

# Analysis of kinship index distributions in Koreans using simulated autosomal STR profiles

In Seok Yang  
Department of Forensic Medicine  
Yonsei University College of Medicine

© TemplatesWise.com

## Short tandem repeat (STR)

- STR markers are highly variable among individuals, thereby enabling individual identification and kinship testing.
- 13 Combined DNA Index System (CODIS) and 7 European Standard Sets were established for the purpose in the late 1990s and have been used in many countries.
- Recently, the number of the STR loci was expanded to obtain accurate analysis results in various kinship testing or forensic casework.

## Statistical approaches

- individual identification and kinship testing

- Match probability (MP)

MP is used to evaluate two DNA profile match in individual identification.

- Likelihood ratio (LR)

LR compares the probabilities of two alternative hypotheses. Kinship testing is largely based on LR-based approach.

LR: kinship index (KI) in kinship testing  
paternity index in paternity testing  
sibling index in sibling testing.

## Limitation of kinship analysis

- Since a parent and a child share at least an allele in a STR locus, it allows ease discrimination between related and unrelated persons in parent/child relationship.
- However, shared allele can not be observed in full-sibling, uncle/nephew, and first cousin relationship, making it difficult to discriminate between relate and unrelated persons in the relationships.
- Due to limitation of family sample collection, simulation approach has been used to present kinship index data in a particular population.

## Aim of this study

- To provide KI data required for guideline establishment on various kinship testing and familial search in Koreans
  - Generation of KI distributions by simulation with defined relationships (parent/child, full-siblings, uncle/nephew, first-cousins) in Koreans
  - Evaluation of the distributions for discrimination between related and unrelated persons

## Simulation method

### ◆ Related genotype pairs

STR allele frequency

250,000 random genotypes

50,000 virtual pedigrees

KI calculation

### ◆ Unrelated genotype pairs

STR allele frequency

50,000 random genotype pairs

KI calculation

Analysis of KI distribution for each relationship

Simulation was performed by using Microsoft Excel Macro functions written with Visual Basic for Application (VBA) language.

# STR loci and allele frequencies

Two sets of 13 CODIS and 20 STRs (13 CODIS STRs + 7 additional STRs) were used to generate random genotypes by simulation.

Allele frequencies											
										No. of random genotypes :	50000
										Apply minimal allele frequency?	No
STR loci	D5S818	D13S317	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D16S539	VWA	
Sample size	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000	
Mutation rate	0.0006	0.0006	0.00097	0.00209	0.00075	0.00112	0.00045	0.00022	0.0006	0.0015	
Alleles	Frequencies										
5											
6	0.0004										
<7	0.0001										
7	0.0127	0.0011			0.0045	0.0008			0.2585		
8	0.0075	0.2689	0.0004			0.1426	0.001	0.0431	0.0041		
9	0.0872	0.1447	0.0023			0.0581	0.0494	0.4814	0.2929		
9.2											
9.3											
10	0.204	0.1426	0.1143			0.1733	0.2362	0.013	0.1482		
11	0.3117	0.2299	0.0932			0.3408	0.2453	0.0001	0.2555		
11.3											
12	0.2292	0.1642	0.1441			0.2374	0.3861	0.0051	0.1954		
12.2											
13	0.1359	0.0374	0.2304			0.0385	0.0706	0.001	0.0916	0.0004	

- Allele frequencies of 13 CODIS STRs, D2S1338, and D19S433 were offered by Supreme Prosecutors' Office.
- Allele frequencies of 5 STRs (D1S1656, D2S441, D10S1248, D12S391, and D22S1045) were adopted from Park et al., 2013.

# Random genotypes and virtual pedigrees

**Genotypes**

View allele frequencies

Start: 2013-03-19 140309

End: 2013-03-20 44423

ID	D5S818		D13S317		D8S1179		D21S11		D7S820		CSF1PO		D3S1358		TH01		D16S539		VWA		
	Allele 1	Allele 2	Allele 1	Allele 2	Allele 1	Allele 2	Allele 1	Allele 2	Allele 1	Allele 2	Allele 1	Allele 2	Allele 1	Allele 2	Allele 1	Allele 2	Allele 1	Allele 2	Allele 1	Allele 2	
01-1	10	12	8	10	10	12	13.2	13.2	8	12	11	12	15	16	9	9	10	14	17		
01-2	9	11	9	13	13	13	30	31	9	12	10	12	15	17	6	7	9	10	14	17	
01-3	13	13	11	12	11	13	30	31	12	12	12	12	14	15	9	9	10	11	18	19	
01-4	11	12	8	12	10	13	29	30	11	12	11	13	12	17	9	9	11	13	14	17	
01-5	11	11	8	11	13	15						14	15	16	7	7	11	12	16	19	
01-6	11	13	8	11	14	14						12	15	16	9	9	11	12	17	18	
01-7	11	12	8	11	10	13						12	15	16	6	9	11	11	14	17	
01-8	10	10	10	13	13	12						12	15	16	9	9	12	14	16	18	
01-9	10	11	10	12	13	14	13	13.2	8	9	10	12	17	17	7	9	12	12	16	19	
01-10	11	13	8	10	12	15	30	30	8	8	12	12	15	17	7	9.3	9	11	14	17	
01-11	10	12	8	10	14	15	29	31.2	11	12	11	12	15	15	9	9	9	12	15	19	
01-12	13	13	10	12	13	16	28	30	10	12	12	12	13	15	16	9	9	12	15	16	
01-13	12	12	12	12	14	14	29	30	9	11	12	14	15	17	7	9	9	13	18	18	
01-14	11	11	8	12	14	15	29	30	11	11	11	12	15	15	6	7	11	13	18	19	
01-15	10	12	10	10	11	13	29	30	8	10	11	12	15	15	7	9	11	12	16	18	
01-16	9	10	9	11	12	13	12	12.2	11	11	11	12	16	17	6	9.3	10	11	16	16	

• Virtual pedigree

**Pedigrees**

Construct pedigrees

Clear pedigree

Family	ID	D5S818		D13S317		D8S1179		D21S11		D7S820		CSF1PO		D3S1358	
		Allele 1	Allele 2	Allele 1	Allele 2	Allele 1	Allele 2	Allele 1	Allele 2	Allele 1	Allele 2	Allele 1	Allele 2	Allele 1	Allele 2
Family 1	Grandfather	10	12	8	10	10	12	13.2	13.2	8	12	11	12	15	16
	Grandmother	9	11	9	13	13	13	30	31	9	12	10	12	15	17
	Father	12	11	8	9	12	13	31.2	30	12	12	11	12	16	17
	Mother	13	13	11	12	11	13	30	31	12	12	12	12	17	18
	Child	11	13	8	12	12	13	31.2	30	12	12	11	12	16	16
	First cousin	12	11	8	9	10	13	31.2	30	11	11	10	12	16	17
Family 2	Grandfather	11	12	8	11	10	13	33.2	33.2	8	12	12	12	15	15
	Grandmother	11	12	8	11	10	13	33.2	33.2	8	12	12	12	15	15
	Father	13	12	11	8	14	13	29	33.2	31	8	12	12	16	15
	Mother	10	10	10	13	13	12	30	30	8	10	12	12	15	16
	Child	13	10	11	10	14	11	29	30	11	10	12	12	16	16
	First cousin	13	12	11	8	14	10	29	33.2	31	12	12	12	16	15

Genotype data of virtual pedigree

# KI calculation

$$KI = \frac{\Pr(G_2 | G_1, H_p)}{\Pr(G_2 | G_1, H_d)} \rightarrow \Pr(G_2 | G_1) = \Pr(G_2 | G_1, Z_2) \times \Pr(Z_2) + \Pr(G_2 | G_1, Z_1) \times \Pr(Z_1) + \Pr(G_2 | G_1, Z_0) \times \Pr(Z_0)$$

• ITO stochastic matrices

$G_2$

	$G_1$				
	AA	AB	BB	AC	
I =	AA	1	0	0	0
	AB	0	1	0	0
	BB	0	0	1	0
T =	AA	$P_A$	$P_B$	0	$P_C$
	AB	$0.5P_A$	$0.5(P_A+P_B)$	$0.5P_B$	$0.5P_C$
	BB	0	$P_A$	$P_B$	0
O =	AA	$P_A^2$	$2P_A P_B$	$P_B^2$	$2P_A P_C$
	AB	$P_A^2$	$2P_A P_B$	$P_B^2$	$2P_A P_C$
	BB	$P_A^2$	$2P_A P_B$	$P_B^2$	$2P_A P_C$

• Probability that two individuals with a given relationship share 0, 1, and 2 pairs of IBD alleles

Relationship	$\Pr(Z_0)$	$\Pr(Z_1)$	$\Pr(Z_2)$
Identical twin	0	0	1
Parent/Child	0	1	0
Full-siblings	1/4	1/2	1/4
Grandparent/Grandchild	1/2	1/2	0
Uncle/Nephew	1/2	1/2	0
Half-siblings	1/2	1/2	0
First cousin	3/4	1/4	0
Unrelated	1	0	0

→ adopted from Nat Rev Genet 2006;7(10):771-80.

→ adopted from Biometrics 1954; 10:347-360.

# Calculated KI values – truly related and unrelated genotype pairs

**Related pairs**  
- First cousins

Calculate likelihood ratio

Start: 2013-02-19 End: 2022-05

Clear likelihood ratio

Family	Ch-Fc	D15S10	D15S17	D6S1179	D21S11	D7S820	C21RF0	D15S198	TH01	D16S19	VWA	D18S51	HLA	IFCN	CEU	Log10(KI)	No. of IBS
family	Ch-Fc	0.95051314	1.25426194	0.75	1.008304156	2.27376623	1.07375024	0.75	1.7171788	0.96318307	1.335480094	1.064950446	1.104710556	0.75	1.35479089	0.5	12
family	Ch-Fc	1.056172549	1.021857329	1.420617008	1.002764073	0.922228015	1.000000000	1.000000000	1.000000000	0.75	1.070523802	0.75	1.848418278	0.879078893	3.550865914	0.6	12
family	Ch-Fc	1.223200923	0.982428432	0.75							1.122467223	1.104710556	1.008157786	1.008157786	0.753402485	-0.1	11
family	Ch-Fc	0.75	1.414556539	1.420617008							1.139900887	1.097029428	1.062900422	2.990234689	0.5	15	
family	Ch-Fc	1.022687609	1.183281227	0.75	1.100336323	0.933230219	0.75	0.75	1.009659327	1.603533629	1.264528549	0.75	1.10471834	0.933823529	0.762009716	-0.1	11

**KI values of related genotype pairs**

---

**Unrelated pairs**  
- First cousins

Calculate likelihood ratio

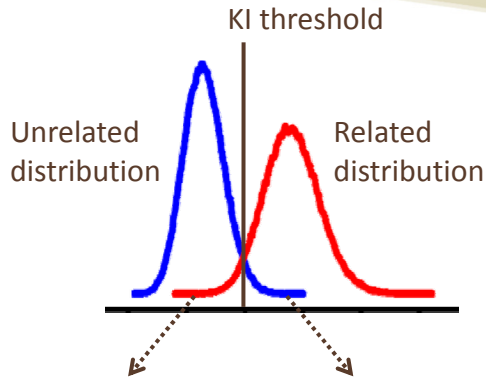
Start: 2013-02-19 End: 2022-05

Clear likelihood ratio

Random GT pair	Random GT 1	Random GT 2	LR values	D15S10	D15S17	D6S1179	D21S11	D7S820	C21RF0	D15S198	TH01	D16S19							
Random GT pair	Random GT 1	7	12	10	11	12	14	29	30	8	9	11	12	15	18	9	10	9	11
Random GT 2	10	10	11	12	10	14	30	30	8	11	10	11	15	16	9	9	9	12	13
LR values	0.75			1.104509359	1.100336323				1.18828892	1.104710556	0.909805676					1.009659327			
Random GT pair	Random GT 1	9	10	8	12	14	14	29	30	11	9	12	17	17	6	8	11	12	12
Random GT 2	11	12	10	10	16	29	2	30	10	12	12	14	14	15	6	9	11	12	12
LR values	0.75			0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	1.53369906			1.334475099	
Random GT pair	Random GT 1	10	11	12											15	16	9	9	11
Random GT 2	7	11	10												15	15	9	9	10
LR values	0.95051314			0.75											1.069611393			1.269318654	
Random GT pair	Random GT 1	9	12	8	11	10	15	30	30	9	11	9	13	16	17	8	9	9	11
Random GT 2	11	11	10												17	9	9	11	12
LR values	0.75			1.298806649	1.450672646				0.75	0.75	0.75	0.75	1.273554575		1.009659327			0.994618395	
Random GT pair	Random GT 1	11	12	9	14	12	13	30	30	10	10	10	13	15	16	9	9	8	9
Random GT 2	12	13	10	11	13	15	30	33.2	11	12	10	10	13	15	15	7	9	10	11
LR values	1.022687609			0.75	1.021267361	1.100336323		0.75	1.279212552	1.069611393			1.009659327					0.75	

**KI values of unrelated genotype pairs**

## KI distribution and its evaluation

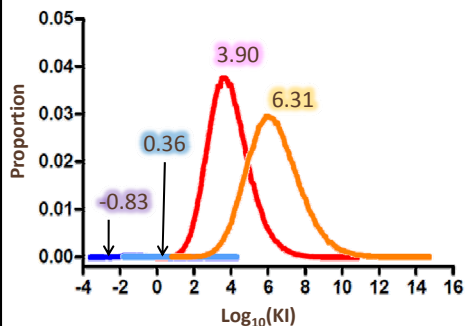


KI values less than threshold on related distribution  
 - The values are falsely excluded.  
 → False negative rate

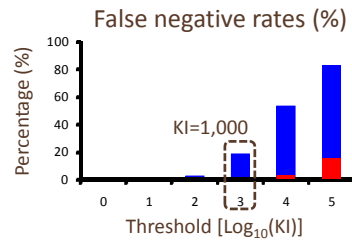
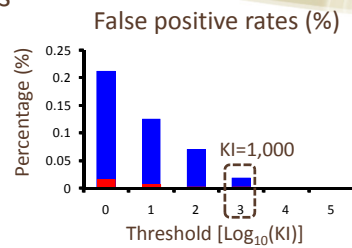
KI values greater than threshold on unrelated distribution  
 - The values are falsely included.  
 → False positive rate

## KI distribution in parent/child relationship

13 STRs → 20 STRs

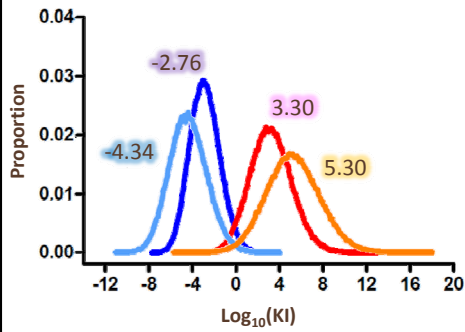


	Related	Unrelated
13 STRs	Red	Blue
20 STRs	Orange	Light Blue

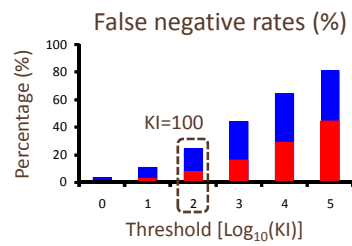
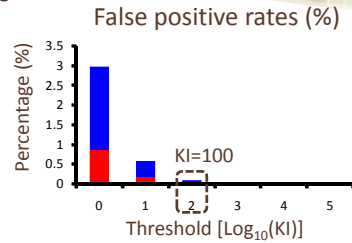


## KI distribution in full-sibling relationship

13 STRs → 20 STRs

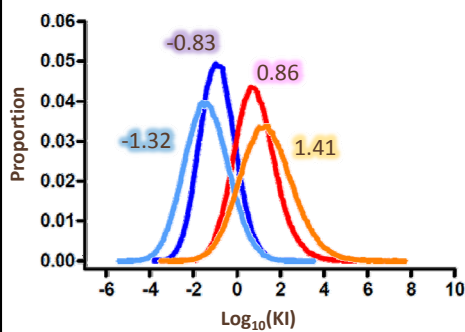


	Related	Unrelated
13 STRs	Red	Blue
20 STRs	Orange	Light Blue

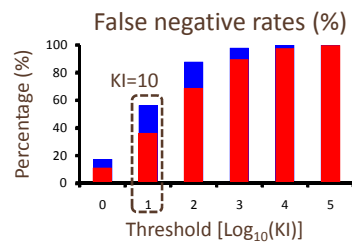
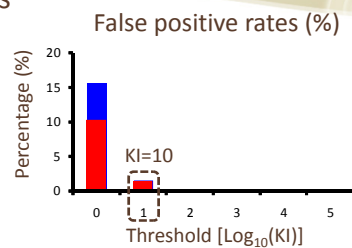


## KI distribution in uncle/nephew relationship

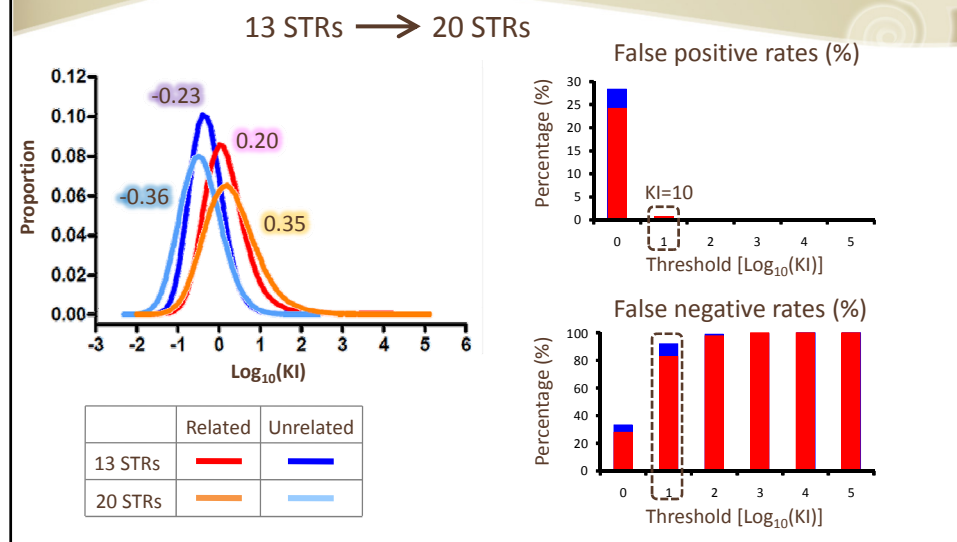
13 STRs → 20 STRs



	Related	Unrelated
13 STRs	Red	Blue
20 STRs	Orange	Light Blue



## KI distribution in first cousin relationship



## Concluding remarks (1)

- Using 13 CODIS STRs, true relatives in parent/child and full-sibling relationships could be discriminated from unrelated persons with KI thresholds of 1,000 and 100, respectively.
- However, the CODIS STRs lacked discrimination power to differentiate between true and unrelated pairs in uncle/nephew and first cousin relationships with KI threshold of even 10.
- By increasing the number of STR to 20, discrimination between true and unrelated pairs was significantly improved in parent/child and full-sibling relationships, but not in uncle/nephew and first cousin relationships.



## Concluding remarks (2)

- To raise discrimination power in uncle/nephew and first cousin relationships, more than 20 STRs should be needed. Alternatively, SNP and lineage markers (Y-STR and mitochondrial DNA) may also be helpful to improve it in the relationships.
- KI has been utilized in various kinship testing and can be directly used to evaluate potential candidate in familial searching.
- Thus, the KI data from this study will help to establish guidelines on various kinship testing and familial searching in Koreans.