

표현형-유전형의 연관 연구: From GWAS to NGS

안성민

가천의과대학대학교 분자외과, 가천의대 길병원 중개외과

Phenotype-Genotype Association Study: From GWAS to NGS

Sung-Min Ahn

Department of Molecular Medicine, Gachon University of Medicine and Science, Department of Translational Medicine, Gachon University Gil Hospital, Incheon, Korea

임상외과와 질병의 이름, 표현형

임상외과의 주제인 질병은 그것을 질병으로 인지하고 이름을 붙이고 대응하기 전까지는 존재하지 않는다[1]. 어떤 의미에서 의학의 역사는 질병의 참 이름을 찾아가는 과정이며 이런 역사는 현존하는 많은 질병의 이름에 그대로 반영되어 있다. 예를 들어, 베체트병은 터키의 의사인 베체트가 일련의 증상들을 하나의 질병으로 인지해서 이름을 붙임으로써 새로운 질병으로 자리를 잡게 되었다[2]. 질병의 참 이름이 찾아진다는 의미에는 그 이름에 질병의 원인이 반영된다는, 즉 질병의 원인이 규명되었다는 의미가 포함된다. 베체트병이나 고혈압과 같이, 대다수의 질환에서는 질병의 이름이 증상만을 반영하고 있다(즉, 원인이 규명되지 않았음). 현대의학에서 질병의 참 이름이 밝혀지고 이로 인해 괄목할 만한 발전이 이루어진 두 분야는 바로 감염성 질환 및 유전질환 분야이다.

감염성 질환에서는 질병의 원인(감염원)을 규명함으로써 진단 및 치료에 획기적인 전기가 마련되었다. 예를 들어, 폐렴과 같이 증상에 초점이 맞추어진 이름이 “pneumococcal infection of the lung” 등과 같은 새로운 이름으로 세분화되어 정의되기 시작했고, 이와 관련한 진단 및 치료가 마련되었다[3]. 마찬가지로, 유전질환에서도 분자생물학의 눈부신 발전, 특별히 인간 게놈과 유전자에 관한 이해에 힘입어 유전질환의 원인을 단일 유전자의 변이 수준에서 정의할 수 있게 되었으며, 이 역시 진단과 치료(치료의 경우 감염성 질환에 비해 훨씬 제한적이기는 하지만)의 발전으로 이어졌다. 예를 들어, 증상에 초점이 맞추어졌던 헌팅턴무도병(Huntington's chorea)은 이제는

“a trinucleotide repeat disorder in Huntington gene”이란 이름으로 정의될 수 있다[4].

유전체학의 발전에 힘입어 막 도약을 시작하는 맞춤의학 또는 유전체학에는 질병을 유전자 수준에서 이해하고 재정의한다는 의미가 담겨 있다. 사실 이러한 개념은 이미 1902년에 알카톤노증을 보고한 영국의 의사인 개로드에 의해 제시되었다. 개로드는 이미 그 때 “질병에 대한 감수성이나 면역성을 부여하는 요소들은 우리 자신의 화학적 구조에 내재되어 있으며 염색체 구조를 이루는 분자 그룹 내에 존재한다”라는 개념을 통해 질병과 건강의 화학적 개체성(chemical individuality)을 주창했다[5]. 이를 좀더 현대적 개념으로 바꾸어 말하면 인간의 표현형(키, 혈액형, 질환 등의 특성)은 유전형과 환경적 요인의 상호작용으로 결정되며, 질병의 경우 이와 연관된 유전형을 이해함으로써 질병을 더욱 잘 예방, 진단, 치료할 수 있다는 의미가 포함된다[6]. 그러면, 표현형(또는 질병)이 유전형(또는 유전자)과 연관되어 있는지를 어떻게 밝힐 수 있을까? 이 종설의 목적은 유전체학, 맞춤의학 구현에 핵심이 되는 이 질문에 간략하게 답변하는 것이다.

표현형과 유전형을 어떻게 연관시키는가?

1. Positional cloning

유전질환의 원인유전자를 규명하는 전통적인 방법으로 이 방법을 이용해서 chronic granulomatous disease의 원인유전자가 1986년 발표된 이후로 많은 유전질환의 원인이 이 방법을 통해 밝혀졌으며 여기에는 헌팅턴무도병과 낭포성섬유증도 포함된다[4,7,8]. 지금은 po-

Corresponding author: Sung-Min Ahn
Department of Molecular Medicine, Gachon University of Medicine and Science,
Department of Translational Medicine, Gachon University Gil Hospital, 7-45
Songdo-dong, Yeonsu-gu, Incheon 406-799, Korea
Tel: +82-32-899-6586, Fax: +82-32-899-6519, E-mail: smahn@gachon.ac.kr

※ This work was supported by National Research Foundation grant funded by the Korea government (MEST) (No. 2009-0081789).

Copyright © 2011 Korean Endocrine Society

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

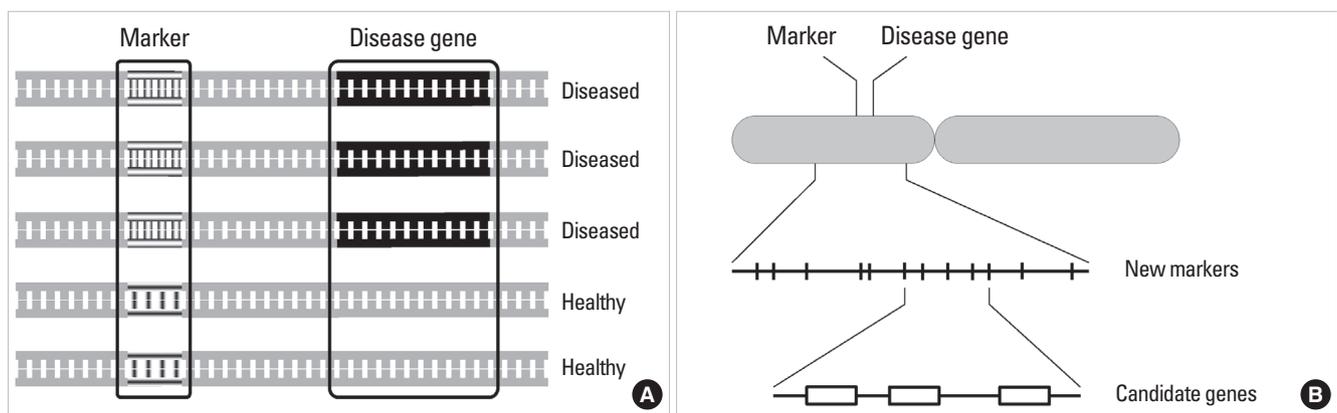
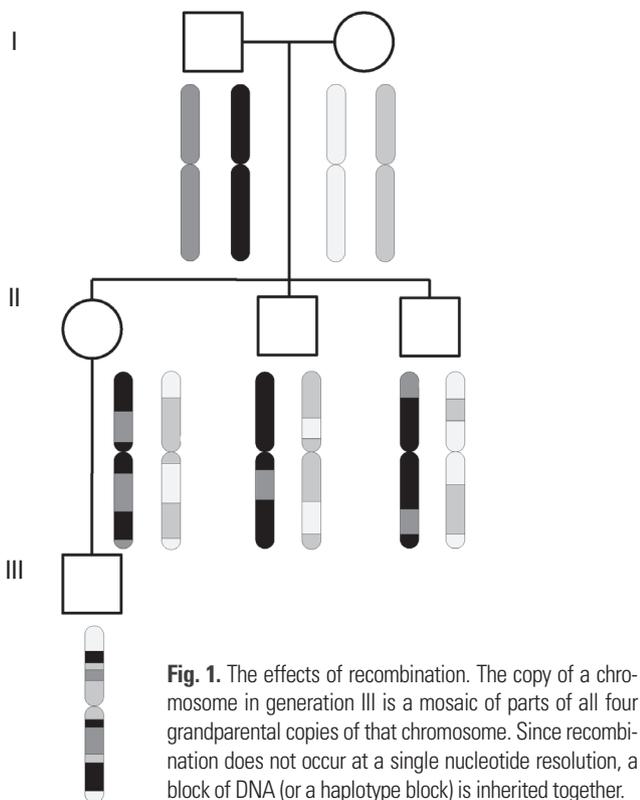
sitional cloning의 활용도가 많지 않지만 마커를 이용해서 마커와 연관된 유전자를 찾는다는 개념은 뒤에 설명할 Genome-wide association study (GWAS)에서도 그대로 적용되기 때문에 이를 이해하는 것은 중요하다. Figure 1에서 보여주는 바와 같이 조부, 조모, 외조부, 외조모의 DNA는 유전자재조합을 거쳐 부모에게로 전달되고 다시 유전자재조합을 거쳐 손자에게 전달된다. 이때 한 가지 중요한 사실은 유전자재조합이 단일염기수준(single nucleotide resolu-

tion)으로 일어나지 않는다는 점이다. 그렇기 때문에 하나의 큰 블록 단위로 유전자재조합이 일어나게 되며 이렇게 함께 유전되는 단위를 일배체형(haplotype)이라 부른다(일배체형의 사전적 의미는 같은 염색체위의 매우 가깝게 연관된 유전자자리에서 한 단위로 유전되는 대립유전자 세트이지만 기전적으로는 유전자재조합의 단위로 이해할 수 있음). Figure 1에서 손자의 염색체 하나에 나타나는 각 조각을 일배체형이라 부를 수 있다.

그렇기 때문에 Figure 2A에서와 같이 우리가 아는 어떤 마커가 있을 경우 해당 마커와 질병유전자가 충분히 가까우며 따라서 같이 유전된다는 전제하에서 우리는 그 마커를 질병유전자를 찾는 길잡이로 활용할 수 있다. Figure 2B는 이 과정을 좀더 구체적으로 보여준다. 이 과정은 마치 현미경을 볼 때 저배율(해상도가 낮은)에서 고배율(해상도가 높은)로 옮겨가며 관찰하는 것과 유사하다. 우선 우리는 해상도(resolution)가 낮은 마커를 이용해 질병유전자의 위치를 저배율로 파악하고, 다시 그 곳에서 새로운 마커군(좀더 촘촘하며 따라서 해상도가 높은)을 이용해 질병유전자의 위치를 좁혀나간다. 이 과정은 질병유전자 후보 군이 충분히 좁혀질 때까지 반복된다[9]. 그러면 마커로는 어떤 것들이 사용되는가? 헌팅턴무도병의 예처럼 초창기에는 제한단편길이다형성(restriction fragment length polymorphism, RFLP)이 많이 사용되었다[10]. RFLP는 제한효소로 DNA를 자를 때 개인 간 염기서열의 차이 때문에 단편의 길이가 달라지는 원리를 이용한 것이다. 아직까지 친자감별에 활용되는 microsatellite라는 반복염기서열도 이러한 마커로 활용된다[11]. 가장 해상도가 높은 마커는 역시 단일염기다형성(single nucleotide polymorphism, SNP)이며 SNP 마커를 통해 GWAS가 가능해진다.

2. Genome-wide association study (GWAS)

GWAS 기법을 이용해 최근 몇 년간 다양한 질환(주로 common,



complex diseases)을 대상으로 대규모의 연구들이 진행되고 있다. 2011년 현재 미국 국립유전체연구소의 GWAS 데이터베이스에 등록된 연구의 수는 총 993개이며 총 4914개의 SNP 마커가 특정 표현형들과 연관되어 있는 것으로 보고되어 있다. 제2형 당뇨병만 해도 2008년부터 2011년까지 29개의 연구가 보고되어 있으며 총 5만 명이 넘는 대규모의 환자-대조군이 GWAS 방법으로 분석되었다[12].

GWAS는 인간유전체사업과 마이크로어레이 기술의 발달이란 두 가지 요소에 의해 가능해졌다[13,14]. 인간유전체사업이 끝나면서 많은 SNP들이 존재하며 이들이 물리적으로 가까운 거리에 있는 유전자들과 연관될 수 있다는 사실을 알게 되었고, 마이크로어레이 기술의 발달은 50만 개, 1백만 개, 현재는 2백만 개 이상의 SNP 마커를 한 번에 스크리닝할 수 있게 해주었다[15].

GWAS의 과학적 정당성은 소위 “common disease, common variant” 가설에 기초하고 있다. 이 가설에 의하면 당뇨병, 고혈압과 같이 흔한 질환, 다시 말하면 단일 요소가 아닌 여러 유전적, 환경적 요소의 영향을 받는 복합 질환(complex diseases)의 경우 빈도가 1-5% 이상 되는 흔한 변이(common variant) 수 개 내지 수십 개만으로 표현형-유전형 연관이 가능하다[16]. 2010년에 유럽과 일본에서 각각 발표된 제2형 당뇨병 GWAS 연구는 각각 12개와 2개의 마커를 당뇨병에 대한 감수성이란 표현형에 연관 짓고 있다[17,18].

하지만, 최근 common variant hypothesis에 기초한 GWAS 연구에

관한 비판도 제기되고 있다. GWAS 연구의 가장 큰 맹점 중 하나는 얻어지는 표현형(또는 질병) 연관 유전형이 결국 마커에 관한 것이라는 사실이다. 어떤 의미에서 GWAS 연구는 위에 설명한 positional cloning (마커를 이용해 질병유전자를 찾는)을 전장유전체적으로 효율적으로 구현한 것이지만 이를 유전자로 연결시켜서 기전적 관계(mechanistic relationship)를 설명하지는 못한다. BRCA1과 유방암과의 연관을 밝힌 연구로 유명한 Mary-Claire King [19]은 질병과 연관되는 것으로 밝혀진 대부분의 common variants들이 질병의 예후나 치료에 생물학적 적절성을 확보하지 못했다고 비판했다[20]. 영국을 대표하는 유전학자인 월터 보드머도 GWAS 연구는 특정 질환에 대해 해당 집단의 유전적 구조를 이해하는 데는 도움이 되겠지만 질병의 유전적 원인을 규명하거나 이를 타겟으로 해 질병을 예방하는 것과 같은(예를 들어, BRCA1 돌연변이가 있을 경우 예방적 유방절제술을 시행하는 것) 실제적인 분야에는 역할이 대단히 제한적일 것이라고 기술한 바 있다[21].

3. 차세대염기서열분석법(Next Generation Sequencing, NGS)과 rare variant hypothesis

1) 염기서열분석의 역사

우리가 일반적으로 알고 있는 DNA 시퀀싱은 생어 시퀀싱을 의미한다. 영국의 프레드릭 생어는 단백질 시퀀싱과 DNA 시퀀싱 방법의

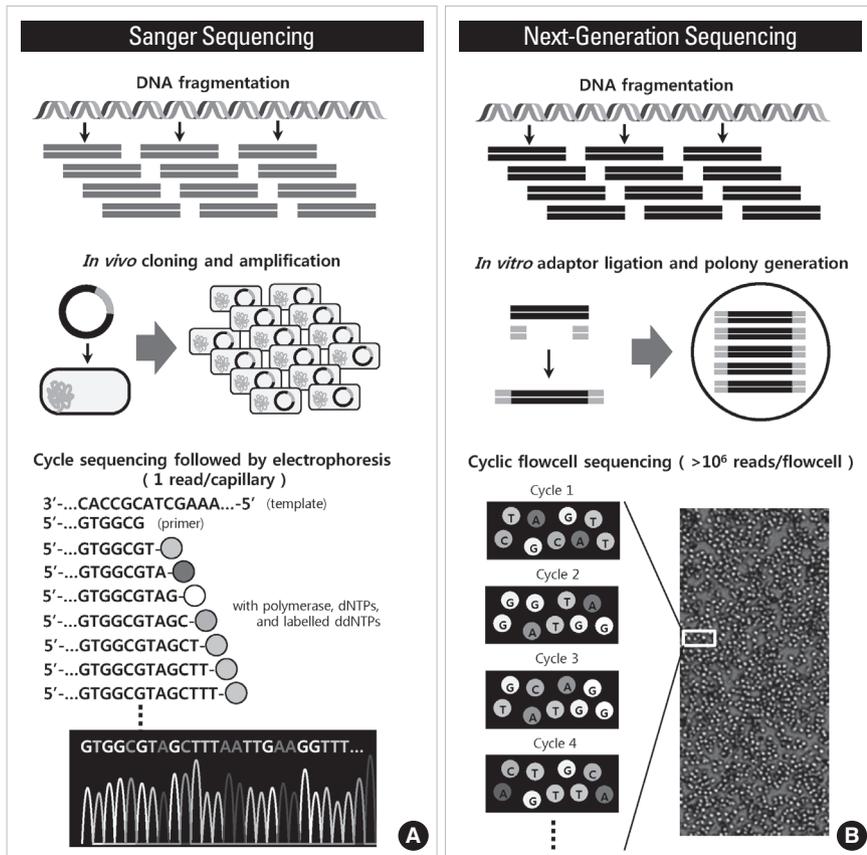


Fig. 3. Sanger sequencing and next-generation sequencing. In the conventional Sanger sequencing using chain-termination method, only one clone can be sequenced per reaction (A). In NGS, millions of clones can be sequenced in parallel, which decreases the sequencing cost dramatically (B).

개발로 1958년 및 1980년에 노벨화학상을 각각 수여하게 된다[22]. 단백질 시퀀싱을 할 때 생어는 단백질의 말단에서 아미노산을 하나씩 잘라나가면서 시퀀싱하는 방식을 택했다. 반면에 DNA (단백질보다 서열이 훨씬 긴) 시퀀싱 방법을 개발할 때는 이와 반대로 DNA를 증폭하면서 시퀀싱하는 “chain termination method”를 택했다[23]. Figure 3A는 DNA를 증폭할 때 dideoxynucleotide에 의해 chain termination이 일어나고 이를 통해 염기서열을 얻는 원리를 보여준다. 생어 시퀀싱 방식에는 두 가지 도약이 있었다. 첫 번째는 방사능 동위원소 대신 형광물질이 사용되면서 DNA 시퀀싱이 자동화될 수 있었고 두 번째는 겔 전기영동 대신 모세관 전기영동(capillary electrophoresis)이 사용되면서 비용절감 및 분석효율이 높아진 것이다. 인간유전체사업과 관련된 시퀀싱 기술 발전의 생생한 이야기는 “크레이그 벤터 게놈의 기적”에 잘 묘사되어 있다[24].

2) 차세대염기서열분석이란?

염기서열분석기술은 인간유전체프로젝트와 맞물려 급속하게 발전했지만, 분석비용은 아직도 높았다. 이 때 분석비용을 낮출 수 있는 가장 확실한 아이디어로 제시된 것이 바로 병렬화(parallelization)이다. 생어 시퀀싱 방식은 한 번에 하나의 클론(clone)만을 시퀀싱할 수 있다. 예를 들어, PCR 증폭을 할 때 프라이머 디자인 상의 결함으로 두 개의 amplicon이 만들어졌고 이를 함께 시퀀싱하게 된다면 그 결과는 읽을 수가 없게 된다. “병렬화”는 만약 한 번에 10개, 100개 혹은 100만 개의 클론을 동시에 시퀀싱할 수 있다면 그 만큼 비용을 낮출 수 있다는 개념을 반영한다. 이 개념은 20세기를 대표하는 분자생물학자이자 2002년 노벨생리의학상 수상자인 시드니 브레너에 의해서

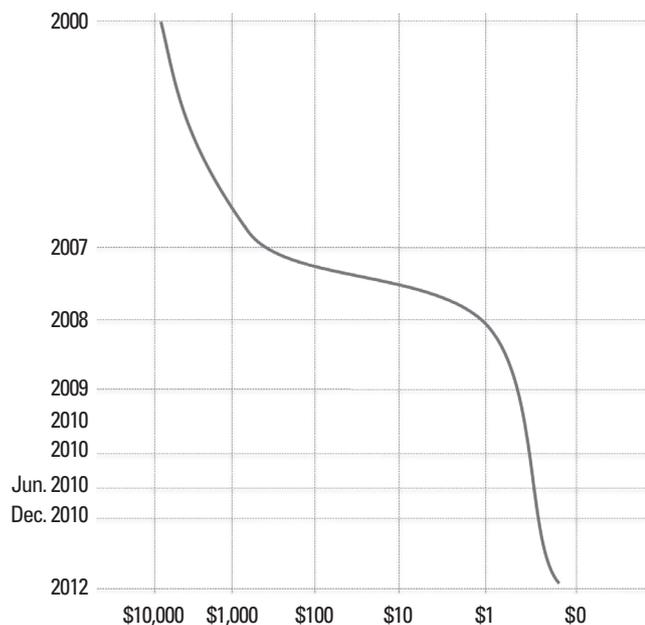


Fig. 4. Sequencing cost per megabase of DNA sequence.

구현되었다[25]. 논문의 제목인 “massively parallel signature sequencing”에도 병렬 개념이 강조되어 있다. 2000년에 발표된 이 방법이 실제 상용화되기까지는 거의 10여 년이 소요되었는데 그 동안 분석법 자체의 진보도 있었지만 실제로 IT분야의 발전이 중요했다. Figure 3B에서 볼 수 있는 것과 같이 병렬시퀀싱의 요체는 고해상도의 이미지를 얻고 이를 분석함으로써 염기서열정보를 병렬로 얻는 것이다. 이 작업은 많은 메모리와 저장공간, 컴퓨팅 파워를 필요로 하며 이것이 일반 실험실에서 구현 가능할 정도로 IT가 발전하고 관련 비용이 내려온 후에야 비로소 상용화가 가능해졌다. Figure 4는 2000년에 비해 현재 1 megabase당 염기서열분석 비용이 얼마나 낮아졌는지를 잘 보여준다. 간단하게 말해서 시퀀싱 비용은 1만배 이상 낮아졌다.

3) Rare variant hypothesis

앞서 GWAS와 common variant hypothesis를 설명한 바 있다. 사실 common variant hypothesis에 대한 반론은 이 가설이 제기된 초창기부터 제기되었다. 1993년에 케네스 바이스와 조셉 터윌리거는 common variant는 아주 적은 생물학적 효과 밖에 가지지 못할 것이며 (만약 그렇지 않다면 자연선택에 의해 이들 변이가 “common”해지지 않았을 것이기 때문에) 복합질환에 대한 감수성은 질환을 촉진시키는 다수의 rare variants와 연관될 것이라고 주장한 바 있다[26].

이 문제를 좀더 실제적인 측면에서 보면 가용한 기술과도 밀접한 관련이 있다. NGS를 통해 아주 적은 비용으로 시퀀싱이 가능한 지금에는 rare variant (빈도가 0.1-1% 미만)를 스크리닝하는 것이 기술적으로 가능하다. 하지만, 그 전에는 common variant hypothesis에 기초해 GWAS 방식으로 접근하는 것이 기술적으로 가능한 최선의 방식이었을 수 있다. Rare variant hypothesis를 이용해 당뇨병과 같은 common disease를 이해하려는 방식에는 다음과 같은 몇 가지 전제들이 있다. 1) rare variants가 common diseases의 일차적인 촉진요소가 된다. 2) common diseases는 처음에 common disease-common variant model에 의해 주장되었던 것보다 멘델유전질환(단일유전자 질환) 모델에 더 가깝다. 3) GWAS 연구를 통해 찾아진 신호의 상당수가 다수의 rare variants의 복합적인 효과가 common variant에 반영된 것일 수 있다[27,28]. 요약하자면, rare variant hypothesis는 우리가 희귀유전질환을 이해하기 위해 사용하던 모델을 common, complex diseases에도 적용해야 한다는 개념을 반영한다.

4) Common variant hypothesis와 rare variant hypothesis에 접점은 없는가?

이 종설에서 지금까지 우리는 표현형과 유전형을 연관시키는 방법을 다루었다. 유전질환의 원인유전자를 찾는 전통적인 positional cloning 방법에서 최근 각광받았던 GWAS (common variant hypothesis에 기초한), 그리고 NGS의 발전에 의해 새롭게 대두되고 있는 rare variant hypothesis, 국내에서도 많은 GWAS 연구가 진행되었고,

이러한 연구들 역시 범부처로 기획되고 있는 유전체사업과 맞물려 새로운 도약을 준비하고 있다.

현재 common variant hypothesis와 rare variant hypothesis를 연결시키는 가장 성공적인 방법은, 즉 GWAS의 결과를 NGS를 이용해 더 깊게 전개해나갈 수 있는 방법은 common variants가 발견된 혹은 주위의 유전자를 시퀀싱해서 rare variants를 찾아나가는 것이다[21]. 요한슨 등[29]은 2010년에 Nature Genetics에 발표한 논문에서 hypertriglyceridemia 환자를 대상으로 한 GWAS 연구에서 찾아진 4개의 common variants가 위치한 유전자 APOA5, GCKR, LPL, APOB를 438명의 환자에서 시퀀싱함으로써 154개의 rare variants를 찾았다는 보고를 한 바 있다. Rare variants는 빈도가 낮기 때문에 GWAS로 연구된 환자군을 다시 시퀀싱할 때, NGS로 분석할 집단에서 rare variants의 빈도를 최대한 높일 수 있도록 GWAS data에 기초해 샘플을 선택할 수 있게 해주는 생물정보학 도구도 등장하고 있다[30].

요 약

표현형과 유전형을 연관시키는 것은 맞춤의학, 유전체의학의 근간을 이룬다. 이를 위해서 최근 대규모로 진행되고 있는 GWAS 연구는 많은 장점이 있지만 common variant hypothesis에 기초한 한계, 기전으로 연결되지 않기 때문에 발견된 결과를 질병의 원인 규명 및 예방에 활용할 수 없는 한계를 안고 있다. NGS의 도입으로 시퀀싱 비용이 낮아졌고, 바야흐로 개인유전체학의 시대가 도래했다. NGS의 발전 덕분에 rare variant hypothesis를 검증하는 것을 기술적으로 검증하는 것이 이제는 가능하다. GWAS에 찾아진 질병연관 common variants들이 사실상 rare variants의 종합적인 효과를 반영하는 것이라는 관찰들이 제기되고 있으며, 이러한 common variants와 연관된 유전자를 다시 환자군에서 시퀀싱함으로써 질병과 연관된 rare variants를 찾는 접근방법이 흥미롭게 대두되고 있다.

참고문헌

1. Rosenberg CE: Disease in history: frames and framers. *Milbank Q* 67 Suppl 1:1-15, 1989
2. Behçet H: Über rezidivierende Aphthose durch ein Virus verursachte Geschwüre am Mund, am Auge, und an den Genitalien. *Dermatol Wochenschr* 105:1152-1157, 1937
3. Tuomanen EI, Austrian R, Masure HR: Pathogenesis of pneumococcal infection. *N Engl J Med* 332:1280-1284, 1995
4. Huntington JL, Dueck A: Handling missing data. *Curr Probl Cancer* 29: 317-325, 2005
5. Nussbaum RL, McInnes RR, Willard HF: *Genetics in Medicine*. 7th ed. p485, Korea, Epublic, 2008
6. Feero WG, Guttmacher AE, Collins FS: Genomic medicine--an updated primer. *N Engl J Med* 362:2001-2011, 2010
7. Royer-Pokora B, Kunkel LM, Monaco AP, Goff SC, Newburger PE, Baeh-

- ner RL, Cole FS, Curnutte JT, Orkin SH: Cloning the gene for an inherited human disorder--chronic granulomatous disease--on the basis of its chromosomal location. *Nature* 322:32-38, 1986
8. Anderson D: Cystic fibrosis of the pancreas and its relation to celiac disease: a clinical and pathological study. *Am J Dis Child* 56:344-399, 1938
9. Hartwell L, Hood L, Goldberg ML, Silver LM, Reynolds A: *Genetics: from genes to genomes*. 2nd ed. pp387-398, McGraw-Hill College, 2000
10. Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, Tanzi RE, Watkins PC, Ottina K, Wallace MR, Sakaguchi AY, et al.: A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* 306: 234-238, 1983
11. Risch NJ: Searching for genetic determinants in the new millennium. *Nature* 405:847-856, 2000
12. Hindorf LA, Junkins HA, Hall PN, Mehta JP, Manolio TA: A catalog of published genome-wide association studies. Available at: www.genome.gov/gwastudies. Cited on Sept. 4th, 2011
13. Lander ES, Linton LM, Birren B, et al.: Initial sequencing and analysis of the human genome. *Nature* 409:860-921, 2001
14. Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ: High density synthetic oligonucleotide arrays. *Nat Genet* 21:20-24, 1999
15. Al-Chalabi A: Ch. 8. Genome-wide association studies. In: Al-Chalabi A, Almsy L, eds. *Genetics of complex human diseases: a laboratory manual*. New York, Cold Spring Harbor Laboratory Press, 2009
16. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM: Finding the missing heritability of complex diseases. *Nature* 461:747-753, 2009
17. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, Zeggini E, Huth C, Aulchenko YS, Thorleifsson G, McCulloch LJ, Ferreira T, Grallert H, Amin N, Wu G, Willer CJ, Raychaudhuri S, McCarroll SA, Langenberg C, Hofmann OM, Dupuis J, Qi L, Segre AV, van Hoek M, Navarro P, Ardlie K, Balkau B, Benediktsson R, Bennett AJ, Blagieva R, Boerwinkle E, Bonnycastle LL, Bengtsson Bostrom K, Bravenboer B, Bumpstead S, Burtt NP, Charpentier G, Chines PS, Cornelis M, Couper DJ, Crawford G, Doney AS, Elliott KS, Elliott AL, Erdos MR, Fox CS, Franklin CS, Ganser M, Gieger C, Grarup N, Green T, Griffin S, Groves CJ, Guiducci C, Hadjadj S, Hassanali N, Herder C, Isomaa B, Jackson AU, Johnson PR, Jorgensen T, Kao WH, Klopp N, Kong A, Kraft P, Kuusisto J, Lauritzen T, Li M, Lieveve A, Lindgren CM, Lyssenko V, Marre M, Meitinger T, Midtthjell K, Morken MA, Narisu N, Nilsson P, Owen KR, Payne F, Perry JR, Petersen AK, Platou C, Proenca C, Prokopenko I, Rathmann W, Rayner NW, Robertson NR, Rocheleau G, Roden M, Sampson MJ, Saxena R, Shields BM, Shrader P, Sigurdsson G, Sparso T, Strassburger K, Stringham HM, Sun Q, Swift AJ, Thorand B, Tichet J, Tuomi T, van Dam RM, van Haefen TW, van Herpt T, van Vliet-Ostapchouk JV, Walters GB, Weedon MN, Wijmenga C, Witteman J, Bergman RN, Cauchi S, Collins FS, Gloyn AL, Gyllenstein U, Hansen T, Hide WA, Hitman GA, Hofman A, Hunter DJ, Hveem K, Laakso M, Mohilke KL, Morris AD, Palmer CN, Pramstaller PP, Rudan I, Sijbrands E, Stein LD, Tuomilehto J, Uitterlinden A, Walker M, Wareham NJ, Watanabe RM, Abecasis GR, Boehm BO, Campbell H, Daly MJ, Hattersley AT, Hu FB, Meigs JB, Pankow JS, Pedersen O, Wichmann HE, Barroso I, Florez JC, Frayling TM, Groop L, Sladek R, Thorstein-

- sdottir U, Wilson JF, Illig T, Froguel P, van Duijn CM, Stefánsson K, Alshuler D, Boehnke M, McCarthy MI: Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet* 42:579-589, 2010
18. Yamauchi T, Hara K, Maeda S, Yasuda K, Takahashi A, Horikoshi M, Nakamura M, Fujita H, Grarup N, Cauchi S, Ng DP, Ma RC, Tsunoda T, Kubo M, Watada H, Maegawa H, Okada-Iwabu M, Iwabu M, Shojima N, Shin HD, Andersen G, Witte DR, Jorgensen T, Lauritzen T, Sandbaek A, Hansen T, Ohshige T, Omori S, Saito I, Kaku K, Hirose H, So WY, Beury D, Chan JC, Park KS, Tai ES, Ito C, Tanaka Y, Kashiwagi A, Kawamori R, Kasuga M, Froguel P, Pedersen O, Kamatani N, Nakamura Y, Kadowaki T: A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at UBE2E2 and C2CD4A-C2CD4B. *Nat Genet* 42:864-868, 2010
 19. Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, King MC: Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* 250:1684-1689, 1990
 20. McClellan J, King MC: Genetic heterogeneity in human disease. *Cell* 141: 210-217, 2010
 21. Bodmer W, Bonilla C: Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet* 40:695-701, 2008
 22. "Frederick Sanger - Autobiography". Nobelprize.org. http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1980/sanger.html, 2011
 23. Hartwell L, Hood L, Goldberg ML, Silver LM, Reynolds A: Genetics: from genes to genomes. 2nd ed. pp302-307, Boston, McGraw-Hill College, 2000
 24. Venter C/노승영: 크레이그 벤터 게놈의 기적(A life decoded). 서울, 추수밭, 2009
 25. Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, Roth R, George D, Eletr S, Albrecht G, Vermaas E, Williams SR, Moon K, Burcham T, Pallas M, DuBridge RB, Kirchner J, Fearon K, Mao J, Corcoran K: Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol* 18:630-634, 2000
 26. Hall SS: Revolution postponed. *Sci Am* 303:60-67, 2010
 27. Cirulli ET, Goldstein DB: Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet* 11:415-425, 2010
 28. Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB: Rare variants create synthetic genome-wide associations. *PLoS Biol* 8:e1000294, 2010
 29. Johansen CT, Wang J, Lanktree MB, Cao H, McIntyre AD, Ban MR, Martins RA, Kennedy BA, Hassell RG, Visser ME, Schwartz SM, Voight BF, Elosua R, Salomaa V, O'Donnell CJ, Dallinga-Thie GM, Anand SS, Yusuf S, Huff MW, Kathiresan S, Hegele RA: Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. *Nat Genet* 42:684-687, 2010
 30. Edwards TL, Song Z, Li C: Enriching targeted sequencing experiments for rare disease alleles. *Bioinformatics* 27:2112-2118, 2011