Large panels of SNPs for human identity typing are feasible with current generation sequencing (CGS) technology

Seung Bum Seo[1], Jonathan L King[1], David H Warshauer[1], Carey P Davis[1], Jianye Ge[1], Bruce Budowle[1,2]

[1] Institute of Applied Genetics, Department of Forensic and Investigative Genetics, University of North Texas Health Science Center, 3500 Camp Bowie Blvd., Fort Worth, Texas 76107
[2] Center of Excellence in Genomic Medicine Research (CEGMR), King Abdulaziz University, Jeddah, Saudi Arabia

Forensic DNA typing can provide useful information on human identification, such as in criminal cases and mass disasters. Short tandem repeats (STRs) are the primary genetic markers used because of their high discrimination power and relatively short amplicon size. However, some evidence samples are highly degraded and may not be characterized well with the current battery of STRs. Single nucleotide polymorphisms (SNPs) are an alternate set of markers that may be applied successfully to degraded samples. Most SNP containing amplicons can be designed to be smaller than 150 bp and, potentially, as short as 50-60 bp in length. However, current instrumentation in a typical crime laboratory is not amenable to typing a large battery of SNPs.

Current Generation Sequencing (CGS) provides a platform for more comprehensive coverage of genetic markers. CGS technologies sequence specified DNA targets in a massively-parallel fashion with high coverage and high throughput. Moreover, with the high-throughput capacity afforded by CGS, many different samples may be sequenced simultaneously by use of barcoding.

The Ion AmpliSeq™ HID SNP panel (Life Technologies), a primer pool of 103 autosomal SNPs and 33 Y-SNPs, was evaluated using the Ion 314™ Chip on the Ion PGM™ Sequencer (Life Technologies) with four DNA samples. Procedures and methods for the SNP typing were described on Ion Community (1). DNA was sequenced at different target quantities from 10 ng to 100 pg to span the range of amounts that might be encountered in databasing and caseworking laboratories. Genotypes were obtained for all 136 SNPs for the three male samples and 103 SNPs for the female sample with 10 ng of template DNA (Fig. 1). With 1 ng of DNA, most SNPs were detected and typed correctly; there was one example of high heterozygote imbalance across the four samples. With 100 pg of DNA, an average of 1.6 SNP loci were not detected and an average of 4.3 SNPs showed a heterozygote imbalance <20% across the samples. All barcoded samples showed high autosomal SNP allele coverage averaging 945X with 10 ng of template DNA, 792X with 1 ng of DNA and 689X with 100 pg of DNA. For Y-SNPs, the samples showed high average allele coverage of 465X with 10 ng of template DNA, 350X with 1 ng of DNA and 257X with 100 pg of DNA. Average heterozygote allele coverage ratios were 89.6%, 70.7% and 61.8% with 10 ng, 1 ng and 100 pg of template DNA, respectively. Successful and accurate typing with approximately 1 ng of initial template DNA is promising and indicates that the sensitivity of detection of CGS technology may reach the sensitivity of detection of current forensic DNA typing methods.

In the analyses using the standard 10 ng of template DNA, all genotypes were the same as those obtained from an in-house GAIIx SNP panel (Illumina, Inc), except for SNP rs1029047, where an incorrect result was induced by adjacent homopolymers (Fig. 2). SNPs residing adjacent to homopolymers may require further scrutiny before placing them in panels and, even more so, before selecting them as core markers. Given that all but one of the SNPs tested were concordant between CGS systems and that hundreds of SNPs may be typable, panels with substantial SNP overlap are possible. Moreover, it is likely that very few SNPs may not be amenable to one particular platform. Thus, SNP marker compatibility should be attainable within the forensic community. Other factors such as simplicity of library preparation, cost, labor, coverage, accuracy, robustness and especially software design may be more meaningful for selecting a particular platform or system.

This study, although limited to four samples, indicated that typing of samples for a large battery of SNPs is feasible and that CGS technology may be a reality for characterizing reference samples for national DNA databases in the near term. In addition, applications of massive SNP typing in forensic genetics go beyond databasing and make feasible testing such as of distant familial relationships. The sensitivity observed also suggests that casework analyses may be possible. Further efforts will focus on an improved SNP panel for human identification, balancing the amplicon yield to obtain similar coverage across more of the SNPs in a panel, elucidating the contributing factors of allele drop-out, and alignment and interpretation guidelines for single-source and mixture samples.

References

1. Ion community (2013) Single Nucleotide Polymorphisms (SNPs) – Identity. http://ioncommunity.lifetechnologies.com/community/applications/hid/snps. Accessed 5 October 2013

Figure legend

Fig. 1. Representative allele coverage chart for the genotyping of 136 SNPs. The typing results generated using 10 ng of genomic DNA are shown. Genotypes of Y-SNPs were not detected in sample no. 1 because this profile was from a female. All other samples were male. Purple bar: C, Green bar: A, Orange bar: G, Red bar: T.

Fig. 2. SNP typing results at rs1029047, which resides adjacent to an A homopolymer and a T homopolymer. The locus appears to be heterozygous, showing a mixture of T and A. However, the correct genotype at this SNP locus is AA, as determined by the allele count and the correction of misaligned SNPs. Black bar, (-): Deletion. Purple bar, (I): Insertion. Shading: low quality base.

Acknowledgement

This manuscript was written based on the paper entitled Single nucleotide polymorphism typing with massively parallel sequencing for human identification (Seo et al., Int J Legal Med, 2013 June 5, DOI: 10.1007/s00414-013-0879-7). This project was supported in part by Award No. 2012-DN-BX-K033, awarded by the National Institute of Justice, Office of Justice Programs, US Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect those of the US Department of Justice. We would like to thank Life Technologies for providing its technical expertise and contributing to the sample preparation and sequencing chemistries used in this study. In addition, we would like to thank Illumina, Inc. for its contributions for generating the in-house SNP panel.