Forensic Science International: Genetics Supplement Series 5 (2015) e363–e364

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

journal homepage: www.elsevier.com/locate/FSIGSS

A Strategy for Characterization of Single Nucleotide Polymorphisms in a Reference Material

Kevin M. Kiesler*, Katherine B. Gettings, Peter M. Vallone

National Institute of Standards and Technology, Gaithersburg, Maryland, USA

ARTICLE INFO

Article history:
Received 26 August 2015
Accepted 7 September 2015
Available online 25 September 2015

Keywords:
Single nucleotide polymorphism
Next generation sequencing
Concordance

ABSTRACT

The advent and adoption of next generation sequencing (NGS) is enabling analysis of single nucleotide polymorphisms (SNPs) at an unprecedented scale, limited primarily by multiplexing during the PCR amplification based enrichment step used for forensic applications. Since only a single nucleotide is assayed, PCR primers may be designed to generate small amplicons, making SNP markers well-suited to forensic DNA typing. Carefully selected panels of SNP markers have been previously established for forensic applications such as one-to-one matching, estimating biogeographical ancestry, and predicting externally visible phenotype [1–6]. To support the implementation of SNPs in forensic DNA analysis, NIST has examined the HID-Ion AmpliSeq Identity Panel and the HID-Ion AmpliSeq Ancestry Panel for the Ion Torrent Personal Genome Machine (PGM). In total, over 289 SNP markers have been typed. NIST intends to use a strategy combining NGS on orthogonal platforms and Sanger sequencing for characterizing the SNP markers for varying levels of confidence. The outcome will be to report only the SNP allele calls, analogous to the mitochondrial sequence variants in NIST Standard Reference Material (SRM) 2392, and not the subsequent application of the information (e.g. prediction of ancestry or phenotype).

Published by Elsevier Ireland Ltd.

1. Introduction

Multiplex amplification of 100 to 200 single nucleotide polymorphisms (SNPs) in a single polymerase chain reaction (PCR) has become a routine endeavour when typing SNP panels via next generation sequencing (NGS). Several panels of SNP loci have been proposed for performing varied functions in forensic DNA analysis such as one-to-one matching [1,2], ancestry estimation [3–5], and prediction of externally observable phenotypes [6]. These panels have been incorporated into commercial multiplex kits for NGS analysis. The likely adoption of SNP typing methodology in forensic genetics warrants consideration of strategies for making well characterized reference materials to ensure the robustness and accuracy of measurements of SNPs. To this end, NIST has performed pilot experiments to evaluate a strategy for characterizing SNP genotypes in a reference material.

NIST has three levels of confidence for reference material values: (1) Certified – NIST has the highest confidence in its accuracy; all known or suspected sources of bias have been investigated or taken into account, (2) Reference – a high confidence estimate of the true value but where all possible sources of bias have not been fully investigated by NIST, and (3) Informational – data that may be of interest and use to the SRM user, but insufficient information is available to assess the confidence of the assignment [7]. Pilot experimental data was examined for evidence of potential sources of bias or inaccuracy.

2. Materials and methods

NIST obtained commercially available versions of two SNP multiplexes from Life Technologies: the HID-Ion AmpliSeq Identity Panel (v.4.0) and the HID-Ion AmpliSeq Ancestry Panel (v.4.0), intended for use on the Ion Torrent Personal Genome Machine (PGM). Nine candidate reference material samples were typed in triplicate with each multiplex SNP assay. All samples are single source DNA extracted from Buffy coat white blood cells or cell cultures. Three samples constitute a family trio of Son, Father, and Mother from the Personal Genome Project (http://www.personalgenomes.org/). All NGS kits were employed according to the manufacturers' recommended procedures with the exception that the Ion Chef from Ion Torrent was used to perform emulsion PCR, enrichment of Ion Sphere Particles (ISPs), and loading of Ion 318 chips. A barcoded pool of 31 libraries (nine candidate reference material samples and one positive control in triplicate and one negative control) was run on a single Ion 318 chip for each SNP panel.

* Corresponding author.
E-mail address: kevin.kiesler@nist.gov (K.M. Kiesler).

http://dx.doi.org/10.1016/j.fsigss.2015.09.144
1875-1768/Published by Elsevier Ireland Ltd.

Table 1
Summary of SNP markers which displayed allelic imbalance (a) with average allele ratio calculated only for imbalanced genotype measurements, and (b) markers which displayed strand bias, with average strand bias calculated for all samples in the dataset for that locus.

(a) Allelic imbalance

Identity panel				Ancestry panel			
	% Ratio		S.D.		% Ratio		S.D.
rs7520386	71.0	+/−	0.7	rs734873	68.2	+/−	4.5
rs4530059	73.9	+/−	3.8	rs7722456	92.2	+/−	1.0
rs430046*	66.3	+/−	0.6	rs3943253	93.2	+/−	1.4
				rs4918664	94.8	+/−	0.5
				rs2899826	90.8	+/−	0.6
*Strand bias				rs7251928*	75.5	+/−	10.8

(b) Strand bias

Identity panel				Ancestry panel			
	% Bias		S.D.		% Bias		S.D.
M479**	17.1	+/−	2.4	rs1040045	39.3	+/−	6.3
rs430046***	70.1	+/−	11.9	rs1760921	73.6	+/−	8.5
rs1463729	63.8	+/−	2.5	rs1871428	35.5	+/−	2.6
rs2111980	39.3	+/−	3.5	rs1871534	36.5	+/−	2.8
rs4364205	63.0	+/−	2.5	rs2986742	46.7	+/−	8.9
				rs6990312	38.3	+/−	10.2
				rs7251928**	40.1	+/−	13.3
**Low coverage				rs9845457	45.2	+/−	5.9
***Allelic imbalance				rs12629908	41.6	+/−	4.6

3. Results and discussion

The Ion AmpliSeq Library Preparation for Human Identification Applications manual (Pub. No. MAN0010640 Rev A.0) suggests a minimum sequencing coverage of 300 X for autosomal SNPs and 150 X for Y-chromosome SNPs. To achieve this minimum coverage for all loci in a panel, recommended average coverage depth for the Identity Panel is 738 X and 594 X for the Ancestry Panel. These recommended values allow for locus-to-locus variations in sequencing coverage balance. This translates to a maximum throughput of 77 samples in a library pool for the Identity Panel or 59 samples for the Ancestry Panel when using an Ion 318 Chip. Average sequencing coverage per sample for the Identity Panel was 1747.3 X +/− 531.5 X for autosomal loci and 849.5 X +/− 236.4 X for Y-chromosome loci. For the Ancestry Panel, the average was 1063.1 X +/− 354.0 X.

Two (2) out of 2430 (0.08%) Identity Panel autosomal SNP measurements (both for locus rs2342747) had sequencing coverage below 300 X. The number of Identity Panel Y-chromosome SNP data points with coverage below 150 X was one (1) out of 597 (0.17%) for the locus M479. In the Ancestry Panel data, 90 SNP genotypes out of a total 4455 (2.02%) were below the suggested minimum 300 X coverage; however, many of these were outliers and not representative of overall marker performance.

Allelic imbalance, calculated as the ratio of sequencing reads from the allele with highest coverage divided by the other allele(s) present, was considered to be an indicator of bias in measurement. If the ratio fell between the arbitrary thresholds of above 65% and below 95%, averaged across three replicate data points for a sample, that genotype was considered uncertain. There were three loci in the Identity Panel where heterozygote genotypes had imbalance >65%, while in the Ancestry Panel there were two loci with imbalanced heterozygotes and four loci with imbalanced homozygotes. These loci are summarized in Table 1(a).

Strand bias, or the ratio of positive strand reads to negative strand reads (+strand/−strand), was also considered an indicator of a low confidence genotype when observed to deviate from the ideal 50%. For initial screening, an arbitrary threshold was set for any one sample with an average bias above 65% or below 35% to identify loci exhibiting strand bias. An overall average value was then calculated across all samples for that locus, shown in Table 1(b). Five loci in the Identity Panel and nine loci in the Ancestry Panel were identified by this screening method with average strand bias above 60% and below 40%. These markers will require comprehensive characterization.

Discordant replicate genotypes were observed in two instances in the Ancestry Panel (rs3943253: AA, AA, NN, and rs7251928: AC, AC, and CC). Both of these loci displayed allelic imbalance which most likely contributed to the discordances. Results will be confirmed by Sanger sequencing.

4. Conclusions

Although no individual sample fell below the recommended average coverage for each panel (data not shown), there were a small number of genotypes which did not achieve the recommended minimum sequence coverage level (0.25% for the Identity Panel and 2.02% for the Ancestry Panel). This indicates that there may be some markers which yield low coverage despite following the manufacturer's recommendations for sample throughput.

Observations of heterozygote imbalance, strand bias, and discordance among replicates warrant further examination of specific genotype calls before any level of information value may be assigned to a reference material.

NIST will continue to evaluate commercial SNP multiplexes for forensic use, with the expectation that once SNP typing becomes widely adopted in forensic genetics, the community may benefit from well characterized reference materials to ensure robust and accurate measurements of SNP genotypes.

Role of funding

This work was funded in part by the Federal Bureau of Investigation (FBI) inter-agency agreement DJF-13-0100-PR-0000080: “DNA as a Biometric”.

Conflict of interest

None.

Acknowledgement

None.

References

[1] E. Musgrave-Brown, D. Ballard, K. Balogh, et al., Forensic validation of the SNPforID 52-plex assay, Forensic Sci. Int. Genet. 1 (2007) 186–190.
[2] A. Pakstis, W. Speed, R. Fang, F. Hyland, M. Furtado, J. Kidd, K. Kidd, SNPs for a universal individual identification panel, Hum. Genet. 127 (3) (2010) 315–324.
[3] T.M. Karafet, F.L. Mendez, M. Meilerman, P. Underhill, S. Zegura, M. Hammer, New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree, Genome Res. 18 (5) (2008) 830–838.
[4] R. Nassir, R. Kosoy, C. Tian, P.A. White, L.M. Butler, G. Silva, R. Kittles, M.E. Alarcon-Riquelme, P.K. Gregersen, J.W. Belmont, F.M. De La Vega, M.F. Seldin, An ancestry informative marker set for determining continental origin: validation and extension using human genome diversity panels, BMC Genet. 10 (1) (2009) 39.
[5] K. Kidd, W. Speed, A. Pakstis, M. Furtado, R. Fang, A. Madbouly, M. Maiers, M. Middha, F. Friedlander, J. Kidd, Progress toward and efficient panel of SNPs for ancestry inference, Forensic Sci. Int. Genet. 10 (2014) 23–32.
[6] S. Walsh, F. Liu, A. Wollstein, L. Kovatsi, A. Ralf, A. Kosiniak-Kamysz, W. Branicki, M. Kayser, The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA, Forensic Sci. Int. Genet. 7 (2013) 98–115.
[7] May, W.E., Gills, T.E., Parris, R., Beck II, C.M., Fassett, J.D., Gettings, R.J., Greenberg, R.R., Guenther, F.R., Kramer, G., MacDonald, B.S., Wise, S.A., Definitions of Terms and Modes Used at NIST for Value-Assignment of Reference Materials for Chemical Measurements; NIST Special Publication 260–136 (2000); http://www.nist.gov/srm/publications.cfm.