23rd Triennial Meeting of the International Association of Forensic Sciences
Nov. 23, 2023

MH-UMIseq system,

unique molecular identifier-based NGS of microhaplotypes,

can improve signal-to-noise ratio

even for sub-nanogram quantities of DNA

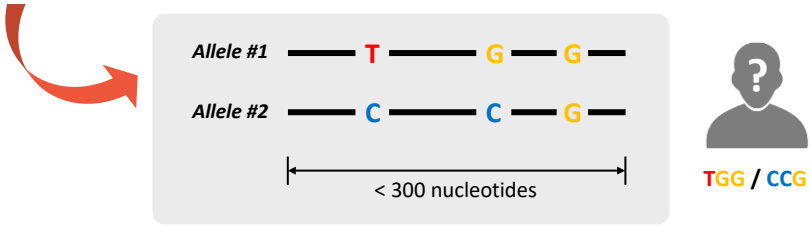
***Ye-Lim Kwon* and Kyoung-Jin Shin**

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

Microhaplotypes (Microhaps, MHs)

NGS
(Next Generation Sequencing)

- High throughput multiplexing
- Phase-known haplotyping



Kidd (2016) FSIG

- ***High heterozygosity*** *(≒ STRs, > SNPs)*
- ***Low mutation rate*** *(< STRs)*
- ***No stutter artifacts*** *(↔ STRs)*

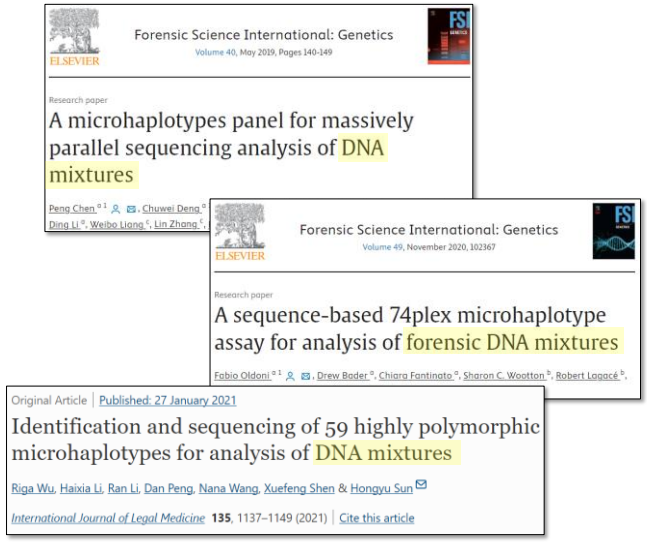
Microhaplotypes for forensic applications

IAFS 2023

- ***Mixture deconvolution*** (Minor DNA)
- ***Human identification***
- ***Relationship identification***
- ***Biogeographic ancestry inference***

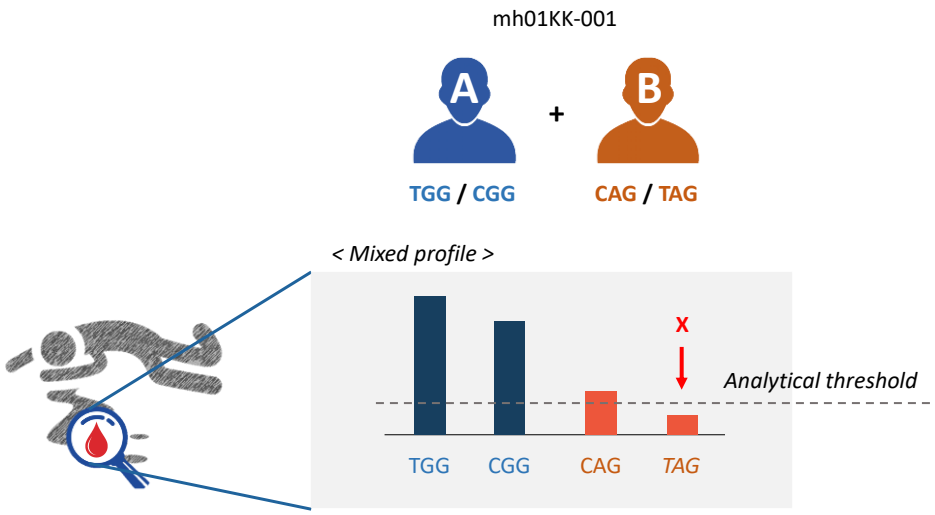
⋮

Oldoni (2019) FSIG



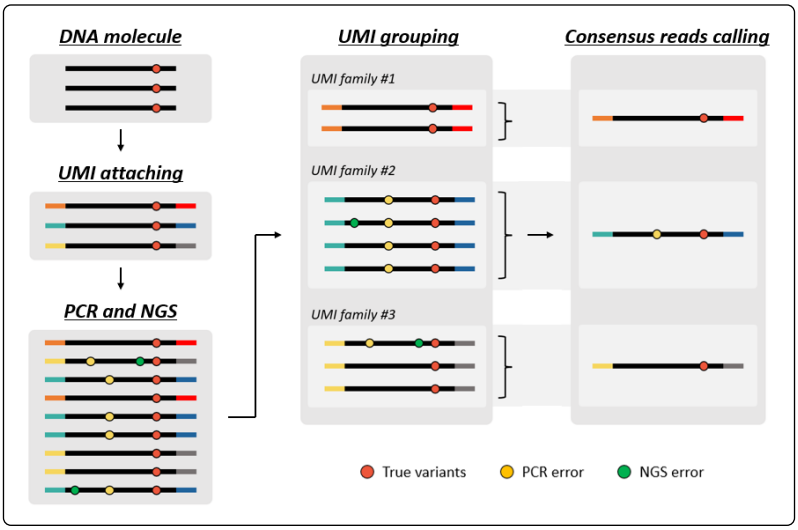
In a crime scene scenario,

IAFS 2023



Unique molecular identifiers (UMIs)

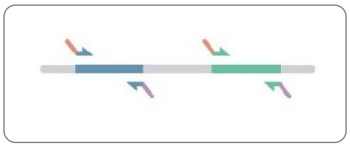
IAFS 2023



Two methods attaching UMIs to DNA molecules

IAFS 2023

❖ PCR-based method



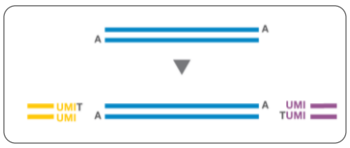
Cleanplex (Paragon Genomics)

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

⇩

Suitable for forensic DNA

❖ Adapter ligation method



TruSight (Illumina)

- ***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.

Research objectives

IAFS 2023

Construct *UMI-based amplicon sequencing system for microhaplotype,* ***MH-UMIseq system,*** *and* ***investigate false-positive proportion*** *of MH-UMIseq according to* ***the amount of input DNA***

❖ *Specific objectives*

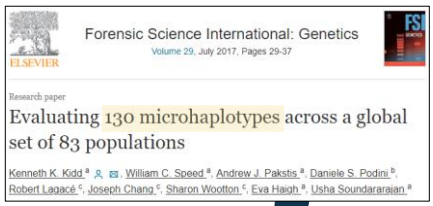
1. Development of **MH-UMIseq** panel
2. Construction of **MH-UMI data analysis pipeline**
3. **Evaluating performance** of MH-UMIseq

Development of MH-UMIseq panel

IAFS 2023

❖ *Marker selection*

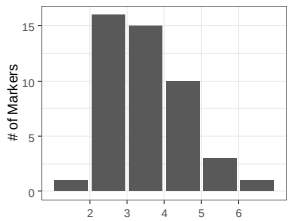
- **Among 130 microhaplotypes**


Forensic Science International: Genetics
Volume 29, July 2017, Pages 29-37
Research paper
Evaluating 130 microhaplotypes across a global set of 83 populations
Kenneth K. Kidd, William C. Speed, Andrew J. Pakstis, Daniele S. Podini, Robert Lagacé, Joseph Chang, Sharon Wootton, Eva Haigh, Usha Soundararajan

Kidd (2017) FSIG

"46 Microhaplotypes"

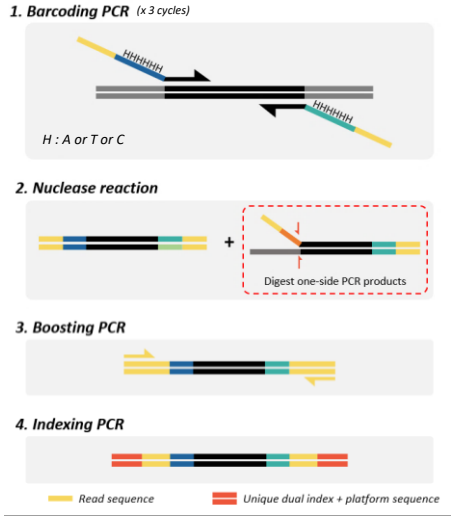
- Consist of 155 SNPs
- Target size: 124 ~ 249 bp



Development of MH-UMIseq panel

IAFS 2023

❖ NGS Library preparation of MH-UMIseq



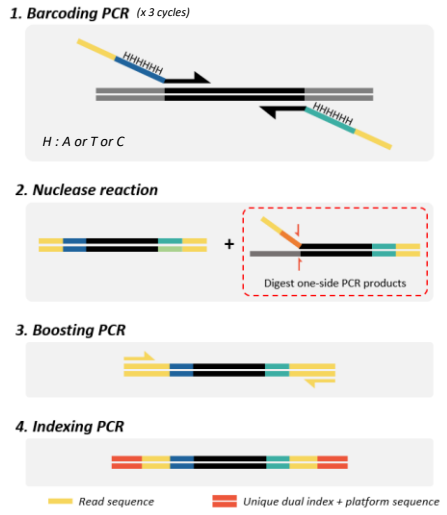
- **Target-specific UMI primer**
- **UMI barcode**
  - Consists of Poly (**H**: A, T, or C)
  - Poly (H_6) x Poly (H_6) = Poly (H_{12}) = 3^{12} = 531 K
- **Remove PCR artifacts by <u>Nuclease</u>**

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

Development of MH-UMIseq panel

IAFS 2023

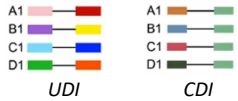
❖ NGS Library preparation of MH-UMIseq



<u>+ To mitigate PCR and NGS error</u>

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

  ↔ Combinatorial dual index (CDI)

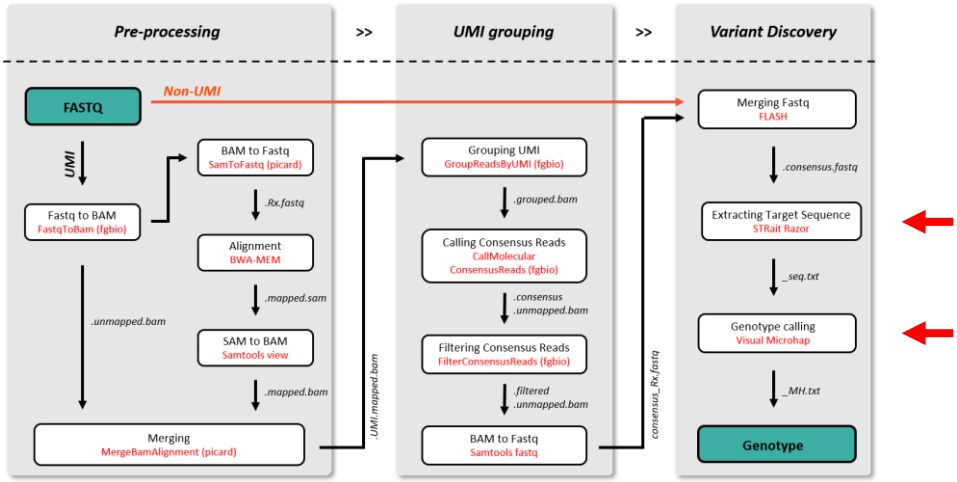


UDI *CDI*

Construction of MH-UMI data analysis pipeline

IAFS 2023

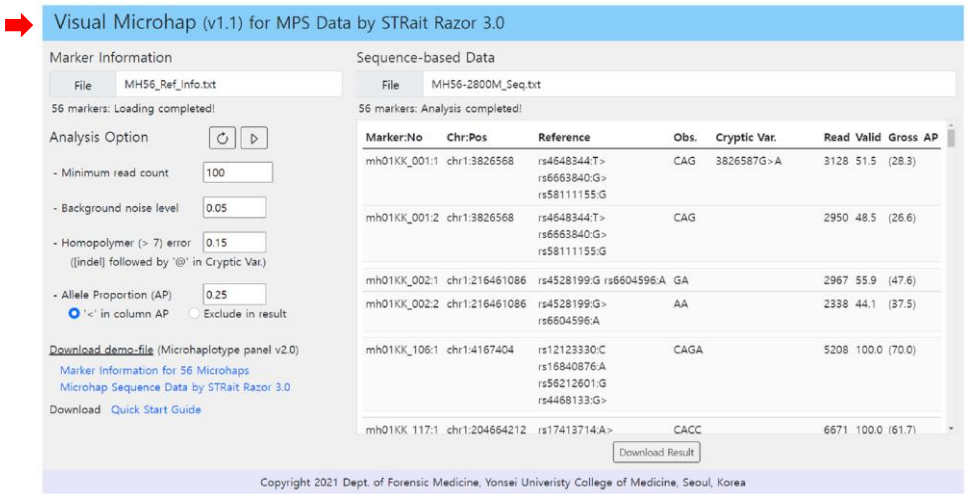
- Linux-based pipeline + ***" STRait Razor + Visual Microhap "***



Construction of MH-UMI data analysis pipeline

IAFS 2023

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



Kwon (2022) FSIG

Evaluating performance of MH-UMIseq

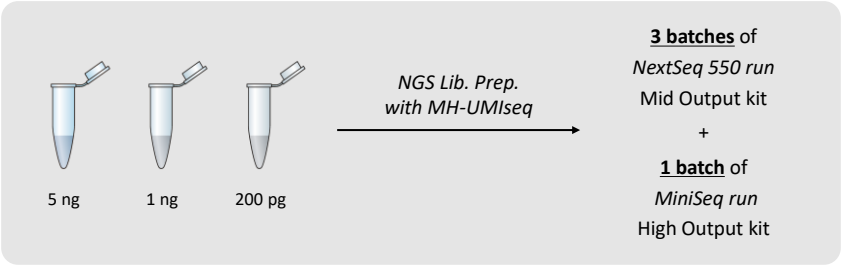
IAFS 2023

❖ Material and Method

- 31 Korean blood samples (20 Male + 11 Female) → *Real-time PCR quantification*

  IRB approval) 4-2022-1534

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

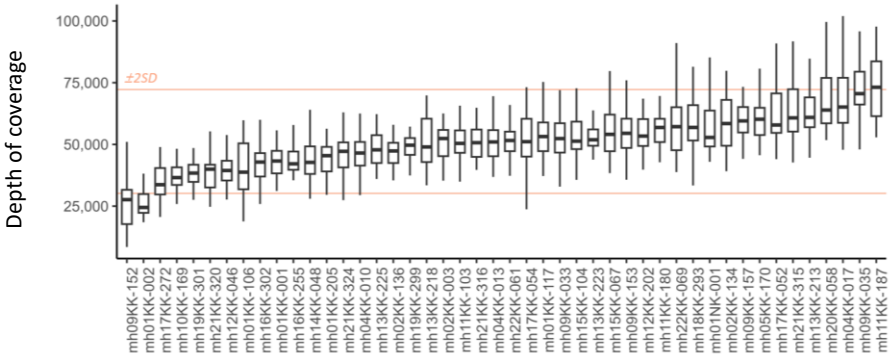


Evaluating performance of MH-UMIseq

IAFS 2023

❖ Depth of Coverage

e.g. 1ng of DNA

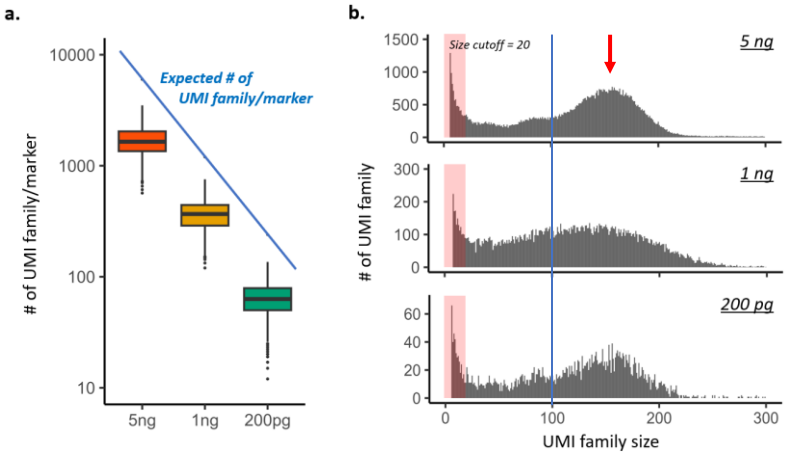


(Manuscript in preparation)

Evaluating performance of MH-UMIseq

IAFS 2023

❖ *UMI family distribution*



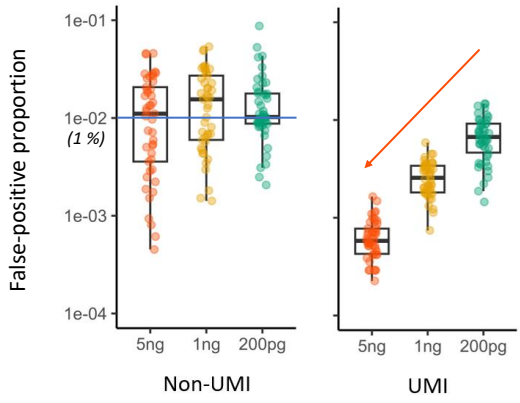
∴ **# of UMI family ∝ Amount of input DNA**

(Manuscript in preparation)

Evaluating performance of MH-UMIseq

IAFS 2023

❖ *False-positive proportion*



- *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 proportion: non-UMI > UMI**

∴ **Amount of input DNA ∝ Resolution of UMI system**

(Manuscript in preparation)

Conclusions

- The **UMI-based 46-plex amplicon sequencing system** and **UMI data analysis pipeline for microhaplotypes** were successfully established.
- The **noise level of MH-UMIseq** is mainly affected by **the number of molecules** (= amount of input DNA).
- The **false-positive proportion was significantly reduced in UMI system**, especially for **higher amounts of input DNA**.

MH-UMIseq is expected to have **higher resolution** for **large amounts of input DNA in mixture analysis**.

Acknowledgements

Kyoung-Jin Shin
Professor

Ye-Lim Kwon
Post-doc
(yelim117@yuhs.ac)

Sumin Joo
Research assistant

Jiwon Kim
M.Sc. student

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