



Sequence Variations of 51 Microhaplotypes in Koreans Detected by Massively Parallel Sequencing

Eun Young Lee · So Yeun Kwon · Jae Ryul Han · Hwan Young Lee · Kyoung-Jin Shin

Department of Forensic Medicine and Brain Korea 21 PLUS Project for Medical Science,

Yonsei University College of Medicine, Seoul, Korea

Introduction

Microhaplotypes (MHs) provide higher genetic diversity than conventional diallelic single nucleotide polymorphisms (SNPs) as well as useful information on ancestry. Microhaplotype loci are promising genetic markers that can aid in mixture deconvolution in forensics while avoiding the issue of stutter from short tandem repeat (STR) markers. To date, microhaplotypes have been studied the most by the Kidd group at Yale University using TaqMan assays. However current massively parallel sequencing (MPS) technology enables loci consisting of 2 or more SNPs within the single-strand sequence reads to be phase-known, thereby facilitating haplotype allele calling at the locus. Here we investigated the sequence variations of 51 microhaplotypes in a Korean population using the sequence data which were generated by an in-house assay designed for MPS.

Results

Table 1. SNPs and No. of haplotypes observed in 51 MHs A total of 196 SNPs and 316 haplotype alleles were observed in 122 Koreans. The SNPs commented by Kidd et al. were written in black and additional variations were written in blue.

Marker	Observed SNPs	SNPs	Haplotypes
ACN9	rs17168174/rs10246622	2	3
ADH7	rs4699748/rs7692081/rs2584461/rs1442492	4	4
ARHGAP27	rs1059504/rs8327	2	4
ATXN1	rs4565296/rs143253642/rs4431439/rs179939/rs79542015	5	5
C14ORF43	rs12717560/rs76446474/chr14:73783873/chr14:73783947/rs12878166/rs12879393	6	9
CCR2	rs4513489/rs6441961	2	3
CDH4	rs10854214/rs10854215	2	3
CEP104	rs4648344/rs114208229/rs376407444/rs6663840	4	5
COG2	rs751171217/rs2296796/rs2296797/rs2296798	4	5
COL4A1	rs1192203/rs1192204/rs1192205/rs3825483/rs3825482/rs3825481	6	6
COL4A3	rs6714835/rs6756898/rs12617010/rs188867039	4	7
COMT	rs4818/rs4680	2	3
CPNE4	rs1225051/chr3:131927154/rs1225050/rs1225049/rs1225048/rs62889461	6	5
D13S169	rs1927847/rs9536429/rs7492234/rs9536430	4	10
D18S1122	rs621320/rs621340/rs678179/rs80093367/rs621766/rs75620451	6	11
D21S1263	rs8126597/rs138895664/rs6517970/rs8131148/rs6517971	5	13
D22S1159	rs763040/rs5764924/rs763041	3	4
D5S1970	rs74865590/rs438055/rs148080283/chr5:2447958/rs370672/rs6555108	6	17
DLEU1	rs806301/rs2066700	2	3
DRD3	rs3732783/rs6280	2	3
EDAR	rs260694/rs11123719/rs11691107	3	4
ESRRG	rs4528199/rs6604596	2	4
FAM99A	rs12802112/rs12360952/rs28631755/rs137953460/rs7112918/rs4752777	6	16
FAT1	rs1280100/rs74986565/rs1280099	3	5
FRMD4A	rs10796164/rs10796165/rs17154765/rs10796166	4	12
GATA4	rs1390950/rs2898295	2	3
GFI1B	rs606141/rs8193001/rs56256724/rs2073578/rs633153	5	6
GNGT2	rs2233360/rs2233361/rs2233362/rs634370	4	4
IGSF21	rs11810587/rs1336130/rs1336131/rs1533623/rs1533622	5	6
ITGB6	rs12469721/rs3101043/rs3111398/rs72623112	4	10
KIF16B	rs6044080/rs17674942/rs532974468/rs6044081/rs16997830	5	6
LINC00111	rs2838081/rs2838082/rs78902658/rs2838083	4	7
LINC01233	rs4932999/rs4932769/rs2361019/rs2860462	4	5
LOC642852	rs6518223/rs7282557/rs2838868/rs7279250/rs8133697	5	8
LRRC2	rs34504121/rs6808142/rs17030627	3	4
LRRC63	rs9562648/rs9534373/rs9562649/rs60234592/chr13:46291986/rs2765614	6	12
LRRN2	rs17413714/rs2772234/rs1610401/rs1610400	4	12
LYPD6B	rs2170607/rs2377509/rs10497052/chr2:149097752	4	4
NCAM1	rs2303377/rs781803207/rs2303378	3	4
NPEPPS	rs3760370/rs3760371	2	3
OR52S1P1	rs2499935/rs10500616/rs2499936	3	4
PAH	rs2247836/rs2133298/rs3817446	3	3
PAPD7	rs870348/rs870347/rs870346	3	4
PLCG2	rs16956011/rs3934954/rs3934955/rs3934956/rs62044947/rs3934957/rs4073828	7	12
PLIN3	rs2271058/rs1055919/rs200918705/rs541039172/rs2271057	5	7
RXRA	rs3118582/rs76875728/rs7037930/rs10776839	4	4
SGCG	rs8181845/rs8181836/rs2152726/rs2152727/rs679482/rs9510616/rs34008986	7	10
SUDS3	rs145955605/rs1503767/rs11068953	3	5
TAS2R1	rs41461/rs41462	2	3
TTC12	rs2288159/rs10891537	2	3
TYRP1	rs1408800/rs1408801	2	3

Materials and Methods

DNA samples and target markers A total of 122 DNA samples from a Korean population were tested in this study. The 51 microhaplotypes were selected with reference to the reports by Kidd et al. [1,2].

Amplification of 51 MHs, library preparation, and NGS analysis A multiplex PCR system was constructed to amplify 51 microhaplotypes simultaneously, and each primer pair was designed to produce amplicons with the size range of 96 to 270 base pairs. The barcoded library for each sample was prepared by a PCR-based library preparation method [3]. The pooled library were analyzed on the MiSeq® System using a MiSeq Reagent Kit v3 (2x300 cycles) (Illumina, Inc., San Diego, CA, USA). The observed SNPs were annotated with frequencies, and the alleles at the microhaplotype locus were determined using an in-house sequence parsing program.

Table 2. Allele frequencies of 196 SNPs in a Korean population

Example: PLCG2 microhaplotype

SNP_1	Freq.	SNP_2	Freq.	SNP_3	Freq.	SNP_4	Freq.
rs16956011		rs3934954		rs3934955		rs3934956	
G	0.6270	A	0.8238	C	0.6434	T	0.5205
A	0.3730	G	0.1762	A	0.3566	C	0.4795
SNP_5	Freq.	SNP_6	Freq.	SNP_7	Freq.		
rs62044947		rs3934957		rs4073828			
C	0.6434	C	0.6434	G	0.6475		
T	0.3566	G	0.3566	A	0.3525		

Table 3. Observed haplotypes and frequencies of 51 MHs in a Korean population

Example: FAM99A microhaplotype

Observed SNPs	Observed haplotypes	Haplotype freq. (n=122x2)
rs12802112/rs12360952/	GCCCCG	0.2459
rs28631755/rs137953460/	ATCCCC	0.4139
rs7112918/rs4752777	ACCCTG	0.0861
	ACCCTC	0.1189
	ATACTG	0.0041
	ATACCC	0.0205
	ATACTC	0.0574
	ACCTCG	0.0041
	ATCCCG	0.0082
	GTCCCC	0.0041
	ATCCTC	0.0041
	ACCCCG	0.0082
	ACCCCC	0.0082
	GCCTCG	0.0082
	GCCCTC	0.0041
	GCCCCC	0.0041

Table 4. Population genetic statistics for 51 MHs in a Korean population

Marker	H _{obs}	H _{exp}	MP	PD	PIC	PE	TPI
D5S1970	0.893	0.883	0.032	0.968	0.868	0.782	4.69
D13S169	0.869	0.875	0.035	0.965	0.857	0.732	3.81
D21S1263	0.811	0.818	0.060	0.940	0.794	0.621	2.65
LRRC63	0.828	0.811	0.066	0.934	0.784	0.652	2.90
LINC00111	0.811	0.819	0.067	0.933	0.790	0.621	2.65
...							
TYRP1	0.344	0.332	0.492	0.508	0.283	0.083	0.76
EDAR	0.246	0.245	0.592	0.408	0.233	0.044	0.66
ATXN1	0.172	0.206	0.667	0.333	0.200	0.023	0.60

H_{obs}, observed heterozygosity; H_{exp}, expected heterozygosity; MP, matching probability; PD, power of discrimination; PIC, Polymorphism Information content; PE, power of exclusion; TPI, typical paternity index

Conclusion

- By MPS analysis of 51 microhaplotypes, a total of 196 SNPs and 316 haplotype alleles were observed from 122 Koreans.
- Each microhaplotype consisted of 2 to 7 SNPs and exhibited 3 to 17 haplotype alleles, and the D5S1970 marker is the most polymorphic.
- Overall match probability (3.98×10^{39}) of the 51 microhaplotypes was much higher than that of the expanded autosomal STR loci (2.73×10^{25} by 22 STRs of PowerPlex Fusion or 3.49×10^{24} by 21 STRs of GlobalFiler) in Koreans.
- Sequence analysis of microhaplotypes using MPS will be of great practical use in forensic DNA typing.

References

- [1] Kidd KK, Pakstis AJ, Speed WC, et al., Forensic Sci Int Genet. (2014) 12:215-224.
- [2] Kidd lab website (<http://medicine.yale.edu/lab/kidd/publications/publications.aspx>)
- [3] Lee EY, Lee HY, Oh SY, et al., Forensic Sci Int Genet. (2016) 22:37-43.

Acknowledgement

- This research was supported by a fund (No. NFS2016DNA02) from the Forensic Research Program of the National Forensic Service, Korea.