

# DNA microbeads array를 이용한 유전자 발현 해석 서비스 Megasort®, MPSS®의 데이터 해석

당사에서는 DNA microbeads array를 이용한 유전자 발현 해석 서비스로서 Megasort®를 이용한 발현 해석 서비스 및 MPSS®를 이용한 유전자 발현 profiling 서비스를 실시하고 있다. Megasort®는 주목할 만한 발현 변이를 나타내는 유전자를 총체적이고 고효율로 얻을 수 있는 발현 변이 유전자 screening법이며, MPSS®는 세포 내에서 발현하는 대부분의 유전자에 대해 저발현 유전자에서 고발현 유전자에 이르기까지 폭넓은 dynamic한 영역에서 발현량을 해석할 수 있는 절대적 유전자 발현 profiling법으로서 큰 호평을 받고 있다.

## Megaclone® · Megasort®의 개요

DNA microbeads array 기술은 Megaclone®과 이를 이용한 2종류의 유전자 발현 해석법인 Megasort®, MPSS® (Massively Parallel Signature Sequence)로 이루어진다<sup>1,2)</sup>. Megaclone®은 1개의 세포에 존재하는 수십만개의 mRNA 분자가 그 발현빈도를 유지한 채로 cDNA로써 beads 위에 cloning한다. 다시 말해 microbeads cDNA library이다.

Megasort®에서는 2종류 시료의 mRNA로부터 각기 다른 형광으로 표식한 probe (RNA)를 조제하여 Megaclone® beads 상에서 경합적으로 hybridization 하여 형광 강도비가 다른 beads를 cell sorter로 분리한다. 이로써 유전자 서열을 이미 알고 있는지에 상관없이, 세포 내에서 발현하는 거의 모든 유전자 중에서 발현 차이가 있는 유전자 단편을 총체적으로 구별할 수 있다<sup>3)</sup> (그림 1).

Megaclone®, Megasort®의 원리 · 기술에 대한 자세한 내용은 Life Science & Biotechnology 19, 21, 24, 28, 30호를 참조하기 바란다.

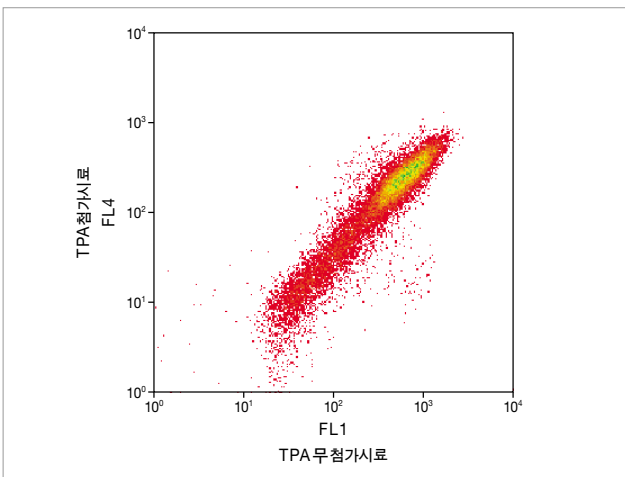


그림 1 Megasort® 해석 예 (Neutrophil 분화된 HL60세포의 TPA 첨가/무첨가)

## Megasort® 해석 표준데이터

Megasort®를 이용한 sorting에서 구별된 beads 위의 유전자는 서열 해석에 의해 동정된다. 통상 1회의 Megasort®당 400~2,000 clone의 서열을 해석한다. 얻어진 서열 데이터로서 clustering, alignment를 작성해, 그 결과 consensus 서열로 얻은 각 유전자 서열에 대한 homology search를 하여 annotation한다. 통상은 NCBI의 non redundant 핵산 데이터베이스 (nr)에 대해 homology search를 하지만, 고객의 요청에 따라 다른 데이터베이스에 대해 실시하는 것도 가능하다. 이렇게 얻어진 데이터를 종합하여 당사에서 개발한 viewer 형태로 제공된다.

Viewer에서는 어떤 유전자가 어떤 발현량의 변동 (up-regulated, down-regulated)으로 Megasort®에서 파악했는가를 간단히 볼 수 있으며, 키워드 검색을 통해 주목할 만한 유전자를 쉽게 찾을 수도 있다 (그림 2). 또한 단순히 서열 뿐 아니라 cluster별 alignment나 각 서열의 정확도를 볼 수도 있다 (그림 3). Megasort®를 이용한 해석을 다시 의뢰할 경우에는 의뢰자가 원하는 대로 지난 데이터를 통합한 해석 · viewer 작성도 제공한다.

## MPSS®의 개요

MPSS®는 Megaclone®에서 얻은 beads를 procell 안에 단층으로 충전한 초고밀도 beads array를 제작하여 각 beads에 cloning된 유전자를 한꺼번에 sequencing하는 기술이다<sup>2)</sup> (그림 4).

그 결과 beads에 cloning된 발현유전자의 signature (17~20 염기의 서열 정보)를 100만개 이상 얻을 수 있다. Signature의 서열정보를 데이터베이스와 대조함으로써 발현유전자를 살펴볼 수 있다 (annotation).

같은 signature의 개수는 시료의 세포내에서 발현하고 있는 상보적인 mRNA의 copy수를 나타내고 있으므로 signature 출현 빈도로 각 유전자의 발현빈도 정보를 얻을 수 있다. 따라서 모르는 것이든 이미 알고 있는 것이든 상관없이 세포 내의 거의 모든 발현유전자에 대해 절대적인 profiling을 할 수 있다. MPSS®에서의 발현 정보는 signature 개수라고 하는 디지털 수치로 얻어지기 때문에 hybridization을 토대로 한 종래의 DNA microarray와는 달리 몇 개의 복제에서 수만 개의 복제라고 하는 폭

check	cluster ID	contigID	S980 D1L1	S980 D1L1	homology search result	score	E value	consensus sequence	cluster overview
<input type="checkbox"/>	4	Contig1	0	3	gb BC025202.1  Mus musculus ryanodine receptor 1, skeletal muscle, mRNA (cDNA clone IMAGE:5029188), with apparent retained intron	858.0	0.0	446	
<input type="checkbox"/>	5	Contig1	1	9	ref XM_126119.3  Mus musculus myosin, heavy polypeptide 4, skeletal muscle (Myh4), mRNA	997.0	0.0	520	
<input type="checkbox"/>	6	Contig1	1	3	ref NM_021285.1  Mus musculus myosin, light polypeptide 1, alkali, embryonic (Myf1), mRNA	844.0	0.0	439	
<input type="checkbox"/>	7	Contig1	0	8	gb BC036292.1  Mus musculus ATPase, Ca++ transporting, cardiac muscle, fast twitch 1, mRNA (cDNA clone MGC:28518 IMAGE:4191741), complete cds	932.0	0.0	487	
<input type="checkbox"/>	7	S980_D1L1_P01_G09.esd	0	1	gb BC036292.1  Mus musculus ATPase, Ca++ transporting, cardiac muscle, fast twitch 1, mRNA (cDNA clone MGC:28518 IMAGE:4191741), complete cds	327.0	1e-86	279	
<input type="checkbox"/>	8	Contig1	0	2	ref NM_009416.2  Mus musculus tropomyosin 2, beta (Tpm2), mRNA	504.0	e-140	262	

그림 2 Viewer에 의한 Megasort® 결과 (서열 clustering data)의 일부

**P010U\_S980.5.C1 Overview**

name = S980\_D1L1\_P01\_F09.esd  
ID = 6  
length = 604 bp  
orient = reverse  
Offset = 128

**CONSENSUS (complementary)**  
CGT GCT GGAT CTTACGGAACTTGGCCAGGTTGACATTGGATTGTTCTCAGCCTCCTCAGCCTGTCTCTTGT  
GCACGACCTAGAAATGCCTTGAACCGGTCCAACCTGTAACCTAACAGGAGTCGGAGGAGTCGGACAGAGAAC

S980\_D1L1\_P01\_A09.esd  
S980\_D1L1\_P01\_C03.esd  
S980\_D1L1\_P01\_C07.esd  
S980\_D1L1\_P01\_G02.esd  
S980\_D1L1\_P01\_A04.esd  
S980\_D1L1\_P01\_F09.esd  
S980\_D1L1\_P01\_D06.esd

그림 3 Viewer에 의한 Megasort® 결과 (서열 alignment data)의 일부

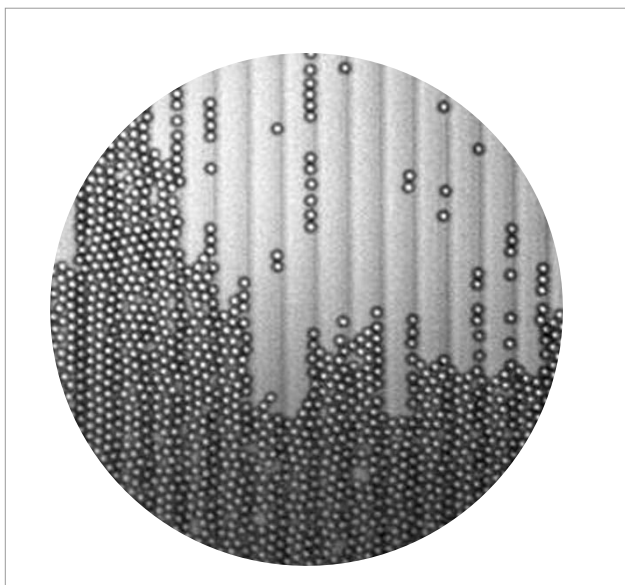


그림 4 MPSS® procel에 충전 중인 Megaclone® beads

넓고 dynamic한 영역에서 고해상도의 결과를 얻을 수 있다. MPSS®를 이용한 해석의 예로는 Human이나 *Arabidopsis thaliana* 등을 중심으로한 논문이 발표되어 있으며<sup>6)</sup>, *Arabidopsis thaliana*에 관해서는 상세한 데이터가 공개되어 있는 web site<sup>7)</sup>도 있다. MPSS®의 원리·기술에 대한 자세한 내용은 LifeScience & Biotechnology 20, 25호를 참조하기 바란다.

### MPSS®의 강점

MPSS® 이외의 gene expression profile을 하는 방법으로서, DNA microarray, SAGE (serial analysis of gene expression), 정량 PCR법 등이 있다. 본 고에서는 DNA microarray법 및 SAGE와 비교하면서 MPSS®의 뛰어난 점에 대하여 설명하고자 한다.

### DNA microarray법과의 비교

DNA microarray법은 hybridization의 원리에 기초해 상대적 발현량을 측정하는 방법이다. 이 방법은 다검체 해석에는 매우 효과적이지만, 발현량이 낮은 유전자의 경우에는 충분한 signal을 얻지 못해 정확한 정량을 못하는 경우가 있다. 출아 효모에 대하여 DNA microarray 해석을 수행한 연

구)에 있어서도 이러한 점이 지적되었다.

특히 포유류 세포처럼 mRNA 종류가 많은 시료에서는 실질적으로 정량이 불가능한 유전자가 많아질 가능성이 있다. 또한 품질이 좋은 DNA Chip을 이용해 신중하게 실험하지 않으면 재현성이나 해석결과의 정밀도가 매우 낮아질 것으로 예상된다.

SAGE와의 비교

한편 MPSS®와 SAGE는 DNA microarray법과는 원리가 완전히 다르며, 유전자 서열을 해석하여, 개개의 유전자 clone수를 계산함으로써 정량적으로 발현량을 측정하는 기술이다. 이러한 방법은 그 원리가 비교적 정확한 유전자 발현량에 관한 정보를 얻을 수 있을 것으로 생각되지만, 발현이 적은 유전자에 관한 정보를 얻기 위해서는 적어도 수십 만개 세포 중의 mRNA 분자 대부분을 읽을 필요가 있다. 그러나 SAGE에서는 sequencing 비용 문제로, 발현이 적은 유전자를 해석하는 것은 사실상 곤란하다.

그러나 MPSS®에서는 한번에 백만개 이상의 order로 해석하기 때문에 발현량이 매우 적은 유전자를 포함하여 하나의세포 속에 발현되어 있는 것으로 여겨지는 수십만개의 mRNA 분자에 대하여 총체적으로 발현을 해석할 수 있다. 이 밖에 해석수를 늘려 해상도를 높여가는 것도 가능하다.

MPSS® signature sequence의 annotation

MPSS®에 의해 얻어지는 signature는 mRNA 서열의 poly(A) 서열에 가장 가까운 GATC 서열 (DpnII 인식서열)에서 downstream에 걸친 17 ~ 20 염기 서열이다.

각 signature가 어느 유전자에서 유래하는가를 알아보기 위한 작업은 사전에 생물 종별로 작성한 annotation 데이터베이스를 이용하여 실시한다. Annotation 데이터베이스는 생물종별로 cDNA 서열 또는 genome 서열

에서 signature가 될 가능성이 있는 GATC로 시작하는 17 ~ 20염기를 받아들이고 있다. 각 signature 후보서열에 그 확실성을 나타내는 class 번호를 붙여, 유전자명 등의 유전자 정보와 합쳐 데이터베이스화하고 있다.

예를 들면, human이나 mouse의 경우에는 genome 서열, NCBI의 Unigene cluster 정보, EST 서열정보를 토대로 위의 annotation 데이터베이스를 작성한다. 이 데이터베이스로부터 MPSS® 해석 결과 얻어진 signature에 대한 정보를 검색함으로써, 각 signature가 어느 유전자에 대응하는가를 알 수 있다. 미지의 유전자인 경우는 genome에 대한 mapping 정보를 얻을 수 있다.

그 밖의 생물종의 경우에는 입수 가능한 서열정보가 genome 서열뿐이므로 mRNA 서열정보가 없을 경우나 반대로 EST 서열은 있지만 genome 정보가 없을 경우 등 다양한 경우를 생각할 수 있다. 이 경우에는 고객의 의견을 물어보며, 각각에 가장 적합한 annotation 데이터베이스를 작성하고 있다.

\*: 알 수 있는 signature 후보서열을 토대로 한 서열에 poly (A) strand 및 poly (A) 부가 signal 서열이 존재하는지의 구조적 특징과 GATC가 3' 말단에 있는지, sense strand 위에 있는지 등의 위치정보를 토대로 signature 후보서열의 신뢰성을 평가하여 분리한 지표를 class라고 부른다.

MPSS® 해석 표준 데이터

2종류의 human 유래 시료를 MPSS®로 해석한 결과의 데이터 일례를 그림 5에 나타내었다. 각 signature ①에 대해 그 서열의 class 번호 ②, 동정된 유전자 정보로서 accession ③ 및 description ④이 표시된다. 또한 각 시료에서의 발현량 ⑤은 100만개의 전사산물당 개수 (tpm; transcript per million)로 표시된다. 또한 시료간 발현차가 비순한지를 판정하기 위한 통계적 지표로서 p-value ⑥도 표시된다\*. 이러한 해석은

①	②	③	④	⑤	⑥			
Signature	Class	Accession	Description	A Counts(tpm)	B Counts(tpm)	P-Val	significant	replication
GATCGCCACTACCGCCGCC	24	Hs.406013	B0876558 in Hs.406013 rep by NM_000224: Homo sapiens keratin 18 (KRT18)	1193	1024	0.013392158	s	yes
GATCGTGACCACTGGAACA	1	Hs.151988	A4837292 in Hs.151988 rep by NM_005923: Homo sapiens mitogen-activated protein kinase 10	0	4	0.07960812	s	yes
GATCGCCACTACCGCCGCC	25	Hs.355214	BM016596 in Hs.355214 rep by NM_000526: Homo sapiens keratin 14 (epidermal)	0	164	0	s	yes
GATCCTTACCTGAAGAATAT	1	Hs.418162	AW411328 in Hs.418162 rep by NM_009905: Homo sapiens amyloid beta precursor protein 2	0	12	0.014305566	s	yes
GATCGTGACCTGAGTAAAT	24	Hs.356721	BE765522 in Hs.356721 rep by NM_031157: Homo sapiens heterogeneous nuclear ribonucleoprotein A1	156	211	0.04664971	s	yes
GATCGTGTGGGATAGCAC	3	Hs.181418	BC000371 in Hs.181418 rep by NM_014730: Homo sapiens KIAA0152 gene product	226	203	0.000125	s	yes
GATCGGCAATGTGGTTCCT	22	Hs.433701	AA742796 in Hs.433701 rep by NM_000998: Homo sapiens ribosomal protein L18	0	30	0.000183	s	yes
GATCGTGTGGCTTCCAAAGC	1	Hs.203910	AL050156 in Hs.203910 rep by NM_003021: Homo sapiens small glutamine-rich nuclear phosphoprotein 1	181	169	0.001198975	s	yes
GATCGTGCAAGATATGCT	3	Hs.76240	BC001116 in Hs.76240 rep by NM_000476: Homo sapiens adenylate kinase 1	25	17	0.067127172	s	yes
GATCGTTTCGCCATCATGA	1	Hs.306122	AA769361 in Hs.306122 rep by NM_138501: Homo sapiens glycoprotein, sialylated	178	140	0.0000298	s	yes
GATCGGGCGGATTGAGGGTG	1	Hs.381126	CD174974 in Hs.381126 rep by NM_005617: Homo sapiens ribosomal protein L18	3434	3779	0.005848773	s	yes
GATCCTGTTGGTGAATGAAC	1	Hs.75431	A1064748 in Hs.75431 rep by NM_021870: Homo sapiens fibrinogen, gamma chain	584	520	3.65E-10	s	yes
GATCGGACATGAAAGGACCC	1	Hs.311765	AA504482 in Hs.311765 rep by NM_003119: Homo sapiens spastic paraplegia 1	0	7	0.083403097	s	yes
GATCGCTGATGATTCGCC	1	Hs.106876	A1885283 in Hs.106876 rep by NM_000469: Homo sapiens ATPase, H+ transporting, lysosomal 1	22	15	0.066785513	s	yes
GATCGCCTTCCAACCCCACT	3	Hs.159	BC010140 in Hs.159 rep by NM_001065: Homo sapiens tumor necrosis factor alpha	133	124	0.004986412	s	yes
GATCGGAGAGATTGAAGGC	24	Hs.93194	T68297 in Hs.93194 rep by NM_000039: Homo sapiens apolipoprotein A-I	0	6	0.083263929	s	yes
GATCACTACACAAATTGTCT	1	Hs.309517	AA056657 in Hs.309517 rep by NM_002276: Homo sapiens keratin 19 (KRT19)	0	137	2E-15	s	yes
GATCGTACCAACCCGACCTCC	1000	chr17	chr17:47639491-47639510	0	9	0.045499653	s	yes
GATCGTGACTGGGTACTCTG	22	Hs.8372	F28332 in Hs.8372 rep by NM_006830: Homo sapiens ubiquinol-cytochrome b reductase complex subunit 1	0	18	0.000455	s	yes
GATCGCACTGACTGACTCCD	2	Hs.434502	NM_152586 in Hs.434502 rep by NM_152586: Homo sapiens hypothetical protein	283	363	1.24E-08	s	yes
GATCCTGCTCATCTCCAC	15	Hs.270845	BC017705 in Hs.270845 rep by NM_138555: Homo sapiens kinesin family member 10	0	6	0.031798383	s	yes
GATCAAAACAACAGACAAA	12	Hs.282883	CB243986 in Hs.282883 rep by NM_032376: Homo sapiens hypothetical protein	0	4	0.07960812	s	yes
GATCAAAAGTGAACACCTG	1	Hs.434953	AL531642 in Hs.434953 rep by NM_002129: Homo sapiens high-mobility group protein 1	0	44	2.97E-08	s	yes
GATCCATGAGTCTAGCCD	1	Hs.411509	AA230277 in Hs.411509 rep by NM_000852: Homo sapiens glutathione S-transferase gamma	3619	3297	0.008196453	s	yes
GATCAAGTACCCACCCAA	1	Hs.436985	AF128625 in Hs.436985 rep by NM_006035: Homo sapiens CDC42 binding protein 1	0	47	0.0000273	s	yes
GATCGCAGAGAAAAAAA	23	Hs.5978	BX282044 in Hs.5978 rep by NM_005358: Homo sapiens LIM domain only 7	0	6	0.083263929	s	yes
GATCAAGAGATTCCAAAGA	24	Hs.409535	CD390030 in Hs.409535 rep by NM_007104: Homo sapiens ribosomal protein L18	0	42	0.0000131	s	yes
GATCATGCTGCCCTGGACCC	22	Hs.387576	BF562494 in Hs.387576 rep by NM_001005: Homo sapiens ribosomal protein L18	13	39	0.01229339	s	yes
GATCACTCTGAGACCCACB	1	Hs.381172	BU536710 in Hs.381172 rep by NM_021104: Homo sapiens ribosomal protein L18	43	29	0.01540021	s	yes
GATCAATGCCACTGACCTCT	1	Hs.446546	AA446119 in Hs.446546 rep by NM_133436: Homo sapiens asparagine synthetase 1	0	14	0.014305566	s	yes
GATCCGAGCCAAAGAGGAC	3	Hs.122908	AF321125 in Hs.122908 rep by NM_030928: Homo sapiens DNA replication origin 1	85	89	0.03927968	s	yes
GATCTCCAGGACTGAACCTG	24	Hs.406013	AA873061 in Hs.406013 rep by NM_000224: Homo sapiens keratin 18 (KRT18)	0	10	0.025346831	s	yes
GATCAACATTCGCAATTGCC	22	Hs.432330	A1127601 in Hs.432330 rep by NM_006570: Homo sapiens Ras-related GTP-binding protein 1	0	9	0.045499653	s	yes

그림 5 MPSS® 데이터 예

Z-test로서 SAGE에서도 이루어지고 있으며<sup>10)</sup>, 통계적으로 신뢰할만한 지표를 얻는 것은 특히 발현이 적은 유전자 해석에서 중요하다. 예를 들면  $p < 0.001$ 이면, 발현량이 30~40 tpm인 유전자에서 2배의 발현차이 정도는 유의한 것으로 간주하고, 발현량이 200 tpm 정도의 유전자에서 40%의 변동은 확실한 것으로 간주할 수 있다.

### Megaclone®, MPSS® 결과를 이용한 해석

당사에서는 Megasort®, MPSS®에 의한 해석 결과를 상기의 표준 데이터로 할 뿐 아니라, 표준 데이터를 토대로 고객의 요청에 따라 다양한 option 해석을 제공하고 있다. 문의가 많은 option 해석은 다음과 같다.

(1) 먼저 Megasort®, MPSS®의 해석 데이터의 표준적인 annotation 정보 (유전자 정보)를 토대로, annotation을 더 확장할 수 있다.

예를 들면, 세포 내에서의 유전자 및 유전자 산물의 기능 및 구조에 관한 용어를 계층적으로 통합한 Gene Ontology<sup>11)</sup> 데이터베이스 정보를 첨가함으로써 기능 일람, 주목할 만한 기능을 지닌 유전자 추출, 유사한 기능을 지닌 유전자 분류 등이 가능하다.

유전자가 code하는 단백질 domain, motive 정보 등을 첨부하거나 서열 정보로부터 세포의 국부존제 성향을 예측하는 것도 가능하다. 확장된 이들 annotation에 의해서 기능과 발현 변동 관련에 대한 고찰이 가능하다.

(2) 이 밖에 MPSS®로 다수의 시료를 해석한 결과로 얻은 데이터로부터, DNA microarray와 같은 gene expression profile 해석기술에 의한 다양한 해석이 가능하다. 예를 들면 다음과 같은 해석방법이 자주 이용된다.

- 개개의 유전자에 착안해 통계적으로 유의한 발현차가 있는 유전자를 선택하는 방법

- 유전자 발현형태에 따라 유사성으로 유전자를 그룹화하여, 유사성에 의해 계층적으로 분석한다「계층적 cluster 분석」(그림 6)
- 가장 먼저 cluster수를 정해두고, 유전자 발현 형태에 따라 분류한다「k-평균 cluster 분석」(그림 7)

이러한 해석은 데이터를 분석하는 도구가 될 뿐 아니라, 전술한 annotation 데이터와 함께 기능 및 유전자의 중요성에 대해 고찰하기 위한 재료가 된다.

### Functional DNA chip

당사에서는 다수의 검체 유전자 발현을 해석하는데 유용한 DNA microarray 제작 서비스도 실시하고 있다. Megasort®, MPSS® 해석결과를 이용하면, 발현이 변동할 가능성이 높은 유전자 (신규 유전자도 포함)만으로 구성된 DNA microarray (Functional DNA chip)를 제작할 수 있다 (Life Science & Biotechnology 23호 참조).

Megasort® 결과를 이용하여 DNA microarray를 제작할 경우에는 각 유전자 cluster 중에서 적합한 clone을 선택해 Functional DNA chip을 설계한다. 또한 Megasort®, MPSS®에서 얻은 유전자 발현정보로부터 IntelliGene® CHIP으로 실제로 얻을 수 있는 300 mer의 cDNA probe 또는 60 mer의 oligo DNA probe를 설계하는 것도 가능하다. Tm값 등의 조건 및 다른 유전자와의 상동성에 대해 검토하고, 독자의 선택기준에 따라 비특이적 hybridization이 최소가 되는 probe 서열을 설계한다. 이 설계정보를 토대로 고품질의 DNA chip을 제작한다. 이 밖에 DNA microarray의 hybridization 실험 및 clustering 해석, Real Time PCR의 primer 설계, 해석 실험도 실시하고 있다.

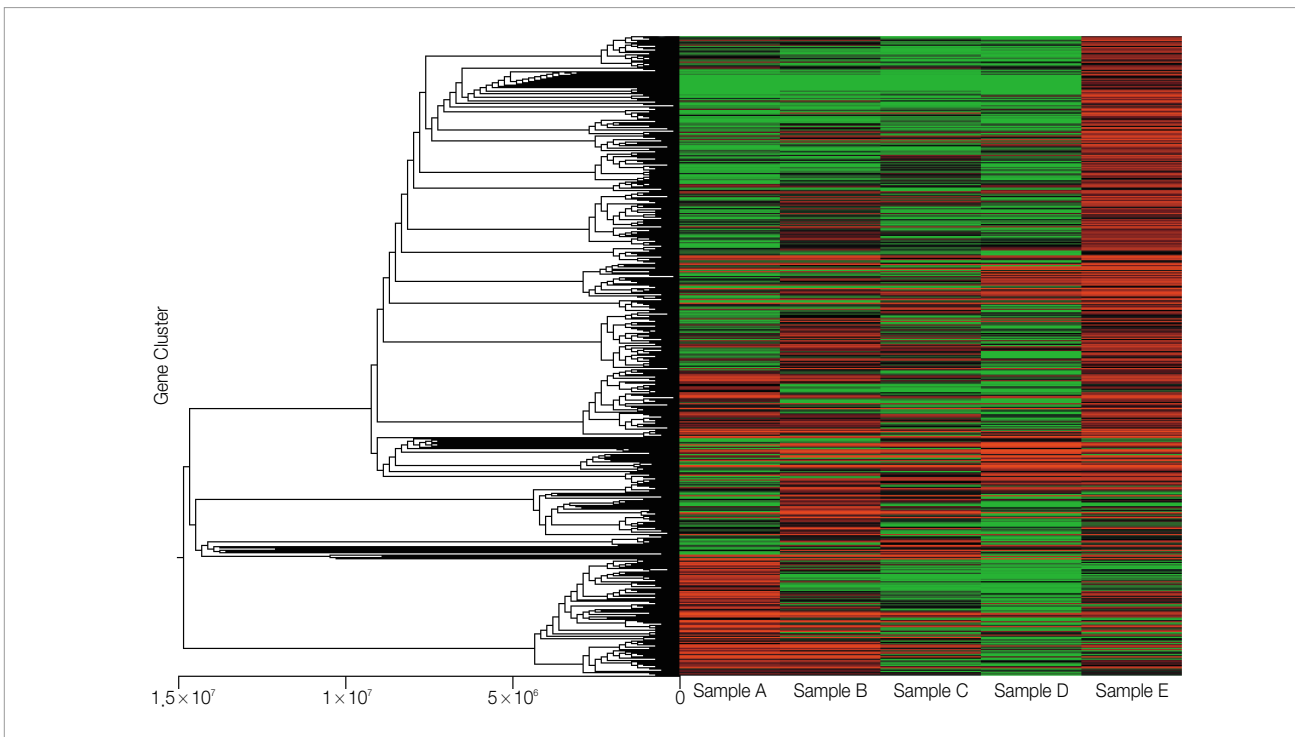


그림 6 MPSS® 데이터를 이용한 계층적 cluster 분석 예

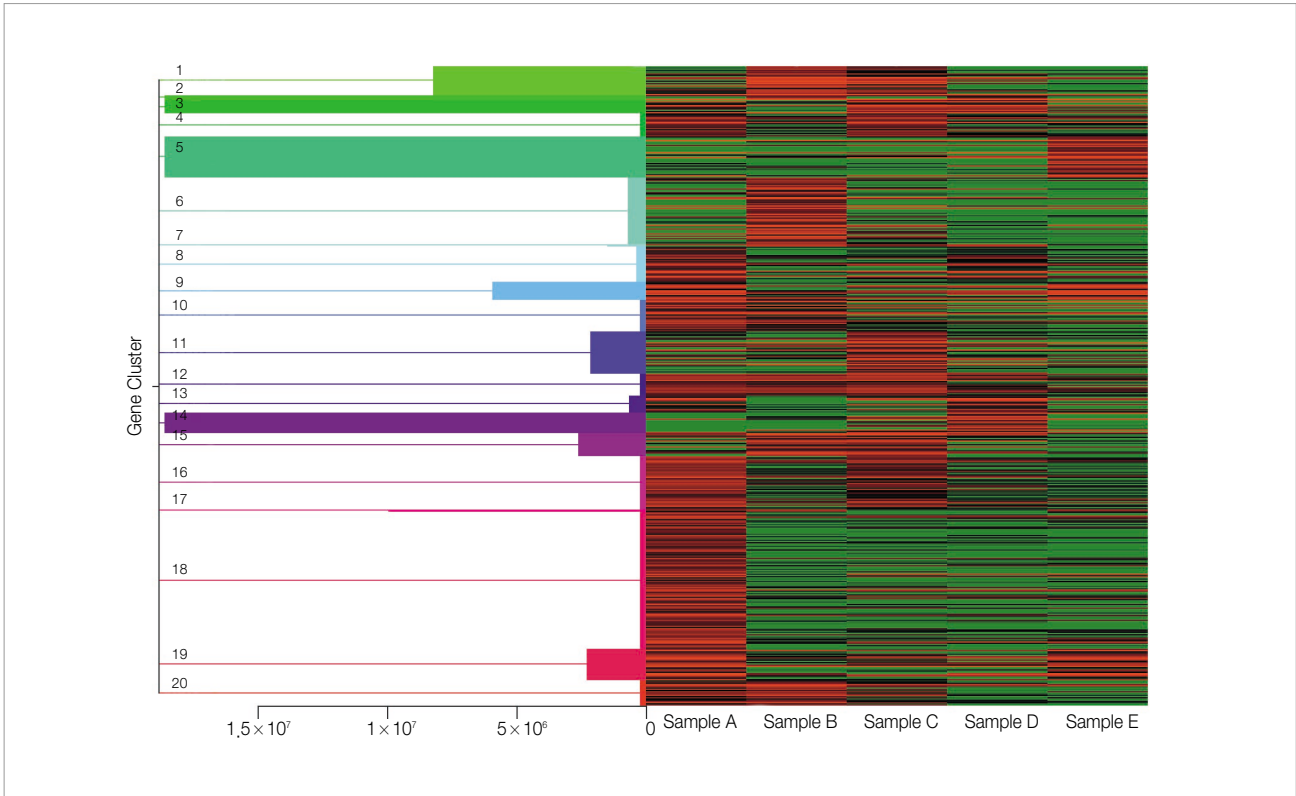


그림 7 MPSS<sup>®</sup> 데이터를 이용한 k-Means clustering 분석 예

### 맺음말

Megasort<sup>®</sup> 및 MPSS<sup>®</sup>는 DNA microarray와는 다른 원리에 의해 새로운 견해를 얻을 가능성이 높은 매우 유용한 유전자 발현 해석방법이다. 당사에서는 Megasort<sup>®</sup>, MPSS<sup>®</sup>를 좀더 적은 노력으로 유용하게 활용하기 위해 Megasort<sup>®</sup>, MPSS<sup>®</sup> 뿐 아니라 RNA 조제에서 DNA microarray 제작까지 폭넓은 menu를 구비하고 있다. 또한 Megasort<sup>®</sup>, MPSS<sup>®</sup>에서 얻은 대량의 데이터 처리 및 각종 해석에 대해서도 고객의 요청에 따라 당사 bioinformation을 지원하고 있으니 자세한 문의는 연구지원사업부 (02-575-7409)로 하시기 바랍니다.

### 참고문헌

- 1) Brenner, S., et al. (2000) *Proc. Natl. Acad. Sci. USA*, **97**, 1665-1670
- 2) Brenner, S., et al. (2000) *Nat. Biotechnol.* **18**, 630-634
- 3) Jongeneel, C. V., et al. (2003) *Proc. Natl. Acad. Sci. USA*, **100**, 4702-4705
- 4) Meyers, B. C., et al. (2002) *Plant. J.* **32**, 77-92.
- 5) Hoth, S., et al. (2002) *J. Cell. Sci.* **115**, 4891-4900.
- 6) Lid, S. E., et al. (2002) *Proc. Natl. Acad. Sci. USA*, **16**, 5460-5468.
- 7) Directory of MPSS data pages: <http://mpss.udel.edu/>
- 8) Holland, M. J., et al. (2002) *J. Biol. Chem.* **227**, 14363-14366

- 9) Reinartz, J., et al. (2002) *Briefings in Functional Genomics and Proteomics*, **1**, 95-104
- 10) Kal, A. J., et al. (1999) *Mol. Biol. Cell*, **10**, 1859-1872.
- 11) An introduction to Gene Ontology: <http://www.geneontology.org/GO.doc.html>