

유전체의학(Genomic Medicine)과 생명정보학(BioMedical Informatics)

글 _ 고 인 송 연구관 · 국립보건연구원 유전체연구부 · insong@ngri.re.kr

들어가며

새로운 과학의 출현은 한 학문분야의 가장자리에서, 즉 다른 학문분야와 만나는 자리에서 생겨난다는 말이 있다. 2000년 6월에 약 10여 년간 지속된 인간유전체계획(Human Genome Project, HGP)의 1차 결과가 발표되면서 소위 포스트지놈시대(Post-Genome Era)를 열었다. HGP 시대의 목표가 인간 유전체를 구성하는 30억 쌍의 모든 DNA 염기서열을 밝히는 것이었다면, 이 포스트지놈시대의 1차 목표는 3만 여개로 밝혀진 인간 유전자 각각의 기능 및 질병과의 관계를 밝혀서 새로운 차원의 의료를 예고하는 유전체의학(Genomic Medicine)을 가능하게 하는 기반을 마련하는 것이다. HGP는 정보기술(IT)과 바이오기술(BT)의 융합을 자연스럽게 가져왔고, 21세기를 IT와 BT가 융합하는 것이

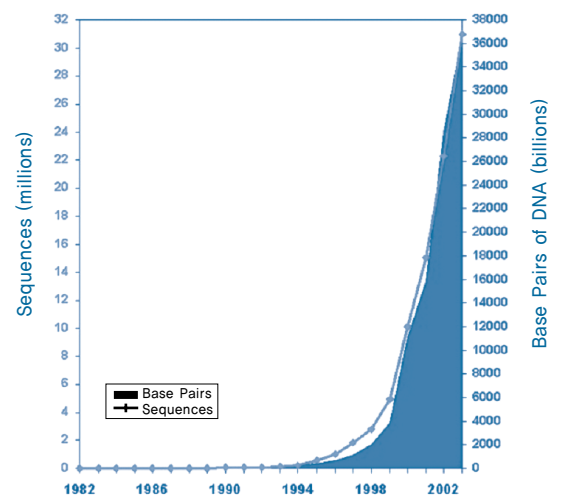
매우 중요한 시대로 만들었다. 정보과학자의 연구와 생물학자의 연구가 연결되는 경향은 날로 커져가고 있고, 새로운 과학적 발견의 기반을 이뤄가고 있다. HGP의 추진에 힘입어 생물학과 정보학이 만나서 태동한 생물정보학(Bioinformatics)은 이제 병원의 데이터베이스에 저장되어있는 환자에 대한 진료기록, 임상검사기록 등의 의료정보를 통합하는 생명정보학(BioMedical Informatics)으로 발전하고 있다. 이것은 개인의 유전체 형에 맞는 차별화된 예방, 진단, 치료를 목표로 하는 소위 맞춤의학을 실현시키기 위해 반드시 가야할 길이다. 본 기고에서는 생물정보학이 유전체의학의 실현을 위해 생명정보학으로 어떻게 발전해야 하는지를 고찰하고자 한다.

정보과학이 되어버린 생물학

과학기술의 발전 그 중에서도 컴퓨터의 발전은 우리 생활을 엄청나게 변화시켰다. 그 발전의 속도는 반도체의 집적도가 2년 내에 2배로 상승하는 것에 따라 기하급수적으로 발전하고 있다. 이제 심지어 생물학 분야에서도 과학기술의 발전으로 인해 인터넷이 세상을 바꾼 것 이상의 변화가 예상된다. 생물학에서 전자공학의 기하급수적 발전과 비슷한 발전속도가 관찰되는 곳이 실험 결과물의 산출양이다. 미국 국립보건원(NIH)의 국립생물학정보센터(NCBI)의 GenBank의 예를 보면 반도체 집적도의 발전을 훨씬 능가하는 속도인 1년에 약 2배로 데이터의 양이 증가한다(그림 1).

이렇게 생물학 실험데이터의 산출량이 기하급수적으로 증가한 데는 전자공학, 기계공학, 컴퓨터과학 등의 발전에 따른 실험과정의 자동화 및 데이터베이스 기술 발전 등의 영향이 매우 컸다. 불과 10여 년 전에 한 시험

Growth of GenBank



<그림 1> Exponential growth of biological data: NCBI GenBank statistics

관에서 한 번에 한 가지 실험을 하던 때의 생물학 연구실은 이제 수백 개의 실험을 동시에 할 수 있는 microwell이나 수천, 수만의 실험을 동시에 할 수 있는 chip을 사용하여 실험하는 연구실로 변했고, 사람은 자동화된 실험장비에 시료를 집어넣고 가끔씩 올바르게 작동하는지를 감시만 하면 되는 식으로 발전했다. 이제 생물학은 이렇게 생산된 많은 양의 데이터로 인해서 더 이상 서술적인 학문이 아니고 정량적인 데이터를 다루는 새로운 생물학으로 변화했다. 따라서 초고속대용량 (high-throughput) 실험방법으로 생산된 방대한 양의 데이터를 활용하고, 데이터의 분석과 그 데이터 간의 상관관계로부터 정보와 지식을 유도할 수 있고, 생명현상을 복잡한 모델로 만들 수 있는 생물학자들이 필요해

졌다. 새로운 생물학의 IT 의존도는 아주 간단히 생물학 실험실에 PC가 없다면 연구할 수 있을까하고 생각해보는 것으로 충분하다. 이미 우리는 실험노트에 기록하고 정리할 수 있는 수준의 생물학적 데이터를 가지고 연구하고 있지 않다. 이제 생물학 연구는 정보를 분석할 수 있는 능력을 가진 사람이 필수적으로 필요하고, 가장 복잡한 시스템이라고 할 수 있는 생명체에 관한 데이터를 정보로 만들어 주는 사람도 필요하다. 즉 IT는 새로운 생물학의 기반이 된 것 뿐만이 아니라 새로운 생물학의 일부가 되었다. 이제 우리는 생물학 데이터를 정보로, 또 지식으로 가공해주는 정보학자, 즉 생물정보학자가 필요하게 되었다.

초고속컴퓨팅(HPC)과 시스템생물학

1999년에 이미 IBM은 물리학이 컴퓨터의 발전을 주도했듯이 생물학이 과학분야의 슈퍼컴퓨팅을 주도할 것이라고 보았다. 새로운 슈퍼컴퓨터의 최대 수요자는 물리학의 거대 프로젝트가 아니라 세월이 갈수록 생물학의 거대 프로젝트로 바뀌고 있었다. 또한 생물학의 컴퓨팅은 염기서열 검색과 비교에 유리한 고도의 병렬컴퓨터 쪽으로 슈퍼컴퓨터의 구조 변화도 가져왔다. 1999년 12월 IBM은 블루진프로젝트(Blue Gene project)를 발표하였다. 이 프로젝트는 본래 2006년까지 새로운 컴퓨터 구조를 가진 petaflop 컴퓨터를 만들겠다는 계획으로, 생명과학의 어려운 문제를 해결하려 하였으나, 현재는 범용 컴퓨터로 추진되고 있다. 그러나 블루진의 가장 중요한 몇 가지 응용프로그램들은 여전히 단백질의 폴딩(folding) 같은 중요한 생물학적 과정을 모의하는 분자적 모델링 등의 생물학 분야를 위한 것이다. 생명과학분야 컴퓨팅의 진화는 컴퓨팅그리드(computing

grid)와 같은 분산컴퓨팅(distributed computing) 모델에도 희망을 줄 것으로 보인다. 미국과 영국에서 건설되고 있는 국가적 그리드(National Grid)와 같이 자기진단이 가능하고, 처리작업의 부하를 자동적으로 분배하고, 사용자가 모르게 틀린 오류를 고치는 등의 컴퓨터 환경을 이를 통해 추구하고 있다.

한편, 최근에 유전체학, 단백질체학, 모델링 및 데이터분석 등의 연구결과가 통합되어 시스템생물학(Systems Biology)이라는 새로운 분야로 발전하고 있다. 궁극적으로 시스템생물학의 목표는 세포, 조직, 기관, 개체의 행동을 컴퓨터모델로 만들려는 것이다. 이러한 컴퓨터 모델링이야말로 슈퍼컴퓨터의 도움 없이는 불가능한 분야일 것이다. 또한 시스템생물학의 시대가 오면 그야말로 생물학은 완전히 수량적인 학문이 되어버릴 것이다.

생물학 데이터베이스의 통합과 데이터마이닝

미국 NIH의 NCBI는 여러 가지 생물학관련 공공 데이터베이스를 운영하고 있다. 그 데이터의 양은 매우 빠른 속도로 기하급수적 증가를 하고 있고, 데이터의 성격도 유전자, 단백질, 화학구조, 관계데이터, 의료영상, 실험데이터영상 등 매우 이질적이다. 또한 산업계 쪽에

도 제한된 접근만을 허용하는 상업적 데이터베이스를 운영하고 있다. 이러한 데이터베이스를 연구자들이 효율적으로 활용하기 위해서는, 데이터베이스들이 상호 연결되고, 동시 접근이 가능해야 하고, 질문에 대한 답을 여러 데이터소스로부터 얻을 수 있어야 한다. 이런

요구를 충족시키려면 접근성과 유연성이 동시에 가능하도록 소프트웨어가 만들어져야 한다. 즉 증가하고 팽창하는 연합된 데이터소스와 데이터라이브리리를 만들 수 있는 유연성이 필요하다. 이러한 노력의 한 가지 예로 IBM은 North Carolina Bioinformatics Consortium과 함께 바이오그리드(BioGrid)를 만들고 North Carolina 지역 16개 대학을 연결하여 데이터의 접근, 슈퍼컴퓨팅, 협력연구환경 등을 제공하는 일을 추진하고 있다. 최근 제약업계는 연구개발 분야의 생산성 저하로 고심하고 있다. 하나의 신약이 성공적으로 개발되어 시장에 나오기까지 약 8억 달러와 15년의 시간이 걸리는 현실에도 불구하고, 매년 시장진입에 성공한 신약의 수는 줄어들고 있다고 한다. 이 문제에 대한 해결책은 유전체학과 단백질학의 연구를 통해 신약후보물질들을 발굴하는 것이며, 이것을 가능하게 하기 위해서는 역시 데

이터의 통합과 접근 및 마이닝을 할 수 있는 강력한 도구의 개발이 필수적이다. 일례로 IBM이 개발하여 Aventis란 제약회사가 사용한 DiscoveryLink라는 도구는, 신약후보물질을 신속히 찾아내는데 매우 희망적이라고 한다. 또한 앞서 언급한 시스템생물학이 성공적으로 기관이나 세포모델을 개발한다면 약물투여에 대한 컴퓨터모의실험을 할 수 있어 신약개발의 과정 중 전임상(preclinical)과 임상실험을 일부 대체할 수도 있을 것이다. 실제로 미국의 Physiome Sciences Inc라는 회사는 심장의 컴퓨터모델을 만들어 제약회사가 개발한 심장치료약의 전임상 실험의 일부를 컴퓨터시뮬레이션으로 대체하는 성과를 올리기도 하였는데 이러한 컴퓨터 모델이 실제와 가장 가까운 모델이 되기 위해서도 여러 가지 관련데이터베이스의 통합을 통한 정보의 생성은 필수적이다.

약물유전체학(Pharmacogenomics)

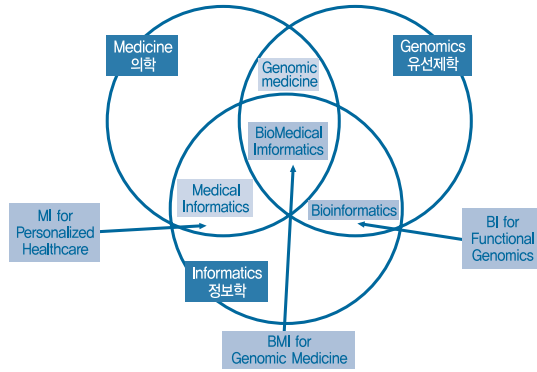
제약업계에서 가장 좋은 약이라 해도 그 약이 처방된 환자의 60%에서 효과가 있는 것이 보통이다. 어떤 약의 임상시험 결과가 성공적이라면 대부분의 환자에서 효과를 보이면서, 소수의 환자에서 약간의 부작용을 보이고, 큰 부작용을 보이는 환자는 한명도 없는 경우일 것이다. 그러나 만약 소수의 환자에서 아주 극적인 효과를 보이지만, 소수의 환자에서는 매우 심한 부작용을 보이는 약이 있다면 이것은 실패한 임상실험결과로 그 약은 허가가 나지 않을 것이다. 그런데 만약 그 약에 대해 극적인 효과를 보이는 환자집단과 심한 부작용을 보이는 환자집단을 구별할 수 있는 간단한 진단방법이

있다면 그 약은 그 진단법과 함께 허가가 날 수도 있을 것이다. 실제로 이런 약의 예가 있는데 그것은 Genentech에서 개발한 전이된 유방암의 치료제인 Herceptin이라는 약이다. 이 약은 Her2라는 유전자 발현 검사를 실시해서 유전자 과발현 환자에게만 투여해야 한다. 이와 같이 미래의 제약업계는 대부분의 환자에게 효과 있는 약을 개발하려는 전략에서 전체 환자는 아니더라도 일부 특정한 유전체형을 가진 집단의 환자에게 극적인 효과를 보이는 약을 개발하려는 쪽으로 전략을 수정할 것이다.

유전체의학과 생명정보학

(그림 2)에서 보는 바와 같이 의학과 정보학의 융합분야인 의료정보학(Medical Informatics)과 유전체학과 정보학의 융합분야인 생물정보학(Bioinformatics)은 또 다른 융합분야의 출현을 가져오고 있는데, 유전체학과 의학의 융합분야인 유전체의학(Genomic Medicine)과 유전체학, 의학, 정보학의 세 분야가 융합한 생명정보학(BioMedical Informatics)이 그것이다. 현재의 의료정

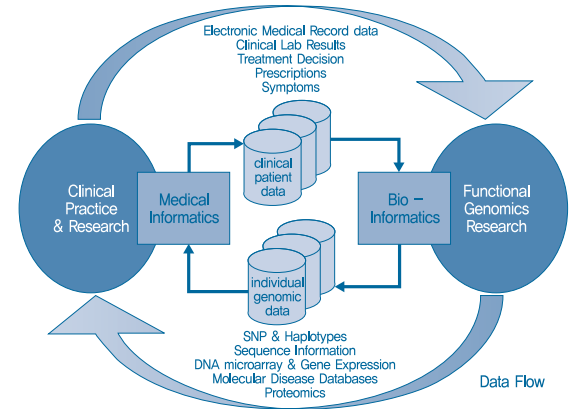
보학은 단순히 환자진료를 위해 그리고 생물정보학은 기능유전체학의 연구를 위해 사용되고 있다. 그러나 포스트지놈시대의 의학으로 발전하면서 임상 의사나 의료정보학자가 환자에 관련된 분자수준의 데이터를 이해하고 의료에 적용하는 것이 더욱 필요하게 되면서 생물정보학도 개별화된 의료를 지원하기 시작한다. 한편 기능유전체학연구도 발전하면서 임상검사결과, 임상시료에



〈그림 2〉 Interdisciplinary nature of newly emerging disciplines in biomedical science fields

대한 주석정보, 환자의 가계도 정보 등 병원 의료정보 시스템의 데이터를 필요로 하기 시작한다. 따라서 의료 정보학도 점차적으로 기능유전체학을 지원하기 시작하게 된다. 따라서 포스트지놈시대의 의료정보학과 생물 정보학은 자연스럽게 만나며 시너지효과를 창출하는 생명정보학으로 융합된다. 따라서 생명정보학은 질병발생과 관련된 분자, 임상, 인구집단 등 각기 다른 수준에서 생산된 정보를 사용하여 질병과 관련된 지식의 분석이나 임상진료의 개별화 방안 등을 목표로 하는 통합적 접근방법이라 할 수 있다.

〈그림 3〉은 의료정보학과 생물정보학 사이의 데이터흐름을 보여주고 있다. 그림 위쪽의 흐름은 일반적인 임상진료와 연구에서 의료정보학자가 수집한 전자의무기



〈그림 3〉 Data flow between the medical informatics and bioinformatics promoting synergy between the two disciplines

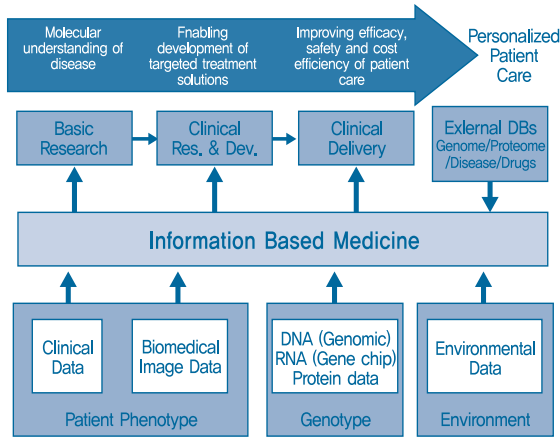
록이나 임상검사결과 등의 데이터가 기능유전체학 연구에 사용될 수 있도록 생물정보학자에게 제공되는 것을 보여주고 있다. 이때 환자의 개인식별정보(이름과 주민등록번호 등)는 개인정보보호를 위해 제거된다. 제공된 의료정보 데이터는 생물정보와 통합되어 질병의 분자적 이해를 증진시키고, 환자의 유전체적 특성과 연결시키는 것을 가능하게 한다. 그림 아래쪽의 흐름은 생물정보학자가 얻은 유전자나 SNP 유형 등 기능유전체학이나 개인대상 유전체학 연구에서 얻어진 정보를 임상정보시스템에 통합하여 환자의 완전한 의무기록의 일부가 되게 하여 환자에 대한 질병예방, 진보된 진단, 개별화된 의료가 가능하도록 하는 것을 보여준다.

정보기반의학(Information-Based Medicine)

유전체학(Genomics)은 고해상도의 유전자 지도를 완성하는 구조유전체학을 HGP의 완성으로 발전시켰고, 유전자의 염기서열과 유전자의 기능, 전사체, 단백질 구조, 생화학적 과정을 연결시키는 기능유전체학을 발전시키고 있다. 기능유전체학 연구의 결과는 유전자와 질병의 관계에 대한 이해를 높이고, 표적기반신약발굴(Target-based drug discovery) 및 개별화된 약의 처방을 추구하는 약물유전체학(Pharmacogenomics)의 발전으로 연결되고 있다. 이러한 유전체학 연구에서 생산된 정보들을 의학에 적용하는 것이 유전체학이며, 그것을 정보시스템의 형태로 구현하여 환자의 유전체정보를 진료에 활용하는 것이 정보기반의학이다. 다시 말해서 정보기반의학은 다양한 임상 및 생명의학 데이터를

정보시스템의 형태로 통합하여 생산된 지식을 가지고 기존의 의학/약학의 진료 방법을 개선하는데 사용하는 것이다. 특히 진단의 정확도를 높임으로써 치료의 효과를 높이는 것이 중심 목표이며, 목표화 된 치료(targeted therapies)와 개별의학(personalized healthcare) 또는 맞춤의학을 지향한다. IBM의 Information Based Medicine(회사이름과 약자가 일치)은 임상유전체학(Clinical Genomics)과 생물·의료영상(Biomedical Imaging)에 중점을 두고 이를 추진하고 있다. 여기서 임상유전체학이란 표현형(Phenotype)과 유전형(Genotype) 데이터를 통합하는 것을 의미한다. 표현형 데이터의 예로는 전자의무기록(EMR)/전자데이터 캡처(EDC) 데이터, 환자치료결과데이터(Patient

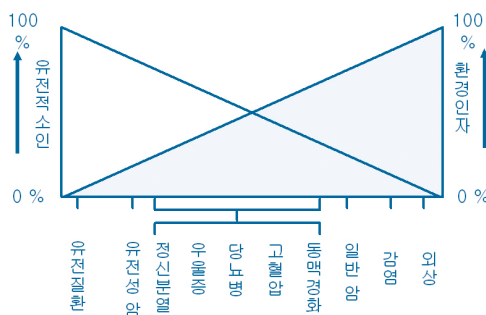
outcome data), 세포영상(Cell Image), 의료영상(Medical Images), 질병진행(Disease Progression)데이터 등이다. 유전형(Genotype) 데이터의 예로는 염기/아미노산 서열(Raw sequence), 단일염기변이(SNP)/Microsatellites, 전사체(Transcriptome) 단백질체(Proteome), 가계도(Genealogy) 등이다.



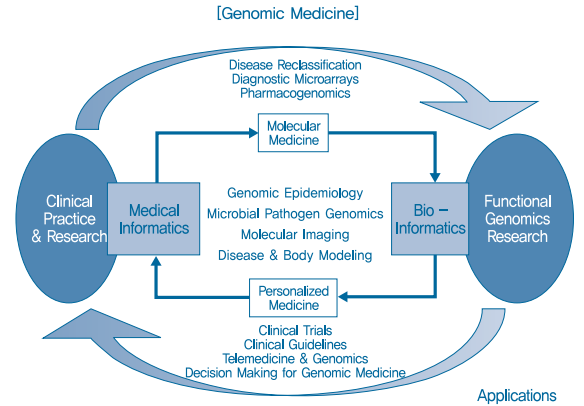
〈그림 4〉 Organization of Information Based Medicine

〈그림 4〉는 정보기반의학의 구성도를 보여주고 있는데 특기할 만한 사항은 환경데이터를 포함시키고 있다는 것이다. 이것은 〈그림 5〉에서 잘 나타나 있듯이 질병의 발생은 유전적 요인만으로 설명되지 않고, 환경적 요인도 고려해야만 완전한 의료정보시스템이 되기 때문이다. 하나의 유전자 돌연변이에 의해 생기는 유전질환의 경우는 유전형에서 그 원인을 찾을 수 있으나, 대부분의 만성질환은 최근에 생활습관질환이라고 불리는 것으로도 알 수 있듯이 운동, 음주, 흡연 등 생활습관을 포함한 환경적 요소가 발병기전의 중요한 요소이기 때문이다.

유전적 요인과 환경적 요인의 상관 관계



〈그림 5〉 각종 질병의 발생에 대한 유전적 요인과 환경적 요인의 기여도



〈그림 6〉 New applications in genomic medicine developed from synergy between molecular medicine and personalized medicine

〈그림 6〉에 표시된 의료정보학과 생물정보학의 시너지를 통해 나타나는 새로운 응용분야의 하나인 유전체역학(Genomic Epidemiology)은 유전체의학에 환경적인 요소를 고려하면서 발전하고 있는 새로운 융합분야이다. 이 분야는 지역사회나 직업군 등 대규모의 인구집단 등에서 생활습관이나 질병발생 양상 등을 수십 년간 관찰하여 필요한 정보들을 데이터베이스화하고, 유전체형 데이터와의 연관성을 연구하기 위해 필요한 혈액 등의 시료도 채취한다. 따라서 유전체역학을 통해 얻은 환경적 데이터는 정보기반의학의 실현을 위한 환경데이터를 제공하는 중요한 원천이 된다.

일반적으로 말해서 현재의 의료는 증상 치료를 목표로 하는 비특이적이고 일반화된 의료이다. 앞으로 전자의 무기록이 사용되고 유전체 정보와의 상관성을 활용하기 시작하면 진단 오류를 줄이고 체계적인 변환의학(Translational Medicine)의 단계로 발전할 것이고, 이것이 성숙되면 인공지능의 도입 등으로 질병의 예방이 가능한 개별화된 의료(Personalized Health Care)로 발전할 것이다.

의사들은 약을 처방할 때 그 약이 임상실험에서 얼마나 좋은 효과를 보였나와 자신의 경험과 지식을 통계적으로 판단하여 처방한다. 유전체연구결과가 새로운 데이터로 등장함에 따라, 앞으로는 의사가 환자에 대한 유전체정보를 고려하여 판단해야할 시기가 곧 올 것이다. 미국의 Mayo Clinic, Emory 대학교 등 대형 병원의 연구센터에서는 이러한 시대를 준비하는 연구를 이미 시작하고 있다. 특히 Mayo Clinic은 의학연구에 사용해도 좋다고 동의한 6백만명의 환자 의무기록을 데이터베이스로 확보하고 있고, 이러한 환자 중에는 유전체 검사


를 할 수 있는 혈액이나 조직 등의 시료가 확보된 환자도 있어, 이들 환자 데이터에 그 시료들을 통해 얻은 분자수준의 유전체 분석정보가 통합되면 의사들은 더 양질의 진료, 즉 환자에 따라 개별화된 진료를 할 수 있게 된다. Mayo Clinic과 IBM은 이를 목표로 협력하여 정보기반의학의 시작품이 될 의료정보시스템을 만들고 있다.

이러한 의료정보시스템이 개발되면 우선 의사는 환자의 질병에 대한 감수성을 조사할 것이다. 예를 들면 환자의 가족력과 환자의 유전체 패턴을 보고 특정한 암에 관련된 유전자나 단백질들이 어떻게 발현될 것인지 또 그 암이 그 환자에서 발생할 확률 등을 예측할 수 있을 것이다. 이런 것을 가능하게 하는 유전체실험기술이나 유전체기능 정보 등은 일부 영역에선 이미 활용이 가능할 정도로 발전되어 있다. 좋은 예는 클러스터링 알고

리즘을 이용하여 DNA chip의 유전자 발현 패턴을 분석하고 백혈병 등 일부 암의 종류를 더욱 세분화하여 진단하려는 시도이다. 암은 일반적으로 한 가지 진단명이 붙더라도 그 발생원인은 아주 여러 가지 일 수 있다. 따라서 같은 진단이 났다고 한 가지 치료법을 사용하면 효과가 없는 경우가 많다. 그런데 DNA chip 등의 유전체 실험결과가 한 가지 진단을 여러 종류의 세부아형(subtype)으로 더 분류해준다면 치료방법도 그 아형별로 달리 해볼 수 있을 것이고, 최적의 치료방법을 찾아낼 수도 있을 것이다. 그러나 문제는 이러한 정보들이 실제 의료에 통합되는 환경이 아직 구현되지 못하고 있는 것이다. 정보기반의학시스템이 개발되면 의사들에게 전산화된 데이터획득과 관리 및 분석을 통해 통찰력을 줄 수 있는 새로운 강력한 도구들을 제공하게 되어, 의견기반진단(opinion-based diagnosis)을 보조하게 될 것이다.

결론

맞춤의학 시대의 가상적인 시나리오를 생각해보자. 아기가 태어나면 혈액으로부터 유전체형을 검사하여 보안 처리된 정보의 형태로 유전체정보은행의 데이터베이스에 수록되게 될 것이다. 주치의사는 아기 부모의 동의 하에 데이터베이스에서 아기의 유전체정보를 검색하여 예방접종 및 각종 질병에 대한 검사 일정을 정하고, 특정 위험질병에 대한 예방적 생활습관 등을 조언해 줄 것이다. 평생 동안 주치의사가 정해준 검진 일정에 맞춰 건강할 때도 위험질환에 대한 예방적인 검사를 하고 그에 따른 조치를 받을 것이다. 질병이 발생할 경우에 주치의사는 그 환자의 유전체형과 분자수준에서 감별 진단된 그 질환의 세부유형의 치료에 탁월한 효과를 보이는 약으로 처방할 것이다. 이런 식으로 예방 중심의 의

학과 개인에게 최적의 약물을 선택하여 치료하는 의학의 혜택을 받음으로써 우리는 건강한 삶을 더 오래 유지할 수 있게 된다. 이 모든 것을 가능하게 하는 하드웨어와 소프트웨어 기술은 이미 IT 분야에서 많이 개발되었다. 문제는 생물정보학과 의료정보학이 많은 교류를 통해서 서로를 이해하고 통합하려는 많은 노력이 있어야만 맞춤의학의 시대라는 꿈을 현실로 실현시킬 수 있다. 가장 중요한 것은 역시 전문인력이다. 생물정보학과 의료정보학을 융합할 수 있는 생명정보학 전문인력을 양성하려는 시도가 본격적으로 이루어져야 21세기 의료시장에서 우리나라가 국가경쟁력을 유지할 수 있을 것이다. 

References

C Kovac (2003) Computing in the Age of the Genome. The Computer Journal, 46(6).

EC-DG IST (2002) Synergy between Research in Medical Informatics and Bioinformatics, white paper.