Massively parallel sequencing of the 124 Precision ID Identity SNPs in a Myanmar population

Su Min Joo[1] · Ye-Lim Kwon[1] · Mi Hyeon Moon[1] · Kyoung-Jin Shin[1, *]

[1]Department of Forensic Medicine, Graduate School of Medical Science, Brain Korea 21 Project, Yonsei University College of Medicine, Seoul 03722, Korea

Introduction

Single nucleotide polymorphisms (SNPs) have been steadily gaining attention in the field of forensic genetics due to low mutation rates and being analyzed with small size amplicons for degraded DNAs. The Precision ID Identity Panel (Thermo Fisher Scientific), consisting of 90 autosomal SNPs and 34 Y-SNPs, was introduced for human identification using massively parallel sequencing (MPS). In this study, we have established new procedure that the Precision ID Identity Panel could be analyzed on a MiSeq (Illumina) by preparing MPS library using in-house TruSeq compatible universal adapters. It makes the Precision ID library preparation process simpler and more flexible than original method. We have also developed custom MPS data analysis pipeline, which will allow forensic geneticists to easily analyze SNP-MPS results. Here, we investigated the allele frequencies of 90 autosomal SNPs and distribution of Y-haplogroups among 96 Myanmar males. In this presentation, we show structure of the in-house TruSeq compatible universal adapters, sequencing performance of the Precision ID Identity Panel, the distribution of Y-haplogroups in Myanmar population, and cryptic SNPs to increase discrimination power.

Materials and Methods

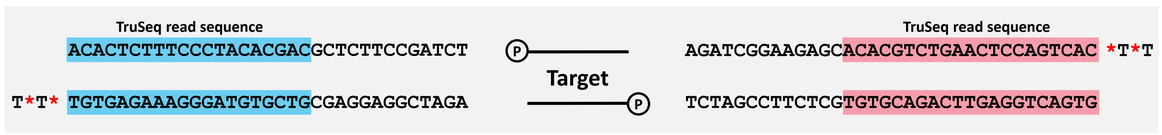
Samples

- A total of 96 unrelated Myanmar male DNA samples from a biobank of the Asian Sample Network with Seoul National University were extracted from FTA card. All extracted DNA was quantified using Qubit 4.0 Fluorometer (Thermo Fisher Scientific), and the concentration of DNA was around 200 pg/µL. 2 µL of DNA was used for MPS library preparation.
- 2800M Control DNA (Promega) was used as a reference in this study.

MPS Library Preparation and Run

- In order to perform the MPS of the Precision ID Identity Panel on a MiSeq system, in-house TruSeq compatible universal adapters (Fig. 1) were designed and diluted to 1.0 µM before ligation.

Fig. 1. Structure of the in-house TruSeq compatible universal adapters



*: Phosphorothioate bond

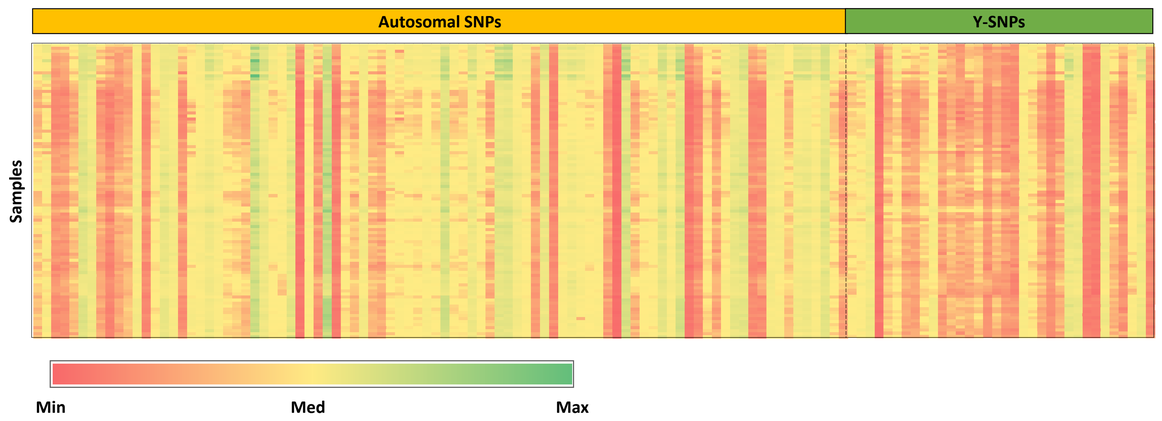
- The MPS libraries were prepared based on the Precision ID Library Kit (Thermo Fisher Scientific) with some modified conditions.
- The MPS libraries were purified with 1.4 × Agencourt® AMPure® XP beads (Beckman Coulter) and quantified using KAPA library quantification kits (KAPA Biosystems) and 2100 Bioanalyzer (Agilent).
- The purified libraries were normalized to 10 nM and then pooled in equal volumes. The pooled library was sequenced on a MiSeq system.

MPS Data Analysis

- FASTQ files were analyzed using two independent methods and we confirmed the genotype concordance between two methods.
  1. STRait Razor v.3 and Visual SNP for MPS (http://forensic.yonsei.ac.kr/VisualSNP/index.html)
     - STRait Razor v.3 identifies and characterizes the sequence-based polymorphism.
     - Visual SNP for MPS processes and visualizes results of STRait Razor v.3 to analyze SNP-MPS data efficiently.
  2. GATK HaplotypeCaller and Integrative Genome Viewer (IGV) 2.8.13.

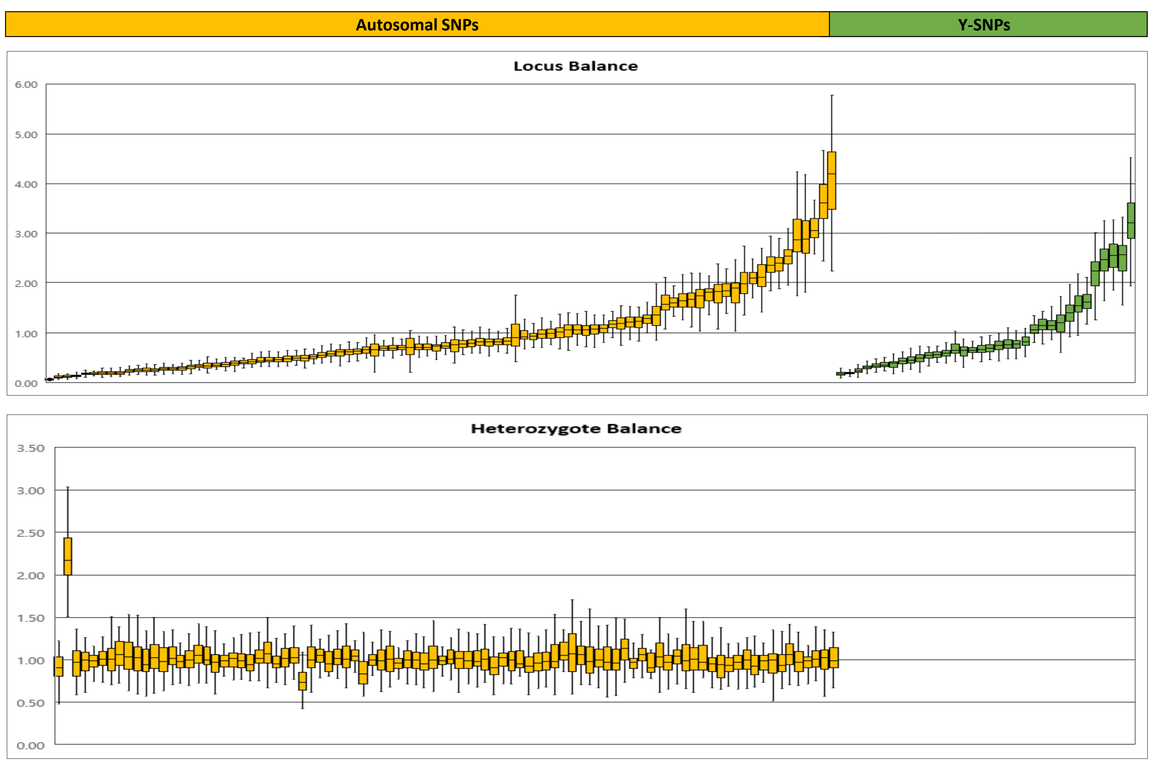
Results

Fig. 2. Read depth heatmap for 124 SNPs in the Precision ID Identity Panel of the Myanmar population



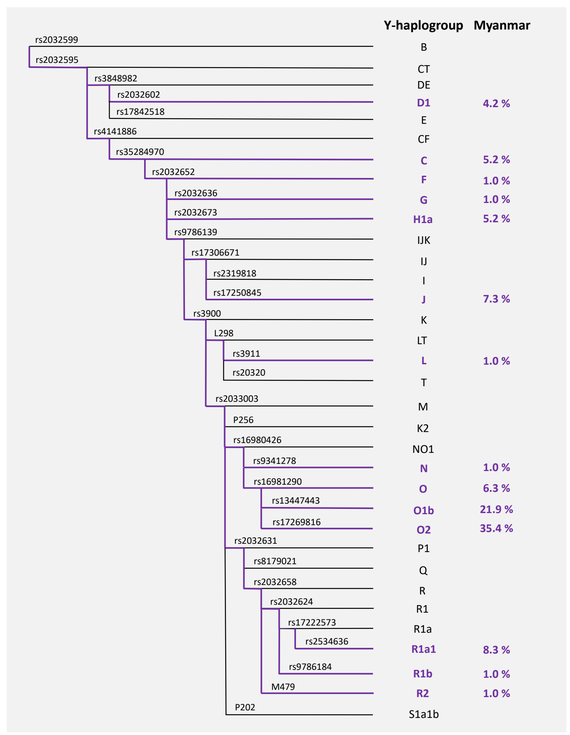
- The mean read depth of autosomal SNPs and Y-SNPs was 2830 and 1643, respectively.
- The top five SNPs with lowest read depth were rs2342747, rs9786139, rs917118, rs2032631, and rs9951171.

Fig. 3. Sequencing performance of the Precision ID Identity Panel



- Sequencing performance of the Precision ID Identity Panel was evaluated by locus balance and heterozygote balance.
- Median locus balance of autosomal SNPs ranged from 0.07 (rs2342747) to 4.19 (rs10092491), and that of Y-SNPs ranged from 0.19 (rs9786139) to 3.22 (rs3848982) with an average value of 1.00 for both.
- While almost all loci had heterozygote balance within the manufacturer's recommended threshold (0.54 ~ 1.86), there was only one outlier, rs7520386 (the highest heterozygote balance at 2.17).

Fig. 4. Y-haplogroup tree determined in 96 Myanmar male samples



A total of 14 haplogroups were shown in **bold purple**.

- 34 upper Y-clade SNPs were used for Y-haplogroup determination following ISOGG Y-DNA Haplogroup Tree 2019-2020 v.15.73.
- A total of 14 haplogroups were observed in 96 Myanmar males.
- The haplogroup O2 (35.4%) and O1b (21.9%) accounted for the majority.

Table 1. Comparison of combined match probability between A-SNPs and A-STRs

	combined match probability
90 autosomal SNPs	6.994×10^{-34}
22 autosomal STRs	3.130×10^{-26}

- For comparison of discrimination power, 22 autosomal STRs are genotyped using PowerPlex® Fusion System (24 loci, Promega).
- The combined match probability of 90 autosomal SNPs was 10^{-8} lower than that of 22 autosomal STRs.

Table 2. Increase in expected heterozygosity by cryptic SNPs

	rs907100-rs11689319	**rs10776839**-rs7037930	**rs1490413**-rs16838591	**rs12997453**-rs72883670
Target SNPs He	0.487	0.496	0.454	0.432
Haplotype He	0.644	0.642	0.52	0.479
% of increase	32.3	29.5	14.6	10.9

Target SNPs are shown in **bold**; He, expected heterozygosity

- Cryptic SNPs are in the flanking region of 124 target SNPs.
- rs11689319, rs7037930 increased expected heterozygosity by about 30% compared to when using the target SNPs only.

Conclusion

- We have established new procedure that the Precision ID Identity Panel could be analyzed on a MiSeq by using in-house TruSeq compatible universal adapters and analyzed MPS data efficiently using STRait Razor v.3 and Visual SNP for MPS, in-house web-based tool.
- The Precision ID Identity Panel was performed successfully on a MiSeq and showed high discrimination power in a Myanmar population. However, careful interpretation is recommended especially for some problematic SNPs which presented low coverage or severe heterozygote imbalance.

Acknowledgement

- This study was supported by a faculty research grant at Yonsei University College of Medicine (2020-2021).