전사체학(Transcriptomics)에 있어서의 황금표준정량법(Gold Standard Method)

신 기 원(포항공과대학교 시스템 생명공학부)
정 규 열(포항공과대학교 화학공학과)

Introduction

전사체학(transcriptomics)이란 mRNA 수준에서 생명 현상을 전체적으로 다루는 학문이라고 정의 할 수 있다. 유전자를 하나의 단백질에 대한 정보를 포함(coding)하고 있는 DNA 서열이라고 한다면 유전체는 모든 유전자의 집합체이다. 마찬가지로 전사체는 모든 mRNA의 집합을 의미한다. 전사체 수준에서의 유전자 발현(gene expression)에 대한 연구는 DNA 마이크로어레이(microarray)와 같은 고 처리량의 분석 기술 발달과 더불어 활발하게 진행되어 왔다. 유전자의 발현을 전사 수준에서 다루는 것은 단백질 활성의 직접 정량보다 분석이 쉬우며 간접적 지표로서의 가치가 있다. 하지만 최근에 시스템스 생물학과 함께 여러 오믹스(omics)에 대한 관심과 연구가 활발해 지면서 전사 수준의 연구 자체가 가지는 의미도 주목을 받고 있다. 즉, 유전체학, 전사체학, 단백체학(proteomics)을 각각으로서의 전체가 아닌 함께 유기적인 연결선상에서의 전체로서 바라보는 것이다. 그런 의미에서 분석 기술의 성숙과 함께 전사체학을 연구하는 것은 그 자체로서 가지는 의미가 점점 커지고 있다고 할 수 있다.[1]

마이크로어레이, massive parallel signature sequence (MPSS), serial analysis of gene expression(SAGE) 등의 고 처리량 분석 기술의 등장은 전사 수준에서의 오믹스가 실현가능하다는 것을 보여 주었다. 이제는 하나의 mRNA가 아닌 전체 mRNA의 변화를 관찰하는 것이 가능하게 되었고 어마어마한 양의 데이터들이 쏟아지고 있다. 하지만 여전히 문제는 남아 있다. 일반적으로 기술적 관점에서 보면 고 처리량(high-throughput)의 분석 기술은 정량적인 측면에서 정확도가 떨어지는 경향이 있으며, 반대로 정확한 정량이 가능한 분석 기술은 그 처리량 면에서 아쉬운 점이 있다. 분석 기술의 처리량과 정확도는 반비례 관계인 것이다. 현재 가장 널리 사용되고 있는 마이크로어레이의 경우에도 그 데이터의 질적 측면에 대한 의문이 있으며, 실제로 여러 마이크로어레이 플랫폼(platform) 간에 서로 상반되는 결과들이 나오기도 한다. 최근 마이크로어레이 실험의 프로토콜을 표준화 하려는 노력들이 진행되고 있다. 대표적인 것이 Minimum Information About Microarray Experiment(MIAME, http://www.mged.org/Workgroups/MIAME/miame.html) standards, External RNA Controls Consortium(ERCC, http://www.cstl.nist.gov/biotech/Cell&TissueMeasurements/GeneExpression/ERCC.htm) 그리고Micro-Array Quality Control(MAQC, http:// www.fda.gov/nctr/science/centers/toxicoinformatics/maqc/) consortium 등이다.

Gene Expression Omnibus(GEO)나 ArrayExpress 등의 데이터베이스에는 현재 어마어마한 양의 마이크로어레이 데이터가 축적되어 있다. 이 데이터가 유용하게 사용되려면 서로 다른 플랫폼 간에 나온 데이터를 비교 종합할 수 있어야 한다. 최근 몇 년간 플랫폼 간 비교연구 결과들이 나오고는 있으나 서로 상반되는 결과들이 나오고 있다. 따라서 한쪽에 치우치지 않은 공정한 결과를 낼 수 있는 비교연구가 절실한 상황이다.[1]

마이크로어레이 플랫폼 간 비교연구에서 가장 중요한 사항 중의 하나로 데이터 질의 기준이 되는 참고

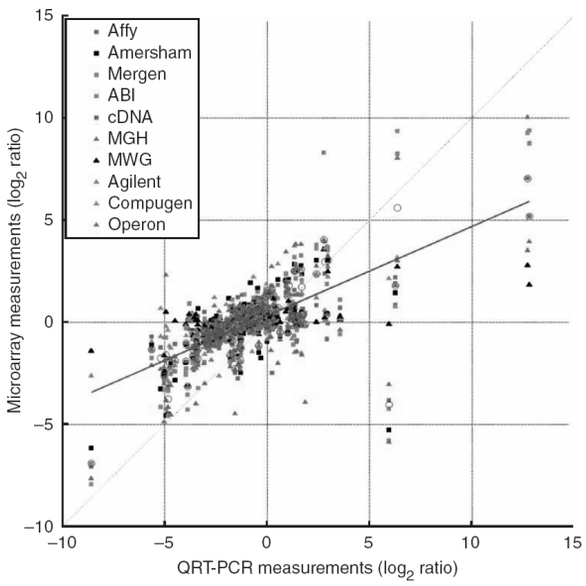

그림 1 TaqMan assay와 마이크로어레이의 비교 분석[2]

값을 결정할 대조군 실험에 사용되는 대체 분석 기술의 결정이다. MAQC 컨소시움은 TaqMan assay, Standardized RT-PCR(staRT-PCR), 그리고 QuantiGene을 대체 플랫폼으로 선정하였다. 이들 플랫폼들은 민감도와 특이성(정확성) 그리고 선형 동적 영역(linear dynamic range) 측면에서 우수하다고 알려져 있다.[2] 한편, 이러한 대조군 실험용 분석 기술들을 "황금표준정량법(gold standard method)"이라고 명명하기도 한다. 그림 1은 TaqMan assay의 결과를 대조군으로 10가지의 서로 다른 플랫폼의 데이터를 산점도(scatter plot)로 나타낸 결과로, TaqMan assay가 황금표준정량법으로서 마이크로어레이 플랫폼 간 비교에 활용된 예를 보여준다.

이 글을 통해 MAQC consortium에서 사용하는 세 가지 황금표준정량법들의 원리와 장단점을 비교하고 이들에 대한 이해를 돕고자 한다.

TaqMan assay

TaqMan assay는 실시간 유전자 증폭 기술(real time PCR system)을 통하여 증폭되기 전 주형(template)의

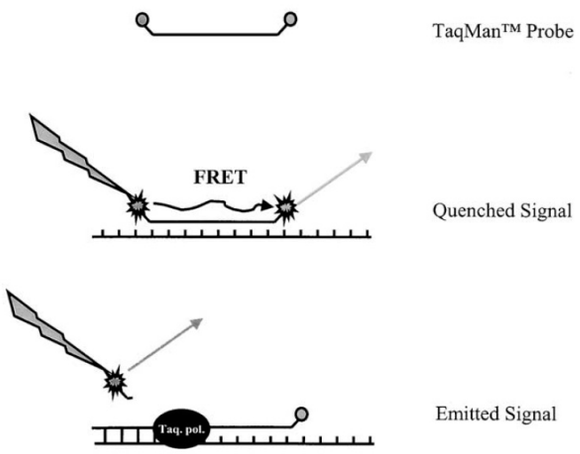

그림 2 TaqMan probe의 작동 원리[4]

양을 정량하는 방법이다. 이 방법의 핵심원리는 *Taq* 중합효소의 5' 뉴클리아제(nuclease) 활성과 FRET(fluorescence resonance energy transfer) 현상이다. *Taq* 중합효소는 5' 뉴클리아제 활성이 있기 때문에 중합과정 중에 비특이적으로 결합한 올리고머(oligomer)의 방해를 받지 않는다. 이점을 이용하여 TaqMan probe(Applied Biosystems)를 디자인함으로써 각 사이클마다 dsDNA의 양을 정량할 수 있다. TaqMan probe는 PCR 프라이머 세트와는 별도로 제작되는 올리고머로서 프라이머가 주형에 결합하는 단계(annealing step)에서 프라이머와 함께 주형에 결합하게 된다(그림 2). 주형에 결합한 상태의 TaqMan probe는 양 끝에 달린 두 개의 형광색소, 즉 reporter dye와 quencher dye의 FRET에 의해 reporter dye가 방출한 파장의 빛을 quencher dye가 흡수하여 긴 파장의 빛을 방출하게 된다. 중합 반응이 진행되면서 중합 효소의 5' 뉴클리아제 활성에 의해 TaqMan probe는 가수 분해되며 이에 따라 reporter dye와 quencher dye가 분리되고, reporter dye의 직접적인 형광색을 측정할 수 있게 된다. 이때 방출되는 빛의 파장은 quencher dye가 방출했던 파장보다 작다. 결과적으로 reporter dye 한 개로부터의 방출된 빛의 세기는 dsDNA 한 개와 상응하는 것이기 때문에 이를 dsDNA의 정량에 이용할 수 있는 것이다.

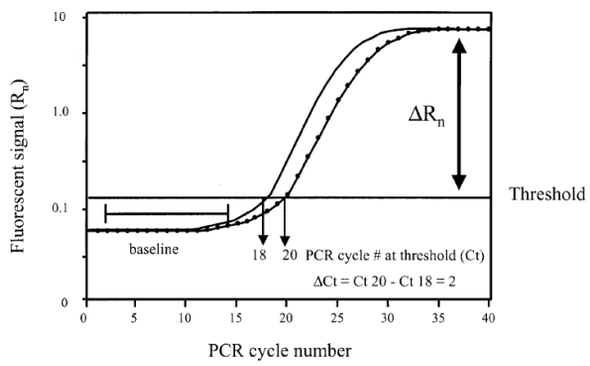

그림 3 PCR 사이클 수에 따른 신호 증가 그래프[4]

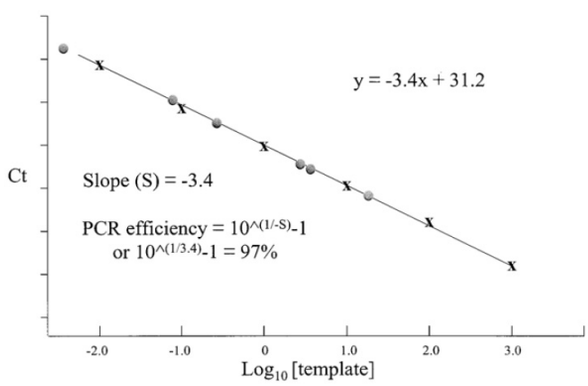

그림 4 threshold 사이클과 주형(template)량의 산점도[4]

실시간(real time)으로 dsDNA의 양을 관찰하면 PCR 사이클에 따라 초기에는 형광의 세기가 잡음(noise)에 묻혀있는 지연 상태(lag phase)가 관찰된다. 일반적으로 10에서 15 사이클이 지나면 형광 신호의 크기가 잡음 신호의 크기를 넘으면서 증가 상태(exponential phase)가 관찰되며, 35 내지는 40 사이클이 지나면 형광 신호 값이 포화 상태(saturated phase)가 된다. 지연 상태와 포화 상태 이후에는 증폭 전 주형의 양에 대한 정보를 얻기가 힘들다. 따라서 증가 상태에서 threshold 선을 긋고 각각의 형광 신호의 그래프와 만나는 지점의 사이클 수를 정량에 이용한다(그림 3). 이 사이클 수를 cycle threshold, Ct라고 정의한다. Ct와 주형의 양에 log를 취한 값을 그래프로 그리면 그림 4와 같이 두 값 사이의 선형적 관계를 관찰 할 수 있다.

Standardized reverse transcription PCR (staRT-PCR)

Standardized reverse transcription PCR은 PCR을 이미 알려진 양의 주형 세트와 함께 진행함으로써 표적 주형(target template)의 양을 정량하는 방법이다. 이 방법은 여러 개의 중합 반응이 한 시험관에서 이루어지기 때문에 최적화된 조건을 잡기가 어려운 점이 있기는 하지만 특별한 추가 장비가 필요 없다는 점에서 장점이 있다. Standard competitive template(CT)는 이미 알려진 양의 주형 세트로서 staRT-PCR의 핵심적인 부분이다. 최적화된 조건의 반응 용액으로 적당한 사이클의 PCR을 수행하고 젤 이미지 분석을 통해 젤에서의 Ct 밴드들의 형광 세기와 Ct의 양을 그래프로 그려 표준 곡선(standard curve)을 얻을 수 있다(그림 5). 따라서 그 양이 알려지지 않은 표적 주형 밴드의 형광 세기를 통해 증폭되기 전의 양을 정량할 수 있다.

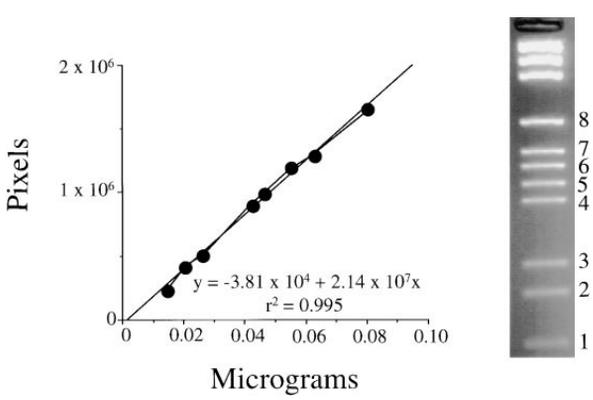

그림 5 Standard Competitive Template(CT)로부터 얻어진 표준곡선[5]

QuantiGene

다중(multiplex) bDNA assay라고도 하는 이 방법은 xMAP 형광 비드(Luminex 社)를 이용하여 기존의 bDNA assay를 다중 assay로 발전시킨 것이다. 우선

bDNA(branched DNA)란 나뭇가지 모양의 probe를 말하며, bDNA assay란 bDNA의 각 가지에 바이오틴을 부착하는 등의 방법으로 하나의 핵산에 다중 라벨을 붙여 정량하는 방법을 말한다. 다시 말하면 하나의 주형에 bDNA를 결합시키고(hybridization) 각 가지마다 다시 바이오틴이 부착된 probe를 결합시키게 되면 한개의 주형 당 증폭된 신호를 얻을 수 있다는 것이다.(그림 6)

xMAP 형광 비드는 앞서 말한 bDNA assay가 다중 표적에 대해 한 번에 수행될 수 없다는 단점을 보완해 주는 물질이다. 각각의 비드가 고유파장의 빛을 방출하며, 따라서 각 비드에 특정 서열의 주형을 결합시켜 각각이 서로 다른 광학적 아이디를 갖게 해준다. (그림 6) 이 과정에서 특히 중요한 것이 특이성의

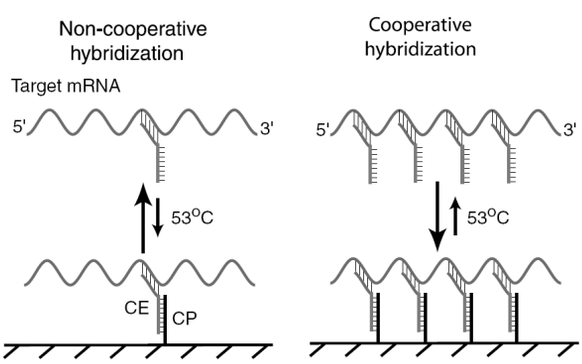

그림 6 협동 결합(cooperative hybridization)으로 인한 반응 평형의 이동[6]

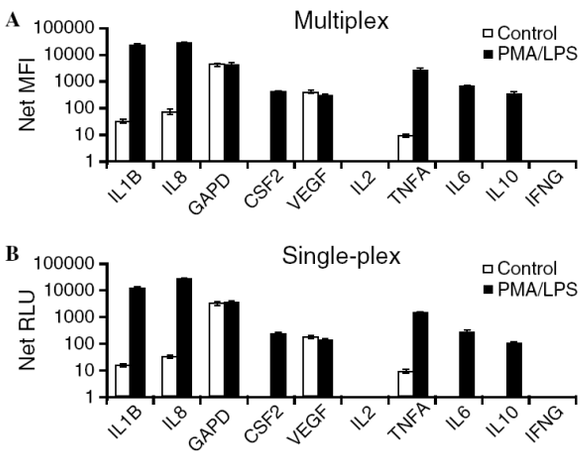

그림 8 (A) 다중 bDNA assay 결과 (B) 개별 bDNA assay 결과[6]

문제인데, 비특이적 결합에 의해서 비드와 주형 간에 엇갈린 결합(cross-hybridization)이 일어나게 되면 주형의 광학적 아이디 시스템의 치명적인 결함이 생길 수 있다는 것이다.

Flagella *et al.*[6]은 이러한 문제점을 협동 결합(cooperative hybridization)으로 해결했다고 주장한다. 협동 결합을 통해 비드 표면과 주형 간의 결합력이 증가하고 더 높은 녹는점(melting temperature)을 갖게 되었다. (그림 7) 따라서 특이적 결합과 비특이적 결합의 에너지 차가 커짐에 따라 보다 특이적인 가능하며, 다중 표적에 고유의 아이디를 부여할 수 있게 된다는 것이다. 이러한 방법으로 10개의 서로 다른 mRNA를 개별 assay와 다중 assay로 분석한 결과를

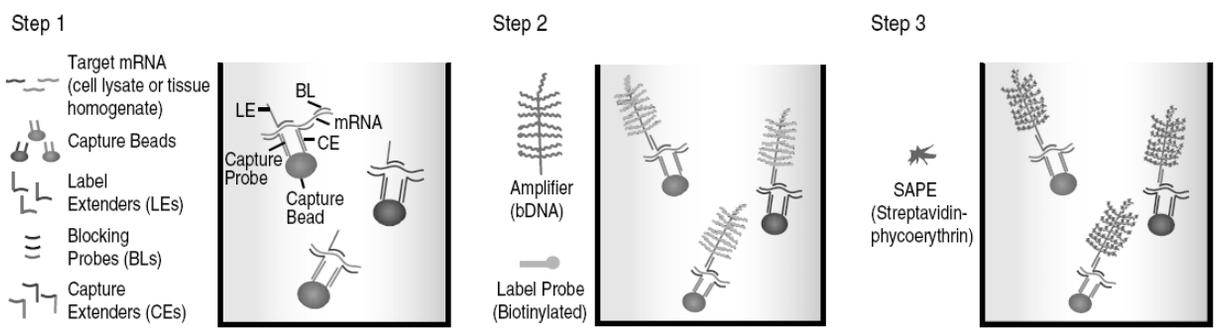

그림 7 다중 bDNA assay의 과정. bDNA는 신호를 증폭하고, 비드는 아이디의 역할을 한다[6]

비교한 것이 그림 8이다. 두 결과가 대조군과 실험군에서의 발현양의 정량적인 차이를 동일하게 보여주고 있음을 알 수 있다.

Comparison of the three methods

앞서 설명한 세 가지의 황금표준정량법은 저마다 각각의 장점과 단점을 지니고 있다. 먼저 TaqMan assay와 staRT-PCR의 경우는 PCR 기반의 정량 방법이기 때문에 조건을 잡는 절차가 까다롭다. 최적화된 반응 조건을 잡는 것이 정확한 assay를 위한 큰 관문이다. 또한 프라이머 세트의 디자인에 있어서도 결합의 특이성이나 주형의 길이 차이에 따른 중합 효율도 차이가 날 수 있다. 하지만 PCR의 특성상 민감도 면에서 뛰어난 점을 보이는데 TaqMan assay의 경우 5개의 분자까지 측정이 가능하다는 보고가 있다.[7] StaRT-PCR의 경우 앞서 언급한 바와 같이 추가 장비의 필요가 없고 누구나 쉽게 구할 수 있는 시약과 장비로 실험이 가능하다는 장점이 있다.

QuantiGene assay의 경우 mRNA에 직접 다중 결합을 통해 라벨을 부착하여 측정이 이루어지기 때문에 역전사(reverse transcription) 과정을 거치지 않는다는 강점이 있다. 하지만 하나의 mRNA 표적을 정량하기 위하여 capture bead, label extender, blocking probe, capture extender 등을 표적마다 특이적으로 합성해야 하기 때문에 probe 간의 비특이적 결합의 문제가 있을 수 있으며, assay당 드는 비용도 만만치 않을 것으로 생각된다. 가장 널리 쓰이는 TaqMan assay의 경우도 TaqMan probe를 특이적인 결합을 하도록 합성하는데 드는 비용이 만만치 않기 때문에 비용의 문제가 있다.

시간적 효율 면에서 각 플랫폼을 비교해 보자면 TaqMan assay와 staRT-PCR의 경우 PCR 조건이 최적화 되었다는 가정 하에 assay당 드는 시간은 3시간 정도이다. QuantiGene assay의 경우 probe를 표적에 결합 시키는 반응에만 드는 시간이 16시간으로 assay당 걸리는 시간이 가장 길다. 하지만 비드의 종류가 100개로 최대 한번에 100여개의 표적을 한 번에 실험한다고 했을 때 96개의 홈에서 실험하면 그 처리량에서 오히려 앞선다고도 할 수 있다.

Carnales *et al.*[3]은 이 세 가지 플랫폼을 비교하는 실험을 하여 정량적인 비교 결과를 보고했다(표 1). 이 결과에 따르면 TaqMan assay의 경우 가장 넓은 선형 동적영역(linear dynamic range)을 갖으며, 재현성(precision)과 정확도(accuracy)면에서는 중간 정도의 성적을 올렸다. QuantiGene assay의 경우 재현성과 정확도 면에서는 단연 돋보이는 데이터를 얻을 수 있었지만, 선형동적영역의 경우 PCR 기반의 다른 두 플랫폼에 비해 떨어지는 면을 알 수 있다. QuantiGene assay가 다른 플랫폼과 비교하여 재현성과 정확도 면에서 뛰어난 이유는 여러 가지로 생각해볼 수 있으나. 역전사

표 1 MAQC의 세가지 플랫폼의 정량적 비교[3]

Platform	Gene list	Sample processing			Detection sensitivity[a]		Dynamic range[b] (log10)	Precision[c] (median)		Accuracy[d] (median)		
Symbol	Number of genes tested	Sample input	Assay replicates	Data presentation	Both A & B above LOD	Both A & B below LOD		All data	>6,000	Linearity[e] (R2)	RA (%median)[f]	RA (%variance)[g]
TAQ	997	cDNA from 10 ng total RNA, one RT reaction	Four replicates of cDNA	Normalized against POLR2A	857 (86%)	38 (3.8%)	8.1	3.46	2.42	0.950	3.6	9.4
GEX	205	cDNA from 10 ng total RNA, one RT reaction	Three replicates of cDNA	Normalized against beta-actin	193 (94%)	4 (2.0%)	6.8	6.26	3.82	0.96[h]	0.4[h]	21.1[h]
QGN	244	500 ng total RNA	Three replicate of RNA directly	Original data	223 (91%)	5 (2.0%)	4.1	2.16	2.12	0.994	1.0	5.0

과정이 없이 probe와 결합(hybridization)시킨 후 바로 신호를 측정한다는 점과, 협동 결합(cooperative hybridization)을 통해 특이적인 결합이 가능해 효과적으로 다중 표적을 구분(identification)할 수 있다는 점을 꼽을 수 있겠다.

Concluding remark

시스템스 생물학은 생명현상에 대한 총체적이고 정량적인 이해를 지향하는 분야이다. 하지만 현재까지 총체적인 이해와 정량적인 이해가 동시에 이루어지고 있는 것 같지는 않다. 총체적인 이해의 측면을 제시해 주는 마이크로어레이와 같은 분석 기술이 있는 반면, 정량적인 이해의 측면을 제시해 주는 TaqMan assay 등과 같은 기술이 있다. 이 두 가지 측면은 각각의 진영에서의 진화를 통해 빠른 시간 안에 서로가 서로를 견제하는 사이가 될 것이다. 하지만 아직까지는 상호 보완적 관계를 유지할 수밖에 없다. 스스로에게 부족한 측면이 상대에게는 충분하기 때문이다. 따라서 현재 시점에서는 보다 양질의 데이터를 얻기 위해 다음과 같은 상호 보완적 분석 과정을 제시할 수 있을 것이다. (그림 9)

마이크로어레이의 강점인 높은 처리량을 이용하여 특이적으로 발현된 유전자(differentially expressed gene, DEG)를 일차적으로 걸러내고, 정량적인 분석 기술로 검증하는 과정으로 분석 과정을 요약할 수 있다. 이러한 검증 단계의 추가가 주는 장점은 일차적으로 DEG를 걸러내는 과정에서의 위양성률(false positive rate)을 줄여줌으로써 보다 더 정확한 데이터의 확보가 가능하다는 점이다. 이렇게 확보된 양질의 데이터를 잘 해석하여 생명 현상을 이해한다면 각각의 분석 기술의 한계로 인해 가려져 있던 진리들을 발견해 내는 데 큰 도움이 될 것이다.

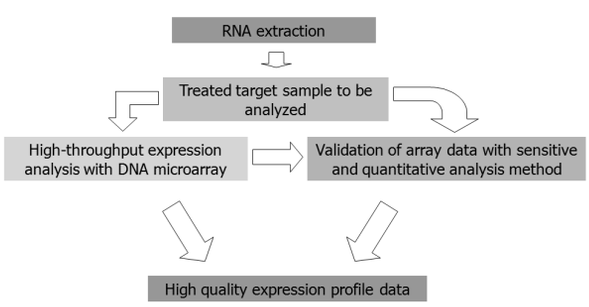

그림 9 양질의 유전자 발현 프로파일을 얻는 과정

References

[1] Ishii, N. *et al.* Multiple High-Throughput Analyses Monitor the Response of *E. coli* to Perturbations *Science* **316**, 593-597(2007)

[2] Kuo, W. P. *et al.* A sequence-oriented comparison of gene expression measurements across different hybridization-based technologies *Nat. Biotechnol.* **24**, 832-840(2006)

[3] Carnales, R. D. *et al*. Evaluation of DNA micro-arrayresults with quantitative gene expression platforms *Nat. Biotechnol.* **24**, 1115-1122(2006)

[4] Ginzinger, D. G. Gene quantification using real-time quantitative PCR: An emerging technology hits the mainstream *Experimental Hematology* **30**, 503-512(2002)

[5] Willey, G. C. *et al.* Expression Measurement of Many Genes Simultaneously by Quantitative RT-PCR Using Standardized Mixtures of Competitive Templates *Am. J. Respir. Cell Mol. Biol.* **19**, 6-17(1998)

[6] Flagella M. *et al.* A multiplex branched DNA assay for parallel quantitative gene expression profiling *Analytical Biochemistry* **352**, 50-60(2006)

[7] Jhao J. R. *et al.* Detection of hepatitis B virus DNA by real-time PCR using TaqMan-MGB probe technology *World J. Gastroenterol* **11**, 508-510(2005)