

# Mixture에서 봉우리 면적을 활용한 유전자 증거의 해석

홍유림<sup>1</sup> · 이효정<sup>2</sup> · 이재원<sup>3</sup>

<sup>1</sup>삼성생명 강북사업부, <sup>2</sup>고려대학교 통계학과, <sup>3</sup>고려대학교 통계학과

(2009년 6월 접수, 2009년 12월 채택)

## 요약

형사사건에서 STR(short tandem repeat)을 이용한 개인 식별은 범죄현장에서 발견된 유전자 증거와 용의자의 유전자 검사 결과를 이용하여 평가하게 된다. 특히 범죄현장에서 발견된 유전자 증거 표본에 기여자가 두 명 이상인 경우를 Mixture라고 하며, 이는 강간 등과 같은 사건에서 흔히 볼 수 있다. 이러한 상황에서 법의학적 추론을 위해 유전자 증거 표본을 해석하고자 하는 연구들이 계속되어 왔으며, 최근에는 Mixture에서 발생한 대립유전자의 봉우리 면적을 활용하여 유전자 증거를 해석하기 위한 노력들이 계속되고 있다. 따라서 본 연구에서는 봉우리 면적을 이용하여 유전자 증거 표본을 해석하기 위한 연구 방법들에 대해 살펴볼 것이며, 또한 유전자 증거 표본에 기여한 사람의 수가 세 명 이상인 표본으로 확장시킨 연구를 실시해 볼 것이다. 마지막으로 사례를 통해 유전자 증거를 해석하는 방법을 살펴보고자 한다.

주요용어: STR(short tandem repeat), 법의학적 추론, Mixture, 봉우리 면적.

## 1. 서론

범죄현장에서 발견된 유전자 증거 표본(DNA profiles)에 기여한 사람이 두 명 이상인 경우를 Mixture라고 하며, 이는 강간 등과 같은 사건에서 흔히 볼 수 있는 경우로 피해자와 용의자의 유전자가 섞여 있거나 또는 다수의 용의자 유전자가 함께 발견되는 경우를 들 수 있다. Mixture에 대한 유전자 증거를 해석하기 위한 방법으로 증거 표본에서 발견된 대립유전자 빈도를 이용한 방법이 먼저 제안되었는데 (Evetts 등, 1991; Weir 등, 1997; Gill 등, 1998b; Curren 등, 1999), 분석방법은 유전자 증거 표본에 기여한 사람과 용의자의 수를 정한 후 유전자 증거 표본에서 관측된 대립유전자의 모든 가능한 유전자형 조합에 대한 확률을 이용하여 우도비(likelihood ratio; LR)를 계산하는 방법이다. 그러나 이러한 방법은 분석에 앞서 유전자 증거 표본에 기여한 사람의 수를 정해야 하며, 또한 알려지지 않은 기여자의 수가 많아질수록 우도비의 값이 작아진다는 한계점을 지닌다 (Brenner 등, 1996). 이러한 한계점을 해결하기 위한 방법으로 Mixture에서 봉우리 면적(peak areas)을 활용하여 우도비를 계산하는 방법에 대한 연구들이 시도되었다. Evetts 등 (1998)는 봉우리 면적을 이용하여 혼합 비율(mixing proportion)을 정의하고 이를 우도비 계산에 포함시키는 방법을 제안하였으며, Gill 등 (1998b)은 유전자 증거 표본에 기여자가 두 명인 경우에 유전자 증거를 해석하는 모형을 제안하였는데, 이 방법은 두 명의 혼합 비율을 추정하기 위한 간단한 모형을 설정한 후, 모의실험을 통해 유전자 증거로부터 나타날 가능성

본 연구는 2008년 고려대학교 특별연구비 지원에 의해 수행되었으며(K0822201), 이효정은 대한민국 과학기술부(MOST)와 한국과학재단(KOSEF)의 특정연구개발사업 프로그램(M10640030002-08N4003-00210)의 지원을 받았다.

<sup>3</sup>교신저자: (136-701) 서울특별시 성북구 안암동 5-1, 고려대학교 통계학과, 교수. E-mail: jael@korea.ac.kr

이 거의 없는 유전자형 조합을 우도비 계산 시 제외하는 방법을 제안하였다. 본 연구에서는 Mixture에서 봉우리 면적을 활용하여 유전자 증거 표본을 해석하기 위한 연구인 Evett 방법 (1998)과 Gill 방법 (1998b)에 대해 자세히 살펴 볼 것이다. 또한 유전자 증거 표본에 기여자 수가 두 명인 경우에 대한 Gill 방법 (1998b)에 대해 세 명 이상인 표본으로 확장시킨 연구를 실시해 보고 사례를 통해 유전자 증거를 해석하는 방법을 살펴보고자 한다.

## 2. 봉우리 면적을 활용한 유전자 증거 해석

### 2.1. Evett 방법

Evett 등 (1998)은 봉우리 면적을 이용하여 유전자 증거를 해석하기 위한 우도비 모형을 제시하였다. 다음과 같은 가설 하에 제시된 우도비를 살펴보면,

$H_0$ : Mixture에 의한 유전자 증거는 용의자와  $(n-1)$ 명의 알려지지 않은 기여자에 의한 것이다.

$H_1$ : Mixture에 의한 유전자 증거는  $n$ 명의 알려지지 않은 기여자에 의한 것이다.

$$LR = \frac{P(E|H_0)}{P(E|H_1)} = \frac{P(A, w, G_1|H_0)}{P(A, w, G_1|H_1)} = \frac{P(A, w|H_0, G_1)}{P(A, w|H_1, G_1)} \cdot \frac{P(G_1|H_0)}{P(G_1|H_1)}$$

으로, 각 가설 하에서 유전자 증거  $E$ 가 관측될 확률의 비이다. 여기서  $A = (A_1, A_2, \dots, A_r)$ 는 대립 유전자의 집합,  $w = (w_1, w_2, \dots, w_r)$ 은 봉우리 면적의 집합,  $f = (f_1, f_2, \dots, f_r)$ 은 대립유전자 빈도,  $G_1$ 은 용의자의 유전자형이다. 위의 우도비 식에서  $P(G_1|H_0)/P(G_1|H_1) = 1$ 로 가정할 때,  $LR = P(A, w|H_0, G_1)/P(A, w|H_1, G_1)$ 이 되고  $G_i$ 를  $n$ 개 유전자형 집합인  $(G_1, G_2, \dots, G_n)$ 이라고 하면 Mixture에서 발생할 수 있는 모든 가능한 유전자형  $G_i$ 에 대해 고려할 때 우도비는 다음과 같다.

$$LR = \frac{\sum_i P(A, w|C, G_i, G_1)P(G_i|H_0, G_1)}{\sum_i P(A, w|C, G_i)P(G_i|H_1)}$$

$m$ 을  $n$ 개 유전자형에 대한 혼합 비율  $(m_1, m_2, \dots, m_n)$ 의 집합이라고 하면, 우도비는  $m$ 의 모든 값에 대하여 적분하여 계산된다.

$$LR = \frac{\int \sum_i P(G_i|H_0, G_1)P(A, w|H_0, G_i, G_1, m)p(m|H_0, G_i, G_1)dm}{\int \sum_i P(G_i|H_1)P(A, w|H_1, G_i, m)p(m|H_1, G_i)dm}$$

여기서  $p(m|H_0, G_i, G_1)$ 와  $p(m|H_1, G_i)$ 는 사전 확률밀도함수이며 다음과 같은 3가지 가정 하에 우도비가 계산된다.

- (가정 1) 사전 확률밀도함수  $p(m|H_0, G_i, G_1)$ 와  $p(m|H_1, G_i)$ 는 유전자형과 독립이며  $p(m|H_0, G_i, G_1) = p(m|H_1, G_i) = p(m)$ 으로 동일하다.
- (가정 2) 분자 및 분모 합의 모든  $i$ 에 대해서  $P(A|H_0, G_i, G_1, m) = P(A|H_1, G_i, m) = 1$ 이다.
- (가정 3) 봉우리 면적은 유전자형과 혼합비율에 의존한다고 할 때,  $P(w|A, H_0, G_i, G_1, m) = P(w|G_i, G_1, m)$ 이고  $P(w|A, H_1, G_i, m) = P(w|G_i, m)$ 이다.

이러한 가정들을 모두 적용하면 우도비는 다음과 같다.

$$LR = \frac{\int \sum_i P(G_i|H_0, G_1)P(w|G_i, G_1, m)p(m)dm}{\int \sum_i P(G_i|H_1)P(w|G_i, m)p(m)dm}$$

이와 같이 유전자 증거 표본이 두 명 이상인 경우인 Mixture에서 봉우리 면적을 이용하여 우도비를 계산한 후 유전자 증거를 해석할 수 있다. 그러나 우도비에서  $p(m)$ 은  $m$ 이 알려지지 않은 경우 사전 분포에 대한 적분을 필요로 하기 때문에 실제 적용에 있어서는 어렵다는 단점을 지닌다.

## 2.2. Gill 방법

Gill 등 (1998b)은 Mixture에서 기여자가 두 명인 경우에 대해 유전자 증거를 해석하는 모형을 제시하였다. 이 방법에서 가장 중요한 것은 혼합 비율(mixture proportion)을 추정하는 것으로 혼합 비율은 다음과 같이 계산된다. 대립유전자의 수가 4개인 유전자 좌에서 대립 유전자를 시각적으로 주 대립유전자(major allele)와 부 대립유전자(minor allele)로 구분하고, 시각적으로 구분이 어려운 경우에는 모든 경우를 고려하여 우도비를 계산하는 방법이 좋은 선택이 될 것이다.

유전자 증거를 주와 부의 대립유전자로 구분할 수 있다면 유전자 증거 표본을 구성하는 요소들의 비율을 계산할 수 있다. 예를 들어 4개의 대립유전자  $A, B, C, D$ 의 봉우리 면적을 각각  $paA, paB, paC, paD$ 라고 하고, 두 명의 기여자 중 첫 번째 사람( $x$ )의 유전자형을  $AB$ , 두 번째 사람( $y$ )의 유전자형을  $CD$ 라고 할 때  $x$ 에 의해 기여된 혼합 비율은 다음과 같다.

$$\widehat{M}_x = \frac{paA + paB}{paA + paB + paC + paD}$$

한편, 유전자 증거 표본에서 2개 혹은 3개의 대립유전자로 구성된 유전자 좌가 존재할 수 있다. 이때  $\widehat{M}_x$ 은 모든 유전자 좌에서 유사한 값을 가지므로 4개의 대립유전자의 유전자 좌에서 계산된  $\widehat{M}_x$ 을 적용할 수 있다.  $\widehat{M}_x$ 이 주어졌을 때 유전자 증거 표본에서 발견된 대립유전자 개수에 따른 각 대립유전자의 기대 비율은 표 2.1과 같이 나타낼 수 있다. 이러한 해석이 올바르다면 대립유전자의 기대 비율과 관측 비율이 유사할 것이고 이들의 차이를 이용하여 다음과 같이 잔차가 계산되며, 각 유전자 좌의 모든 유전자형 조합에서 잔차가 가장 작은 조합, 즉 해당 유전자 좌에서 나타날 가능성이 가장 높은 유전자형 조합을 선택하게 되는 방법이다.

$$\text{잔차} = \sum (\text{관측 대립유전자 비율} - \text{기대 대립유전자 비율})^2$$

이 연구에서 봉우리 면적 차이에 대한 영향력을 평가하기 위한 모의실험 모형이 개발되었는데, 먼저 봉우리 면적 초기값의  $\pm 20\%$  범위 내에서 독립적으로 봉우리 면적을 생성하고, 혼합 비율은 실제 관측된  $\widehat{M}_x$ 의 평균과 표준편차를 가지는 정규분포에서 생성한다. 생성된 봉우리 면적과 혼합 비율을 이용하여 기대 및 관측 대립유전자 비율을 계산한 후 모든 가능한 유전자형 조합에 대해 최소 잔차를 갖는 조합을 선택하는 방법이다. 이 과정을 1,000번 반복하여 유전자형 조합이 나타날 가능성이 적은 유전자형 조합은 우도비 계산 시 고려하지 않는 방법이다.

Gill 방법은 Mixture에 기여자가 두 명일 때 유전자 증거 표본을 해석할 수 있는 간단한 방법이다. 그러나 Mixture에서 보통 두 명 내지 세 명으로 구성된 유전자 증거가 많이 존재함을 고려할 때 세 명 이상의 기여자로 구성된 표본에서는 적용이 어렵다는 한계점을 지닌다. 따라서 3장에서는 Mixture의 기여자가 세 명 이상으로 구성된 경우 봉우리 면적을 활용하여 유전자 증거를 해석할 수 있는 방법에 대해 살펴볼 것이다.

표 2.1. 대립유전자의 개수에 따른 기대 비율

대립유전자 개수	유전자형 조합	대립유전자			
		A	B	C	D
4개인 경우	AB, CD	$\widehat{M}_x/2$	$\widehat{M}_x/2$	$(1 - \widehat{M}_x)/2$	$(1 - \widehat{M}_x)/2$
	AC, BD	$\widehat{M}_x/2$	$(1 - \widehat{M}_x)/2$	$\widehat{M}_x/2$	$(1 - \widehat{M}_x)/2$
	AD, BC	$\widehat{M}_x/2$	$(1 - \widehat{M}_x)/2$	$(1 - \widehat{M}_x)/2$	$\widehat{M}_x/2$
	BC, AD	$(1 - \widehat{M}_x)/2$	$\widehat{M}_x/2$	$\widehat{M}_x/2$	$(1 - \widehat{M}_x)/2$
	BD, AC	$(1 - \widehat{M}_x)/2$	$\widehat{M}_x/2$	$(1 - \widehat{M}_x)/2$	$\widehat{M}_x/2$
	CD, AB	$(1 - \widehat{M}_x)/2$	$(1 - \widehat{M}_x)/2$	$\widehat{M}_x/2$	$\widehat{M}_x/2$
3개인 경우		A	B	C	
	AA, BC	$\widehat{M}_x$	$(1 - \widehat{M}_x)/2$	$(1 - \widehat{M}_x)/2$	
	BB, AC	$(1 - \widehat{M}_x)/2$	$\widehat{M}_x$	$(1 - \widehat{M}_x)/2$	
	CC, AB	$(1 - \widehat{M}_x)/2$	$(1 - \widehat{M}_x)/2$	$\widehat{M}_x$	
	AB, AC	0.5	$\widehat{M}_x/2$	$(1 - \widehat{M}_x)/2$	
	BC, AC	$(1 - \widehat{M}_x)/2$	$\widehat{M}_x/2$	0.5	
	AB, BC	$\widehat{M}_x/2$	0.5	$(1 - \widehat{M}_x)/2$	
	BC, AC	$1 - \widehat{M}_x$	$\widehat{M}_x/2$	$\widehat{M}_x/2$	
	AC, BB	$\widehat{M}_x/2$	$1 - \widehat{M}_x$	$\widehat{M}_x/2$	
	AB, CC	$\widehat{M}_x/2$	$\widehat{M}_x/2$	$1 - \widehat{M}_x$	
	AC, AB	0.5	$(1 - \widehat{M}_x)/2$	$\widehat{M}_x/2$	
	AC, BC	$\widehat{M}_x/2$	$(1 - \widehat{M}_x)/2$	0.5	
BC, AB	$(1 - \widehat{M}_x)/2$	0.5	$\widehat{M}_x/2$		
2개인 경우		A	B		
	AA, AB	$\widehat{M}_x/2 + 0.5$	$(1 - \widehat{M}_x)/2$		
	AB, AB	0.5	0.5		
	AA, BB	$\widehat{M}_x$	$1 - \widehat{M}_x$		
	AB, AA	$1 - \widehat{M}_x/2$	$\widehat{M}_x/2$		
	BB, AA	$1 - \widehat{M}_x$	$\widehat{M}_x$		
	AB, BB	$\widehat{M}_x/2$	$1 - \widehat{M}_x/2$		
	BB, AB	$(1 - \widehat{M}_x)/2$	$\widehat{M}_x/2 + 0.5$		

3. Mixture 기여자가 세 명 이상인 경우 유전자 증거 해석

Gill 등 (1998b)이 제안한 방법은 봉우리 면적을 이용하여 혼합 비율을 구하는 것이 중요하며 혼합 비율을 직접 구하지 못한다면 이를 추정하는 과정이 중요하다. 또한 이 방법은 유전자 증거 표본의 기여자가 두 명인 경우에만 적용할 수 있는 방법으로 세 명 이상인 경우에는 적용이 어렵다는 한계를 지닌다. 대부분의 유전자 증거 표본이 두 명 또는 세 명의 기여자로 구성되어 있음을 고려할 때 세 명의 기여자로 구성되어 있는 유전자 증거 표본을 해석할 수 있는 방법을 제안해 보기로 한다.

3.1. 유전자 증거 해석 방법

두 개의 대립유전자가 동일한 사람에게서 나온 것이라면 대립유전자의 봉우리 면적 비는 1:1이 되어야 한다. 그러나 유전자 증폭(polymerase chain reaction; PCR) 과정에서 나타나는 측정 오차로 인하여 관측된 대립유전자의 봉우리 면적들 간에는 약간의 차이가 존재하게 된다. 표 3.1은 6개 유전자 좌에서 대립유전자의 봉우리 면적 비(작은 봉우리 면적/큰 봉우리 면적)를 계산하여 분석한 것으로 (Gill 등, 1997) 평균과 중앙값이 모든 유전자 좌에서 0.9 이상의 값을 가짐을 볼 수 있다. 또한 동일한 사람에게

표 3.1. 6개 유전자 좌에서 대립유전자 봉우리 면적 분석

	평균	중앙값	표준편차	95% 하한값	최소값	관측수
D18S51	0.90	0.92	0.08	0.74	0.61	190
D21S11	0.92	0.93	0.05	0.83	0.70	195
D8S1179	0.91	0.92	0.07	0.77	0.64	102
HUMFIBRA/FGA	0.90	0.90	0.06	0.79	0.77	128
HUMTH01	0.92	0.95	0.07	0.80	0.71	46
HUMVWA	0.91	0.94	0.07	0.77	0.68	93

서 나온 두 개의 대립유전자는 다른 사람에게서 나온 두 개의 대립유전자와 비교할 때 유사한 봉우리 면적 값을 보이므로 동일한 사람의 것으로 알려진 유전자형의 봉우리 면적 비를 구해 보면 그렇지 않은 경우보다 큰 값을 가지게 될 것이다. 이러한 성질을 유전자 증거 표본 해석에 이용해 보고자 한다.

### 3.2. 모의실험 방법 및 결과

모의실험의 주요 방법은 한 개의 유전자 좌에서 모든 가능한 유전자형 조합에 대하여 봉우리 면적 비를 계산한 후 이를 비교하여 식별하는 것이다. 세 명으로 구성된 증거 표본을 고려할 때 각 유전자형 조합에 대해 세 개의 봉우리 면적 비를 얻을 수 있으며, 모든 가능한 유전자형 조합에서 봉우리 면적 비를 계산한 후 각 조합에서 최소값을 선택한다. 이 값들을 비교하여 신뢰할 수 있는 범위 내에 포함되는 값을 가지는 유전자형 조합을 선택하도록 하는 방법이다. 분석 과정을 단계별로 살펴보면 다음과 같다.

(단계 1) 유전자 증거를 구성하고 있는 대립유전자를 식별한 후 각 대립유전자의 봉우리 면적을 산한다.

(단계 2) 식별된 대립유전자를 이용하여 가능한 유전자형 조합을 구성한다.

(단계 3) 봉우리 면적을 이용한 유전자 증거 해석을 위한 모의실험을 수행한다.

(단계 4) 유전자 증거로 나타날 수 있는 가능성이 있는 유전자형 조합을 선택한다.

(단계 3)의 모의실험 방법은 먼저 봉우리 면적의 관측 오차를 고려하여 봉우리 면적 초기값의  $\pm 20\%$  범위 내에서 독립적으로 생성하는 Gill 등 (1998b)의 방법과 동일하게 적용하였다. 봉우리 면적을 생성한 후 고려해야 할 유전자형 조합에 대해 각각의 봉우리 면적 비를 계산한다. 기여자가 세 명인 경우 봉우리 면적 비는 각 유전자형 조합마다 세 개씩 나오게 될 것이며, 이 중 가장 작은 값  $R = \min(\text{ratio } 1, \text{ratio } 2, \text{ratio } 3)$ 을 선택한다. 이 과정이 1,000번 반복되며 대립유전자의 수는 4, 5, 6개인 경우를 고려하였다. (단계 4)에서 가능성 있는 유전자형 조합의 선택 기준은  $R$ 이 95% 신뢰구간의 하한값 이상으로 나타나는 경우로 정의하였으며(즉 D18S51 유전자 좌의 경우에는 유전자형 조합의 선택을 위한 기준 값은 0.74임), 이 선택기준을 만족하지 못하는 유전자형 조합은 우도비 계산 시 고려 대상에서 제외되는 것이다.

본 연구에서는 D18S51 유전자 좌의 유전자 증거 표본을 이용하여 모의실험을 수행하였으며, 1,000번의 모의실험 결과에서 봉우리 면적 비가 0.74 이상의 값을 가지는 유전자형 조합을 계수하였다. 표 3.2는 대립유전자 개수 및 초기 봉우리 면적에 따른 모의실험 결과이다. 결과를 살펴보면, 대립유전자의 개수가 6개이고 초기 봉우리 면적이 A: 1339 B: 1662, C: 2580, D: 2830, E: 3480, F: 3765인 경우 (AB, CD, EF) 유전자형 조합이 547번으로 가장 많이 선택되어 가능성이 가장 높은 조합으로 판단되었으며 그 다음으로 (AB, CE, DF), (AB, CF, DE) 조합이 가능성이 높은 조합이었다. 한 번도 선택되지 못한 유전자형 조합은 우도비 계산 시 제외되는 방법이다.

표 3.2. Mixture에서의 대립유전자 개수/초기 봉우리 면적에 따른 모의실험 결과

대립유전자 개수가 6개인 경우											
초기 봉우리 면적: A: 1339, B: 1662, C: 2580, D: 2830, E: 3480, F: 3765											
AB CD EF	547	AC BD EF	2	AD BC EF	0	AE BC DF	0	AF BC DE	0		
AB CE DF	178	AC BE DF	0	AD BE CF	0	AE BD CF	0	AF BD CE	0		
AB CF DE	158	AC BF DE	0	AD BF CE	0	AE BF CD	0	AF BE CD	0		
초기 봉우리 면적: A: 1339, B: 1662, C: 1543, D: 1860, E: 3540, F: 3675											
AB CD EF	474	AC BD EF	658	AD BC EF	368	AE BC DF	0	AF BC DE	0		
AB CE DF	0	AC BE DF	0	AD BE CF	0	AE BD CF	0	AF BD CE	0		
AB CF DE	0	AC BF DE	0	AD BF CE	0	AE BF CD	0	AF BE CD	0		
대립유전자 개수가 5개인 경우											
초기 봉우리 면적: A: 1339, B: 1465, C: 1378, D: 1480, E: 2895											
AB AC DE	0	AB AD CE	0	AB AE CD	0	AC AD BE	0	AC AE BD	0		
AD AE BC	0	AA BC DE	5	AA BD CE	0	AA BE CD	5	AB BC DE	0		
AB BD CE	0	AB BE CD	0	AC BD BE	0	AE BC BD	0	AD BC BE	0		
AC BB DE	6	AD BB CE	0	AE BB CD	0	AC BC DE	0	AC CD BE	0		
AC CE BD	0	AB CD CE	0	AD BC CE	0	AE BC CD	0	AB CC DE	4		
AD CC BE	4	AE CC BD	0	AD BC DE	0	AC CD BE	0	AD CE BD	0		
AB CD DE	0	AC BD DE	0	AE BD CD	0	AB CE DD	4	AC BE DD	4		
AE BC DD	0	AE BC DE	773	AE BE CD	753	AE BD CE	784	AB CE DE	789		
AC BE DE	837	AD BE CE	779	AB CD EE	806	AC BD EE	876	AD BC EE	817		
초기 봉우리 면적: A: 1339, B: 1465, C: 1378, D: 1480, E: 1425											
AB AC DE	0	AB AD CE	0	AB AE CD	0	AC AD BE	0	AC AE BD	0		
AD AE BC	0	AA BC DE	847	AA BD CE	864	AA BE CD	829	AB BC DE	2		
AB BD CE	0	AB BE CD	1	AC BD BE	0	AE BC BD	0	AD BC BE	0		
AC BB DE	849	AD BB CE	799	AE BB CD	802	AC BC DE	0	AC CD BE	0		
AC CE BD	0	AB CD CE	0	AD BC CE	0	AE BC CD	0	AB CC DE	809		
AD CC BE	800	AE CC BD	841	AD BC DE	0	AC CD BE	0	AD CE BD	0		
AB CD DE	0	AC BD DE	0	AE BD CD	0	AB CE DD	809	AC BE DD	800		
AE BC DD	841	AE BC DE	0	AE BE CD	0	AE BD CE	0	AB CE DE	0		
AC BE DE	0	AD BE CE	0	AB CD EE	783	AC BD EE	861	AD BC EE	786		
대립유전자 개수가 4개인 경우											
초기 봉우리 면적: A: 1335, B: 3020, C: 2995, D: 2790											
AB AC BD	0	AD AB BC	0	CD AB AB	0	AB BC CD	18	AC BC BD	24		
AD BC BC	0	AB CD CD	0	AC CD BD	0	AD BC CD	0	AB CD AD	0		
AC AD BD	0	AD AD BC	0	AA BC BD	0	BB AC CD	10	CC AD BD	0		
DD AB AC	0	AA BB CD	898	BB CC AD	0	CC DD AB	0	AA DD BC	931		
초기 봉우리 면적: A: 1325, B: 1465, C: 2995, D: 2790											
AB AC BD	0	AD AB BC	0	CD AB AB	764	AB BC CD	0	AC BC BD	0		
AD BC BC	0	AB CD CD	764	AC CD BD	735	AD BC CD	768	AB CD AD	0		
AC AD BD	0	AD AD BC	0	AA BC BD	0	BB AC CD	14	CC AD BD	844		
DD AB AC	0	AA BB CD	894	BB CC AD	0	CC DD AB	860	AA DD BC	0		
초기 봉우리 면적: A: 2537, B: 3020, C: 2995, D: 2790											
AB AC BD	0	AD AB BC	0	CD AB AB	700	AB BC CD	3	AC BC BD	2		
AD BC BC	820	AB CD CD	700	AC CD BD	0	AD BC CD	0	AB CD AD	0		
AC AD BD	0	AD AD BC	820	AA BC BD	2	BB AC CD	6	CC AD BD	0		
DD AB AC	0	AA BB CD	903	BB CC AD	889	CC DD AB	773	AA DD BC	925		

4. 사례연구

Mixture에서 기여자가 세 명인 경우 우도비를 계산하는 과정을 사례를 통해 적용하였다. 범칙 현장에서 혈흔이 발견되었으며 사건의 정황으로 볼 때 이 혈흔은 세 명의 것으로 생각할 수 있다. 이때 다음의 가설 하에 유전자 좌 HUMTH01의 검사 결과 5개의 대립유전자가 조사되었으며 각 대립유전자에서

표 4.1. 모의실험 결과

AB AC DE	0	AB AD CE	0	AB AE CD	0	AC AD BE	0	AC AE BD	0
AD AE BC	0	AA BC DE	2	AA BD CE	104	AA BE CD	0	AB BC DE	0
AB BD CE	0	AB BE CD	0	AC BD BE	0	AE BC BD	0	AD BC BE	0
AC BB DE	0	AD BB CE	0	AE BB CD	0	AC BC DE	627	AC CD BE	0
AC CE BD	0	AB CD CE	0	AD BC CE	0	AE BC CD	0	AB CC DE	648
AD CC BE	0	AE CC BD	0	AD BC DE	627	AC CD BE	0	AD CE BD	0
AB CD DE	0	AC BD DE	0	AE BD CD	0	AB CE DD	648	AC BE DD	0
AE BC DD	0	AE BC DE	0	AE BE CD	175	AE BD CE	0	AB CE DE	58
AC BE DE	0	AD BE CE	0	AB CD EE	750	AC BD EE	0	AD BC EE	0

관측된 봉우리 면적은 A: 1339, B: 1465, C: 2890, D: 3229, E: 4160이었으며 용의자의 유전자형은 AB로 밝혀졌다.

$H_0$  : Mixture에 의한 유전자 증거는 용의자와 두 명의 알려지지 않은 기여자에 의한 것이다.

$H_1$  : Mixture에 의한 유전자 증거는 세 명의 알려지지 않은 기여자에 의한 것이다.

유전자 증거 해석을 위한 우도비는 각 가설 하에서 확률 비로 계산될 수 있으며, 봉우리 면적 정보를 이용하기 위해 3.1절의 방법에 따라 모의실험을 실시한 결과 표 4.1과 같다.

모의실험 결과 9개의 유전자 조합이 선택되었으며 (AB, CD, EE) 조합이 750번으로 가장 많이 선택되었다. 선택된 9개 유전자형이 우도비의 계산 시 고려될 것이다. 하에서 용의자의 유전자형이 로 밝혀졌으므로 고려해야 할 유전자 조합은 (AB, CE, DE), (AB, CC, DE), (AB, CE, DD), (AB, CD, EE)이며  $H_1$  하에서 9개 모든 유전자 조합을 고려할 때 우도비는 다음과 같다( $G_S$ : 용의자의 유전자형,  $E_C$ : 범죄현장 흔적의 유전자 증거,  $p_a$ : 대립유전자 A의 빈도).

$$LR = \frac{P(E_C|G_S, H_0)}{P(E_C|G_S, H_1)} = \frac{4p_c p_a p_e (2p_c + p_d + 3p_e)}{24p_a p_b p_c p_d p_e (2p_a + 3p_c + 3p_d + 5p_e)} = \frac{2p_c + p_d + 3p_e}{6p_a p_b (2p_a + 3p_c + 3p_d + 5p_e)}$$

만일 봉우리 면적을 고려하지 않았을 경우 우도비를 계산하면 다음과 같다.

$$LR = \frac{P(E_C|G_S, H_0)}{P(E_C|G_S, H_1)} = \frac{12p_c p_a p_e (2p_a + 2p_b + p_c + p_d + p_e)}{360p_a p_b p_c p_d p_e (p_a + p_b + p_c + p_d + p_e)} = \frac{2p_a + 2p_b + p_c + p_d + p_e}{30p_a p_b (p_a + p_b + p_c + p_d + p_e)}$$

예를 들어 각 대립유전자의 확률을 5%로 가정하면 봉우리 면적을 고려한 경우의 우도비는 30.77이며, 봉우리 면적을 고려하지 않은 경우의 우도비는 18.67로 봉우리 면적을 고려한 경우의 우도비가 더 큰 값을 가지는 것을 볼 수 있다.

## 5. 결론 및 토의

범죄현장에서 발견된 유전자 증거가 피해자의 것과 섞여 있거나 다수의 용의자가 관련되어 Mixture가 발생한 경우에 있어 이를 해석하기 위한 연구들이 진행되었다. 유전자 증거 표본에서 관측된 대립유전자의 모든 가능한 유전자형 조합에 대한 확률을 이용해 우도비(likelihood ratio; LR)를 계산하는 방법이 먼저 시도되었으며, 최근에는 대립유전자 빈도 외에 봉우리 면적을 활용한 연구들이 진행되고 있다.

본 논문에서는 봉우리 면적을 활용하여 Mixture의 유전자 증거 해석을 위한 연구인 Evett 방법(1998)과 Gill 방법(1998b)에 대해 자세히 살펴보았으며, 유전자 증거 표본의 기여자 수가 두 명인 경우의 Gill 방법(1998b)에 대해 세 명 이상인 표본으로 확장시킨 연구를 실시하였다. 본 연구에서는

동일한 사람의 것으로 알려진 대립유전자의 봉우리 면적 비(작은 봉우리 면적/큰 봉우리 면적)가 그렇지 않은 경우보다 큰 값을 가지게 되는 성질을 이용하여 유전자 증거 표본의 해석을 시도한 것으로 모의 실험을 통해 모형을 제시하였다.

본 연구에서 소개된 방법 외에도 봉우리 면적을 활용하여 Mixture를 해석하기 위한 연구들이 많이 진행되고 있다. 표본을 남긴 사람들의 DNA 양에 따라 유전자형을 구분하여 각자의 유전자형을 밝히고, 유전자 증폭 과정에서 발생하는 분석 오류인 여분의 대립유전자 형태로 가장 흔히 발생하는 스테터 밴드(stutter band)를 고려하여 Mixture를 해석하기 위한 연구가 이루어졌다 (Clayton 등, 1998; Gill 등, 1998a). 또한 봉우리 면적에 대한 혼합비율을 추정하여 확률적 모형이 아닌 수치적 접근 방법으로 유전자 증거를 해석하는 방법이 제안되었으며 (Perlin과 Szabady, 2001; Wang 등, 2002), 최근에는 Probabilistic expert systems(PES)의 네트워크 모형(network model)에 기초하여 봉우리 면적 정보를 이용한 Mixture의 식별 및 해석을 위한 연구들이 많이 실시되고 있다 (Dawid 등, 2002; Mortera 등, 2003; Cowell 등, 2007). 이처럼 Mixture에서 대립유전자 빈도 뿐 아니라 봉우리 면적을 활용하여 유전자 증거 표본을 해석하기 위한 다양한 방법들이 제안되고 있으며, Mixture가 흔히 발생할 수 있는 강간 사건 뿐 아니라 난투 등과 같은 더 복잡한 상황에 대해 보다 정확한 법의학적 추론이 가능할 것으로 판단된다.

## 참고문헌

- Brenner, C. H., Fimmers, R. and Baur, M. P. (1996). Likelihood ratios for mixed stains when the number of donors cannot be agreed, *International Journal of Legal Medicine*, **109**, 218-219.
- Clayton, T. M., Whitaker, J. P., Sparkes, R. and Gill, P. (1998). Analysis and interpretation of mixed forensic stains using DNA STR profiling, *Forensic Science International*, **91**, 55-70.
- Cowell, R., Lauritzen, S. and Mortera, J. (2007). Identification and separation of DNA mixtures using peak area information, *Forensic Science International*, **166**, 28-34.
- Curran, M., Triggs, J. C. M., Buckleton, J. and Weir, B. S. (1999). Interpreting DNA mixtures in structured populations, *Journal of Forensic Science*, **44**, 987-995.
- Dawid, J., Mortera, V. L., Pascali and Boxel, D. V. (2002). Probabilistic expert systems for forensic inference from genetic markers, *Scandinavian Journal of Statistics*, **29**, 577-595.
- Evett, I. W., Buffery, C., Willot, G. and Stoney, D. A. (1991). A guide to interpreting single locus profiles of DNA mixture in forensic cases, *Journal of Forensic Science*, **31**, 41-47.
- Evett, I. W., Gill, P. and Lambert, J. A. (1998). Taking account of peak areas when interpreting mixed DNA profiles, *Journal of Forensic Science*, **43**, 62-69.
- Gill, P., Sparkers, R. and Buckleton, J. (1998a). Interpreting of simple mixture of when artefacts such as stutters are present - with special reference to multiplex STRs used by the forensic Science Service, *Forensic Science International*, **89**, 185-197.
- Gill, P., Sparkers, R., Pinchin, R., Clayton, T., Whitaker, J. and Buckleton, J. (1998b). Interpreting of simple STR mixtures using allele peak areas, *Forensic Science International*, **91**, 41-53.
- Gill, P., Sparkers, R., Pinchin, R. and Kimpton, C. (1997). Development of guidelines to designate alleles using an STR multiplex system, *Forensic Science International*, **89**, 185-197.
- Mortera, J., Dawid, A. P. and Lauritzen, S. L. (2003). Probabilistic expert systems for DNA mixture profiling, *Theoretical Population Biology*, **63**, 191-205.
- Perlin, M. W. and Szabady, B. (2001). Linear mixture analysis: A mathematical approach to resolving mixed DNA samples, *Journal of Forensic Science*, **46**, 1372-1378.
- Wang, T., Xue, N. and Wickenheiser, R. (2002). Least square deconvolution (LSD): A new way of resolving STR/DNA mixture samples, in: *Proceedings of the 13th International Symposium on Humana Identification*, October 7-10, Phoenix, AZ.
- Weir, B. S., Triggs, C. M., Starling, L., Stowell, L. I., Walsh, K. A. J. and Buckleton, J. (1997). Interpreting DNA mixtures, *Journal of Forensic Science*, **42**, 213-222.



# Interpreting Mixtures Using Allele Peak Areas

Yu-Lim Hong<sup>1</sup> · Hyo-Jung Lee<sup>2</sup> · Jae Won Lee<sup>3</sup>

<sup>1</sup>Samsung Life Insurance Co.,LTD.; <sup>2</sup>Department of Statistics, Korea University

<sup>3</sup>Department of Statistics, Korea University

(Received June 2009; accepted December 2009)

---

## Abstract

Mixture is that DNA profiles of samples contain material from more than one contributor, especially common in rape cases. In this situation, first, the method based on enumerating a complete set of possible genotype that may have generated the mixed DNA profile have been studied for interpreting DNA mixtures. More recently, the methods utilizing peak area information to calculate likelihood ratios have been suggested. This study is concerned with the analysis and interpretation of mixed forensic stains using quantitative peak area information and the method of forensic inference for extension of material from more than or equal to three contributors. Finally, the numerical example will be outlined.

Keywords: STR(short tandem repeat), forensic inference, mixture, peak areas.

---

Jae Won Lee was supported by a Korea University Grant(K0822201). Hyo-Jung Lee was supported by the grant(M10640030002-08N4003-00210) from National R&D program of Ministry of Science and Technology(MOST) and Korea Science and Engineering Foundation(KOSEF).

<sup>3</sup>Corresponding author: Professor, Department of Statistics, Korea University, Anam-Dong, Seongbuk-Gu, Seoul 136-701, Korea. E-mail: jael@korea.ac.kr