Application Note

Evaluation of QIAGEN's universal next generation sequencing workflow and SNP panel including the Ion PGM™ system for identity and relationship testing

A.G. Kinnane[1], N. Jones[2], Z. Wu[3], M.E. Gahan[1], R. Daniel[4], Julie Chmielowski[2], Ivonne Peterman[2], Keith Elliott[5] and D. McNevin[1]

[1] National Centre for Forensic Studies, Faculty of Education, Science, Technology & Mathematics (ESTeM), University of Canberra, Australia

[2] QIAGEN Australia, 169 Chadstone Centre VIC 3148, Australia

[3] QIAGEN Sciences, Inc., Biological Research Content, 6951 Executive Way, Frederick, MD 21703, USA

[4] Office of the Chief Forensic Scientist, Victoria Police Forensic Services Department, Victoria, Australia

[5] QIAGEN Ltd, Skelton House, Lloyd Street North, Manchester, M15 6SH, UK

Introduction

QIAGEN's next generation sequencing portfolio is designed to support users of all major sequencing platforms, providing the end-user with more choice in both marker selection and workflow optimization. An example of QIAGEN's "universal" approach for human identification is the Individual Identity (II) SNP panel previously analyzed by Grandell and coworkers (1). The 140 SNP panel includes markers not available in panels from other providers and was originally developed for paternity applications as part of a workflow using the Illumina MiSeq® FGx platform. However, to be truly universal the panel must also perform on the other commonly used NGS platforms in HID: the Ion PGM System and the Ion S5™ System (Thermo Fisher Scientific). Here we present a preliminary evaluation of the 140 SNP panel as part of a workflow including the Ion PGM System, demonstrating high coverage and sensitivity as well as reproducible results across a range of sample types and upstream sample prep methods.

Methods

Panel design

The 88 bi-allelic Individual Identity (II) SNPs from the set characterized and published by Pakstis and coworkers (2) were combined with the 52 SNPs characterized and published by Sanchez and coworkers (3) in a 140 SNP panel. Each SNP in the panel is covered by two pairs of (forward and reverse) primers thereby providing a level of redundancy in case of primer binding site mutations or other primer related issues. The combined 140 SNP panel was previously described and evaluated on the MiSeq platform by Grandell and coworkers.

Samples

A sample set was prepared to enable comparison of a number of key variables within a concise study. Variables evaluated were body fluid, sensitivity, extraction method and donor. The sample set is described in Table 1.

Table 1. Samples analyzed

Donor	Source	Template amount (ng)	Extraction	PCR cycles	Replicates
A	Saliva	20	QIAsymphony®	20	3
A	Saliva	10	QIAsymphony	21	3
A	Saliva	5	QIAsymphony	22	3
A	Saliva	2.5	QIAsymphony	23	3
A	Saliva	1.25	QIAsymphony	24	3
A	Saliva	1	QIAsymphony	24	6
A	Saliva	0.625	QIAsymphony	25	3
A	Saliva	0.3125	QIAsymphony	26	3
A	Saliva	0.2	QIAsymphony	26	3
NCT		0	QIAsymphony	26	3
A	Blood	1	QIAsymphony	24	1
A	Semen	1	QIAsymphony	24	1
B	Saliva	1	QIAsymphony	24	1
C	Saliva	1	QIAsymphony	24	1
A	Saliva	1	QIAamp®	24	3
A	Saliva	1	ChargeSwitch	24	3
A	Saliva	1	Phenol/chloroform	24	1

Target enrichment

Purified DNA was quantified using the Investigator® Quantiplex® HYres Kit on the Applied Biosystems® 7500 Real-Time PCR System (Thermo Fisher Scientific) before target enrichment using the GeneRead™ DNAseq V2 140 ID SNP panel and the GeneRead DNAseq PCR Kit V2 (comprising oligonucleotides, enzymes and buffers) according to manufacturer's instructions. For

sensitivity studies both template DNA amount and PCR cycle number were varied from the recommendations in the manufacturer's instructions (20 cycles, 20 ng input DNA) as described in Table 1. The Applied Biosystems GeneAmp® PCR system 9700 (Thermo Fisher Scientific) was used for amplification. Confirmation of successful amplification was performed using the Bioanalyzer High Sensitivity DNA chip (Agilent). This QC step enables confirmation that amplification of fragments of the expected size was achieved.

Library preparation

Library preparation was then conducted on the amplified DNA using QIAGEN's GeneRead library preparation workflow designed for sequencing on the Ion PGM System: GeneRead DNA Library L Core Kit, GeneRead DNA L Amp Kit, and GeneRead Adapter L Set 12-plex, followed by cleanup using AMPure® XP beads (Agencourt). The Bioanalyzer was again used after library preparation and cleanup to verify the size shift associated with addition of sequencing adapters and to ensure the absence of adapter dimers. This step is also used to calculate the molarity of the library prior to library quantification. Library quantification was achieved with the GeneRead Library Quant Kit. All steps were conducted according to the manufacturers' instructions. Table 2 summarizes the pre-sequencing steps and the products used.

Table 2. Pre-sequencing products used

Step	Chemistry	Manufacture	QC
Quantitation	Investigator Quantiplex HYres Kit	QIAGEN	
PCR	GeneRead DNAseq IISNP Panel GeneRead DNAseq Panel PCR Kit	QIAGEN	
Cleanup	QIAGEN	Agencourt	✓
Library construction	AMPure XP beads	QIAGEN	
Cleanup	GeneRead DNA Library L Core Kit	Agencourt	
Library amplification		QIAGEN	
Cleanup	QIAGEN	Agencourt	✓
Library quantitation	AMPure XP beads	QIAGEN	

Sequencing

Sequencing was carried out on the Ion PGM System using the Ion OneTouch® II System for template preparation, the Ion OneTouch ES for target enrichment and Ion 314™ v2 chips for sequencing. Six barcoded libraries were pooled on each Ion 314 v2 chip used. All systems were operated using the manufacturer's instructions.

Analysis

Raw data was analyzed using the Torrent Suite 5.0.2 analysis software, followed by variant calling using the Variant Caller Plugin 5.0.2.1, both from Thermo Fisher Scientific. Coverage of reference and alternate alleles was extracted from variant call (VCF) files. ▷

The minimum observed coverage for any amplicon was 40x, so no coverage threshold was applied. The minimum allele frequency was arbitrarily defined as 20%. This resulted in genotypes deemed as homozygous if one allele had >80% frequency and heterozygous with two alleles having frequencies between 20% and 80% (see Figure 4). Heterozygote balance was analyzed based on allele read frequency (ARF) values. ARF was defined, for each marker, as the number of reads for a reference allele divided by the total number of reads for the specific marker.

Results were evaluated for coverage, heterozygous balance, sensitivity, reproducibility across sample prep methods, and genotype concordance.

The impact of increased PCR cycle number (from the manufacturers recommended 20 cycles) to counter the reduced template input (from the recommended 20 ng) at target enrichment was evaluated.

Results

Quality checks

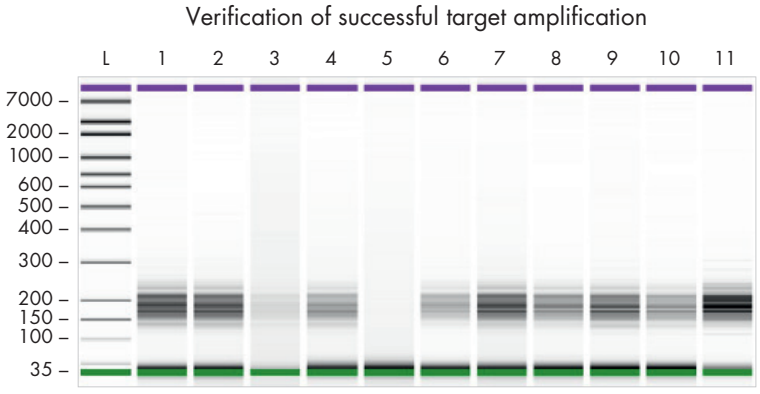

Figure 1. Bioanalyzer result for post-target amplification. The band between 150 and 250 bp indicates successful PCR amplification in the expected size range.

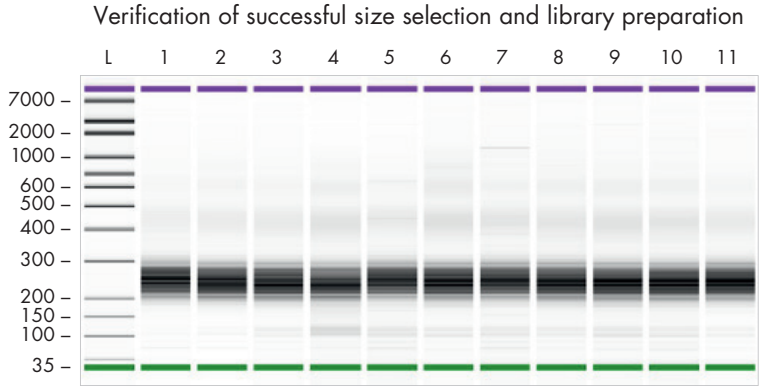

Figure 2. Bioanalyzer result for post-library preparation. The band has shifted in size due to the addition of sequencing adapters. The absence of strong bands below 160 bp confirms successful removal of any remaining adapter artefacts, e.g., dimers.

Coverage

Median coverage for all SNPs was over 100x with the exception of rs1360288 which has previously demonstrated poor coverage in NGS panels (Chris Phillips, personal communication), including in the same panel when sequenced on the MiSeq (1). In general, the markers exhibited high and uniform coverage.

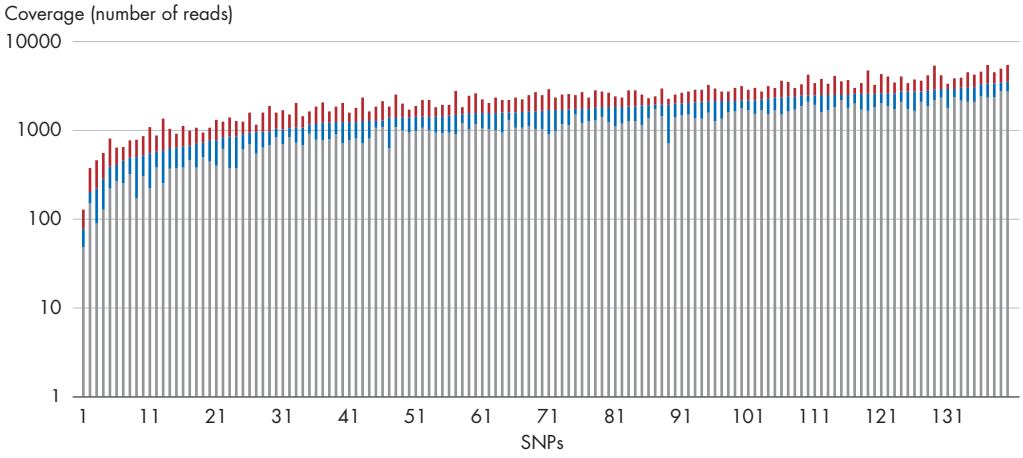

Figure 3. Coverage distribution for all 140 SNPs ordered from lowest median coverage (left) to highest median coverage (right). Coverage range above and below the median for each SNP is shown in red and blue, respectively.

Allele frequencies and heterozygous balance

Allele frequency values were plotted against number of reads indicating distinct bands representing homozygous and heterozygous genotypes (Figure 4). Such distinct bands can be used to define heterozygous calling thresholds and are therefore important to distinguish heterozygous samples from mixed or contaminated samples. Plots of ARF values for this panel as described by Gandell and coworkers (1) show a similar distribution when analyzed on the MiSeq, with distinct bands demarcating homozygous (values >0.9 and <0.1) and heterozygous (values between 0.4 and 0.6) genotypes.

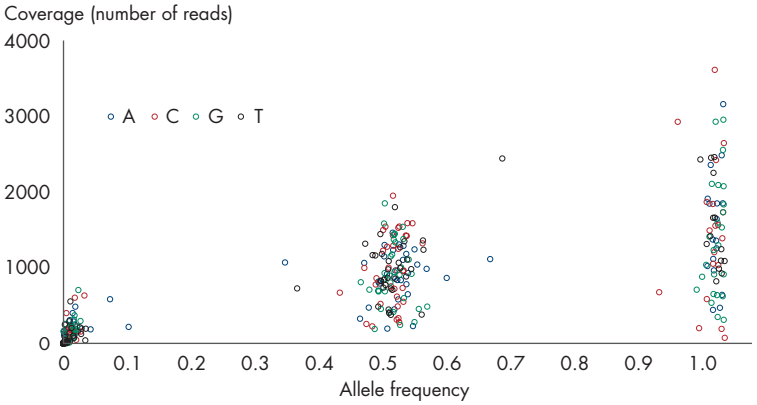

Figure 4. Mean allele coverage (vertical axis) as a function of mean allele frequency (horizontal axis) displaying homozygote range (90–100% allele frequency), heterozygote range (40–60% allele frequency) and non-consensus alleles calls (0–10% allele frequency).

Sensitivity

Sensitivity results are shown in Figure 5. On average 138 of the 140 SNPs represented in the panel were identified down to 0.2 ng template DNA. For most samples, rs1360288 proved the most susceptible to drop out. This SNP is described above and previously as problematic in NGS enrichment panels and will likely be removed from future iterations of the QIAGEN ID SNP Panel. In no sample were less than 98% of SNPs identified, down to 0.2 ng input DNA. These results suggest that, with careful readjustment of PCR cycles, a high level of sensitivity can be achieved with this panel, making it applicable to forensic samples containing low levels of amplifiable DNA. More detailed validation studies are required, however, to confirm optimal cycling conditions for use in casework.

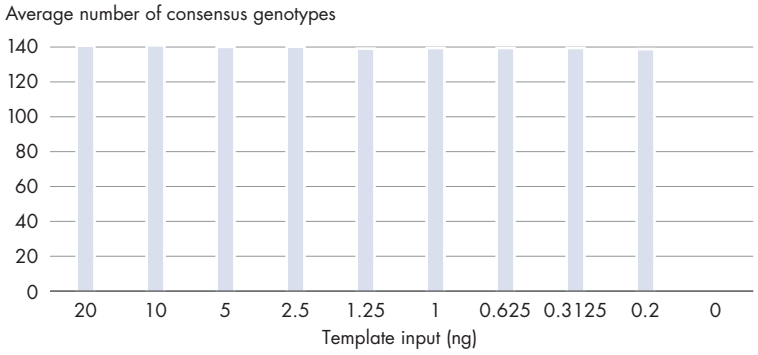

Figure 5. Sensitivity data showing numbers of consensus genotypes (140 genotypes being the maximum). Template input in nanograms varies from 20 ng down to 0 ng.

Source and extraction method

To observe any impact of DNA sample preparation and of body fluid type on the sequencing data, several sample prep methods and body fluids were compared in the study (Figure 6). Blood, semen and saliva performed comparably with the same input volume of DNA while sample prep method was also demonstrated to have no impact on the number of consensus genotypes.

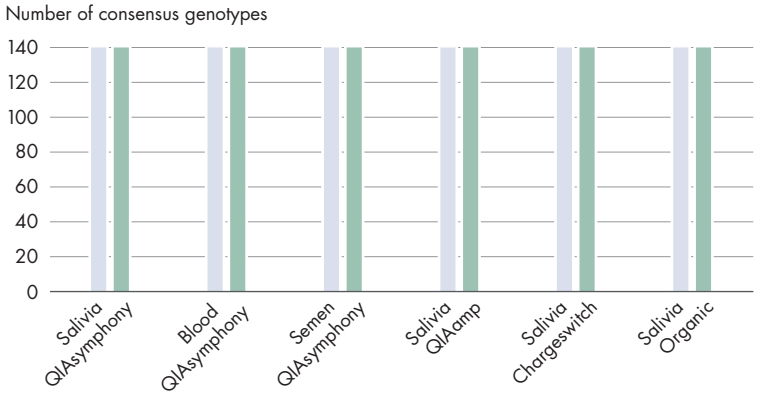

Figure 6. Comparison of tissue (body fluid) source and sample preparation methods on the number of consensus genotypes.

To provide a more qualitative assessment of the impact of sample prep method on sequencing, performance coverage for replicates of the three commercial sample prep kits was analyzed. Results are shown in Figure 7, demonstrating consistent coverage across the three DNA extraction methods.

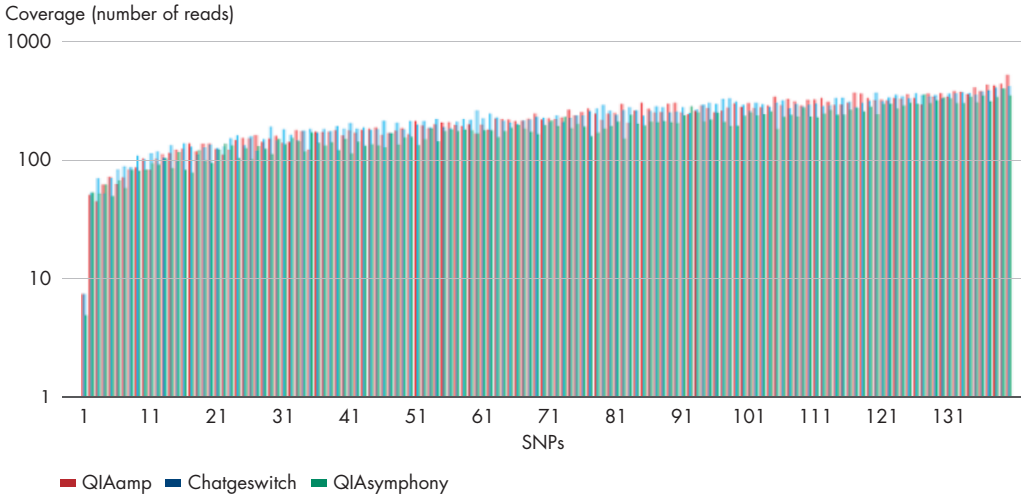

Figure 7. Coverage across all 140 SNPs for the three sample prep methods compared: QIAsymphony, Chargeswitch and QIAamp.

Conclusion

The QIAGEN 140 Individual Identity SNP panel was evaluated for the first time on the Ion Torrent™ PGM instrument. Results demonstrate high, consistent coverage, allelic balance and sensitivity comparable to other similar panels on this platform (in our experience [4]), and to the same panel when analyzed on the MiSeq. These results provide a practical example of the universal workflow approach from QIAGEN, which aims to make NGS solutions applicable to all major sequencing platforms, providing customers with choice and flexibility in their NGS work in HID.

Summary

- The GeneRead DNAseq IISNP panel can be integrated with the Ion PGM.
- It is sensitive to 0.2 ng DNA (with extra PCR cycles).
- Genotypes are concordant across blood, semen and saliva.
- Genotypes are concordant across extraction methods.

References

1. Grandell, I., Samara, R., and Tillmar, A.O. (2016) A SNP panel for identity and kinship testing using massive parallel sequencing. Int. J. Legal Med. **130**, 905–914. doi: 10.1007/s00414-016-1341-4
2. Pakstis, A.J., Speed, W.C., Fang, R., Hyland, F.C., Furtado, M.R., Kidd, J.R., and Kidd, K.K. (2010) SNPs for a universal individual identification panel. Hum. Genet. **127**, 315–324. doi:10.1007/s00439-009-0771-1
3. Sanchez, J.J. et al. (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification. Electrophoresis **27**, 1713–1724. doi:10.1002/elps.200500671
4. Daniel, R. et al. (2015) A SNaPshot of next generation sequencing for forensic SNP analysis. Forensic Sci. Int. Genet. **14**, 50–60. doi:10.1016/j.fsigen.2014.08.013

Ordering Information

Product	Contents	Cat. no.
Investigator Quantiplex HYres Kit (200)	Reaction Mix FQ, Primer Mix IC YQ, Control DNA Z1, QuantiTect Nucleic Acid Dilution Buffer	387116
GeneRead Custom DNAseq Gene Panels	Primer sets for targeted enrichment of genomic regions of interest	181902
GeneRead DNAseq Panel PCR Kit V2 (12)	GeneRead HotStarTaq polymerase, GeneRead DNAseq Panel 5x PCR Buffer, Nuclease-free Water; enough for 12 samples (4-pool panels) or 24 samples (1-pool panels)	181940
GeneRead DNAseq Panel PCR Kit V2 (96)	GeneRead HotStarTaq polymerase, GeneRead DNAseq Panel 5x PCR Buffer, Nuclease-free Water; enough for 96 samples (4-pool panels) or 192 samples (1-pool panels)	181942
GeneRead DNA Library L Core Kit (12)	For 12 reactions: Buffers and reagents for end-repair, ligation and nick-repair, for use with Ion Torrent instruments	180462
GeneRead DNA L Amp Kit (100)	For 100 reactions: Ready-to-use library amplification master mix and primer mix, for use with Ion Torrent instruments	180485
GeneRead Adapter L Set 12-plex (72)	For 72 reactions: 12 barcoded adapters for ligation to DNA library, for use with Ion Torrent instruments	180994
GeneRead Library Quant Kit	qPCR assays for quantifying Illumina libraries	333314
QIAsymphony SP	QIAsymphony sample prep module: includes 1-year warranty on parts and labor	9001297
QIAsymphony SP System	QIAsymphony sample prep module: includes installation and training, 1-year warranty on parts and labor	9001751
QIAsymphony Cabinet SP	Accessory for correct positioning of the QIAsymphony SP instrument	9020244

For up-to-date licensing information and product-specific disclaimers, see the respective QIAGEN kit handbook or user manual. QIAGEN kit handbooks and user manuals are available at **www.qiagen.com** or can be requested from QIAGEN Technical Services or your local distributor.