A pilot study of microhaplotype analysis for degraded DNA and mixed DNA using in-house next generation sequencing panels

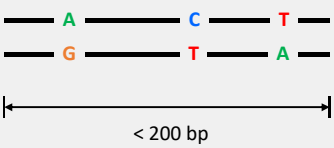
Kyoung-Jin Shin, Ph.D.

Dept. of Forensic Medicine

Yonsei University College of Medicine

Seoul, Republic of Korea

➢ Microhaplotypes



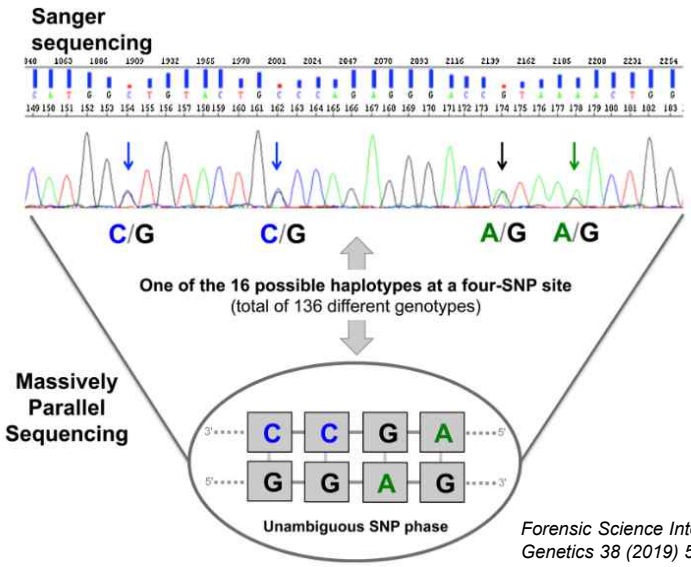
- Multi-allelic marker
- High heterozygosity
- ID and/or ancestry

❖ NGS (Next Generation Sequencing)

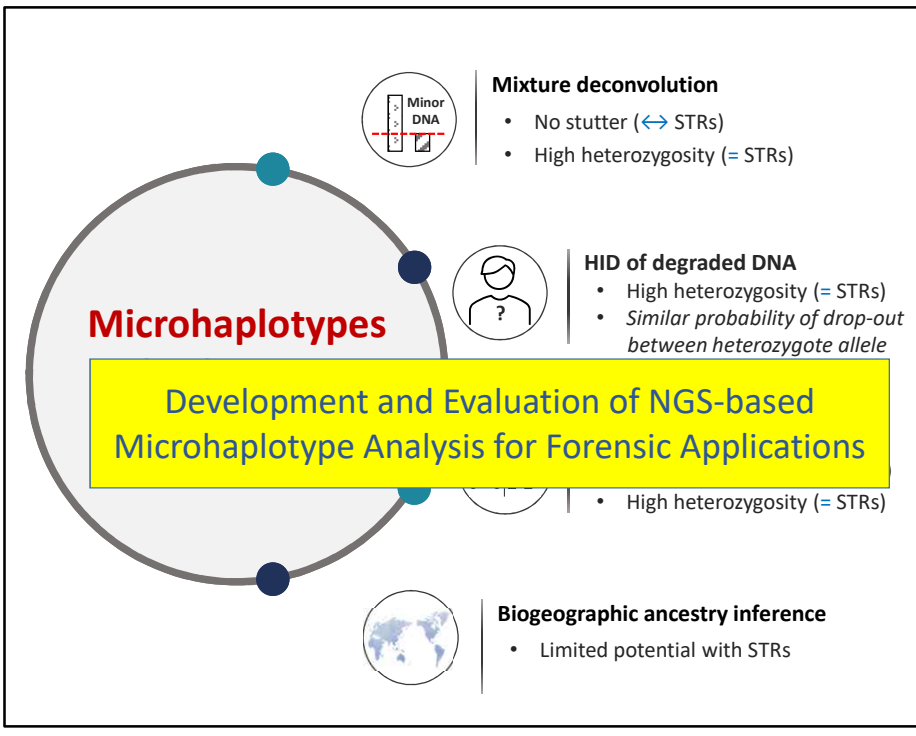
- High-level multiplexing
- Phase-known haplotyping

Kidd *et al.*, FSIG, 2014

Phase-known haplotyping of microhaplotype marker



Forensic Science International: Genetics 38 (2019) 54–69



Materials & Methods

❖ **DNA samples in duplicate**

- 1ng of Mixed DNAs with ratio of 1:1, 1:3, 1:6, 1:9, 1:19, 1:29, 1:49 and 1:99
- A total of 20 degraded DNAs from more than 50-years-old skeletal remains
  - by using total demineralization and silica column method
- 100pg, 50pg and 33pg of diluted 2800M control DNA

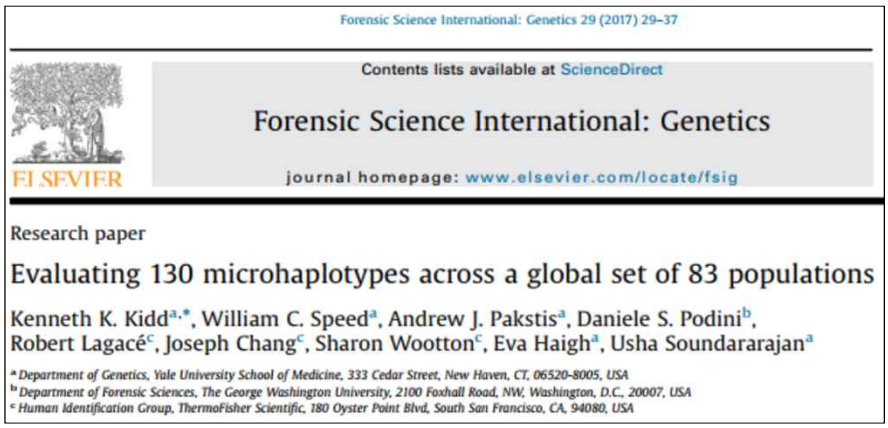
❖ **Molecular characteristics of DNA samples**

- Assessed by the Quantifiler® Trio DNA Quantification kit (Thermo Fisher Scientific)

❖ **Capillary electrophoresis-based STR genotyping**

- The PowerPlex® Fusion System for autosomal STR genotyping of Mixed DNAs
- LCN 2800M control DNA and 2μl of degraded DNA from old skeletal remains
- Analyzed with an AB 3130 Genetic Analyzer and GeneMapper®ID v3.2 software

Methods



Forensic Science International: Genetics 29 (2017) 29–37

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Research paper

Evaluating 130 microhaplotypes across a global set of 83 populations

Kenneth K. Kidd[a,*], William C. Speed[a], Andrew J. Pakstis[a], Daniele S. Podini[b], Robert Lagacé[c], Joseph Chang[c], Sharon Wootton[c], Eva Haigh[a], Usha Soundararajan[a]

[a] Department of Genetics, Yale University School of Medicine, 333 Cedar Street, New Haven, CT, 06520-8005, USA
[b] Department of Forensic Sciences, The George Washington University, 2100 Foxhall Road, NW, Washington, D.C., 20007, USA
[c] Human Identification Group, ThermoFisher Scientific, 180 Oyster Point Blvd, South San Francisco, CA, 94080, USA

➢ **Criteria for microhaplotype selection**

- Ae rank : 1 ~ 51, In rank : 1 ~ 25

32 MH markers for Degraded DNA

Group A	Marker	Primer	Target
1	COG2	F190/R340	151
2	ITGB6	F201/R358	158
3	D18S1122	F222/R390	169
4	GFI1B	F230/R424	195
5	D21S1263	F271/R468	198
6	D5S1970	F254/R463	210
7	LOC642852	F216/R435	220
8	COL4A1	F206/R431	226
9	IGSF21	F208/R434	227
10	RXRA	F231/R463	233
11	SGCG	F212/R447	236
12	LINC0111	F280/R519	240
13	LRRN2	F227/R472	246
14	CPNE4	F222/R473	247/252

Group B	Marker	Primer	Target
1	GNGT2	F158/R314	157
2	COL4A3	F216/R382	167
3	SUDS3	F215/R394	180
4	D13S169	F277/R469	193
5	PLCG2	F212/R416	205
6	D22S1159	F220/R435	216
7	KIF16B	F227/R447	221
8	ADH7	F213/R437	225
9	C14ORF43	F215/R444	230
10	FAM99A	F282/R517	236
11	FRMD4A	F205/R443	239
12	OR52S1P	F160/R402	243
13	ARHGAP27	F222/R470	249
14	LRRC63	F257/R510	254

+ 4 additional MH markers : USH2A, LINC01233, EDAR and CEP104

56 MH markers for Mixed DNA

Group C	Marker	Primer	Target
1	KLK5	F221/R335	115
2	USH2A	F167/R290	124
3	D13S1320	F226/R379	154
4	SEMA6D	F224/R395	172
5	MYO5C	F215/R392	178
6	TOM1L1	F215/R404	190
7	HERC1	F228/R422	195
8	DRD2NCAM	F219/R421	195
9	ELK2B	F225/R432	208
10	FRMD3	F193/R412	220
11	CEBPB	F220/R443	224
12	LINC01233	F148/R384	237
13	STATP1	F221/R472	252
14	RBFOX1-1	F201/R455	255

Group D	Marker	Primer	Target
1	NELFA	F203/R324	122
2	ZC3H7B	F213/R357	145
3	EDAR	F230/R395	166
4	KANK1	F172/R347	176
5	RBFOX1	F228/R416	189
6	PFKP	F199/R391	193
7	LPPR1	F228/R421	194
8	CYYR1	F228/R431	204
9	HRH4	F202/R416	215
10	LOC28716	F225/R451	227
11	D12S290	F222/R453	232
12	TENM4	F222/R458	237
13	CNTN5	F228/R481	254
14	CEP104	F230/R492	263

❖ Two-step PCRs for Microhyplotype Library Preparation using Degraded DNA

1st PCR Amplification

PCR mixture	Volume
dH_2O	3.0 µl
10 X Gold ST*R Buffer	2.0 µl
5 X Primer Mix¥	12.0 µl
AmpliTaq Gold (5U/µl)	1.1 µl
Template DNA*	2.0 µl
Fill up to with dH_2O	20.0 µl

Thermal Cycling		
95℃	11 min	
94℃	20 sec	
59℃	60 sec	**X 29 cycles**
72℃	45 sec	
72℃	5 min	
4℃	Soak	

2nd PCR Amplification

PCR mixture	Volume
dH_2O	3.5 µl
10 X Gold ST*R Buffer	2.0 µl
Index 1 (i7)	2.0 µl
Index 2 (i5)	2.0 µl
AmpliTaq Gold (5U/µl)	0.5 µl
Purified 1st PCR product	10.0 µl
Fill up to with dH_2O	20.0 µl

Thermal Cycling		
95℃	15 min	
94℃	20 sec	
61℃	30 sec	**X 12 cycles**
72℃	45 sec	
72℃	5 min	
4℃	Soak	

* Template; 2µl of degraded DNA from old skeletal remain DNA

Simple Workflow of MPS on MiSeq system

Step 1. PCR amplification

- **Template DNA**
  ; 1 ng DNA samples
- **Amplicon 1/10 dilution or Bead purified amplicon**

Step 2. Library preparation

- **Nextera XT Index kit (Illumina)**
- **Agilent BioAnalyzer**

Step 3. Library pooling, purification and QC

- **Library pooling with equal amount (10ng/ul)**
- **Beads purification**
  ; X1.1 ~ 1.2X beads ratio to remove non specific amplicons
- **Library Quantification**
  ; KAPA library quantification kit for Illumina platforms

Step 4. Sequencing

- **Cluster generation and sequencing on a MiSeq**

Methods

➢ **NGS run**

- Library pooling; final conc. to 10 nM
- NGS run on an MiSeq system (Illumina)
- MiSeq Reagent Kit v3, 600 Cycles (2x300 bp)

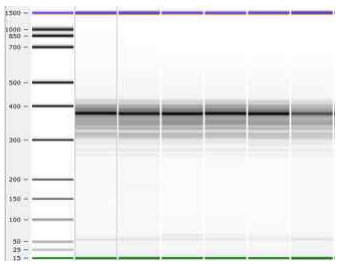
➢ **Data analysis**

- STRait Razor v3.0 and Microsoft Excel

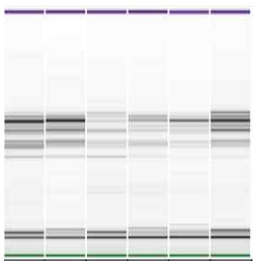
Results

➢ **NGS Library QA using Bioanalyzer 1000 chip**

Mixed DNA (1:1 ~ 1:29)



Degraded DNA



Examples of NGS data from Mixed DNAs

Read > 50

3:1 mixture

```
COG2:0  112 bases TGGTATGAAGTACCTATTAAACGTTATTTCTGAATGCTATATGTATTTGATGTTTATCCAAACACCTGGGAGATAGTGTCATGTAAAATTGTGCGTGGCATGAAGGATGTGG 9364  0
COG2:0  112 bases TGGTATGAAGTACCTATTAAACGTTATTTCTGAATGCTATATATATTTGATGTTTATCCAAACACCTGGTAGATAGTGTCATGTAAAATTGTGCGTGGCATGAAGGATGTGG 870  0
COG2:-5  107 bases TGGTATGAAGTACCTATTAAACGTTATTTCTGAATGCTATATGTATTTGATGCCAAACACCTGGGAGATAGTGTCATGTAAAATTGTGCGTGGCATGAAGGATGTGG  50  0
COG2:0.0 0 bases  SumBelowThreshold 2044  0
ITGB6:0  118 bases ACCCTCTCTACCTAAGGATGGGCAATGGCTTATGAGTGAGAAACATGGAGCCGTGGGAACTCAGAATGACATGCTACCTGGAGATTGTGGTAACGCCCTGTTTTTTTGTGGGCATATC  4639  0
ITGB6:0  118 bases ACCCTCACTACCTAAGGATGGGCAATGGCTCATGAGTGAGAAACATGGAGCCGTGGGAACTCAGAATGACATGCTACCTGGAGATTGTGGTAACGCCCTGTTTTTTTGTGGGCATATC  4421  0
ITGB6:0  118 bases ACCCTCACTACCTAAGGATGGGCAATGGCTTATGAGTGAGAAACACGGAGCCGTGGGAACTCAGAATGACATGCTACCTGGAGATTGTGGTAACGCCCTGTTTTTTTGTAGGCATATC  1109  0
ITGB6:0  118 bases ACCCTCACTACCTAAGGATGGGCAATGGCTCATGAGTGAGAAACATGGAGCCGTGGGAACTCAGAATGACATGCTACCTGGAGATTGTGGTAACGCCCTGTTTTTTTGTGGACATATC  59  0
ITGB6:0.0 0 bases  SumBelowThreshold 2506  0
D18S1122:0  127 bases GAACTGGAGAGCAGGTGGATTAAATCTGGGGGGTGACTCCAGCACATCTCTAATGAACACTTCTTAACATTTAATTTCAAAGGGCCTGGTGACCCTCGGATGTGTCAGTGTGGAGAAGAGATAGGTA  7319  0
D18S1122:0  127 bases GAACTGGAGAGCAGGTGGATTAAATCTGGGTGGTGACTCCAGCACATCTCTAATGAACACTTCTTAACGTTTAATTTCAAAGGGCCTGGTGACCCTCGGATGTGTCAGTGTGGAGAAGAGATAGGTA  2155  0
D18S1122:0  127 bases GAACTGGAGAGCAGGTGGATTAAATCTGGGGGGTGACTCCAGCACATCTCTAATGAACACTTCTTAACATTTAATTTCAAAGGGCCTGGTGATCCTCGGATGTGTCAGTGTGGAGAAGAGATAGGTA  62  0
D18S1122:0.0  0 bases  SumBelowThreshold 2093  0
```

6:1 mixture

```
COG2:0  112 bases TGGTATGAAGTACCTATTAAACGTTATTTCTGAATGCTATATGTATTTGATGTTTATCCAAACACCTGGGAGATAGTGTCATGTAAAATTGTGCGTGGCATGAAGGATGTGG 8280  0
COG2:0  112 bases TGGTATGAAGTACCTATTAAACGTTATTTCTGAATGCTATATATATTTGATGTTTATCCAAACACCTGGTAGATAGTGTCATGTAAAATTGTGCGTGGCATGAAGGATGTGG 367  0
COG2:0.0 0 bases  SumBelowThreshold 1601  0
ITGB6:0  118 bases ACCCTCACTACCTAAGGATGGGCAATGGCTCATGAGTGAGAAACATGGAGCCGTGGGAACTCAGAATGACATGCTACCTGGAGATTGTGGTAACGCCCTGTTTTTTTGTGGGCATATC  3561  0
ITGB6:0  118 bases ACCCTCTCTACCTAAGGATGGGCAATGGCTTATGAGTGAGAAACATGGAGCCGTGGGAACTCAGAATGACATGCTACCTGGAGATTGTGGTAACGCCCTGTTTTTTTGTGGGCATATC  3294  0
ITGB6:0  118 bases ACCCTCACTACCTAAGGATGGGCAATGGCTTATGAGTGAGAAACACGGAGCCGTGGGAACTCAGAATGACATGCTACCTGGAGATTGTGGTAACGCCCTGTTTTTTTGTAGGCATATC  392  0
ITGB6:0.0 0 bases  SumBelowThreshold 1779  0
D18S1122:0  127 bases GAACTGGAGAGCAGGTGGATTAAATCTGGGGGGTGACTCCAGCACATCTCTAATGAACACTTCTTAACATTTAATTTCAAAGGGCCTGGTGACCCTCGGATGTGTCAGTGTGGAGAAGAGATAGGTA  5324  0
D18S1122:0  127 bases GAACTGGAGAGCAGGTGGATTAAATCTGGGTGGTGACTCCAGCACATCTCTAATGAACACTTCTTAACGTTTAATTTCAAAGGGCCTGGTGACCCTCGGATGTGTCAGTGTGGAGAAGAGATAGGTA  958  0
D18S1122:0  127 bases GAACTGGAGAGCAGGTGGATTAAATCTGGGGGGTGACTCTAGCACATCTCTAATGAACACTTCTTAACATTTAATTTCAAAGGGCCTGGTGACCCTCGGATGTGTCAGTGTGGAGAAGAGATAGGTA  59  0
D18S1122:0.0  0 bases  SumBelowThreshold 1307  0
```

9:1 mixture

```
COG2:0  112 bases TGGTATGAAGTACCTATTAAACGTTATTTCTGAATGCTATATGTATTTGATGTTTATCCAAACACCTGGGAGATAGTGTCATGTAAAATTGTGCGTGGCATGAAGGATGTGG 8575  0
COG2:0  112 bases TGGTATGAAGTACCTATTAAACGTTATTTCTGAATGCTATATATATTTGATGTTTATCCAAACACCTGGTAGATAGTGTCATGTAAAATTGTGCGTGGCATGAAGGATGTGG 159  0
COG2:0.0 0 bases  SumBelowThreshold 1699  0
ITGB6:0  118 bases ACCCTCTCTACCTAAGGATGGGCAATGGCTTATGAGTGAGAAACATGGAGCCGTGGGAACTCAGAATGACATGCTACCTGGAGATTGTGGTAACGCCCTGTTTTTTTGTGGGCATATC  4525  0
ITGB6:0  118 bases ACCCTCACTACCTAAGGATGGGCAATGGCTCATGAGTGAGAAACATGGAGCCGTGGGAACTCAGAATGACATGCTACCTGGAGATTGTGGTAACGCCCTGTTTTTTTGTGGGCATATC  3587  0
ITGB6:0  118 bases ACCCTCACTACCTAAGGATGGGCAATGGCTTATGAGTGAGAAACACGGAGCCGTGGGAACTCAGAATGACATGCTACCTGGAGATTGTGGTAACGCCCTGTTTTTTTGTAGGCATATC  223  0
ITGB6:0.0 0 bases  SumBelowThreshold 2068  0
D18S1122:0  127 bases GAACTGGAGAGCAGGTGGATTAAATCTGGGGGGTGACTCCAGCACATCTCTAATGAACACTTCTTAACATTTAATTTCAAAGGGCCTGGTGACCCTCGGATGTGTCAGTGTGGAGAAGAGATAGGTA  8087  0
D18S1122:0  127 bases GAACTGGAGAGCAGGTGGATTAAATCTGGGTGGTGACTCCAGCACATCTCTAATGAACACTTCTTAACGTTTAATTTCAAAGGGCCTGGTGACCCTCGGATGTGTCAGTGTGGAGAAGAGATAGGTA  543  0
D18S1122:0.0  0 bases  SumBelowThreshold 1954  0
```

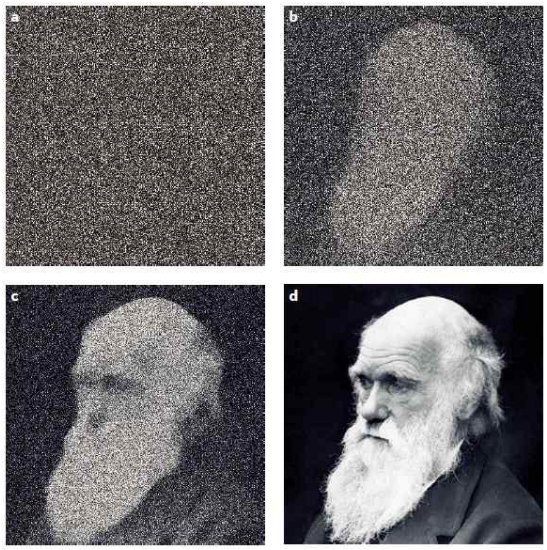
A Result from Degraded DNA

Read > 100

Marker	Old Skeletal Remain		
	1st	2nd	# of matched allele
COG2	a, b, c	a, b, c, d	**1**
ITGB6	a, b	a, b	**0**
D18S1122	-	-	**0**
GFI1B	-	a, b	**0**
D21S1263	1	-	**0**
D5S1970	1	a, b	**1**
LOC642852	1	-	**0**
COL4A1	1	a, b	**1**
IGSF21	-	-	**0**
RXRA	-	1	**0**
SGCG	-	-	**0**
LINC0111	-	1	**0**
LRRN2	1	1	**0**
CPNE4	1	-	**0**
GNGT2	a, b	a, b	**1**
COL4A3	1	a, b	**0**
SUDS3	1	-	**0**
D13S169	a, b	-	**0**
PLCG2	1	-	**0**
D22S1159	-	1	**0**
KIF16B	1	-	**0**
ADH7	1	1	**1**
C14ORF43	-	a, b, c	**0**
FAM99A	1	a, b	**1**
FRMD4A	-	1	**0**
OR52S1P	-	1	**0**
ARHGAP27	-	1	**0**
LRRC63	a, b	a, b	**0**
USH2A	1	1	**1**
LINC01233	a, b, c	1	**1**
EDAR	1	1	**1**
CEP104	-	-	**0**
# of typed marker	20	21	9

```
COG2:0  112 bases TGGTATGAAGTACCTATTAAACGTTATTTCTGA … GGATGTGG 3540  0
COG2:0  112 bases TGGTATGAAGTACCTATTAAACGTTATTTCTGA … AGGATGTGG 1002  0
COG2:0  112 bases TGGTATGAAGTACCTATTAAATGTTATTTCTGA … GGATGTGG 614  0
COG2:0  112 bases TGGTATGAAGTACCTATTAAACGTTATTTCTGA … GGATGTGG  558  0
COG2:0.0 0 bases  SumBelowThreshold 696  0
ITGB6:0  118 bases ACCCTCACTACCTAAGGATGGGCAATGGCTTA … TGTTTTTTTGTGGGCATATC  1841  0
ITGB6:0  118 bases ACCCTCTCTACCTAAGGATGGGCAATAGCTTA … GTTTTTTTGTGGGCATATC  945  0
ITGB6:0.0 0 bases  SumBelowThreshold 415  0
D18S1122:0  127 bases GAACTGGAGAGCAGGTGGATTAAA … CTCGGATGTGTCAGTGTGGAGAAGAGATAGGTA  978  0
D18S1122:0  127 bases GAACTGGAGAGCAGGTGGATTAAA … CCTCGGATGTGTCAGTGTGGAGAAGAGATAGGTA  905  0
D18S1122:0  127 bases AAACTGGAGAGCAGGTGGATTAAA … CTCGAATGTATCAGTGTGGAGAAGAGATAGGTA  895  0
D18S1122:0.0  0 bases  SumBelowThreshold 414  0
GNGT2:0  117 bases GACCCGCTGGAGGAGCAGAAGTTAAGGGAC … AGAGAGGAGGCCCGCCGT  1700  0
GNGT2:0  117 bases GACCCGCTGGAGGAGCAAAAGTTAAGGGAC … AGAGAGGAGGCCCGCCGT  1010  0
GNGT2:0  117 bases GACCCGCTGGAGGAGCAAAAGTTAAGGGAC … AGAGAGGAGGCCCGCCAT  908  0
GNGT2:0  117 bases GACCCGCTGGAGGAGCAAAAGTTAAGGGAT … AGAGAGGAGGCCCGCCGT  805  0
GNGT2:0  117 bases GATCCGCTGGAGGAGCAGAAGTTAAGGGATT … AGAGAGGAGGCCCGCCGT  597  0
GNGT2:0  117 bases GACCCACTGGAGGAGCAAAAGTTAAGGGAC … AGAGAGGAGGCCCGCCGT  290  0
GNGT2:0.0  0 bases  SumBelowThreshold 634  0
COL4A3:0 126 bases TAGGTTTCTCATCTAGTAGATGGAGCTTATAAT … CAAAACATACTGAGATCATTACAG 3990  0
COL4A3:0 126 bases TAGGTTTCTCATCTATTAGATGGAGCTTATAAT … AAAACATACTGAGATCATTACAG 1989  0
COL4A3:0 126 bases TAGGTTTTTCATCTAGTAGATGGAGCTTATAAT … CAAAACATACTGAGATCATTACAG 1253  0
COL4A3:0 126 bases TAGGTTTCTCATCTATTAGATGGAGCTTATAAT … AAAACATATTGAGATCATTACAG 802  0
COL4A3:0 126 bases TAGGTTTCTCATCTAATAGATGGAGCTTATAAT … CAAAACATACTGAAATCATTACAG 209  0
COL4A3:0 126 bases TAGGTTTCTCATCTAGTAGATGGAGCTTATAAT … CAAAACATACTGAGATCATTATAG 100  0
COL4A3:0.0  0 bases  SumBelowThreshold 1166  0
SUDS3:0  137 bases TCATTTAAAAATCTGGTAGCCTGGTGGAAGGG … GAAATAAAGGTTCAGAAAAAAGCTATACCTTTACAA  828  0
SUDS3:0  137 bases TCATTTAAAAATCTGGTAGCCTGGTGGAAGGG … AAATAAAGGTTCAGAAAAAAGCTATACCTTTATAA  618  0
SUDS3:0.0  0 bases  SumBelowThreshold 213  0
```

The signal-to-noise problem

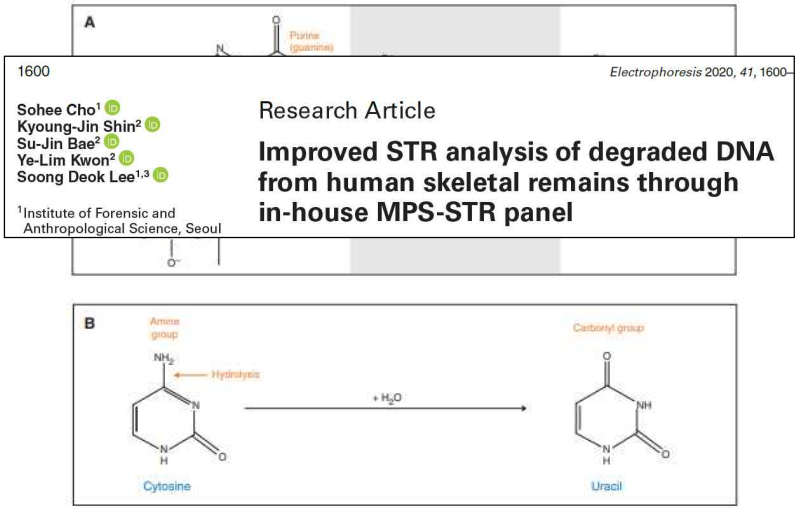


Nature Reviews Genetics 2018, 269–285.

Causes of Noise in NGS data

- DNA Damage
- PCR and Sequencing Error
- NGS Multiplexing Error

DNA Damage

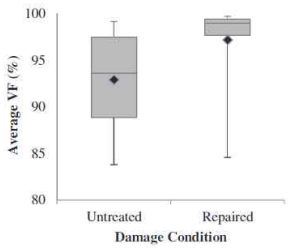


Cold Spring Harb Perspect Biol doi: 10.1101/cshperspect.a012567

Method to Reduce Noise in MPS (1)

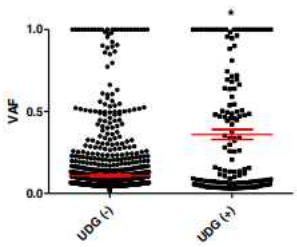
- DNA Repair Enzyme

  Forensic Science International: Genetics 2018, 257-264



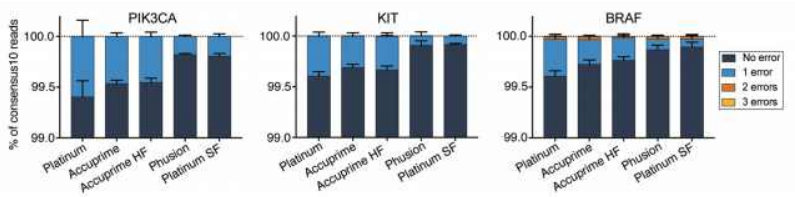
- UDG Treatment

  Applied Cancer Research 2019, Article number: 7



Method to Reduce Noise in MPS (2)

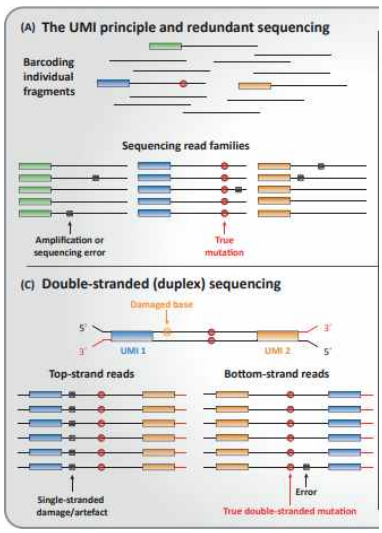
- Use High Fidelity PCR Enzyme



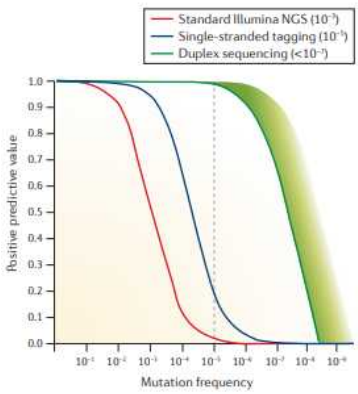
Scientific Reports 2019, Article number: 3503

- Adoption of Unique Molecular Identifiers (UMIs)

Detecting Rare Mutations and DNA Damage

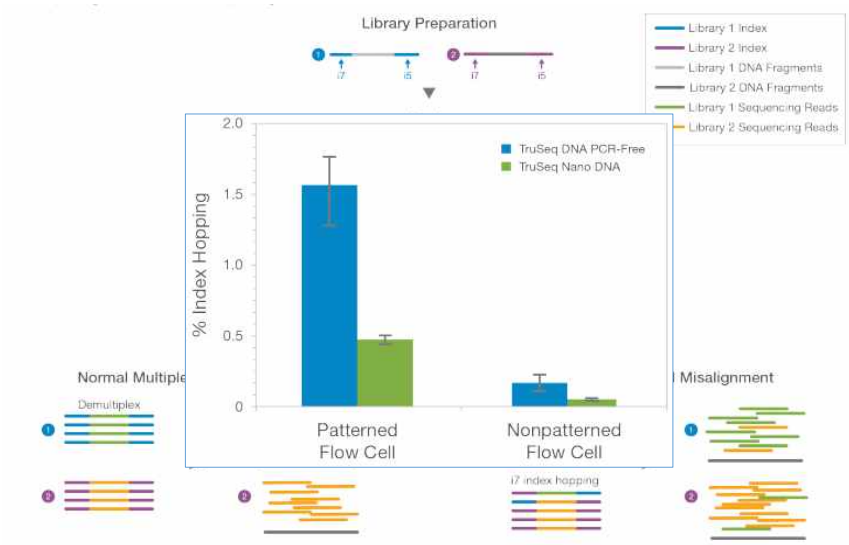


Trends in Biotechnology, 2018, 729-740.



Nature Reviews Genetics 2018, 269–285.

Overview of Indexing Hopping



https://www.illumina.com

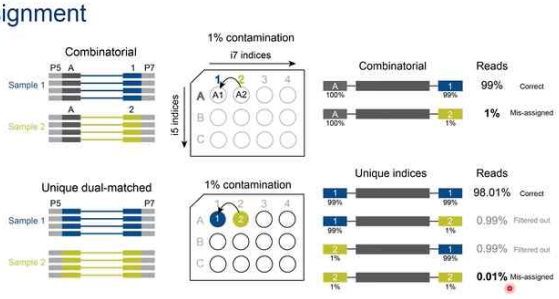
Tips and Best Practices

To minimize the level and effect of index hopping, customers should follow these library preparation best practices:

- Remove free adapters from library preps
- Store libraries individually at -20°C
- Pool libraries prior to sequencing
- Use unique dual indexing pooling combinations (unique i5 and i7 indexes)

https://www.illumina.com

Unique, dual matched indices reduce contamination mis-assignment



https://www.idtdna.com

Suggestion to Minimize Noise in MPS for Analysis of Forensic Challenging Samples

- Establish Experimental Procedure
  - Pre-treatment of Damaged DNA
  - Use High Fidelity PCR Enzyme for Library Preparation
  - Adoption of Unique Molecular Identifiers
  - Use Unique Dual Indexing for Multiple Samples
- Need Bioinformatic Pipeline for Forensic Applications

Thank you for your attention!

kjshin@yuhs.ac
http://forensic.yonsei.ac.kr