Nemours

Thomas Jefferson University

Evaluation of STR-PCR and SNP-Pyrosequencing for Forensic and Biomedical Monitoring

Deborah L. Stabley[1], Adeline Richez[1], Donald Kristt[2] and Katia Sol-Church[1*]

[1]Alfred I. duPont Hospital For Children, Wilmington DE; [2]Rabin Medical Center, Petach Tikvah Israel

[*] Corresponding Author: ksolchur@nemours.org

Center for Pediatric Research

ABSTRACT

When cells of two individuals are mixed in a single sample, identifying each source will have implications, for example, in criminal identification, prenatal testing, and stem cell post-transplantion monitoring. The most expeditious approach to this type of analysis relies on identifying the unique DNA markers of each of the individuals represented in the sample. For this reason, the choice of DNA genotyping technology should have an impact on the accuracy and sensitivity of such an evaluation. To examine this possibility we undertook a comparison of the genotyping performance of two techniques that utilize highly polymorphic markers, namely, STR-PCR and SNP-Pyrosequencing. This study is particularly timely, since SNP-based genotyping has been touted as a more sensitive alternative to STRs.

For the present study, two types of DNA mixtures were investigated: (1) artificial chimeric mixtures, fabricated in varying proportions from DNA samples derived from two healthy unrelated individuals; (2) chimeric clinical samples.

We found that STR-based genotyping was still the most reliable and sensitive method. This was particularly true for the detection of low levels of chimerism. SNP-based genotyping using pyrosequencing did not support the contention that it is the wave of the future for these applications.

INTRODUCTION

Human identification based on genotyping technology has a broad range of applications in both the biomedical and forensic sciences. The goal of this study was to identify the most sensitive and accurate technology for genotyping by comparing two in-house techniques, Short Tandem Repeats (STRs) and Single Nucleotide Polymorphisms (SNPs). STRs are highly polymorphic, widely distributed, short repetitive sequences that can be detected using polymerase chain reaction (PCR). For this study the AmpFISTR ® Profiler Plus™ PCR Amplification Kit from Applied Biosystems was used to amplify a set of 9 STRs as well as the Amelogenin gender marker. In this method, one primer for each of the loci amplified is labeled with either the Fam, Joe or Ned dyes which are detected as blue, green and yellow respectively on a genetic analyzer. The panel of SNP biomarkers consisted of 14 high allele frequency SNPs, as described in Hochberg *et al* (2003). In the current study, the technique of pyrosequencing was selected for SNP genotyping. During pyrosequencing, immobilized biotinylated PCR products containing the SNP of interest are sequenced using real time chemistry coupled with an enzyme-cascade system.
In clinical studies, such as Bone Marrow Transplantation (BMT), sensitivity for a minor component in a DNA mixture is critical for patient care. After BMT, persistence of recipient hematopoietic cells, referred to as *mixed chimerism*, is often associated with disease relapse. A more sensitive technique to detect chimerism will hopefully improve the outcome of the patient by determining the onset of relapse or remission.

We report here our preliminary evaluation of two genotyping techniques, STR-PCR and SNP-pyrosequencing, to identify cellular chimerism in clinical samples as well as artificial DNA mixtures. STR-PCR, which is the current standard in mixed cell monitoring was compared with SNP-pyrosequencing because the latter is reputed to be more sensitive in this application.

This work is supported in part by grant 1P20RR20173 from the National Institute of Health NCRR and by translational research funds from the Nemours' Children's Clinic to KSC.

RESULTS

Analysis Methods: Genotyping considerations in cases of mixed DNA samples

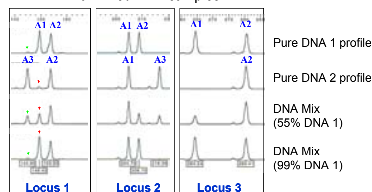

Figure 1: STR-PCR profiles of DNA 1 and 2 and mixtures of these DNAs. The chromatograms display three different loci used to exemplify methods for data analysis discussed in Table 1. Arrowheads mark stutter peaks which are the same size (in bp) as major informative alleles. These stutter peak may create interference in the data.

Table 1: Allelic Combinations in Mixed DNA samples

	STR-PCR DNA 1 alleles		DNA 2 alleles		Equation used for calculation of percent DNA 1 present in mix
	A1	A2	A3	A4	%=A1+A2/(A1+A2+A3+A4)
	A1	A2	A3	A3	%=A1+A2/(A1+A2+A3)
	A1	A1	A2	A3	%=A1/(A1+A2+A3)
	A1	A1	A2	A2	%=A1/(A1+A2)
Locus 1 →	A1	**A2**	A3	**A2**	%=A1/(A1+A3)
Locus 2 →	**A1**	A2	**A1**	A3	%=A2/(A2+A3)
	A1	**A1**	**A1**	A2	Non-informative or Normalize
Locus 3 →	A1	**A2**	**A2**	**A2**	Non-informative or Normalize

Locus 1 and locus 2 use the general equation % DNA 1 = Sum of the peak area of DNA 1 alleles / Sum of peak area of DNA 1 and DNA 2 alleles. Locus 3 may be either declared non-informative (due to allele sharing) or normalized using %= Mixed (A1/A1+A2) / DNA 1 (A1/A1+A2).

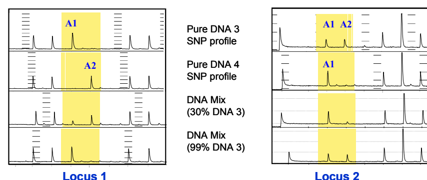

Figure 2: SNP-pyrosequencing profiles of DNA 3 and 4 genotypes at two loci. The SNP regions are highlighted in yellow. The pyrograms of 2 mixtures of these DNAs are also presented. For data analysis see Table 2.

Table 2: Analysis of SNPs in Mixed DNA samples

	SNP-Pyrosequencing DNA 3 Alleles		DNA 4 Alleles		Percent DNA 3 present in mix
Locus 1 →	A1	A1	A2	A2	%=A1/(A1+A2)
	A1	**A1**	**A1**	A2	Non-informative or Normalize
Locus 2 →	**A1**	A2	**A1**	**A1**	Non-informative or Normalize

Calculation of % DNA 3 in mixed samples uses a simplified equation in cases when the 2 DNAs are homozygous polymorphic (as in locus 1). For locus 2, when only 1 DNA carries an informative allele (A2), normalization can be applied using the % of A2 in the mixed DNA / % of A2 in the pure DNA carrying the informative allele (%= Mixed (A2/A1+A2) / DNA 3 (A2/A1+A2)), where A1 and A2 represent peak heights.

Artificial Chimerism: Genotyping by STR-PCR and SNP-pyrosequencing of mixed DNA populations

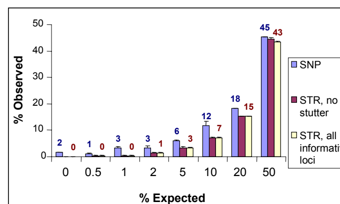

Figure 3: Artificial mixtures (DNA 1 ranging from 0 to 50%) were genotyped using SNP and STR biomarkers. STR results are calculated 2 ways, with loci having no stutter peak involvement or with all informative loci. There was no significant difference between including or excluding loci with stutter peak interference.

Clinical Chimerism: Testing of clinical samples collected from patients undergoing therapeutic bone marrow transplantation (BMT)

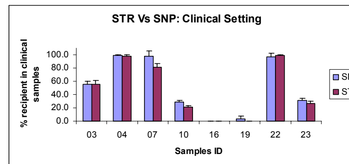

Figure 4: Clinical samples were genotyped using SNP and STR Biomarkers. % of recipient DNA was calculated and the average of all informative STR and SNP loci were plotted side by side. Details of % recipient for each loci are included in the table below. Overall, there is a great concordance between the 2 sets of data. Exceptions were samples 07 and 19 for which greater data variability was observed (see Table 3).

Table 3: Compilation of % Recipient calculated at each of the loci

ID #	SNP 1	SNP 2	SNP 3	SNP 4	SNP 5	STR 1	STR 2	STR 3	STR 4	STR 5	STR 6
03	60.9	53.7	53.0			56.2	53.6	55.1	50.5	64.5	
04	99.4	100.4	98.0			99.2	96.5	100.0	96.1		
07	92.0	103.6				86.3	82.7	81.8	74.1		
10	31.7	26.9	28.8			20.7	18.5	20.2	24.7		
16	0.0	0.0	0.0			0.0	0.0	0.0			
19	4.5	0.0	0.0	9.0		0.0	0.0	0.0	0.0	0.0	0.0
22	88.1	100.0	98.7	99.8		97.8	96.2	97.6	100.0	100.0	100.0
23	35.1	27.2	32.8	32.9	28.6	28.3	26.7	30.4	22.3	29.7	24.3

In general there were more informative loci using the STRs than the SNPs to detect chimerism in recipients receiving bone marrow from HLA matched donors. Samples showing discordance between the 2 methods are shaded.

METHODS

Artificial mixtures: Genomic DNAs 1 and 2 were isolated from saliva collected from two unrelated healthy volunteers using the Oragene saliva kit (DNA Genotek), and their concentrations determined by Pico green. DNA 1 was mixed with DNA 2 in different proportions : 0%, 0.5%, 1%, 2%, 5%, 10%, 20%, and 50% DNA 1, at a final concentration of 10 ng/ul total DNA. Subsequent dilution of this artificial set was performed (1 ng/ul) prior to STR testing.
Clinical DNA samples: Residual DNA samples from BMT patients treated at the Rabin medical center were collected after diagnosis, de-identified and sent to the dupont Lab for comparative testing. DNA concentrations were assessed by picogreen and/or spectrophotometry.
SNP-Pyrosequencing: For each SNP, 100ng of DNA was used to PCR amplify the region of interest. In this method, one of the primers was biotinylated to allow subsequent immobilization of a single stranded PCR product onto Streptavidine Sepharose HP (Amersham Pharmacia Biotech, Uppsala, Sweden). Specific sequencing primers, located adjacent to the SNP, were then annealed to the single-stranded template and samples subjected to pyrosequencing using the PSQ96 Pyro Gold SNP Reagent Kit (Pyrosequencing AB, Uppsala, Sweden), following the manufacturer's recommendation.
STR-PCR: Ten different markers were co-amplified in a single reaction using the AmpF/STR Profiler Plus PCR amplification kit. The PCR was performed using 2 ng of DNA. Separation of the PCR products and fluorescence detection was performed on an ABI Prism 310 Genetic Analyzer.

DISCUSSION

In this initial study, using artificial mixtures enabled us to compare side by side two techniques for the quantitation of mixed chimerism, STR-PCR and SNP-pyrosequencing. We found that, overall, both platforms are equally suited to detect mixed DNA (Figure 3), and that SNP-pyrosequencing does not appear to offer additional sensitivity for the detection of low level chimerism despite expectations to the contrary. We observed that SNP-pyrosequencing appeared more unreliable for detection of low level contamination due to background noise introduced by the pyrosequencing chemistry. Furthermore, we found STRs to offer more choices of polymorphic loci from which to determine an average value for the % chimerism in clinical samples (Figure 4).

Optimized Genotyping of Mixed Chimerism:

➢ Quantiative chimerism analysis must rely on informative loci when available:
• For SNPs the DNAs should be homozygous polymorphic (eg. A/A vs. G/G). The % recipient or % chimerism in the mixed sample is determined automatically using the PSQ96 allelic quantification (AQ) software.
• For STRs both DNAs should have at least one non-shared informative allele, preferably with no stutter peak involvement. The % chimerism or % DNA1 = Sum of informative alleles of DNA 1 / Sum of all informative alleles in DNA1+DNA2.
• To prevent bias created by locus-specific differential allele amplification it is important to use the average of all informative loci for data reporting for both SNP and STR genotyping analysis.

➢ When few informative loci are available (e.g. Figure 1, locus 3):
• Normalization protocol can be applied to the data. In this case, allelic amplification differential must remain constant in independent PCR and show minimal technical variations.
• For both STRs and SNPs the % chimerism = % of the informative allele in the mixed DNA / % of the informative allele in the pure DNA.

➢ In the clinical setting post-BMT, a longitudinal evaluation of sequential samples is the only way to detect trends in disease progression. This is routinely feasible with *ChimerTrack* software (*Leukemia* 2004b;**18**:909-911).

Finally, it is worth noting that for both platforms the observed values for the artificial mixtures are significantly lower than the expected. This suggests that either an error in DNA concentration evaluation occurred prior to the making of the artificial mixtures or that one of the DNA amplified better under our PCR conditions.

Nemours

Nemours is one of the largest pediatric group practices in the United States, serving children in Delaware, Maryland, New Jersey, Pennsylvania, Florida, and Georgia. Visit us online at www.Nemours.org and www.PedsEducation.org.