

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

## Forensic Science International: Genetics Supplement Series

journal homepage: [www.elsevier.com/locate/fsigss](http://www.elsevier.com/locate/fsigss)

## Population data for 94 identity SNPs in four U.S. population groups

K.M. Kiesler<sup>\*</sup>, L.A. Borsuk, C.R. Steffen, K.B. Gettings, P.M. Vallone

US National Institute of Standards and Technology, NIST 100 Bureau Drive, Gaithersburg, MD 20899-8314, USA

## ARTICLE INFO

## Keywords:

Single nucleotide polymorphism (SNP)  
Identity informative SNP  
DNA sequencing

## ABSTRACT

The U.S. National Institute of Standards and Technology (NIST) sequenced 1036 human DNA samples from four United States population groups (African American, Asian, Hispanic, and Caucasian) using the ForenSeq DNA Signature Prep Kit with Primer Mix B (DPMB) on a MiSeq FGx instrument. In addition to STR markers, DPMB includes amplification primers for single nucleotide polymorphisms (SNPs) used for individual identification (iiSNPs,  $n = 94$ ), ancestry inference (aiSNPs,  $n = 56$ ), and phenotype prediction (piSNPs,  $n = 22$ ). Resulting sequencing coverage information was interpreted for the 94 iiSNP markers. Here we present performance characteristics of the ForenSeq DNA Signature Prep Kit in the population studied.

## 1. Introduction

With the arrival of bench-top scale next generation DNA sequencing (NGS) platforms suitable for applications typical of forensic human identification (HID), the community has witnessed the development of commercial multiplexes for various HID applications. Many of these kits address the need for backward compatibility with length-based measurement of short tandem repeat (STR) markers, the standard analytical technique in use for DNA-based HID for over 20 years. However, NGS technology offers advantages over length-based measurement in terms of multiplexing capacity. Each individual marker is recognized by its unique sequence, eliminating both the need for fluorescent dye labelling of PCR fragments and the constraint of non-overlapping amplicon size range for electrophoretic separation. Free of these restrictions, kit manufacturers have optimized amplicon size and added non-traditional marker types such as single nucleotide polymorphisms (SNPs) selected for specific uses in human identification. SNPs with evenly distributed allele frequency across many global populations are suitable for one-to-one matching [1], whereas those with alleles distributed uniquely to specific populations may be used to infer biogeographic ancestry in an unknown analyte [2]. SNPs in certain genes may also allow prediction of externally visible characteristics such as hair, eye, and skin color [3]. The ForenSeq DNA Signature Prep Kit (Verogen, San Diego CA, USA) is a ‘megaplex’ containing a combination of STR and SNP markers. In addition to autosomal, Y-chromosome, and X-chromosome STR content, the kit includes two options for SNP content: DNA primer mix A (DPMA) contains primers for amplification of 94 identity informative SNPs (iiSNPs) while DNA primer mix B (DPMB) adds 56 ancestry informative

SNPs (aiSNPs) and 22 phenotype informative SNPs (piSNPs) to the same 94 iiSNPs in DPMA. To characterize technical performance of the iiSNP content found in DPMB, we interpret and summarize sequencing coverage for these loci herein.

## 2. Materials and methods

All work has been reviewed and approved by the National Institute of Standards and Technology Research Protections Office. Briefly, 1036 human DNA samples were analyzed with the ForenSeq DNA Signature Prep Kit using DPMB on a MiSeq FGx instrument (Verogen) as previously described [4]. Data analysis was performed using Universal Analysis Software (UAS) (Verogen). Genotypes and sequencing read depth data were exported to Excel (Microsoft, Redmond WA, USA). Allele coverage ratio (ACR) was calculated for each heterozygous genotype by dividing the coverage depth value from the allele with lower coverage by the value from the allele with higher coverage, resulting in an ideal value of one for perfectly balanced coverage for each heterozygous genotype. Total coverage for each locus was calculated as the average value for all genotypes meeting defined genotype calling criteria: minimum read depth of  $31\times$  for homozygote, minimum of  $11\times$  for each allele for heterozygote, and allele coverage ratio of  $\geq 0.2$  for heterozygotes,  $< 0.2$  for homozygotes. Genotypes not meeting these thresholds were omitted from calculations.

## 3. Results and discussion

Plots of heterozygote balance and average sequencing depth for 94

<sup>\*</sup> Corresponding author.

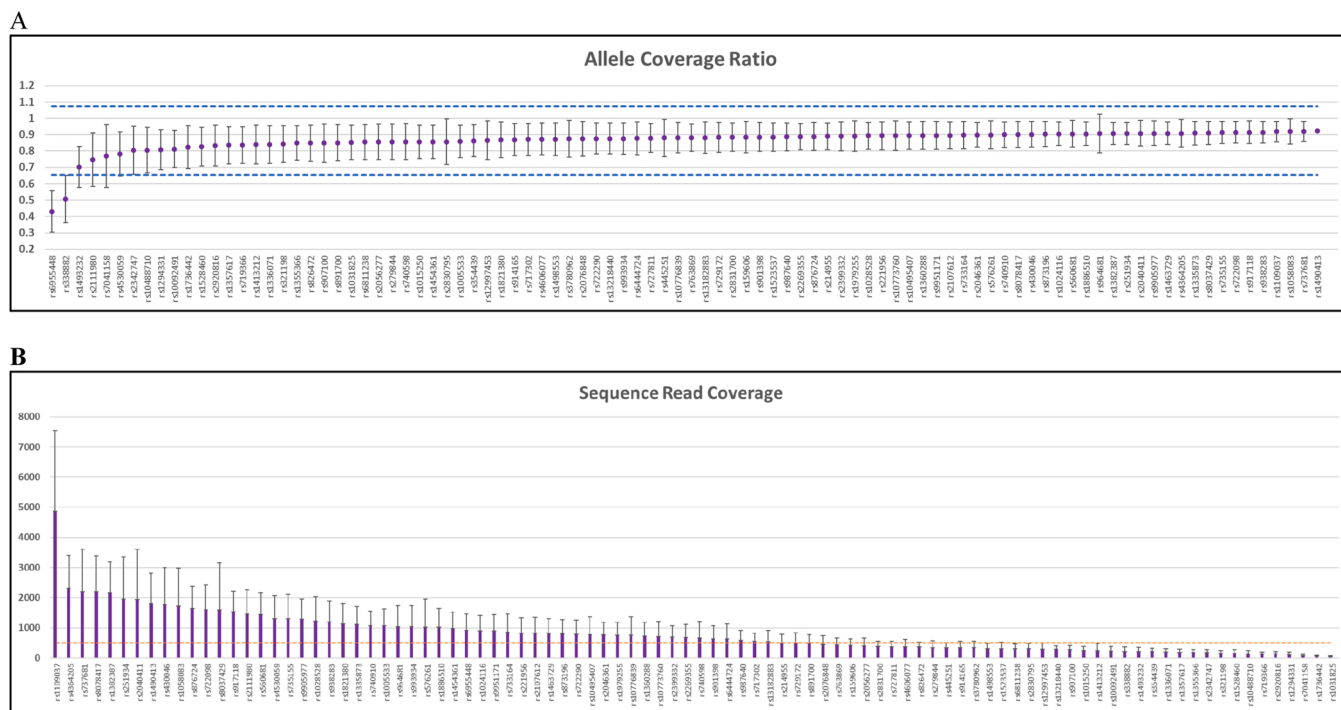
E-mail address: [kevin.kiesler@nist.gov](mailto:kevin.kiesler@nist.gov) (K.M. Kiesler).

<https://doi.org/10.1016/j.fsigs.2022.09.003>

Received 20 September 2022; Accepted 22 September 2022

Available online 23 September 2022

1875-1768/Published by Elsevier B.V.



**Fig. 1.** Allele coverage ratio (A) for 94 iSNPs; average values are represented by closed circles with standard deviation denoted by error bars, dotted reference lines represent three standard deviations of the ACR values. (B) Sequencing read coverage for 94 iSNPs; the average value is shown in the histogram and standard deviation is represented by the error bar; a reference line is drawn depicting the mean coverage value (508×).

iSNPs in DPMB are presented in Fig. 1. Two SNP loci had notably low ACR values. Locus rs6955448 had the lowest ACR with an average value of 0.43 and range of 0.20–0.99 with the Alternate (“Alt”) allele, T, receiving lower coverage in most cases (455 instances out of 460 heterozygote calls). Locus rs338882 had an average ACR of 0.51 with a range of 0.23–0.99 with the Reference (“Ref”) allele, C, receiving lower coverage in most cases (507 out of 513 heterozygote calls).

**4. Conclusions**

Two SNP loci were observed to have a sequencing coverage imbalance exceeding three standard deviations of the ACR values of the 94 iSNPs. The consistent nature of the allele exhibiting lower coverage for these two loci could be an indication that an additional SNP associated with either the “Ref” allele or “Alt” allele lies within the primer binding site for the locus. Overall, sequencing coverage depth was variable by two orders of magnitude in the population study.

**Role of funding**

This work was supported in part by the NIST Special Programs Office: Forensic Genetics and in part by an interagency agreement with the U.S. Federal Bureau of Investigation: DNA as a biometric.

**Conflict of interest**

None.

**Acknowledgements**

Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Commerce. Certain commercial equipment, instruments, and materials are identified to specify experimental procedures clearly. In no case does such identification imply a recommendation or endorsement by NIST, nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose.

**References**

- [1] A.J. Pakstis, W.C. Speed, J.R. Kidd, K.K. Kidd, Candidate SNPs for a universal individual identification panel, *Hum. Genet.* 121 (3–4) (2007) 305–317, <https://doi.org/10.1007/s00439-007-0342-2>.
- [2] K.K. Kidd, et al., Progress toward an efficient panel of SNPs for ancestry inference, *Forensic Sci. Int. Genet.* 10 (2014) 23–32, <https://doi.org/10.1016/j.fsigen.2014.01.002>. Epub 2014 Jan 15.
- [3] L. Chaitanya, K, et al., The HirisPlex-S system for eye, hair and skin colour prediction from DNA: introduction and forensic developmental validation, *Forensic Sci. Int. Genet.* 35 (2018) 123–135, <https://doi.org/10.1016/j.fsigen.2018.04.004>.
- [4] K.B. Gettings, L.A. Borsuk, C.R. Steffen, K.M. Kiesler, P.M. Vallone, Sequence-based U.S. population data for 27 autosomal STR loci, *Forensic Sci. Int. Genet.* 37 (2018) 106–115, <https://doi.org/10.1016/j.fsigen.2018.07.013>.