

엑솜 염기서열 분석 방법을 이용한 단일유전자질환의 원인 유전자 발굴

울산대학교 의과대학 서울아산병원 아산생명과학연구소

이 종 극

Exome Sequencing in Mendelian Disorders

Jong-Keuk Lee

Asan Institute for Life Sciences, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea

More than 7,000 rare Mendelian diseases have been reported, but less than half of all rare monogenic disorders has been discovered. In addition, the majority of mutations that are known to cause Mendelian disorders are located in protein-coding regions. Therefore, exome sequencing is an efficient strategy to selectively sequence the coding regions of the human genome to identify novel genes associated with rare genetic disorders. The "exome" represents all of the exons in the human genome, constituting about 1.5% of the human genome. Exome sequencing is carried out by targeted capture and intense parallel sequencing. After the first report of successful exome sequencing for the identification of causal genes and mutations in Freeman Sheldon syndrome, exome sequencing has become a standard approach to identify genes in rare Mendelian disorders. Exome sequencing is also used to search the causal genes and variants in complex diseases. The successful use of exome sequencing in Mendelian disorders and complex diseases will facilitate the development of personalized genomic medicine.

Key Words : Mendelian disorders, Exome sequencing, Next-generation sequencing (NGS)

최근 개인의 모든 유전체 염기서열을 분석할 수 있는 차세대 염기서열 분석(next-generation sequencing) 기술의 발전으로 다양한 분야에서 학문적 발전이 가속화되고 있다. 특히 환자의 유전체에 존재하는 단백질을 코딩하는 모든 엑손(exon) 부위를 선별하여 대량의 염기서열을 분석하는 엑솜 염기서열 분석(exome sequencing) 방법은 다양한 유전질환의 원인 유전

자를 발굴하는데 사용되고 있다. 엑솜 염기서열 분석 방법은 특정 유전모형으로 다음 세대에 전달되는 질병 가계도 시료에서 환자와 정상인의 모든 엑손 부위를 차세대 염기서열 분석기술로 해독하여 질병 특이적인 원인 유전변이형을 발굴하는 방법이다(Fig. 1). 본 논문에서는 단일유전자질환(mendelian disorders)의 원인 유전자 발굴을 위한 엑솜 염기서열 분석 방법에 대한 전반적인 내용을 고찰해 보고자 한다.

접 수: 2010년 11월 22일
수정본접수: 2010년 12월 10일
게재승인일: 2010년 12월 13일
게 재 일: 2010년 12월 31일
책 임 저 자: 이종극

우138-736 서울특별시 송파구 풍납동 388-1
울산대학교 의과대학 서울아산병원 아산생명과학연구소
Tel: 02)3010-4142, Fax: 02)486-3312
E-mail: cookie_jklee@hotmail.com

본 논문의 내용은 2010년 11월 18일 제13차 대한의학유전학회 임상유전학 추계연수강좌에서 강연되었음.

엑솜 염기서열 분석의 대상 질환으로써 단일유전자 질환군

지금까지 약 7,000개 가까운 단일유전자질환이 보고되어 있으며, 이 중에서 원인 유전자가 밝혀진 질병은 2,893개이고 원인이 밝혀지지 않은 질병은 1,771개이다(Table 1). 그

Table 1. The Number of Mendelian Disorders

Classification	Number
Gene known	2,893
Gene unknown	1,771
Suspected disorders	1,977
Total	6,641

리고 추가적으로 단일유전자질환일 것으로 추정되는 질병도 약 1,977개 정도가 존재하는 것으로 알려져 있다¹⁾.

원인 유전자가 밝혀진 단일유전자질환의 경우, 대부분의 원인 유전변이형들은 단백질을 코딩하는 엑손 부위에 존재하고 있다. 더욱이 인간 유전체의 단백질 코딩 부위는 질병의 원인 유전변이형의 약 85%를 차지할 것으로 예측되고 있다²⁾. 이러한 사실은 인간의 유전체에서 약 1.5%를 차지하고 있는 엑손 부위만을 판별하여도 단일유전자에 의하여 발생하는 대부분의 원인 유전자 및 원인 유전변이형 검색이 가능함을 시사한다. 따라서 단일유전자질환의 원인 유전자 및 원인 유전변이형 발굴을 위해서는 엑솜 염기서열 분석방법이 적은 비용으로 매우 효과적으로 사용할 수 있는 유전 분석법이 될 것이다.

엑솜 염기서열 분석 방법의 실험적 절차 및 방법

엑솜 염기서열 분석 방법은 우선 약 3 ug의 genomic DNA에서 상업적으로 이용 가능한 엑솜 분리 키트(exome capture kits)를 사용하여 전체 유전체 중 단백질을 코딩하는 엑손 부위만을 선별적으로 분리하게 된다. 이렇게 분리된 DNA 조각들을 Illumina 회사의 Genome Analyzer 또는 Life Technologies 회사의 SOLiD와 같은 차세대 염기서열 분석기를 이용하여 대량의 염기서열을 분석하고 분석된 염기서열은 인간의 표준 염기서열(reference genome sequence)과 비교하여 유전변이형의 종류와 빈도를 조사하게 된다(Fig. 2). 따라서 엑솜 염기서열 분석 방법은 1) 엑솜 분리, 2) 차세대 염기서열 분석, 3) 변이형 발굴을 위한 염기서열 데이터 분석으로 구성된 일련의 과정을 통해 진행된다.

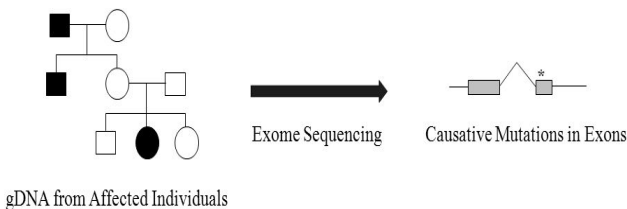


Fig. 1. Exome sequencing in mendelian disorders for disease gene discovery.

1. 엑솜 분리(exome capture)

전체 유전체 염기서열 중에서 단백질을 코딩하는 엑손 부위만을 선별하기 위해서는 다음의 순서를 거치게 된다. 우선 약 3 ug의 genomic DNA 시료를 사용하여 특정 크기(예, 150 bp)로 genomic DNA를 자른 후 엑손 부위의 염기서열에 상보적인 DNA 또는 RNA probes와 결합을 시킨다. 엑손 부위만을 선별적으로 분리하기 위해 biotin으로 표지되어 있는 probes를 사용한다. Probe에 결합한 엑손 부위의 genomic DNA 조각은 streptavidin으로 코팅되어 있는 bead로 분리한 후 증폭하여 차세대 염기서열 분석기로 염기서열을 분석한다(Fig. 3).

2. 차세대 염기서열 분석

차세대 염기서열 분석기는 주로 Roche 회사의 GS-FLX, Illumina 회사의 Genome Analyzer 및 Life Technologies 회사의 SOLiD 방법이 주로 사용되고 있다. 그러나 작은 크기로 구성된 엑손을 분석하는 목적으로는 Genome Analyzer 또는 SOLiD가 가장 적합한 분석 기술이다. 엑솜 염기서열 분석을 위해 사용되는 SOLiD의 경우에는 양쪽에서 약 50 bp 씩을 읽어서 총 30 Gb 정도의 염기서열을 분석하게 되고 Genome Analyxer IIx의 경우에는 다양한 크기로 읽을 수 있으며 주로 36 bp, 76 bp 또는 100 bp를 양쪽에서 읽어서 총 50 Gb의 염

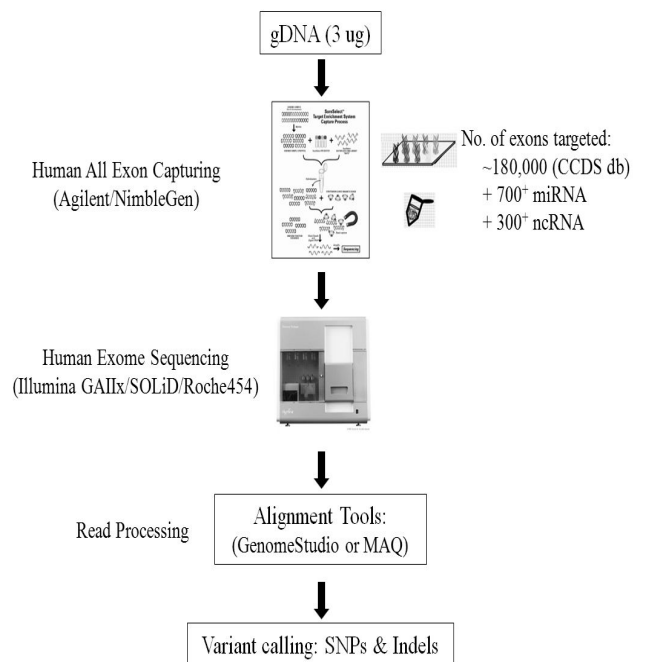


Fig. 2. Workflow for exome sequencing.

Table 2. Comparison of Next-generation Sequencing Technologies for Exome Sequencing

	GS-FLX (454)	SOLiD (V4)	GAllx
	Roche	Life Technologies	Illumina
Technology	Synthesis	Ligation	Synthesis
Read length (bp)	~400	50x(2)	36x(2), 76x(2), 100x(2)
Throughput	~500Mb	~30Gb	up to ~50Gb
Run time (50bp or 100bpx2)		12 days (50bpx2)	10 days (100bpx2)
Multiplexing	yes	yes	yes
Main applications	genome sequencing, metagenomics, transcriptomics	genome sequencing, gene expression, ChIP-sequencing	genome sequencing, ChIP-sequencing, mRNA-sequencing

기서열 분석이 가능하다(Table 2). 최근 SOLiD 4hq의 최신 버전은 기존보다 약 3배 이상의 염기서열 분석이 가능하고 Illumina 회사의 HiSeq2000의 경우에는 기존의 시스템 보다 약 4배 이상의 염기서열을 분석할 수 있는 능력을 가지고 있다.

3. 변이형 발굴을 위한 염기서열 데이터 비교 분석

생산된 대량의 염기서열 정보는 인간 표준 유전체 염기서열과 비교하여 새로운 유전변이형을 탐색하게 된다. 이때 사용하는 표준 염기서열은 NCBI build 36.3 (UCSC의 hg18와 동일함)을 이용하거나 최근에 개정된 버전(NCBI build 37.1 또는 UCSC hg19)을 사용할 수도 있다. 두 개의 염기서열 간의 비교

분석 프로그램으로는 Maq, Bowtie, SOAP, GSNAP 등의 다양한 프로그램이 이용 가능하다. 비교 분석 프로그램을 선정하는 기준으로는 전산 분석 수행 속도와 정확도가 가장 중요한 척도로 사용되고 있다. 또한 염기서열 비교 분석을 통해서 발굴된 유전변이형 정보는 기존의 database single nucleotide polymorphism (dbSNP)에 기록된 유전변이형과 비교함으로써 새로운 유전변이형인지 또는 이미 밝혀진 유전변이형인지를 확인하게 된다. NCBI의 dbSNP 데이터베이스는 주로 build 130을 사용하고 있으며 최근에 개정된 build 131을 기준으로 분류할 수도 있다. 이러한 분류는 아주 희귀하게 발생하는 단일 유전자질환의 원인 유전자 발굴을 위하여 매우 중요한 판단 기준으로 사용된다.

엑솜 염기서열 분석으로 생산된 정보로 부터 질병의 원인 유전변이형 발굴 전략

2009년에 처음으로 Freeman-Sheldon 증후군을 가진 환자 4명과 정상인 8명의 유전체 시료를 사용하여 엑솜 염기서열 분석을 수행하여 MYH3 유전자의 돌연변이가 이 질병의 원인 유전자임을 밝힌 이후로 매우 많은 종류의 단일유전자질환의 원인 유전자 발굴을 위해 엑솜 염기서열 분석 방법이 사용되고 있다³⁾. 한 개인의 유전체에서 엑솜 염기서열 분석에 의하여 발굴된 코딩 SNP (cSNP)은 약 17,000개 정도 발견되고 있으며 기존의 dbSNP에 보고된 것과 대조군의 엑솜 염기서열 분석으로 발견된 유전변이형을 제외하고 나면 개인별로 약 350개 정도가 단백질의 변화를 유발하는 유전변이형으로 남게 된다. 특히, 동일한 단일유전자질환의 환자 4명에서 공통적으로 발견되는 것을 선별하면 최종적으로 질병의 원인 유전변이형 하나만이 선별된다는 사실은 아주 작은 수의 단일유전자질환 환자의 유전체 시료를 이용한 엑솜 염기서열 분석만으로도 질병의

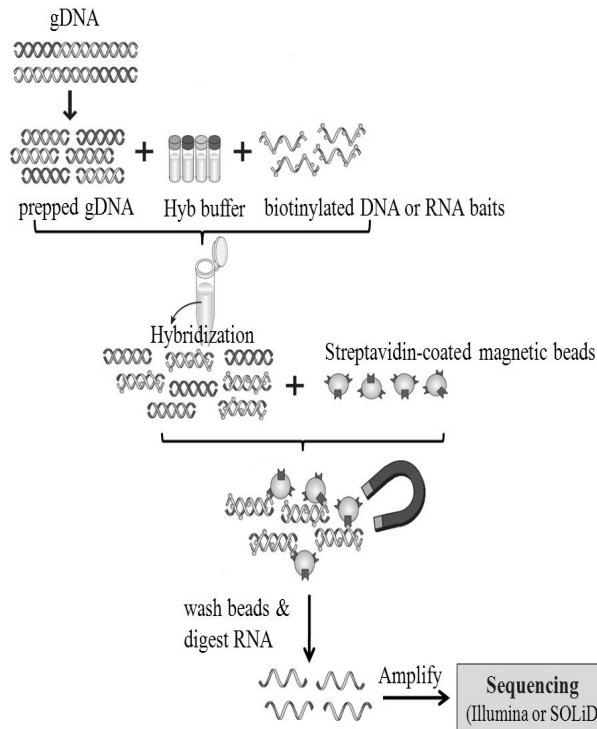


Fig. 3. The principles of exome capture. This figure was modified from an Agilent brochure (<http://www.genomics.agilent.com/>).

Table 3. Filtering Strategy for the Identification of Causal Variants in Mendelian Disorders Using Exome Sequencing

	Ind. 1	Ind. 1 Ind. 2	Ind. 1 Ind. 2 Ind. 3	Ind. 1 Ind. 2 Ind. 3 Ind. 4	Any 3 of 4: Ind. 1 Ind. 2 Ind. 3 Ind. 4
cSNPs	~17,000				
NS/SS/I	~ 4,500	~ 3,300	~ 2,800	~2,500	~ 3800
NS/SS/I not in dbSNP	~ 500	~ 130	~ 70	~ 50	~ 120
NS/SS/I not in controls	~ 800	~ 170	~ 50	~ 20	~ 160
NS/SS/I Neither in dbSNP nor control exomes	~ 350	~ 40	~ 8	1	~ 20
and Predicted to be damaging	~ 160	~ 10	~ 2	1	~3

Abbreviations: SNP, single nucleotide polymorphism; cSNP, coding SNP; NS, nonsynonymous cSNP; SS, splice-site SNP; I, coding Indel

원인 유전자 발굴이 매우 용이함을 증명한다(Table 3). 따라서 단일유전자질환의 원인 유전자 발굴을 위해서 사용되는 엑솜 염기서열 분석 방법은 적은 수의 환자 시료에서 엑솜 염기서열 분석을 수행하고 기존에 dbSNP에 보고된 common SNP과 대조군 시료에서 엑솜 염기서열 분석으로 발견된 유전변이형 정보를 제외하면 쉽게 단일유전자질환의 원인 유전자 발굴이 가능하게 된다.

단일유전자질환에서 엑솜 염기서열 분석으로 발굴된 원인 유전자 정보

지금까지 엑솜 염기서열 분석 방법에 의하여 단일유전자질환의 원인 유전자가 약 10여개 발굴되었다(Table 4). 엑솜 염기서열 분석에 사용된 엑솜 분리 방법은 주로 Agilent 또는 NimbleGen 제품이 사용되고 있다. 염기서열 분석은 비용이 많이 들고 생산된 염기서열 정보가 적은 Roche 454는 유용하지 못하며, 작은 크기의 엑손을 대량으로 분석하기에 적합한 Illumina 회사의 Genome analyzer 또는 Life Technologies 회사의 SOLiD 방법이 주로 사용된다. 그리고 유전변이형 발굴을 위한 프로그램도 연구그룹에 따라 다양한 종류가 사용되고 있음을 확인할 수 있다.

현재까지 엑솜 염기서열 분석에 의해 발굴된 연구들의 내용을 종합해 보면 다음과 같은 내용을 파악할 수 있다. 엑솜 염기서열 분석을 통해 발굴된 질병의 원인 유전자는 모두 유전모형이 밝혀진 질병으로써 1-10개 정도의 질병 환자 유전체 시료가 엑솜 염기서열 분석에 사용되었다. 염기서열 분석 이후에 유전모형과 돌연변이형의 정보를 이용하여 선별된 후보 유전변이형들은 다른 독립된 많은 질병군에서 검정 작업을 거치게 된다. 이러한 내용을 바탕으로 단일유전자질환의 원인 유전자 발

굴은 유전적으로 동일한 배경을 가진 환자 시료를 우선적으로 선별하여 엑솜 염기서열 분석을 수행하여 유전모형과 돌연변이형의 정보를 이용함으로써 소수의 질병 원인 유전변이형 정보를 선별하도록 한다. 선별된 질병 원인 후보 유전변이형은 다수의 독립된 질병시료에서 검정하는 과정을 거치는 것이 필수적이다(Fig. 4).

한편 엑솜 염기서열 분석에 사용되는 유전체 시료는 열성(recessive)의 유전모형을 따르는 경우에는 분석에 사용되는 시료의 수도 우성(dominant)의 유전모형을 따르는 질병의 시료보다는 적게 사용하더라도 질병의 원인 후보 유전자 선별이 가능한 것으로 밝혀졌다. 일반적으로 열성 유전을 따르는 질병은 한 개인에서 엑솜 염기서열 분석에 의해 발굴된 질병 원인 후보 유전자는 수 개로 제한적이지만, 우성의 유전을 따르는 경우에는 한 개인의 엑솜 염기서열 분석에 의하여 수 백 개의 후보 유전자가 존재하는 것으로 알려져 있다. 또한 환자의 엑솜 염기서열 분석을 통해 생성된 유전변이형 정보에서 정상 대조군으로 사용하는 SNP (약 5명의 데이터로부터 dbSNP에 보고된

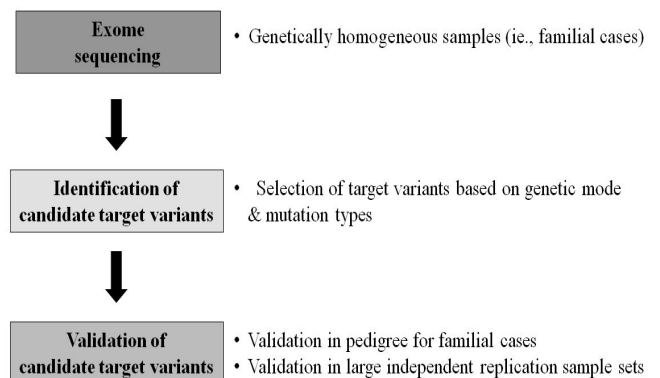


Fig. 4. Strategy for the identification of causal genes and causal variants in mendelian diseases using exome sequencing.

Table 4. Lists of Causal Genes Identified in Mendelian Disorders According to Exome Sequencing

Target disease*	Genetic mode	Causal gene	No. of samples			Methods for Exome seq				Reference
			Seq.	Val. (case/ctl)	mutation sites (#)	Capture	Seq	Fold-X (build#)	Alignment (dbSNP#)	
Freeman Sheldon synd.	autosomal dominant	<i>MYH3</i>	4+8(ctl)	- (known)	2	Agilent	GAll	51X (hg18)	Eland&Maq (build 129)	Ng et al. ³⁾
Miller synd.	autosomal recessive	<i>DHODH</i>	4+8(ctl)	3fam/100	11	Agilent	GAll	40X (hg18)	Eland&Maq (build 129)	Ng et al. ⁴⁾
Bartter synd.	congenital (germline)	<i>SLC26A3</i>	5	39/95	5	NimbleGen	GA	44X (hg18)	Maq (build130)	Choi et al. ²⁾
Kabuki synd.	autosomal dominant	<i>MLL2</i>	10	43/95	33	Agilent	GAll	40X (hg18)	Eland&Maq (db129)	Ng et al. ⁵⁾
Sensenbrenner synd.	autosomal recessive	<i>WDR35</i>	2	6/105	4	Agilent	SOLiD	63X (hg18)	BioScope (build130)	Gilissen et al. ⁶⁾
HPMR (=Mabry) synd.	autosomal recessive	<i>PIGV</i>	3 sblings	3fam/200	4	Agilent	SOLiD	46X (hg18)	iMAP (db130)	Krawitz et al. ⁷⁾
Schinzel-Giedion synd.	Autosomal dominant	<i>SETBP1</i>	4	9/94	3	Agilent	SOLiD	43X (hg18)	BioScope (build130)	Hoischen et al. ⁸⁾
Severe brain malformation	autosomal recessive	<i>WDR62</i>	2 (homo map) → 1	30/1290	6	NimbleGen	GAllx	44X (hg18)	Maq &BWA (build130)	Bilguver et al. ⁹⁾
Cleft palate (=TARP synd.)	X-linked	<i>RBM10</i>	X-linked → 2	2 families	2	Agilent	GAll	110X (hg18)	28Mb (~200 genes)	Johnston et al. ¹⁰⁾
Retinal-renal ciliopathy	autosomal recessive	<i>SDCCAG8</i>	(homo map) → 5	11 fam/270	12	NimbleGen	GAll	200X (hg18)	CLC (build130)	Ott et al. ¹¹⁾
Charcot-Marie-Toot	autosomal recessive	<i>SH3TC2</i>	1fam → 1	1 fam	2	Whole genome	SOLiD	30X	(build126)	Lupski et al. ¹²⁾

Abbreviations: synd, syndrome

common SNP)을 제외할 경우, 개체 특이적인 유전변이형은 매우 작은 숫자만이 남는다. 따라서 대조군으로 사용하는 시료의 숫자는 5개 이상 정도면 단일유전자질환의 원인 유전자 발굴에 사용하기에는 충분할 것으로 사료된다. 한편 엑솜 염기서열 분석에 사용되는 시료가 한 가계에서 분리된 시료라면 동일한 유전변이형에 의하여 질병이 발생할 수 있지만 다른 가계에서 분리된 혈연관계가 없는 질병 시료라면 다른 유전변이형이 동일한 종류의 질병 원인으로 작용할 수 있기 때문에 이러한 경우에는 한 유전자 내에서 다수의 유전변이형이 동일한 질병을 일으킬 수 있는 것을 가정하고 분석하는 것이 필요하다.

엑솜 염기서열 분석 방법의 한계 및 향후 발전 방향

엑솜 염기서열 분석 방법은 단일유전자질환의 원인 유전자 발굴을 위한 매우 효율적인 방법이지만 다음과 같은 몇 가지 제

한성을 가지고 있다. 첫 번째는 최근 출시된 엑솜 분리 키트는 전체 단백질을 코딩하는 영역이 좀 더 확대되었지만 일반적으로 지금까지 발표된 연구 결과에 사용된 엑솜 분리는 전체 대상 엑손의 약 75%만을 분리함으로써 약 25%의 단백질 코딩 영역에 대한 염기서열 분석은 불가능한 한계를 가지고 있다¹⁾. 또한 엑솜만을 검색하는 방법은 일부의 질병에서 원인으로 작용하는 유전체의 구조변화와 유전자 발현을 조절하는 유전자 조절 부위에 대한 정보는 확인할 수 없다. 그러나 향후에는 유전체 염기서열 분석 기술의 발전으로 모든 유전체 전체에 대한 염기서열 분석이 가능하게 되면 이러한 문제점들은 해결될 것이다. 한편 지금까지는 주로 단일유전자질환의 원인 유전자를 발굴하기 위하여 엑솜 염기서열 분석 방법이 이용되었다. 그러나 최근에는 특정 표현형(콜레스테롤 농도 등)에서 극단적으로 낮은 값을 가진 개체의 가계도 시료에서 엑솜 염기서열 분석 방법을 사용하여 질병 원인 유전자를 발굴함으로써 복합형질

(complex traits)의 원인 유전자 발굴도 가능함을 증명하였다¹³⁾. 따라서 앞으로는 다양한 유전자와 환경에 의하여 결정되는 복합질병의 원인 유전자 발굴에도 엑솜 염기서열 분석 방법이 매우 유용하게 활용될 것으로 기대된다. 또한 개인 유전체 염기서열 분석 방법은 다양한 질병의 원인 유전변이형 정보를 탐색할 수 있는 의학적 진단을 혁신적으로 발전시킬 것이다. 그리고 궁극적으로는 개인 유전체 염기서열 분석 기술을 통해 개인의 질병을 조기에 진단하고 예방하기 위한 기술 개발로 이어져 개인 맞춤의학의 실현을 앞당기는데 크게 기여할 것이다.

감사의 글

본 연구는 한국연구재단의 2010년도 소규모연구회지원사업의 과제(E00044)로부터 지원을 받아 수행됨.

국문초록

약 7,000 여개의 단일유전자질환이 보고되어 있지만 보고된 질환의 절반도 아직 원인 유전자가 밝혀지지 못한 상황이다. 그리고 기존에 밝혀진 원인 유전자의 돌연변이형들은 대부분 단백질을 코딩하는 부위의 돌연변이에 의하여 발생하고 있어서 인간 유전체에서 단백질을 코딩하는 엑손 부위만을 선별적으로 분리하여 염기서열을 분석하는 엑솜 염기서열 분석 방법은 희귀한 유전질환의 신규 원인 유전자 발굴을 위한 매우 효과적인 유전 분석법이 될 것이다. 엑솜은 전체 유전체의 약 1.5% 정도를 차지하고 있어서 매우 경제적으로 분석이 가능하다. 그리고 엑솜 염기서열 분석 방법은 엑솜 부위를 선별하는 기술과 대용량 염기서열 분석기술로 수행된다. Freeman-Sheldon 증후군의 원인 유전자를 엑솜 염기서열 분석 방법으로 발굴한 이후로 단일유전자질환의 원인 유전자 발굴을 위한 표준 분석 방법으로 엑솜 염기서열 분석 방법이 사용되고 있다. 향후에는 엑솜 염기서열 분석 방법이 다양한 복합질병의 유전분석에도 활용되어 개인 맞춤의학의 실현을 앞당기는데 크게 기여할 것으로 기대된다.

References

1) Kaiser J. Affordable 'exomes' fill gaps in a catalog of rare diseases. *Science* 2010;330:903.

- 2) Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci USA* 2009;106:19096-101.
- 3) Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 2009;461:272-6.
- 4) Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, et al. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* 2009;42:30-5.
- 5) Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* 2010;42:790-3.
- 6) Gilissen C, Arts HH, Hoischen A, Spruijt L, Mans DA, Arts P, et al. Exome sequencing identifies WDR35 variants involved in Sensenbrenner syndrome. *Am J Hum Genet* 2010;87:418-23.
- 7) Krawitz PM, Schweiger MR, Rodelsperger C, Marcelis C, Kolsch U, Meisel C, et al. Identity-by-descent filtering of exome sequence data identifies PIGV mutations in hyperphosphatasia mental retardation syndrome. *Nat Genet* 2010;42:827-9.
- 8) Hoischen A, van Bon BWM, Gilissen C, Arts P, van Lier B, Steehouwer M, et al. De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nat Genet* 2010;42:483-5.
- 9) Bilguvar K, Oztürk AK, Louvi A, Kwan KY, Choi M, Tatli B, et al. Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. *Nature* 2010;467:207-10.
- 10) Johnston JJ, Teer JK, Cherukuri PF, Hansen NF, Loftus SK, NIH Intramural sequencing center, et al. Massively parallel sequencing of exons on the X chromosome identifies RBM10 as the gene that causes a syndromic form of cleft palate. *Am J Hum Genet* 2010;86:743-8.
- 11) Ott EA, Hurd TW, Airik R, Chaki M, Zhou W, Stoetzel C, et al. Candidate exome capture identifies mutation of SDCCAG8 as the cause of a retinal-renal ciliopathy. *Nat Genet* 2010;42:840-50.
- 12) Lupski JR, Reid JG, Gonzaga-Jauregui C, Deiros DR, Chen DCY, Nazareth L, et al. Whole-genome sequencing in a patient with charcot-marie-tooth neuropathy. *N Engl J Med* 2010;362:1181-91.
- 13) Musunuru K, Pirruccello JP, Do R, Peloso GM, Guiducci C, Sougnez C, et al. Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia. *N Engl J Med* 2010;363:2220-7.