

# ToMAS : 마이크로어레이 이미지 분석용 컴포넌트 도구

천봉경<sup>o</sup> 장철진 진희정 이평준 김혜정 조환규

부산대학교 컴퓨터공학과, ALGORIGENE 연구실

{bkchun<sup>o</sup>, cijang, hijin, pjlee, hjkim, adagio}@pearl.cs.pusan.ac.kr

## ToMAS : Component Tools for Microarray Image Analysis

BongKyung Chun<sup>o</sup> ChulJin Jang Jin H.J. Lee P.J. Kim H.J. Cho HwanGue

Dept. of Computer Engineering, Pusan National University

### 요 약

마이크로어레이 칩을 이용한 실험은 한 번에 수백 혹은 수십만 개의 유전자 발현 정보나 유전자형을 얻을 수 있기 때문에 유전자 비교 분석에 있어서 획기적인 방법이라 할 수 있다. 점차 마이크로어레이 칩을 이용한 실험이 활발히 진행되면서 마이크로어레이 이미지 분석을 위한 소프트웨어가 필요로 하게 되었으며, 이를 위해 많은 소프트웨어들이 개발되었다. 하지만 지금까지 개발되어온 소프트웨어들은 필요한 기능을 하는 모듈을 재사용하여 소프트웨어를 쉽게 업데이트(update)하거나, 다른 소프트웨어 개발 시에 일부 필요한 모듈(module)들을 쉽게 재사용하기가 힘들었다. 본 논문에서는 이러한 문제점을 극복하기 위해서 컴포넌트 기반으로 마이크로어레이 이미지 분석 소프트웨어 개발에 대해서 제안한다. 개발된 컴포넌트들은 사용자 편의에 맞게 새로운 마이크로어레이 이미지 분석 소프트웨어를 다시 개발하거나, 단백질 칩과 같은 다른 바이오 이미지 분석 소프트웨어 개발 시에서도 쉽게 재사용될 수 있어 개발 기간 및 개발 비용을 줄일 수 있다.

### 1. 연구 동기

마이크로어레이(microarray)는 기계 자동화와 전자 제어 기술 등을 이용하여 적게는 수백 개부터 많게는 수십만 개의 DNA를 유리 박편에 매트릭스형태로 심어 놓은 것이다. 그림 1은 마이크로어레이 실험과정을 나타낸다. 그림에서 DNA 클론(clone)들은 여러 가지 장기기관이나 개체에서 추출된다. 그러나 그 양이 매우 적어 PCR(Polymerase Chain Reaction)이라는 방법을 통해 증폭하여 많은 양의 같은 유전정보를 가진 개체를 만들어 DNA 프로브(probe)로 심는다. 그리고 두 개의 다른 환경(test, reference)에서 얻은 세포들의 mRNA를 추출하여 각각의 mRNA로부터 cDNA를 만들어 초록색(Cy3-dUTP)와 빨강색(Cye5-dUTP)로 형광 처리를 한다. 이후에 테스트 샘플과 레퍼런스 샘플을 마이크로어레이 위의 DNA 프로브와 실험을 한다. 실험은 시간변화나 온도변화 등의 다양한 조건으로 하며 실험 후에는 광학 스캐닝 단계를 거쳐서 마이크로어레이 이미지 파일을 얻어낸다. 이 후 마이크로어레이 이미지 분석 소프트웨어를 이용하여 생성된 이미지의 형광 양을 측정하여 유전자 발현 정도를 알아낸다. 이러한 마이크로어레이 실험은 한번에 수백 혹은 수십만 개의 유전자 발현 정보나 유전자형(genotype)을 얻을 수 있기 때문에 유전자 비교분석에 있어서 획기적인 방법이라 할 수 있다[1].

현재 마이크로어레이를 이용한 실험이 활발히 진행되면서, 마이크로어레이 이미지 분석을 위한 소프트웨어가 필요하게 되었다. 그러나 현재 국내에서 제대로 개발된 마이크로어레이 이미지 분석 소프트웨어가 없어서, 마이크로어레이 실험을 하는 국내 대부분의 기업이나 연구소에서는 GenePix[2]나 ImaGene[3]과 같은 가격이 비싼

국외 상용 소프트웨어를 이용하고 있다. 설사 국내에서 개발된 마이크로어레이 소프트웨어가 있다 하더라도, 컴포넌트 기반으로 개발되어 있지 않아 일부 필요한 기능의 컴포넌트들을 재사용하여 소프트웨어를 쉽게 업데이트(update)하거나 다른 소프트웨어 개발 시에 필요한 컴포넌트들을 쉽게 재사용하기 힘든 문제점이 있었다. 본 논문에서는 이러한 문제점을 극복하기 위해서 컴포넌트 기반으로 마이크로어레이 이미지 분석 소프트웨어 개발에 대해서 제안한다.

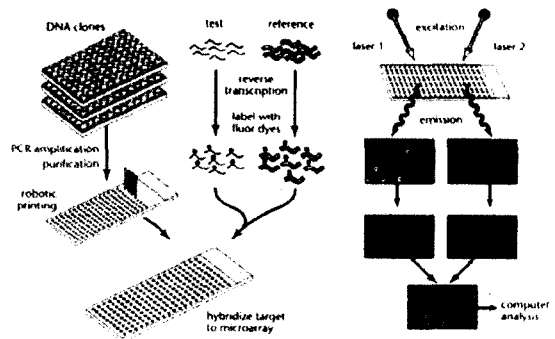


그림 1. 마이크로어레이 실험과정 [1]

### 2. 마이크로어레이 이미지 분석 단계

마이크로어레이 이미지의 분석단계는 크게 그리딩(griding), 세그멘테이션(segmentation) 그리고 강도 추출의 3가지 단계로 나눌 수 있다.

• 그리딩(griding)

그리딩 단계는 마이크로어레이 이미지에서 스팟(spot)이 있을만한 영역을 그림 2와 같이 그리드(grid) 모양으로 구역을 나누고 각 구역의 인덱스(index)를 붙여 주는 단계이다. 보통 마이크로어레이 이미지는 수십만 개의 스팟들로 구성되어 있기 때문에 이와 같은 작업을 수동으로 할 경우 이미지 분석 시 많은 시간이 소요된다. 본 논문에서 개발한 컴포넌트에서는 본 연구실에서 개발한 방법으로 자동으로 그리딩 작업을 수행해 준다[4, 5].

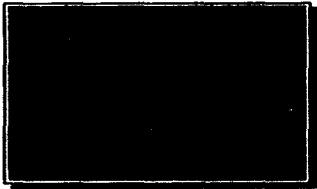


그림 2. 그리딩 된 이미지 모습

• 세그멘테이션(segmentation)

세그멘테이션 단계는 그리딩된 각 영역에서 스팟 영역인 포그라운드(foreground)와 스팟 외 영역인 백그라운드(background)를 나눠주는 역할을 한다. 이론적으로 이미지의 백그라운드의 강도는 0이 되어야 하지만 스캐너 상태 및 발현 간섭 등 여러 가지 환경 요인으로 백그라운드의 강도가 0이 나오는 경우가 거의 없다. 그래서 정확한 스팟 영역을 얻어내기 위해 포그라운드와 백그라운드를 구별해 줘야 한다. 이 영역은 이미지 분석 결과에 직접 영향을 미치기 때문에 상당히 중요하다. 본 논문에서는 로컬(local) 백그라운드[6]와 Mann-Whitney test[7]를 이용한 방법을 사용하였다.

• 강도 추출

강도 추출은 마이크로어레이 이미지의 최종 결과인 각 스팟의 R과 G의 비(ratio) 값 및 기타 다른 값들을 추출하는 단계이다. 본 논문에서는 강도 추출뿐만 아니라 마이크로어레이 실험결과와 품질평가 결과도 알려준다[8].

3. 마이크로어레이 이미지 분석용 컴포넌트

본 장에서는 현재 개발 중인 ToMAS(Toolkits of Microarray Analysis System)라는 마이크로어레이 이미지 분석용 컴포넌트에 대해서 소개한다. ToMAS는 마이크로어레이 이미지 분석 시 자주 사용되는 컴포넌트들로서, 크게 이미지 분석 컴포넌트, 정규화 컴포넌트, 클러스터링 컴포넌트, 파일 포맷 변환 컴포넌트 그리고 그래프 컴포넌트로 구성된다.

• 이미지 분석 컴포넌트(Image Analysis Component)

이미지 분석 컴포넌트는 ToMAS에서 가장 핵심인 컴포넌트로서, 이미지 입출력(IO), 이미지 프로세싱(image processing), 이미지 분석과 같은 기능을 담당하는 컴포넌트이다. 이미지 분석 컴포넌트는 COM기반의 ActiveX로 구현하였으며, 만약 다른 소프트웨어 개발 시에 재사용하기 위해서는 윈도우 폼(window form)에 추가하는

식으로 간단히 재사용할 수 있다.

• 정규화 컴포넌트(Normalization Component)

정규화는 마이크로어레이 발현 데이터를 보정하여 주는 작업을 의미한다. 정규화 단계에서는 실험 데이터의 종류, 스캐너 특성, 실험자의 숙련도 등에 따라서 정규화 방법이 달라지므로, 한 시스템에서 다양한 환경, 다양한 실험자들을 위한 정규화 방법을 제공해주기가 힘들다. 이러한 문제점을 해결하기 위해서 본 논문에서 개발하는 정규화 컴포넌트는 R 라이브러리인 BioConductor[9]를 사용하여 다양한 정규화 소스를 제공하며, 사용자가 구현한 R 소스를 플러그인(plug-in)하여 사용할 수 있도록 개발 중이다.

• 클러스터링 컴포넌트(Clustering Component)

클러스터링 컴포넌트는 마이크로어레이 이미지의 시간별 분석 파일들을 읽어 각 유전자의 시간별 발현 값을 그래프로 보여주며, 각 유전자와 가장 비슷한 발현 패턴을 가지는 유전자 혹은 가장 대조적인 패턴을 가지는 유전자들을 찾아 주는 기능을 하는 컴포넌트이다.

• 파일 포맷 변환 컴포넌트(File Format Transformation Component)

파일 포맷 변환 컴포넌트는 현재 마이크로어레이 이미지 분석 시 가장 많이 사용되는 소프트웨어인 GenePix의 파일 포맷과 현재 마이크로어레이 실험의 표준으로 자리 잡아 가는 MAGE-ML 파일 포맷[10]으로 변환해주는 기능을 하는 컴포넌트이다.

• 그래프 컴포넌트(Graph Component)

그래프 컴포넌트는 흔히 사용되는 2D, 3D의 히스토그램(histogram) 또는 스캐터(scatter) 그래프를 그려주는 컴포넌트이다. 이 컴포넌트는 전체 이미지의 픽셀(pixel) 값, 각 스팟의 픽셀 값들을 2D 혹은 3D그래프로 보여주는데 사용된다.

4. 활용 예 : ArrayShop의 구현

현재 ToMAS를 이용하여 마이크로어레이 용 통합 소프트웨어인 ArrayShop을 개발 중이다. ArrayShop은 크게 그림 3에서 보듯이 4가지 작은 프로그램으로 구성되며, 각 프로그램은 현재 ToMAS의 컴포넌트들을 사용하여 개발하고 있다.

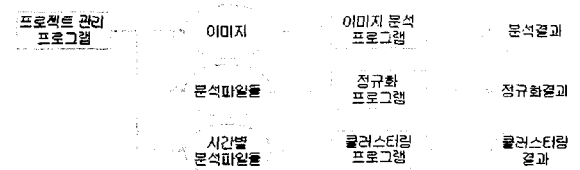


그림 3. 마이크로어레이 이미지 통합 소프트웨어인 ArrayShop의 각 프로그램 간의 관계도

• 프로젝트 관리 프로그램

프로젝트 관리 프로그램은 마이크로어레이 이미지들을

프로젝트 단위로 구성하여 이미지들과 그 분석 결과들을 체계적으로 관리해주는 프로그램이다. 그림 4와 같이 Project 파일은 Experiment, Work, Shot의 요소들로 구성되며, 이 프로그램은 파일 포맷 변환 컴포넌트, 그래프 컴포넌트를 이용하여 개발하고 있다.

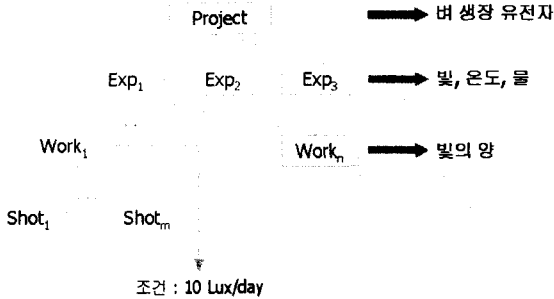


그림 4. 프로젝트 관리 프로그램의 프로젝트 계층도

• 이미지 분석 프로그램

이미지 분석 프로그램은 프로젝트 관리 프로그램에서 선택된 이미지들을 분석해주는 프로그램이다. 그림 5는 마이크로어레이 이미지 분석 프로그램의 실행 화면이다. 이미지 분석 프로그램에서는 그림 5에서 보듯이 이미지 분석 컴포넌트, 그래프 컴포넌트 그리고 파일 포맷 변환 컴포넌트를 사용한다.

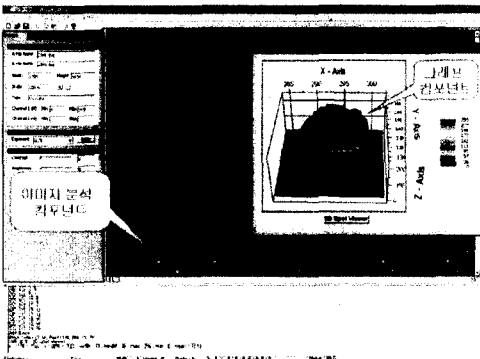


그림 5. 이미지 분석 프로그램의 실행 모습과 사용된 각 컴포넌트

• 정규화 프로그램

정규화 프로그램은 프로젝트 관리 프로그램에서 선택된 다수의 분석 결과 파일들을 정규화 컴포넌트를 사용하여 정규화해주는 프로그램이다.

• 클러스터링 프로그램

클러스터링 프로그램은 시간별 분석 파일들을 입력으로 받아, 클러스터링 컴포넌트와 그래프 컴포넌트들을 이용하여 구현된다.

5. 결 론

현재 국내에서는 마이크로어레이 칩을 이용한 실험이

활발히 진행되고 있음에도 불구하고, 아직까지 이렇다할 국내 기술로 개발된 마이크로어레이 용 분석 소프트웨어 또는 마이크로어레이 용 컴포넌트들이 없는 실정이다.

본 논문에서는 마이크로어레이 이미지 분석 용 컴포넌트인 ToMAS를 개발하였고, 이를 이용하여 마이크로어레이 분석 프로그램인 ArrayShop 구현에 대해 소개하였다. ArrayShop은 컴포넌트 기반으로 구현되므로, 업그레이드 시 필요한 부분의 ToMAS 컴포넌트를 업그레이드함으로써 ArrayShop의 업그레이드가 쉽게 이루어 질 수 있다. 또한 사용자 편의에 맞게 새로운 마이크로어레이 이미지 분석 소프트웨어를 다시 개발하거나, 단백질 칩과 같은 다른 바이오 이미지 분석 소프트웨어 개발 시에서도 쉽게 재사용될 수 있어 개발 기간 및 개발 비용을 줄일 수 있다. 현재 개발 진행 과정 및 개발된 프로그램들은 ToMAS 홈페이지인 <http://pearl.cs.pusan.ac.kr/~tomas>에서 확인 할 수 있다.

5. 감사의 글

본 연구는 한국과학기술기획평가원 국책연구개발사업 (M1-0339-00-002)의 지원으로 수행되었습니다.

6. 참고 문헌

- [1] David J. Duggan et al, Expression profiling using cDNA microarray, *Nature genetics supplement*, 21:10-14, 1999
- [2] GenePix, [http://www.axon.com/gn\\_GenePixSoftware.html](http://www.axon.com/gn_GenePixSoftware.html)
- [3] ImaGene, <http://www.biodiscovery.com/imagene.asp>
- [4] 진희정, A New Method of Block Indexing with Maximal  $\epsilon$  Regularity Point Set for Microarray Image Analysis, *Technical Report*, Graphics Application Lab, Pusan National University, 2003.
- [5] Ho-Youl Jung and Hwan-Gue Cho. "An Automatic Block and Spot Indexing with k-Nearest Neighbors Graph for Microarray Image Analysis". *Bioinformatics(Suppl.)*. 18 : 141-151, 2002
- [6] Yee Hwa Yang et al, Comparison of Methods for Image Analysis on cDNA Microarray Data, *Journal of Computational and Graphical Statistics*, Volume 2 Number 1 : 108-136, 2002.
- [7] Yidong Chen et al, Ratio-based Decisions and the Quantitative Analysis of cDNA Microarray Images, *Journal of Biomedical Optics*, 2 : 364-374, 1997
- [8] 김판규, 진희정, 조환규, "Quality Measures for Microarray Design and Experiment", *KSBI*, 2 : 155p, 2003.
- [9] BioConductor, <http://www.bioconductor.org/>
- [10] MAGE-ML, <http://www.mged.org/>