ThermoFisher SCIENTIFIC
The world leader in serving science

TECHNICAL NOTE

Maximize Forensic Genetic Analysis Potential Using an STR + SNP Combination NGS Workflow

The purpose of this Technical Note is to introduce forensic laboratories to the capabilities of combining Precision ID panels into a single workflow to maximize genetic information obtained from forensic samples. The flexibility of the Ion Chef System offers a simple, mixed content workflow to interrogate a total of 324 genetic markers with the commercially available Precision ID STR and SNP panels. This STR + SNP combination panel workflow includes STR's (31 autosomal STR's, Y indel, SRY, Amelogenin and DYS391),165 ancestry informative SNP's (aiSNP), 90 identity SNP's (iiSNP) and 34 Y-SNP's to achieve increased genetic representation suitable for both identity and investigative applications.

The Technical Note informs the community of the following attributes of the STR + SNP combination workflow:

- Benefits of mixed STR + SNP content to provide biogeographic ancestry and identity information in a single assay
- Workflow, genetic analysis requirements and system compatibility
  - Ion Chef™ and Ion S5™/S5 XL or Ion GeneStudio™ NGS Systems
  - Torrent Suite™ Software v5.10
  - Converge™ Software v2.2
- Performance data for the STR + SNP panel workflow including the following:
  - Sequencing run metrics (ISP loading, total useable reads, etc.)
  - Marker read coverage
  - Genotyping and tertiary analysis concordance for STR + SNP versus single panel workflows
  - Intralocus balance, stutter and artifact assessment

STR + SNP NGS Workflow

One of the many benefits of NGS applications is the ability to target diverse sets of genetic markers in a single assay. With the introduction of SNP analysis on NGS platforms, Human Identification (HID) laboratories can now combine STR identity applications with SNP-based analysis to ascertain biogeographic ancestry, Y lineage, phenotypic traits[1,2] and identity using large, informative NGS panels. Ion AmpliSeq library preparation technology combined with clonal amplification and massively parallel sequencing (MPS) greatly expands the pool of available genetic information to interrogate in forensic applications while preserving DNA samples that may be limited in both quantity and quality.

While modular, standalone panels may be better suited in focused investigations with highly degraded, challenging specimens containing limited DNA amounts – as fewer, small amplicons will ensure more complete genetic profiles – routine forensic samples may benefit from mixed genetic marker content, especially when investigations require identity (STR or SNP-based identity, Y or mtDNA haplogroup analysis) and investigative information (phenotypic traits or ancestry inference). One recent case involving DNA testing of a carbonized corpse discovered in France highlights the efficacy of targeting a range of genetic markers for investigative purposes when traditional STR analysis is unable to conclusively identify an individual[3].

As the forensic community refines the NGS application use cases, researchers and practitioners may benefit from experimenting with various marker combinations to determine targeted approaches to investigative questions when challenging, mixed and degraded specimens may be more suitable for NGS versus CE-based testing methods. The automated flexibility of the Ion Chef System is well-positioned to offer this type of *mix and*

match panel approach to answer a range of forensic inquiries. Table 1 provides a summary of the NGS offerings available from Thermo Fisher Scientific in a modular panel format.

Table 1. Summary of commercially available Thermo Fisher Scientific NGS panels

Panel	Catalog Number	Application	Marker Type	Genetic Markers
Precision ID mtDNA Whole Genome	A30938	Identity	Maternal lineage	mtGenome (1-16569)
Precision ID mtDNA Control Region	A31443	Identity	Maternal lineage	mtDNA control region (16024-576)
Precision ID Identity	A25643	Identity	Autosomal + paternal lineage	90 autosomal SNP's + 34 Y-SNP's
Precision ID Ancestry	A25642	Investigative leads	Autosomal	165 autosomal SNP's
Precision ID GlobalFiler™ NGS STR v2	A33114	Identity	Autosomal	31 autosomal STR's + Y indel, SRY, AMEL and DYS391
Ion AmpliSeq DNA Phenotyping*	NA	Investigative leads	Autosomal	23 autosomal SNP's
Ion AmpliSeq HID Y-SNP Research v1*	NA	Identity/Investigative leads	Paternal lineage	781 Y-SNP's

*Panel was developed in collaboration with forensic genomics researchers and available on Ampliseq Designer. Refer to citations[4],[5] below for details on panel performance and development studies performed.

For the panel list above, assay conditions were optimized during development with Precision ID chemistry and Torrent Suite Software to maximize overall performance and workflow compatibility for each individual panel. As such, run conditions for mtDNA and SNP-based panels are more similar than NGS STR Panel v2, which has different PCR cycling for library preparation as well as templating conditions, nucleotide flow order and the total number of flows in the sequencing run. While the Ion GeneStudio S5 Systems offer a high degree of scalability to support large NGS panels, some combinations (e.g., aiSNP + iiSNP) may perform better under similar conditions than others (e.g., STR + ai/iiSNP) when one considers chemistry and optimal run conditions through the various steps in the NGS workflow. Given that, we sought to assess performance for the STR + SNP combination workflow approach to include STR's (31 autosomal STR's, SRY, Amelogenin and DYS391) in combination with 165 aiSNP's, 90 iiSNP's and 34 Y-SNP's using standard procedures already developed for the suite of commercially available Precision ID panels listed above.

Methods

Performance of the STR + SNP combination workflow

For these studies, genomic DNA (gDNA) extracts from 12 known reference samples (8 males and 4 females) and 4 control DNA samples (HC-1018, NA24149, 9947A and 007) were used for library preparation for a total of 16 samples. As the Ion Chef System is designed to accommodate one or two-pool library preparation strategies, two configurations were tested as shown in Table 2. Commercially available primer pools were added to the DL8 reagents cartridge as described[6].

Table 2. STR + SNP Primer Pool Combinations for the Chef DL8 Run Setup

Combination	Panel Pool 1	Panel Pool 2	Total Targets per Pool	Primer Additions
A	STR (2X)	Ancestry + Identity w/ Y-SNP (2X)	35 / 289	For 8 samples on Chef, combine the following: 1. Pipet 150 uL of STR Panel to Position A tube in DL8 reagent cartridge. 2. Add 75 uL of Ancestry primer pool with 75 uL of Identity primer pool to Position B tube. 3. Vortex well and centrifuge before adding to DL8 reagent cartridge.
B	STR + Identity w/ Y-SNP	Ancestry SNP (2X)	159 / 165	For 8 samples on Chef, combine the following: 1. Add 75 uL of STR primer pool with 75 uL of Identity primer pool to Position A tube.

2. Vortex well and centrifuge before adding to DL8 reagent cartridge.

3. Pipet 150 uL of Ancestry panel to Position B tube in DL8 reagent cartridge.

Sample setup used for library preparation with the Ion Chef followed the protocol outlined in the *Precision ID SNP Panels with the HID Ion S5/HID Ion GeneStudio S5 System Application Guide (MAN0017767)* with the Precision ID DL8 Kit. Cycling parameters for library preparation on the Ion Chef System were modified to accommodate the two-pool system setup (Table 3).

Table 3. Library Preparation Parameters

Library Preparation Cycle Number	24 cycles
Anneal & Extension Time	4 minutes

Library pools were quantified using the Ion Library TaqMan® Quantitation Kit, diluted and pooled to 50 pM for templating on an Ion 530™ Chip using the Ion S5 Precision ID Chef & Sequencing kit. The effects of multiplexing 12 and 16 samples per Ion 530 chip on marker performance were assessed. Template plans were created using the '*Precision ID GlobalFiler STR Panel – S5'* template on Torrent Suite Software v5.10 (Figures 1 & 2). Custom target and hotspot region BED files[7] containing entries for STR's and SNP's were created to designate templating plans (Table 4).

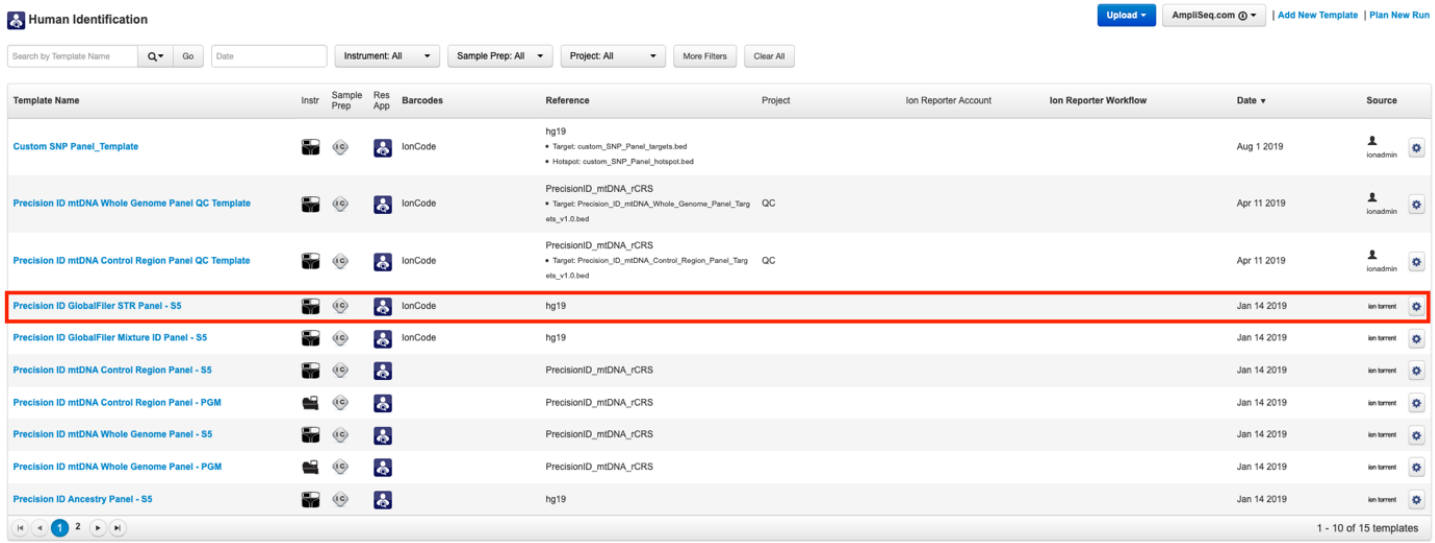


Figure 1. Templating plan setup on Torrent Suite Software for the STR + SNP combination panel workflow.

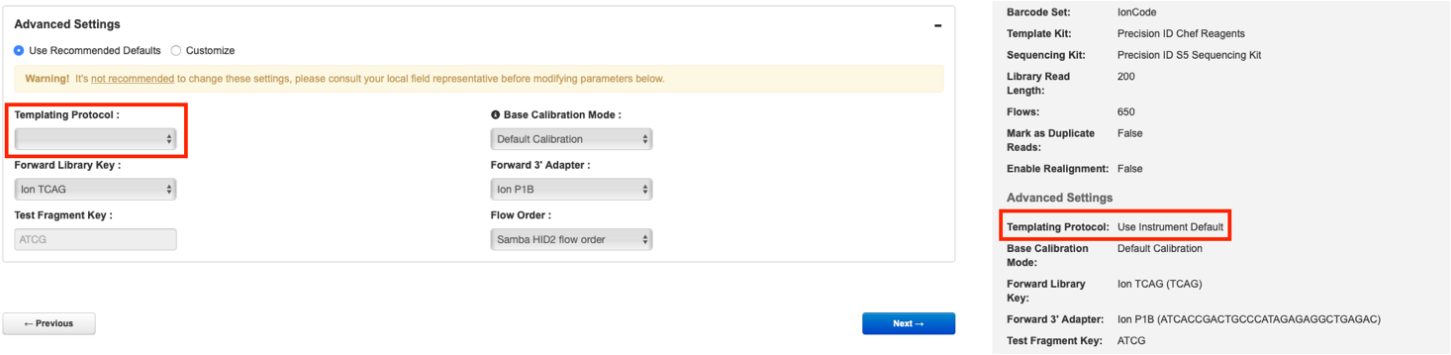


Figure 2. Ensure that the default instrument templating protocol (outlined in red) is selected for the templating plan

Sequencing was performed on an Ion S5 XL System on Torrent Suite Software v5.10. HID Genotyper v2.2 plugin was initiated twice for STR and SNP analysis using the respective target region and hotspot bed files (Table 4). Converge Software v2.2 with NGS Data Analysis module v1.2 was utilized for genotyping, Y-haplogrouping and ancestry prediction results. Pre-loaded frequency template files for the Precision ID Identity and Ancestry SNP Panels were used for Y-haplogrouping and ancestry analysis.

Table 4. Templating Run Plan Set-Up and Analysis Files

Templating Plan Set-Up		
Templating Plan	Target File	PrecisionID_ComboPanel_Targets.bed
	Hotspot File	PrecisionID_ComboPanel_Hotspot.bed
HID Genotyper v2.2 Plugin Analysis		
STR Analysis	Target File	Precision_ID_GlobalFiler_NGS_STR_Panel_Target_v1.1.bed
	Hotspot File	Precision_ID_GlobalFiler_NGS_STR_Panel_Hotspot_v1.1.bed
SNP Analysis	Target File	PrecisionID_ComboPanel_Targets_SNP.bed
	Hotspot File	PrecisionID_ComboPanel_Hotspot_SNP.bed

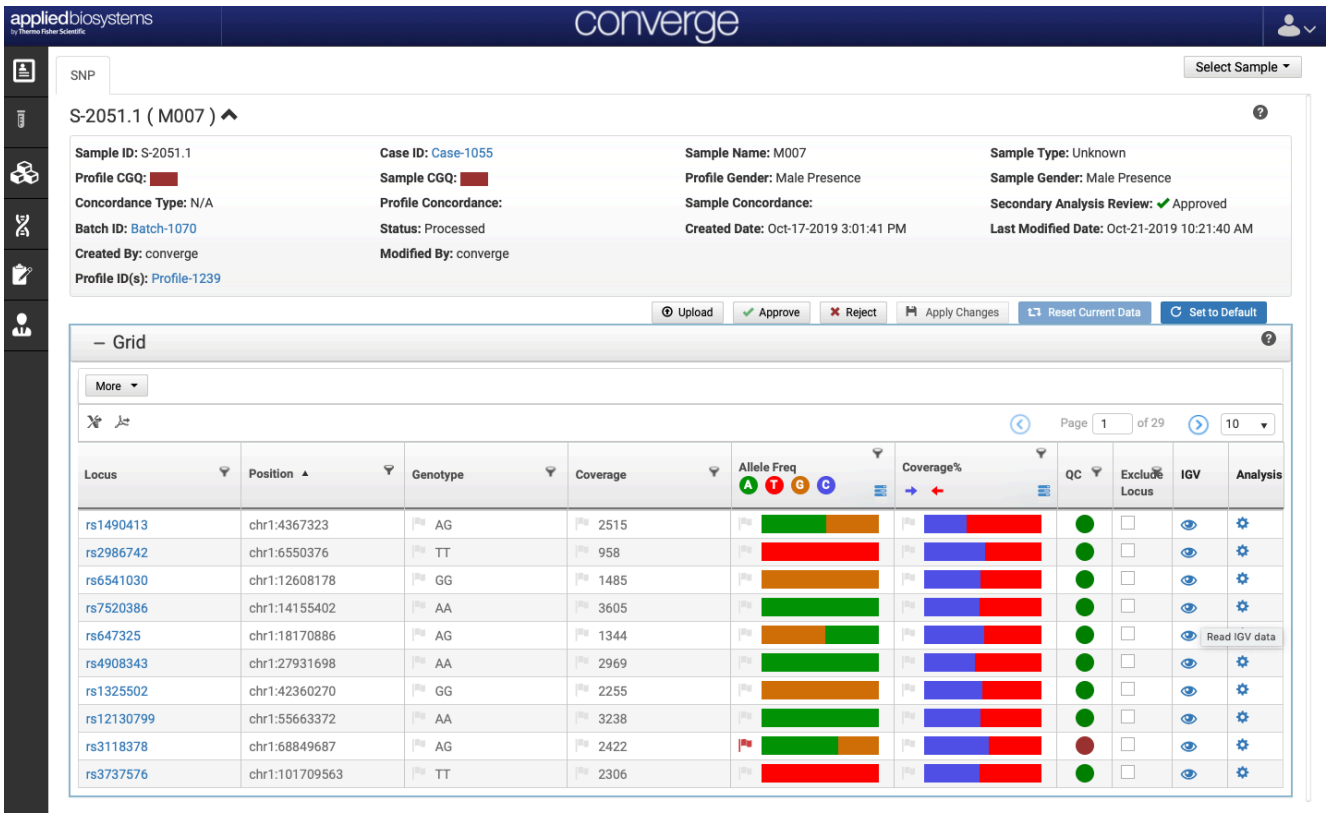


Figure 3. SNP genotyping results in Converge 2.2 for aiSNP, iiSNP and Y chromosomal SNPs in the STR + SNP Panel configuration. Sample genotypes will be displayed in the custom SNP panel display ('SNP' tab of the user interface).

Results

Assay performance

In general, both primer pool combinations performed well using the standard Ion Chef run protocols outlined above with throughputs of 12, 16 and 32 samples/chip. For simplicity, results presented in the following tables and figures will focus primarily on the Combination B workflow and 12 samples/chip format as that formulation performed slightly better than Combination A for reasons that will be discussed further in the text. Table 5 highlights sequencing run metrics obtained for Combination B approach and the STR and SNP panels run individually.

Table 5. Sequencing run metrics for NGS STR Panel v2 and Combination B using Ion 530 Chip

Panel Workflow	# of Samples Multiplexed	ISP Loading (%)	% Useable Reads	Total # of Useable Reads	Median bp	Aligned Reads (%)
Combination Panel B	12	69%	32%	8,002,033	99	97%
GlobalFiler NGS STR Panel v2	32	44%	32%	5,227,053	129	97%
Identity Panel	16	95%	61%	21,654,109	111	100%
Ancestry Panel	16	94%	56%	19,579,008	102	99%

*Note that only Chef workflow was used in this study; manual library construction was not performed.

As shown above, the primary sequencing run metrics for Combination B are similar to those obtained for the NGS STR Panel v2 run alone in terms of ISP loading, useable reads and % aligned reads. This result is expected as the STR markers exhibit lower ISP loading densities and total useable reads than SNP markers due to more stringent chip loading and sequencing run conditions. For all panels tested, an high percentage of reads (97-100%) aligned properly to the hg19 genome, which indicates successful mapping to the reference sequence. All runs performed in this study (N = 6) were sequenced successfully without any workflow issues.

Coverage

As the number of samples processed per chip will impact read coverage at the marker level, we analyzed the performance of the combination panel workflow to assess average coverage, markers with low read coverages (i.e., <50X and <100X) and potential for marker dropout. We compared sequence data output for each of the panels (STR v2, Identity and Ancestry) run individually versus the combination workflow to note any drops in performance with the expanded 324 marker multiplex. Finally, for the STR markers, we also measured intralocus balance, marker stutter percentages and sequence artifacts.

Table 6 highlights the expected trend of decreasing average marker coverage with increasing sample throughputs (12, 16 and 32 samples per chip). The results show a greater than 4-fold difference in average coverage within a given marker set (e.g., STR marker coverage for 12 samples/chip vs. 32 samples/chip). For the combination panel workflow, STR markers demonstrate the highest average coverage, followed by autosomal SNP's and Y-SNP's. The lowest average marker coverage (162 ± 72) was obtained for the Y-SNP markers at 32 sample/chip throughputs.

Table 6. Coverage obtained with Combination B for STR, autosomal SNP and Y-SNP markers

Samples per Chip	STR Coverage	Autosomal SNP Coverage	Y-SNP Coverage
12	2631 ± 1241	1727 ± 781	738 ± 344
16	1561 ± 743	1013 ± 506	407 ± 180
32*	650 ± 300	413 ± 204	162 ± 72

*Note that 16 libraries each were taken from Combination A and Combination B preparations and mixed prior to templating to make up the 32 samples/run.

As users may choose to target <1ng gDNA input in this assay, we analyzed low performing markers with coverages <50X and <100X to assess the potential for marker dropout with lower DNA inputs. When sequencing 12 samples per chip at 1ng gDNA input, only three iiSNP markers (rs2269355, rs1523537 and rs2342747) provided coverages <100X. Six markers (four iiSNP's and two aiSNP's) dropped below 100X with 16 samples/chip; 32 samples/chip throughput resulted in a total of 26 markers <100X (Table 7). No STR or Y-SNP markers fell below 100X coverage for 12 or 16 sample/chip throughputs in our testing for Combination B.

Table 7. Combination B markers detected with low coverage (<50X and <100X)

Samples per Chip	# of Markers <50X	% of Markers ≥50X	# of Markers <100X	% of Markers ≥100X
12	2	99.31%	3	98.96%
16	4	98.62%	6	97.92%
32	6	97.92%	26	91.00%

Figure 3 shows the STR marker coverage (2631 ± 1241) in Combination B obtained with a throughput of 12 samples/chip. Table 8 displays the low performing markers observed in the study (Penta D and E, D2S1338, FGA and D1S1677) had <60% coverage relative to the mean. This marker performance was comparable to running the Precision ID GlobalFiler NGS STR Panel v2 individually with standard conditions. Slightly lower Y-marker (AMELY, DYS391, Yindel, and SRY) coverage was observed with the combination panel approach when compared to the NGS STR Panel v2 alone; however, all Y-Markers exhibited coverage >800X.

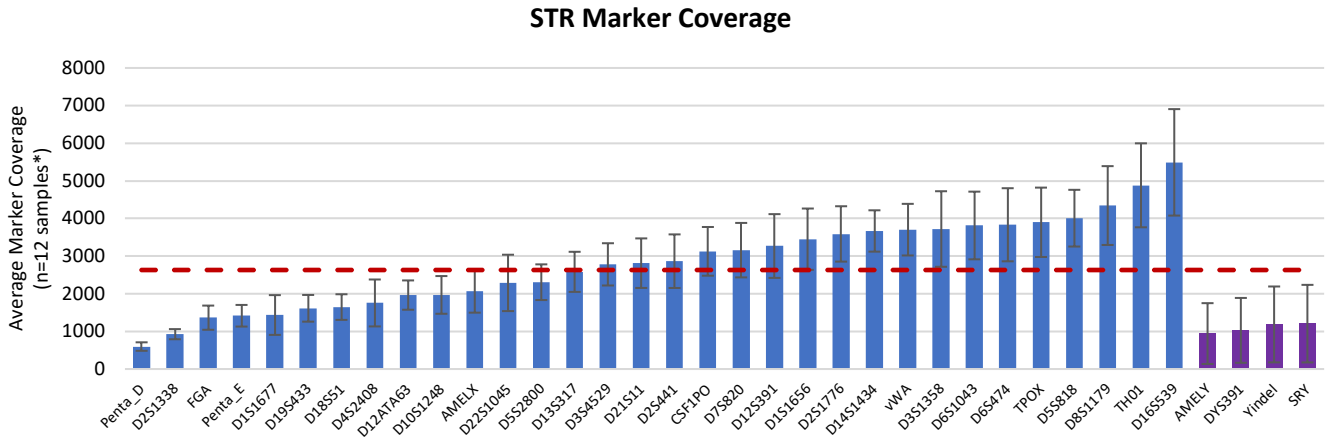


Figure 3. STR marker coverage in Combination B with a 12 sample/chip throughput shows a 9.2-fold coverage difference from lowest to highest performing markers. The red-dotted line represents the average coverage across the markers in the Precision ID GlobalFiler NGS STR v2 Panel when run individually. Sex-determining markers (AMELY, DYS391, Y indel and SRY) are roughly 2.6-fold less than the average autosomal STR marker coverage. *Of the 12 samples tested, 8 samples were males.

Table 8. Low performing autosomal STR markers (Combination B vs. GlobalFiler NGS STR Panel v2)

Combination B		Precision ID GlobalFiler NGS STR Panel v2	
Marker	**Average Coverage**	**Marker**	**Average Coverage**
Penta D	597 ± 113	Penta D	710 ± 116
D2S1338	925 ± 135	D2S1338	1519 ± 258
FGA	1364 ± 321	FGA	1649 ± 242
Penta E	1416 ± 288	Penta E	1950 ± 335
D1S1677	1435 ± 528	D1S1677	1830 ± 210

Table 9, Figures 4 and 5 show the aiSNP, iiSNP and Y-SNP marker coverages obtained for Combination B testing. Low performing aiSNPs were <20% relative to the mean coverage.

Table 9. SNP marker coverage results in Combination B panel workflow

Marker Set	Average Coverage	Low Performing SNP's	High Performing SNP's
aiSNP	1852 ± 699	rs37369 rs1296819 rs2504853 rs2306040	rs7657799 rs12544346 rs4951629
iiSNP	1498 ± 871	rs2269355 rs1523537 rs2342747	rs722098 rs1821380 rs10092491 rs354439
Y-SNP	748 ± 344	rs2032599 rs13447443 rs3911 P202	rs9786184 rs2032652 rs17222573

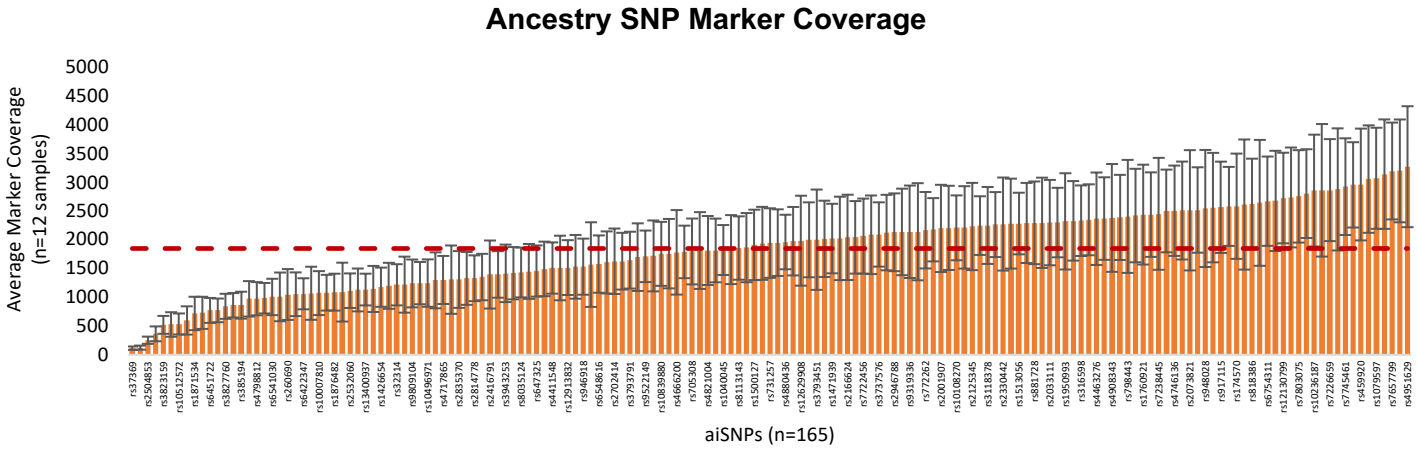


Figure 4. Average marker coverage of aiSNPs in Combination B with a throughput of 12 samples/chip. The red-dotted line represents the average marker coverage.

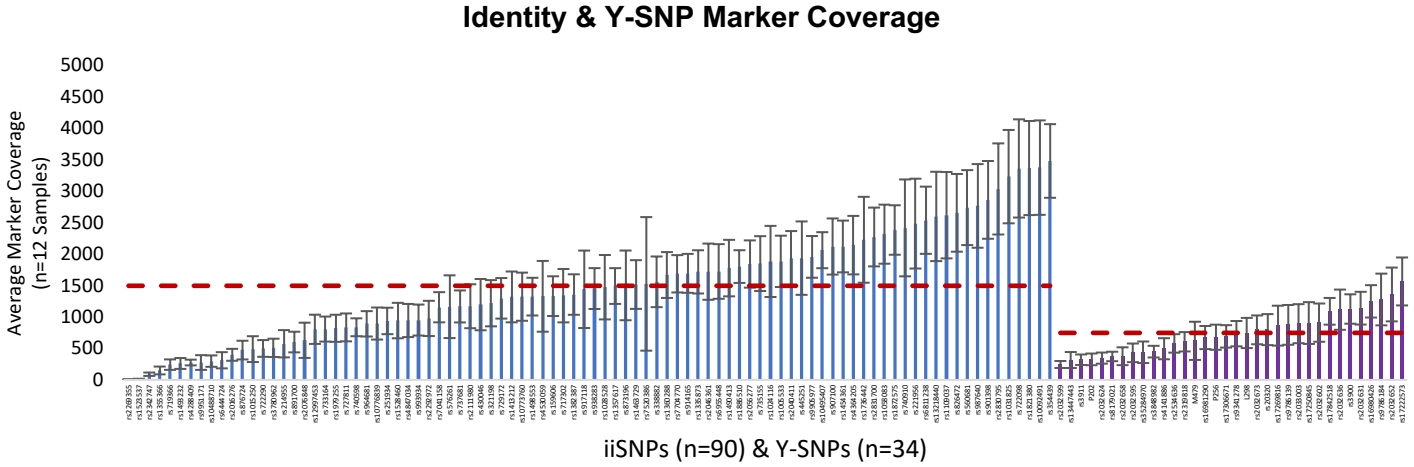


Figure 5. Identity SNP (blue) and Y-SNP (purple) marker coverage in Combination B with a throughput of 12 samples/chip. Red dotted lines represent the average marker coverage of iiSNP's and Y-SNP's.

Figure 6 shows intralocus balance (ILB) for the autosomal STR markers. 28 of 31 (90%) markers had a <10% ILB difference for Combination B versus the STR panel run alone with just under half of the markers (48.4%) showing a 5% or less difference in ILB. For Combination B, ILB ranged from 60.4% - 86.0%, whereas STR Panel v2 alone had a slightly larger ILB range of 64.3% - 95.3%. The largest % differences with the STR panel alone were observed for D2S441 (10.5%), D3S1358 (11.2%) and D2S1776 (11.5%). Five markers (i.e., D12S391, D22S1045, D16S539, D1S1656 and D4S2408) actually had improved ILB in Combination B (ranging from 1.1% - 6.9% ILB increases).

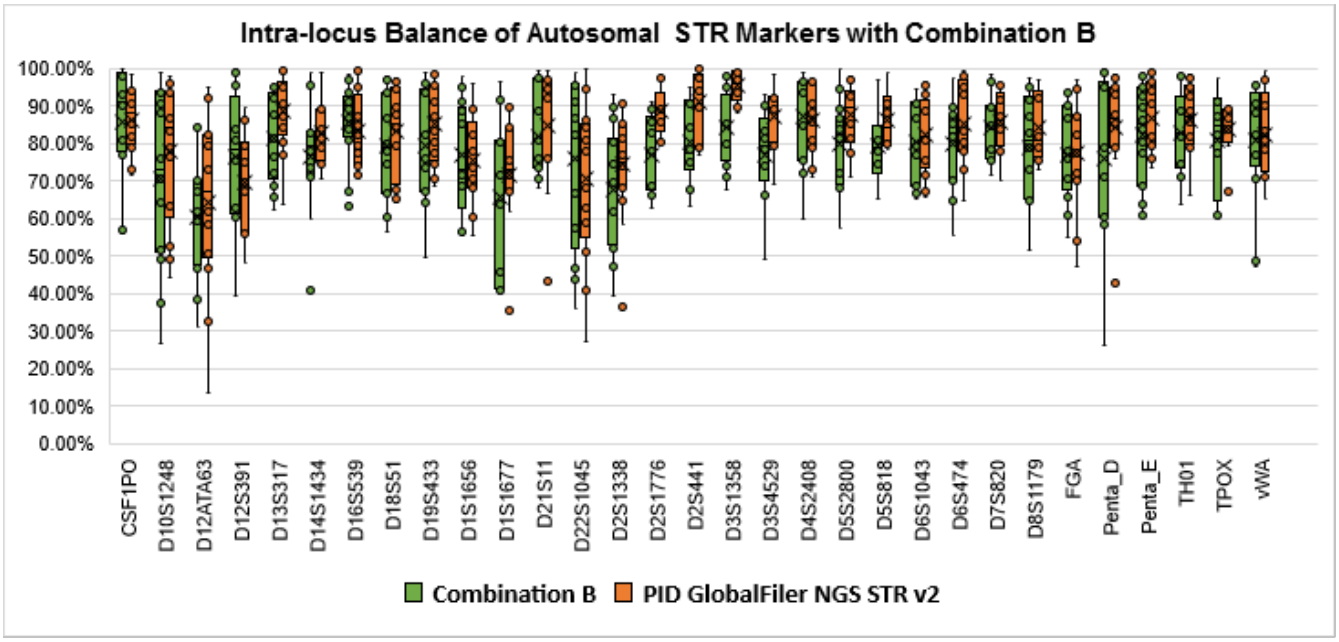


Figure 6. ILB was calculated for heterozygote pairs using the 1 ng DNA input test samples in the study (ratio of minimum reads over maximum reads obtained in a heterozygote pair). Boxes show the middle 50% or interquartile range (IQR). "Whiskers" indicate 1.5 IQR from the upper and lower margins of the IQR.

Concordance

Genotyping concordance was determined for all the genetic marker sets in Combination B. NGS data was either compared to CE truth data or previous NGS in-house data sets. STR markers were 100% concordant for the 12 sample/chip throughput. Autosomal SNP genotyping concordance was 99.13%. Discordances were either true genotyping miscalls (rs7520386, rs214955, and rs310644) or N calls which result from instances where the SNP did not meet quality thresholds for genotyping (e.g., low coverage or poor quality reads). Discordant calls occurred in SNP's exhibiting allelic imbalance in individual panel performance, which is further exacerbated in the combination approach due to the increased number of amplicons and slightly lower coverage for some markers. These imbalanced SNP's result in false homozygous genotypes (where the reference SNP genotype is heterozygous). Users should exercise caution when interpreting these SNPs; visual confirmation of the read pile-up in these regions is recommended.

Combination B also provided consistent admixture predictions for biogeographic ancestry results compared to the single panel workflow approach. Figures 7a-7d show representative examples of ancestry predictions for some of the major root populations including European (IB-0523), African (IB-0663), East Asian (HC-1017) and Southwest Asia (IB-0908).

Figure 7a. IB-0663 Ancestry Prediction

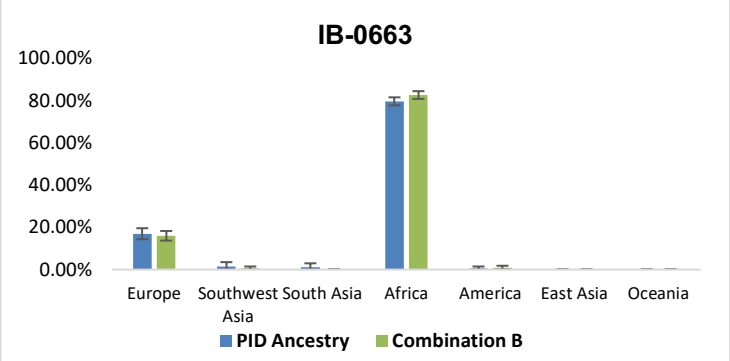


Figure 7b. IB-0908 Ancestry Prediction

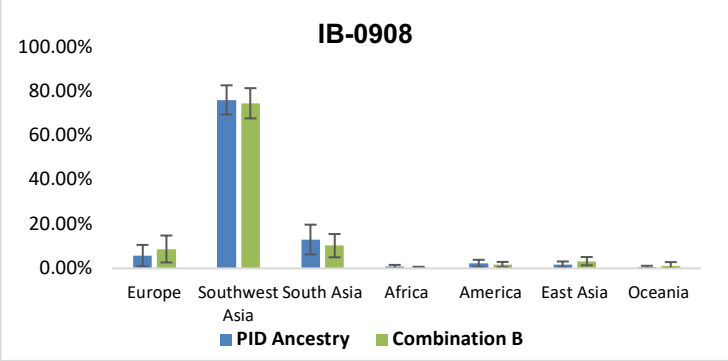


Figure 7c. HC-1017 Ancestry Prediction

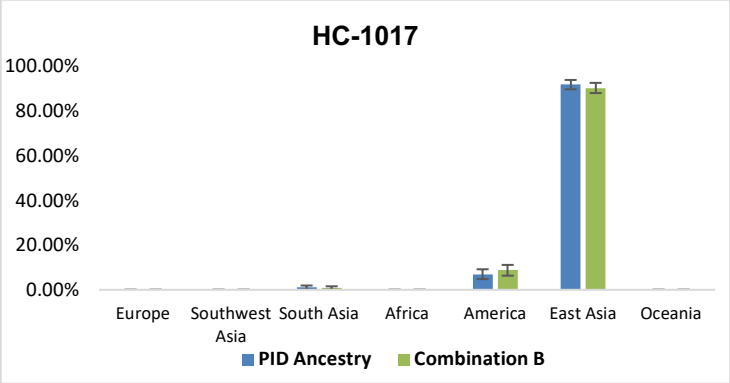
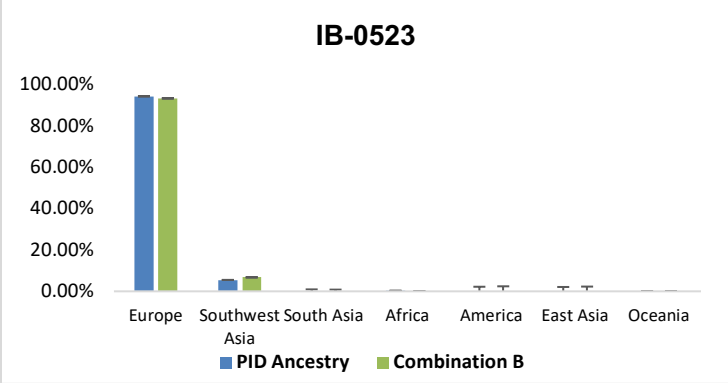


Figure 7d. IB-0523 Ancestry Prediction



Figures 7a-7d. Four representative samples analyzed for biogeographic ancestry (N = 16) utilizing the Converge v2.2 bootstrapping algorithm. Graphs show the similarities in admixture predictions for the aiSNP's in the Ancestry panel alone versus Combination B. The v2.2 analysis settings included 40 bootstrap replications (with replacement) and 50% SNP resampling size. Only slight differences were observed in the admixture results among these test samples.

Similarly, Y haplogrouping of the 34 Y-SNP's demonstrated 100% concordance for of the eight samples analyzed in this study (R1b haplogroup was obtained for all samples except for O2 haplogroup for sample HC-1017). Y-haplogroup predictions were 100% concordant between the combination panel approach and the Identity SNP panel for 12 samples and 16 samples multiplexed per 530 chip.

STR Artifacts & Stutter

STR artifacts and stutter products were also compared between Combination B and individual panel workflows. Artifacts similar to those previously characterized in the recent STR Application Note[8] were present in D12S391, D10S1248, D18S51 and D3S4529 (.1 or .2 bp sequence artifacts) either in the STR Panel v2 or Combination B (Table 10). However, when sample throughput was increased from 12 to 16 samples/chip for Combination B, three additional instances of sequence artifacts (5 total artifacts) were detected, including artifacts at D12S391, D10S1248, D18S51, D3S4529 and Penta E. An increase in the number of elevated stutter peaks was observed in Combination B results; however, the stutters were not classified as artifacts in this analysis. Users may consider adjusting thresholding parameters after performing internal empirical studies to minimize the impact of sequence artifacts.

Table 10. STR artifacts detected above the stochastic threshold of 5%.

Configuration	Sample	Marker	Artifact Allele	% of Truth Peak	Short Sequence
Single	IB-0533	D12S391	25.1	53.3	[AGAT]15 [AGAC]9 AGAGT
Single	M007	D10S1248	12.2	17.1	GA [GGAA]12
Combination B	Cor419	D18S51	16.2	20.0	[AGAA]11 GA[AGAA]1 GA[AGAA]1 GA[AGAA]2
Combination B	IB-0635	D3S4529	14.3	12.4	[ATCT]10 ATT[ATCT]4

Conclusions

This study demonstrates the feasibility of combining the commercially available Precision ID STR and SNP panels in a two-pool library preparation strategy on the Ion Chef System to generate additional genetic content for biogeographic ancestry and identity testing. For this STR + SNP combination workflow, multiplexing 12 samples on an Ion 530 Chip provided sufficient chip real estate to generate adequate marker coverages for the suite of 324 genetic markers without risk of marker dropout or additional sequence artifacts using standard run conditions. Additional considerations should be given when adjusting genomic DNA input amounts (<1ng) or sample throughputs (>12 samples/chip) which may lead to STR and/or SNP marker dropouts and possibly an increase in sequence artifacts for a subset of the STR markers. Users are advised to perform empirical studies to support any modifications to the protocol described above and to assess the inherent variability of the workflow which may further impact these guidelines.

Combination B's simultaneous analysis of 31 autosomal STR's (plus Y indel, SRY, Amelogenin and DYS391) with 165 aiSNP's, 90 iiSNP's and 34 Y-SNP's enables significantly more genetic intelligence – for both identity and investigative applications – in a more efficient workflow than analyzing individual panels in a modular format alone. Comparable sequence run results and the high percent concordance obtained in this study validate the feasibility of combined marker sets in the STR + SNP workflow described. Comparative data generated for the combination panel workflow and the Precision ID panels run alone were similar enough to justify this approach, and the increases in stutter and artifacts for the STR markers in Combination B did not hinder the performance of these markers. Similarly, the minor discordances detected with a small number of SNP markers did not impact Y-haplogroup or ancestry tertiary analysis results overall.

References

[1] Precision ID NGS System for Human Identification.
[2] Ion AmpliSeq HID community panels are available at ampliseq.com.
[3] Hollard, C. & Keyser, Christine & Delabarde, Tania & Gonzalez, A. & Lamego, C. & Zvenigorosky, Vincent & Ludes, Bertrand. (2016). Case report: on the use of the HID-Ion AmpliSeq™ Ancestry Panel in a real forensic case. International Journal of Legal Medicine. 131. 10.1007/s00414-016-1425-1.
[4] Walsh S, Liu F, Wollstein A, Kovatsi L, Ralf A, Kosiniak-Kamysz A, Branicki W, Kayser M. The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. Forensic Sci Int Genet. 2013 Jan;7(1):98-115. https://doi: 10.1016/j.fsigen.2012.07.005.
[5] Ralf, A., Oven, M., Gonzalez, O., de Kniff, P., der Beek, K., Wootton, S., Lagace, R., Kayser, M. Forensic Y-SNP analysis beyond SNaPshot: High-resolution Y-chromosomal haplogrouping from low quality and quantity DNA using Ion AmpliSeq and targeted massively parallel sequencing. Forensic Sci Int Genet. Volume 41, July 2019, Pages 93-106. https://doi.org/10.1016/j.fsigen.2019.04.001
[6] Precision ID SNP Panels with the HID Ion S5/HID Ion GeneStudio S5 System Application Guide (MAN0017767)
[7] Analysis files can be found at Thermo Fisher Scientific's Converge Software Downloads page.
[8] STR Application Note: *Get more information from challenging samples with next-generation sequencing of short tandem repeats* (COL32762 1118) can be found at thermofisher.com/hid-ngs.