

DNA Microarray 분석 기술

김종원*, 김종대**

*(주) 바이오메드랩, **한림대학교

I. 서 론

Human Genome Project가 성공적으로 마치면서, 생명과학계는 이에 대한 활용에 관심이 집중되었다. 이러한 생명과학계의 관심 속에서 방대한 유전자 정보를 처리 및 해독할 수 있는 신기술로서 DNA microarray 기술이 개발되었다. 1990년대 초반에 기존의 유전자 분석 방법의 하나인 dot blot 방법을 개선한 microarray 방법이 소개되기 시작하였다^[1]. DNA microarray 기술은 소위 DNA의 상보적인 반응성을 이용하여, 알고 있는 유전자 정보와 알고자 하는 유전자를 hybridization시킴으로써 유전자 정보를 분석하는 방법의 한 가지이다. DNA microarray를 이용하여 알아낼 수 있는 유전자 정보는 대체적으로 유전자의 발현 정도(gene expression level)와 유전자의 염기 서열로 나눌 수 있으며, 이 목적에 따라 DNA microarray의 제작 방법은 물론, 분석 방법도 달라지게 된다. 이러한 DNA microarray 방법은 기존의 유전자 분석 방법보다 분석 단가의 측면이나 분석에 소요되는 시간적 측면으로 볼 때, 훨씬 유리한 것으로 평가되고 있어, 계속적으로 관심이 집중되고 있다. 따라서 최근 DNA microarray 기반 기술 및 응용에 대한 연구 개발이 폭발적으로 이뤄지고 있는 실정이다. 또한 최근 추정되는 DNA microarray 중에서 상품화된 DNA microarray(pre-made Microarray)의 세계 시장 규모는 2004년에 10억 달러 규모로 발전할 것으로 예상하고 있으며, 국내의 시장 규모도 500억원 이상으로

추정되고 있다.

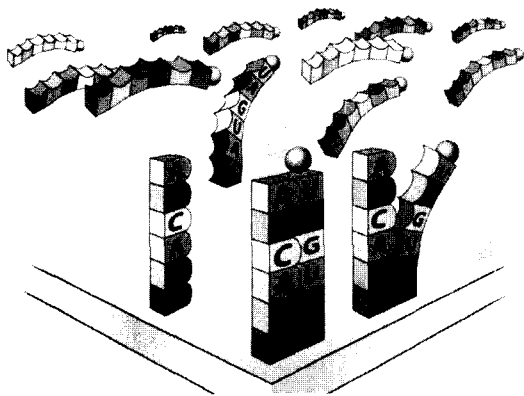
DNA microarray 기술은 나날이 변화하고 있으며, 다양한 신기술이 소개되고 있다. 본 글에서는 DNA microarray 기술과 이의 분석에 대한 내용만을 다루고자 하며, 최근 소개되고 있으며, DNA microarray 기술과 함께 융합이 가능한 Microfluidics나 Lab-on-chip에 관한 기술은 다루지 않고자 한다. 따라서 보편화된 DNA microarray의 원리 및 제작 방법, 종류 및 특징 등외에 Microarray 영상을 기반으로 하는 분석 방법에 대하여 기술하고자 한다.

II. DNA microarray

1. DNA microarray 원리

DNA microarray는 알고 있는 기본적으로 유전자를 고체 기판(solid substrate) 위에 고정시키고, 고정된 위치 정보와 유전자 정보를 이용하여 알고 싶은 유전자의 정보를 알아 내는 방법이다. DNA microarray가 제공할 수 있는 유전자 정보는 크게 유전자 염기 서열(gene sequence)과 유전자 발현 정도(gene expression level)로 나뉘지며, 이 목적에 DNA microarray를 설계, 제작한다. 가장 간단한 개념으로 <그림 1>에서와 같이, 특정 유전자 서열을 갖는 DNA 염기 서열(DNA sequence)을 고체 기판에 고정한 후, 형광 물질로 labeling 한 미지의 유전자를 hybridization을 시키게 된다.

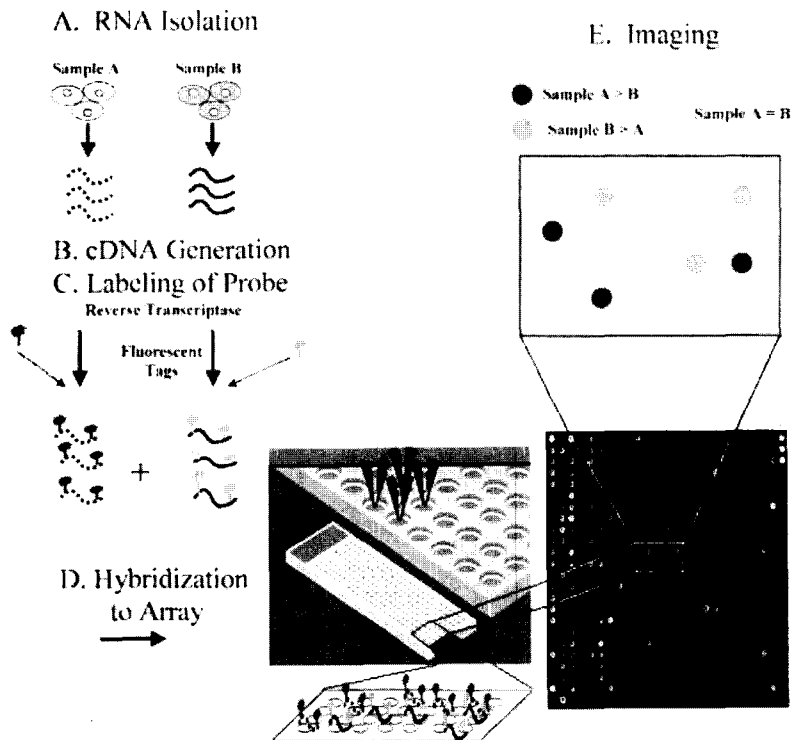
DNA hybridization이 상보적으로 결합된다



〈그림 1〉 DNA microarray 원리에 대한 모식도
(미국, Affymetrix사의 그림)

는 원리를 바탕으로, 반응된(형광물질을 포함하게 된) 유전자에서 형광 영상을 얻고, 영상으로부터 위치 정보를 확인하게 되면, 반응된 유전자의 염기 서열을 유추할 수 있다. 이러한 방법과 같이 hybridization을 이용한 염기 서열 분석

방법은 DNA microarray 기술 이전에 소위 dot blotting 방법으로 개발되어 사용되고 있었다. 반면, 유전자 발현 정도를 확인하기 위하여서는 이 방법을 좀 더 변형하여야 한다. 유전자의 발현 정도를 측정한다는 말은 세포 내의 특정 유전자가 생산하는 mRNA를 측정한다는 의미이며, 이러한 현상은 세포가 정상 상태에서와는 달리, 특정한 자극을 줄 경우, 정상 세포 상태의 경우에 비하여 상대적으로 mRNA이 적거나 많이 생산한다는 의미이다. 이 경우, 알고 있는 특정 유전자(실제로는 mRNA를 가지고 상보적으로 만든 complimentary DNA)를 고체 기판에 고정화 하여 DNA microarray를 제작한다. 따라서 이를 특별히 cDNA microarray라 하며, 유전자의 크기가 대략 500개-2000개의 염기로 구성된다. 정상적이지 못한 세포와 정상 세포로부터 mRNA를 추출하여, 소위 역전사(reverse transcription) 방법으로 각각의 cDNA를 만든



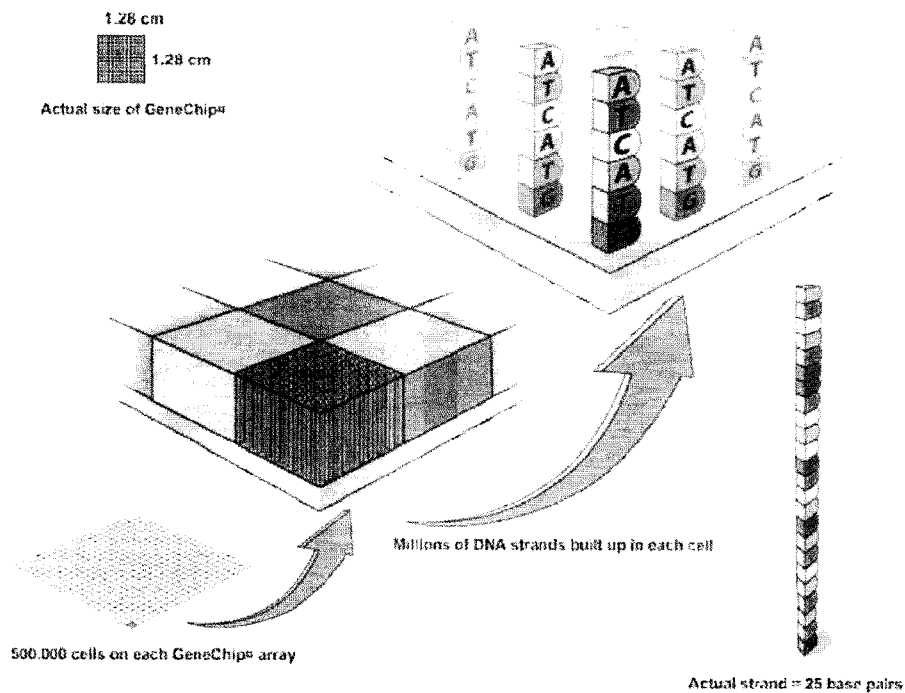
〈그림 2〉 cDNA microarray를 이용한 유전자 발현 정도 추정하는 실험 과정에 대한 모식도.

다. 이 역전사 과정 동안, 각기 서로 다른 형광 물질(파장이 달라서 형광 필터를 이용하여 서로 파장으로 영상을 구성할 수 있음)로 labeling 한다. <그림 2>에 설명된 것과 같이 sample A의 세포에서 한 유전자가 많이 발현되었을 경우, 그 유전자는 sample A로부터 나온 것이 sample B로부터 나온 유전자보다 상대적으로 많게 된다. 따라서 같은 유전자일지라도, 상대적으로 많은 sample A로부터 유래된 유전자가 cDNA microarray에 고정된 유전자와 많이 반응한다. 또한 같은 유전자라 하여도, sample A와 sample B의 모든 유전자를 각각 다른 형광물질로 염색하였기 때문에, 각각의 형광물질에 알맞은 형광필터를 이용하여 hybridization 반응이 끝난 cDNA microarray의 영상을 얻을 경우, 두 가지의 다른 영상을 얻게 된다. 최종적으로 얻은 두 가지 영상은, <그림 2>의 E. Imaging 부분에 나타난 것과 같이 각 sample로부터 유래된 유전자

의 양에 따라, 해당 형광 물질의 파장의 형광 밀도(Fluorescent Intensity)가 높아진다. 형광 밀도에 대하여 유사색으로 변환(pseudo-coloring)하여 두 가지 색의 영상을 만든 후, 후속 영상 처리를 한다. 만약 같은 양의 유전자가 반응될 경우, 동일한 형광밀도를 보이게 됨으로 두 영상을 합할 경우, 중간색으로 나타난다. 각각의 형광물질에 의한 영상에서, 각각의 색이 많은 해당 유전자가 많이 발현된 것으로 해석하게 된다.

2. DNA microarray의 제작 기술

DNA microarray는 근본적으로 고체 기판에 유전자를 고정화하는 과정이 필요한 데, 두 가지 과정을 생각할 수 있다. 하나는 고정하고자 하는 유전자를 미리 합성하여 고정화하는 방법이 있고, 다른 한 방법은 고체 기판 위에서 유전자를 합성하는 방법이 있다. cDNA의 경우 화학적



<그림 3> Affymetrix사 genechip의 유전자 밀도에 관한 설명도

로 DNA 합성 과정(chemical DNA synthesis process)으로 제작할 수 없기 때문에 모든 cDNA microarray는 cDNA를 mRNA로부터 역전사하여, 이를 고체 기판에 고정화하는 과정으로 제작하게 된다. 화학적으로 DNA 합성 과정을 통하여 합성할 수 있는 DNA 서열의 크기가 현재까지 100base 이하이며, 고체 기판 위에서 직접 합성하는 방법으로는 이보다 작은 크기인 30base 이하이다. <그림 3>은 이러한 고체 기판 위에서 직접 DNA 서열을 합성하는 방법으로 DNA microarray를 제작하는 Affymetrix사의 genechip 모식도입니다. 이 합성 방법은 반도체의 photolithography 기술을 이용하여 제작하며, 고정된 합성 유전자의 길이는 약 25개로 소위 oligonucleotide DNA microarray로 분류된다. Photolithography 기술을 이용하기 때문에, oligonucleotide의 밀도를 높일 수 있다는 장점이 있다²⁾. 현재 상용화된 Affymetrix Genechip의 oligonucleotide 밀도는 1.28cm×1.28cm의 면적안에 50만 가지의 oligonucleotide를 고정하였다. 이 밀도는 한번에 처리가 가능한 유전자의 정보량을 의미하게 되어 조만간 이 밀도는 훨씬 높아질 것으로 예상된다.

이와 달리, cDNA microarray든지 합성된 oligonucleotide든지 완성된 염기 서열을 고체

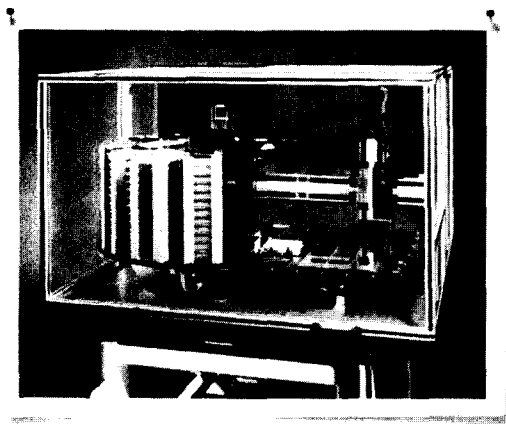
기판에 고정하는 방법은 유전자를 소위 spotting 하는 방법이 있다. 이 방법은 고체 기판의 표면을 적절히 화학적으로 처리하고, 유전자를 표면에 spotting 한 후, 적절한 후처리로 유전자를 고정한다. 따라서 어떤 방식으로 유전자를 spotting 하느냐에 따라 차이가 있다. 다양한 형태의 microarray spotter가 상용화되어 있다. 따라서 유전자 밀도는 최대 표준슬라이드 glass 표면(2.5cm×7.5cm) 안에 수 만개의 유전자를 spot하고 있다.

III. DNA microarray 분석 기술

DNA microarray 분석은 hybridization 반응을 마친 microarray 영상을 받아, 적절한 영상 처리 후, 영상내의 위치 정보로부터 대상 유전자의 정보를 알아내고, 궁극적으로는 기존 유전자 데이터뱅크(예를 들면, Gene Bank)로부터 생물학적 해석을 하는 부분까지 포함된다. 따라서 이 부분적으로 Bioinformatics 분야의 내용도 포함된다. 본 글에서는 영상 분석과 영상을 데이터 베이스화 하는 부분까지 기술하고, 좀더 Bioinformatics 분야의 내용은 다른 분의 글을 참조하길 기대한다.

1. 염기서열 분석을 위한 microarray 영상 분석

염기서열 분석은 기본적으로 대상 DNA와 microarray상에 고정화된 DNA가 hybridization을 할때 상보적으로 반응한다는 가정에서, 대상 DNA 정보를 위치 정보로부터 유추할 수 있다는 가정이다. 따라서 microarray 영상으로부터 형광이 밝혀진 dot의 위치를 찾아내는 것이 가장 중요한 일이며, 정확한 위치를 찾아내는 것은 비교적 쉬운 일이다. 그러나 Affymetrix의 고밀도 oligonucleotide DNA microarray의 대개 25base 염기로 이뤄졌으며, 이 짧은 염기 서열을 이용하여 대상 DNA를 분석하기 위한

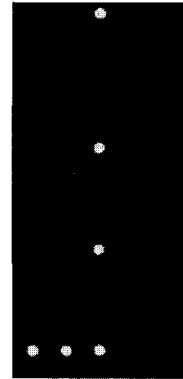


<그림 4> 상용화된 DNA microarrayer의 모습 (Genemachine사의 OmniGrid Microarrayer)

microarray 표면 고정화한 DNA sequence의 유사성을 이하에 자신 긴 대상 DNA 서열을 분석할 수 있다. 마치 유사한 조각을 가지고 전체 모습을 유추할 수 있는 알고리즘이 필요하다는 것을 알 수 있다^[3]. Affymetrix의 DNA microarray와 달리 spotting 방법에 의한 oligonucleotide의 경우는 microarray 위에 고정화하려는 oligonucleotide를 잘 설계하여, oligonucleotide 자체가 대상 DNAdp 특이하게 반응할 수 있도록 설계를 한다. 이러한 예로 저자의 회사에서 개발한 자궁경부암 진단용 DNA microarray는 아래와 같이 설계가 되었다. 임상적으로 자궁경부암을 일으키는 원인균으로 인유두종바이러스(human Pappilomavirus; HPV)가 알려져 있다. 그런데 HPV는 실제로 100여종의 바이러스가 있으며, 이들 중 약 20여종의 바이러스가 자궁경부암의 일으키는 아주 위험한 바이러스로 밝혀졌다. 많은 임상 의사는 바이러스의 감염 여부와 함께 어떤 종류의 바이러스 즉 위험군의 바이러스에 감염 여부를 중요한 임상 정보로 알고 싶어 하며 이의 진단적 가치는 향후 자궁경부암 발병과도 긴밀한 관계가 있음을 알 수 있다. 따라서 HPV의 유전형을 구별할 수 있는 DNA microarray는 HPV의 DNA 정보

58	● ●	○	○ ○	44
56	● ●		○ ○	43
52	● ●		○ ○	42
51	● ●		○ ○	40
45	● ●	○	○ ○	34
39	● ●		○ ○	11
35	● ●		○ ○	6
33	● ●	○	● ●	69
31	● ●		● ●	68
18	● ●		● ●	66
16	● ●	○	● ●	59

〈그림 5〉 on-off 형 진단용 HPV DNA microarray의 구성



〈그림 6〉 HPV 16번에 감염된 환자의 실험 결과

로부터 각 바이러스마다 특이하게 보유하고 있으며, 구별이 가능한 부위를 찾아 내고, 이에 적절하게 hybridization이 가능한 oligonucleotide를 설계한다. 이를 바탕으로 〈그림 5〉와 같이 DNA microarray를 구성한다^[4].

환자의 자궁경부로부터 얻은 세포진으로부터 DNA를 추출한 후, HPV DNA Chip과 hybridization을 할 경우, 〈그림 6〉과 같은 결과를 얻을 수 있다

이 경우, 비교적 간단한 영상 처리를 통하여 임상진단용으로 응용할 수 있다. 이러한 HPV-DNA Chip에서는 결과 영상으로부터 진단 결과에 대한 자동인식에 관한 연구를 수행하고 있다^[5].

2. 유전자 발현 정도 분석을 위한 Microarray 데이터 처리

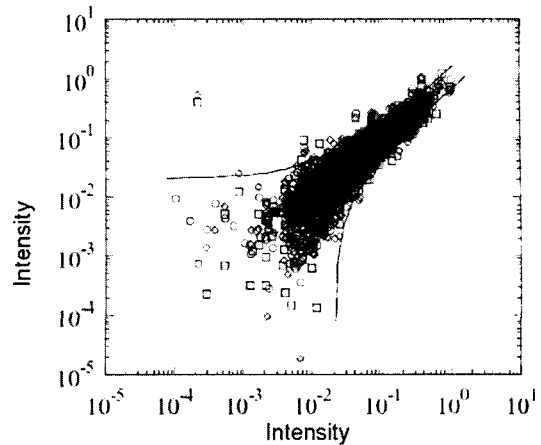
신약 개발이나 질병 역학 조사 등에서 이용되는 유전자 발현 정보 분석용 DNA microarray는 고집적도를 요구된다. 상술한 바와 같이 염기 정보를 알고 있는 기준 표본과 염기정보를 분석하고자 하는 목적 표본에 각기 다른 색의 형광물질로 labeling하여 고집적도의 DNA microarray에 hybridization 반응시킨 다음, 고분해능의 스캐너를 사용하여 다른 색에 대한 두 개의 영상을 얻는다. 이 영상들을 분석하여 염기정보를 알고자 하는 표본과 기준 표본과의 차이를 분석한다.

이와 같은 응용분야에 있어서는 정확한 분자량을 측정해야 하기 때문에 칩의 각 probe의 위에서 반응한 부분을 정확히 찾아내야 하고 반응정도를 각 영상의 밝기 정보를 이용하여 비를 구하는 것이 중요하다. 그러므로 형광영상의 취득에서부터 발현비(expression ratio)을 알아낼 때까지 각종 영상처리 기법을 동원하고 있다. 형광영상은 매우 작은 신호를 취득하기 위하여 dark current, spectral crosstalk, integration time, camera gain 등의 영상취득 장치의 모든 부분에 대한 고려가 필수적이다^[7]. 영상 취득 후에도 지역적 불균일을 해소하기 위하여 가우시안 모델이나 모폴로지를 이용하여 배경교정(background correction)을 하게 된다. 정확한 영상취득을 위한 기능은 스캐너의 기능에 포함되기도 한다.

디지털 영상을 취득한 후에 필요한 영상처리 기능은 크게 세 부분으로 나눌 수 있는데, array 격자의 위치를 알아내고, probe spot에서 hybridization된 부분을 정밀하게 분리한 후에, 각 probe spot에서 발현비를 알아낸다^[6-9, 10]. 대부분 상당히 많은 probe spot을 포함하고 정확한 arrayer(혹은 dotter)를 사용하여 공간적으로 규칙적으로 배열하기 때문에 probe의 근사 위치를 알아내는 것은 간단한 편이다^[6, 8, 10]. 예를 들면 원형 템플릿을 사용한 템플릿 정합^[6], 2-D periodogram^[8], Median Projection^[5] 등의 영상처리 기법을 단독 혹은 조합하여 사용한다.

그러나 probe spot에서 hybridization된 부분을 찾아낼 때에는, 신호대 잡음비를 높여야 하기 때문에 비교적 복잡한 방법을 사용한다^[7-9, 10]. 여기서는 Thresholding, 모폴로지, Projection, Hugh transform, 외에 Mann-Whitney 영상분할법 등 각종 방법들이 시도되었다.

영상처리의 최종 목적이라고 할 수 있는 발현비를 구하기 위하여, 일반적으로 영상분할이 잘 되었다는 가정하여 평균 밝기의 로그비(log ratio)를 사용하지만^[10], 영상분할이 불완전한 경우를 보완하기 위하여 함수모델을 사용하기도 한다^[6]. 이들은 실제로 형광 밝기가 분자량과 정확



〈그림 7〉 probe들의 형광 각 색의 밝기의 분산도

하게 관련되어 있다는 가정을 하고 있으나 실제로는 여러 가지 잡음의 영향을 받는다. 만일 기준표본과 목적 표본이 같은 염기로 구성되어 있다면 분자량에 관계없이 그 형광 밝기의 비가 일정하여, 〈그림 1〉과 같이 각 형광 색의 밝기를 축으로 하는 그래프에서 각 probe의 밝기 측정값의 분산도는 45° 선에 모일 것이다. 그러나 실제로는 〈그림 7〉에서 보는 바와 같이 상당히 오차가 있으며, 형광 밝기가 낮은 부분에서 더욱 심하다. 발현비의 용도의 하나로 일부 염기 서열에서 차이를 찾아내려면, 그 probe spot이 있는 곳에서의 밝기 비는 45° 선에서 많이 떨어질 것이다. 그러므로 상당히 많은 probe 위치 각각에서 구한 ratio들의 분포를 분석하는 것도 중요한 작업 중의 하나가 된다^[9, 10]. 이밖에 상기와 같은 분석을 통하여 얻은 정보를 여러 연구기관에서 공유하기 위한 정보의 자료화, 데이터 접근 방법 개발, 대단위 처리를 위한 병렬처리 등 또한 필수적이 되고 있다^[8].

3. DNA microarray의 데이터 표준화

바이오 정보의 표준화는 지금 중요한 과제로 대두되어 있으며, DNA microarray 또한 예외일 수 없다. 이는 XML을 기반으로 DNA microarray 영상 분석 소프트웨어에 내재되어 있어 사용될 예정이며, 이에 대한 표준화의 한 방

향으로 Microarray Gene Expression Data Society(MGED society)의하여 제안된 Minimum Information About Microarray Experiments(MIAME) 안으로 유력한 표준안이 되고 있다. MIAME의 경우, 6가지 중요한 부분으로 나뉘어져 있다; (1) 실험 설계, (2) 칩의 레이아웃, (3) 표준 추출 및 labeling, (4) hybridization, (5) normalization과 조절, (6) 측정. 향후 추진 방향을 살펴보아야 하겠지만, 이러한 표준화는 당연한 과제이며, 아주 빠른 속도로 진행될 것으로 예상된다.

IV. DNA microarray 향후 발전 방향

본 글에서 주로 영상 분석을 바탕으로 하는 DNA microarray에 관한 분석기술을 다뤘지만, 형광 DNA microarray scanner의 가격이 비싸고, 형광을 통한 측정의 한계, 이동성이 없다는 점등이 이러한 DNA microarray 기술이 분자유전진단 분야에 폭 넓게 수용되고 있지 못하다는 이유로 들 수 있다. 따라서 다른 종류의 검출 방법을 이용하는 DNA microarray들에 대한 연구도 활발히 진행되고 있다. 이들 중 유력한 경쟁 기술은 microelectrode를 이용한 DNA microarray, bead를 이용한 array 기술, microcapillary electrophoresis를 이용한 DNA 분석 기술 등을 들 수 있다. 그러나 이들 기술 또한 각각이 제한 요소를 지니고 있다. 현재 DNA microarray 기술은 대용량 검출 기술(High throughput Screening)로 발전하고 있으며, cDNA보다는 oligonucleotide를 이용한 microarray의 개발이 훨씬 우세하게 진행되고 있다.

V. 결 론

본 글에서는 DNA microarray의 분석 기술

을 중심으로 기본적이고 상용화가 진행된 분야에 관련한 기술에 대하여 소개하였다. 초기에 DNA를 유리기판에 고정화하고, 이에 대한 상보적 Hybridization을 이용한 유전정보 분석에 부분이 주요 내용이었다. 그러나 이 분야의 연구 개발 속도가 엄청나게 빠르게 진행되고 있는 만큼 소개된 내용이 최신 기술이 아닐 수 있다고 평가한다. 이 분야의 연구는 생물학, 컴퓨터공학, 화학, 전자공학, 기계공학 등 모든 분야의 기술이 융합되어 발전하고 있는 분야 중에 하나이기 때문에, 끝없는 노력이 필요하다고 생각한다.

참 고 문 헌

- [1] Maskos, U. and Southern, E. M., "A novel method for the analysis of multiple sequence variants by hybridization to oligonucleotide", *Nucleic Acids Res.*, Vol. 21, pp.2267-2268, 1993
- [2] Stephen PA Fodor, J Leighton Read, Michael C. Pirrung, Lubert Stryer, Amy Tsai Lu, Dennis Solas, "Light-directed, spatially addressable Parallel Chemical Synthesis", *Science*, Vol. 251, pp.767-771, Feb., 1991
- [3] Mark Chee, Robert Yang, Earl Hubbell, Anthony Berno, Xiaohua C Huang, David Stern, Jim Winkler, David J. Lockhart, Macdonald S. Morris, Stephen PA Fodor, "Accessing Genetic Information with high-density DNA Arrays", *Science*, Vol. 274, pp.610-614, Oct. 1996
- [4] Nam Hoon Cho, Hee Jung An, Jeongmi Kim Jeong, Suki Kang, Jae Wook Kim, Young Tae Kim, and Tchan Kyu Park, "Genotyping of 22 human papillomavirus types by DNA chip in Korean women with cytologic

- diagnosis”, Am J Obstet Gynecol, 188, pp. 56-62, 2003.
- [5] J. D. Kim, S. K. Kim, J. S. Cho and J. Kim, “Knowledge-based image processing for on-off type DNA microarray,” Proceedings of SPIE vol. 4623, pp. 38-46, 2002.
- [6] N. Brandle, H. Bischof and H. Lapp, “A generic and robust approach for the analysis of spot array images,” Proceedings of SPIE vol. 4266, pp. 1-12, 2001.
- [7] L. M. Kegelmeyer, L. Tomsascik-Cheeseman, M. S. Burnett, P. van Hummelen and A. J. Wyrobek, “A groundtruth approach to accurate quantitation of fluorescence microarrays,” Proceedings of SPIE vol. 4266, pp. 35-45, 2001.
- [8] Z. Z. Zhou, J. A. Stein and Q. Z. Ji, “GLEAMS: A novel approach to high throughput genetic micro-array image capture and analysis,” Proceedings of SPIE vol. 4266, pp. 13-23, 2001.
- [9] T. Bergermann, F. Quiaoit, J. Delrow and L. P. Zhao, “Statistical issues in signal extraction from microarrays,” Proceedings of SPIE vol. 4266, pp. 24-33, 2001.
- [10] Y. Chen, E. R. Dougherty, and M. L. Bittner, “Ratio-based decisions and the quantitative analysis of cDNA microarray images,” Journal of Biomedical Optics 2(4), pp. 364-374, Oct. 1997.

저자 소개



김종원

1985년 2월 서울대학교 자연대학 물리학과 학사, 1987년 2월 서울대학교 자연대학 물리학과 석사, 1992년 8월 서울대학교 공과대학 의공학협동과정 박사, 1987년 8월~1994년 6월 : 서울대학교병원 의공학과 연구원, 1994년 9월~현재 : (주) 바이오메드랩, 대표이사, <주관심 분야 : 바이오칩, Lab-on-Chip, 인공장기>



김종대

1982년 2월 서울대학교 공과대학 전자공학과 학사, 1984년 2월 한국과학기술원 전자공학과 석사, 1990년 8월 한국과학기술원 전자공학과 박사, 1988년 3월~2000년 2월 : 삼성전자주식회사 중앙연구소 수석연구원, 1994년 12월~1995년 12월 : UCSB 방문연구자, 1996년 1월~1997년 2월 : UCLA 방문연구자, 2000년 3월~현재 : 한림대학교 정보통신공학부 조교수, <주관심 분야 : 신호 및 영상처리, 멀티미디어 통신, 바이오영상처리>