

Clinical Implication of Functional Genomics in Cancer

연세의대, 암전이 연구센터

노 재 경

Introduction

세포는 똑같은 유전정보를 가지고 있지만 발생단계에서 자신의 결정된 역할들에 의하여 특정 유전자들만 발현함으로써 다른 세포들과 구별되어진다. 즉 인간의 몸을 형성하는 각각의 세포군들에게는 그들만의 독특한 유전자들이 발현되는 것이다. 이들은 서로 아주 긴밀한 관계를 유지하고 있으며, 만약 이들 유전자들에 돌연변이가 생기거나 변화가 일어나면 질병이 발생한다. 이와같이 생명현상과 질병에 있어서 유전자의 중요한 역할에 대한 관심과 biotechnology의 발달 결과로, 인간 genome의 DNA 염기서열을 해독하고자 하는 Human Genome Project (HGP)가 미국과 유럽 일부 국가들의 협력으로 1990년 시작되었다. Celera Genomics (Craig Venter, Ph.D.)사의 가세로 그 연구 속도에 가속을 더하게 되어, 2001년 2월 HGP consortium은 Nature지에(2001:409:860-921: "Initial sequencing and analysis of the human genome"), 또한 Celera Genomics사는 Science에 (2001:291:1305-1351: "The sequence of the human genome") "인간 유전체 지도 완성"을 발표하였다. 전체 유전체의 99%가 분석되어 생명과학자들에게 많은 유용한 정보를 줄 것으로 기대되나, 29억여 개의 염기서열이 확인된 유전자 지도는 실제 사람의 유전자수는 예측한 바와 달리 30,000~35,000 개로 알려지고, 이들 중 오직 1.1% 만이 실제 단백을 coding하는 exon (intron 24%, intergenic DNA 75%)임이 밝혀지게 되었다. 이들 유전자 중 기능을 알고 있는 유전자는 일만 개 이하이며, 그 기능의 복잡성으로 인해, 처음 HGP가 시작되었을 때의 기대처럼 유전자 해독이 끝나면 거의 모든 유전자 관련 질병의 진단과 치료 및 생명체의 생성, 성장 및 노화에 이르기까지의 비밀이 풀릴 수 있기에는 더욱 많은 유전자의 기능적 연구가 진행되어야 함을 시사하고 있다.

21세기 생명과학의 핵심이 되고있는 게놈연구 결과 21세기에는 생명체로서 활동하는 개체에 대한 근본 염기서열에 대한 정보가 모두 밝혀지게 될 예정이다. 그러나 단순히 엄청난 DNA 서열에 대한 정보 만으로는 생명현상과 질병을 이해하는 유용한 가치를 가지지 못한다. 즉, 1) 염기서열로부터 새로이 밝혀진 유전자들의 기능은 무엇인가? 2) 새로운 유전자들은 어떻게 발현되어지고 통제되어지는가? 3) Genome 구조의 비밀은 무엇인가? 라는 문제들이 제시되는 "Functional Genomics" 시대에 이르게 된 것이다. 지금까지는 한 두 개의 유전자를 클론해서 이들의 성질을 규명하는 것이 주 연구주제이었다. 그러나 유전자의 염기서열이 이미 다 밝혀지게 된 현 상황에서는 원하는 유전자를 어렵게 클론할 필요가 없어지고, 서열정보로부터 연구하고자 하는 유전자만을 골라내어 그 기능연구를 하면 될 것이다. 이러한 시대에 우리는 어떻게 대처해야 할 것인가? 첫째는, 유전자 수준에서 유전체 수준으로 연구 방식이 변함에 따라 알려진 유전자 서열정보를 이용하여 아직 그 기능을 모르고 있는 많은 유전자들의 기능을 신속히 밝히는 것이다. 그 대표적인 예들로 cDNA microarray를 이용한 유전자의 발현정도를 평가하고 molecular pathway의 연관성을 밝히며,

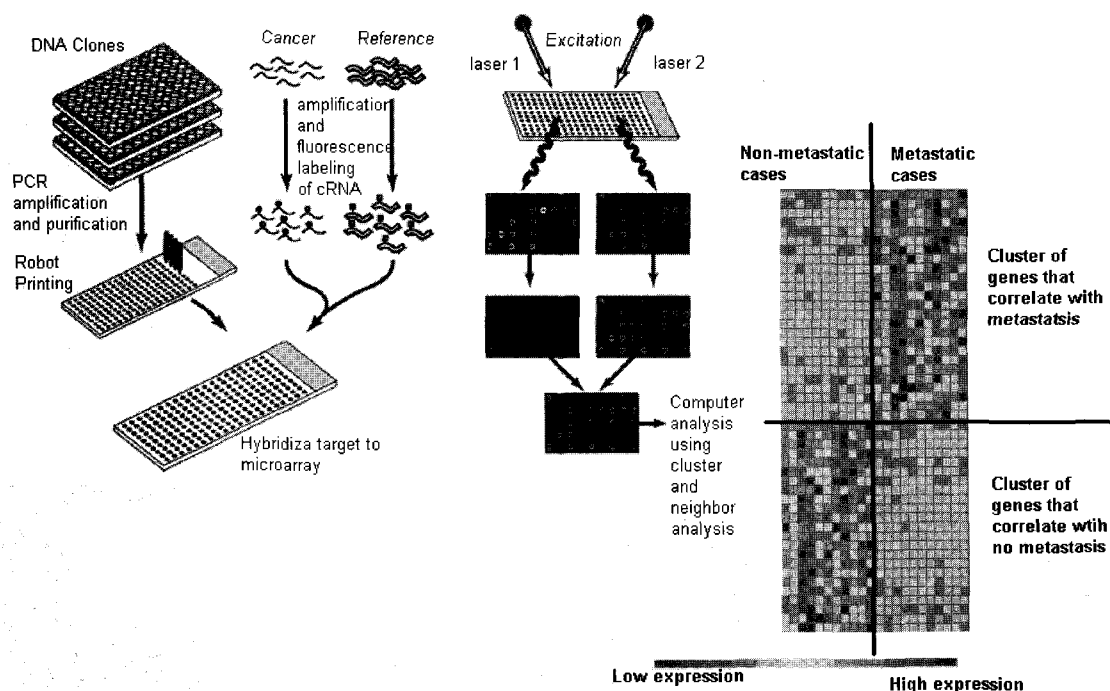
proteomics 및 tissue array를 이용하여 실제 기능을 유전체와 공동 연구함으로써 생물학적 현상과 질병현상을 더욱 유기적으로 이해하고자 하는 것이다. 둘째는 쏟아져 나오는 정보들을 효율적으로 이용할 수 있는 생물정보학의 개발이 필수적이다. 이를 위하여 연구 정보들의 효율적 통합과 검색, 새로운 분석 알고리즘의 개발, genetic network의 구성 및 이해, data mining을 통한 새로운 생물학 지식의 창출을 위해 많은 노력들을 기울이고 있다. 결과적으로 이러한 연구 결과들은 급속히 성장하고 있는 생물산업에 직접적으로 응용되어 신약 개발, 유전병의 진단 등 많은 관련 산업들의 발전에 기여할 수 있으리라 생각한다.

cDNA Microarray

유전자의 발현과 그 유전자에 의한 단백질 발현은 유전자의 기능연구의 기본이다. 발달된 분자생물학 기술로 특정 유전자들의 질병과의 관련성 및 치료에의 응용이 가능하게 되었고, 암 등의 유전학적 질환의 분자 유전적 분류에 대한 연구가 활발히 진행되고 있다. 특히 유방암에선 이미 estrogen receptor와 HER2 oncoprotein이 각각 tamoxifen과 adriamycin의 치료효과를 예측하는 데 쓰이고 있어서 분자 유전적 분류의 임상적 중요성을 알 수 있다. 하지만 지금까지 쓰여진 방법, 즉 미리 정해진 가설에 의해 한가지 유전자나 단백질을 보는 방법에는 한계가 있으며, 지금까지 몇 가지 지표이외에는 임상적 가치가 있는 유전자를 찾아내지 못한 것으로 그 한계가 증명되었다고 볼 수 있다. 그리고 이러한 한계는 최근에 보편화된 pathway concept에 의해서 더욱 명백해졌다. 즉 암세포의 성장 대사를 지배하는 성분들은 서로 복잡한 pathway에 의해서 연관되어 있으므로 어느 한 성분만 보아서는 전체 암세포의 행동을 예측할 수 없다는 것이다. 5년여 전부터 개발된 cDNA microarray를 이용하여 수천 개의 유전자 발현상태를 한 번의 실험으로 볼 수 있게 되기 시작함으로써 이런 학설이 증명되었으며, 이와 같은 high throughput screening method는 유전자의 발현과 기능을 근거로 하는 임상 의학의 발달을 가능하게 하였다.

DNA chip/microarray란 작은 유리 slide나 membrane 위에 수천, 수만 개의 유전자를 얹어놓고 검사하고자 하는 시료에서 추출해 낸 RNA의 발현 정도를 한 번의 실험으로 조사할 수 있는 high throughput screening 방법이다. 1995년 미국 스탠포드 대학의 Pat Brown 등에 의하여 개발된 이 DNA chip은 마이크로프로세서를 이용하여 얇은필의 조그만 면에 이미 염기순서가 알려진 여러 유전자의 DNA를 붙여 놓은 것이다. 반도체 기술과 화학기술의 발달, 그리고 bioinformatics의 발달로 방대한 연구결과의 분석이 가능하게 되었으며, 이를 통하여 질병의 진단과 치료 개념의 변화가 나타나고 있다. 이와 같은 DNA chip의 장점으로는 동일한 chip의 다량 제작이 가능하며 수백 개 이상의 유전자를 동시에 검사 가능하게 되어 기존의 방법보다 수백 배 이상의 시간과 비용을 절감할 수 있을 뿐 아니라, 극히 미량의 DNA로도 chip 제작이 가능한 점 등을 들 수 있다.

DNA chip의 응용 가능성은 무궁무진하여, 이러한 DNA chip이 21세기 생명공학의 시대를 주도할 것으로 예견하고 있다. DNA chip은 유전자 발현연구, 인간유전자 기능 분석연구, 유전 질환의 진단 및 치료, 암 및 질병 관련 유전자발굴에 의한 진단과 치료에의 응용, 신약개발, 유전자치료, 돌연변이 분석, 재조합 동/식물 연구, pharmacogenomics (약물 유전체학) 등 많은 분야에서 응용이 가능하다. 또한 이의 제작과 관련된 반도체 산업 등 공학의 발전과 함께, 결과 분석 과정에서 statistics, computer science의 발전에 따른 bioinformatics라는 새로운 학문이 꽃을 피우는 데 지대한 공헌을 하리라 생각된다.



Proteomics

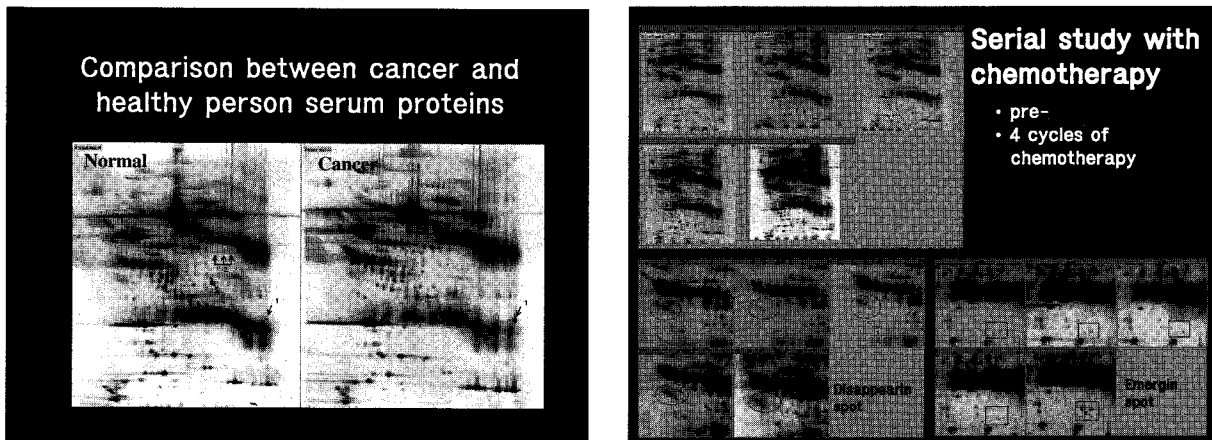
인간 게놈의 전체 염기서열이 다 밝혀졌다고 해도, 이 정보만으로는 유전자 산물의 생체 내의 기능을 파악하는 데는 제한이 있다. 즉, 유전자의 전사 및 번역을 조절하여 단백질의 생성이 조절 되어도, 이 단백질이 세포 내에서 기능을 수행하기 위해서는 단백질-단백질간의 상호작용 및 단백질합성 후 변형과정 등이 정확히 일어나야 한다. 따라서 유전자의 기능을 이해하기 위해서는 세포 내 유전체(genome)의 기능과 함께, 이들이 만들어 내는 단백질체(proteome)의 합성조절과 함께 단백질간의 상호작용, 합성 후 변형과정 등에 대한 연구가 필수적이므로 세포기능연구는 이 두 분야를 연구할 수 있는 유전체학(Genomics)와 단백질체학(Proteomics), 그리고 생물정보학(Bioinformatics)을 반드시 병행하여야 구체적인 연구결과를 산출할 수 있다.

Proteome은 특정 시점에서 수천개의 단백질로 구성된 세포나 조직을 이루는 전체 단백질체(그룹)를 말하는 것이고, proteomics는 단백질 발현 profile, 단백질 변형 및 단백질-단백질간 결합을 총칭한다. Proteome은 실제로 post-genomic 시대에 가장 적합한 연구대상으로 표적물질을 발견하는데 활용되고, 새로운 단백질과 그의 유전자를 검색할 수 있으며, 질환관련 단백질을 식별하여 신약개발의 표적으로 만들 수 있으며, 세포 내 질환발생과정을 규명할 수 있다. Proteomics의 적용분야 역시 매우 폭넓게 넓어지고 있으며 진단, 치료제 개발, 곡물증산, 보호에 활용되고 있다.

HGP 후 인간의 유전자수가 35,000여개임이 밝혀졌고, 이는 초파리나 벌레들의 유전자수인 14,000~18,000개의 3배가 되지 않는다. 이는 복잡한 인간활동을 위해서 한 유전자가 한 단백을 생성하여 기능을 하리라는 기존의 가설을 뒤집는 결과라 하겠다. 즉, 한 개의 유전자가 10개 이상의 단백질을 생성하고 이들의 post-translational modification이나 조합의 차이로 단백질의 다양성을 유도하고 결과적인 인간의 복잡성을 가능하게 되리라고 생각하게 되었다. 즉 한가지 유전자나 한 개의 단백질에 대한 연구를 하는 것은 매우 단편적으로 생명현상을 이해하는데 단편적인 정보 밖에 제공하지 못

하고있다. 이를 해결하기위해 개발된 proteomics는 한번에 조직내의 모든 단백을 있는 그대로 펼쳐서 질병과의 유기적인 관계를 설명하게 해주는 중요한 정보를 제공하고 있다. 이 또한 biotechnology와 bioinformatics의 발달로 가능하게 되었다.

Proteomics를 이용한 임상연구의 한 예로 정상과 암환자의 혈액에서의 단백 발현 양상을 비교하여 암 특이 단백을 찾을 수 있고 이를 tumor marker나 예후인자, 또는 치료의 목표로 이용할 수 있는 가능성의 제시하였다. 또한 항암제 치료를 받고 있는 환자의 혈액에서 치료과정에 따른 단백 변화를 관찰하여 항암제 감수성 및 저항성 단백을 찾아낼 수 있었다. 아직은 다량의 정보를 어떻게 효율적으로 분석하고, 유전자와의 관계 및 pathway 발견, 아미노산 염기서열의 확인 등 해결하여야 할 문제점이 많으나, 향후 계속 연구를 통하여 Proteomics도 Genomics와 같이 질병을 이해하는데 매우 중요한 역할을 할 수 있으리라 생각한다.



Bioinformatics

서열이 밝혀진 수많은 유전자들 중에서 생물학적으로 중요하고 그 임상가치의 가능성이 있는 유전자들을 빨리 선정하여 그 기능을 연구하고, 그들간의 밀접한 연관관계를 규명하는 것이 postgenome시대에서 생물학자들의 나아갈 연구방향이다. 즉, 공개된 서열정보로부터 Bioinformatics의 기술을 효율적으로 이용하여 기능을 예측할 수 있는 유전자들을 가려내어 그 유전자들의 기능 연구를 추진해야 한다. 이를 위해 생명과학자들의 computer science를 이용하여 접근하여야 우리 앞에 무한히 쏟아져 나오는 유전자정보를 제대로 이용할 수 있고, 유전자의 생물학적 기능을 쉽게 밝힐 수 있게 되기 때문이다. 염기서열 정보를 이용할 data mining, annotation, DNA chip image 분석프로그램 개발 등 유전자 정보로부터 우리가 해결해야 할 문제들이 산적해 있다. 이 분야는 전 세계적으로 이제 막 활발히 추진되고 있는 상황이나 이에 대한 전문인력이 매우 부족하므로 생명과학자들도 생각을 전환하여 computational biology 관련기술을 습득하여 생물학과 정보학을 동시에 이용하며 또한 생물정보학자와의 긴밀한 협조가 필요한 시기가 되었다.

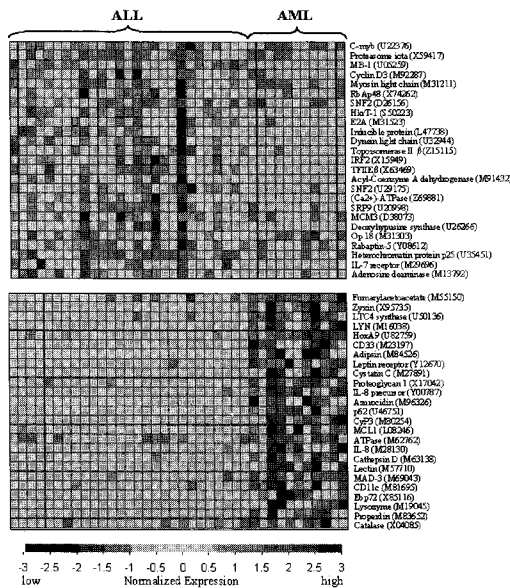
전술한 바와같이 생물학적 현상이 한 두 유전자로 이루어지는 것이 아니며, 또한 여러 유전자가 일률적인 방향이 아닌 유기적인 연관관계를 가지고 기능을 하고 있으므로 이들의 기능을 분석하는 bioinformatics의 tool 개발이 매우 중요하다. 즉 각 유전자들간의 coss-talk가 어떻게 이루어지며

유전자들 사이에 어떤 network를 통해 서로 연관을 가지고 기능하느냐를 밝혀야 할 시대가 된 것이다. 이와 같이 bioinformatics는 1차적인 생물학 실험 결과와 데이터베이스로부터 많은 지식 기반 방법들을 이용하여 새로운 생물학적 지식을 창출해 내기 위해 끊임없이 발전하고 있다. 더욱이 실험방법들이 점점 더 자동화되고, 결과 산출 속도가 가속화됨에 따라 그 중요성은 점점 더 부각할 것으로 예측한다. 특히 post genome 시대의 주역인 Functional Genomics 연구는 신약 개발, 유전병 진단 및 치료 등 기술 집약적인 미래 산업과 밀접한 관계를 가지고 있으므로, bioinformatics는 미래 생명과학을 선도할 가장 핵심 분야로 발전할 것으로 기대한다.

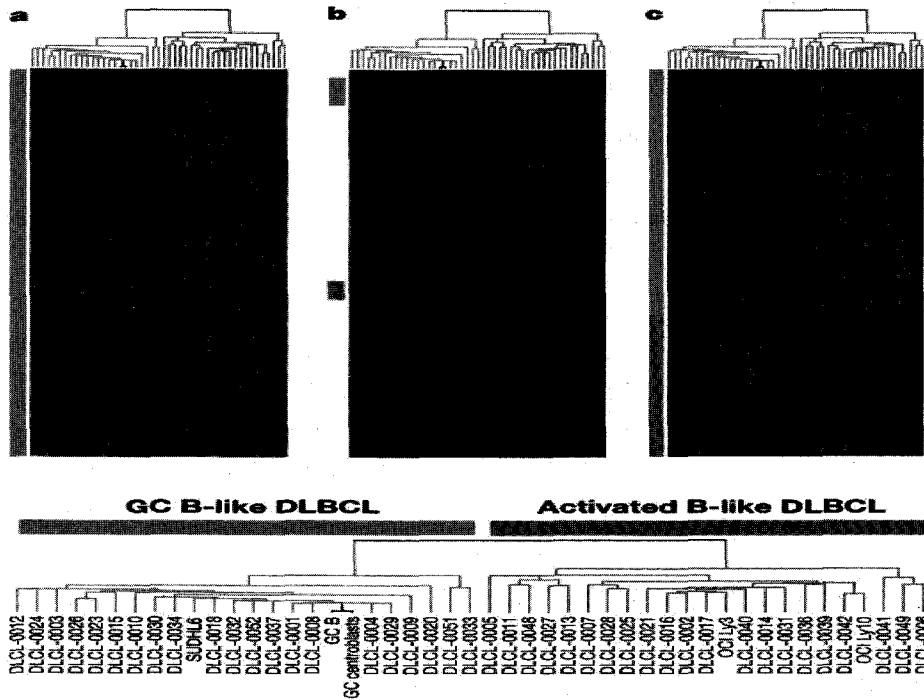
Translational Research: Tailored Therapy Based on Molecular Classification

과거에는 암을 주로 임상 및 형태병리학적으로 분류하였으나, 종양 이형성(tumor heterogeneity)로 인한 다양한 임상경과 차이 때문에 이러한 분류법의 미비한 점이 많이 지적되었다. 특히 환자의 예후와 치료 효과를 예측하는 데는 많은 문제점이 제기되어, 이를 극복하기 위해 암의 분자 유전적 분류에 대한 연구가 활발히 진행되고 있다. 또한 위에서 언급한 바와 같이 1~2개의 유전자만으로 병태생리를 설명할 수 없음을 명백한 사실이기 때문에, DNA chip을 이용하여 많은 유전자들의 발현 pattern을 이용한 분자 유전학적인 분류가 시도되고 있다. 이와 같은 분자유전학적인 분류는 정확한 진단, 치료내성과 효과 예측에 따른 적절한 치료 방법 선택, 예후 예측 등 임상적 모든 분야에 적절하게 이용됨으로써 현재 의학계의 목표인 개인별 맞춤 치료를 가능하게 하는 기본 개념으로 생각된다.

대표적인 예로 MIT Whitehead Institute의 Golub 등은 AML과 ALL을 구별할 수 있는 유전자군을 발견하였으며, 실제로 감별진단이 불확실한 환자들에서 이들 유전자군을 이용하여 분자적 진단이 가능함을 보고하였다.



또한 Stanford University의 Alizadeh 등은 diffuse large B-cell lymphoma 환자들의 유전자 발현을 비교한 후 임상 예후를 예측할 수 있는 유전자 군을 발견하였다. 그리고 이를 근거로 하여 동일한 병리소견을 보이는 환자들을 두 군으로 나눌 수 있게 되어, 진정한 의미의 분자 유전학적인 분류가 가능함을 보고하였다. 또한 B-cell lymphocyte의 maturation 등 생리적 기능을 이해하는 데 도움이 되는 유전자군들도 밝혀짐에 따라 향후 연구에 많은 도움을 주고 있다.



Conclusion

암치료를 높이기 위해서 새로운 항암제의 개발과 함께 암세포의 생물학적 활성화에 근거한 새로운 개념의 치료제가 집중적으로 개발되고 있으며, 특히 개발된 약제의 투여방법에 대한 연구도 매우 활성화되고 있다. 현재 임상에서 효과를 인정받고 있는 대표적인 생물학적 치료제로는 monoclonal Ab로 성장인자 수용체인 Her-2/neu를 억제하는 Herceptin과, 만성골수성 백혈병에서 획기적인 효과를 보고하고있는 Gleevec 등이 있다.

아직 어떤 환자를 대상으로, 어떠한 치료 schedule로 얼마동안 치료를 할 것인지, 어떻게 효과관정을 하여야 하는지 등, 풀어야할 많은 문제점들이 남아있으나 현재까지의 결과로는 생물학적 치료제의 항암제로의 가능성을 확인시켜주고 있다. 또한 최근 밝혀진 인간의 유전자 염기서열을 기초로 하여 암의 발생과 전이에 관련된 유전학적인 정보가 늘어남에 따라 각 개인의 특성에 맞는 암의 정확한 진단 및 치료 약제 선정이라는 맞춤형치료가 가능해지리라 기대해본다. 이와같이 맞춤형 치료의 근간을 조성하기위한 기초연구로 DNA chip, proteomics, tissue array의 사용은 현대과학의 발달의 큰 성과로, 그 결과를 bioinformatics를 이용하여 효과적으로 분석하여 질병의 원인을 이해하고 그를 치료할 수 있는 가능성을 보여주고 있다. 즉, 이들 새로운 기술들은 임상 뿐 아니라 기초 연구와 법적 효능성 등 인류 사회의 모든 분야에 적용이 가능하며, 많은 영향을 끼칠 수 있는

가능성 역시 제시되고 있다.

하지만 지금까지의 기술로는 아직 고가의 chip 제작비나 실험결과 판독의 어려움, 유전정보와 단백질과의 관계를 통한 기능 연구, tissue array 등을 이용한 결과 validation 등 해결해야 할 과제가 많다. 또한, 모든 유전자의 기능이 확인되지 않은 현 시점에서 chip에 넣고자 하는 유전자의 종류나 수의 결정, 이들의 patent 문제 등이 임상 응용까지의 중요한 문제점이다. 이를 해결하기 위해 protein chip과 proteomics, full length cDNA library 제작, chip 제작 기술 개발, 결과 해석을 위한 bioinformatics의 확립과 database 구축, 결과 validation 등의 분야에서 많은 연구가 시행되고 있어, 멀지 않은 시기에 우리가 기대하는 유전자정보에 의한 개인 평가 및 맞춤 치료가 가능해지리라 생각한다.

참 고 문 헌

1. A gene expression database for the molecular pharmacology of cancer. Scherf U, Ross DT, Waltham M, Smith LH, Lee JK, Tanabe L, Kohn KW, Reinhold WC, Myers TG, Andrews DT, Scudiero DA, Eisen MB, Sausville EA, Pommier Y, Botstein D, Brown PO, Weinstein JN. *Nat Genet.* 2000;24:236-44.
2. Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. AJ. Butte, P Tamayo, D Slonim, TR Golub, IS. Kohane. *PNAS USA*, 2000;97: 12182-12186.
3. Drug target validation and identification of secondary drug target effects using DNA microarrays. Marton MJ, DeRisi JL, Bennett HA, Iyer VR, Meyer MR, Roberts CJ, Stoughton R, Burchard J, Slade D, Dai H, Bassett DE Jr, Hartwell LH, Brown PO, Friend SH. *Nat Med.* 1998;11:1293-301.
4. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling, *Nature*, 2000;403:503-511.
5. Molecular portraits of human breast tumours. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D. *Nature* 2000;406:747-52.
6. Molecular Classification of Cutaneous Malignant Melanoma by Gene Expression Profiling. M. Bittner, P. Meltzer, Y. Chen, Y. Jiang, E. Seftor, M. Hendrix, M. Radmacher, R. Simon, Z. Yakhini, A. Ben-Dor, E. Dougherty, E. Wang, F. Marincola, C. Gooden, J. Lueders, A. Glatfelter, P. Pollock, E. Gillanders, D. Leja, K. Dietrich, C., M. Berens, D. Alberts, V. Sondak, N. Hayward, and J. Trent *Nature* 2000;406: 536-540.
7. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. T.R. Golub, D.K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J.P. Mesirov, H. Coller, M. Loh, J.R. Downing, M.A. Caligiuri, C.D. Bloomfield, and E.S. Lander. *Science* 1999;285:531-537.
8. Proteomics in human disease: cancer, heart and infectious diseases. PR Jungblut, U Zimny-Armdt, E Zeindl-Eberhart, J Stulik, K Koupilova, KP Pleissner, A Otto, EC Muller, W Sokolowska-Kohler, G Grabher, G Stoffler. *Electrophoresis* 1999;20:2100-2110
9. Proteomics: quantitative and physical mapping of cellular proteins. WP Blackstock, MP Weir. *Trends Biotechnol* 1999;17:121-127
10. The molecular biology database collection: an online compilation of relevant database resources. AD Baxevanis. *Nucleic Acids Res* 2000;28:1-7
11. The sequence of the human genome. J.C. Venter et al. *Science* 2001;293:1304-1351.