

For performing DNA Microarray experiments from beginning to end

권성우, 한종훈 (sw74@postech.ac.kr)

포항공과대학교 화학공학과 지능공정시스템 연구실

The introduction of DNA Microarray

DNA chip 은 기존의 분자 생물학적 지식에 현대에 엄청난 발전을 한 기계 및 전자공학의 기술을 접목해서 만들어 졌다. 기계, 자동화와 전자 제어 기술등을 이용하여 적게는 수백개부터 많게는 수십만개의 DNA 를 아주 작은 공간에 집어 넣을 수 있게 만든 것이다. 즉 DNA chip 이란 유전자 검색용으로서 엄청나게 많은 종류의 DNA 를 고밀도로 붙여 놓은 것을 말한다. 이러한 DNA chip 이 대체 할 수 있는 기존의 대표적인 분자생물학 방법으로는 Southern 과 Northern blot, 돌연변이 검색 그리고 DNA sequencing 등이 있다. 이와 같은 방법들과 가장 차이점은 동시에 최소한 수 백개 이상의 유전자를 빠른 시간 안에 검색할 수 있다는 것이다. 더욱이 이전 분자 생물학적인 실험 도구와 가장 큰 차이는, 측정 가능한 유전자의 양적 증가 뿐만 아니라 세포내 환경 변화에 따른 오케스트라 앙상블 연주와 같이 시간에 따라 변화하는 유전자의 상호 작용 까지도 DNA Microarray 을 통하여 측정할 수 있다. 또 하나의 다른 점은 Southern 이나 Northern blot 의 경우 유전물질을 붙이는 매체로 nitrocellulose 막을 사용하는데 반하여 DNA chip 에서는 유리와 같은 고형체를 사용한다는 것이다. 그럼으로서 DNA chip 은 아주 적은 양의 유전물질을 고밀도로 붙일 수 있게 되었고 동시에 많은 수를 검색할 수 있게 된 것이다. DNA chip 은 붙이는 유전물질의 크기에 따라 cDNA chip 과 oligonucleotide chip 으로 나누어 질 수 있다. 이들 DNA chip 의 이름에서도 알 수 있듯이 cDNA chip 에는 최소한 500bp 이상의 유전자 (full-length open leading frame 또는 EST)가 붙여져 있고, oligonucleotide chip 에는 약 15 ~ 25 개의 염기들로 이루어진 oligonucleotide 가 붙여져 있다.

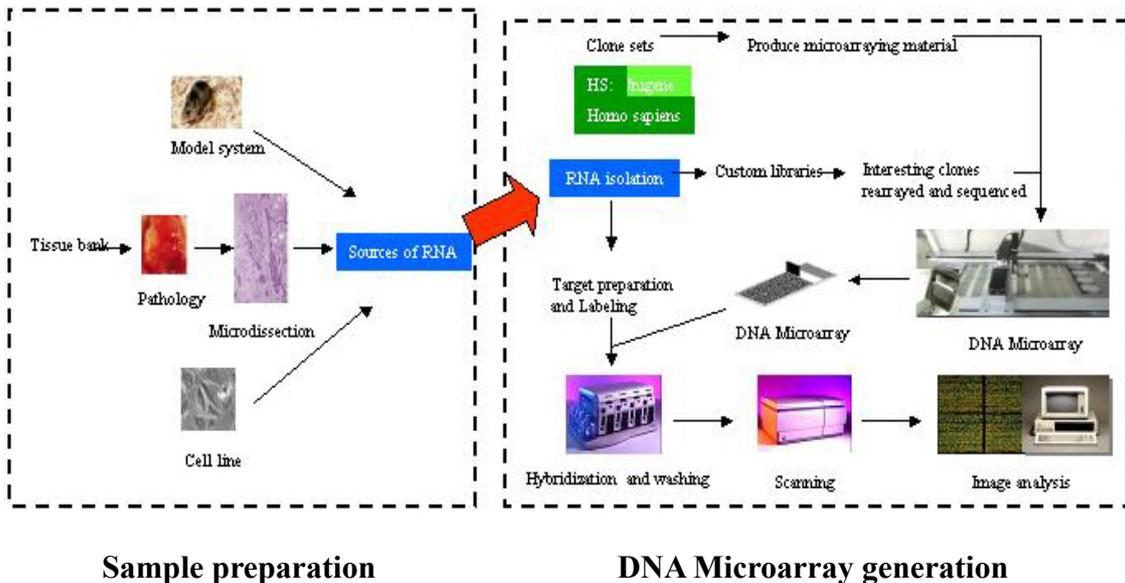


Fig 1, cDNA clone set으로부터 PCR products를 robotic arraying을 이용하여 DNA Microarray를 만든다. 만들어진 DNA Microarray를 사용하여 gene expression profile을 측정하는 과정도.

DNA Microarray의 실험은 일반적으로 세 단계로 나누어 볼 수가 있는데 (Fig.1), 첫 단계는 sample preparation이며 둘째 단계는 DNA Microarray generation이고 마지막 단계는 data handling과 interpretation이다. 본 보고서에서는 첫 단계와 둘째 단계에 초점을 맞추어서 작성이 되어 있다. 마지막 단계는 다음 번 보고서에서 다루어질 예정이다.

Sample preparation

cell line을 이용하는 경우는 homogeneous cell population으로부터 RNA를 얻기 때문에 비교적 쉬우며 이때 density, PH와 inducers의 영향에 주의 해야 한다.

cell line의 경우와 달리 tissue로부터 RNA preparation하거나 환자의 tissues에서 RNA preparation하는 것은 그 절차가 까다롭고 윤리적인 문제도 따름으로 주의해야만 한다. 수술 시 환자의 몸에서 tissues를 수집하는 것은 사전에 반드시 환자의 동의를 얻어야만 하며 somatic cell analysis가 아닌 germline cell analysis인 경우는 특히나 더 주의해야만 한다. Human tissues preparation에 따른 윤리적인 문제의 보다 상세한 내용은 다음 사이트에 있다 (<http://www.health.gov.au/nhmrc/issues/index.htm>) . 또한 환자로부터 동의를 얻는 과정에 대한 모델은 <http://www.pmc.unimelb.edu.au/>에 상세히 기술 되어 있다. 한편 tissue bank로부터 얻을 때 주의 할 점은 각 sample

환자의 정확한 정보가 있는가 이다. 대표적인 tissue bank로는 National Cancer Institute (NCI)가 있으며 웹사이트 주소는 <http://www-chn.ims.nci.nih.gov/>이다. 또한 LifeSpan BioScience와 같은 회사의 tissue bank에서도 여러 가지 질병의 tissue를 제공하며 그 웹 사이트 주소는 <http://www.lsbio.com/>이다. 이렇게 얻어진 tissue는 normal tissue, inflammatory cell, necrotic tissue외에도 여러 등급의 암세포와 여러 가지 종류들의 세포가 포함 되어 있다. 따라서 이들은 여러 가지 세포들의 heterogeneous RNA가 합쳐지기에 복잡한 RNA expression profile을 가지게 된다. 그러므로 DNA Microarray 실험에 microdissection은 매우 중요하게 되는데 Microdissection이란 tissue에서 pure cell을 추출하는 방법으로 보통 Laser capture microdissection이 많이 사용된다. 이에 대한 자세한 설명은 <http://mecko.nichd.nih.gov/lcm/lcm.htm>와 <http://www.arctur.com>에 되어 있다.

DNA Microarray generation

The first step: obtaining clone set

DNA Microarray를 만들거나 아니면 구입 하는 방법이 있는데 보통 cost, speed와 human resources의 이용가능성을 고려해서 판단한다.

보통 DNA Microarray를 만드는데 있어서 가장 큰 문제들 중 하나는 clone set을 얻는 것이다. 연구하고자 하는 organism의 전체 gene이 identification 된 경우는 비교적 쉽게 얻을 수가 있지만 아직까지 *Saccaromyces cerevisiae*, *Caenorhabditis elegans*와 몇몇 unicellular microbe를 제외 하고는 full genome sequencing이 되지 않았다. 따라서 이러한 경우에는 DNA Microarray의 제작에 한계가 있을 수 밖에 없다.

Gene discovery를 하는 데 있어 전통적으로 mutation등의 방법을 사용했지만 이와 달리 large-scale expressed sequence tag (EST) sequencing이 최근 널리 사용이 되고 이 방법을 통해 많은 수의 gene 이 발견되었다. 그 예로 Washington university와 IMAGE (Integrated Analysis of Genomes and their Expression: <http://rsb.info.nih.gov/nih-image/>) 콘소시움과 Cancer Genome Anatomy Project (CGAP)로 인하여 100만개가 넘는 human EST가 발견되었다. EST sequence는 dbEST (<http://www.ncbi.nlm.nih.gov/dbEST/>)에 있다. EST와 달리 Gene bank의 sequence를 중복되는 서열을 한 개의 cluster로 나누어 놓은 Unigene (<http://www.ncbi.nlm.nih.gov/UniGene/>)이 있으며, 많은 수의 EST가 여기에 속하기도 한다. 이러한 각 cluster를 대표하는 유전자의 clone set은 여러 회사에서 판매 하고 있다. 또한 mouse, human, zebrafish, rat와 plant의 clone set은 institute for Genomic Research (<http://www.tigr.org/>)에서 이용할 수가 있다. 한편 알려진 유전자의 open reading frame (ORF)는 American Tissue Culture Collection (<http://www.atcc.org/>)에서 얻을 수가 있다.

대표적으로 clone sets을 판매하고 있는 회사로는 Incyte (<http://www.incyte.com/>)와 Research genetics (<http://www.resgen.com/index.php3>)가 있다. 이들 두 곳에서는 판매하는 clone은 restreaking과 resequencing을 통해서 clone validation을 한 것이며, clone set외에도 PCR product 형태의 purified한 DNA를 제공한다. 이 것은 DNA Microarray의 제작 시, bacterial colonies에 비해 T1 phage의 오염 및 plasmid의 분리 과정이 필요하지 않은 장점이 있다.

Human 이외의 다른 organism의 EST sequence의 경우 mouse (http://genome.wustl.edu/est/mouse_esthmpg.html)와 drosophila (<http://www.celera.com>)등이 있다.

The second step: fabrication

위의 단계로부터 확보된 clone set을 PCR에 의해 증폭하고 gel-filtration이나 precipitation을 통해서 detergent, PCR primer, protein을 제거한다. 이렇게 정제된 cDNA들은, 슬라이드 위에 robot에 의해서 slide glass등의 matrices에 심어진다. 보통 정제된 cDNA를 100-500 µg/ml 정도로 하여 칩을 제작한다. 이때 slide의 hydrophobicity와 침전된 cDNA의 부착력을 높이기 위해서, slide glass 표면에 poly-L-lysine, amino silanes나 amino-reactive silane을 코팅 한다. 또한 코팅을 하면 slide 표면에 있는 DNA droplet이 퍼지려고 하는 것을 방지할 수가 있다. 그 다음으로 ultraviolet을 이용하여 cDNA를 slide glass에 고정시키고, slide 표면의 양전하를 감소시키기 위해 succinic anhydride를 처리 하여 남아있는 amine과 반응 시킨다. 마지막으로 열이나 알칼리 처리를 통해 single-strand cDNA Microarray를 완성한다. cDNA Microarray의 경우 oligonucleotide Microarray에 비해서 probe의 접근이 어렵기 때문에 hybridization이 잘안되는 단점이 있다. 그러나 specific hybridization을 할 수가 있는 장점이 있다. 한편 oligonucleotide 칩인 경우에는 슬라이드 칩 위에 디자인된 oligonucleotide를 합성하며 칩을 제작하고 자세한 과정은 <http://www.affymetrix.com/>에 있다

The third step: target labeling and hybridization

실험하고자 하는 test 시료와 reference 시료에서 total RNA를 각각 분리한다. 이후 oligo-dT primer를 이용한 RT-PCR (reverse transcription polymerase chain reaction) 과정을 거치게 된다. 이 과정을 통하여 두 가지 형광 색소인 빨간색(Cy5)이나 녹색(Cy3)로 각각 표지 하여 시료의 cDNA를 합성 한다. 하나의 DNA Microarray당 보통 형광 색소를 50-200µg사용하며, 이때 유의 해야 되는 점은 cellular protein, lipid나 carbohydrate가 남아 있는 경우, 표지 된 cDNA가 slide 표면에 non-specific binding하기 때문에 반드시 충분히 정제된 RNA를 사용해야만 한다. 또한 시료로부터 RNA를 추출하는 방법은 비교적 간단하지만, DNA Microarray의 probe로

사용되는 cDNA 제조에 필요한 현재의 형광물질 표지 방법은 비교적 많은 양의 RNA를 요구하기 때문에, 시료의 양이 적을 경우 이를 이용하기 어렵다. 따라서 이 경우에는 ^{33}P -dCTP를 이용한 radioactive detection 방법이 사용하기도 한다. 이렇게 제작해 놓은 DNA Microarray에 두개의 시료 cDNA를 같은 양으로 섞어서 hybridization을 하고, hybridization이 안된 cDNA들을 씻어 낸다.

The forth step: Image analysis

DNA Microarray를 스캐닝 과정으로 레이저 형광 스캐너를 이용하여 각 spot의 형광정도를 읽어 들인다. 각각 유전자의 형광 정도는 그 유전자의 발현 정도를 알려주는 것으로 이들 정보는 DNA Microarray image analysis 프로그램에 의하여 분석되어 진다. Image analysis에 대한 보다 자세한 내용은 <http://www.scanalytics.com/product/hts/microarray.html#analysis>에 나와 있다.

Data management and mining

DNA Microarray의 gene의 정보를 관리하기 위해서는 데이터베이스 구축이 필수적이다. Hybridization 결과와 여러 번의 DNA Microarray 실험 결과와의 비교를 위한 알고리즘 또한 필요하다. 이들 데이터베이스 구축에 대한 실제 예들은 <http://www.biologie.ens.fr/en/genetiqu/puces/bddeng.html>에 있다.

한편 DNA Microarray data analysis의 기본은 correlation에 기반 하여 다변량 통계적인 방법이나 neural network 등의 machine learning 기법이 많이 사용되고 있다. 이러한 data mining 기술을 이용하여 Genetic network의 구축이나 여러 질병의 molecular classification과 pathogenesis등의 새로운 생물학적, 의학적 사실들을 밝힌다. 보다 자세한 Data management and mining 내용은 다음번 보고서에서 다루도록 한다.

The useful site

1. <http://www.gene-chips.com/> :

DNA Microarray- Leming Shi, Ph.D.

2. <http://www.bio.davidson.edu/courses/genomics/chip/chip.html>

DNA Microarray Methodology (flash animation)- Davidson College

3. <http://www.genechip.co.kr/> :

Genechip Research Center

4. <http://industry.ebi.ac.uk/~alan/MicroArray/IntroMicroArrayTalk/index.htm>:

An introduction DNA chip technology-Alan J. Robinson

Reference

1. Adams, M., et al. (1991). "Complementary DNA sequencing: expressed sequence tags and Human Genome Project". *Science*. 252,1651-1656.
2. Basset, D.E., et al. (1999). "Gene expression informatics-it's all in your mine". *Nature Genet.* 21, 51-55.
3. Boguski, M.S., et al. (1993). "dbEST-database for "expressed sequence tags". *Nature Genet.* 4, 332-333.
4. Cheung, V.G., et al. (1999). "Making and reading microarrays". *Nature Genet.* 21,15-19.
5. Cole, K. A., et al. (1999). "The genetics of cancer-a 3D model". *Nature Genet.* 21, 38-41.
6. Cowtell, D.L. (1999). "Options available-from starts to finish-for obtaining expression data by Microarray". *Nature Genet.* 21,25-32.
7. Duggan, D.J., et al. (1999). "Expression profiling using cDNA Microarrays". *Nature Genet.* 21, 10-14.
8. Ermolaeva, O., et al. (1998). "Data management and analysis for gene expression arrays", *Nature Genet.* 21, 19-23.
9. Friemert, C., et al. (1989). "Preparation of radiolabeled cDNA probes with high specific activity for rapid screening of gene expression". *Methods Mol. Cell Biol.* 1, 143-153.
10. Kallioniemi, A., et al. (1992). "Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors". *Science*. 258, 818-821.
11. Lennon, G., et al. (1996). "The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression". *Genomics*. 33,151-152.
12. Simone, N.L., et al (1998). "Laser-capture microdissection: opening the microscopic frontier to molecular analysis". *Trends Genet.* 14,272-276.