International Conference of the Genetics Society of Korea
Oct. 20, 2023

MH-UMIseq:
improvement of signal-to-noise ratio in NGS of microhaplotype with unique molecular identifiers

Ye-Lim Kwon, Ph.D.

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

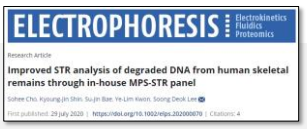
Main challenges in forensic DNA profiling

- Main topics of our research group

1

Degraded DNA

- Short amplicon
- 25-plex A-STR / 24-plex Y-STR NGS panel



2

Low copy number DNA ***(Ongoing study)***

- PTA-based whole genome amplification (WGA)
  → Hybrid-capture for NGS

3

Mixture deconvolution ***(Ongoing study)***

- Microhaplotypes NGS

ICGSK 2023

Microhaplotypes

T — G — G

C — G — G

< 300 nucleotides

Genotype : TGG / CGG

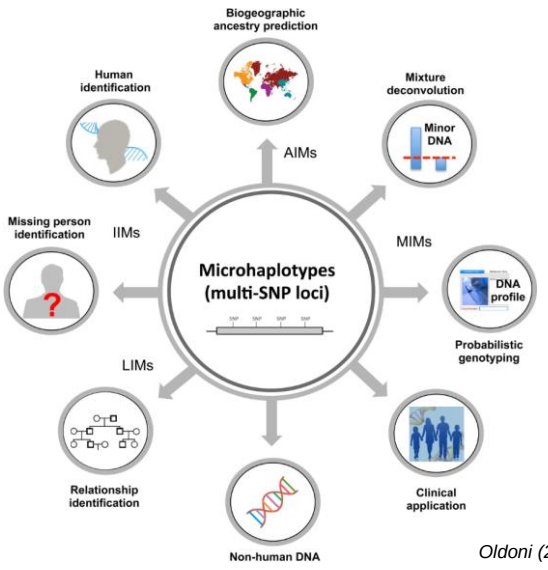
Next generation sequencing (NGS)

- High throughput multiplexing
- Phase-known haplotyping

- Low mutation rate
- High heterozygosity
- No stutter artifacts

ICGSK 2023

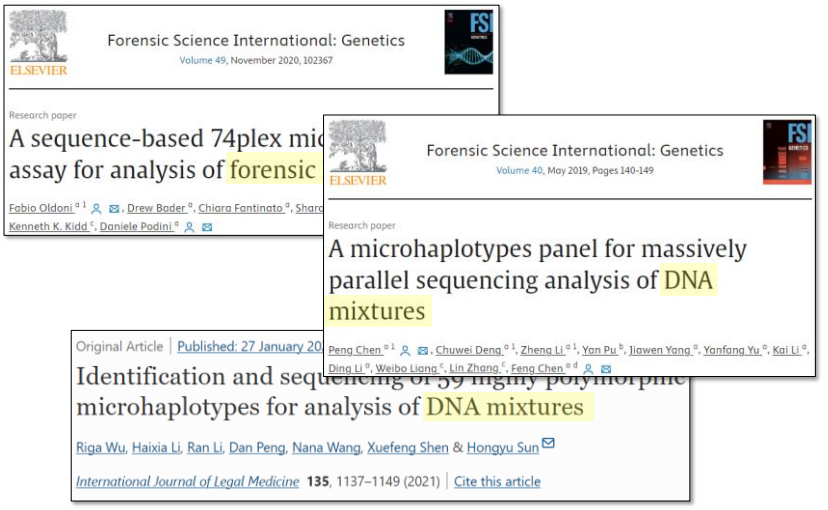
Microhaplotypes



Oldoni (2019) FSIG

ICGSK 2023

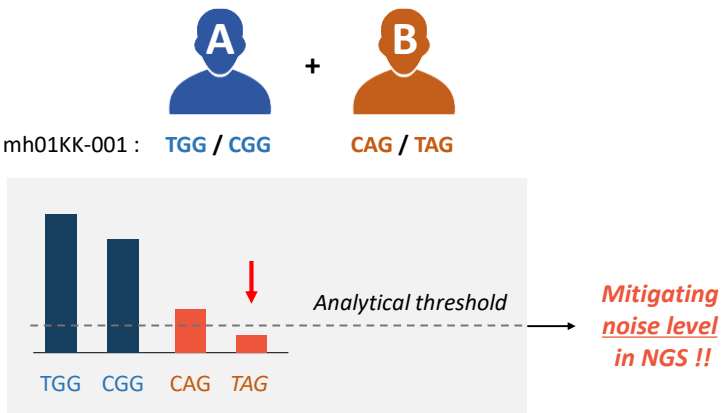
Microhaplotypes



ICGSK 2023

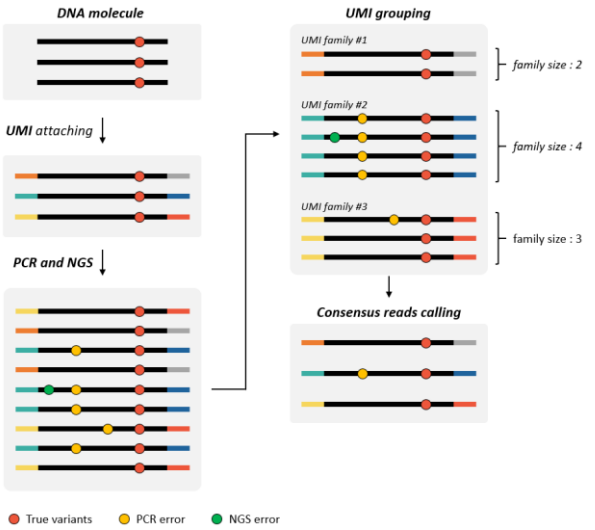
Minor contributor or Noise in Mixture?

> In a crime scene scenario,



ICGSK 2023

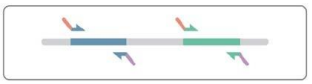
Unique molecular identifiers (UMIs)



ICGSK 2023

Attaching UMIs to DNA molecules

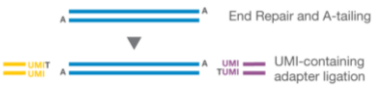
1. PCR-based method



- **Pros** : High sensitivity (≒ 1 ng)
- **Cons** : Small number of targets
  (Difficult to design primer)

Suitable for forensic DNA

2. Adapter ligation method



- **Pros** : Large-scale multiplexing
  (> 100 kb)
- **Cons** : Low sensitivity (> 10 ng)

e.g. QIAseq Targeted DNA Panel (Qiagen)

Allele- / locus- drop out, < 2 ng DNA

Truelsen (2021) Sci. Rep.

ICGSK 2023

Research objectives

General objective

Construct **UMI-based amplicon sequencing system for microhaplotype** and investigate **false-positive rate** of UMI data according to **the amount of input DNA**.

Specific objectives

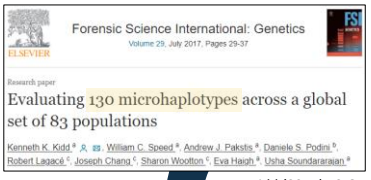
1. Development of **MH-UMIseq** Panel
2. Construction of **MH-UMI data analysis pipeline**
3. **Evaluating performance** of MH-UMIseq

ICGSK 2023

1. Development of MH-UMIseq Panel

Marker selection

- **130 Microhaplotypes**

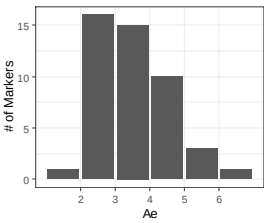


Kidd (2017) FSIG

- ***46 Microhaplotypes***

  - Consist of 155 SNPs

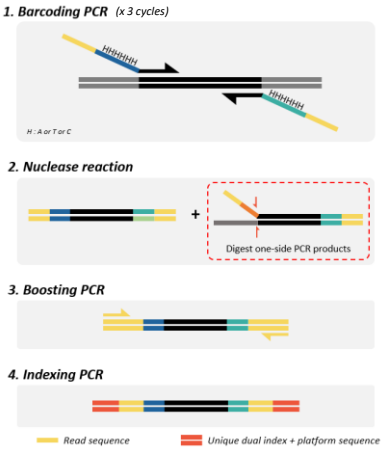
  - Target size: 124 ~ 249 bp



ICGSK 2023

1. Development of MH-UMIseq Panel

MH-UMIseq



- **UMI barcode**
  - Poly (H_6) x Poly (H_6) = Poly (H_{12})

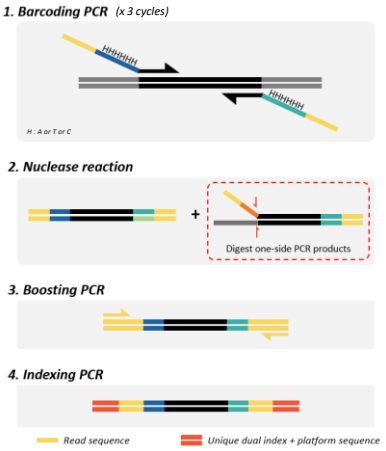
    = 3^{12} = 531 K
- **Remove PCR artifacts by <u>Nuclease</u>**

  by CleanPlex (Paragon Genomics)
- **UMI-attached target amplification by <u>Boosting PCR</u>**

ICGSK 2023

1. Development of MH-UMIseq Panel

MH-UMIseq



➢ ***<u>To mitigate PCR and NGS error</u>***

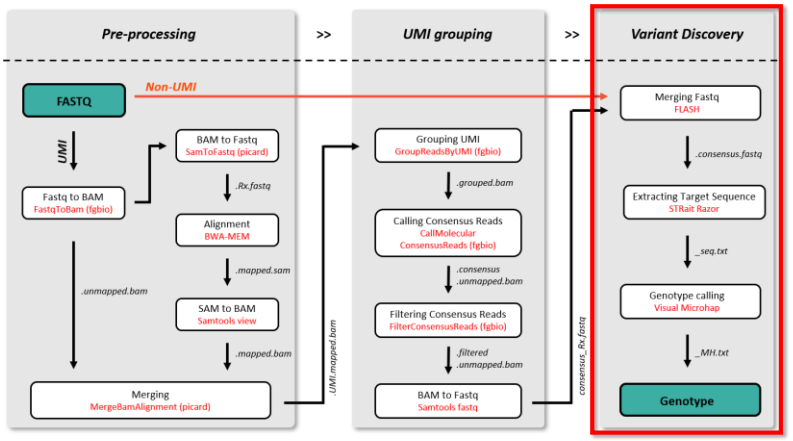
- **High-fidelity enzyme** (> x 300)
- ***Unique dual index (UDI)***

  ↔ Combinatorial dual index (CDI)

ICGSK 2023

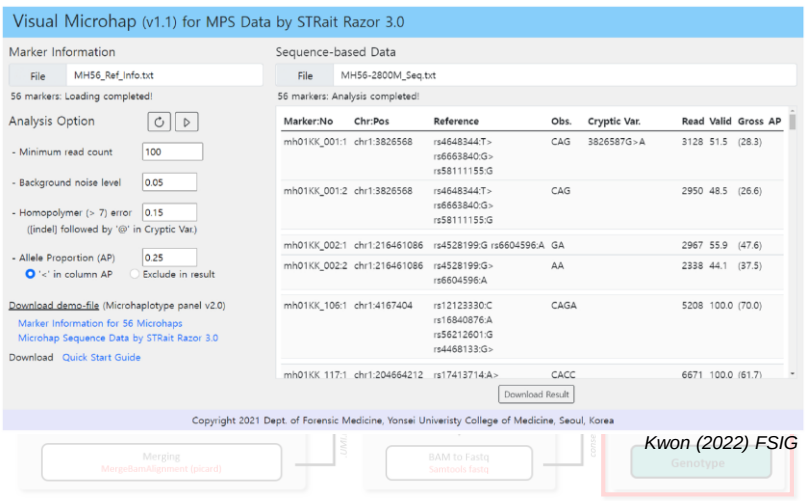
2. Construction of MH-UMI data analysis pipeline

- Linux-based pipeline + ***STRait Razor + Visual Microhap***



ICGSK 2023

2. Construction of MH-UMI data analysis pipeline



Kwon (2022) FSIG

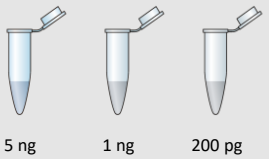
ICGSK 2023

3. Evaluating performance of MH-UMIseq

Material and Method

- 31 Korean blood samples (20 Male + 11 Female) → *Real-time PCR quantification*
  Severance IRB. 4-2022-1534

- 31 samples x 3 types of input DNA = 93 samples



NGS Lib. Prep. with MH-UMIseq

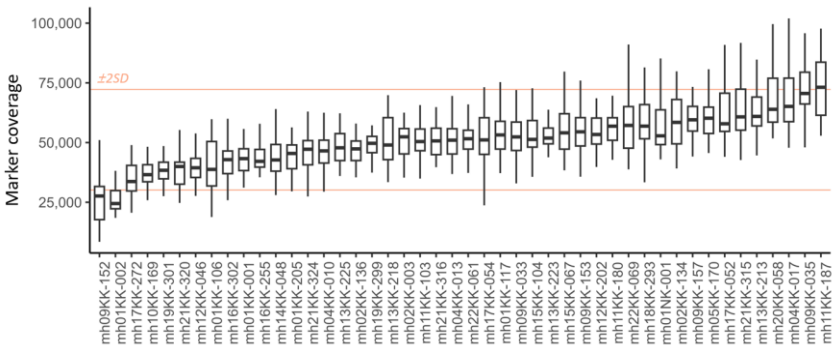
3 batches of *NextSeq 550 run* Mid Output kit
+
1 batch of *MiniSeq run* High Output kit

ICGSK 2023

3. Evaluating performance of MH-UMIseq

Marker coverage

e.g. 1ng of DNA

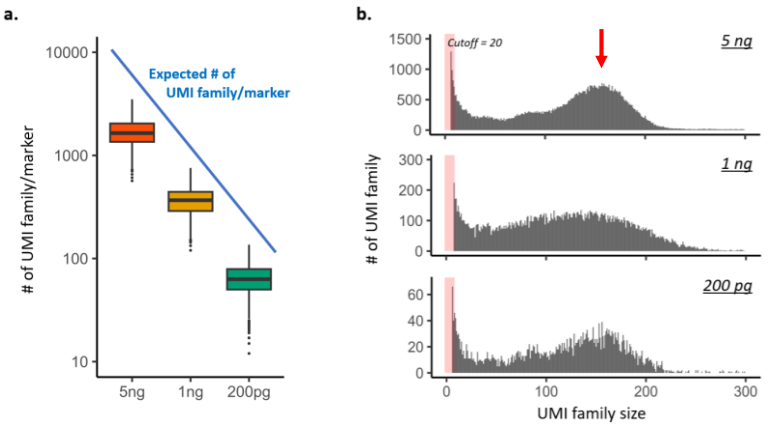


(Manuscript in preparation)

ICGSK 2023

3. Evaluating performance of MH-UMIseq

UMI family distribution



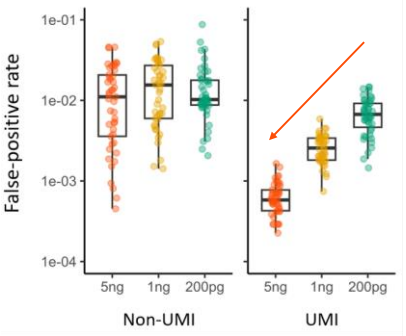
∴ **# of UMI family ∝ amount of input DNA**

(Manuscript in preparation)

ICGSK 2023

3. Evaluating performance of MH-UMIseq

False-positive rate



- *Based on Minimum analytical options*

UMI type	Min. read cov. / UMI family size	noise cutoff
Non-UMI	x 20	0.001
UMI	x 20	0.000

∴ **False-positive rate: non-UMI > UMI**

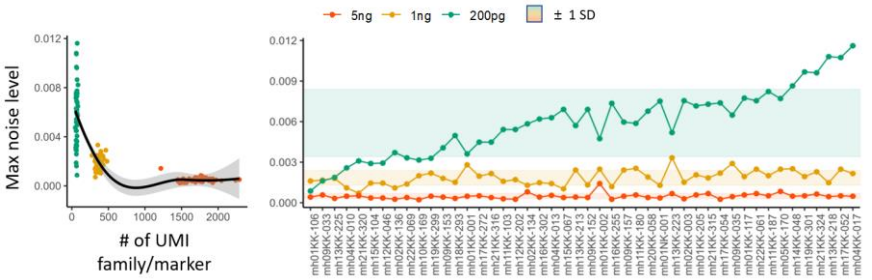
∴ **amount of input DNA ∝ efficiency of UMI system**

(Manuscript in preparation)

ICGSK 2023

3. Evaluating performance of MH-UMIseq

Max noise level of MH-UMIseq



➢ ***Expected resolution of MH-UMIseq***

	5ng	1ng	200pg
Minor component ratio	1:231	1:64	1:13

(Manuscript in preparation)

ICGSK 2023

Summary

- The **UMI-based 46-plex amplicon sequencing system** and **UMI data analysis pipeline for microhaplotypes** were well established.
- The **false-positive rate was significantly reduced in UMI system**, especially for **higher amounts of input DNA**.
- The **noise of MH-UMIseq** is mainly affected by **the number of molecules** (= amount of input DNA).
  - The **noise level decreases** as the **input increases**.
- To obtain a **resolution of 1:50 or higher**, **≥ 1 ng of DNA must be used**.

ICGSK 2023

Future application of UMI system

- It is expected to achieve high accuracy and sensitivity (resolution) in **mixed DNA analysis**, especially for high amount of DNA.

+ DNA Methylation analysis

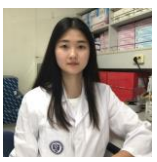
- It is expected to improve the accuracy and sensitivity of **age, body fluid prediction** by **DNA methylation analysis**.
- It enables simultaneous **human identification** and **methylation analysis** of **mixed samples**.

ICGSK 2023

Acknowledgements

➢ ***Yonsei DNA profiling group***


Kyoung-Jin Shin
Professor


Ye-Lim Kwon
Post-doc


Sumin Joo
Research assistant


Jiwon Kim
M.Sc. student

This work was supported by the Korean National Police Agency
(Project Number: 0411-20230029)

ICGSK 2023