

Denoising of microhaplotype MPS data by computational correction and its application to mixed DNA analysis

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



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



Microhaplotypes

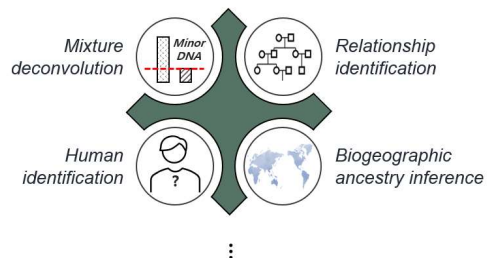
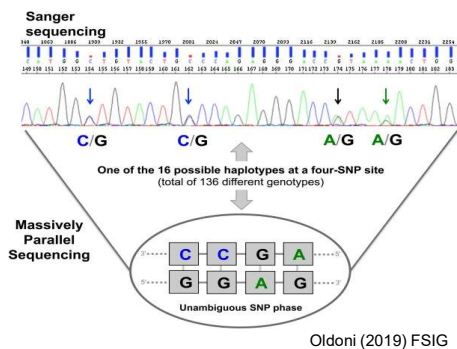
Allele #1 — T — G — G —

Allele #2 — C — C — G —

< 300 nucleotides

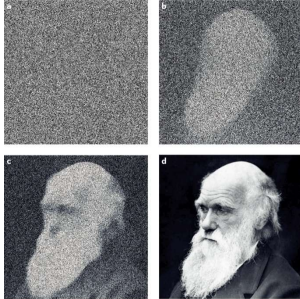
- No stutter artifacts (↔ STRs)
- No preferential amplification (↔ STRs)
- High heterozygosity (≅ STRs, > SNPs)
- Low mutation rate (< STRs)

Kidd (2016) FSIG



Noise in Mixed DNA Analysis

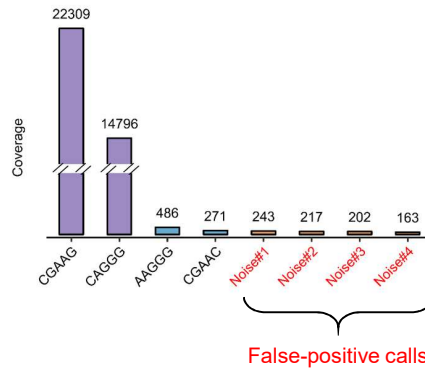
The signal-to-noise problem



Salk (2018) Nature Reviews Genetics

❖ 1:100 two-person mixed profile for MH

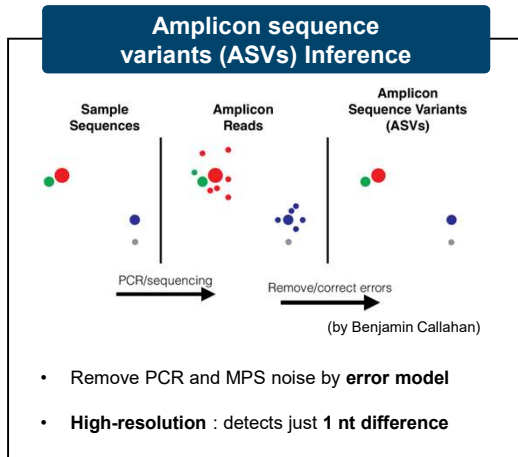
→ 1% of minor DNA



- Minor allele?
- Number of contributors?
- Mixture ratio?

Denoising MPS data by computational correction

Amplicon sequence variants (ASVs) Inference



DADA2: High-resolution sample inference from Illumina amplicon data



Benjamin J Callahan¹, Paul J McMurdie², Michael J Rosen³, Andrew W Han², Amy Jo A Johnson² & Susan P Holmes¹

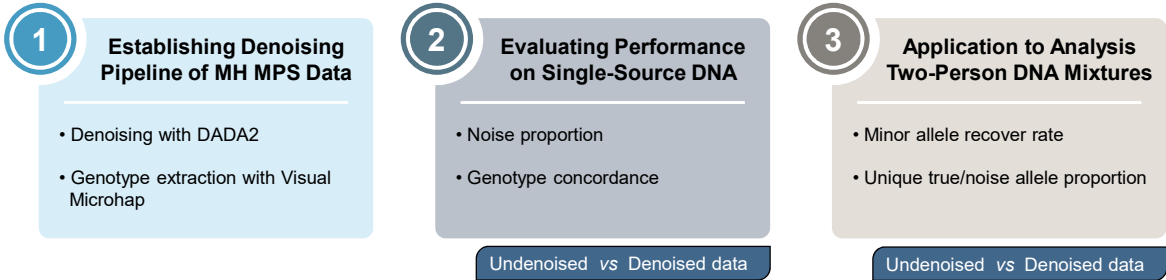
NATURE METHODS | VOL.13 NO.7 | JULY 2016 | 581

- Denoising algorithms to detect ASVs
- **Open-source** R package
- Best practice in **microbiome research**

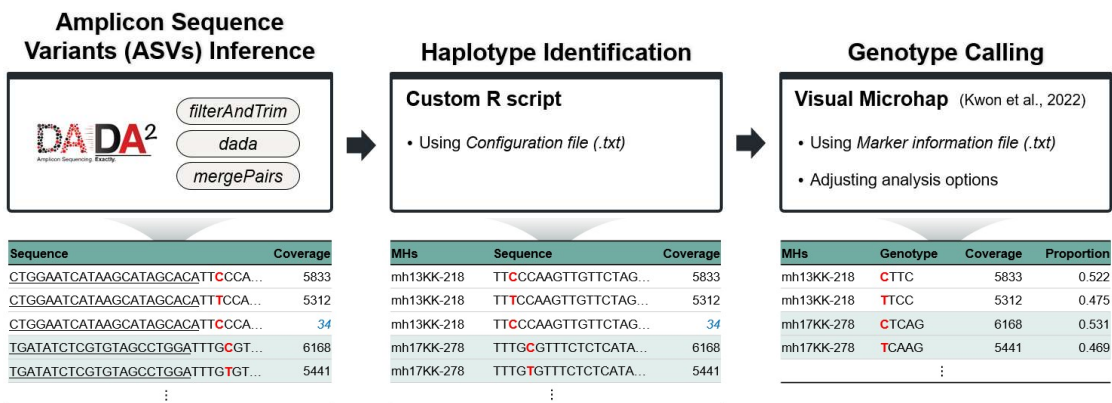
Research Objectives



Denoising Microhaplotype MPS data using computational correction



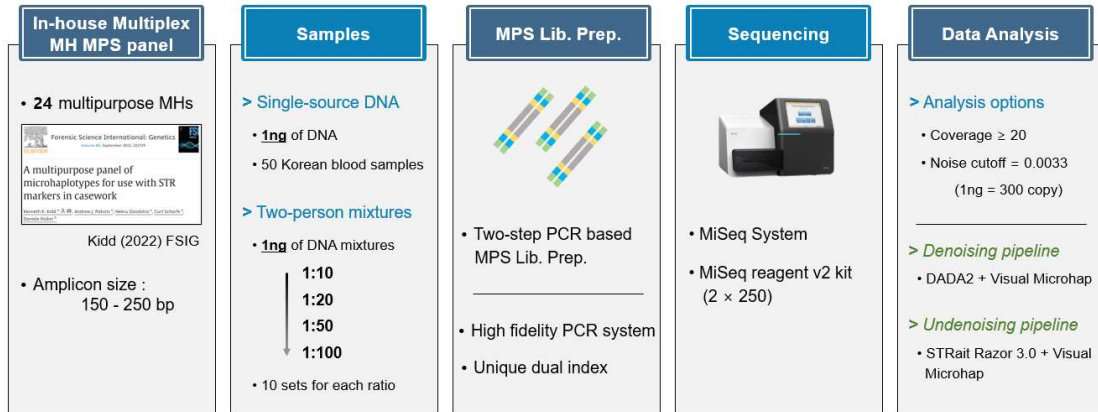
Establishing Denoising Pipeline of MH MPS Data



(Manuscript in preparation)



Multiplex MH MPS Analysis Workflow Overview

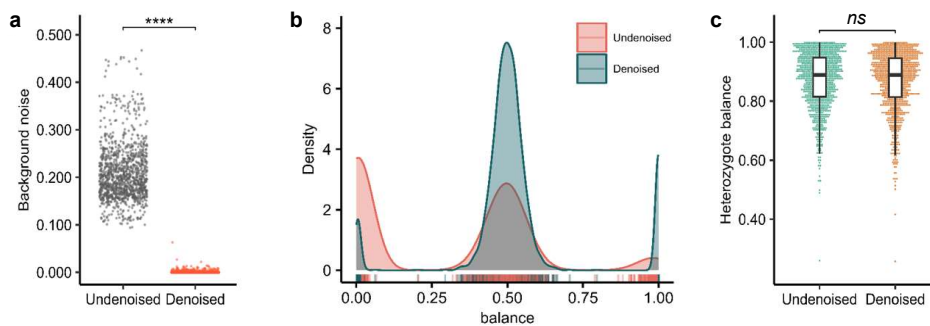


* Approved by the IRB of Severance Hospital, Yonsei University, Seoul, Korea (IRB No. 4-2022-1534)

(Manuscript in preparation)



Evaluating Performance on Single-Source DNA



- Denoised method : **significant reduction in noise**

removes only the noise haplotypes without affecting the true haplotypes

(Manuscript in preparation)

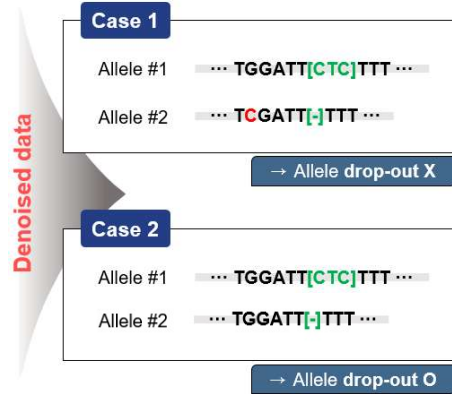


Evaluating Performance on Single-Source DNA

❖ Genotype concordance btw. *undenoisised* and *denoised* data : **99.5 %** (= 1194/1200)

- 6 genotype discordance

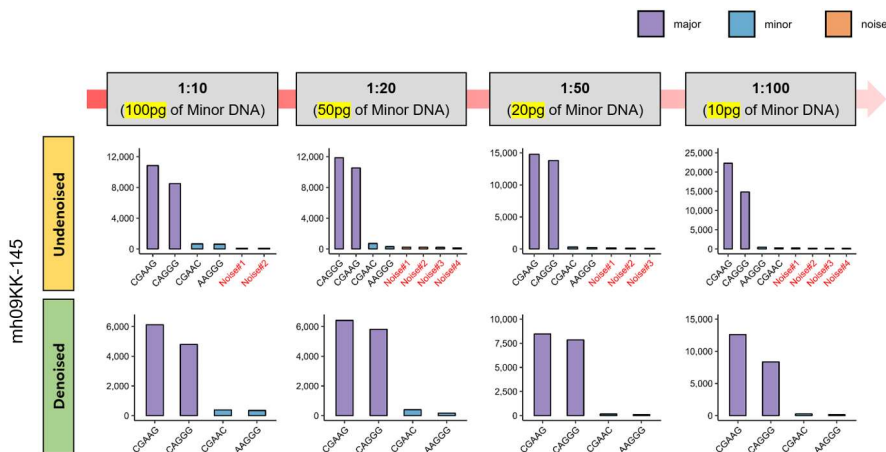
#	MHs	Genotype with cryptic variation	
		Undenoised	Denoised
1	mh02KK-022	AGGAA	AGGAA
		AGGAA 3172897-[CTC]	
2	mh02KK-029	GCCATA	GCCATA
		GCCATA 69138981+[C]@	
3	mh07KK-009	CAGGAT 18861141+[ATGGTGG]	CAGGAT 18861141+[ATGGTGG]
		CGGGAT	
4	mh09KK-153	TGAC	TGAC
		TGAC 103969780+[T]@	
5	mh11KK-183	ATC-T	ATC-T
		ATCGT	
6	mh16KK-011	CCTCACA 84285885+[G]	CCTCACA 84285885+[G]
		CCTCACA	



(Manuscript in preparation)



Application to Two-Person DNA Mixtures



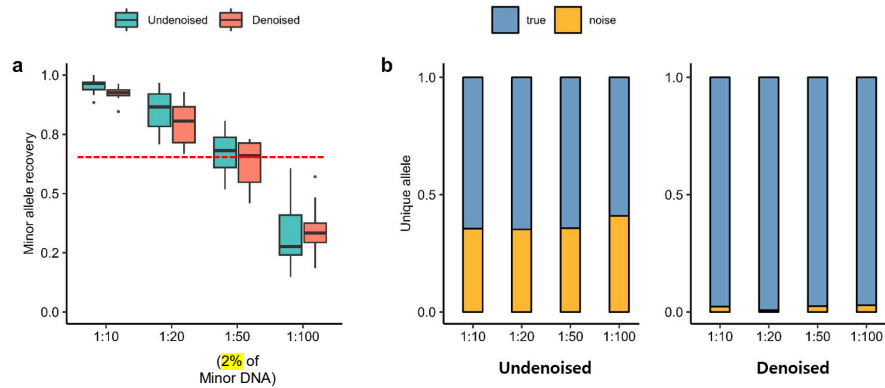
- Identifying the minor allele profile
- Estimating the exact number of contributors



(Manuscript in preparation)



Application to Two-Person DNA Mixtures



- 1:50 mixture (2% of minor DNA) : > 60% of the minor alleles were recovered
- Remarkably low proportion of unique noise even in the 1:100 mixture (1% of minor DNA)

(Manuscript in preparation)

Concluding remarks

- Successfully established a denoising pipeline for MH MPS data using **DADA2** and **Visual Microhap**.

❖ Single-source DNA

- **MPS noise levels** were **significantly reduced** in the **denoised method** compared to the undenoised method.
- The **genotype concordance** between undenoised and denoised data was **99.5 %**.
- **DADA2 has limitations** in handling a **single indel** when **all other sequences** except the indel **were identical**.

❖ Two-person DNA mixtures

- In the **1:50 mixture**, > **60% of the minor alleles** were successfully recovered in the denoised method.
- The **proportion of unique noise alleles** was **significantly lower** in the denoised method, even in the **1:100 mixture**.
→ particularly enhance the **estimation of the exact number of contributors!!**

Thank you !



The Yonsei Profiling Group

Prof. **Kyoung-Jin Shin**

Dr. **Ye-Lim Kwon**
(yelim117@yuhs.ac)

MSc. **Sumin Joo**

Jiwon Kim



This study was supported by the **Korean National Police Agency** (Project Number: PR10-01-000-21) and a faculty research grant of **Yonsei University College of Medicine (2024)**, Republic of Korea.



Conflicts of Interest: None declared.