

기/획/특/집

마이크로어레이 실험과 이미지를 위한 품질제어

김판규¹, 조환규^{1,2}

¹부산대학교 전자계산학과, ²부산대학교 정보컴퓨터공학부

Quality Control for Microarray Experiments and Images
Pan-Gyu Kim¹, Hwan-Gue Cho^{1,2}

¹Department of Computer Science, Pusan National University,
Busan 609-735, Korea

²Department of Computer Science and Engineering, Pusan
National University, Busan 609-735, Korea

Key Words : microarray, quality measure, quality control

1. INTRODUCTION

대용량의 유전자 발현 마이크로어레이를 이용한 대용량 처리(high-throughput) 실험 데이터의 분석은 현대 생명과학에서 점점 더 중요해지고 있다. 마이크로어레이는 한번에 수만 혹은 수십 만개의 유전자 발현 정보나 유전자형(genotype)을 얻을 수 있다. 마이크로어레이 연구가 활발해지면서, 분석 되어야 할 마이크로어레이 이미지의 수가 급속히 증가하며 또한 마이크로어레이 제조 기술의 발달로 인해서 하나의 슬라이드가 포함하고 있는 스팟의 수도 점점 증가하고 있다. 이에 따라 슬라이드 내의 각 스팟에 대한 강도를 측정함으로써 유전자의 발현값을 조사하는 마이크로어레이 실험 및 결과에 대한 검증이 필요하게 되었다.

마이크로어레이를 통한 유전자 발현실험이 많이 이루어지고 있지만, 아직까지 마이크로어레이 실험은 여러 가지 실험환경이나 조건에 따라 적지 않은 오류를 가지게 된다. 이러한 오류는 마이크로어레이 실험 후 이미지 분석 시 많은 어려움을 주게 되며, 또한 조절 유전자(control gene)가 특정 영역에 집중되어 있거나, 강하게 발현되는 유전자들이 밀집되어 있으면, 서로간의 간섭으로 인하여 분석 시 어려움이 따른다. 마이크로어레이를 통한 유전자 발현 실험은 실험 환경과 조건에 따라 다양한 형태의 결과를 나타내고, 실제로 실험이 제대로 이루어지지 않아서 이미지 분석 결과가 전혀 의미가 없는 경우도 있다. 이 문제점을 해결하기 위해서 마이크로어레이

이 실험의 완성도를 측정하는 평가함수가 필요하다.

본 논문에서는 각 마이크로어레이 이미지의 몇 가지 통계량을 이용하여 이미지들 사이의 품질을 비교하고, 분석된 마이크로어레이 이미지 결과에 대하여 그 품질을 측정하는 Quality Measure를 정의하여 각각의 마이크로어레이 실험에 대한 품질을 정량화하여 나타낸다. 전체 실험과 한 이미지에 대한 품질을 평가하기 위해서 메타그리드를 이용하여 이미지내의 블럭들과 스팟들간의 간격에 대한 통계량을 측정하고, 스팟들의 발현값에 대한 정확도를 측정하기 위한 품질평가함수들을 정의한다. Hautaniemi 등은 마이크로어레이 실험의 정확도에 영향을 미치는 여러 가지 요소들을 가정하고, 베이지안 네트워크를 통해서 실험의 정확도에 가장 많은 영향을 주는 요소들을 알아보는 모델을 제시하였다(Hautaniemi et al., 2003). 그러나, 이 모델은 각각의 스팟에 대한 품질평가를 측정해보는 것으로 이미지 전체에 대한 품질평가를 나타내기에는 미흡하다.

일반적으로 마이크로어레이는 설계, 실험, 분석과 같은 세가지 단계를 거친다. 이러한 세가지 과정이 올바르게 이루어져야만 좋은 quality의 마이크로어레이라고 한다. 즉, 설계가 잘 되어야만 각각의 스팟의 위치를 쉽게 알 수 있어서 그리딩이 쉽고, foreground와 background부분의 분리시에 간섭효과가 적다. 또, 실험이 잘 되어야 각 스팟의 크기와 모양이 일정하게 생성되어 정확한 결과의 발현값을 얻을 수 있다. 이때 스팟들의 위치가 일정하지 않을 경우, 이미지 분석시에 스팟의 그리드 위치가 올바르게 구해지지 않게 되므로 잘못된 background강도를 추출할 수도 있다.

좋은 마이크로어레이를 만들어서 실험하기 위해서는 앞에서 이야기한 세가지 과정이 잘 되어야하며, 이것은 다음과 같은 4가지 요소에 의해서 결정된다 : microarray designer, spotter, experimenter, scanner.

① Microarray Designer

마이크로어레이 디자인은 이미지 분석시에 각각의 블럭과 스팟의 그리드구조를 정확하게 파악할 때 굉장히 중요한 요소로 작용한다. 예를 들어서, 블럭간의 간격이 스팟간의 간격과 별 차이가 없다면 자동 그리딩이 매우 힘들게 되며, 사용자가 일일이 직접 블럭의 위치를 설정한다 하더라도 그 때 필요한 노력이 매우 많다. 이와 더불어 control gene의 위치도 마이크로어레이 품질에 매우

기/획/특/집

중요하게 작용한다. control gene은 일반적으로 그 위치를 알고 있고 항상 100%발현하는 유전자로 선정되는데, 이를 이용하여 다른 유전자의 그리드상 위치를 쉽게 알 수 있으며 슬라이드간, 혹은 슬라이드 내에서의 표준화 작업을 하는 기준값을 제공하게 된다. 그러므로 control gene의 배치를 잘못하게 되면 스팟의 그리딩 작업이 매우 어려울 수 있으며, control gene의 선정이 잘못될 경우에는 표준화작업이 잘못되어 발현결과값에 심각한 영향을 미칠 수 있다.

② Spotter

각각의 스팟을 슬라이드에 심어주는 기계인 spotter는 최근 많은 기술발전과 개발로 인해서 그 성능이 향상되고 있다. Spotter가 올바르게 작동하지 않을 경우, 스팟의 위치가 일정하지 않고 그 모양이 이상적인 형태인 원형의 모양을 이루지 못할 수 있다. 이와 더불어서 실제의 foreground영역이외의 오염된 영역을 만들 수 있기 때문에, spotter의 능력은 마이크로어레이 품질에 많은 영향을 준다.

③ Experimenter

실험자는 유리슬라이드에 붙일, 테스트할 유전자들을 선정하고, 각각의 샘플을 준비하게 된다. 만일 실험자가 control gene과 test gene을 적절히 선정하지 못하면 슬라이드 내의 거의 모든 유전자가 발현되거나 혹은 미 발현될 수 있다. 발현된 유전자와 미 발현된 유전자의 비는 어느 정도 일정한 비율을 유지하는 것이 좋으며, 어느 한 쪽으로 치우치면 그 결과에 대한 신뢰성이 떨어지고 표준화 작업이후에 더 많은 에러값을 포함할 수 있다.

④ Scanner

마이크로어레이의 실험결과는 이미지상의 각 스팟에 대한 색의 강도를 측정하여 얻는다. 이것은 결국 슬라이드를 스캐닝하는 스캐너와도 밀접한 관계를 가진다. 일반적으로 마이크로어레이 스캐너는 스캐닝시에 빛의 강도 측정 레벨을 조정할 수 있고, 각각의 슬라이드에 알맞은 레벨은 슬라이드마다 다르다. 스캐닝시에 스캐너의 정확도와 함께 이러한 레벨을 결정하는 것도 스팟의 foreground와 background, 발현값을 결정하는 중요한 요소가 될 수 있다.

2. 마이크로어레이 구성요소들의 통계량

마이크로어레이의 품질을 정량화하는 quality measure를 정의하기 전에, 이미지가 포함하고 있는 스팟과 블록들의 여러 가지 통계량을 측정해보는 것이 중요하다. 이러한 통계량은 이후 정의할 quality measure에 앞서 전체 이미지의 구조를 파악하는데 매우 중요하며 실제로 이미지의 quality가 이러한 값에 의존적일 수 있다. 일반적으로, 한 실험을 통해서 생성되는 여러 이미지들이 포함하고 있는 그리드구조는 모두 동일하다. 그러므로 한 실험을 통해서 얻어지는 이미지들이 포함하고 있는 블록과 스팟들의 규칙성은 통계적 자료를 통해서 간단하게 알 수 있다. 또한 이러한 정보들은 똑같은 그리드 구조를 가지는 다른 칩 이미지들의 분석 시에 쉽게 그리드구조를 파악하여 분석할 수 있는 좋은 정보가 된다. 블록과 스팟들의 그리드구조를 알아보기 위해서 다음과 같은 통계량을 측정한다.

- ① IBA = Avg(Inter Block) ④ ISs = Std(Inter Spot)
 ② IBs = Std(Inter Block) ⑤ SDA = Avg(Spot Diameter)
 ③ ISA = Avg(Inter Spot) ⑥ SDs = Std(Spot Diameter)

일반적으로 스팟과 블록간 간격이 크고, 각 간격의 표준편차가 적을수록 마이크로어레이의 quality가 좋고, 스팟의 반지름이 크고 표준편차가 작을수록 quality가 좋다고 할 수 있다. 하지만, 이러한 관계가 항상 만족하는 것은 아니며 이것은 정확한 quality확인을 위한 또 하나의 검증자료로 활용할 수 있다. 또한, 향후 quality값들의 표준화를 위해서 이러한 통계량을 사용할 수 있다.

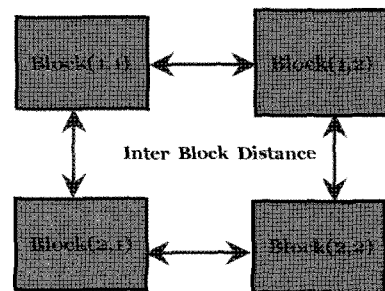


Figure 1. 2 X 2 블록구조를 가질 때의 블록간 거리의 계산. 각각의 블록에 대해서 이웃한 4-연결(connected) 인접 블록과의 거리를 블록간의 거리로 나타내고 모든 블록간 거리의 평균을 IBA, 표준편차를 IBs라 한다.

Korea Genome Organization

기/획/특/집

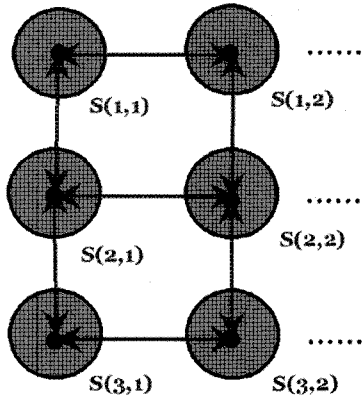


Figure 2. 3 X 3 이상의 스팟구조를 가질 때의 스팟간 거리의 계산. 각각의 스팟에 대해서 이웃한 4-연결(connected) 인접 스팟과의 거리를 스팟간의 거리로 나타내고 모든 스팟간 거리의 평균을 ISa, 표준편차를 ISs라 한다.

3. 마이크로어레이 이미지품질평가함수

마이크로어레이 실험의 목적은 유전자들의 발현율을 측정하는 것이다. 이것은 이미지상에서 각 스팟들의 위치에 해당하는 색의 강도를 조사하여 알 수 있다. 하지만, 마이크로어레이 실험은 실험자의 능숙도나 마이크로어레이어의 성능, 염료의 양 등에 따라 다양한 결과를 얻을 수 있고, 실제로 실험한 결과가 실제 반응을 제대로 나타내지 못할 수도 있다.

우리는 스팟의 크기나 모양, 위치 등의 정보를 고려하여 유전자의 발현 값이 얼마나 정확한지를 측정할 수 있는 품질평가함수들을 다음과 같은 5가지로 정의한다 : Signal Noise, Background Noise, Scale Invariant, Size Regularity, Spot Alignment.

① Signal Noise

스팟의 영역에서 블랍(blob) 혹은 여러가지 다른 이물질이 이물질이 묻게 되면, 발현값을 측정할 때 오류가 나타내게 된다. 우리는 이러한 평가함수를 전체 스팟의 영역과 오염된 영역의 비로 정의한다.

$$\text{Signal_Noise} = \frac{\text{Ignored_Signal_Area}}{\text{Signal_Area}}$$

② Background Noise

마이크로어레이 이미지 분석시에 배경의 강도값도 실험

결과에 많은 영향을 미친다. 그러므로 배경에 포함되어 있는 오염된 영역도 중요한 평가함수 중 하나이다. 배경이 포함하고 있는 잡음이 얼마나 되는지를 확인하여 이미지를 평가할 수 있다. 마이크로어레이 실험시, 각각의 스팟의 발현율은 Foreground와 background강도를 이용하여 측정된다. 만일 이 두가지 값중 하나라도 많은 에러를 포함하고 있다면 그 결과는 신뢰할 수 없다. 그러한 측면에서 Signal Noise와 Background Noise를 매우 중요하다.

$$\text{Background_Noise} = \frac{\text{Ignored_Background_Area}}{\text{Background_Area}}$$

③ Scale Invariant

하나의 스팟에서 유전자들이 hybridization할 때, 그 모양의 원형에 가까울 수록 실험의 질이 좋다고 할 수 있다. 우리는 Scale Invariant와 Size Regularity를 이용하여 스팟의 모양이 이상적인 형태인 원형에 얼마나 가까운지를 체크한다. 먼저 스팟이 포함하고있는 컴포넌트영역의 무게중심을 그 스팟의 중심이라 가정한다. 스팟의 중심이 원의 중심이고 스팟내에서 발견된 유효 컴포넌트영역을 포함하는 원들중 최소의 반지름을 가지는 원을 Co라 하고, 컴포넌트 영역이 포함할 수 있는 원들중 최대 반지름을 가지는 원을 Ci라 하면, Scale Invariant는 Co의 반지름과 Ci의 반지름의 비로 정의된다(Figure 3). Size Regularity와 비교할 때, 원의 크기에 대한 고려가 적고, 원의 모양이 얼마나 원형인지를 확인할 수 있다.

$$\text{Scale_Invariant} = \frac{\text{Radius(Inner_Circle_of_Spot)}}{\text{Radius(Outer_Circle_of_Spot)}}, \in [0,1]$$

④ Size Regularity

Scale Invariant와 비슷한 의미를 지니지만 때때로 다른 값을 가질 수 있다. 시그널 영역과 시그널 영역으로 만들어질 수 있는 이상적인 원의 넓이의 비로 나타낸다 (Figure 3). 넓이는 원의 반지름의 제곱과 비례하므로 반지름의 크기에 따라 많은 변화를 가질 수 있고, Scale Invariant에 비해서 실제 Signal의 영역에 의존하는 measure이다.

$$\text{Size_Regularity} = \frac{\text{Signal_Area}}{\text{Inner_Circle_Area}}, \in [0,1]$$

기/획/특/집

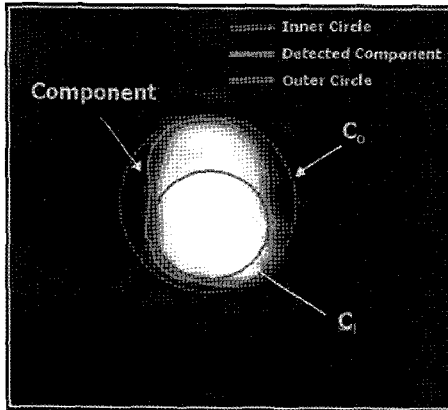


Figure 3. Scale Invariant와 Size Regularity를 설명한 그림. Scale Invariant는 하나의 스팟으로 인식된 컴포넌트에서 컴포넌트를 포함하는 최소원과 컴포넌트가 포함하고 있는 최대원사이의 반지름비로 정의되고, Size Regularity는 컴포넌트를 포함하는 최소원의 넓이와 컴포넌트의 넓이의 비로 정의된다.

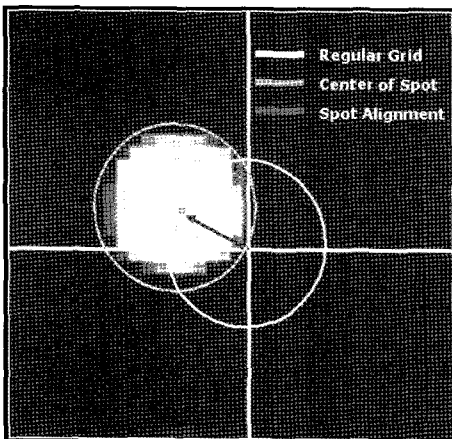


Figure 4. Spot Alignment measure. 전체 블록 이미지를 일정한 거리의 그리드로 분할하였을 때, 실제 스팟과 가장 가까운 그리드 센터 사이의 거리값을 이용하여 계산한다. 값이 1이면 실제 스팟의 중심과 그리드 센터가 동일한 점이며, 0이면 가까운 두 그리드 센터 사이의 중간에 스팟이 놓이게 된다.

⑤ Spot Alignment

블록내 각 스팟의 위치는 모두 똑같은 거리의 그리드 형태로 놓여 있어야 한다. 각 스팟이 이상적인 그리드 형태로 놓였을 때의 위치와 실제 위치의 차이를 스팟위치 에러로 정의한다. 만일 Spot Alignment의 값이 안좋은 경우, 자동 그리딩이 힘들며 auto adjust시 많은 시간이

소요된다. 이러한 현상은 대부분 spotting혹은 scanning 시의 잘못으로 인해서 발생한다. 앞에서 설명한 4가지 measure들은 모두 정규분포를 따르지만, Spot Alignment는 Poisson분포를 따른다(Figure 4).

Spot Alignment =

$$1 - \frac{\text{Avg}(\text{Real_Position_of_Spot_Regular_Position_of_Spot})}{\text{Avg}(\text{Inter_Spot_Distance}/2)}, \in [0,1]$$

4. EXPERIMENT AND RESULT

우리는 마이크로어레이 이미지의 품질평가함수의 테스트를 위해서 6종류의 이미지를 사용하였다. A부터 E까지는 국내 대학 연구실 혹은 국내 기업 등을 통해서 얻은 실험 데이터들이고, F는 스탠포드대학에서 Yeast의 세포주기를 연구할 때 사용한 마이크로어레이 이미지이다. 먼저 각 이미지들이 포함하고 있는 칩의 특성을 알아보기 위해서 블록과 스팟에 대한 기본적인 통계량을 Table 1과 같이 알아보았다. 이러한 정보는 실험에 대한 정보를 알 수 있을 뿐만 아니라, 향후 똑같은 형태로 설계된 마이크로어레이의 발현값을 분석할 때 메타그리드에서 사용할 수 있는 유용한 정보가 될 수 있다. 또한 Table 2에서 보여주는 품질평가측정 값이 실험의 1차 품질평가함수라면 Table 1은 2차 품질평가함수가 될 수 있다. 각각의 이미지는 발현실험을 한 유전자들의 종류에 따라 그 발현율이 다르다. 각각의 스팟에 대하여 발현율이 높을 수록 이미지 분석시에 그리딩이 쉬우며, 나아가 자동그리딩이 가능하게 된다. 6종류의 이미지에 대하여 품질평가함수를 적용한 결과는 Table 2와 같다.

Scale Invariant는 다른 4가지 measure에 비해서 비교적 차이가 크지 않았고, 나머지 4종류의 Measure에 대해서는 뚜렷한 차이를 보였다. 우리는 이러한 차이를 통해서 각 이미지가 가지는 품질을 확인할 수 있다. 예를 들어서 A이미지는 Background Noise에 대해서는 매우 안좋은 결과를 나타내었는데, 이것은 발현값을 측정할 때 심각한 오류를 나타낼 수 있음을 나타낸다. 이에 비해서 D의 이미지는 Background Noise와 Signal Noise에 대해서 아주 좋은 결과를 보여주며 이는 발현값을 계산할 때 잡음이 거의 포함되지 않은 값을 이용하여 계산할 수 있다는 것을 나타낸다.

기/획/특/집

Chip	B	S	Image Resolution	BD	SD	Radius	C(%)
A	16	16×100	1886×2321	102.41	35.97	8.16	141(8.8%)
B	16	16×324	1900×1860	29.73	23.03	6.63	995(19.2%)
C	16	16×324	1808×1800	31.33	23.02	7.14	1650(31.8%)
D	16	16×168	1984×2004	122.79	24.89	7.49	1289(48.0%)
E	16	16×342	2000×2000	22.17	22.93	8.82	656(12.0%)
F	4	4×1596	1024×1024	23.17	10.43	4.13	715(11.2%)

Table 1. Chip Specification. 각 이미지에 포함되어 있는 블록과 스팟의 그리드 구조와 발현을 등을 나타낸다. 이 정보는 메타그리드의 중요한 정보로서 향후 똑 같은 형태로 설계된 마이크로어레이 이미지에서는 이러한 정보를 이용하여 쉽게 분석 가능하다. B : number of blocks, S : number of spots, BD : block distance, SD : spot distance, C : number of component detected.

Chip	Background Noise	Signal Noise	Scale Invariant	Size Regularity	Spot Alignment
A	51.94%	0.53%	0.5648	0.6931	0.9223
B	36.39%	2.34%	0.5238	0.5534	0.9016
C	21.35%	0.85%	0.5575	0.7315	0.9456
D	9.85%	0.85%	0.5608	0.6211	0.9526
E	30.13%	1.58%	0.5411	0.4512	0.8721
F	46.24%	5.92%	0.4714	0.5076	0.8445

Table 2. Quality Measure 실험결과. 6개의 데이터에 대한 품질평가표. 대체적으로 Signal Noise와 Spot Alignment는 좋은 결과를 보이지만, Background Noise, Scale Invariant, Size Regularity에서는 편차를 가진다.

전체적으로 C와 D가 가장 좋은 결과를 나타내었으며, F는 나머지 5종의 테스트 이미지 결과와는 많은 차이를 나타내었다. F와 같은 이미지는 자동 그리딩이 거의 불가능한 이미지라고 할 수 있다.

5. DISCUSSION

Figure 5는 실험한 6종류의 그레이 스케일 이미지에 대해서 color mapping한 모습을 보여준다. A의 이미지는 foreground와 background의 구분이 명확하지 않기 때문에 Background Noise measure에 대해서 매우 안 좋은 값을 나타내었으며 이것은 향후 정확한 발현값을 얻어내기 힘들 것이다. B의 경우에는 다른 이미지들에 비해서 Signal

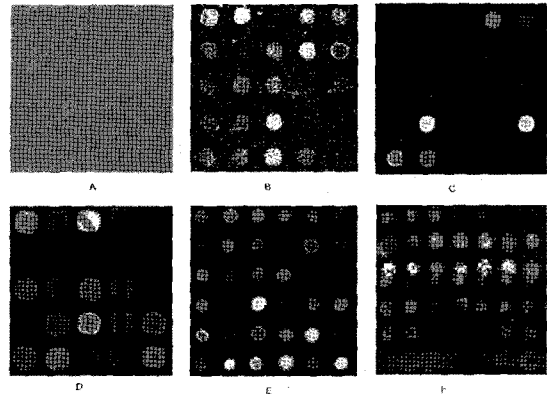


Figure 5. 테스트 이미지 비교. A의 이미지는 foreground와 background의 구분이 명확하지 않기 때문에 Background Noise measure에 대해서 매우 안 좋은 값을 나타내었으며, B의 경우에는 다른 이미지들에 비해서 Signal Noise가 비교적 안 좋게 나타났다. E는 Size Regularity가 좋지 않고, Spot Alignment값도 비교적 낮다. 이것은 스팟의 모양이 원형을 이루지 못하는 것이 많고, 그리드의 위치가 다른 것들에 비해서 일정하지 못하다는 것을 나타낸다. C와 D는 전체적으로 양호한 결과를 보여준다.

Noise가 비교적 좋지 않게 나타났다. Background값에 비해서 Signal값에 해당하는 foreground값은 그 편차가 매우 크기 때문에 잡음의 비율은 그리 높지 않지만, 발현값에 영향을 줄 가능성이 있다. E는 Size Regularity가 좋지 않고, Spot Alignment값도 비교적 낮다. 이것은 Figure 5의 E에서 보는 바와 같이 스팟의 모양이 원형을 이루지 못하는 것이 많고, 그리드의 위치가 다른 것들에 비해서 일정하지 못하다는 것을 나타낸다. C와 D는 전체적으로 양호한 결과를 보여주지만, F는 모든 종류의 measure에 대해서 가장 안 좋은 결과를 나타내었고, Figure 5에서 보는 바와 같이 유전자의 발현결과값이 거의 신뢰할 수 없을 것으로 보인다.

6종류의 이미지에 대하여 테스트한 결과 신호잡음과 스팟 위치에러는 비교적 큰 편차를 보이지 않았다. 초기 마이크로어레이 실험 시에는 이러한 부분이 많은 문제를 야기시켰지만, 마이크로어레이어의 발달과 실험자의 주의 및 실험환경 개선 등으로 최근에는 이러한 부분이 많이 향상되었다. 이와 달리 배경잡음은 큰 편차를 나타내었는데, 이것은 유전자의 발현값을 측정할 때 정확도의 차이를 야기시킬 수 있다. 똑같은 슬라이드에 대해서

기/획/특/집

도 레이저로 얻은 이미지와 CCD카메라로 얻은 이미지는 배경잡음과 신호잡음의 차이를 나타내는데, 특히, 미세하게 구분되어야 할 배경에서 더 큰 차이를 나타내게 되는 것이다. Scale Invariant와 Size Regularity는 실험시 환경에 많은 영향을 받는다. 특히, 슬라이드 구성시에 얼마만큼의 염료를 사용하는가에 따라 똑같이 발현되더라도 그 모양이 틀려질 수 있다. 일반적으로 넉넉한 양의 염료를 사용할수록 스팟의 모양이 원에 가까워지며 이 두가지 평가함수에서도 좋은 결과를 나타낼 것이다.

6. CONCLUSION AND FUTURE WORK

본 보고서에서는 발현실험이후의 마이크로어레이 이미지에 대하여 실험이 얼마나 잘 이루어졌는지를 예측할 수 있는 품질제어평가함수(quality control measure)를 제안하였다. 또한 마이크로어레이 전체 실험과 한 이미지에 대한 품질을 평가하기 위해서 이미지내의 블러들과 스팟들 간의 간격에 대한 통계량을 이용하여 정의하고, 스팟들의 발현값에 대한 정확도를 측정하기 위한 품질평가함수들을 정의하였다. 이러한 품질평가함수들은 실험을 통해서 얻어지는 동일한 디자인의 여러 마이크로어레이 이미지들 간의 품질을 평가할 수 있어서, 실험의 완성도나 실험자의 능숙도 그리고 마이크로어레이의 성능을 측정할 수 있다. 본 보고서에서 제안한 5가지 quality measure(Signal Noise, Background Noise, Scale Invariant, Size Regularity, Spot Alignment)들은 마이크로어레이 이미지의 quality를 잘 반영하였으며, 마이크로어레이 실험결과가 나타내야 할 특징들을 정량화할 수 있었다.

앞으로 개선하거나 연구해야 할 과제는 다음과 같다.

① 본 논문에서 정의한 마이크로어레이 이미지 품질평가함수는 각 스팟들의 크기는 고려하지 않고, 단순하게 평가함수를 통해서 얻어지는 값만을 계산한다. 만일 어떤 스팟은 반지름이 아주 크고, 다른 스팟은 반지름이 아주 작는데, 두 스팟에 대한 평가가 비슷하게 나왔다면 반지름이 작을수록 더 잘 된 것이라 할 수 있다. 왜냐하면 스팟의 크기가 크면 클수록 더 정확한 실험 결과를 얻기 때문이다. 향후 한 이미지내에 포함되어 있는 스팟의 수, 혹은 스팟의 반지름의 크기에 따라 품질평가값을 정규화하는 작업이 필요하다.

② 각각의 measure가 전체 quality에 영향을 주는 가중치가 다르고, 실험자에 따라 그 중요성이 달라질 것이다. 1차적으로 범용으로 사용할 수 있는 통합 quality measure model의 구성이 필요하며, 각각의 실험 특징에 맞는 quality measure model에 대한 연구가 필요하다.

③ 마이크로어레이 실험은 슬라이드에 붙이는 유전자나, 그리드구조, 반지름크기, 염료, 이미지 추출방법 등 여러 가지 조건에 따라 많은 영향을 받는다. 그러므로, 마이크로어레이 실험의 완성도를 측정하는 품질평가 함수는 상황에 따라 다른 모델로 구성되어 질 수도 있다. 각 변수들 간의 상관관계를 잘 파악하여 특정 상태나 상황에 맞는 마이크로어레이 품질평가함수의 개발이 필요하다.

④ 정의한 5가지 measure에 대해서 사용자가 수용가능한 임계값을 결정할 필요가 있다. 예를 들어서 Figure 5의 A에서는 Background Noise가 50%를 초과하였는데, 이때 수용 임계값이 50%로 정해져있다면 A의 이미지 분석 결과는 신뢰할 수 없다는 것을 나타낸다.

<http://pearl.cs.pusan.ac.kr/~arrayzer/index.html>에서 본 연구에서 사용한 프로그램과 문서 및 설명서를 다운받거나 확인할 수 있다.

7. REFERENCES

- Hautaniemi S, Edgren H, Vesanen P, Wolf M, Jarvinen AK, Yli-Harja O, Astola J, Kallioniemi O and Monni O (2003) A novel strategy for microarray quality control using Bayesian networks. *Bioinformatics* 19: 2031-2038.
- Jung HY and Cho HG (2002) An automatic block and spot indexing with k-nearest neighbors graph for microarray image analysis. *Bioinformatics* 18: 141S-151S.
- Kuklin APA and Shams S (2001) Quality control in microarray image analysis. *Microscopy Research*.
- Pan-Gyu Kim, Hee-Jeong Jin and Hwan-Gue Cho (2003) Quality measures for microarray design and experiments. *KSBI* 155-163.