

Microarray를 이용한 유전자 발현 연구

김 선 주

경상대학교 의과대학 임상병리과학교실

Microarray Gene Expression Analysis

Sunjoo Kim, M.D.

Department of Clinical Pathology, Gyeongsang National University, College of Medicine, Chinju, Korea

Arrays are used for many purposes, most prominently to measure level of gene expression for tens of thousands of genes simultaneously. Microarrays consist of probe DNA attached to the solid surface (chips), which is hybridized with fluorescent labeled cDNA samples. Microarrays can be classified into two categories. Oligonucleotide array uses 25 bp oligonucleotide probe DNA that is synthesized in situ on the solid surface photolithographically. Avidin-phycoerythrin is used to develop fluorescence. Compared to oligonucleotide array, cDNA microarray is more flexible to customize. The size of cDNA produced by RT-PCR, ranges from 250 to 750 bp. Spotter and two-color scanner are needed for cDNA microarray. The ratio of Cy5/Cy3 is plotted and clustered for each gene to see the expression pattern. With the fast growing computer technologies and improved understanding for the genes, microarray is a promising tool to enhance our knowledge about genetically complicated diseases. We can choose the most effective drug to the specific pathogen for each person according to the expression pattern. Gene therapy would be applied prophylactically for the patients who have disease marker genes in the future. In spite of high price and technical subtleness of microarray, we had better to be accustomed to this system because we are living in the post-genomic era. (*Korean J Clin Microbiol* 2001;4(2):82-86)

Key words : Microarray, Oligonucleotide array, cDNA microarray

서 론

인간의 세포에는 약 30,000개의 mRNA가 있지만 이들이 같은 빈도로 존재하는 것이 아니다. 이 중 99%는 매우 드물게 발현되고 약 300개의 유전자가 세포 mRNA의 절반을 차지한다. 세포나 세균은 실험조건에 따라 수 백 개 혹은 수 천 개의 mRNA가 동시에 증가하거나 감소할 수 있을 것이다. mRNA의 증감을 보기 위해 Northern blot이 많이 사용되어 왔고, 최근 RT-PCR, competitive RT-PCR, real-time PCR 등이 이용되고 있다.

그러나 이들 방법은 몇 개의 mRNA의 증감은 관찰할 수는 있어도, 세포내의 수 백 개 혹은 그 이상의 mRNA의 증감을 관찰하는 데는 한계가 있다. 이를 위해 개발된 방법들이 1) SAGE [1], 2) differential display [2], 3) oligonucleotide array [3], 4) cDNA microarray [4] 등이다. 이 글에서는 후자 두 가지에 대해서만 설명하기로 한다. 현재는 고가의 장비와 칩, 소프트웨어, 시약 때문에 검사실에서 일상적으로 사용할 수는 없지만, 앞으로 이들의 가격이 점차 낮아져 몇 년 후에는 보편적인 검사가 될 수도 있다. 인간 유전자 게놈 연구가 거의 완성 단계에 와 있어 각 유전자의 기능과 상호 연관성에 대한 연구가 광범위하게 이루어질 것이다. 이에 microarray에 대한 이해를 높이고자 그 원리와 특징을 고찰하고자 한다.

접수번호 : CM 4-02-12

교신저자 : 김선주

(660-702) 경남 진주시 칠암동 90

경상대학교 의과대학 임상병리과학교실

Tel : 055) 750-8239 Fax : 055) 762-2696

E-mail : sjkim@nongae.gsnu.ac.kr

Microarray란 무엇인가?

Microarray란 유전자 조각들이 유리표면(칩)에 부착하여 배열된 것을 말한다. 이 칩은 형광표지된 cDNA 검체와 보합을 이루고 각 spot의 형광강도를 측정하여 수치화 된 자료를 얻게 된다. 미세유체학(microfluidics)의 발달로 작은 칩 공간에 수 만개의 유전자 조각들을 집적할 수 있어서 전체 유전자 양상을 관찰할 수 있게 되었다. 그러나 한 실험에서 너무 많은 자료가 수집되어 이를 분석하는데 오랜 시간이 걸리는 단점이 있어, 실험 고안을 잘 하는 것이 중요하다. Microarray는 크게 oligonucleotide array와 cDNA microarray 두 가지로 분류할 수 있는데, Fig. 1에 그 차이점을 도식화 하였다.

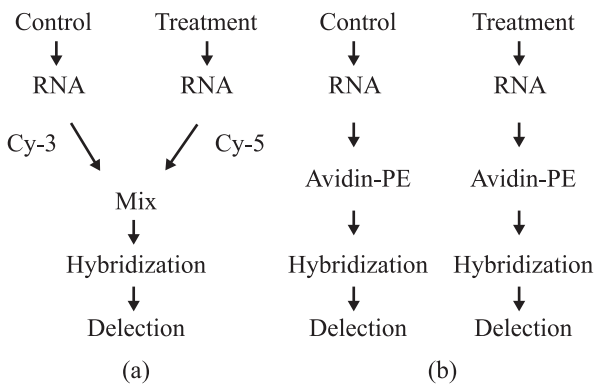


Fig. 1. flow chart of (a) spotted microarray, (b) oligonucleotide array.

가. Oligonucleotide array

Fodor 등[5]에 의해 개발된 이 방법은 Affymetrix 회사에 의해 상품화되었다. Photolithography에 의해 염기를 부착하여 1개의 oligonucleotide를 합성한다. 아직은 수요자가 원하는 유전자만 든 칩은 사용할 수 없지만 이것도 곧 상품화될 것으로 보인다.

나. Spotted microarray

이 유전자 칩은 실험자가 자가 제작할 수 있어 원하는 유전자만 관찰할 수 있는 유연성이 있다. 탐식 유전자를 펜이나 핀에 묻혀 로봇 시스템을 이용해 array 표면에 분주한다. 탐식 유전자로는 cDNA 클론이나 DNA 유전자 탐식자, PCR 증폭산물 등을 사용한다.

다. 명명 및 정의

용어에 혼동이 있는 경우가 많다. Duggan 등[6]은

array 표면에 부착된 유전자 조각을 '탐식자(probe)'로 하고, 검체로부터 분리되어 표지되고 보합반응에 사용되는 유전자를 '검체(sample)' 혹은 '목표물(target)'로 정의하였다. 하지만 많은 논문에서 검체에서 분리된 유전자를 탐식자로 혼용하고 있는 것을 발견할 수 있다.

reference probe - target 염기서열과 상보적인 탐식자(perfect match)

partner probe - reference 탐식자와 1개의 염기가 다른 탐식자(mismatch)

feature - 한 탐식자가 10^6 - 10^7 개 든 array의 구역(Fig. 2)

Microarray를 구축하기 위해 고려해야 할 사항들

1. 장비

1) Spotter

일정한 크기의 spot이 array 표면에 정확히 찍혀야 하고 carryover가 최소화되어야 한다. 384-well microplate로부터 nanoliter 이하의 양을 운반하기 때문에 자동화된 로봇 분주기를 이용한다. 습도가 일정해야 하고 먼지가 들어가지 않아야 한다. 재현성은 변이계수가 15% 이내면 적합하다고 본다.

2) 스캐너

형광 검출은 microarray 실험에서 매우 중요하다. 두 가지 형광염색시약을 사용하는 경우 서로 다른 형광파장을 검출할 수 있어야 한다. 약간의 변화도 감지할 수 있도록 역치가 낮을 뿐 아니라, 반대로 큰 변화도 측정할 수 있도록 넓은 측정 범위(linear dynamic range)를 갖고 있어야 한다. FITC (fluorescein thiocyanate)는 Cy (Cyanin)에 비해 스캔 후 형광강도가 급속히 감소하므로 부적합하다. 두 파장을 읽을 때 서로 다른 파장에 대해 간섭할 수 있는데, 특히 낮은 파장(Cy3, 550 nm)에서 높은 파장(Cy5, 650nm)에 영향을 미친다.

2. Array 표면 화학

Microarray에 사용하는 현미경 슬라이드는 표면이 매우 편평하고 형광이 없으며 질산에 의해 깨끗이 세척되고, silane에 의해 화학 처리되어야 DNA가 균질하게 부착될 수 있다. 슬라이드는 열, 산, 소독제 등에 내구성을 보여야 한다.

3. DNA 부착

cDNA나 PCR 증폭 산물을 silane 도포된 슬라이드에 부착하기 위해 sodium thiocyanate (NaSCN)을 사용한다. 탐식자 유전자는 통상 250-750 bp 크기이고, 양은 1 nl, 농도는 100 pg/spot이다. 반면 Affymetrix 시스템(oligonu-

celotide array)은 spotter를 사용하지 않고 photolithography법으로 직접 고형 표면에서 유전자를 합성한다.

4. 검체 유전자 분리

Microarray 실험은 고가의 비용이 들기 때문에 실험 전 과정에서 신중을 기해 재현성이 높도록 해야 한다. 특히 검체 유전자 분리 과정은 수작업으로 하기 때문에 검사자마다 다른 결과를 보일 수 있다. 만약 분리된 RNA가 파괴되거나 오염이 되어 있다면 실험 전 과정에 큰 영향을 미칠 수 있다. 파괴되지 않은 RNA를 분리하기 위해서는 세포 내 ribonuclease를 신속히 불활성화 시켜야 한다. DNA는 크기가 크기 때문에 분리 효율이 낮다. 검체 유전자 분리의 재현성을 높이기 위해서는 동일한 회사의 동일한 lot 시약을 사용하고 같은 사람이 일관된 방법에 의해 실험을 하는 것이 좋다.

5. 검체 유전자 표지

순수하게 분리된 RNA 혹은 DNA를 형광물질로 표지한 후 표지가 안된 염기 조각이나 작은 조각(<200 bases)은 제거한다. Cy5는 스캔 후 형광이 약해지고 동결-해동에도 약하므로 주의해서 다루어야 한다. 표지와 보합 반응이 비특이적인 경우 전혀 다른 결과를 보일 수 있으므로 주의해야 한다. Cy3 (green)는 Cy5 (red)에 비해 양자(quantum) 회수율이 낮기 때문에 강도가 약하게 나올 수 있다.

6. 교잡

Array 탐식자 유전자는 표지된 검체 유전자에 비해 충분한 양(20-30 pg)이 있어야 한다. Membrane을 사용하는 경우 유리 슬라이드에 비해 많은 양의 DNA가 부착할 수 있지만, 교잡 강도가 낮고 고밀도 array를 제작할 수 없으며 2색 형광이 불가능하여 많이 사용되지 않는다. 교잡반응은 42℃에서 포르말린이 함유된 완충액에서 반응시킨다.

7. 자료 분석

Microarray는 방대한 자료가 생성되기 때문에 생명과학과 전산, 정보를 동시에 다룰 수 있는 '생명정보학(bioinformatics)'에 대한 수요가 커질 것으로 기대된다. 자료는 그래프로 나타나는데(16비트, .tif 파일), 10⁴배까지 볼 수 있어 넓은 측정 범위를 표시할 수 있다. 2색 형광인 경우 Cy3/Cy5의 비율로 표시되어 대조군에 비해 실험군에서 각 유전자의 증감 정도를 알 수 있으며, 증감 패턴에 따라 유전자를 군집(cluster)화 할 수

있다. 자료를 얻거나 분석하는 데는 상품화된 소프트웨어를 구입하여 사용해야 한다.

Oligonucleotide array

Affymetrix 회사에서 GeneChip™이라는 이름으로 상품화 되어 있다. 탐식자가 광화학적으로 칩에서 직접 생성되기 때문에 클로닝이나 spotting, PCR 과정이 필요 없다. 매우 높은 밀도로 수 십만 개의 탐식자를 부착할 수 있고, 탐식자의 염기서열(25 bp)도 정확히 알고 있다. 반응 강도를 측정하여 보합 양상을 분석하는 소프트웨어가 많이 개발되어 있다. Reference 탐식자와 partner 탐식자가 붙어 있어서 실험 조건에 따른 교차 반응에 의한 비특이적 반응 여부를 판단할 수 있다. 탐식자 염기 부착은 de-protection, chemical coupling, capping 과정에 의해 이루어지며, 1.28×1.28 cm array의 경우 260,000개의 feature, 약 7,000개 이상의 유전자 혹은 EST(expressed sequence tags, 기능이 밝혀지지 않은 유전자)를 분석할 수 있다(Fig. 2). Feature 1개의 크기는 14×14 μm이다. 칩은 플라스틱으로 덮여 있기 때문에 다루기 편하고, 일정한 양(200 μl)만 chamber내로 들어가기 때문에 보합반응 과정에서 에러를 줄일 수 있다.

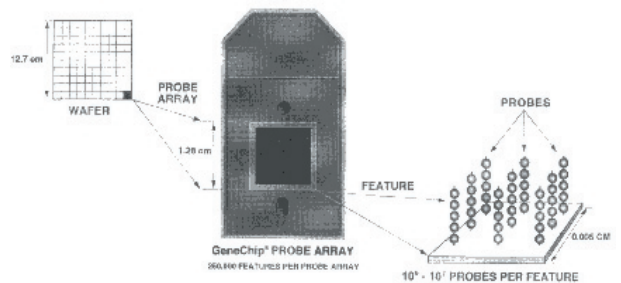


Fig. 2. Schematic picture of GeneChip™ probe array.

유전자 발현 연구를 위해 전체 RNA나 mRNA를 사용할 수 있다. mRNA는 cDNA나 cRNA로 전환하고 이 과정에서 biotin이나 형광물질로 표지하고 약간의 증폭도 이루어진다. 만약 돌연변이를 보기 위한 실험이라면 유전자 양이 문제가 안되므로 DNA를 분리하여 증폭한 PCR 산물을 sample DNA로 사용한다. Biotin이 표지된 유전자인 경우 streptavidin-phycoerythrin conjugate를 반응시킨다. 레이저 confocal 형광 스캐너를 사용하여 형광정도를 측정한다. 교잡반응의 강도는 mRNA 양과 잘 일치한다[7,8]. 전 실험과정 중 검체 RNA를 분리하는 과정에서 에러가 가장 크다.

Oligonucleotide array의 적응증은 크게 1) 유전자 발현

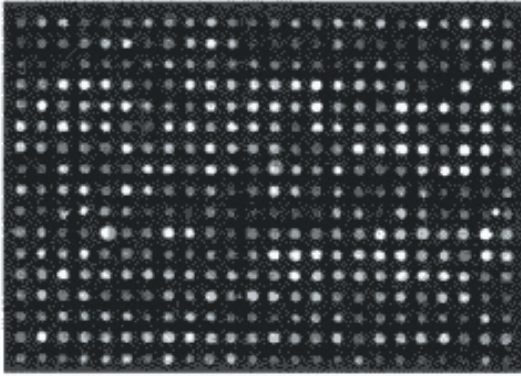


Fig. 3. Fluorescence image of oligonucleotide array produced by confocal scanner.

분석, 2) 유전자 변이 이다. 유전자 발현은 어떤 유전자의 발현이 다른 유전자와 동시에 되는지, 혹은 어떤 조직에서 더 잘되는지, 근접한 loci에 있는 유전자들이 동시에 발현되는지 등을 살피는 것이다. 유전자 변이는 진단과 치료에 많이 사용될 전망이다. 어떤 질환에서 어떤 변이가 잘 생기는지 알 수 있고, SSCP나 염기서열 분석에 비해 대량의 검체를 분석하는데 유리하다[9].

cDNA microarray

PCR에 의해 증폭된 cDNA를 로봇 시스템을 이용해 유리 슬라이드에 고밀도로 분주한다. 검체로부터 분리한 유전자를 두 가지 형광 물질로 염색하여 보합 시킨 후 두 형광의 비를 구한다(Fig. 4).

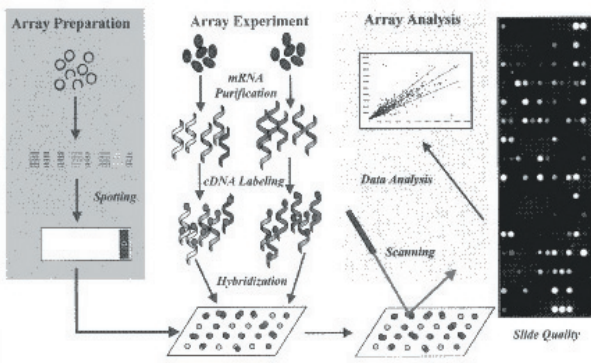


Fig. 4. Flow chart of two-color microarray experiments.

1. 탐식자 유전자 spotting

poly-L-lysine 혹은 aminosilane으로 코팅 된 유리 슬라이드에 양이온 교환 컬럼을 통과시킨 PCR 증폭 산물을

spotting한다. 일정한 양을 옮기는 핀 혹은 펜을 사용하는데 spot 완충용액은 재현성에 영향을 미친다. Spot의 크기나 모양은 온도나 습도의 영향을 받는다.

2. 검체 유전자 준비

1) RNA 분리

Trizol™ (Life Technologies)이나 RNeasy System (Qiagen), ToTALLY RNA Isolation Kit (Ambion) 등을 사용해 RNA를 분리한다.

2) RNA 표지

RNA로부터 cDNA를 합성하는 과정에서 형광물질이 표지 된다. Total RNA 4 µg 혹은 poly (A+) RNA 1 µg이 역전사반응에 필요하고, cyanine 3-, 혹은 cyanine 5-dUTP가 cDNA 합성시 사용된다. 하지만 이 두 가지 형광물질이 같은 효율로 삽입되지 않기 때문에 형광물질 자체에 의한 오차를 배제할 수 없다. 이들 형광물질은 빛에 예민하기 때문에 이 표지 전 과정에서 빛에 노출되는 기회를 줄여야 한다.

3) 교잡

교잡과 세척시 교차 반응을 줄여 특이도를 높여야 한다. 교잡 전 1% BSA (bovine serum albumin)으로 아민기를 불활성화 해야만 cDNA의 비특이적 부착을 예방할 수 있다.

3. 자료 수집, 정상화 및 분석

1) 자료 수집

Cy3와 Cy5의 형광 강도를 confocal 레이저 스캐너로 측정하여 TIFF 이미지를 얻는다. 교잡이 잘 이루어 졌다면 전 영역이 고르게 형광을 낼 것이다. 탐식자가 침전되거나 먼지에 의해 오차가 생겨서는 안 된다. 비특

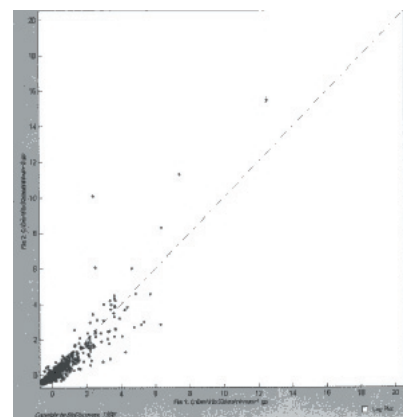


Fig. 5. Two-dimensional scattergram of Cy5 and Cy3 fluorescence intensities.

이적 반응(background)은 슬라이드 전체보다는 각 spot 주위의 배경 형광을 계산하여 판정한다. Spot 주위의 비특이적 형광을 뺀 교잡 반응 강도의 평균 혹은 중앙값을 구한다.

2) 정상화(normalization)

유전자 발현을 분석하기에 앞서 상대적인 형광 강도를 '정상화' 해야 한다. 정상화는 1) 두 형광의 표지 및 검출 효율이 다르고, 2) 두 검체 간 RNA 양이 다를 수 있기 때문에 필요하다. 정상화 방법에는 세 가지가 있다. 첫 째 Cy3와 Cy5에 의해 표지 된 RNA의 양이 같다는 전제 하에 전체 형광 강도를 비교하는 것이다. 각 spot의 형광 강도는 조금씩 다르더라도 수 천 개 spot의 형광 합이 같도록 조절한다면 그 오차가 줄어 들 것이다. 둘 째 선형 회귀 분석을 이용해 두 형광 강도 분포의 기울기가 1이 되도록 한다. 셋 째 house-keeping 유전자는 검체 종류와 상관없이 일정하게 유지된다는 가정 하에, Cy3와 Cy5에 의해 표지 된 house-keeping 유전자의 비를 1로 조절한다. 주로 많이 사용되는 house-keeping 유전자로는 rRNA와 GABDH (glyceraldehyde-3-phosphate dehydrogenase)가 있다. 낮게 발현되는 유전자는 변동폭이 커지기 때문에 정상화 과정에서 주의해서 다루어야 한다[10].

3) 분석

정상화 과정이 끝나면 유전자들의 발현이 2배 이상 의미 있게 증가하였는지, 반대로 감소하였는지 관찰하고 그 양상을 군집 분석한다[11]. 실험 결과를 신뢰할 수 있기 위해서는 적어도 3회 이상 처음부터 array 실험을 반복해야 한다.

Array 제한점과 전망

염기서열이 비슷한 경우 교차 교잡 반응이 생길 수 있고, 칩간에 다른 결과를 보일 수 있다. 수 천 개의 유전자를 다루기 때문에 약간의 오차에도 수 십 개 혹은 수 백 개의 유전자 결과가 다르게 나올 수 있다. 또 유전자의 상대적인 비를 측정하는 것이지 절대 양을 측정하는 것은 아니며, 유전자 비가 크다고 해서 생물학적 의미도 그 만큼 크다고 볼 수 없다. 탐색자 유전자의 대부분은 기능이 밝혀지지 않은 EST이므로 이들에 대한 기능 연구가 뒷받침되어야 한다. 유전자 발현이 조금밖에 안 된 경우 검출이 안 될 수 있고, 생명정보학에 대한 이해가 부족해 자료를 분석하고 해석하는데 어려움이 있다.

DNA array는 앞으로 암, 당뇨병, 고혈압, 백혈병 등 복합 유전자 질환의 진단, 약물 치료 효과의 예측, 돌

연변이와 다형성(SNP), 군주 분석[12] 등에 이용될 전망이다. 질환 가능성 유전자를 미리 찾아내 예방 의학적으로 대처할 수 있지만, 의료 윤리적인 문제가 수반될 수 있다. 질환이 더욱 세분화 될 것이며 감염균에 대해 가장 효과적인 약을 선택할 수 있을 것이다.

참 고 문 헌

1. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. (1995). Serial Analysis of Gene Expression. *Science* 1995;270: 484-7.
2. Liang P, Pardee AB. Differential display. A general protocol. *Mol Biotechnol* 1998;10:261-7.
3. Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, et al. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 1996;14:1675-80.
4. Lou XJ, Schena M, Horrigan FT, Lawn RM, Davis RW. Expression monitoring using cDNA microarrays. A general protocol. *Methods Mol Biol* 2001;175:323-40.
5. Fodor SP, Rava RP, Huang XC, Pease AC, Holmes CP, Adams CL. Multiplexed biochemical assays with biological chips. *Nature* 1993;364: 555-6.
6. Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM. Expression profiling using cDNA microarrays. *Nat Genet* 1999;21(Suppl. 1): 10-4.
7. Wodicka L, Dong H, Mittmann M, Ho MH, Lockhart DJ. Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nat Biotechnol* 1997;15: 1559-67.
8. de Saizieu A, Certa U, Warrington J, Gray C, Keck W, Mous J. Bacterial transcript imaging by hybridization of total RNA to oligonucleotide arrays. *Nat Biotechnol* 1998;16: 45-8.
9. Lockhart DJ, Winzeler EA. Genomics, gene expression and DNA arrays. *Nature* 2000;405:827-36.
10. Chen, V., E.R. Dougherty and M.L. Bittner. Ratio-based decisions and the quantitative analysis of cDNA microarray images. *J Biomed Optics* 1997;24:364-74.
11. Eisen, M.B., P.T. Spellman, P.O. Brown and Botstein. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 1998;95:14863-8.
12. Gingeras TR, Ghandour G, Wang E, Berno A, Small PM, Drobniewski F, et al. Simultaneous genotyping and species identification using hybridization pattern recognition analysis of generic mycobacterium DNA arrays. *Genome Res* 1998;8:435-48.