

## Similarity Analysis between Total RNA and Amplified RNA Using Entropy Measure<sup>†</sup>

### 엔트로피 척도를 이용한 전체 RNA와 증폭 RNA의 유사성 분석

Chanho Park<sup>1\*</sup>, Sung-Bae Cho<sup>1</sup>, Ji Hye Shin<sup>2</sup>, Sang Cheol Kim<sup>2</sup>, Min Young Seo<sup>2</sup>, Sang Hwa Yang<sup>2</sup>, Sun Young Rha<sup>2</sup>, Hyun Cheol Chung<sup>2</sup>

<sup>1</sup> Department of Computer Science, Yonsei University, Seoul, Korea

<sup>2</sup> Cancer Metastasis Research Center, College of Medicine, Yonsei University, Seoul, Korea

\*To whom correspondence should be addressed. E-mail: cpark@sclab.yonsei.ac.kr

#### Abstract

암의 조기 발견 및 예후 예측을 위하여 마이크로어레이 데이터를 이용할 수 있다. 하지만 이를 분석하기 위해서는 40 $\mu$ g 이상의 RNA 샘플이 필요한데, 실제 임상 시료를 사용하는 경우 요구되는 충분한 양을 얻기가 어려운 단점이 있다. 따라서 소량의 RNA 샘플을 채취한 후 PCR 증폭 과정을 통하여 요구되는 양의 샘플을 얻을 수 있는 RNA 증폭 방법이 시도되고 있고, 이를 마이크로어레이 실험에 이용하기 위해서는 증폭 전후의 유사성이 보장되어야 한다. 본 논문에서는 증폭 RNA와 전체 RNA의 유사성을 비교하기 위한 새로운 방법으로 엔트로피 기반의 방법을 제시한다. 아울러 다양한 조건에 따라서 엔트로피 값을 측정하여 세포주와 조직에서 엔트로피 값이 어떻게 사용될 수 있는지 체계적인 분석을 하였다.

#### Introduction

암은 인간이 사망하는 주된 원인 중 하나이지만, 조기에 발견할 수 있다면 빠르게 대응하여 생존율을 높일 수 있다. 근래에 개발된 마이크로어레이(microarray) 기술은 동시에 수천 개 이상의 유전자 발현 정도를

측정할 수 있게 해 주었고, 이렇게 얻어진 마이크로어레이 데이터로 암의 조기 발견 및 예후 예측이 가능하게 되었다[1, 2, 3].

한편 마이크로어레이 데이터를 분석하기 위해서는 양질의 전체 RNA(total RNA) 샘플을 40 마이크로그램( $\mu$ g) 이상 채취해야 하는데, 진단을 위한 조직 검사와 같이 실제 임상 시료를 사용하는 경우에는 모든 샘플에서 원하는 충분한 양을 얻기가 힘들다. 따라서 소량의 RNA 샘플을 채취하여 PCR(중합효소연쇄반응, polymerase chain reaction) 증폭 과정을 통하여 얻어진 증폭 RNA(amplified RNA)로 마이크로어레이를

<sup>†</sup> This study was supported by the grant of the Korea Health 21 R&D Project, Ministry of Health & Welfare, Republic of Korea(02-PJ1-PG1-CH04-0001) and the Korea Sciences and Engineering Fund through the Cancer Metastasis Research Center at Yonsei University.

수행하는 방법이 제시되고 있다[4, 5, 6]. 이렇게 얻어진 증폭 RNA를 사용하는 경우, 증폭을 위한 PCR과정에서 RNA transcript의 질적 혹은 양적 변화가 발생할 수 있으므로, 그 결과가 원래의 전체 RNA와 얼마나 비슷한 양상을 보이느냐 하는 것이 매우 중요한 관건이다. 현재까지는 전체 RNA와 증폭 RNA와의 유사성을 비교하기 위하여 마이크로어레이 결과의 상관계수를 측정하거나 유의한 유전자들을 찾아내서 비교하기, 군집(clustering) 분석을 통한 비교 등 몇 가지 분석이 이루어져 왔으나, 이것만으로는 그 유사성을 완벽하게 설명하기 어렵다[7].

한편, 기계학습 분야에서 널리 사용되어 오던 척도인 엔트로피(entropy)는 집합의 무질서한 정도를 측정하기 위한 방법이다. 한번의 실험으로 수천 개 이상의 결과를 얻게 되는 마이크로어레이 데이터는 각 실험결과를 하나의 집합으로 고려할 수 있으므로 엔트로피를 적용할 수 있다. 본 논문에서는 이 엔트로피 척도를 기반으로 한 전체 RNA와 증폭 RNA를 비교하는 방법을 제안하고자 하며, 이를 통하여 전체 RNA와 증폭 RNA가 얼마나 유사한지에 대한 체계적인 분석을 시행하였다.

## Method

그동안 마이크로어레이에서 전체 RNA와 증폭 RNA를 비교하기 위한 많은 방법이 연구되었고 적용되어 왔다[4, 6, 7]. 이를 크게 나누어 보면 생물학적인 방법과 통계/전산학적(이하 전산학적)인 방법의 두 가지 범주로 나눌 수가 있다.

### 생물학적 방법

**시료 및 전체 RNA 추출** : 정상과 종양조직은 대장암 환자로부터 수술을 통해 얻은 후, -80℃에 보관하였다. 4개의 종양 세포주(AGS, HL-60, HT29, HT1080)는 37℃, 5%,

CO<sub>2</sub>하에서 MEM 배지를 첨가하여 배양하였다. 조직과 세포주로부터 Trizol을 이용하여 전체 RNA를 추출한 후, 4개의 세포주는 동량으로 섞었다. 전체 RNA의 양과 질은 Gene spec I spectrophotometer(Hitachi, inc.)와 Bioanalyzer(Agilent, Inc.)를 이용하여 확인하였다.

**RNA 증폭** : 2μg의 전체 RNA에 T7-oligo(dT) primer와 Superscript II를 첨가하여 역전사 반응시킴으로서 double strand cDNA를 합성하였다. 생성된 double strand cDNA에 T7 MEGAscript™ kit (Ambion, Inc.)로 in vitro transcription을 수행하여 증폭된 mRNA를 합성하였다. 증폭된 mRNA의 양과 질은 Gene spec I spectrophotometer와 Bioanalyzer를 이용하여 확인하였다.

**cDNA 마이크로어레이** : 4μg의 증폭된 mRNA에 random primer와 Superscript II를 첨가하여 역전사시킴으로서 single strand cDNA를 합성하고, test sample은 Cy3-dUTP로, reference sample은 Cy5-dUTP로 labeling하였다. 각각 Cy3-dUTP와 Cy5-dUTP로 labeling된 cDNA를 정제하고 혼합한 후, 17,000개의 유전자가 점적되어 있는 마이크로어레이 (GT-CMRC, 연세의대)을 사용하여 hybridization하였다. Hybridization이 끝난 마이크로어레이는 washing하여 GenePix 4000B scanner(Axon, Inc.)로 이미지를 스캔하고, 데이터를 얻었다.

**마이크로어레이 데이터** : 얻어진 마이크로어레이 데이터는 Within-pin Lowess algorithm[8]을 이용하여 표준화한 후, 계층적 군집분석 (Eisen software, Stanford Univ.)[3]을 통해 전체적인 유전자의 발현 양상을 관찰하였다. 데이터는 세포주와 정상 및 암 조직에 대하여 각각 전체 RNA 샘플과 여러 가지 증폭 조건에 따른 증폭 RNA 샘플들이 있다. 총 샘플의 수는 세포

주가 다섯 개이며 정상 및 암 조직의 경우는 다섯 개의 쌍으로 총 10개였다.

각 실험의 이름은 세포주의 경우 C, aC4, aC2, aC1, 2aC이고, 정상 조직의 경우는 N, aN4, aN2, aN1, 2aN, 암 조직의 경우는 T, aT4, aT2, aT1, 2aT이며, 각 기호의 의미는 표 1과 같다.

표 1. 실험에서 각 기호의 의미

기호	의미	기호	의미
C	cell line (세포주)	a	amplified (증폭)
N	normal (정상조직)	2a	2번 증폭
T	tumor (암조직)	숫자	증폭에 사용된 양( $\mu\text{g}$ )

### 전산학적 방법

**상관계수 측정** : 상관계수분석이란 변수간의 관련성을 분석하기 위해 사용하는 방법으로, 본 논문에서는 피어슨 상관계수, 스피어맨 상관계수 및 코사인 계수를 사용하였다[9]. 상관계수  $r$ 은 -1부터 +1까지의 값을 갖는다.  $r$ 의 절대값이 클수록 두 변수간의 관련 정도가 높은 의미를 가지며 0에 가까울수록 관련 정도는 낮다. 부호는 관련성의 방향을 나타낸다. 전체 RNA의 유전자 벡터인  $X$ 와 증폭 RNA 유전자 벡터인  $Y$ 사이의 피어슨 상관계수는 식 (1)과 같이 정의된다.

$$r_{pearson} = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{(\sum X^2 - \frac{(\sum X)^2}{N})(\sum Y^2 - \frac{(\sum Y)^2}{N})}} \quad (1)$$

한편, 스피어맨 상관계수는 변수의 순위 배열을 사용하여 변수간의 상관관계를 분석하는 방법으로, 두 벡터  $X$ 와  $Y$ 의 순위배열  $Dx$ 와  $Dy$ 를 사용하여 식 (2)와 같이 나타낼 수 있다( $N$ 은 전체 유전자의 수).

$$r_{spearman} = 1 - \frac{6 \sum (Dx - Dy)^2}{N(N^2 - 1)} \quad (2)$$

전체 RNA와 증폭 RNA의 유사성은 두 벡터 사이의 각으로도 나타낼 수 있는데 각이 작을수록 같은 방향을 가리키며 서로 유사하다. 각을 직접 구하기보다는 그 각에 대한 코사인 값을 구하는 것이 쉬우므로 유사도는 코사인 값으로 나타낼 수 있다. 두 벡터  $X, Y$  사이의 코사인 계수는 식 (3)과 같이 나타낼 수 있다.

$$r_{cosine} = \frac{\sum XY}{\sqrt{\sum X^2 \sum Y^2}} \quad (3)$$

**엔트로피 측정** : 엔트로피는 식 (4)와 같이 구할 수 있으며, 집합에 속하는 특정 부집합의 차지하는 비율이 0에 가깝거나 1에 가까울수록 작아지는 특성을 지닌다[10]. 엔트로피 값이 0에 가까울수록 그 집합은 질서가 잡혀있다는 의미가 되며, 집합이 무질서할수록 이 값이 커진다.

$$Entropy = - \sum_{i=1}^c (p_i \log_2 p_i) \quad (4)$$

여기서  $p_i$ 는  $c$ 개의 class가 있을 때, 그중  $i$ 번째 class에 속하는 원소의 비율이며, 엔트로피의 최대값은  $\log_2 c$ 가 된다.

전체 유전자들을 증폭 전후의 변화가 큰 것과 작은 것의 두 개의 집합으로 구분할 경우 엔트로피 값의 분포는 그림 1과 같다. 즉, 전체 유전자에서 전체 RNA와 증폭 RNA간의 발현 비율의 변화가 큰 유전자의 비율을  $p$ 로 놓으면 이 값이 0에 가까울 경우 엔트로피 값은 낮아지고, 이는 증폭 전후에 유전자들의 패턴 변화가 거의 없다는 것을 의미한다. 그리고 대부분 유전자의 발현비율이 크게 변한 경우에도 엔트로피 값은 0에 가까워지는데, 이는 전체적인 발현 패턴 변화가 유사하다는 것을 의미한다. 마지막으로 발현 비율의 변화가 큰 유전자들이 반 정도 있고, 나머지에서 변화가 작을 때 엔트로피는 높은 값을 나타내는데, 이 경우 전체 RNA와 증폭 RNA의 유사성은

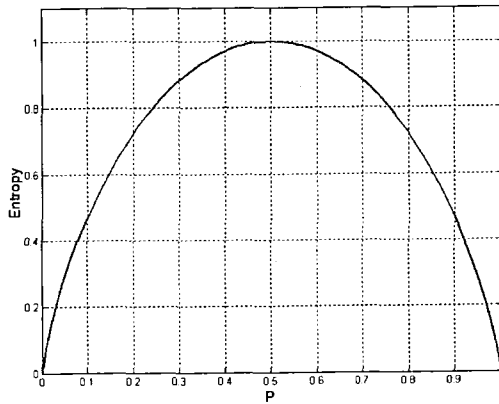


그림 1.  $p$ 의 변화에 따른 엔트로피 값.  $p$ 는 전체 유전자 중 두 샘플 사이에서 변화가 낮은 유전자의 비율

거의 없다고 할 수 있다.

한편,  $p$ 값이 0에서 0.1로 혹은 0.9에서 1로 0.1만큼 변하는 동안 엔트로피 값은 0.47 정도 변하기 때문에, 특히  $p$ 값이 매우 작거나 큰 경우에 엔트로피 값을 이용하면 그 변화를 민감하게 살펴볼 수 있다.

## Results

### 생물학적 실험결과

전체 RNA의 input 양에 따른 amplification yield(증폭 효율) : 전체 RNA의 입력 양에 따른 증폭 RNA의 산출량을 표 2와 같이 구하였다. 1회와 2회 증폭한 경우 각각  $1\sim 3 \times 10^3$ 배,  $6.6\sim 7.1 \times 10^5$ 배의 증폭 효율을 얻었다.

표 2. Total RNA의 input 양에 따라 증폭된 mRNA의 양(단위 :  $\mu\text{g}$ )

증폭회수	1회		2회		
입력 양	4	2	1	0.5	0.5
세포주	46.6	38.1	26.1	10.0	65.8
정상 조직	49.2	35.6	29.9	9.2	77.5
암 조직	56.6	44.5	21.0	10.0	65.8

전체 RNA와 증폭 RNA를 이용한 마이크로어레이 데이터의 scatter plot : 그림 2에서 알 수 있듯이 마이크로어레이 데이터를

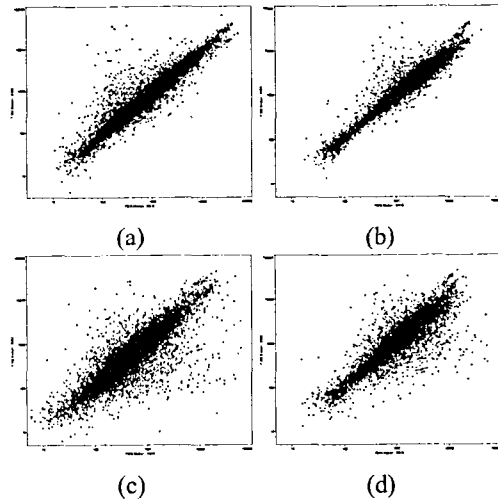


그림 2. 마이크로어레이 데이터의 scatter plot (a) 세포주-전체 RNA, (b) 세포주-증폭 RNA, (c) 조직-전체 RNA, (d) 조직-증폭 RNA

scatter plot으로 나타내었을 때, 조직(c, d)이 세포주(a, b)보다 분포가 퍼져있는 모습이 관찰되었으며, 에서 전체 RNA(a, c)에 비하여 증폭 RNA(b, d)를 사용한 경우에 유전자 분포의 분산이 감소하였다.

전반적인 발현 양상 비교 : 전체 RNA와 증폭 RNA간의 전반적인 발현 양상을 그림 3의 마이크로어레이 비교와 그림 4의 계층적 군집 분석을 통하여 비교하였다.

전체 RNA보다 증폭 RNA를 사용하였을 때, 전체적으로 cy5 및 cy3로 염색한 색이 또렷하게 나타난 것을 그림 3을 통하여 확인할 수 있었다. 그림에서 한번 증폭한 것과 두 번 증폭한 것의 차이는 거의 없어보였으며, 전체 RNA와 증폭 RNA의 염료 세기의 차이는 조금 있었지만 전체적인 패턴은 비슷하게 나타났다. 그림 4는 정상조직과 암조직에서 차이 나게 발현하는 256개의 유의한 유전자들을 선별하여 계층적 군집분석을 한 결과이다. 그림 3의 결과와 마찬가지로 1번 증폭한 것과 2번 증폭한 것의 차이는 거의 없어보였으며, 증폭 RNA와 전

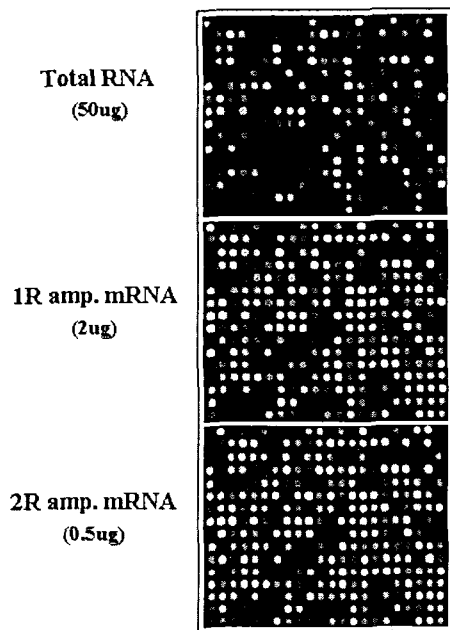


그림 3. 1개의 동일 암조직을 이용하여 전체 RNA, 1회 및 2회 증폭한 RNA를 사용한 마이크로어레이 이미지 비교  
 체 RNA는 비슷한 가운데 약간의 차이를 보여주었다.

### 전산학적 실험결과

전산학적인 실험으로는 기존 방법인 상관계수를 비교하는 것과 새롭게 제시한 엔트로피를 측정하는 방법을 사용하였다. 상관계수를 측정하는 경우 두 샘플의 발현 값만 알고 있으면 수식에 대입하여 구할 수 있지만, 엔트로피를 구할 경우에는 변화의 유무를 가릴 수 있는 임계값이 필요하다. 아직까지 어떤 값이 임계값으로 가장 적당하다고 알려진 바가 없기 때문에, 본 논문에서는 임계값  $k$ 를 1.5, 2, 3, 4로 변화시켜가며 측정하였다. 임계값  $k$ 의 의미는 특정 유전자에서 전체 RNA와 증폭 RNA간의  $cy5/cy3$  비율의 변화가  $k$ 보다 작으면 그 유전자는 변화가 없는 그룹에 속한다는 것이고  $k$  이상일 때에는 변화가 큰 그룹에 속한다는 것이다.



그림 4. 유의한 유전자들에 대한 계층적 군집 분석 결과

상관계수 비교 : 세포주와 정상 및 암 조직에 대하여 전체 RNA와 증폭 RNA간의 상관계수를 측정하였고, 평균값을 표 3과 같이 얻었다.

표 3. 전체 RNA와 증폭 RNA간의 평균 상관계수 값

	세포주	정상 조직	암 조직
pearson	0.80	0.89	0.86
spearman	0.61	0.84	0.78
cosine	0.80	0.89	0.86

피어슨 상관계수와 코사인 계수의 경우 소수 둘째자리는 같은 값이 나왔으며, 발현 값 대신 순위 벡터를 사용하는 스피어맨 상관관계의 경우 가장 낮은 값을 보여주었다. 전체적으로 조직에서 세포주보다 높은 상관관계를 보여주었으며, 세포주에서 각 상관계수의 변화 정도가 가장 컸다.

*k* 값에 따른 평균 엔트로피 비교 : 세포주와 정상 및 암 조직의 *k* 값에 따른 엔트로피는 표 4와 같았다.

표 4. *k* 값에 따른 평균 엔트로피

<i>k</i>	세포주	정상 조직	암 조직
1.5	0.53	0.73	0.75
2	0.19	0.40	0.36
3	0.05	0.18	0.15
4	0.02	0.10	0.08

전체적으로 정상 조직과 암 조직은 비슷한 결과를 보여주었고, 세포주에서는 조직에서보다 낮은 엔트로피 값을 보여주었다. 이는 세포주에서 조직보다 전체 RNA와 증폭 RNA 사이의 변화가 작다는 것을 의미한다. 한편 *k* 값이 커질수록 임계값 내에서 변하는(즉, 변화가 적은 것으로 판별된) 유전자의 비율이 압도적으로 우세해지기 때문에 엔트로피 값이 감소하는 경향을 보여주었다.

input RNA의 양에 따른 엔트로피 측정 : 증폭에 사용된 RNA의 양을 1μg, 2μg, 및 4μg로 각각 달리해가며 실험한 후 엔트로피 값을 측정한 결과는 표 5와 같다.

표 5. *k*와 input RNA의 양에 따른 엔트로피(단위: μg)

<i>k</i>	세포주			조직		
	1	2	4	1	2	4
1.5	0.52	0.46	0.64	0.73	0.71	0.75
2	0.21	0.18	0.22	0.37	0.35	0.40
3	0.05	0.05	0.05	0.16	0.15	0.18
4	0.03	0.02	0.02	0.09	0.08	0.11

표에서 알 수 있듯이 *k*가 1.5일 때는 세포주에서 증폭에 사용된 전체 RNA의 양에 따른 엔트로피 값의 변화가 컸다. 그러나 *k*가 2 이상일 때는 변화가 많이 감소하였다. 반면, 조직에서는 각 *k* 값에 대하여 전체적으로 비슷한 엔트로피 값을 얻었다.

증폭 회수에 따른 엔트로피 비교 : RNA의 증폭 회수에 따른 엔트로피 측정 실험결과

는 표 6과 같다.

표 6. 증폭 회수에 따른 엔트로피

<i>k</i>	세포주		조직	
	1st	2nd	1st	2nd
1.5	0.52	0.49	0.73	0.74
2	0.21	0.16	0.37	0.38
3	0.05	0.05	0.16	0.16
4	0.03	0.02	0.09	0.09

세포주와 조직 모두 한번 증폭했을 때와 두 번 증폭했을 때의 엔트로피 값의 차이는 관찰되지 않았다.

*p*에 따른 *cy5/cy3* 비율 차이의 임계값 : 표 7은 전체 RNA와 증폭 RNA 사이의 *cy5/cy3* 비율 차이가 작은 유전자들의 비율을 80%, 85%, 90%, 95%, 99%로 정하였을 때의 비율 차이 및 이를 밑이 2인 로그값으로 변환시킨 값을 정상 조직, 암 조직, 세포주에 대해서 구한 결과이다.

표 7. *p*가 정해진 상태에서의 *cy5/cy3* 비율 차이의 임계값

<i>p</i>	정상 조직		암 조직		세포주	
	차이	log <sub>2</sub>	차이	log <sub>2</sub>	차이	log <sub>2</sub>
80%	1.52	0.60	1.52	0.60	1.37	0.45
85%	1.64	0.71	1.62	0.70	1.43	0.52
90%	1.83	0.87	1.79	0.84	1.51	0.59
95%	2.31	1.21	2.22	1.15	1.79	0.84
99%	4.38	2.13	4.06	2.02	2.64	1.40

세포주에서는 비율의 차이가 1.37배 이하인 유전자의 비율이 80%이고, 정상 및 암 조직에서는 비율 차이가 1.52배 이하인 유전자의 비율이 80%로 나왔는데, 이는 앞의 결과와 마찬가지로 세포주에서 전체 RNA와 증폭 RNA간의 차이가 적다는 것을 의미한다. 비율을 높게 잡을수록 세포주와 조직에서 사용될 수 있는 임계값의 차이가 벌어지는 모습을 관찰 할 수 있었는데, 이는 증폭에 따라 *cy5/cy3* 비율의 값이 크게 변하는 유전자의 수가 세포주보다 조직에서 많다는 것을 의미한다.

## Discussion

전체 RNA와 증폭 RNA의 연관관계를 마이크로어레이 실험의 결과를 토대로 엔트로피 척도를 사용하여 분석하였고, 그 결과 기존의 분석방법과는 다른 방법으로 유사성을 관찰하였다. 또한 엔트로피는 전체 RNA와 증폭 RNA가 유사할 때뿐만 아니라 전체적으로 cy5/cy3 비율이 높게 증가하거나 감소한 경우에도 0에 가까운 값을 나타내기 때문에 유용한 척도임을 확인하였다.

엔트로피를 사용할 경우, 변화의 크고 작음을 구분할 수 있는 임계값  $k$ 의 설정이 매우 중요한데, 지금까지 특별히 알려진 바는 없었다. 한편, 상관계수의 경우 조직에서는 전체적으로 고른 값이 나왔으나 세포주에서는 변화가 컸다. 이 결과와 비교하여 조직에서는 엔트로피 값의 변화가 작고, 세포주에서는 큰 경우에 대한  $k$ 값을 찾아본 결과  $k = 1.5$ 일 때가 가장 유사한 경우임을 표 5를 통하여 알 수 있었다. 따라서 1.5배를 변화의 기준으로 고려할 수 있다.

전체적으로 엔트로피 값은 세포주에서 작고 조직에서 커서 세포주에서의 전체 RNA와 증폭 RNA가 유사하게 판별된 반면, 상관계수는 반대의 결과가 관찰되었다. 일단 세포주와 조직에서 서로 다른 값들이 나온 것은, 세포주와 조직을 연구할 때 다른 기준을 적용시키는 것이 필요하다는 것을 유추할 수 있다. 그리고 둘 사이에 엔트로피와 상관계수에서 다른 결과가 나온 것은 세포주는 증폭에 따른 cy5/cy3 값의 변화가 크지 않지만, 각 유전자에 대하여 값의 순위를 매겼을 때 그 차이가 크기 때문이고, 조직에서는 그 값의 변화가 큰 반면, 전체적인 값의 순위 차이는 크지 않기 때문이라고 추정할 수 있다.

하지만 현재의 기준으로는 엔트로피와 상관계수의 절대적인 비교는 매우 힘들다.  $k$  값을 결정하고 엔트로피를 마이크로어레이의 분석에 이용함에 있어 생물학적 검증과

이에 근거한 두 군 간의 함수 관계에 대한 추후 연구가 필요하다.

## References

- [1] T. R. Golub, D. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. Mesirov, H. Coller, M. Loh, J. Downing, M. Caligiuri, C. Bloomfield and E. Lander, Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring, *Science*, 286, pp. 531-537, 1999.
- [2] J. Khan, J. S. Wei, M. Ringer, L. H. Saal, M. Ladanyi, F. Westermann, F. Berthold, M. Schwab, C. R. Antonescu, C. Peterson and P. S. Meltzer, Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nature Medicine*, 7, pp. 673-679, 2001.
- [3] M. B. Eisen, P. T. Spellman, P. O. Brown and D. Bostein, Cluster analysis and display of genome-wide expression patterns, *Proc. Natl. Acad. Sci.*, 95, pp. 14863-14868, 1998.
- [4] C. Pabon, Z. Modrusan, M. V. Ruvolo, I. M. Coleman, S. Daniel, H. Yue, L. J. Arnold, Jr. and M. A. Reynolds, Optimized T7 Amplification System for Microarray Analysis, *BioTechniques*, 31(4), pp. 874-879, 2001.
- [5] J. Phillips and J. H. Eberwine, Antisense RNA amplification: a linear amplification method for analyzing the mRNA population from single living cells, *Methods*, 10, pp. 283-288, 1996.
- [6] E. Wang, L. D. Miller, G. A. Ohnmacht, E. T. Liu and F. M. Marincola, High-fidelity mRNA amplification for gene profiling. *Nat.*

- Biotechnol.*, 18, pp. 457-459, 2000.
- [7] H. Zhao, T. Hasie, M. L. Whitfield, A. L. Borresen-Dale and S. S. Jeffrey, Optimization and evaluation of T7 based RNA linear amplification protocols for cDNA microarray analysis. *BMC Genomics*, 3, pp. 31-45, 2002.
- [8] Y. H. Yang, S. Dudoit, P. Luu, D. M. Lin, V. Peng and J. Ngai, Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation, *Nucl. Acids. Res.* 30, 2002.
- [9] C. Park and S. -B. Cho, Genetic Search for Optimal Ensemble of Feature-Classifer Pairs in DNA Gene Expression Profile, *Int. Joint Conf. on Neural Networks*, pp. 1702-1707, 2003.
- [10] T. M. Michell, *Machine Learning*. Carnegie Mellon Univ., 1997.