	-	-	-	-
	-	-	17	17	15	18	-	Y	2409%	N	51% 1:1	N	-	-	-	-	-	-	-	-
	-	-	15	17	17	18	95%	Y	95%	Y	3% >10:1	Y	Include	~~15~~	~~17~~	~~17~~	~~18~~	-	-	-
	-	-	15	18	17	18	2%	N	2%	N	50% 1:1	N	-	-	-	-	-	-	-	-
	-	-	15	15	17	18	-	Y	2394%	N	51% 1:1	N	-	-	-	-	-	-	-	-
	-	-	15	18	17	17	2409%	N	-	Y	49% 1:1	N	-	-	-	-	-	-	-	-
	-	-	15	17	18	18	101%	Y	-	Y	2% >10:1	Y	Include	15	17	18	F	-	-	-
	Database Consolidation for D3S1358													15	17	18	F	-	-	-
TH01	8	53	9.3	9.3	8	8	-	Y	-	Y	3% >10:1	Y	Include	9.3	9.3	8	F	-	-	-
	9.3	1830	8	9.3	8	8	3%	N	100%	Y	-94% >10:1	N	-	-	-	-	-	-	-	-
	-	-	8	9.3	8	9.3	3%	N	3%	N	-	Y	-	-	-	-	-	-	-	-
	-	-	9.3	9.3	8	9.3	100%	Y	3453%	Y	6% 10:1	Y	Include	~~9.3~~	~~9.3~~	~~8~~	~~F~~	-	-	-
	-	-	8	8	9.3	9.3	-	Y	-	Y	97% <1:10	N	-	-	-	-	-	-	-	-
	-	-	8	8	8	9.3	100%	Y	3%	N	194% <1:10	N	-	-	-	-	-	-	-	-
	-	-	8	9.3	9.3	9.3	3453%	Y	100%	Y	94% 1:10	N	-	-	-	-	-	-	-	-
	Database Consolidation for TH01													9.3	9.3	8	F	-	-	-
D21S11	30	1817	30.2	30.2	30	30	-	Y	-	Y	50% 1:1	N	-	-	-	-	-	-	-	-
	30.2	1852	30	30.2	30	30	98%	Y	100%	Y	-1% >10:1	Y	Include	~~30~~	~~30.2~~	~~30~~	~~F~~	-	-	-
	-	-	30	30.2	30	30.2	98%	Y	98%	Y	-	Y	Include	30	30.2	F	F	-	-	-
	-	-	30.2	30.2	30	30.2	100%	Y	102%	Y	99% <1:10	N	-	-	-	-	-	-	-	-
	-	-	30	30	30.2	30.2	-	Y	-	Y	50% 1:1	N	-	-	-	-	-	-	-	-
	-	-	30	30	30	30.2	100%	Y	98%	Y	101% <1:10	N	-	-	-	-	-	-	-	-
	-	-	30	30.2	30.2	30.2	102%	Y	100%	Y	1% >10:1	Y	Include	~~30~~	~~30.2~~	~~30.2~~	~~F~~	-	-	-
	Database Consolidation for D21S11													30	30.2	F	F	-	-	-
D18S51	12	2692	13	15	12	12	1399%	N	-	Y	53% 1:1	N	-	-	-	-	-	-	-	-
	13	2239	13	15	12	13	79%	Y	79%	Y	100% <1:10	N	-	-	-	-	-	-	-	-
	15	160	12	15	12	13	112%	Y	112%	Y	91% 1:10	N	-	-	-	-	-	-	-	-
	-	-	15	15	12	13	-	Y	120%	Y	97% <1:10	N	-	-	-	-	-	-	-	-
	-	-	12	13	12	15	112%	Y	112%	Y	9% 10:1	Y	Include	12	13	12	15	-	-	-
	-	-	13	15	12	15	3%	N	3%	N	59% 1:1	N	-	-	-	-	-	-	-	-
	-	-	13	13	12	15	-	Y	1683%	N	56% 1:1	N	-	-	-	-	-	-	-	-
	-	-	12	13	13	15	79%	Y	79%	Y	0% >10:1	Y	Include	12	13	13	15	-	-	-
	-	-	12	15	13	15	3%	N	3%	N	41% 1:1	N	-	-	-	-	-	-	-	-
	-	-	12	12	13	15	-	Y	1399%	N	47% 1:1	N	-	-	-	-	-	-	-	-
	-	-	12	15	13	13	1683%	N	-	Y	44% 1:1	N	-	-	-	-	-	-	-	-
	-	-	12	13	15	15	120%	Y	-	Y	3% >10:1	Y	Include	12	13	15	15	-	-	-

Figure 9: Sample output of possible genotype combinations and mixing proportions from FSS-i[3]. A sample from the ASHI proficiency test set with no reference profiles applied.

DNA_DataAnalysis is the mixture deconvolution program which possesses all the features most amenable to bone marrow engraftment monitoring. Once the donor and recipient reference profiles are applied to the mixture sample, an output file of only one possible correct genotype combination and the corresponding mixing proportion is generated for each locus, as

seen in Figure 10. In addition, the feature to remove locus-specific stutter may be very useful in bone marrow engraftment monitoring. It is very important to determine the mixing proportion of the two contributors with as much precision as possible. The addition of stutter to an allele can falsely inflate the mixing proportion of the donor or the recipient, depending on which allele is in the stutter position. In addition, DNA_DataAnalysis is the program which requires the least amount of time to perform mixture analysis; the average time to determine the mixing proportions of the 16 loci in one sample was approximately 1 minute, 10 seconds.

E05 ASHI68-PB LTFU2295 2006-05-04.fsa

D3S1358: For a 2-contributor 3-allele mixture of types AA & BC: 3-combinations; Alleles (RFUs): 15 (1539), 17 (1386), 18 (218)
Only combinations including the following reference profiles are included: (15, 18) (15, 17) ; all peak height ratios are >= (0.4);

D3S1358: For a 2-contributor 3-allele mixture of types AB & BC: 3-combinations; Alleles (RFUs): 15 (1539), 17 (1386), 18 (218)
Only combinations including the following reference profiles are included: (15, 18) (15, 17) ; all peak height ratios are >= (0.4);
15, 17 (phr = 0.96; p = 0.86); 15, 18 (phr = 0.96; p = 0.14)

E05 ASHI68-PB LTFU2295 2006-05-04.fsa

TH01: For a 2-contributor 3-allele mixture of types AA & BC: 3-combinations; Alleles (RFUs): 8 (192), 8.3 (1462), 9.3 (1820)
Only combinations including the following reference profiles are included: (8, 9.3) (8.3, 9.3) ; all peak height ratios are >= (0.4);

TH01: For a 2-contributor 3-allele mixture of types AB & BC: 3-combinations; Alleles (RFUs): 8 (192), 8.3 (1462), 9.3 (1820)
Only combinations including the following reference profiles are included: (8, 9.3) (8.3, 9.3) ; all peak height ratios are >= (0.4);
8, 9.3 (phr = 0.91; p = 0.12); 8.3, 9.3 (phr = 0.91; p = 0.88)

E05 ASHI68-PB LTFU2295 2006-05-04.fsa

D18S51: For a 2-contributor 3-allele mixture of types AA & BC: 3-combinations; Alleles (RFUs): 12 (1930), 13 (1646), 15 (207)
Only combinations including the following reference profiles are included: (12, 15) (12, 13) ; all peak height ratios are >= (0.4);

D18S51: For a 2-contributor 3-allele mixture of types AB & BC: 3-combinations; Alleles (RFUs): 12 (1930), 13 (1646), 15 (207)
Only combinations including the following reference profiles are included: (12, 15) (12, 13) ; all peak height ratios are >= (0.4);
12, 13 (phr = 0.96; p = 0.89); 12, 15 (phr = 0.96; p = 0.11)

E05 ASHI68-PB LTFU2295 2006-05-04.fsa

D13S317: For a 2-contributor 2-allele mixture of types AA & BB: 1-combination; Alleles (RFUs): 10 (270), 12 (2509)
Only combinations including the following reference profiles are included: (10, 12) (12, 12) ; all peak height ratios are >= (0.4);

D13S317: For a 2-contributor 2-allele mixture of types AA & AB; (2-combinations for AB: phr = 1 or calc based on set phr); Alleles (RFUs): 10 (270), 12 (2509)
Only combinations including the following reference profiles are included: (10, 12) (12, 12) ; all peak height ratios are >= (0.4);
12, 12 (p = 0.81); 10, 12 (phr = 1; p = 0.19)

D13S317: For a 2-contributor 2-allele mixture of types AB & AB: 1-combination; Alleles (RFUs): 10 (270), 12 (2509)
Only combinations including the following reference profiles are included: (10, 12) (12, 12) ; all peak height ratios are >= (0.4);

E05 ASHI68-PB LTFU2295 2006-05-04.fsa

D16S539: For a 2-contributor 3-allele mixture of types AA & BC: 3-combinations; Alleles (RFUs): 8 (197), 11 (1510), 12 (1308)
Only combinations including the following reference profiles are included: (8, 12) (11, 12) ; all peak height ratios are >= (0.4);

D16S539: For a 2-contributor 3-allele mixture of types AB & BC: 3-combinations; Alleles (RFUs): 8 (197), 11 (1510), 12 (1308)
Only combinations including the following reference profiles are included: (8, 12) (11, 12) ; all peak height ratios are >= (0.4);
8, 12 (phr = 0.77; p = 0.12); 11, 12 (phr = 0.77; p = 0.88)

Figure 10: Sample output from DNA_DataAnalysis of mixing proportions at four loci. A sample from the ASHI proficiency test set with two reference profiles applied. Note that only one possible genotype combination is given per locus. The minor contributor mixing proportion is circled and the major contributor mixing proportion is boxed.

TrueAllele® Casework was unavailable for my use during the project. However, it is

likely that TrueAllele® Casework would not be appropriate for engraftment monitoring because the program reports an average mixing proportion, not the mixing proportions at the individual loci. It is important to know the mixing proportions at each locus to determine if outliers exist that should be removed from the average mixing proportion calculation. In addition, TrueAllele® Casework determines the overall mixing proportion using a random shuffling method, not the traditional method using peak height or peak area calculations. Therefore, it is unlikely bone marrow engraftment laboratories would adopt this mixture deconvolution program.

Evaluation of NEST Project Control Samples

Thirteen National Institute of Justice Expert System Testbed (NEST) Project samples of know mixing proportion were processed using GeneMapper® *ID-X*, FSS-i^3, and DNA_DataAnalysis as a defined dataset to ensure that the mixture deconvolution programs were able to properly assign the correct genotype combinations and calculate the correct mixing proportions. All three programs generated similar mixing proportions which were close to the expected values. As expected, the 30:1 mixture is the least accurate calculation. This is not due to the performance of the individual deconvolution programs, but due to stochastic amplification of low-level DNA during PCR, which results in unequal sampling of the alleles. When excluding the 30:1 and 1:30 samples, the percent error between the expected value and software calculated value is 7.7% for GeneMapper ID-X, 6.0% for FSS-i3, and 6.2% for DNA_DataAnalysis. However, it is unknown if this error is due to pipetting during the preparation of the mixture, stochastic amplification error, or error in the mixture calculation.

Sample	Locus	GMIDX	FSS-i[3]	DNA_DA	Sample	Locus	GMIDX	FSS-i[3]	DNA_DA
A1 (30: 1)	D8S1179	0.042	0.060	0.040	H1 (30:1)	D8S1179	0.052	0.050	0.050
A1 (30: 1)	D21S11	0.036	0.040	0.040	H1 (30:1)	D21S11	0.090	dropout	0.090
A1 (30: 1)	D7S820	dropout	dropout	dropout	H1 (30:1)	D7S820	0.056	0.080	0.060
A1 (30: 1)	CSF1PO	U/I	U/I	U/I	H1 (30:1)	CSF1PO	dropout	dropout	dropout
A1 (30: 1)	D3S1358	U/I	U/I	U/I	H1 (30:1)	D3S1358	0.049	0.050	0.050
A1 (30: 1)	TH01	0.042	0.040	0.040	H1 (30:1)	TH01	dropout	dropout	0.040
A1 (30: 1)	D13S317	dropout	0.030	dropout	H1 (30:1)	D13S317	N/A	0.040	0.040
A1 (30: 1)	D16S539	0.045	0.030	0.050	H1 (30:1)	D16S539	U/I	U/I	U/I
A1 (30: 1)	D2S1338	dropout	dropout	dropout	H1 (30:1)	D2S1338	dropout	dropout	dropout
A1 (30: 1)	D19S433	dropout	dropout	dropout	H1 (30:1)	D19S433	U/I	U/I	U/I
A1 (30: 1)	vWA	0.049	0.050	0.050	H1 (30:1)	vWA	dropout	dropout	dropout
A1 (30: 1)	TPOX	N/A	0.040	0.040	H1 (30:1)	TPOX	U/I	U/I	U/I
A1 (30: 1)	D18S51	0.033	0.030	0.030	H1 (30:1)	D18S51	dropout	dropout	dropout
A1 (30: 1)	Amel	N/A	dropout	dropout	H1 (30:1)	Amel	N/A	dropout	dropout
A1 (30: 1)	D5S818	dropout	dropout	dropout	H1 (30:1)	D5S818	0.070	dropout	0.070
A1 (30: 1)	FGA	U/I	U/I	U/I	H1 (30:1)	FGA	dropout	dropout	dropout
Avg. Mx		**0.041**	**0.040**	**0.041**	**Avg. Mx**		**0.063**	**0.055**	**0.057**

Sample	Locus	GMIDX	FSS-i[3]	DNA_DA	Sample	Locus	GMIDX	FSS-i[3]	DNA_DA
B1 (10:1)	D8S1179	0.111	0.080	0.110	J1(3:1)	D8S1179	0.329	0.330	0.330
B1 (10:1)	D21S11	0.094	0.090	0.090	J1(3:1)	D21S11	0.298	0.260	0.300
B1 (10:1)	D7S820	0.130	0.110	0.130	J1(3:1)	D7S820	0.279	0.260	0.280
B1 (10:1)	CSF1PO	U/I	U/I	U/I	J1(3:1)	CSF1PO	0.285	0.270	0.280
B1 (10:1)	D3S1358	U/I	U/I	U/I	J1(3:1)	D3S1358	0.322	0.320	0.320
B1 (10:1)	TH01	0.086	0.090	0.090	J1(3:1)	TH01	0.315	0.320	0.320
B1 (10:1)	D13S317	0.073	0.070	0.070	J1(3:1)	D13S317	N/A	0.240	0.240
B1 (10:1)	D16S539	0.075	0.040	0.080	J1(3:1)	D16S539	U/I	U/I	U/I
B1 (10:1)	D2S1338	0.117	0.120	0.110	J1(3:1)	D2S1338	0.340	0.340	0.340
B1 (10:1)	D19S433	0.094	0.100	0.090	J1(3:1)	D19S433	U/I	U/I	U/I
B1 (10:1)	vWA	0.103	0.100	0.100	J1(3:1)	vWA	0.276	0.280	0.280
B1 (10:1)	TPOX	N/A	0.050	0.050	J1(3:1)	TPOX	U/I	U/I	U/I
B1 (10:1)	D18S51	0.091	0.090	0.090	J1(3:1)	D18S51	0.271	0.270	0.270
B1 (10:1)	Amel	N/A	0.120	0.110	J1(3:1)	Amel	N/A	0.290	0.290
B1 (10:1)	D5S818	0.092	0.120	0.110	J1(3:1)	D5S818	0.286	0.290	0.290
B1 (10:1)	FGA	U/I	U/I	U/I	J1(3:1)	FGA	0.273	0.290	0.270
Avg. Mx		**0.097**	**0.091**	**0.095**	**Avg. Mx**		**0.298**	**0.289**	**0.293**

Sample	Locus	GMIDX	FSS-i[3]	DNA_DA	Sample	Locus	GMIDX	FSS-i[3]	DNA_DA
C1 (3:1)	D8S1179	0.195	0.170	0.200	K1 (1:1)	D8S1179	0.480	0.480	0.480
C1 (3:1)	D21S11	0.206	0.210	0.210	K1 (1:1)	D21S11	0.504	0.510	0.500
C1 (3:1)	D7S820	0.264	0.230	0.260	K1 (1:1)	D7S820	0.485	0.480	0.490
C1 (3:1)	CSF1PO	U/I	U/I	U/I	K1 (1:1)	CSF1PO	0.443	0.450	0.440
C1 (3:1)	D3S1358	U/I	U/I	U/I	K1 (1:1)	D3S1358	0.487	0.490	0.490
C1 (3:1)	TH01	0.283	0.280	0.280	K1 (1:1)	TH01	0.421	0.420	0.420
C1 (3:1)	D13S317	0.201	0.200	0.200	K1 (1:1)	D13S317	N/A	0.390	0.390
C1 (3:1)	D16S539	0.296	0.310	0.300	K1 (1:1)	D16S539	U/I	U/I	U/I
C1 (3:1)	D2S1338	0.226	0.230	0.230	K1 (1:1)	D2S1338	0.442	0.440	0.440
C1 (3:1)	D19S433	0.301	0.300	0.300	K1 (1:1)	D19S433	U/I	U/I	U/I
C1 (3:1)	vWA	0.255	0.260	0.250	K1 (1:1)	vWA	0.465	0.460	0.460
C1 (3:1)	TPOX	N/A	0.250	0.250	K1 (1:1)	TPOX	U/I	U/I	U/I
C1 (3:1)	D18S51	0.213	0.210	0.210	K1 (1:1)	D18S51	0.500	0.500	0.500
C1 (3:1)	Amel	N/A	0.280	0.280	K1 (1:1)	Amel	N/A	0.470	0.470
C1 (3:1)	D5S818	0.257	0.270	0.260	K1 (1:1)	D5S818	0.527	0.520	0.530
C1 (3:1)	FGA	U/I	U/I	U/I	K1 (1:1)	FGA	0.436	0.440	0.440
Avg. Mx		**0.245**	**0.246**	**0.248**	**Avg. Mx**		**0.472**	**0.465**	**0.465**

Sample	Locus	GMIDX	FSS-i[3]	DNA_DA	Sample	Locus	GMIDX	FSS-i[3]	DNA_DA
D1 (1:1)	D8S1179	0.481	0.480	0.480	L1 (1:3)	D8S1179	0.234	0.240	0.230
D1 (1:1)	D21S11	0.491	0.490	0.490	L1 (1:3)	D21S11	0.303	0.320	0.300
D1 (1:1)	D7S820	0.413	0.410	0.410	L1 (1:3)	D7S820	0.220	0.250	0.220
D1 (1:1)	CSF1PO	U/I	U/I	U/I	L1 (1:3)	CSF1PO	0.265	0.270	0.260
D1 (1:1)	D3S1358	U/I	U/I	U/I	L1 (1:3)	D3S1358	0.198	0.200	0.200
D1 (1:1)	TH01	0.453	0.450	0.450	L1 (1:3)	TH01	0.212	0.210	0.210
D1 (1:1)	D13S317	0.531	0.530	0.530	L1 (1:3)	D13S317	N/A	0.310	0.310
D1 (1:1)	D16S539	0.561	0.570	0.560	L1 (1:3)	D16S539	U/I	U/I	U/I
D1 (1:1)	D2S1338	0.528	0.530	0.530	L1 (1:3)	D2S1338	0.178	0.180	0.180
D1 (1:1)	D19S433	0.510	0.510	0.510	L1 (1:3)	D19S433	U/I	U/I	U/I
D1 (1:1)	vWA	0.499	0.500	0.500	L1 (1:3)	vWA	0.259	0.260	0.260
D1 (1:1)	TPOX	N/A	0.310	0.490	L1 (1:3)	TPOX	U/I	U/I	U/I
D1 (1:1)	D18S51	0.491	0.490	0.490	L1 (1:3)	D18S51	0.247	0.240	0.250
D1 (1:1)	Amel	N/A	0.540	0.450	L1 (1:3)	Amel	N/A	0.280	0.280
D1 (1:1)	D5S818	0.416	0.590	0.420	L1 (1:3)	D5S818	0.221	0.220	0.220
D1 (1:1)	FGA	U/I		U/I	L1 (1:3)	FGA	0.217	0.250	0.220
Avg. Mx		**0.489**	**0.492**	**0.495**	**Avg. Mx**		**0.232**	**0.248**	**0.242**

Sample	Locus	GMIDX	FSS-i[3]	DNA_DA	Sample	Locus	GMIDX	FSS-i[3]	DNA_DA
E1 (1:3)	D8S1179	0.729	0.750	0.730	M1 (1:10)	D8S1179	0.921	0.920	0.920
E1 (1:3)	D21S11	0.728	0.730	0.730	M1 (1:10)	D21S11	0.861	0.890	0.860
E1 (1:3)	D7S820	0.743	0.730	0.740	M1 (1:10)	D7S820	0.879	0.870	0.880
E1 (1:3)	CSF1PO	U/I	U/I	U/I	M1 (1:10)	CSF1PO	0.952	0.980	0.950
E1 (1:3)	D3S1358	U/I	U/I	U/I	M1 (1:10)	D3S1358	0.939	0.940	0.940
E1 (1:3)	TH01	0.703	0.700	0.700	M1 (1:10)	TH01	0.922	dropout	0.920
E1 (1:3)	D13S317	0.765	0.760	0.760	M1 (1:10)	D13S317	N/A	0.920	0.920
E1 (1:3)	D16S539	0.723	0.710	0.720	M1 (1:10)	D16S539	U/I	U/I	U/I
E1 (1:3)	D2S1338	0.734	0.730	0.730	M1 (1:10)	D2S1338	0.894	0.890	0.890
E1 (1:3)	D19S433	0.839	0.840	0.840	M1 (1:10)	D19S433	U/I	U/I	U/I
E1 (1:3)	vWA	0.715	0.720	0.720	M1 (1:10)	vWA	dropout	dropout	dropout
E1 (1:3)	TPOX	N/A	0.820	0.810	M1 (1:10)	TPOX	U/I	U/I	U/I
E1 (1:3)	D18S51	0.730	0.730	0.730	M1 (1:10)	D18S51	dropout	dropout	dropout
E1 (1:3)	Amel	N/A	0.870	0.860	M1 (1:10)	Amel	N/A	0.890	0.880
E1 (1:3)	D5S818	0.729	0.720	0.730	M1 (1:10)	D5S818	0.934	0.930	0.930
E1 (1:3)	FGA	U/I	U/I	U/I	M1 (1:10)	FGA	dropout	0.880	dropout
Avg. Mx		**0.740**	**0.755**	**0.754**	**Avg. Mx**		**0.913**	**0.911**	**0.912**

Sample	Locus	GMIDX	FSS-i[3]	DNA_DA	Sample	Locus	GMIDX	FSS-i[3]	DNA_DA
F1(1:10)	D8S1179	0.932	0.960	0.930	N1 (1:30)	D8S1179	0.952	0.950	0.950
F1(1:10)	D21S11	0.874	0.870	0.870	N1 (1:30)	D21S11	dropout	dropout	dropout
F1(1:10)	D7S820	0.881	0.890	0.880	N1 (1:30)	D7S820	dropout	dropout	dropout
F1(1:10)	CSF1PO	U/I	U/I	U/I	N1 (1:30)	CSF1PO	0.968	0.970	0.970
F1(1:10)	D3S1358	U/I	U/I	U/I	N1 (1:30)	D3S1358	dropout	dropout	dropout
F1(1:10)	TH01	0.905	0.900	0.900	N1 (1:30)	TH01	dropout	dropout	dropout
F1(1:10)	D13S317	0.871	0.870	0.870	N1 (1:30)	D13S317	N/A	0.970	0.970
F1(1:10)	D16S539	0.834	0.920	0.830	N1 (1:30)	D16S539	U/I	U/I	U/I
F1(1:10)	D2S1338	0.874	0.870	0.870	N1 (1:30)	D2S1338	dropout	dropout	dropout
F1(1:10)	D19S433	0.930	0.930	0.930	N1 (1:30)	D19S433	U/I	U/I	U/I
F1(1:10)	vWA	0.849	0.850	0.850	N1 (1:30)	vWA	dropout	dropout	dropout
F1(1:10)	TPOX	N/A	0.910	0.920	N1 (1:30)	TPOX	U/I	U/I	U/I
F1(1:10)	D18S51	0.897	0.900	0.900	N1 (1:30)	D18S51	dropout	dropout	dropout
F1(1:10)	Amel	N/A	0.990	0.930	N1 (1:30)	Amel	N/A	0.940	0.940
F1(1:10)	D5S818	0.902	0.900	0.900	N1 (1:30)	D5S818	0.965	0.960	0.960
F1(1:10)	FGA	U/I	U/I	U/I	N1 (1:30)	FGA	dropout	dropout	dropout
Avg. Mx		**0.886**	**0.905**	**0.891**	**Avg. Mx**		**0.962**	**0.958**	**0.963**

Sample	Locus	GMIDX	FSS-i^3	DNA_DA
G1 (1:30)	D8S1179	dropout	dropout	dropout
G1 (1:30)	D21S11	dropout	dropout	dropout
G1 (1:30)	D7S820	0.967	0.960	0.970
G1 (1:30)	CSF1PO	U/I	U/I	U/I
G1 (1:30)	D3S1358	U/I	U/I	U/I
G1 (1:30)	TH01	dropout	dropout	dropout
G1 (1:30)	D13S317	0.949	dropout	0.950
G1 (1:30)	D16S539	dropout	dropout	dropout
G1 (1:30)	D2S1338	dropout	dropout	dropout
G1 (1:30)	D19S433	dropout	0.970	dropout
G1 (1:30)	vWA	dropout	dropout	dropout
G1 (1:30)	TPOX	N/A	dropout	0.980
G1 (1:30)	D18S51	0.962	0.960	0.960
G1 (1:30)	Amel	N/A	0.930	0.930
G1 (1:30)	D5S818	0.903	dropout	0.900
G1 (1:30)	FGA	U/I	U/I	U/I
Avg. Mx		**0.945**	**0.955**	**0.948**

Table 4: Average mixing proportions and mixing proportion per locus as reported by GeneMapper® *ID-X*, FSS-i^3, and DNA_DataAnalysis for all NEST Project Samples. U/I is an uninformative locus. N/A is an uninformative locus according to Applied Biosystems.

Evaluation of Proficiency Test Samples

Forty-three proficiency test samples were obtained from Seattle Cancer Care Alliance (SCCA) and processed using GeneMapper® *ID-X*, FSS-i^3 and DNA_DataAnalysis. The average mixing proportions calculated by all three programs were concordant. The mixing proportion of each sample was calculated blindly, then compared to the results published by the administering agency, American Society for Histocompatabiity and Immunogenetics (ASHI) and to the results calculated by SCCA. The results demonstrate that the deconvolution software and SCCA calculations are very similar. The mixing proportions vary slightly from the ASHI mean, an

average reported by 40 laboratories, which indicates that a greater variation arises from the amplification and electrophoresis conditions of the samples, not from the mixture calculation.

Test set	Sample	GMIDX	FSS-i3	DNA_DA	ASHI mean	SCCA lab
LTFU1805	EMO-08	85.7	86.2	86.6	87.0	85.0
	EMO-09	77.5	78.6	78.9	79.4	78.0
	EMO-10	93.9	96.0	94.0	94.7	97.0
LTFU1875	EMO-13	31.9	32.1	32.1	38.6	34.0
	EMO-14	9.0	9.0	8.7	10.5	5.0
	EMO-15	15.2	15.2	14	21.4	15.0
LTFU1976	EMO-03	79.4	77.9	78.9	78.5	79.0
	EMO-04	74.2	73.9	74.4	73.2	72.0
	EMO-05	24.0	23.9	23.9	20.3	23.0
LTFU2053	EMO-08	30.6	31.8	31.7	30.3	31.0
	EMO-09	32.9	31.5	31.5	30.1	34.0
	EMO-10	33.3	32.7	32.3	30.9	35.0
LTFU2124	EMO-13	63.2	63.5	63.3	66.8	63
	EMO-14	72.0	72.0	72.3	75.7	73.0
	EMO-15	51.5	51.5	51.7	59.2	52.0
LTFU2241	EMO-03	68.4	67.7	67.5	71.8	68.0
	EMO-04	70.8	70.8	70.9	76.2	72.0
	EMO-05	62.8	63.6	63.1	67.1	63.0
LTFU2295	EMO-08	12.3	13.4	13.1	11.2	12.0
	EMO-09	0.04	0.04	0.06	0.8	0.0
	EMO-10	7.2	7.0	8.3	5.7	5.0
LTFU2378	EMO-13	49.1	51.4	50.5	49.2	52.0
	EMO-14	56.7	54.4	55.9	51.2	56.0
	EMO-15	64.2	62.8	63.7	60.8	64.0
LTFU2475	EMO-03	68.9	69.3	68.8	72.9	68.0
	EMO-04	100.0	100.0	100.0	99.4	100.0
	EMO-05	42.2	42	41.9	48.1	42.0
LTFU2548	EMO-08	40.9	40.9	41.2	42.6	40.0
	EMO-09	28.1	27.9	28.4	30.4	28.0
	EMO-10	59.7	60.1	60.3	57.2	60.0
LTFU2619	EMO-13	62.3	63.7	62.6	60.8	63.0
	EMO-14	90	89	89.2	89.8	90.0
	EMO-15	84.1	84.6	84.4	84.7	84.0
LTFU2896	EMO-21	88.2	87.2	88.2	89.3	88.0
	EMO-22	21.2	21.7	21.9	22.3	21.0
	EMO-23	100.0	99.0	96.0	98.1	100.0
	EMO-24	75.1	75.3	75.6	74.8	75.0
	EMO-25	25.2	25.5	25.4	26.2	25.0
LTFU3029	EMO-26	23.1	23.2	23.2	21.8	21.0
	EMO-27	29.3	29.2	29.2	27.13	29.0
	EMO-28	32.9	33	33.5	31.92	32.0
	EMO-29	100.0	100.0	100.0	100.0	100.0
	EMO-30	79.8	79.2	79.5	80.13	80.0

Table 5: Comparison of average mixing proportion across 16 loci for the proficiency test samples as calculated by GeneMapper® *ID-X*, FSS-i^3, DNA_DataAnalysis, ASHI, and SCCA.

Seen below is a demonstration of the mixing proportion output generated by GeneMapper® *ID-X*, FSS-i^3 and DNA_DataAnalysis for the same proficiency test sample, ASHI 2241-EMO-5 at the D3S1358 locus. All three programs gave concordant results for the sample.

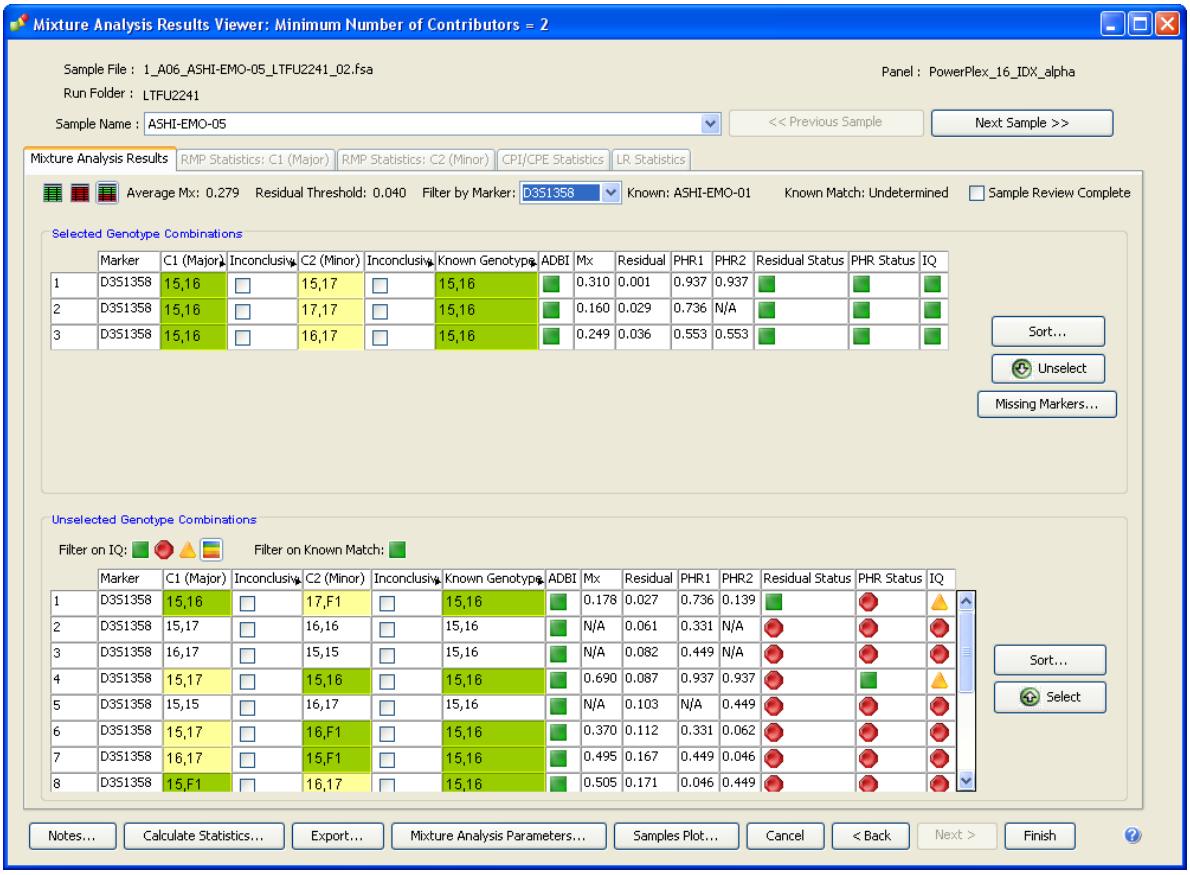

Figure 11: Sample output from GeneMapper® *ID-X* for ASHI sample 2241-EMO-5 at the D3S1358 locus.

Gel Number	Lane Number	Barcode	Case Number	Item Number	Item and Area
-	-	-	-	-	-

Pref Amp Tolerence	Mixing Proportion Tolerence	Homozygote
50%	20%	150

Weight: Maximum	Weight: Minimum	Weight: Mean
44% 1:1	33% 2:1	40% 1:1

-			Possible Contributors				Pref Amp Rule				Mix Prop Rule		-				
Locus	Allele	Height/Area	Contributor 1		Contributor 2		Contributor 1		Contributor 2		Mix Est		RC	Contributor 1		Contributor 2	
D3S1358	15	1649	16	17	15	15	217%	N	-	Y	48% 1:1	Y	-	-	-	-	-
	16	1210	16	17	15	16	55%	Y	55%	Y	82% 1:5	N	-	-	-	-	-
	17	558	15	17	15	16	93%	Y	93%	Y	69% 1:2	N	-	-	-	-	-
	-	-	17	17	15	16	-	Y	136%	Y	84% 1:5	N	-	-	-	-	-
	-	-	15	16	15	17	93%	Y	93%	Y	31% 2:1	Y	Include	**15**	**16**	**15**	**17**
	-	-	16	17	15	17	20%	N	20%	N	63% 1:2	N	-	-	-	-	-
	-	-	16	16	15	17	-	Y	296%	N	65% 1:2	N	-	-	-	-	-
	-	-	15	16	16	17	55%	Y	55%	Y	18% 5:1	N	-	-	-	-	-
	-	-	15	17	16	17	20%	N	20%	N	37% 2:1	Y	-	-	-	-	-
	-	-	15	15	16	17	-	Y	217%	N	52% 1:1	Y	-	-	-	-	-
	-	-	15	17	16	16	296%	N	-	Y	35% 2:1	Y	-	-	-	-	-
	-	-	15	16	17	17	136%	Y	-	Y	16% 5:1	N	-	-	-	-	-
	Database Consolidation for D3S1358													**15**	**16**	**15**	**17**

Figure 12: Sample output from FSS-i^3 for ASHI sample 2241-EMO-5 at the D3S1358 locus.

1 A06 ASHI-EMO-05 LTFU2241 02.fsa

D3S1358: For a 2-contributor 3-allele mixture of types AA & BC: 3-combinations; Alleles (RFUs): 15 (1611), 16 (1186), 17 (533)
Only combinations including the following reference profiles are included: (15, 16) (15, 17) ; all peak height ratios are >= (0.4);

D3S1358: For a 2-contributor 3-allele mixture of types AB & BC: 3-combinations; Alleles (RFUs): 15 (1611), 16 (1186), 17 (533)
Only combinations including the following reference profiles are included: (15, 16) (15, 17) ; all peak height ratios are >= (0.4);
15, 16 (phr = 0.94; p = 0.69); 15, 17 (phr = 0.94; p = 0.31)

Figure 13: Sample output from DNA_DataAnalysis for ASHI sample 2241-EMO-5 at the D3S1358 locus.

Evaluation of Post-Transplant Bone Marrow Engraftment Samples

Post-transplant bone marrow engraftment samples were obtained from City of Hope National Medical Center from two patient-donor pairs and were evaluated using the available mixture deconvolution software programs. All three software programs were able to quickly determine the mixing proportion of the recipient profile. For Case #1, a comparison of the mixing proportion of the recipient whole blood and CD3 T-cell subpopulation show higher recipient cells in the T-cell fraction, which may be an indication of relapse.

Locus	COH-05369 (whole blood)			COH-05370 (CD3 cells)		
	GMIDX	DNA_DA	FSS-i3	GMIDX	DNA_DA	FSS-i3
D8	0.60	0.60	0.60	0.57	0.57	0.56
D21	N/A	0.50	0.50	N/A	0.51	0.51
D7	U/I	U/I	U/I	U/I	U/I	U/I
CSF1PO	0.55	0.55	0.55	0.58	0.58	0.58
D3	U/I	U/I	U/I	U/I	U/I	U/I
TH01	N/A	0.47	0.47	N/A	0.60	0.61
D13	0.48	0.48	0.48	0.61	0.61	0.61
D16	0.53	0.53	0.52	0.56	0.56	0.55
D2	U/I	U/I	U/I	U/I	U/I	U/I
D19	0.52	0.52	0.52	0.61	0.61	0.62
vWA	0.52	0.52	0.52	0.59	0.59	0.58
TPOX	0.57	0.57	0.57	0.60	0.60	0.60
D18	0.47	0.47	0.47	0.50	0.50	0.50
Amel	N/A	0.63	0.62	N/A	0.60	0.59
D5	N/A	0.50	0.51	N/A	0.57	0.57
FGA	U/I	U/I	U/I	U/I	U/I	U/I
AVG	**0.529**	**0.528**	**0.528**	**0.575**	**0.575**	**0.573**

Table 6: Case #1, proportion of recipient cells present in post-transplant peripheral whole blood and CD3 T-cell sub-population.

Case #2 follows the chimerism monitoring of one patient at varying time points post-transplant. At day 1 of monitoring, the patient has mixed chimerism of approximately 50%. At day 34, there is a small decrease in the recipient cell percentage. At day 60, the patient would be categorized as having increasing mixed chimerism, because the recipient cells have increased by approximately 6% from the previous sample. At this step, the physician would likely take preventive action, such as a donor lymphocyte infusion to prevent graft rejection. The GeneMapper® *ID-X* average mixing proportion deviates slightly from those calculated by FSS-i^3 and DNA_DataAnalysis, especially on day 1 and day 87. Because GeneMapper® *ID-X* does not calculate a mixing proportion for a locus of 2 peaks with one shared allele, this reduces the number of loci used to calculate the average mixing proportion to seven. FSS-i^3 and DNA_DataAnalysis both use eleven loci, causing the observed variation between the programs.

Locus	COH09-00346 (day 1)			COH09-00413 (day 34)		
	GMIDX	FSSi^3	DNA_DA	GMIDX	FSSi^3	DNA_DA
D8S1179	0.49	0.49	0.49	0.44	0.44	0.44
D21S11	0.55	0.55	0.55	0.49	0.50	0.49
D7S820	U/I	U/I	U/I	U/I	U/I	U/I
CSF1PO	U/I	U/I	U/I	U/I	U/I	U/I
D3S1358	0.55	0.55	0.55	0.52	0.52	0.52
TH01	N/A	0.34	0.34	N/A	0.64	0.65
D13S317	N/A	0.53	0.54	N/A	0.4	0.4
D16S539	U/I	U/I	U/I	U/I	U/I	U/I
D2S1338	N/A	0.40	0.40	N/A	0.45	0.45
D19S433	0.53	0.52	0.53	0.46	0.46	0.46
vWA	0.46	0.47	0.46	0.47	0.47	0.47
TPOX	N/A	0.53	0.53	N/A	0.48	0.48
D18S51	0.567	0.58	0.57	0.45	0.43	0.44
Amel	U/I	U/I	U/I	U/I	U/I	U/I
D5S818	U/I	U/I	U/I	U/I	U/I	U/I
FGA	0.54	0.52	0.54	0.51	0.51	0.51
AVG	**0.528**	**0.498**	**0.500**	**0.478**	**0.4818**	**0.483**

Locus	COH09-00492 (day 60)			COH09-05505 (day 87)		
	GMIDX	FSSi^3	DNA_DA	GMIDX	FSSi^3	DNA_DA
D8S1179	0.46	0.46	0.46	0.49	0.49	0.49
D21S11	0.58	0.57	0.58	0.47	0.47	0.47
D7S820	U/I	U/I	U/I	U/I	U/I	U/I
CSF1PO	U/I	U/I	U/I	U/I	U/I	U/I
D3S1358	0.56	0.57	0.56	0.623	0.62	0.63
TH01	N/A	0.68	0.67	N/A	0.49	0.49
D13S317	N/A	0.50	0.51	N/A	0.38	0.38
D16S539	U/I	U/I	U/I	U/I	U/I	U/I
D2S1338	N/A	0.49	0.49	N/A	0.39	0.39
D19S433	0.49	0.49	0.49	0.51	0.51	0.51
vWA	0.52	0.51	0.52	0.54	0.53	0.54
TPOX	N/A	0.54	0.54	N/A	0.50	0.50
D18S51	0.61	0.61	0.61	0.50	0.50	0.50
Amel	U/I	U/I	U/I	U/I	U/I	U/I
D5S818	U/I	U/I	U/I	U/I	U/I	U/I
FGA	0.61	0.60	0.61	0.53	0.52	0.53
AVG	**0.546**	**0.547**	**0.549**	**0.523**	**0.491**	**0.494**

Table 7: Case #2, proportion of recipient cells present in peripheral whole blood at varying times post-transplant.

Evaluation of the Effect of Stutter on Mixing Proportions

Stutter was carefully studied to determine the impact it has on the calculation of mixing proportions. Eighteen samples were randomly chosen to be processed using DNA_DataAnalysis with the stutter removal feature. The amount that stutter affects the average mixing proportion will depend on the number of loci which have an allele in a stutter position and whether the major or minor allele is in a stutter position. Stutter had a minimal effect on the average mixing proportion of samples which are closer to a 1:1 mixing proportion, as seen in Table 8. However, samples with a lower mixing proportion, such as 1875-EMO-15 and 1805-EMO-8, are more affected by stutter. When the minor contributor's allele is in a stutter position to the major contributor's allele, this will have a much greater impact on the mixing proportion of the minor

Test Set	**Sample**	**With Stutter**	**With Stutter Removed**	**Difference**
1875	EMO-13	0.321	0.306	0.015
	EMO-15	0.152	0.131	0.021
2241	EMO-3	0.323	0.309	0.014
	EMO-4	0.292	0.276	0.016
	EMO-5	0.364	0.355	0.009
1805	EMO-8	0.138	0.112	0.026
	EMO-9	0.214	0.194	0.020
2053	EMO-8	0.318	0.303	0.015
	EMO-9	0.315	0.302	0.013
	EMO-10	0.328	0.312	0.015
2548	EMO-8	0.409	0.405	0.004
	EMO-9	0.279	0.271	0.008
	EMO-10	0.399	0.394	0.005
2378	EMO-13	0.514	0.511	0.003
	EMO-14	0.436	0.432	0.005
	EMO-15	0.372	0.361	0.011
2475	EMO-3	0.307	0.304	0.003
	EMO-5	0.420	0.408	0.013

Table 8: Comparison of average mixing proportion at 16 loci before and after the removal of stutter using DNA_DataAnalysis.

contributor, as seen in Figure 14. In the example, the stutter from the 18 allele is contributing approximately 270 RFU to the 17 allele. This artificially inflates the mixing proportion of the minor contributor to 21% of the mixture. Once the effect of stutter is removed, the mixing proportion of the minor contributor becomes 11%; this equates to a 48% decrease in the mixing proportion of the minor contributor at this locus. Determining whether a patient is experiencing increasing or decreasing mixed chimerism requires great accuracy when calculating the mixture proportions because an increase in as little as 5% recipient cells from the previous sample is the definition of increasing mixed chimerism. Because stutter has the ability to so greatly skew the mixing proportion of a minor contributor, the DNA_DataAnalysis stutter removal tool could prove very useful in engraftment monitoring. To avoid the effect of stutter, currently, some engraftment monitoring laboratories choose to exclude affected loci from the average mixing proportion calculation when the minor contributor is 10% or less of the mixture. When the major contributor's allele is in the stutter position, the mixing proportion is not impacted significantly.

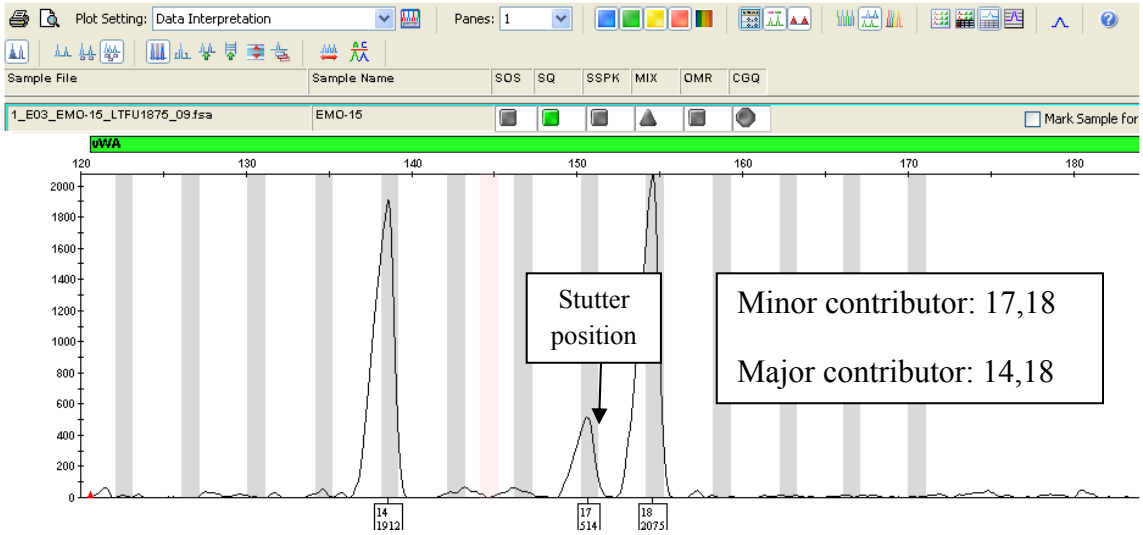

Figure 14: The effect of stutter on a minor allele for ASHI sample 1875-EMO-15.

CHAPTER IV

CONCLUSIONS

The goal of this project was to evaluate available mixture deconvolution software programs to determine their suitability for bone marrow engraftment monitoring. GeneMapper® *ID-X*, FSS-i^3 and DNA_DataAnalysis were all able to process mixed DNA samples to obtain correct mixing proportions consistently and efficiently. TrueAllele® Casework was unable to be evaluated at this time, but should also be considered in a later study. DNA_DataAnalysis possesses the features which would make it most suitable to bone marrow engraftment monitoring, such as ease of use, speed, three-person mixture handling capability, and the ability to subtract stutter. Overall, the three programs were able to deconvolute two-person post-transplant samples with consistency, accuracy, and speed. Implementation of any of the programs into engraftment monitoring laboratories would provide a time-savings and would minimize inter- and intra-laboratory inconsistencies.

Stutter is a phenomenon which must be accounted for when calculating the mixing proportion of a post-transplant sample, especially when a minor contributor's allele is in a stutter position. The effect of stutter becomes greater on the minor allele when the mixing proportion of the minor contributor is small. As such, it is recommended that when the mixing proportion of the minor contributor is less than 10%, any locus with a minor allele in a stutter position not be

used in the average mixing proportion calculation or that the effect of stutter be removed. While many components of forensic short tandem repeat (STR) analysis is regulated, such as how many and which loci must be tested and which kits can be used, bone marrow engraftment monitoring has no such oversight. Many inconsistencies between laboratories due to the lack of uniformity in analysis and mixture calculations; laboratories use anywhere from 2-16 loci to calculate mixing proportion, some laboratories exclude loci if stutter is present in the minor allele while some do not, while some laboratories use peak height and other use peak area when calculating mixing proportions. If bone marrow engraftment laboratories had recommendations set forth regarding how many and which loci must be tested, as well as how to deal with stutter, this would further minimize the amount of intra-laboratory inconsistencies when calculating mixing proportions.

Future of Bone Marrow Engraftment Monitoring

A number of articles have been published recently which focus on the use real-time PCR as a method for bone marrow engraftment monitoring [21, 22]. While STR-PCR chimerism analysis has been the gold standard for over a decade, its use is limited by a 1-5% sensitivity. Detection of low levels of recipient cells is especially important for early detection of leukemic relapse. The real-time PCR method is based on the detection of biallelic single nucleotide polymorphisms (SNPs). One primer is designed to specifically amplify each allele in the polymorphic region, and a second primer binds to a common region. During the extension phase of real-time PCR, the exoculease activity of the DNA-polymerase cleaves the bound fluorogenic probe, releasing the reporter dye from the proximity of the quencher, thereby increasing the fluorescent signal. The main advantage of the real-time PCR method over the traditional STR-

PCR method for chimerism monitoring is the absence of PCR competition and plateau biases, which results in greater accuracy and a sensitivity of 0.1%. In addition, a final result can be obtained in several hours. However, when using 10 SNP loci, the SNP-PCR method is only able to discriminate approximately 90% of patient and donor cells, whereas the STR-PCR method is able to discriminate greater than 99% of donor and patient cells when using 11 loci [21].

It is not uncommon for bone marrow transplant patient to receive an infusion of hematopoietic stem cells from two donors. This creates a three-person mixture which can be difficult to interpret and determine the percent chimerism of each individual. Although FSS-i^3 and GeneMapper® *ID-X* do not have the capability to deconvolute three-person mixtures, DNA_DataAnalysis does have this capability. A potential future study would be to evaluate three-person mixtures of known mixing proportions using DNA_DataAnalysis to determine if the program is able to identify the correct genotype combinations of the three contributors and calculate the correct mixing proportions.

REFERENCES

1. Scharf SJ, Smith AG, Hansen JA, McFarland C, Erlich HA. 1995. Quantitative determination of bone marrow transplant engraftment using fluorescent polymerase chain reaction primers for human identity markers. *Blood*, 85(7): 1954-63.
2. Kristt D, Israeli M, Narinski R, Or H, Yaniv I, Stein J, Klein T. 2005. Hematopoietic chimerism monitoring based on STRs: quantitative platform performance on sequential samples. *Journal of Biomolecular Techniques*, 16: 378-389.
3. Paul WE. 2008. Fundamental Immunolgy, 6th edition. Philadelphia, Pennsylvania: Lippincott, Williams, & Wilkins.
4. Cutler C, Antin JH. Peripheral blood stem cells for allogenic transplantation: a review. 2001. *Stem Cells*, 19(2): 108-117.
5. Woronzoff-Dashkoff KP, McGlennen RC. 2001. Monitoring of bone marrow transplant engraftment. Methods in Molecular Medicine, vol 49: Molecular Pathology Protocols. Totowa, New Jersey: Humana Press.
6. Bader P, Niethammer D, Willasch A, Kreyenberg H, Klingebiel T. 2005. How and when should we monitor chimerism after allogenic stem cell transplantation? *Bone Marrow Transplantation*, 35(2): 107-19.
7. Atkinson K. 2004. Clinical bone marrow and blood stem cell transplantation. Cambridge, UK: Cambridge University Press.

8. Sufliarska S, Minarik G, Horakova J, Bodova I, Bojtarova E, Czako B, Mintrik M, Drgona L, Lakota J, Krivosikova M, Kovacs L. 2007. Establishing the method of chimerism monitoring after allogenic stem cell transplantation using multiplex polymerase chain reaction amplification of short tandem repeat markers and Amelogenin. *Neoplasma*, 54(5): 424-30.
9. College of American Pathologists ME-B & ME-C Monitoring Engraftment Survey. 2008.
10. Chalandon Y, Vischer S, Helg C, Chapuis B, Roosnek E. 2003. Quantitative analysis of chimerism after allogenic stem cell transplantation by PCR amplification of microsatellite markers and capillary electrophoresis with fluorescence detection: the Geneva experience. *Leukemia*, 17(2): 228-31.
11. Nollet F, Billiet J, Selleslag G, Criel A. 2001. Standardization of multiplex fluorescent short tandem repeat analysis for chimerism testing. *Bone Marrow Transplantation*, 28: 511-518.
12. Young L, Bostwick V, Lovelace R, Carney C, Leach T, Miles C, Roby R. A focused study on one major benefit of expert system implementation: the time savings. Proceeding from the Nineteeth Symposium on Human Identification. 2008, Hollywood, CA.
13. NIJ Expert Systems Testbed Project Mixture Deconvolution Workshop, Marshall University, Huntington, West Virginia. May 11-12, 2009.
14. ASHI proficiency testing program: www.ashi-hla.org/lab-center/pt-program/
15. Applied Biosystem's GeneMapper® *ID-X*: http://marketing.appliedbiosystems.com/mk/get/GENEMAPPER_IDX_MLANDING
16. Promega's FSS-i^3: http://www.promega.com/profiles/802/ProfilesinDNA_802_08.pdf

17. CyberGenetic's TrueAllele® Casework: http://www.cybgen.com/systems/cw_features.shtml
18. Okoye DI, Wecht CH. 2007. Forensic Investigation and Management of Mass Disasters. Tuscon, Arizona: Lawyers & Judges Publishing Inc.
19. Dolph AM. DNA_DataAnalysis v.2.1 User's Manual. U.S. Army Criminal Investigation Laboratory.
20. Gill P, Brenner CH, Buckleton JS, Carracedo A, Krawczak M, Mayr WR, Morling N, Prinz M, Schnieder PM, Weir BS. 2006. DNA commission of the International Society of Forensic Genetics: recommendations on the interpretation of mixtures. *Forensic Sciencce International*. 160: 90-101.
21. Alizadeh M, Bernard M, Danic B, Dauriac C, Birebent B, Lapart C, Lamy T, Le Prisé P, Beauplet A, Bories D, Semana G, Quelvennec E. 2002. Quantitative assessment of hematopoietic chimerism after bone marrow transplantation by real-time quantitative polymerase chain reaction. *Blood*, 99(12): 4618-25.
22. Koldehoff M, Steckel NK, Hlinka M, Beleen DW, Elmaagacli AH. 2006. Quantitative of chimerism after allogenic stem cell transplantation by real-time polymerase chain reaction with single nucleotide polymorphisms, standard tandem repeats, and Y-chromosome-specific sequences. *American Journal of Hematology*, 81:735-746.