Whole genome amplification and hybrid capture-based massively parallel sequencing of forensic SNPs for trace amounts of DNA

Su Min Joo[1,2] · Ye-Lim Kwon[1,2] · Kyoung-Jin Shin[1,2]

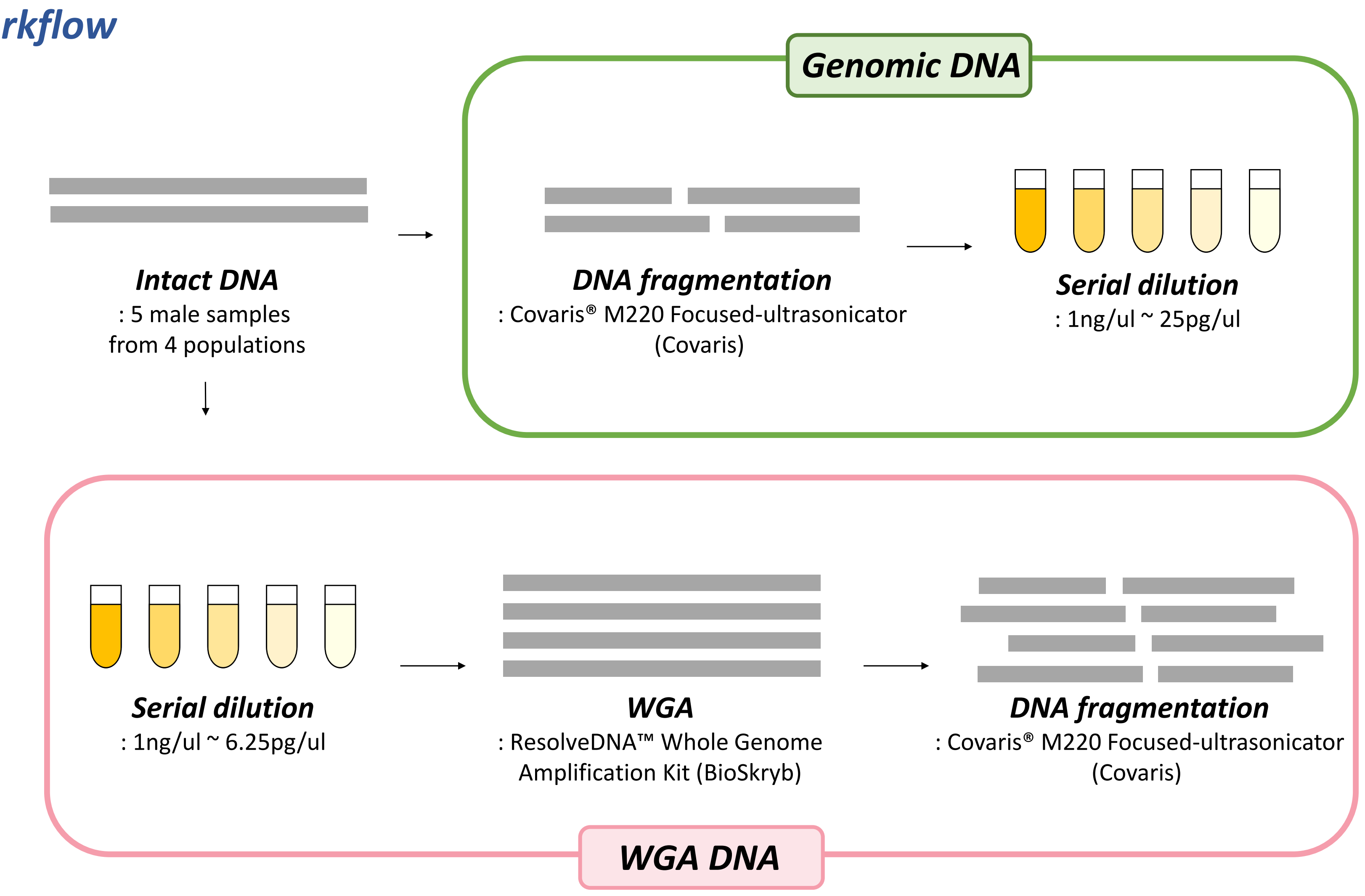
[1]*Department of Forensic Medicine, Yonsei University College of Medicine, Seoul, Korea*
[2]*Graduate School of Medical Science, Brain Korea 21 Project, Yonsei University, Seoul, Korea*

Introduction

DNA recovered from crime scenes is frequently in low quantities, limiting subsequent genetic analyses and leading to unreliable results. With the introduction of massively parallel sequencing (MPS) into forensic genetics, MPS of single nucleotide polymorphisms (SNPs) for trace amounts of DNA has been attempted using low-input DNA library preparation kits. However, this requires a certain amount of input DNA, and as the input DNA quantity decreases, so does library complexity. To overcome these limitations, whole genome amplification (WGA) can be an alternative. Nonetheless, the effects of WGA on the trace amounts of DNA in hybrid capture-based MPS have not been thoroughly investigated. In this study, we developed a hybrid capture-based MPS system for 136 identity-informative SNPs and applied it to five male samples serially diluted to sub-nanogram level. In this presentation, we show the WGA yield, On-SNP rate of hybrid capture-based MPS system, genotype recovery rates, and the characteristics of each marker.

Materials and Methods

Workflow



Genomic DNA

Intact DNA
: 5 male samples
from 4 populations

DNA fragmentation
: Covaris® M220 Focused-ultrasonicator
(Covaris)

Serial dilution
: 1ng/ul ~ 25pg/ul

WGA DNA

Serial dilution
: 1ng/ul ~ 6.25pg/ul

WGA
: ResolveDNA™ Whole Genome
Amplification Kit (BioSkryb)

DNA fragmentation
: Covaris® M220 Focused-ultrasonicator
(Covaris)

Library Preparation
:ThruPLEX® DNA-Seq Kit
(TaKaRa Bio Inc.)

Hybrid Capture
:KAPA HyperChoice probe
(Roche)

Library Quality Control
:KAPA Library Quantification Kits
(KAPA Biosystems)
Agilent DNA 7500 Kit (Agilent)

MPS
: MiniSeq system (Illumina)
(2 × 150 cycles)

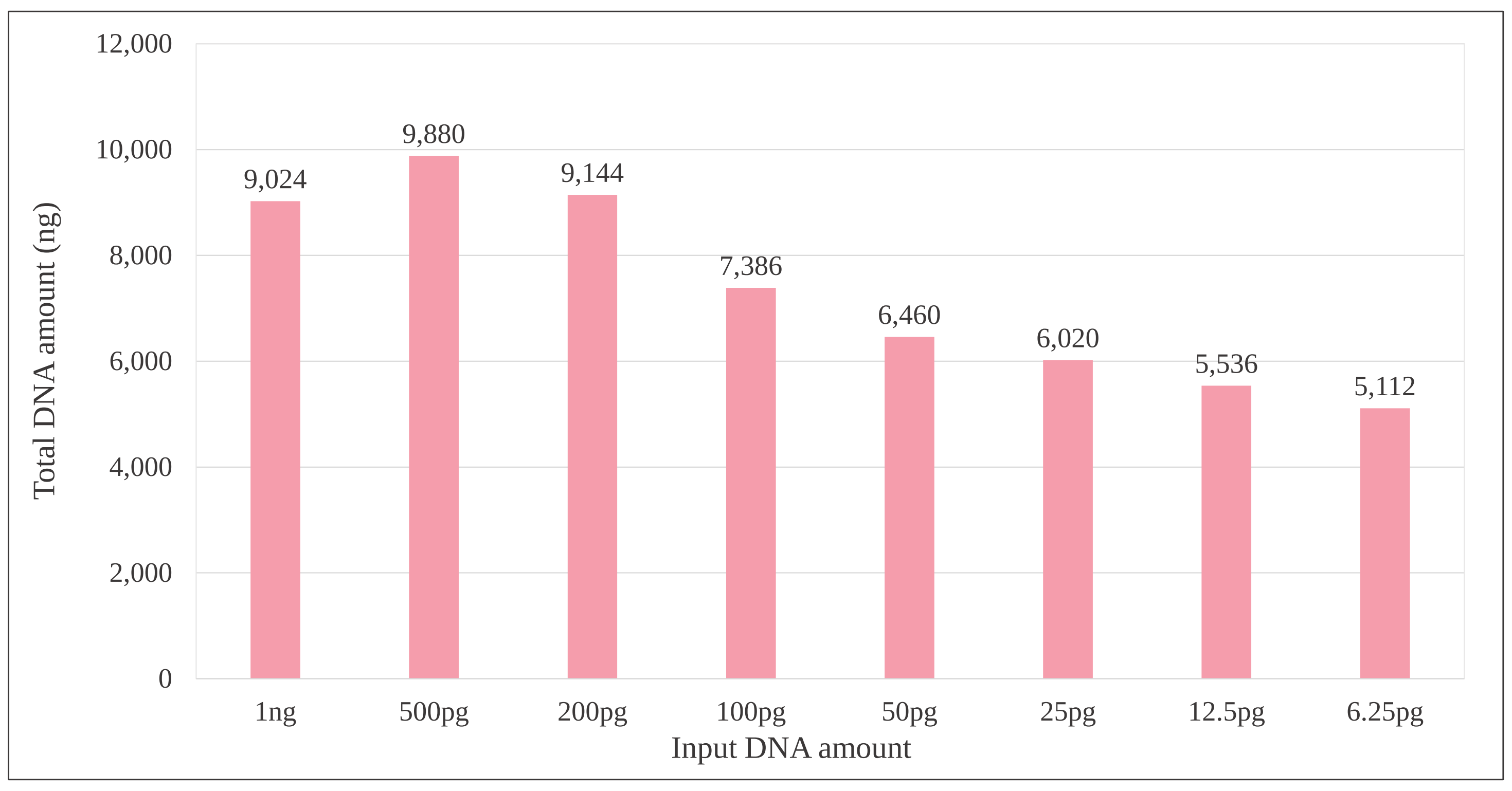
Data Analysis
: Genome Analysis Tool Kit (GATK)

Probe Design for Hybrid Capture

- 102 autosomal SNPs from *Kidd et al.(2006), Pakstis et al.(2010), Phillips et al.(2007), and Sanchez et al.(2006).*
- 34 Y-chromosomal SNPs from the Precision ID Identity Panel (Thermo Fisher Scientific).
- 120 bp length including the target SNPs using KAPA HyperDesign (Roche)

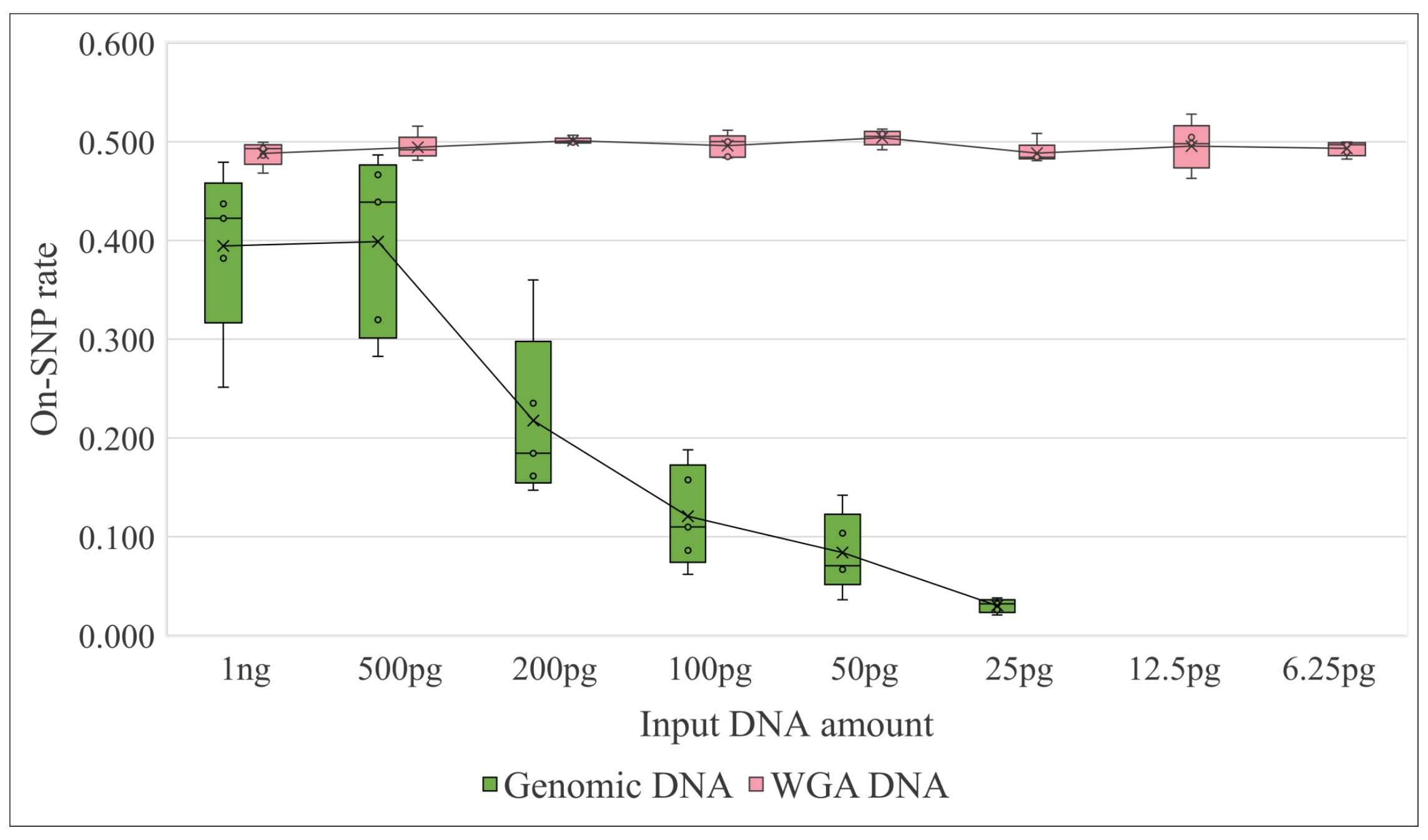
Results

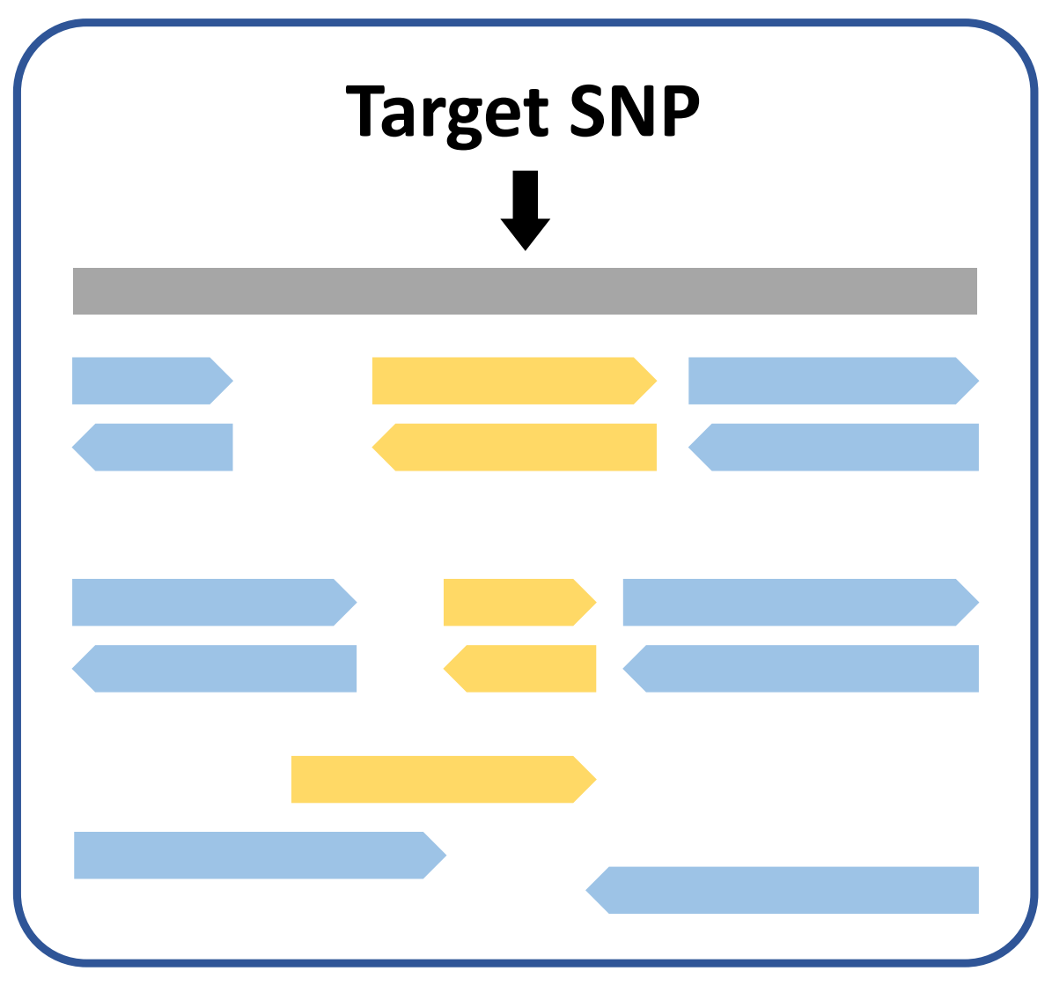
Fig. 1. The WGA yield according to the amount of input DNA



- WGA DNA was quantified on a Qubit 4.0 Fluorometer (Thermo Fisher Scientific).
- The WGA yield was positively correlated with the amount of input DNA and reach the microgram level for all sub-nanogram samples.

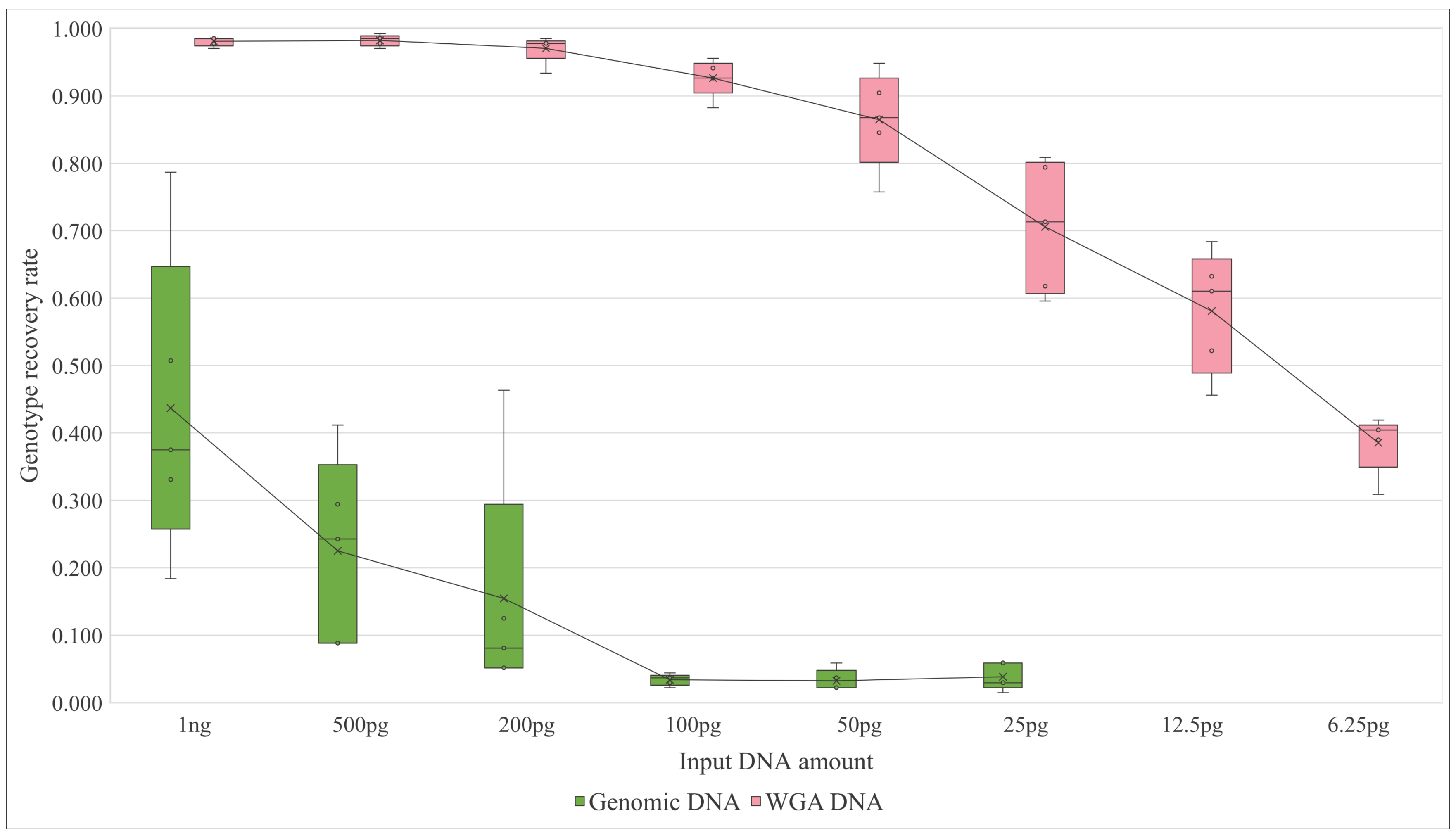
Fig. 2. On-SNP rate of hybrid capture-based MPS system





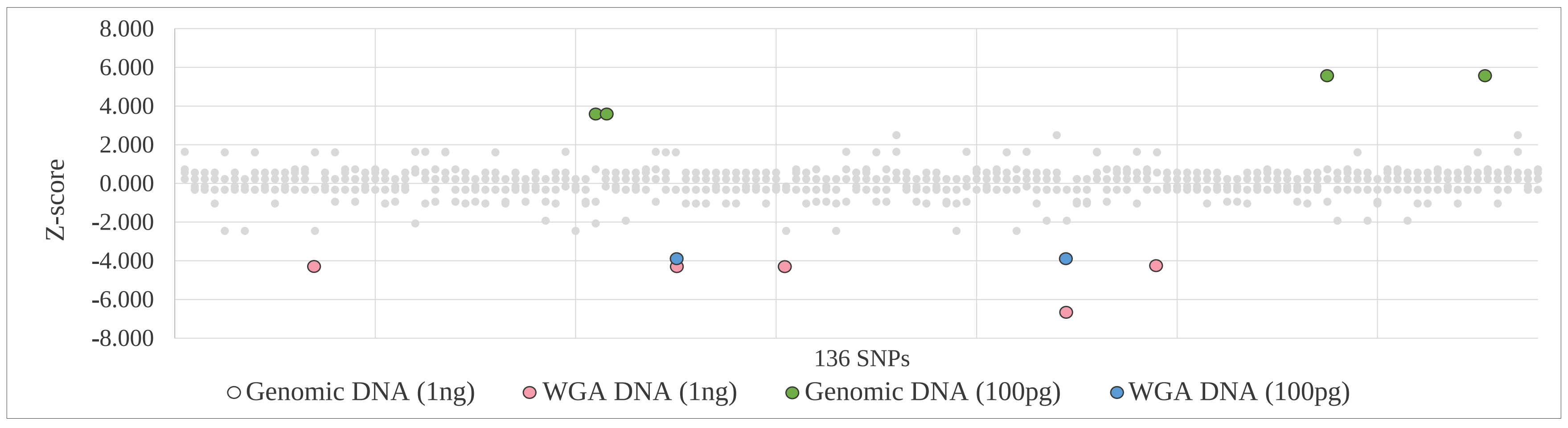
- Reads aligned exclusively to the target SNPs
- Target SNP reads / Aligned reads Aligned reads
- In the case of ***genomic DNA***, it decreased sharply as the amount of input DNA decreased.
- In the case of ***WGA DNA***, it was kept at 50% regardless of the amount of input DNA.

Fig. 3. The genotype recovery rate according to the amount of input DNA



- Genotype reference: 10 ng of genomic DNA
- In the case of ***genomic DNA***, the genotype recovery rate decreased rapidly as the amount of input DNA decreased.
- In the case of ***WGA DNA***, the genotype recovery rate was maintained over 90% even at 100 pg.
- For 50% genotype recovery rate, 1 ng for ***genomic DNA*** and 12.5 pg for ***WGA DNA*** were required.

Fig. 4. Z-scores based on the genotype recovery rate of 136 SNPs



- Based on the genotype recovery rate for each SNP

●● Low capture efficiency: rs1058083, rs2016276, rs2033003, rs4141886, and rs576261
● High capture efficiency: rs1769816, rs17306671, rs763869, and rs9786139

Conclusion

◆ We have developed a hybrid capture-based MPS system for 136 identity-informative SNPs and applied it to five male samples serially diluted to sub-nanogram level.
◆ WGA DNA showed higher genotype recovery rates in all sub-nanogram samples.
◆ We demonstrated the usefulness of WGA for trace amounts of DNA and provided the information for the amount of input DNA required to obtain reliable results.

Acknowledgement

- This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1A2C101306312).