

---

# Analysis of **1.2 K** identity-informative **SNPs** for **low template DNA** by performing **whole genome amplification** and **hybrid capture-based NGS**

Su Min Joo, Ye-Lim Kwon, and Kyoung-Jin Shin

Dept. of Forensic Medicine  
Yonsei University College of Medicine  
Seoul, Republic of Korea



# Backgrounds

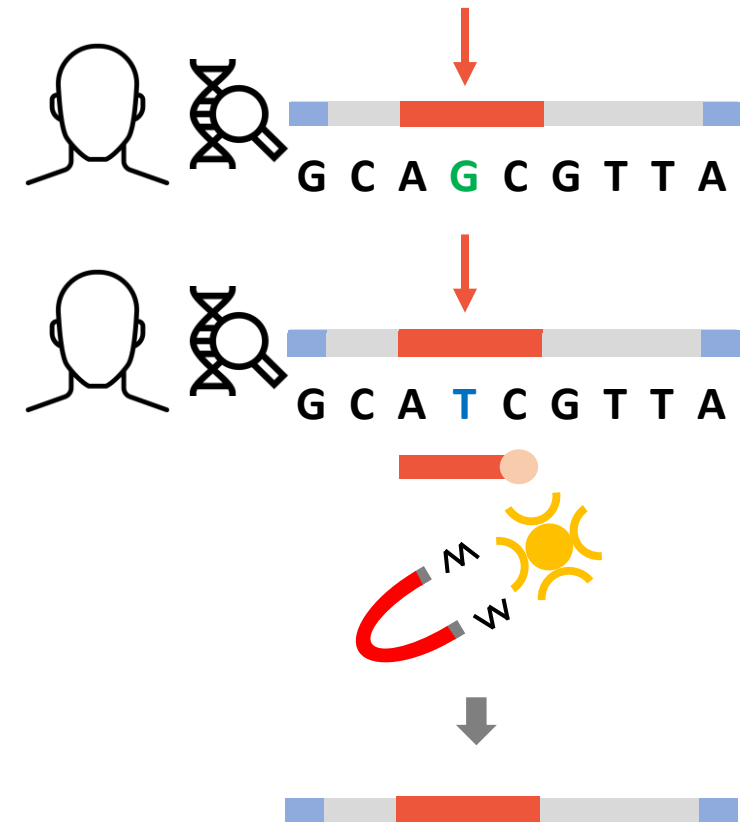
## Single nucleotide polymorphisms (SNPs)

- Small target region (only one base)
- Low mutation rate
- Additional genetic information (ancestry, phenotype, kinship ...)



## Next generation sequencing (NGS)

- Amplicon-based target enrichment
- Hybrid capture-based target enrichment





# Backgrounds (continued)

## Hybrid capture-based target enrichment




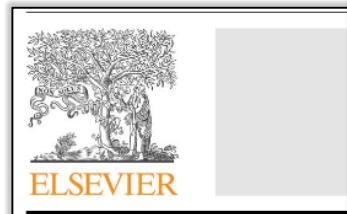
- Flexible expansion of the SNP panel
- Advantageous for degraded DNA

Article

**Applications of Probe Capture Enrichment Next Generation Sequencing for Whole Mitochondrial Genome and 426 Nuclear SNPs for Forensically Challenging Samples**

Shelly Y. Shih <sup>1</sup> , Nikhil Bose <sup>1,2</sup>, Anna Beatriz R. Gonçalves <sup>1,3</sup>, Henry Erlich <sup>1,2,\*</sup>, and Cassandra D. Calloway <sup>1,2,\*</sup>





Research paper

**Target capture enrichment of nuclear SNP markers for massively parallel sequencing of degraded and mixed samples**



Nikhil Bose <sup>a,b</sup>, Katie Carlberg <sup>a,c</sup>, George Sensabaugh <sup>b,d</sup>, Henry Erlich <sup>a</sup>, Cassandra Calloway <sup>a,b,\*</sup>



Article

**The FORCE Panel: An All-in-One SNP Marker Set for Confirming Investigative Genetic Genealogy Leads and for General Forensic Applications**

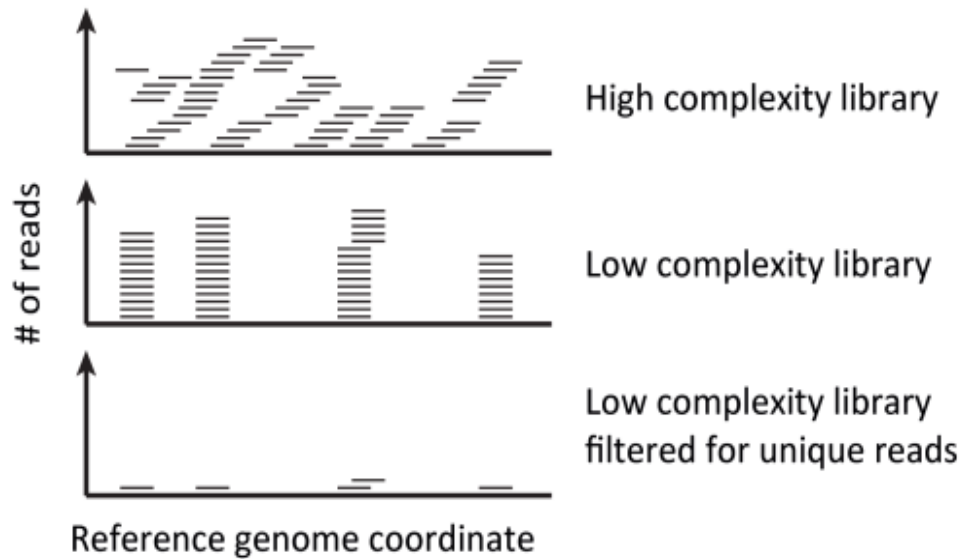
Andreas Tillmar <sup>1,2,\*</sup>, Kimberly Sturk-Andreaggi <sup>3,4,5</sup> , Jennifer Daniels-Higginbotham <sup>3,4</sup>, Jacqueline Tyler Thomas <sup>3,4</sup> and Charla Marshall <sup>3,4,6,\*</sup> 

# Backgrounds (continued)

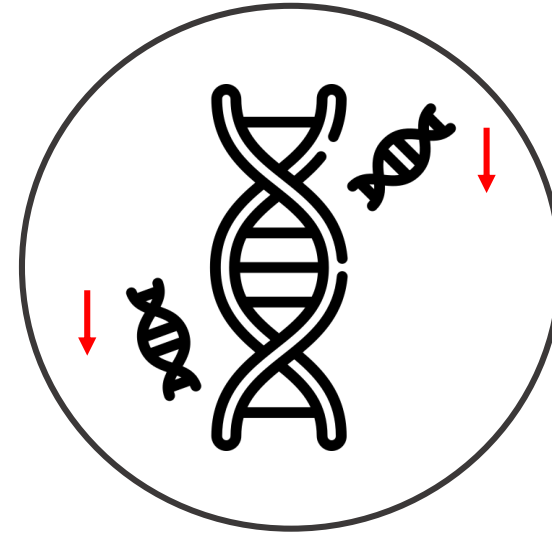
## Low template DNA (LT DNA)

### Low-input DNA Library Preparation Kits

- Low library complexity
- Require at least 100pg



Fossli (2020) WIREs Syst Biol Med



**Whole genome amplification (WGA)**

# Research objectives

---

## ■ General objective

Construct **hybrid capture-based NGS** system for **1.2K SNPs** using **whole genome amplification** and apply it to **low template DNA** to investigate its **sensitivity** and **accuracy** according to the amount of input DNA.

## ■ Specific objectives

1. Development of **1.2K identity-informative SNP** panel
2. Construction of **hybrid capture-based NGS** system using **WGA** for **low template DNA**
3. Assessment of the efficacy of **whole genome amplification** and **duplicate analysis**

# Materials and Methods

## Development of 1.2K SNP panel for hybrid

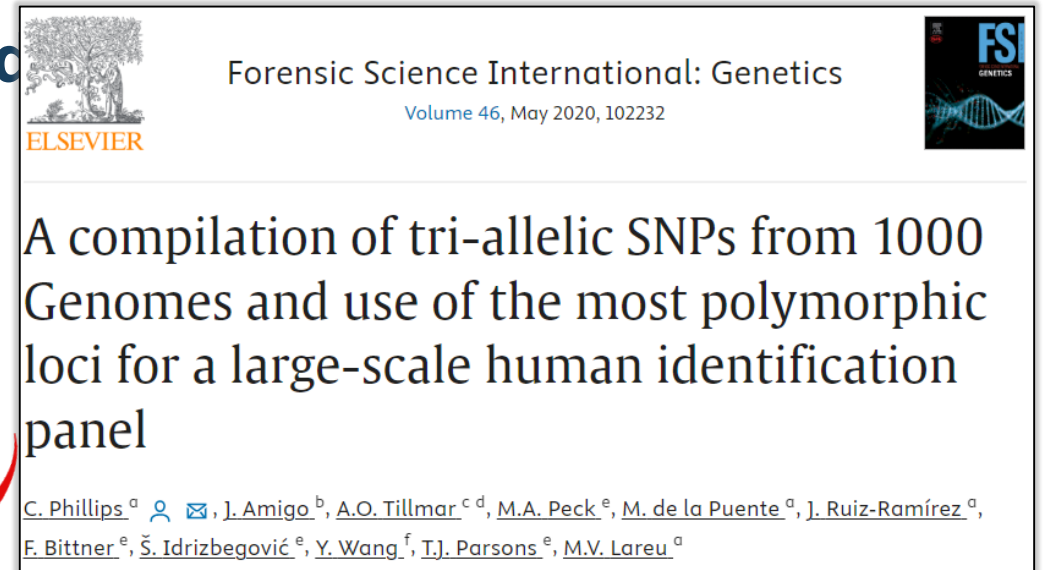
### - 1,225 identity-informative SNPs

- 1,164 autosomal tri-allelic SNPs
- 23 X-chromosomal tri-allelic SNPs

- 38 Y-chromosomal SNPs (including the Precision ID Identity Panel)

### - Hybrid Capture

- KAPA HyperChoice probe (Roche)
- 120bp including target SNPs
- No tiling approach (one probe per one SNP)



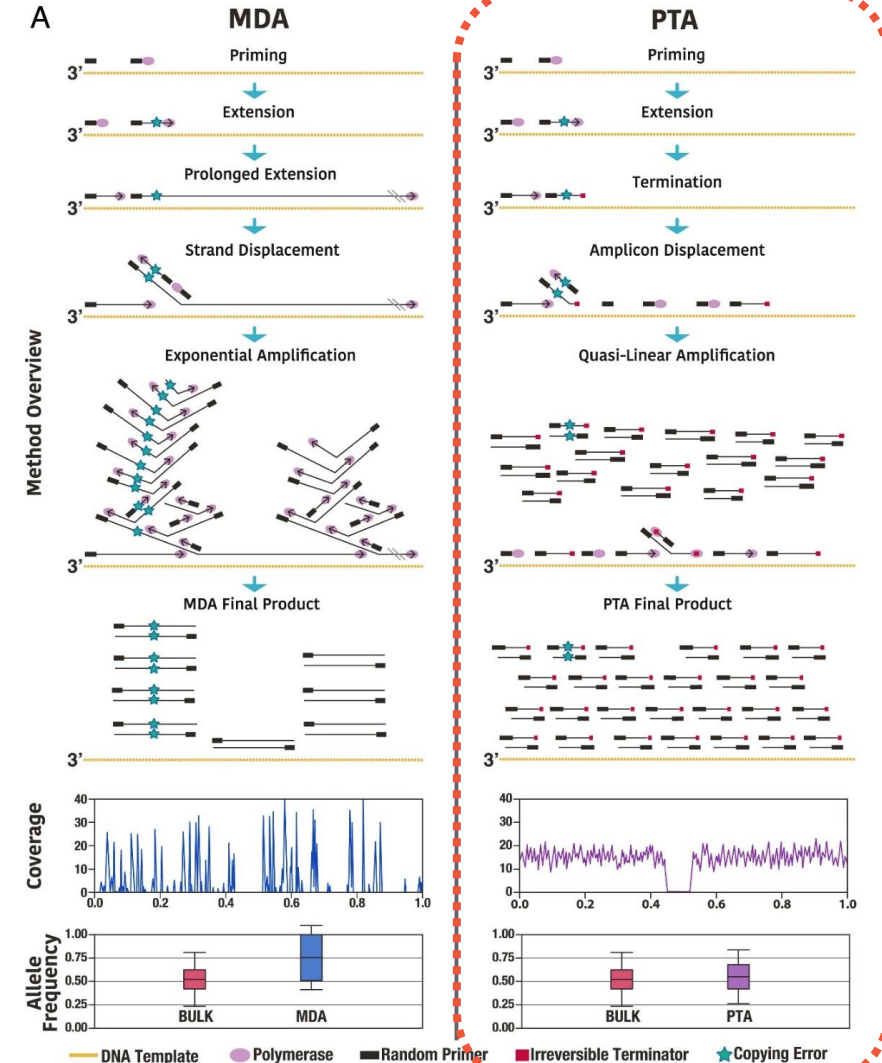
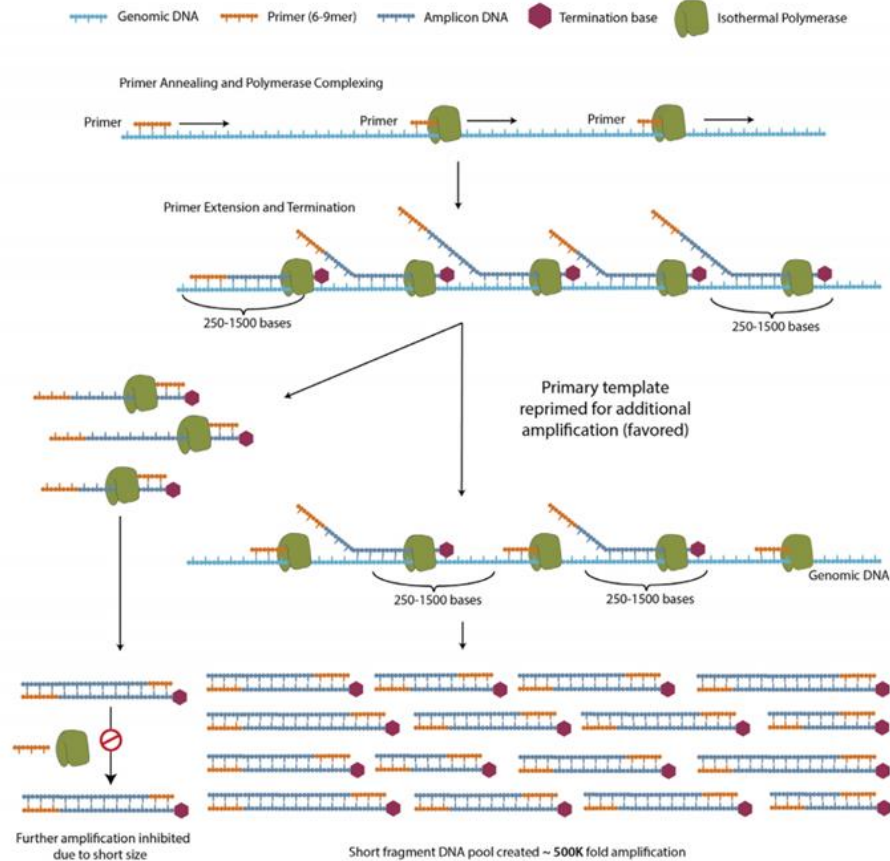
*Phillips (2020) FSIG*

# Materials and Methods (continued)

Gonzalez-Pena (2021) PNAS

## Primary template-directed amplification (PTA)

- ResolveDNA™ Whole genome Amplification Kit



# Workflow

Every experiment was independently  **duplicated** .

**Genomic DNA**

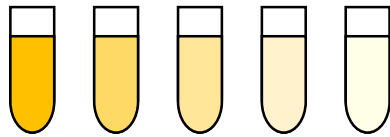
**PTA DNA**

**Intact DNA**

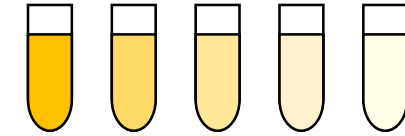
: 5 samples (4 males + 1 female)  
from 4 populations

*Ref.*

**10ng  
1ng  
300pg**



**Serial dilution**   
: 1ng ~ 300pg



**Serial dilution**   
: 300pg ~ 6.25pg

**300pg  
100pg  
25pg  
6.25pg**

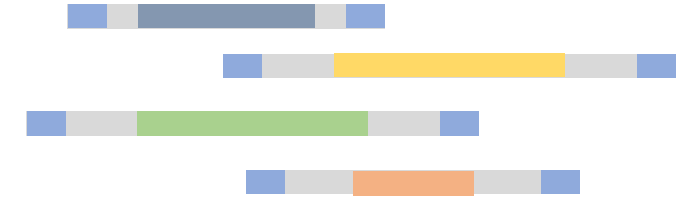
**Whole genome amplification**

: ResolveDNA™ Whole Genome Amplification Kit  
(PTA-based)

# Workflow (continued)

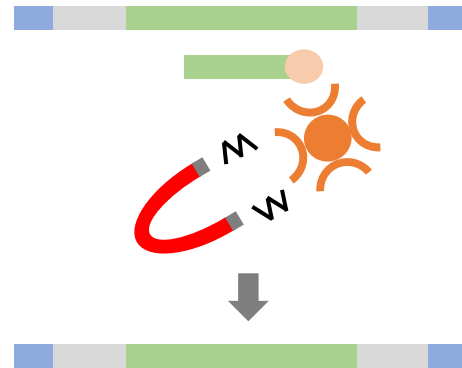
## Library preparation

:xGen™ DNA EZ Library Prep Kit, UDI (Unique Dual Index)



## Library QC (Quality Control)

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

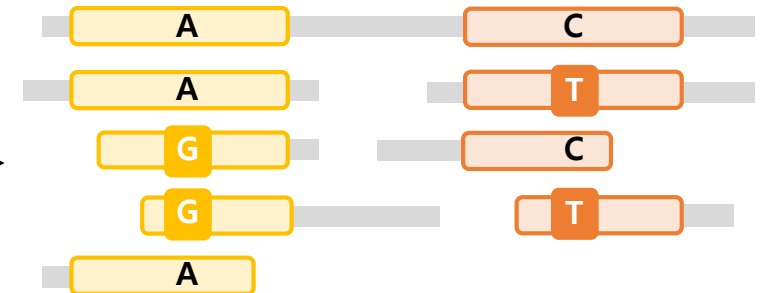


Lib.  
QC



## Next generation sequencing

: NextSeq 550 system (150bp, paired-end)



## Data analysis

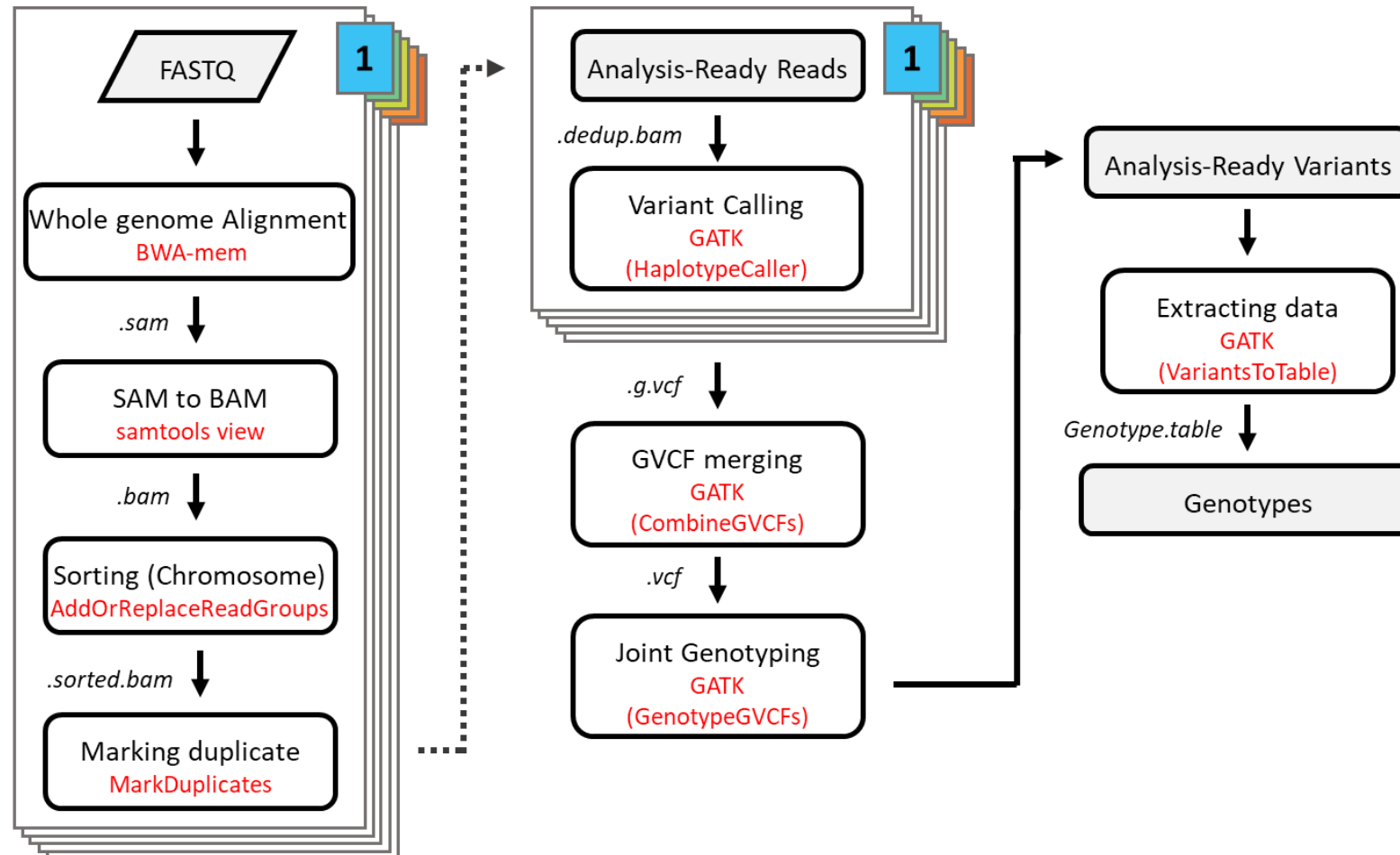
: Genome Analysis Tool Kit (GATK) pipeline  
In-house R script

## Hybrid capture

:1,225 identity-informative SNPs  
KAPA HyperChoice probe  
High fidelity PCR system (> x300)

# Workflow (continued)

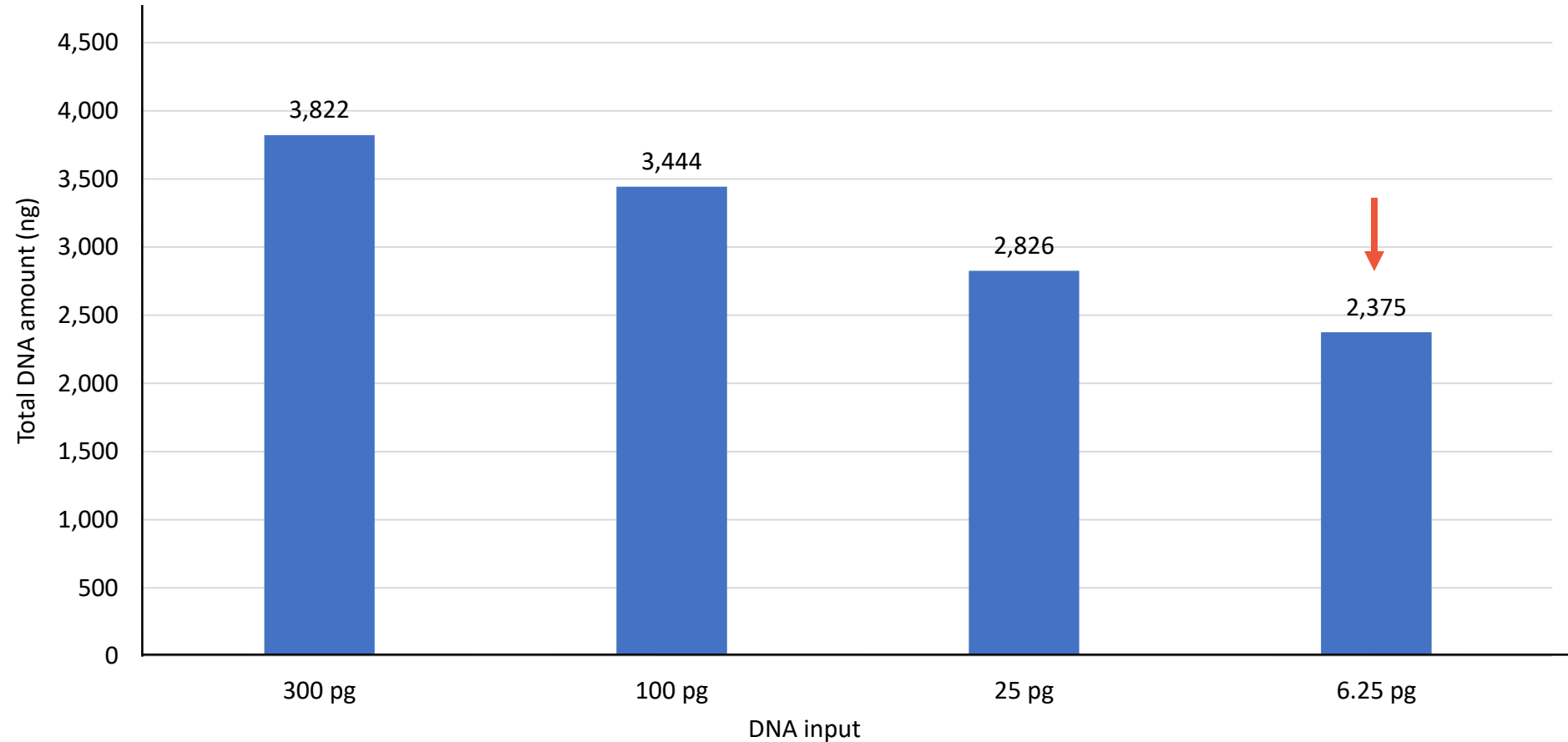
## NGS Data analysis pipeline



- **Read threshold: > 20**
- **Allele ratio: > 0.20**

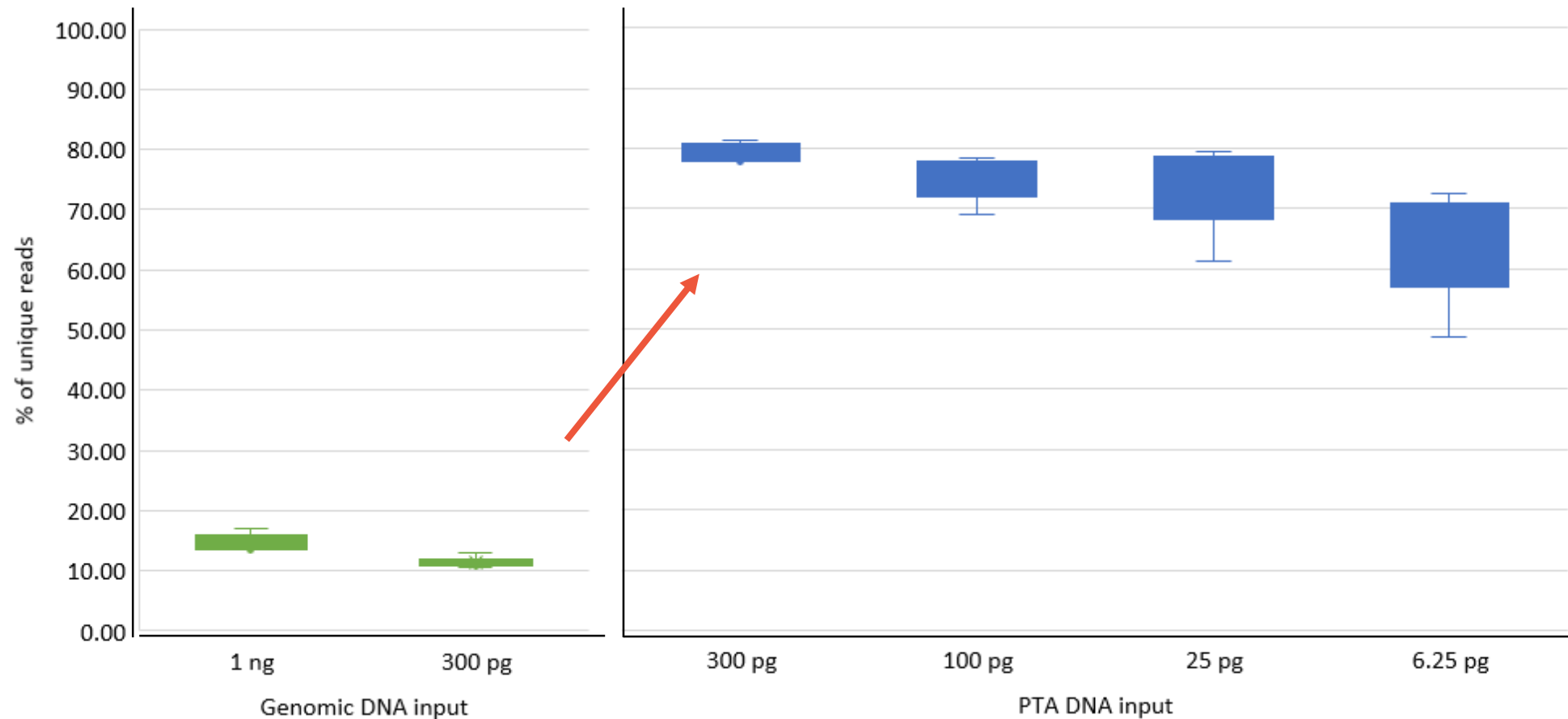
# Result & Discussion

## Performance of primary template-directed amplification



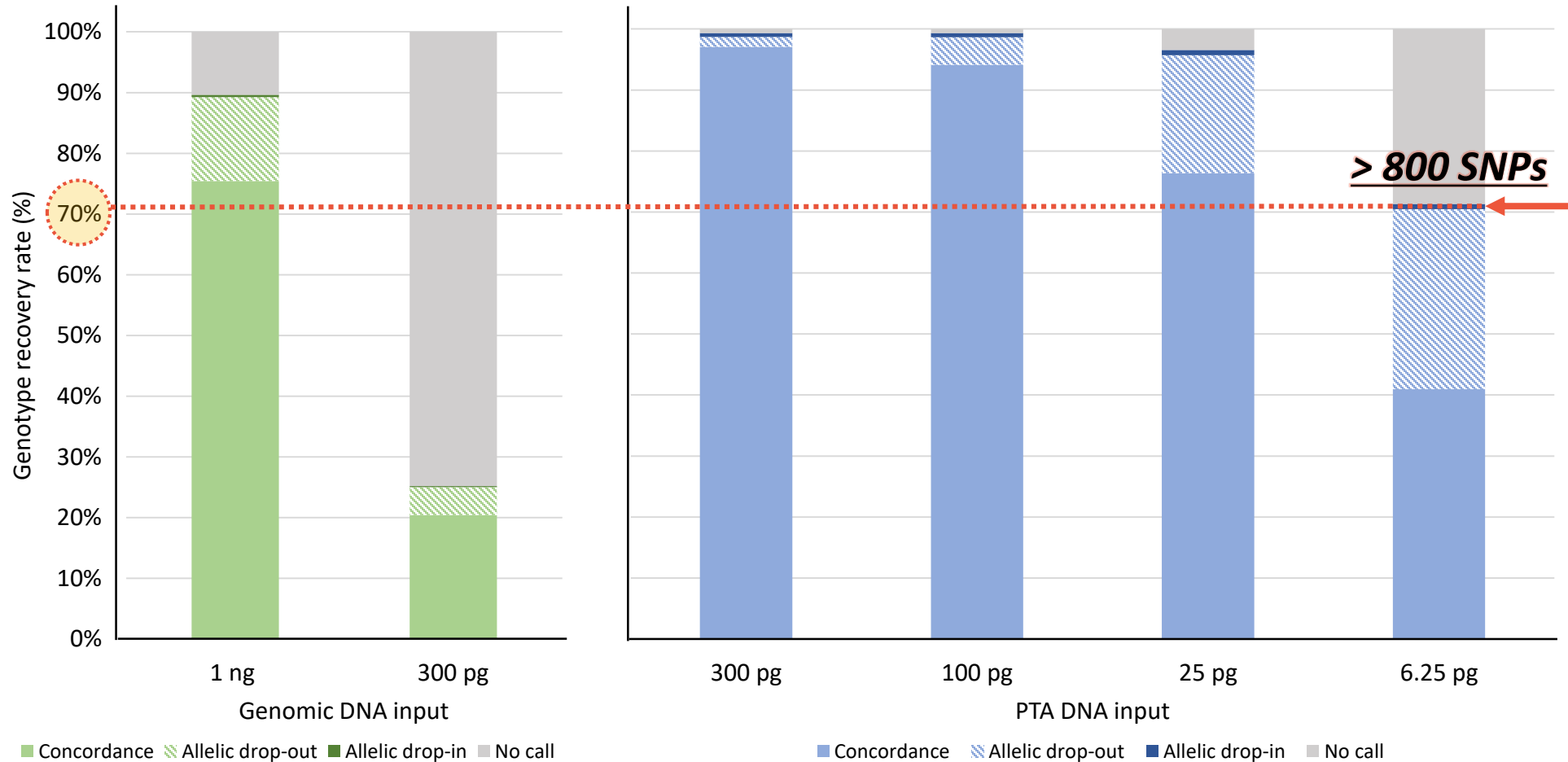
# Result & Discussion (continued)

## Library complexity



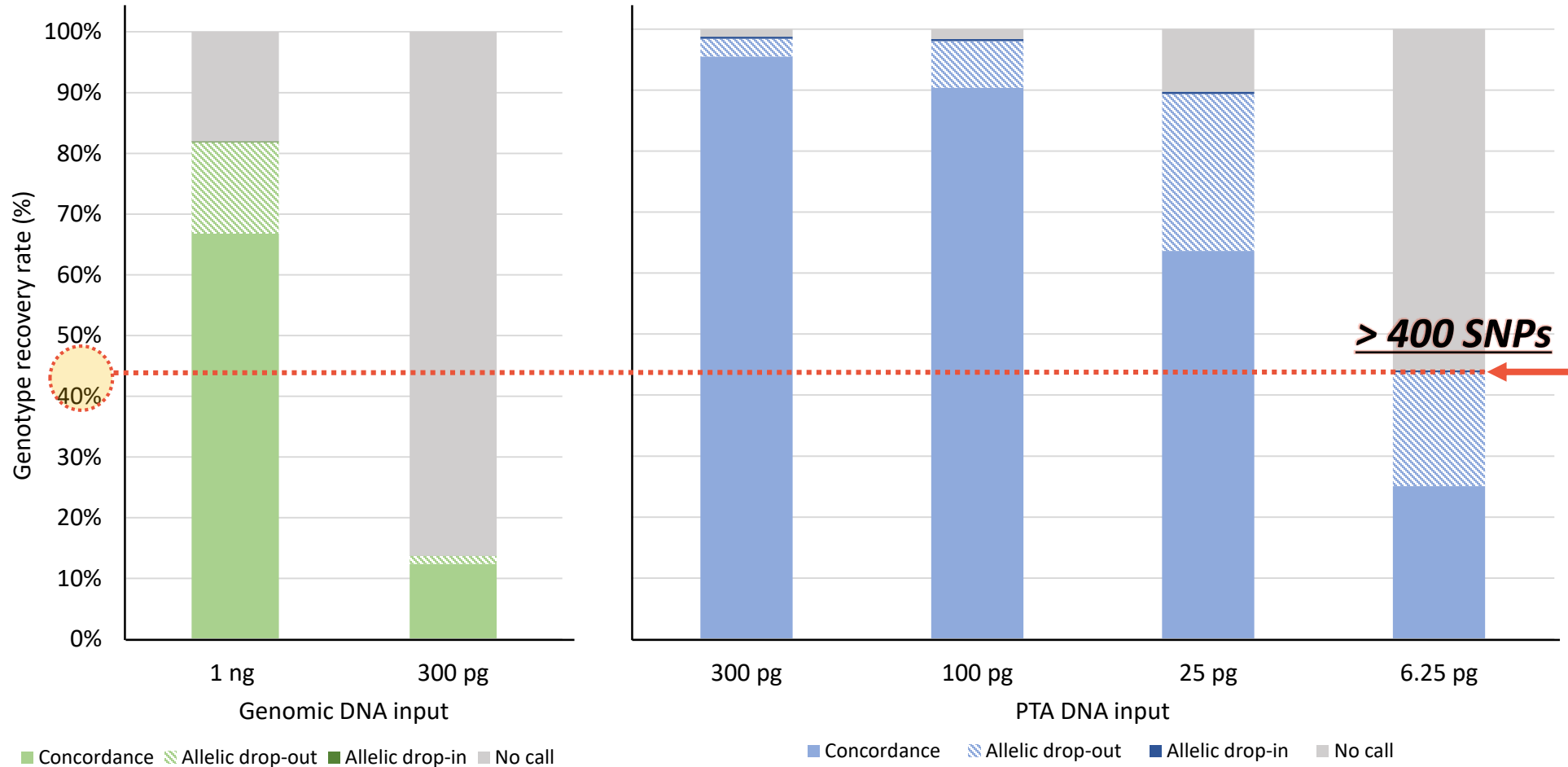
# Result & Discussion (continued)

## Genotype recovery rate (*Non-consensus method*)



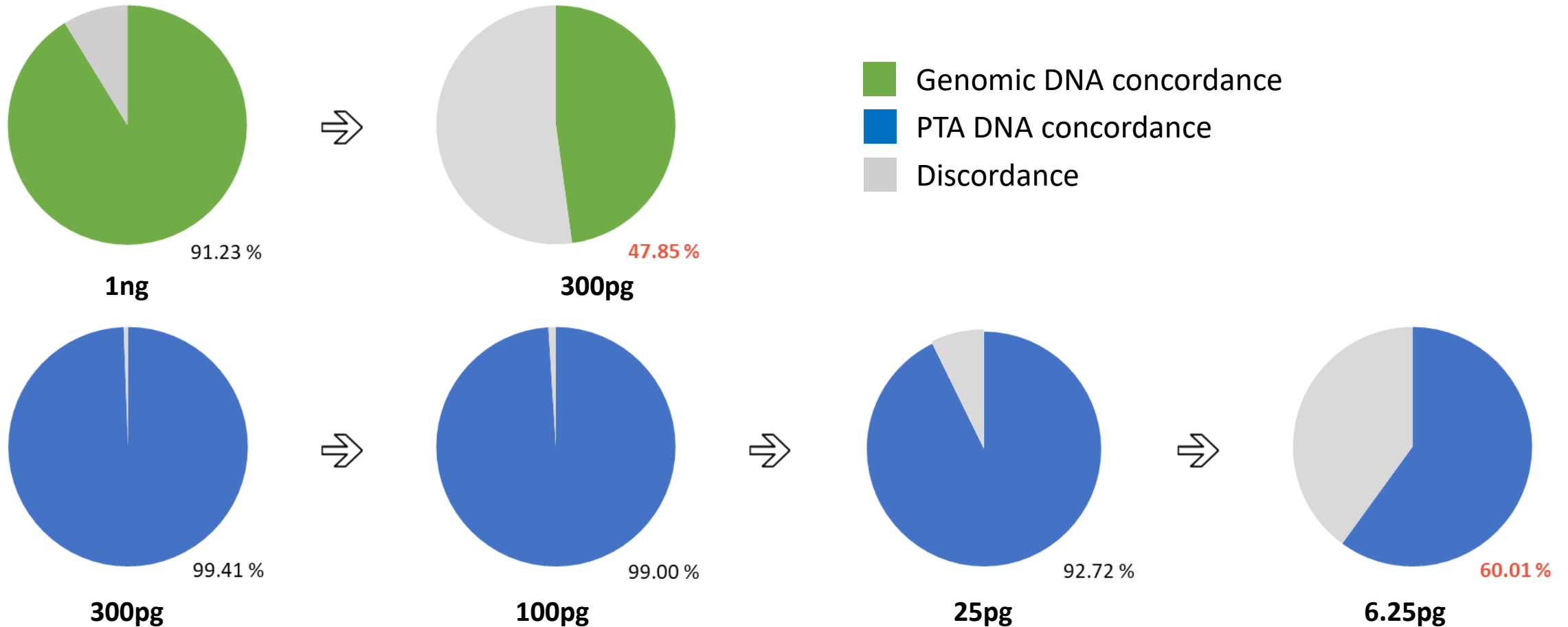
# Result & Discussion (continued)

## Genotype recovery rate (*Consensus method*)



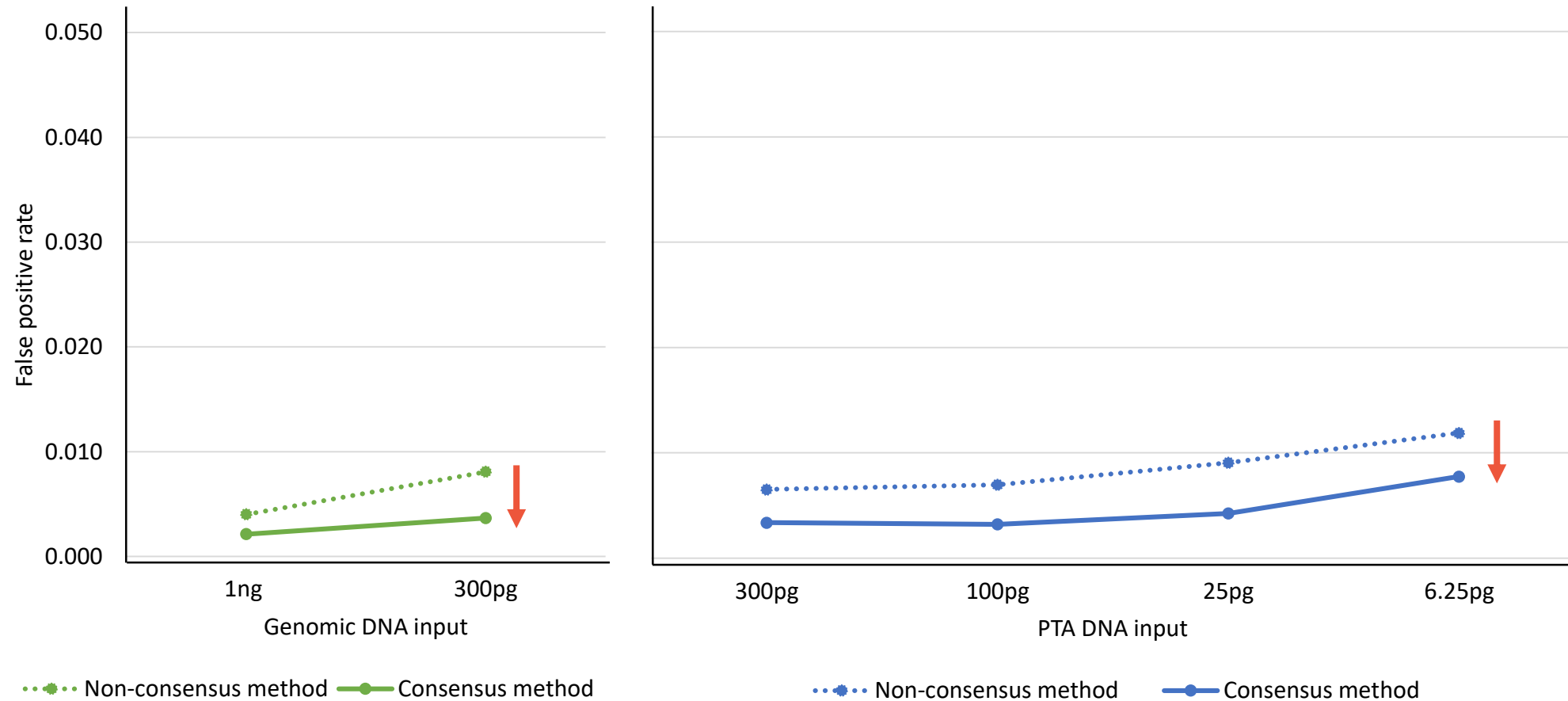
# Result & Discussion (continued)

## Concordance rate between duplicate pairs



# Result & Discussion (continued)

## False-positive rate

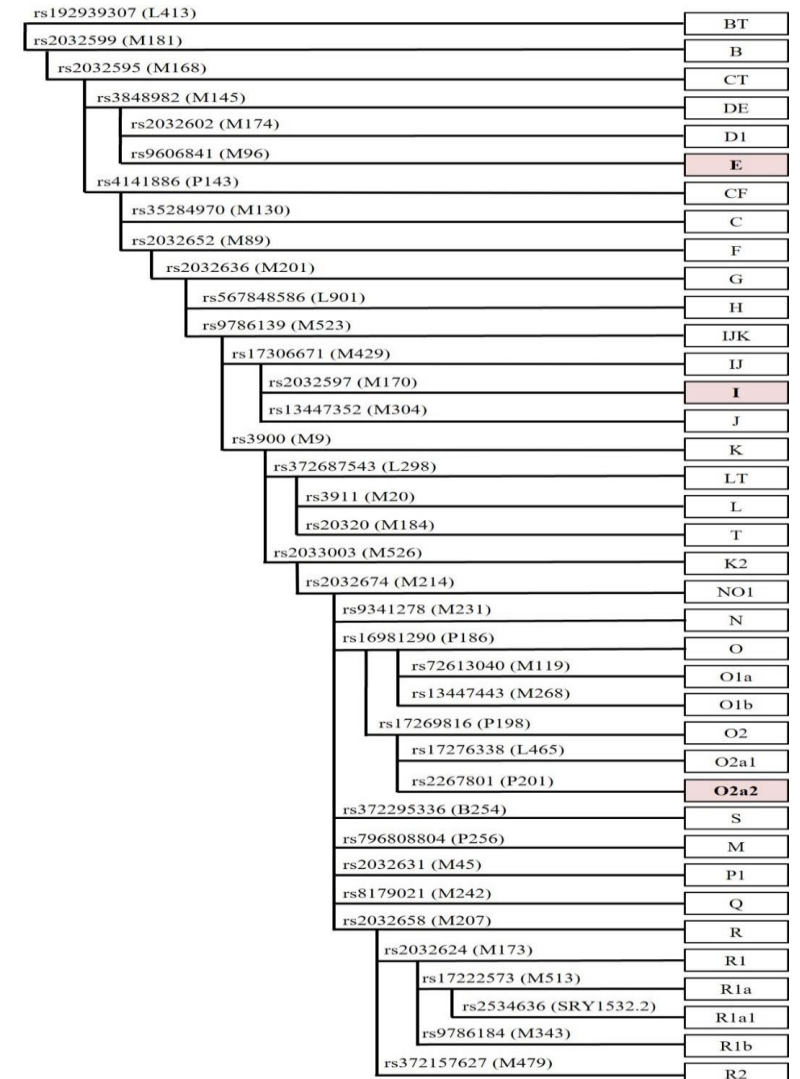


# Result & Discussion (continued)

## Y-DNA Haplogroup

- **38** Y-chromosomal SNPs
- International Society of Genetic Genealogy (ISOGG) Y-DNA Haplogroup Tree 2019–2020 v15.73

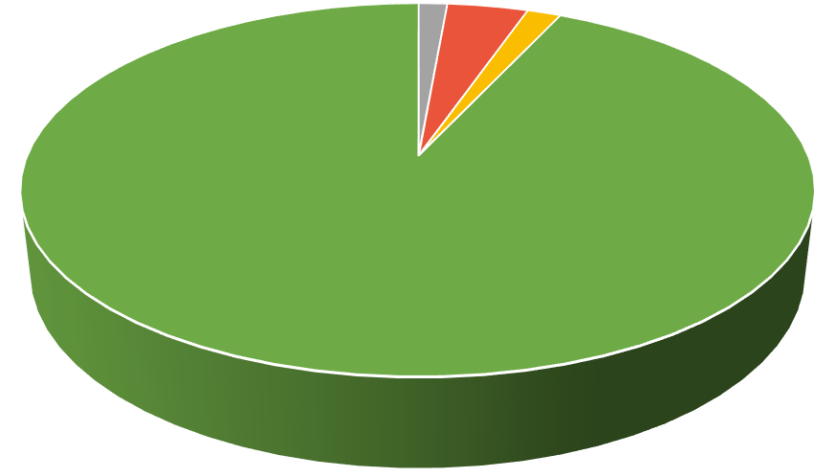
Sample	Population	Haplogroup
NA17102	African-American	E
NA17288	Caucasian	I
NA17661	Hispanic	I
ST65	Korean	O2a2



# Result & Discussion (continued)

## Poorly performing SNPs

- **Non-called SNPs** in reference samples
- **Mis-matched SNPs** in reference samples
- **3<sup>rd</sup> allele SNPs** in all samples



e.g.

rs number	Chr	Position	Allele
rs76617082	12	64141458	C/G/T

⇒ **GATK HaplotypeCaller genotype !**

Classification	Count (%)	Remark
Non-called SNPs	17 (1.4)	rs10911577, rs7427284, rs2709880, rs1693937
Mis-matched SNPs	48 (3.9)	-
3 <sup>rd</sup> allele SNPs	20 (1.6)	rs76617082, rs2657061, rs28621711
<b>Total</b>	<b>71 (5.8)</b>	<b>14 SNPs overlap.</b>

# Summary

---

- ◆ **PTA could produce sufficient microgram level DNA from sub-nanogram level DNA and increase the NGS library complexity of low template DNA.**
- ◆ **Consequently, we obtained 90% genotype recovery using 25pg of input DNA and 70% using only 6.25pg of input DNA (non-consensus method).**
- ◆ **Low false positive rates (approximately 1%) showed that the system achieved high accuracy and robustness.**

We successfully analyzed **1.2K SNPs** from **sub-nanogram level DNA** using **PTA and hybrid capture-based NGS**, demonstrating its **potential for LT DNA analysis in forensic practice.**

# Acknowledgements



***Kyoung-Jin Shin***

Professor

***Ye-Lim Kwon***

Post-doc

***Sumin Joo***

Research assistant  
(juicemin@yuhs.ac)

***Jiwon Kim***

M.Sc. student



This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (Project Number: **2021R1A2C101306313**)

