

# 약동학적 파라미터를 이용한 시간경로 마이크로어레이 자료의 군집분석

이효정<sup>1</sup> · 김별아<sup>2</sup> · 박미라<sup>3</sup>

<sup>1</sup>고려대학교 통계학과, <sup>2</sup>식품의약품안전청 신약연구팀, <sup>3</sup>울지대학교 예방의학교실

(2011년 4월 접수, 2011년 6월 채택)

## 요약

시간경로 마이크로어레이 자료 분석의 주요 목적 중의 하나는 유전자들의 시간에 따른 발현수준의 변화를 고려함으로써 발현패턴에 기초한 유전자들의 그룹을 찾기 위한 것으로, 군집분석을 위한 다양한 알고리즘들이 제안되었다. 본 연구에서 시간경로 마이크로어레이 자료에 대한 군집분석을 위해 두 약물체계 간 생물학적 동등성을 평가하기 위한 약동학 시험에서 사용되는 약동학적 파라미터 값에 기초한 군집분석을 제안하였으며 이를 실제 데이터 및 모의실험 자료에 적용하여 유용성을 검토하였다.

주요어: 시간경로 마이크로어레이, 약동학적 파라미터, 군집분석.

## 1. 서론

DNA 마이크로어레이 분석의 주요 목적 중 하나는 대량의 유전자 발현상황을 총체적으로 탐색하여 서로 다른 조직이나 조건의 차이에서 다르게 발현되는 유전자들의 발현패턴을 확인하고 유전자들 사이의 상호작용과 기능의 연계성을 파악하는데 있다 (Chu 등, 1998; Spellman 등, 1998). 시간경로 마이크로어레이 자료(time-course microarray data)의 경우에는 유전자들의 시간에 따른 발현수준의 변화를 고려함으로써 발현패턴에 기초한 유전자들의 그룹을 찾기 위한 것이다. 유사한 유전자 그룹을 찾기 위한 방법으로 계층적 군집분석(hierarchical clustering),  $k$ -평균 군집분석( $k$ -means clustering), 자기조직도(self-organizing map; SOM) 등의 통계적 군집분석 방법이 널리 사용되었다. 그러나 이러한 군집분석 방법들은 각 시점에서 얻어진 자료들이 동일한 개체로부터 발생된 것임을 무시하고 각 시점을 독립적인 변수로 간주한다는 문제점이 있다. 또한 자료가 얻어지는 시점의 순서나 시점 간 간격이 무시되는 결과를 나타낼 수 있다. 이를 보완하기 위해 시점의 특성을 고려하여 기존의 군집분석 방법보다 더 적절한 유전자 그룹을 찾기 위한 방법들이 연구되었다. Hoon 등 (2002)은 선형 스플라인(linear splines)을 이용한 군집분석 방법을 제안하였으며, Peddada 등 (2003)은 순서제한 추론방법(order-restricted inference method)을 이용한 분석 방법을 제안하였다. Schliep 등 (2003)은 시간경로 자료의 시점 의존성을 고려하기 위해 Hidden Markov Models(HMM)를 사용했으며 Luan과 Li (2003)은 B-스플라인 함수에 기초한 혼합효과모형(mixed-effects model)을 이용한 군집방법, Yeung 등 (2001)은 가우시안 혼합

이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No.2011-0004376).

<sup>3</sup>교신저자: (301-832) 대전시 중구 용두동 143-5, 울지대학교 의과대학 예방의학교실, 교수.

E-mail: mira@eulji.ac.kr

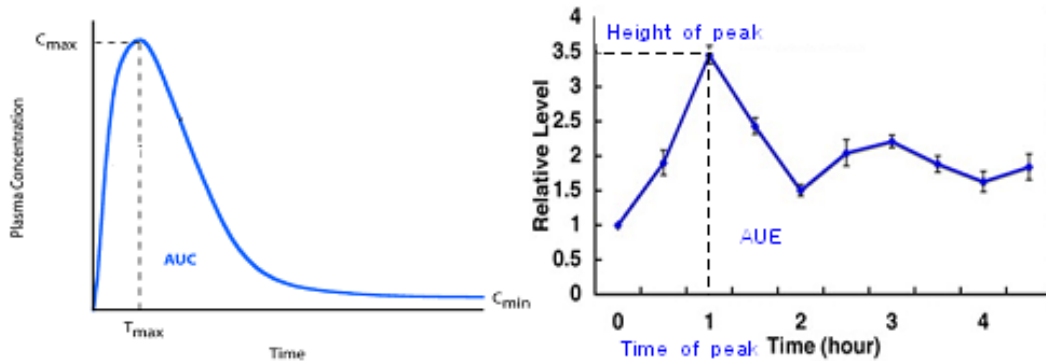


그림 2.1. 시간경로 마이크로어레이 자료에서 PI 정의

모형(Gaussian mixture model)을 이용한 모형 기반 군집방법, Song 등 (2007)은 마이크로어레이 자료 대신 시점의 구간에서 정의된 함수를 이용한 군집방법을 제안하였다. 또한 시간경로 마이크로어레이 자료의 백분위수 또는 표준편차에 기초하여 자료를 순위로 변환한 후 이를 이용하여 군집분석을 실시하는 방법이 제안되었다 (Yi 등, 2009).

본 연구에서는 시간경로 마이크로어레이 자료의 군집분석을 위한 보다 간단하고 유용한 방법으로서, 두 약물제제 간 생물학적 동등성을 평가하기 위한 약동학 시험(pharmacokinetics study)에서 사용되는 분석방법을 벤치마킹하였다. 즉, 마이크로어레이 자료를 약동학 시험에서 사용되는 약동학적 파라미터(pharmacokinetic parameter)들에 대응되는 요약통계량으로 대체하고, 대체된 값들을 군집분석에 적용하는 방법을 제안하였다. 마이크로어레이 연구에서는 실험에 따라 약간의 시차를 두고 반응할 수 있으므로 전반적인 발현패턴의 유사성이 더 중요하다고 할 수 있는데, 고전적인 군집분석으로는 이러한 점을 반영하지 못하게 된다. 여기서는 특정한 확률적 모델을 가정하지 않으면서, 시간에 따른 패턴을 고려할 수 있는 방안을 제시하였다. 먼저 약동학적 파라미터를 이용한 방법과 유용성 평가 방법을 소개하고, 제안된 방법을 실제 데이터 및 모의실험 자료에 적용하여 방법의 타당성과 유용성을 살펴볼 것이다.

## 2. 유전적 연관성 분석방법

### 2.1. 약동학적 파라미터를 이용한 자료변환

두 약물제제 간 생물학적 동등성을 평가하기 위한 약동학 시험에서 사용되는 대표적인 약동학적 파라미터는 약물의 시간 별 혈중 농도 곡선으로부터 얻을 수 있는 최고혈중농도( $C_{max}$ ), 최고혈중농도에 도달하는 시간인 최고혈중농도 도달시간( $T_{max}$ ), 그리고 혈중농도 곡선의 면적인 혈중농도-시간곡선하면적(area under the plasma or blood concentration-time curve; AUC)이다 (그림 2.1).  $C_0, C_1, \dots, C_k$ 를 시간  $t_0, t_1, \dots, t_k$ 에서 얻어진 혈중농도라고 할 때 혈중농도-시간곡선하면적은 다음 공식에 의해 얻어질 수 있다.

$$AUC(0 - k) = \sum_{i=1}^k \frac{C_i + C_{i-1}}{2} \times (t_i - t_{i-1}).$$

본 연구에서는 시간경로 마이크로어레이 자료에서 각 약동학적 파라미터  $C_{max}, T_{max}, AUC$ 에 대응되는 값인 마이크로어레이 자료의 peak 높이(height of peak), peak가 발생한 시점(time at peak), 발현

표 2.1. 군집 결과 비교를 위한 이차원 분할표

Cluster1	Cluster2				Sums
	$v_1$	$v_2$	$\dots$	$v_C$	
$\mu_1$	$n_{11}$	$n_{12}$	$\dots$	$n_{1C}$	$n_{1.}$
$\mu_2$	$n_{21}$	$n_{22}$	$\dots$	$n_{2C}$	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\mu_R$	$n_{R1}$	$n_{R2}$	$\dots$	$n_{RC}$	$n_{R.}$
Sums	$n_{.1}$	$n_{.2}$	$\dots$	$n_{.C}$	$n_{..} = n$

패턴에 따른 면적(area under expression profile; AUE)을 계산한 후, 이들 3가지 요약통계량을 패턴지수(Pattern Indices; PI)로 정의하고, 이렇게 구성된 PI 자료에 군집분석을 적용하고자 한다.

2.2. 유용성 평가

PI를 이용한 군집분석 결과의 유용성을 평가하기 위한 방법으로 두 개의 군집 알고리즘에 의한 군집 결과의 일치도를 평가하기 위한 방법인 Rand Index (Rand, 1971)와 Adjusted Rand Index (Hubert와 Arabie, 1985)를 사용하였다. 또한 모의실험에서는 마이크로어레이 자료의 여러 패턴에 따라 모의자료를 생성함으로써 각 패턴에 대한 그룹을 알 수 있는 경우이므로 분류기법의 타당성 평가를 위해 가장 많이 사용되는 오분류율(misclassification error rate)도 계산하였다.

Rand Index(RI)의 계산방법은 다음과 같다.  $n$ 개의 유전자 관측치의 집합  $S = \{O_1, \dots, O_n\}$ 으로 가정하고 두개의 군집 알고리즘에 의한 군집의 집합을  $U = \{u_1, \dots, u_R\}$ ,  $V = \{v_1, \dots, v_C\}$ 라고 하면 군집 결과를 표 2.1에 의해서 나타낼 수 있다(단,  $S = \bigcup_{i=1}^R u_i = \bigcup_{j=1}^C v_j$  and  $u_i \cap u_{i'} = \phi = v_j \cap v_{j'}$  for  $1 \leq i \neq i' \leq R, 1 \leq j \neq j' \leq C$ ).

이 때 한 쌍의 유전자 관측치가  $U$ 와  $V$ 에서 같은 군집에 포함될 경우의 수를  $a$ 라고 하고,  $U$ 에서는 같은 군집에 속하면서  $V$ 에서 다른 군집에 포함될 경우의 수를  $b$ ,  $U$ 에서는 다른 군집에 속하면서  $V$ 에서 같은 군집에 포함될 경우의 수를  $c$ ,  $U$ 와  $V$  모두에서 다른 군집에 포함될 경우의 수를  $d$ 라고 하자. 여기서  $a$ 와  $d$ 는 군집 결과의 일치수로 해석될 수 있으며,  $b$ 와  $c$ 는 군집 결과의 비일치수로 해석될 수 있다. RI는 전체 유전자 관측치의 쌍의 수에 대한 일치한 유전자 쌍의 수로,  $(a + d)/(a + b + c + d)$ 로 나타낼 수 있다. 즉 RI는 0과 1 사이의 값을 가지게 되며 두 개의 군집 결과가 완전히 일치할 때 RI는 1의 값을 가진다. 2개의 임의 군집에 대한 RI의 기대값이 상수값이 되지 못하는 문제점을 보완하여 일반화 초기하분포(generalized hypergeometric distribution)의 가정 하에 Adjusted Rand Index(ARI)가 제안되었다. RI 및 ARI는 다음 공식에 의해 계산될 수 있으며, 값이 클수록 군집 알고리즘 간에 일치도가 높다고 할 수 있다 (Rand, 1971; Hubert와 Arabie, 1985).

$$RI = \frac{\binom{n}{2} + \sum_{i=1}^R \sum_{j=1}^C n_{ij}^2 - \frac{1}{2} \left[ \sum_{i=1}^R n_{i.}^2 + \sum_{j=1}^C n_{.j}^2 \right]}{\binom{n}{2}},$$

$$ARI = \frac{\sum_{i=1}^R \sum_{j=1}^C \binom{n_{ij}}{2} - \left[ \sum_{i=1}^R \binom{n_{i.}}{2} \sum_{j=1}^C \binom{n_{.j}}{2} \right] / \binom{n}{2}}{\frac{1}{2} \left[ \sum_{i=1}^R \binom{n_{i.}}{2} + \sum_{j=1}^C \binom{n_{.j}}{2} \right] - \left[ \sum_{i=1}^R \binom{n_{i.}}{2} \sum_{j=1}^C \binom{n_{.j}}{2} \right] / \binom{n}{2}}.$$

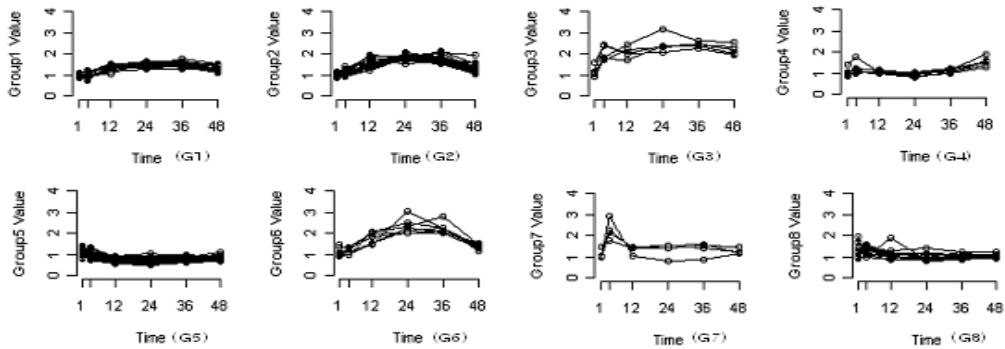


그림 3.1. 원시자료의 군집분석 결과

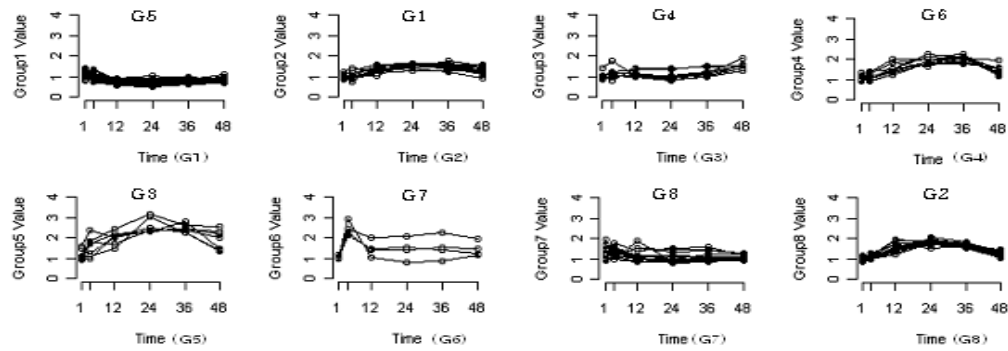


그림 3.2. PI의 군집분석 결과

원시자료를 이용한 군집분석 결과와 PI를 이용한 군집분석 결과의 일치도를 평가하는 데 RI 및 ARI를 사용할 것이다. 또한 모의실험에서는 PI의 군집분석 결과에 따른 그룹과 실제 그룹과의 일치도를 평가하여 PI를 이용한 군집분석의 유용성을 평가할 것이다.

### 3. 분석 결과

#### 3.1. 유방암자료

유방암 자료는 17 베타-에스트라디올( $\beta$ -estradiol) 또는 에탄올로 처리된 MCF-7 Breast Cancer cell line 연구 데이터로 Lobenhofer 등 (2002)에 의해 수행된 결과이다. 전체 1901개의 유전자로 구성되어 있으며 샘플들은 각 처리 후 1, 4, 12, 24, 36, 48시간에서 배양되어 각 시점에서 8번 반복되었다. Lobenhofer 등 (2002)의 연구에서는  $1901 \times 48$  구조로 이루어진 표준화된 자료를 사용하였다. 본 연구에서는 실제 데이터 분석에 앞서 유전자 선택을 실시하였다. 유전자 선택 방법으로는 일반적으로 많이 사용되고 있는 PAM(Prediction Analysis for Microarrays; Tibshirani 등, 2002) 방법을 이용하였으며 전체적인 오류율이 충분히 낮은 기준(threshold = 2.0)에 해당되는 113개의 유전자를 추출하여 본 논문의 분석에 사용하였다. 또한 각 시점에서 8번 반복 자료의 평균을 이용하였으며, 평균 자료에 대해 로그변환(자연로그)을 실시한 후 분석에 사용하였다. PI 자료는 로그변환 자료에 대한 3가지 요약통계량으로 정의되었다.

선택된 113개의 유전자에 대해서 원시자료 및 PI를 이용하여 군집분석을 실시하였으며, 군집분석 방법

표 3.1. 원시자료 및 PI의 군집분석 결과에 대한 이차원 분할표

Rawdata	PI								Rand Index
	G2	G8	G5	G3	G1	G4	G6	G7	
G1	12			2				2	RI = 0.9504 ARI = 0.8349
G2	1	15				6			
G3			4				1		
G4				7					
G5					38				
G6						4			
G7							3	1	
G8	1							13	

표 3.2. 모의자료 생성시 고려된 발현패턴

Pattern 1a	Convex profiles with the high peaks
Pattern 1b	Right shifting profiles of Group 1a
Pattern 2	Convex profiles with the low peaks
Pattern 3	Steady-increasing-decreasing profiles
Pattern 4	Steady-increasing profiles

은  $k$ -평균 군집분석을 이용하였고 그룹 수는 8개로 가정하였다. 그림 3.1과 그림 3.2는 각각 원시자료 및 PI의 군집분석 결과에 대한 각 그룹 별 유전자들에 대한 프로파일을 나타낸다. PI의 군집분석 결과에서 동일한 유전자 프로파일을 가장 많이 포함하는 원시자료의 유전자 그룹을 그림 3.2의 그래프 상단에 'G#'의 형식으로 표기하였다. 표 3.1은 군집분석 결과 비교를 위한 이차원 분할표이다.

원시자료 및 PI의  $k$ -평균 군집분석 결과는 대체적으로 유사한 결과를 보였으며, RI는 0.9504, ARI는 0.8349로 군집분석 결과의 일치도가 높았다. 즉 마이크로어레이 자료의 3가지 요약통계량인 peak 높이, peak 시점, AUE의 PI를 이용하여 군집분석을 실시할 경우에도 원시자료를 이용한 결과와 유사한 결과를 보였다. 각 그룹별 그래프를 살펴보면 원시자료의 군집분석 결과에서는 그룹 3과 그룹 6의 그래프에서 몇몇 프로파일들의 형태가 다른 프로파일들과 다르게 보이는 경향이 있었는데, 그룹 3에서 가장 높은 peak를 보이는 프로파일과 그룹 6에서 가장 높은 peak를 보이는 2개의 프로파일이 PI를 이용한 군집분석에서는 동일한 그룹 5로 분류되었음을 알 수 있다.

### 3.2. 모의자료

모의자료는 시간경로 마이크로어레이 자료의 대표적인 5가지 발현패턴을 가정하였고 (표 3.2), 각 발현 패턴마다 20개의 유전자를 구성하여 총 100개의 유전자 자료를 생성하였다. Sample들은 0.2, 1, 1.5, 2, 3, 4, 5의 7개 시점에서 반복된 것으로 가정했으며, 각 시점에서의 발현값은 각 발현패턴에 따라 최소 0에서 최대 5의 값을 가정하였으며, 첫 번째 시점에서의 발현값은 0.1의 표준편차, 첫 번째 시점을 제외한 모든 나머지 시점에서는 0.6의 표준편차를 가정하였다.

패턴 1a는 시점 1에서 높은 peak(peak의 높이 4 가정)를 보인 후 점점 감소하는 블록 형태의 프로파일이며, 패턴 1b는 패턴 1a와 프로파일 형태 및 peak 높이는 동일하지만 오른쪽으로 이동된 형태의 프로파일이다(시점 1.5에서 peak의 높이 4 가정). 즉 peak가 일어나는 시점이 패턴 1a보다 약간 늦게 발생하는 패턴으로 패턴 1a와 동일한 형태의 유전자 프로파일이라 할 수 있다. 패턴 2는 패턴 1a와는 다르게 낮은 peak를 보인 후 점점 감소하는 블록 형태의 프로파일이며(패턴 1a의 모든 시점에서의 발현값 대비 70%를 가정함) 패턴 3은 발현값이 안정적으로 증가하다가 시점 4에서 peak가 발생한 후 감소하

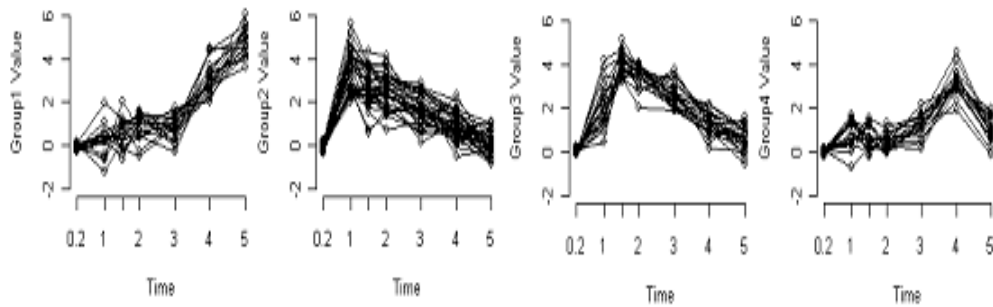


그림 3.3. 원시자료의 군집분석 결과

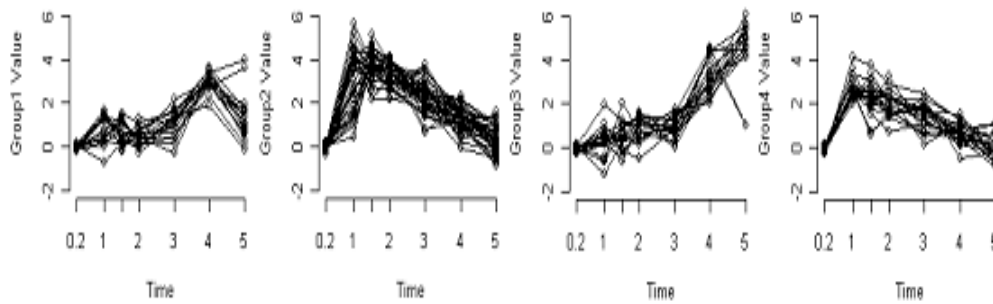


그림 3.4. PI의 군집분석 결과

표 3.3. 실제그룹 vs. 원시자료 군집결과

Real	Rawdata				Rand Index
	3	2	4	1	
1(a, b)	23	17			오분류율 = 17% RI = 0.8523 ARI = 0.6233
2		20			
3			20		
4				20	

는 형태(시점 4에서 peak의 높이 3 가정), 패턴 4는 안정적인 증가패턴을 가지는 형태의 프로파일(시점 1에서 최소 발현값 0, 시점 5에서 최대 발현값 5 가정)이라 할 수 있다. 패턴 1a와 패턴 1b를 그룹 1, 패턴 2를 그룹 2, 패턴 3을 그룹 3, 패턴 4를 그룹 4로 정의하고, 원시자료와 PI자료의 각 군집분석 결과에 따른 군집을 실제그룹과 비교하였으며, 오분류율과 RI, ARI를 계산하였다.

그림 3.3과 그림 3.4는 원시자료 및 PI의 군집분석 결과에 대한 각 군집별 유전자들에 대한 프로파일을 나타내며, 표 3.3과 표 3.4는 실제 그룹과 군집분석 결과에 대한 이차원 분할표이다. 먼저 원시자료의 군집분석 결과를 살펴보면, 실제 그룹 1(패턴 1a, 1b)에 해당되는 유전자 프로파일의 약 42.5%(17/40개)가 군집 2로, 나머지 57.5%(23/40개)가 군집 3으로 분류되어, 이를 다른 그룹으로 분류하고 있음을 알 수 있다. 반면, 패턴 1a와 패턴 2는 peak 높이가 다른 유형의 유전자 프로파일이나 원시자료의 군집분석 결과에서는 대부분 동일한 군집 2에 포함되었다. 패턴 2와 패턴 3, 그리고 패턴 4에 해당되는 유전자 프로파일들의 경우에는 군집 2와, 군집 4, 그리고 군집 1로 각각 100% 분류되어, 전체 오분류율 17%, RI = 0.8523, ARI = 0.6233를 나타내었다. PI를 이용한 군집분석 결과에서는, 패턴 1, 2, 3, 4에 해당되는 90% 이상의 유전자 프로파일이 각 군집 2, 4, 1, 3에 포함되었으며 오분류율 = 9%,

표 3.4. 실제그룹 vs. PI 군집결과

Real	PI				Rand Index
	2	4	1	3	
1(a, b)	37	3			
2	2	18			오분류율 = 9%
3			18	2	RI = 0.9154
4			2	18	ARI = 0.7857

RI = 0.9154, ARI = 0.7857로 나타났다. 즉 PI의 군집분석 결과가 원시자료를 이용한 군집분석 결과 보다 실제 그룹과의 일치도가 높았다.

#### 4. 결론 및 토의

본 연구에서는 시간경로 마이크로어레이 자료에 대한 군집분석을 위해 두 약물제제 간 생물학적 동등성을 평가하기 위한 약동학 시험에서 사용되는 분석방법을 벤치마킹하였다. 약동학 시험에서 사용되는 대표적인 약동학적 파라미터  $C_{max}$ ,  $T_{max}$ , AUC에 대응되는 값인 마이크로어레이 자료의 peak 높이, peak가 발생한 시점, 발현패턴에 따른 면적을 PI로 정의하고, PI를 군집분석에 적용하여 이에 대한 유용성을 검토하였다.

실제 유방암 데이터의 PI를 이용하여 군집분석을 실시한 결과 원시자료를 이용한 결과와 마찬가지로 좋은 군집 결과를 보이고 있음을 알 수 있었다. 또한 모의실험을 통해 유전자 프로파일의 형태는 유사하나 시점이 다른 경우, 즉 프로파일 간 거리 간격이 크나 형태는 유사한 경우에 있어서 원시자료를 이용한 군집분석 방법보다 PI를 이용한 군집분석 방법이 더 정확한 군집 결과를 보이는 것을 알 수 있었다. 이와 같이 보다 적은 변수를 사용하게 되는 PI 군집분석 결과가 모든 변수를 그대로 이용하는 군집분석 결과보다 오분류율 기준에서 더 좋은 결과를 나타낼 수 있었던 이유는, 유전자 프로파일의 형태는 유사하나 시점이 약간 차이 나게 발현되는 경우에 있어서 단순히 원자료를 이용하게 되는 경우 밀린 시점이 다른 변수로 정의되면서 결과에 큰 영향을 미치기 때문이다. 이러한 상황은 마이크로어레이 유전자 실험에서 흔히 발생될 수 있는 것이라 할 수 있다.

PI는 원시자료를 3차원으로 축소, 요약한 자료라고 할 수 있으며, 차원 축소로 인하여 정보의 손실이 발생할 수 있다는 것이 본 연구방법의 제한점이라고 할 수 있다. 차원 축소로 인한 정보의 손실을 최소화하기 위해서는 유전자 자료의 성격에 따라 유전자 프로파일의 패턴을 대표할 수 있는 값들을 추가적으로 정의할 수도 있을 것이다.

#### 참고문헌

Chu, S., DeRisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, P. O. and Herskowitz, I. (1998). The transcriptional program of sporulation in budding yeast, *Science*, **282**, 699–705.

Hoon, D., Imoto, S. and Miyano, S. (2002). Statistical analysis of a small set of time-ordered gene expression data using linear splines, *Bioinformatics*, **18**, 1477–1485.

Hubert, L. and Arabie, P. (1985). Comparing partitions, *Journal of the Classification*, **2**, 193–218.

Lobenhofer, E. K., Bennett, L., Cable, P. L., Li, L., Bushel, P. R. and Afshari, C. A. (2002). Regulation of dna replication fork genes by 17beta-estradiol, *Molecular Endocrinology*, **16**, 1215–1229.

Luan, Y. and Li, H. (2003). Clustering of time-course gene expression data using a mixed-effects model with B-splines, *Bioinformatics*, **19**, 474–482.

- Peddada, S. D., Lobenhofer, E. K., Li, L., Afshari, C. A., Weinberg, C. R. and Umbach, D. M. (2003). Gene selection and clustering for time-course and dose-response microarray experiments using order-restricted inference, *Bioinformatics*, **19**, 834–841.
- Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods, *Journal of American Statistical Association*, **66**, 846–850.
- Schliep, A., Schönhuth, A. and Steinhoff, C. (2003). Using hidden Markov models to analyze gene expression time course data, *Bioinformatics Supplement*, **19**, i255–263.
- Song, J. J., Lee, H. J., Morris, J. S. and Kang, S. (2007). Clustering of time-course gene expression data using functional data analysis, *Computational Biology and Chemistry*, **31**, 265–274.
- Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Botstein, D. and Futcher, B. (1998). Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization, *Molecular Biology of the Cell*, **12**, 3273–3297.
- Tibshirani, R. J., Hastie, T. J., Narasimhan, B. and Chu, G. (2002). Diagnosis of multiple cancer types by shrunken centroids of gene expression, *Proceedings of the National Academy of Sciences*, **99**, 6567–6572.
- Yeung, K. Y., Fraley, C., Murua, A., Raftery, A. E. and Raftery, W. L. (2001). Model-based clustering and data transformations for gene expression data, *Bioinformatics*, **17**, 977–987.
- Yi, S.-G., Joo, Y.-J. and Park, T. (2009). Rank-based clustering analysis for the time-course microarray data, *Journal of Bioinformatics and Computational Biology*, **7**, 75–91.



# Clustering of Time-Course Microarray Data Using Pharmacokinetic Parameter

Hyo-Jung Lee<sup>1</sup> · Peol-A Kim<sup>2</sup> · Mira Park<sup>3</sup>

<sup>1</sup>Department of Statistics, Korea University

<sup>2</sup>Pharmaceuticals & Medical Devices Research Department, KFDA

<sup>3</sup>Department of Preventive Medicine, Eulji University

(Received April 2011; accepted June 2011)

---

## Abstract

A major goal of time-course microarray data analysis is the detection of groups of genes that manifest similar expression patterns over time. The corresponding numerous cluster algorithms for clustering time-course microarray data have been developed. In this study, we proposed a clustering method based on the primary pharmacokinetic parameters in the pharmacokinetics study for assessment of pharmaceutical equivalents between two drug products. A real data and a simulation data was used to demonstrate the usefulness of the proposed method.

**Keywords:** Time-course microarray data, pharmacokinetic parameter, clustering.

---

---

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the ministry of Education, Science and Technology (2011-0004376).

<sup>3</sup>Corresponding author: Associate Professor, Department of Preventive Medicine, Eulji University, Daejeon 301-832, Korea. E-mail: mira@eulji.ac.kr