

안과 영역에서의 차세대 염기서열 분석

Next-generation Sequencing in Inherited Eye Diseases

한진우

Jinu Han, MD

연세대학교 의과대학 강남세브란스병원 안과학교실 시기능 개발연구소

Institute of Vision Research, Department of Ophthalmology, Gangnam Severance Hospital, Yonsei University of College of Medicine, Seoul, Korea

Next-generation sequencing is widely used in inherited diseases and cancer genetics fields. Next-generation sequencing technology provides accurate diagnosis in genetically heterogeneous disorders such as retinitis pigmentosa, Leber congenital amaurosis, or cone-rod dystrophy. However, the precise interpretation of variants produced by massively parallel sequencing is somewhat difficult to most of ophthalmologists, and misinterpretation of these variants lead to unwanted devastating consequences to the patients and their family. The molecular genetic findings need to be carefully evaluated in the context of the clinical findings to avoid misdiagnosis. Gene therapy trials are already in the market for specific forms of Leber congenital amaurosis. We are in the middle of exiting era of effective treatment for patients with inherited eye diseases, which was considered as incurable in the past. To success such a treatment, molecular diagnosis will become essential.

Ann Optom Contact Lens 2018;17(4):89-96

Key Words: Inherited eye disease; Inherited retinal disease; Next-generation sequencing

서 론

1977년대 Sanger et al¹에 의해 개발된 DNA 염기서열 분석 기술은 이후 2007년경 차세대 염기서열 분석(next-generation sequencing)으로 이어져 현대의학 발전에 많은 기여를 하고 있다. 차세대 염기서열 분석은 병렬적(massively parallel)으로 십억 이상의 리드(reads, 염기서열 단편)를 처리하여, 염기서열 분석에 걸리는 시간 및 비용을 획기적으로 줄여주었으며, 이에 따라 유전체 연구의 효과적인 방법

으로 자리 잡고 있다. 이러한 DNA 염기서열 분석 기술이 단백질을 전사하는 부위인 exon과 exon-intron의 경계만 염기서열 분석하는 exome sequencing, 혹은 intron을 포함한 전체 염기서열을 분석하는 genome sequencing이 상대적으로 적은 비용에 가능하게 되었으며 종양의 조직에서 돌연변이를 찾는 종양 유전체학 그리고 유전적으로 다양한 유전자에 의해 질환이 발생하는(genetically heterogeneous) 유전질환의 진단 등에 현재 이용이 되고 있다. 안과적인 질환은 한 개의 유전자가 그 돌연변이의 위치에 따라 다양한 표현형을 나타내기도 하고(Fig. 1), 여러 개의 서로 다른 유전자의 변이가 유사한 표현형을 보이는 경우도 많다. 안과의 분야에서 Leber congenital amaurosis, rod-cone dystrophy, congenital cataract 등의 유전질환에서 원인 돌연변이를 찾는 데 이용되고 있으며, 이렇게 유전학적으로 여러 개의 유전자가 병을 일으키는 유전질환에서 진단율을 획기적으로 높였으며, 현재 여러 유전성 안질환에서의 진단율은 약 50-75% 정도로 보고가 되고 있다.²⁻⁴ 또한, uveal melanoma

■ Received: 2018. 12. 5. ■ Revised: 2018. 12. 18.

■ Accepted: 2018. 12. 18.

■ Address reprint requests to **Jinu Han, MD**
Institute of Vision Research, Department of Ophthalmology,
Gangnam Severance Hospital, #211 Eonju-ro, Gangnam-gu,
Seoul 06273, Korea
Tel: 82-2-2019-3445, Fax: 82-2-3463-1049
E-mail: jinuhan@yuhs.ac

* This research was supported by a fund (2018-ER6902-00) by
Research of Korea Centers for Disease Control and Prevention.

Copyright © 2018, The Korean Optometry Society
The Korean Contact Lens Study Society

© Annals of Optometry and Contact Lens is an Open Access Journal. All articles are 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.

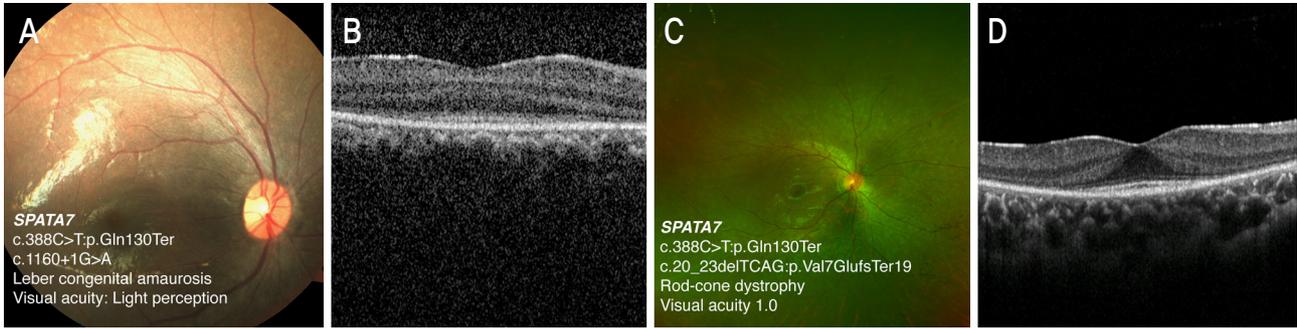


Figure 1. Leber congenital amaurosis and rod-cone dystrophy phenotypes according to different *SPATA7* mutations. (A, B) Fundus photograph showed marbled fundus, and optical coherence tomography revealed degeneration of photoreceptor layer with minimal remaining at the fovea. (C, D) Wide fundus photographs showed pigmentary retinopathy and optical coherence tomography revealed that photoreceptor layer was remained in the central fovea.

와 같은 안종양학에도 차세대 염기서열 분석이 이용이 되고 있으며,⁵ 최근에는 ribonucleic acid (RNA)-seq과 single cell sequencing 기법으로 cell-level 단위에서의 expression을 볼 수 있는 방법이 등장하여 질병의 매커니즘(mechanism) 규명에 다양하게 응용이 되고 있다.⁶ 2015년 미국에서 Precision Medicine Initiative 프로젝트가 시작되면서 백만 명의 환자들의 유전체 정보를 얻어 환자 개개인에 맞는 맞춤 치료를 제공하는 것을 목표로 하고 있으며, 그에 따라 환자 개개인별로 그에 해당되는 임상검사와 적절한 약이 제공되는 것이 요구되는 시대일 것으로 생각된다.^{7,8} 그러므로 차세대 염기서열 분석은 현대의학의 필수적인 부분으로 자리잡고 있으나 아직 이 부분에 대한 안과 의사들의 이해가 부족한 실정이다. 따라서, 차세대 염기서열 분석에 대한 기초적인 이해가 필요할 것으로 생각되며, next generation sequencing (NGS) 방법의 장단점과 한계, 변이 해석에 있어 유의할 점이 어떠한 것이 있는지 자세하게 알아볼 필요가 있을 것이다. 본고에서는 NGS 데이터가 원시 데이터에서 시작하여 유전변이 추출에 이르는 단계들의 기본적인 개념을 소개하고, 안과질환에서의 적용, NGS 방법의 한계점 및 변이의 해석을 어떻게 할 것인지, 안과질환에서 어떻게 응용을 할 수 있을지에 대하여 살펴보고자 한다.

Sequencing 장비의 선택

차세대 염기서열 분석의 업계 선두는 단연 illumina이다. MiSeq, NextSeq550, HiSeq 등을 비롯하여 최근에 출시가 된 NovaSeq으로 염기서열 분석 비용이 획기적으로 낮아졌다. Targeted panel sequencing, whole exome 혹은 whole genome에 따라 장비를 선택해서 염기서열 분석을 하면 가장 낮은 비용으로 원하는 throughput을 얻을 수 있다. 대개 약 1,000개 미만의 원하는 유전자를 구성하여 염기서열 분석

하는 targeted panel 염기서열 분석은 임상적으로 현재 적용이 가능한 분야이며, retinitis pigmentosa, congenital hearing loss, Charcot-Marie-Tooth병에서 보험 적용이 되고 있으며, benchtop 시퀀서인 MiSeq, NextSeq550으로 가능하다. 하지만 whole exome 혹은 whole genome 염기서열 분석은 생산되는 정보의 양이 많으므로 대개는 HiSeq 혹은 NovaSeq을 이용하여 염기서열 분석을 하게 된다. MiSeq과 HiSeq은 4-Dye system으로 네 종류의 염기마다 서로 다른 형광을 사용하지만 NextSeq550과 NovaSeq은 C (cytosine)은 빨강, T (thymine)은 초록, A (adenine)는 두 색깔의 혼합 그리고 G (guanine)는 형광이 없는 형태로 표지한다. 사용되는 형광의 개수를 절반으로 줄임으로서 염기서열 분석 원가를 절감할 수 있으며, 데이터의 quality도 기존의 제품과 유사하다고 보고되고 있다. Long reads 시퀀서인 PacBio와 Oxford Nano는 아직까지는 error rate가 높다고 여겨져 환자의 임상적인 유전 진단에 널리 사용되고 있지는 않다. 따라서, 각각 연구나 임상 목적에 맞는 염기서열 분석 기계를 선택하면 될 것으로 생각된다(Table 1).

NGS 데이터의 data pre-processing 및 SNP/INDEL calling

NGS의 여러 데이터를 분석하고 논문을 이해하기 위해서는 여러 가지 용어들을 우선 알아둘 필요가 있다(Table 2). NGS 염기서열 분석 방법을 간단히 설명하면 환자에서 추출된 DNA는 초음파 분쇄기에 의해 300-700 bp의 작은 조각들로 잘라지며, 분쇄된 DNA를 adapter sequence에 붙여 라이브러리를 생성한다. 이 adapter에 붙은 작은 DNA 조각들이 flow cell에서 클러스터를 형성하며 증폭 과정을 통하여 보통 100-150개 정도의 염기로 구성된 짧은 서열 조각인 리드(reads)를 생성하게 되며 이러한 리드들의 정보

Table 1. Technical specification of sequencing platform

Sequencing platform	Total output (bases per run)	Total reads (million per run)	Read length (bases)	Run time (days)	Purpose	
Illumina	MiSeq	300 Mb -15 Gb	50 M	2 × 300 bp	0.2-2.7	Targeted panel or amplicon sequencing
	NextSeq 550	25-120 Gb	260 M-800 M	2 × 75 bp	1.5-3	Midthroughput to highthroughput versatile use
	HiSeq X	1.6-1.8 Tb	6,000 M	2 × 150 bp	<3	Only for large scale whole genome sequencing
	NovaSeq	0.2-3 Tb	1,300-20,000 M	2 × 150 bp	<2	Scalable platform, flexible performance
ThermoFisher	Ion Torrent	0.3-50 Gb	3-130 M	2 × 200 bp	0.2-1	Diverse sequencing data with as little as 10ng sample
PacBio	Sequel system	500 Mb-16 Gb	55-880 M	Up to 60 Kb	<0.1-0.3	Useful in the studies of de novo assembly Long reads sequencing
Nanopore	MinION	Up to 42 Gb	Up to 4.4 M	230-300 kb	2	Portable sequencing instrument

Mb = mega base; Gb = giga base; M = million; bp = base pair; Tb = tera base; Kb = kilo base.

Table 2. Terms used in next-generation sequencing

Term	Explanations
BAM	SAM의 binary version인 BAM 파일이며, SAM이 아닌 BAM 파일로 저장하면 용량을 줄일 수 있음
Coverage	Sample DNA에서 평균 혹은 각각의 base에 align 되는 중복되지 않은 다른 리드의 숫자(혹은 평균 갯수)
FASTA/FASTQ	FASTA 란 nucleic acid sequence나 protein sequence를 저장하는 format을 말하며, FASTQ는 nucleotide sequence 정보와 그에 해당하는 quality score가 같이 있는 저장파일을 말함
GATK	Genome Analysis ToolKit의 약자로 broad institute에서 제공하는 SNP/indel calling 시 사용하는 표준 bioinformatics analysis tool임
SAM	Sequence alignment/map 포맷으로서 nucleotide sequence 정보와 reference sequence에 정렬되어 있는 정보가 함께 있는 파일 포맷을 말함
Throughput	염기서열 분석 머신으로 생산되는 base의 양을 말하며 데이터의 크기를 뜻함(예, 7Gb으로 whole exome sequencing 정보를 생산했다고 하면, 7 Giga base의 정보가 생산되었다는 뜻)
VCF	Variant call format의 약자로 해당 샘플에서 reference genome에 대비하여 어떤 변이 정보를 가지고 있는지 나타내는 파일 포맷

BAM = binary alignment map; SAM = sequence alignment map; GATK = Genome Analysis Toolkit ; SNP = single nucleotide polymorphism; VCF = variant caller format.

가 있는 파일을 FASTQ 형식의 파일로 저장하게 된다. 수많은 DNA 단편 조각인 리드들을 burrows-wheeler aligner를 통하여 참조 유전체(reference genome)에 리드를 정렬(alignment/mapping)하게 되고,⁹ 중복되는 리드를 표시하는 작업을 거치고 (mark duplicate), mismatch되는 부분의 indel realignment와 base quality score에 따라 recalibration 과정을 거쳐 분석에 이용될 수 있는 BAM 파일을 생성하게 된다. 분석에 이용될 수 있는 BAM 파일에서 reference genome과는 다른 흔하지 않은 변이가 있는 부분을 베이지안 분석과 local de novo assembly 등의 과정을 거치는 haplotypcaller를 통하여 변이가 존재하는 부분에 대하여 보다 정확한 통계 방법으로 genotype에 대한 정보를 생산한다.¹⁰ 이러한 분석 과정은 Genome Analysis ToolKit의 best practice가 현재 표준

분석 방법으로 쓰이고 있다.¹⁰ 또한, 이렇게 생성된 variant call format file을 여러 명의 sample들을 합쳐서 분석함으로써 false-positive와 염기서열 분석 에러 부분들을 필터링하는 후처리 과정을 거쳐 분석한 유전체에서 어떠한 부분에 변이가 있으며 이 부분의 변이가 heterozygote인지 homozygote 상태인지 결정이 된다.

Variant annotation/filtering

이렇게 여러 가지의 처리 과정을 거쳐 나온 변이들이 어떤 기능을 하는 변이들인지 표시하는 과정을 variant annotation이라고 한다. 한 사람의 exome sequencing 결과로 약 2-3만 개의 변이가 나오게 되는데 이러한 변이들이 synon-

ymous, missense, nonsense, splice site, 혹은 intronic 변이인지 표지를 하고, 일반인 집단에서 이러한 변이가 흔하게 발견되는 변이인지 드문 변이인지 minor allele frequency를 통하여 확인한다. 1000 Genomes Project, ESP6500, ExAC, gnomAD 등의 많은 대규모의 유전체 연구에서 일반인 집단에서의 변이의 minor allele frequency를 확인할 수 있으며, 특히 이런 minor allele frequency를 확인할 때 같은 인종에서 해당 변이의 minor allele frequency를 확인하는 것이 중요하다(Fig. 2). 예를 들어 어떠한 변이를 관찰하였는데 해당 질환이 매우 드문 질환이고 우성유전하는 질환이라면 정상 집단에서 해당 변이가 없거나 아주 드문 경우여

야 할 것이다. 여기서 고려되어야 할 부분이 병의 평균적인 발현 시기, 변이를 가지고 있을 때 얼마나 표현형이 나타나는지(penetrance), 해당 질환이 얼마나 드문 병인지 등을 종합적으로 고려하여 판단한다. 흔히 쓰는 database로 gnomAD browser와 ExAC Browser가 있으며, 한국인 1,722명의 유전체에 존재하는 한국인 참조 유전체 데이터베이스 정보가 공개되어 있다(Table 3). 또한, 이러한 변이가 여러 다른 종간에서 잘 보존되어 있는지, 그리고 변이가 단백질의 기능에 어떠한 영향을 미치는지 예측하는 여러 가지 도구들로 표지를 하게 된다(Table 4). PolyPhen2와 SIFT 스코어는 missense 변이만 prediction할 수 있으며, 최근에는 functional

GnomAD browser

Variant: 1-10042628-C-T: *NMNAT1* c.709C>T

Population Frequencies				
Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
Overall	4	19952	0	0.0002005
Korean	4	3818	0	0.001048
Japanese	0	152	0	0.000
East Asian	0	14424	0	0.000
Other East Asian	0	14424	0	0.000
Female	3	9866	0	0.0003041
Male	1	10086	0	0.00009915
South Asian	5	30616	0	0.0001633
European (non-Finnish)	4	129102	0	0.00003098
Latino	1	35438	0	0.00002822
African	0	24962	0	0.000
Ashkenazi Jewish	0	10366	0	0.000
European (Finnish)	0	25122	0	0.000
Other	0	7222	0	0.000
Total	14	282780	0	0.00004951

Figure 2. Minor allele frequency of *NMNAT1* c.709C>T variant in genome aggregation dataset (gnomAD) browser. The web browser displayed minor allele frequency in different ethnic population.

Table 3. Population, disease-specific, and sequence databases

Type of database	Name of database	Website
Population database	Exome aggregation consortium	http://exac.broadinstitute.org/
	GnomAD browser	http://gnomad.broadinstitute.org/
	Exome server project	http://evs.gs.washington.edu/EVS/
	Korean reference Genome database	http://152.99.75.168/KRGDB
Inherited disease database	ClinVar	https://www.ncbi.nlm.nih.gov/clinvar/
	HGMD Professional	https://portal.biobase-international.com/cgi-bin/portal/login.cgi?redirect_url=/hgmd/pro/start.php?
	Leiden open variation database	http://www.lovd.nl/3.0/home
	OMIM	http://www.omim.org

HGMD = human gene mutation database; OMIM = online mendelian inheritance in man.

analysis through hidden Markov models (FATHMM) 혹은 combined annotation dependent depletion (CADD) score가 missense 변이 외의 nonsense 변이와 regulatory 변이 등도 모두 scoring이 가능하여 pathogenicity를 판단할 때 유용하게 쓰인다.^{11,12} 하지만 이러한 *in silico* (=computational) 도구들에 의한 변이의 예측은 여러 가지의 도구를 쓰지만, 그 자체로 강력한 pathogenicity의 증거가 되는 것은 아니며 단지 하나의 요소로 참조하여 결정한다.

Interpretation of variants

American College of Medical Genetics Standard and Guideline에서는 변이들을 분류하고 해석할 때 5가지의 pathogenic, likely pathogenic, uncertain significance, likely benign, benign으로 분류하는 가이드라인을 제시하고 있다.¹³ 이 가이드라인은 변이 해석에 있어 표준지침으로 자리잡고 있으며, NGS 변이를 해석하고 리포트를 받는 임상자들은 반드시 알아야 할 부분이며, 해당 논문을 정독하여 숙지할 필요가 있다. 어떤 변이가 있을 때 해당 질환과 연관되어 있다고 보고가 된 변이인지 null variant인지, 인구 집단에서 얼마나 흔하게 관찰되는 변이인지 등에 따라 해당 변이가 큰 의미가 없는 benign으로 분류를 할 것인지 pathogenicity를 가지는지 점수를 매기게 되며 각각의 변이를 pathogenic인지 아닌지 판단하는데 많은 시간이 걸리게 된다. InterVar라는 프로그램으로 이런 변이들에 대하여 자동으로 classification을 해주기도 하나 결국 수작업으로 확인해야 되는 부분이

존재한다(<http://wintervar.wglab.org>).¹⁴ 이러한 방법으로도 많은 변이들이 uncertain significance로 남게 되며 이러한 변이들을 임상적으로 해석하는데 주의를 요하게 되며, 몇몇의 경우 변이의 잘못된 해석으로 법정 소송까지 이어지는 경우도 있음을 알아야 한다.^{15,16} 또한, NGS 방법으로는 상염색체 열성유전하는 경우 2개의 변이가 cis로 존재하는지 trans로 존재하는지 확인할 수 없다. 왜냐하면 대부분의 2개의 변이가 서로 다른 exon에 존재하는 경우가 많고 NGS는 100-150 bp의 짧은 염기서열의 조각들을 단편으로 환자의 genotype을 추정하는 것이기 때문이다. 따라서, 많은 경우에서 가족검사를 하여 유사한 표현형을 가진 가족에서 같은 변이가 발견되는지 혹은 표현형이 없는 가족에서 병적이라고 간주했던 변이가 나오지 않는지 확인하여야 하며, 상염색체 열성 유전하는 질환의 경우 부모나 자녀 검사로 두 개의 변이가 cis로 존재하는지 trans로 존재하는지 확인을 해야 된다. 드물게 같은 exon에 100-150 bp 내에 2개의 변이가 존재하는 경우도 있는데 이러한 경우 integrative genomics viewer로 확인하면 서로 다른 리드에 두 개의 변이가 있는지 확인하여 상염색체 열성유전하는 질환에서 2개의 변이가 cis로 존재하는지 trans로 존재하는지 확인할 수 있다(Fig. 3).¹⁵

자세한 표현형의 기록 및 조사

유전성 안질환의 유전자검사를 함에 앞서 환자의 증상 및 다른 전신적인 동반 질환에 대한 조사를 철저히 하는 것

Table 4. In silico prediction algorithms

Category	Name	Website	Basis
Missense prediction	SIFT	http://sift.jcvi.org	Evolutionary conservation
	PolyPhen-2	http://genetics.bwh.harvard.edu/pph2	Protein structure/function and evolutionary conservation
Any substitution	FATHMM	http://fathmm.biocompute.org.uk	Evolutionary conservation
Indel	FATHMM-Indel	http://indels.biocompute.org.uk	
All including indel	CADD	http://cadd.gs.washington.edu	Contrasts annotations of fixed/nearly fixed derived alleles in humans with simulated variants
Splice site prediction	Human Splicing Finder	http://www.umd.be/HSF/	Position-dependent logic
	MaxEntScan	http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html	Maximum entropy principle
	NNSplice	http://www.fruitfly.org/seq_tools/splice.html	Neural networks
Nucleotide conservation prediction	GERP	http://mendel.stanford.edu/sidowlab/downloads/gerp/index.html	Genomic evolutionary rate profiling
	PhyloP	http://compgen.bscb.cornell.edu/phast/	Conservation scoring and identification of conserved elements

SIFT = sorting intolerant from tolerant; FATHMM = functional analysis through hidden Markov models; CADD = combined annotation dependent depletion; GERP = genomic evolutionary rate profiling.

이 중요하다.¹⁶ 안과적으로는 망막전위도, 안저사진, 눈떨림 유무, 시야검사, 전안부 이상 등을 철저히 조사해야 되며 동반된 신장 이상이나 청력소실, 발달지연 등이 없는지 확

인하고 해당 과에 의뢰하도록 한다. 물어보지 않으면 환자나 보호자 대부분 자발적으로 이야기하지 않으므로 안과적 증상 외에도 다른 여러 가지 동반된 전신질환들을 확인하

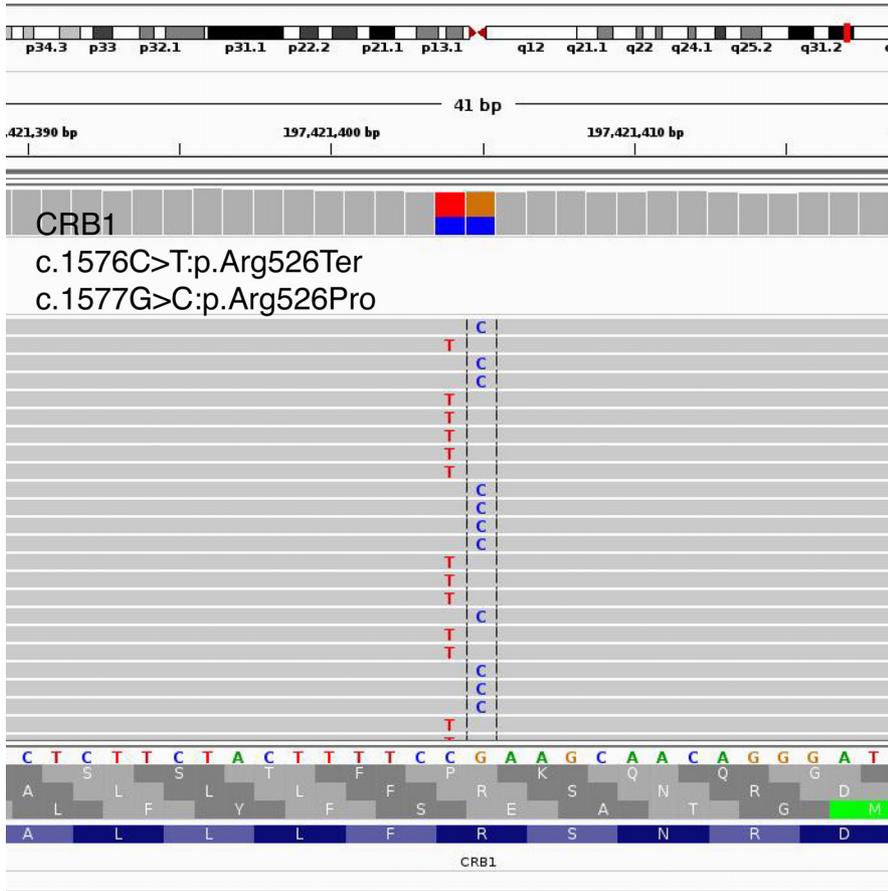


Figure 3. *CRB1* c.1576C>T:p.Arg526Ter and c.1577G>C:p.Arg526Pro variants were presented in different reads, visualized by Integrative genomics viewer. These 2 variants were present in trans, so segregation analysis is not needed. bp = base pair.

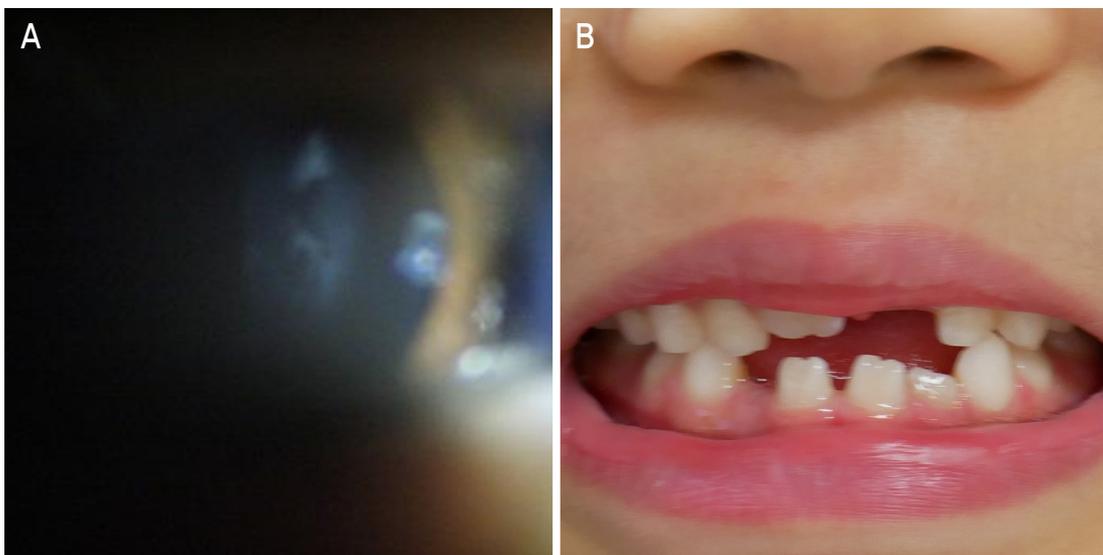


Figure 4. A 2-year-old girl presented at our clinic for bilateral congenital cataract. Targeted next-generation sequencing revealed *NHS* c.1117C>T:p.Arg373Ter nonsense mutation. Nance-Horan syndrome (NHS) is inherited as X-linked dominant trait. (A, B) A mutation in *NHS* gene causes congenital cataract and Hutchinsonian teeth.

는 것이 필수적이다. 왜냐하면 많은 경우에 있어서 안과적 질환 외에도 발달 지연 등의 전신적인 증상이 동반되는 경우가 있으며 이런 전신적인 증상을 잘 확인해 두어야 병적인 변이에 해당되는 표현형이 있는지 확인해 볼 수가 있기 때문이다(Fig. 4). 만약 표현형의 기록이 부족하였으면 유전자검사 결과가 나온 후 다시 철저하게 의심되는 유전 변이에 해당되는 표현형을 확인하도록 한다.

안과에서의 적용

가장 흔하게 안과에서 적용할 수 있는 분야가 retinitis pigmentosa, Leber congenital amaurosis 등의 유전성 망막 질환 분야이다. 진단율은 Leber congenital amaurosis의 경우 약 80% 정도이며,^{17,18} 양안 선천백내장의 경우도 진단율이 약 70-80%로 알려져 있다.³ 양안 선천백내장의 경우 동반된 전신질환에 대한 정보와 식단 조절로 병의 진행을 늦출 수 있는 부분이 있기 때문에 NGS가 환자의 진료에 큰 도움이 될 것으로 생각된다.¹⁹ 영아안진(infantile nystagmus) 환자도 안진을 일으키는 질환이 매우 다양하여 NGS로 분자유전 진단이 선행된다면, 필요한 환자에게서만 자가공명영상검사를 시행할 수 있는 등의 전체적인 진료비를 절감

할 수 있는 기회를 제공할 수 있을 것으로 생각된다.⁴ 또한 선천녹내장과 unexplained optic neuropathy에서도 적용해 볼 수 있을 것이다. 선천녹내장의 경우는 NGS 유전진단율이 약 30-40% 정도로 알려져 있으며, 특징적인 Leber hereditary optic neuropathy나 dominant optic atrophy가 아닌 optic neuropathy의 경우 NGS를 적용하면 atypical Wolfram syndrome이나 Bosch-Boonstra-Schaaf optic atrophy syndrome 등 진단이 어렵고 드문 질환을 정확히 진단할 수도 있을 것으로 생각한다.

NGS의 한계점

NGS의 방법은 대개 100-150 bp의 짧은 염기서열로 이루어진 수많은 리드들에서 정보를 얻으며 polymerase chain reaction (PCR) 증폭 과정을 거친다. 또한 targeted panel 염기서열 분석과 exome sequencing의 경우 exon과 exon-intron boundary만 분석이 되기 때문에 deep intronic 변이와 regulatory 변이는 검출이 안 된다. Hybridization capture와 PCR 증폭 과정을 거치게 되기 때문에 GC-rich region에 대한 coverage가 떨어지며, 반복되는 염기서열이 100-150 bp 이상으로 길게 존재하는 경우에 해당 지역에 대한 mapping

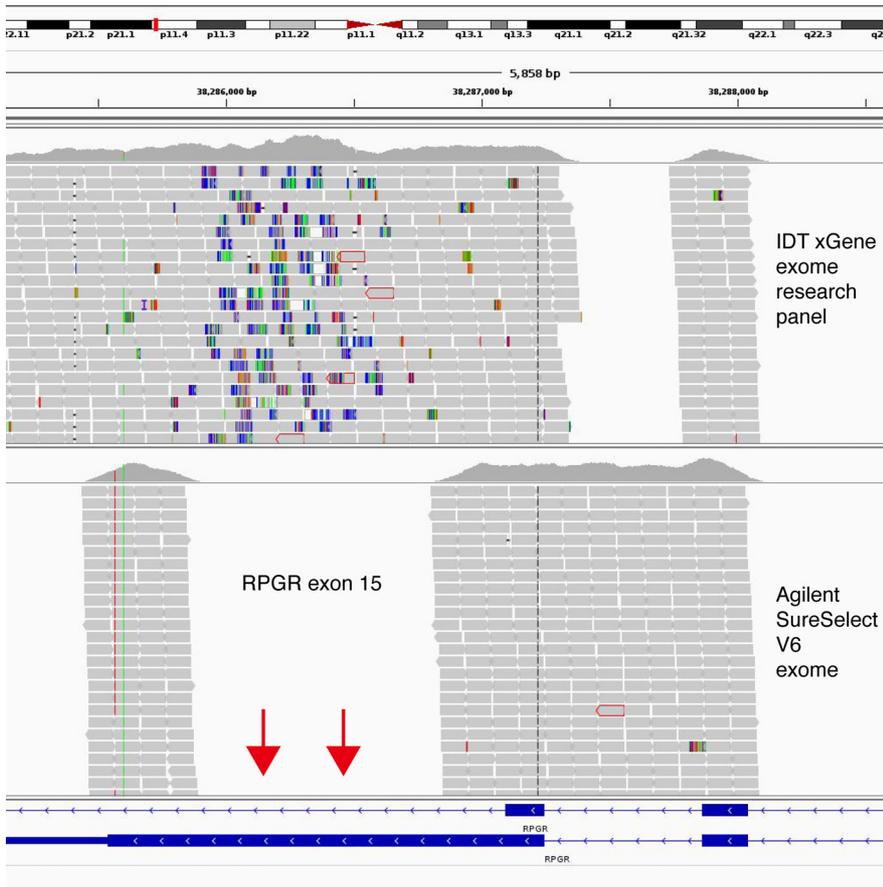


Figure 5. Poor coverage and low quality mapping in *RPGR* exon 15 due to low-complex repetitive sequence. No coverage with SureSelect V6 exome on *RPGR* exon 15 (red arrows). IDT = integrated DNA technologies; bp = base pair.

quality가 떨어지게 된다. 예로 X-linked recessive retinitis pigmentosa의 원인 유전자인 *RPGR* 유전자의 exon 15에 이러한 부위가 존재하는데 이 부위는 mutation hotspot으로 해당 부위의 염기서열 분석은 NGS 방법으로는 정확하게 염기서열 분석하기 어렵다(Fig. 5).²⁰ 또한, segmental duplication이 존재하는 부위도 정확한 염기서열 해독이 어렵다. 또한, 세 개의 염기서열이 반복 증폭되는 triplet expansion이 원인이 되는 Fuch's endothelial dystrophy (CTG expansion in *TCF4* gene), spinocerebellar ataxia type 7 (CAG expansion in *ATXN7* gene)의 경우 NGS로 정확한 진단이 어렵게 된다는 점을 유념에 둘 필요가 있다.^{21,22} 따라서, 이러한 환자에서는 기존의 고식적인 NGS 방법을 이용하면 진단을 놓치게 된다는 사실을 알아야 한다. 마지막으로, 선천성 안질환 및 유전성 안질환의 대부분은 진단이 임상적 검사만으로 충분한 경우가 많고, 유전 진단이 치료로 이어지지 않는 한계점이 있다.

결 론

차세대 염기서열 분석이 상용화된 지 10년이 지났지만 아직도 국내의 안과에서 소수의 임상에만 유전 진단 및 연구에 적용하고 있으며 원시 데이터 분석의 기초에 대하여 무지한 실정이다. 본 리뷰가 안과 임상의가 NGS 데이터와 논문을 접할 때 조금 더 이해가 쉬울 수 있도록 도움이 될 것으로 생각한다. 임상적으로 안과질환에 대한 패넬을 구성할 때는 해당 질환 예를 들어 시신경, 망막, 전안부 등으로 나누어서 구성을 할 수도 있으나 전체를 묶어서 구성하는 것도 좋은 방법이 될 것으로 생각한다. 왜냐하면 하나의 유전자가 다양한 안과적 표현형을 나타내는 경우가 많고, 유전자의 개수가 많아지면 off-target analysis로 여러 다른 copy number variation 등을 같이 확인할 수 있기 때문이다. 유전성 안질환에서 차세대 염기서열 분석은 정확한 유전 진단을 가능하게 하였으나 아직까지는 소수의 환자에서만 치료 방침이 달라지고 대부분의 환자에서는 시력예후, 가족 상담 정도만 할 수 있는 실정이다. 하지만 희귀 질환의 분자유전학적인 매커니즘이 차차 밝혀짐에 따라 많은 환자에게 개개인의 맞춤 치료가 가능하게 될 미래가 올 것으로 기대한다.

REFERENCES

- 1) Sanger F, Air GM, Barrell BG, et al. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* 1977;265:687-95.
- 2) Chaitankar V, Karakulah G, Ratnapriya R, et al. Next generation sequencing technology and genomewide data analysis: perspectives

- for retinal research. *Prog Retin Eye Res* 2016;55:1-31.
- 3) Gillespie RL, O'Sullivan J, Ashworth J, et al. Personalized diagnosis and management of congenital cataract by next-generation sequencing. *Ophthalmology* 2014;121:2124-37.e1-2.
- 4) Rim JH, Lee ST, Gee HY, et al. Accuracy of next-generation sequencing for molecular diagnosis in patients with infantile nystagmus syndrome. *JAMA Ophthalmol* 2017;135:1376-85.
- 5) Luscan A, Just PA, Briand A, et al. Uveal melanoma hepatic metastases mutation spectrum analysis using targeted next-generation sequencing of 400 cancer genes. *Br J Ophthalmol* 2015;99:437-9.
- 6) Grün D, van Oudenaarden A. Design and analysis of single-cell sequencing experiments. *Cell* 2015;163:799-810.
- 7) Sankar PL, Parker LS. The Precision Medicine Initiative's All of Us Research Program: an agenda for research on its ethical, legal, and social issues. *Genet Med* 2017;19:743-50.
- 8) Ashley EA. Towards precision medicine. *Nat Rev Genet* 2016;17:507-22.
- 9) Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754-60.
- 10) McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297-303.
- 11) Shihab HA, Gough J, Cooper DN, et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat* 2013;34:57-65.
- 12) Kircher M, Witten DM, Jain P, et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014;46:310-5.
- 13) Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015;17:405-24.
- 14) Li Q, Wang K. InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. *Am J Hum Genet* 2017;100:267-80.
- 15) Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol* 2011;29:24-6.
- 16) Wiggs JL. Progress in diagnostic genetic testing for inherited eye disease. *JAMA Ophthalmol* 2017;135:1385-6.
- 17) Thompson JA, De Roach JN, McLaren TL, et al. The genetic profile of leber congenital amaurosis in an Australian cohort. *Mol Genet Genomic Med* 2017;5:652-67.
- 18) Han J, Rim JH, Hwang IS, et al. Diagnostic application of clinical exome sequencing in Leber congenital amaurosis. *Mol Vis* 2017;23:649-59.
- 19) Gillespie RL, Urquhart J, Anderson B, et al. Next-generation sequencing in the diagnosis of metabolic disease marked by pediatric cataract. *Ophthalmology* 2016;123:217-20.
- 20) Stone EM, Andorf JL, Whitmore SS, et al. Clinically focused molecular investigation of 1000 consecutive families with inherited retinal disease. *Ophthalmology* 2017;124:1314-31.
- 21) Soliman AZ, Xing C, Radwan SH, et al. Correlation of severity of fuchs endothelial corneal dystrophy with triplet repeat expansion in *TCF4*. *JAMA Ophthalmol* 2015;133:1386-91.
- 22) Wallace SE, Bird TD. Molecular genetic testing for hereditary ataxia: what every neurologist should know. *Neurol Clin Pract* 2018;8:27-32.