Exome sequencing identified *USH2A* mutations in Korean patients with retinitis pigmentosa

SeungHee Jung[1], YoungChan Park [1,2] , JongYoung Lee[2]*, InSong Koh[1]*

[1]Department of Biomedical Informatics, Hanyang University, Seoul, Korea

[2]One Omics, Seoul, Korea

Abstract

Retinitis pigmentosa (RP) is a rare hereditary retinal disease. It is characterized by progressive degeneration of photoreceptors. The worldwide prevalence of RP is about 1 in 4000. The clinical manifestations of RP are usually confined to the eyes, but some patients with RP have syndrome affecting non-ocular organs. RP is genetically heterogeneous and can be inherited in autosomal dominant, autosomal recessive, or X-linked modes. More than 60 genes are identified as causes of RP. Mutations in *USH2A* are most frequently reported as causes of non-syndromic autosomal recessive RP and Usher's syndrome type II which is one of the most common syndromes of RP. The *USH2A* gene is located on chromosome 1q42 with 72 exons. *USH2A* gene encodes for usherin which is a basement membrane protein in the retina. It is known that this protein is crucial for maintenance of mammalian photoreceptors. Since *USH2A* was first reported in 1998 as a cause of visual or hearing loss, many studies were done to reveal the *USH2A* mutation-phenotype correlation. We performed a whole exome sequencing analysis to identify *USH2A* variants in 157 Korean patients with non-syndromic RP. We identified total 92 variants, including 7 deleterious variants, and 49 missense variants.

Materials and Methods

1. Data collection

157 Korean RP patient's genomics materials were extracted from blood. Sequencing libraries were prepared along with capture using xGen Exome Research Panel V2 (Integrated DNA Technologies). The quantitation of libraries was measured by 2100 BioAnalyzer (Agilent).The exome libraries were paired-end sequenced on the Illumina Novaseq 6000 sequencing system aiming throughput at least 100x coverage.

2. Bioinformatics analysis

Bioinformatics analysis involves three stages (Fig 1). (1) Pre-processing; processing raw sequencing data for variant discovery, (2) Variant Discovery; identification of variants and quality of the consensus calls,(3) Variant Prioritization; filtering variants according to purpose of study.

(1) Pre-processing

bcl2fastq2 from the CASAVA software suite was performed to convert from BCL data to fastQ. The Sanger/phred33/illumine 1.8+ was used for the quality scale. To detect sequencing errors, FastQC was used. The remaining adapters were trimmed using cutadapt. Preprocessed sequencing reads were aligned against human reference genome GRCh37 using Burrow Wheeler Aligner(BWA) mem algorithms. To correct duplication that occurs from the PCR stage, mapped reads were processed using GATK's MarkDuplicate. then, base quality score recalibration was conducted to avoid exaggerated quality scores for each base.

(2) Variant Discovery

After preprocessing, GATK HaplotypeCaller was conducted to discover variants.

(3) Variant Prioritization

To prioritize confident variants, we filtered variants with QUAL < 30. Annotation was performed using variants annotation software SnpEff and SnpSift. To predict effect of novel variants, we use in silico prediction tools, Polyphen2, MutationTaster. Classification of variants was done according to ACMG/AMP guidelines.

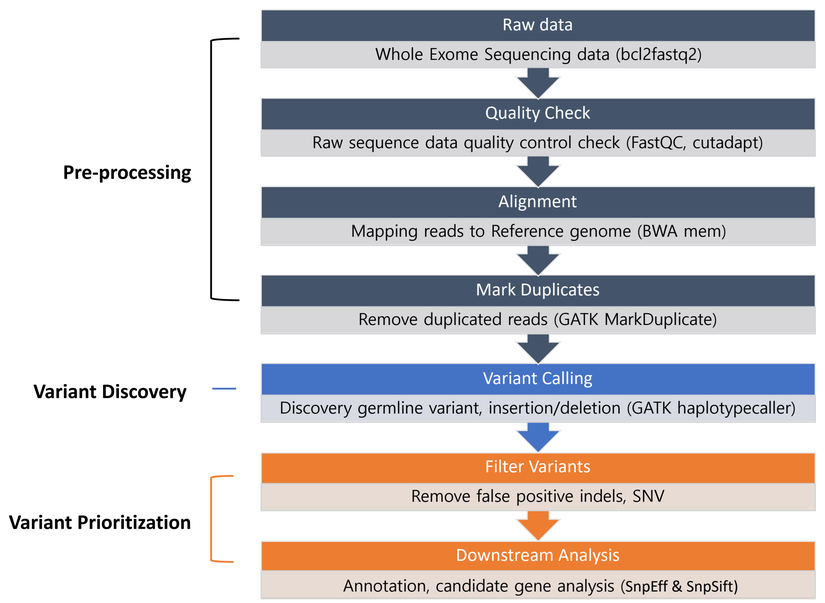

Figure 1. Summary of the 3 stages of WES analysis and tools used in this study

Results

Total 92 variants, including 14 variants known to be deleterious (4 splice site variants, 5 nonsense variants, 5 missense variants) were identified in 157 Korean RP patients (Table. 1). 10 novel variants were first identified in this study.

Nucleotide exchange	Translation effect	Polyphen-2	MutationTaster	state	Classification	Reported
c.2373T>G	p.Asn791Lys	Benign	Polymorphism	Het	Likely_benign	Novel
c.3428C>A	p.Thr1143Lys	Probably damiging	Disease causing	Het	Likely_pathogenic	Novel
c.3947C>G	p.Ala1316Gly	Possibly damiging	Disease causing	Het	Likely_pathogenic	Novel
c.5789G>C	p.Arg1930Pro	Probably damiging	Disease causing	Het	Pathogenic	Novel
c.10019C>A	p.Ser3340Tyr	Possibly damiging	Disease causing	Het	Likely_pathogenic	Novel
c.10613G>C	p.Arg3538Pro	Probably damiging	Disease causing	Het	Uncertain_significance	Novel
c.11657A>C	p.Lys3886Thr	benign	Polymorphism	Het	Likely_benign	Novel
c.13523C>T	p.Thr4508Ile	Possibly damiging	Disease causing	Het	Uncertain_significance	Novel
c.8650C>T	p.Gln2884*	NA	Disease causing	Het	Pathogenic	Novel
c.6326-1G>T	p.(?)	NA	NA	Het	Pathogenic	Novel
c.486-14G>A	p.(?)	NA	NA	Het	Likely_pathogenic	D Baux (2014)
c.820C>T	p.Arg274*	NA	Disease causing	Het	Pathogenic	D Baux (2007)
c.848+5G>C	p.(?)	NA	NA	Hom	Likely_pathogenic	T Sun (2018)
c.4217C>A	p.Ser1406*	NA	Disease causing	Het	Likely_pathogenic	MS Kim (2019)
c.4732C>T	p.Arg1578Cys	Possibly damiging	Disease causing	Het	Pathogenic	W jinda (2014)
c.8559-2A>G	p.(?)	NA	NA	Het	Pathogenic	Dai H (2008)
c.12708T>A	p.Cys4236*	NA	Disease causing	Hom	Pathogenic	H Nakanishi (2011)
c.11328T>G	p.Tyr3776*	NA	Disease causing	Het	Pathogenic	H Nakanishi (2011)

Table 1. Characteristics of *USH2A* variants identified in this study.

Discussion

We identified Korean patient's *USH2A* variants by analyzing genomic data. As a result, we identified 10 novel variants. Most deleterious mutations were found to be heterozygous. Since variants in *USH2A* are known to be responsible for the autosomal recessive RP, the mutations found as heterozygous are likely to have a weak or compound effect on phenotype.

Acknowledgement

This research was supported by NRF-2017M3C9A6047623.

HANYANG UNIVERSITY