의학박사 학위논문

Molecular Diagnosis of Hereditary Spherocytosis by Multi-gene Target Sequencing in Korea

다중 유전자 타깃 염기서열분석법을 이용한 한국인 유전구형적혈구증 환자의 분자진단

2016 년 2 월

서울대학교 대학원
임상의과학과
최 규 태

A thesis of the Degree of Doctor of Philosophy

다중 유전자 타깃 염기서열분석법을 이용한 한국인 유전구형적혈구증 환자의 분자진단

Molecular Diagnosis of Hereditary Spherocytosis by Multi-gene Target Sequencing in Korea

February 2016

The Department of Clinical Medical Sciences,

Seoul National University

College of Medicine

Kyu Tae Choi

ABSTRACT

Background: Hereditary spherocytosis (HS) is the most common cause of hereditary hemolytic anemia in Europe and North America. Current tests used to diagnose HS focus on the detection of hemolysis or indirectly assess protein defects. Direct methods to detect protein defects are complicated and difficult to implement. Recent next-generation sequencing (NGS) methods enable large-scale gene mutation analyses to be used for such diagnoses. In this study, we investigated the patterns of genetic variation associated with HS to determine the molecular mechanisms underlying the condition. Specifically, we analyzed mutations in red blood cell membrane protein-encoding genes and enzyme-encoding genes in Korean HS patients using NGS.

Methods: In total, 61 patients with HS were enrolled in this study. Targeted sequencing of 43 genes (17 membrane protein-encoding genes, 20 enzyme-encoding genes, and 6 additional candidate genes) was performed using the Illumina HiSeq platform and variants were called according to a data-processing pipeline.

Results: Of the 61 patients, 50 (82%) had one or more significant variants in a membrane protein gene. A total of 54 significant variants (8 previously reported and 46 novel) were detected in 6 membrane protein-encoding genes, i.e., *SPTB, ANK1, SPTA1, SLC4A1, EPB41*, and *EPB42*. The most variants (28) were detected in *SPTB*. Four significant variants, all of which were previously reported, were detected in genes encoding enzymes (*ALDOB, G6PD, GAPDH*, and *GSR*). Additionally, 5 previously reported variants were

detected in *UGT1A1*. These results suggest 35 primer sets that can be used to diagnose HS.

Conclusions: This was the first large-scaled genetic study of Korean HS patients. These results clarify the pattern of genetic variation associated with HS in Korean patients. They will enable easier, more rapid diagnosis of HS.

Keywords: Hereditary spherocytosis, Next generation sequencing, RBC membrane protein gene

Student number: 2012-30790

CONTENTS

Abstract ..... **i**

Contents ..... **iii**

List of tables ..... **v**

List of figures ..... **vi**

List of abbreviations ..... **viii**

1. Introduction **1**

1.1. Epidemiology **1**

1.2. Clinical manifestations **1**

1.3. Pathophysiology **1**

1.4. Diagnostic criteria **2**

1.5. Laboratory tests: Indirect methods **3**

1.6. Laboratory tests: Direct methods **6**

1.7. SOP suggested by Korean society of hematology **8**

1.8. Next Generation Sequencing **9**

1.9. Aims of this study **10**

2. Materials and Methods **12**

2.1. Overall study design **12**

2.2. Patients 14

2.3. DNA extraction 18

2.4. Targeted sequencing 19

2.5. Data analysis 23

2.6. Simulation of mutated gene effect on the protein structure 25

3. Results 27

3.1. Summary of targeted sequencing 27

3.2. Variant profile characteristics 31

3.3. Variants profile characteristics for membrane protein genes 35

3.4. Variant profile characteristics for enzyme genes 46

3.5. Variants profile characteristics for other candidate genes 54

3.6. The development of primer sets for the diagnosis of HS by using the mutational profile 57

4. Discussion 60

References 79

Abstract in Korean 87

LIST OF TABLES

Table 1 Baseline characteristics of 61 patients.....16

Table 2 Gene panel for targeted sequencing 20

Table 3 List of protein simulation templates.....26

Table 4 List of significant variants detected in patients.....32

Table 5 List of significant variants detected in membrane protein genes 43

Table 6 List of significant variants detected in enzyme genes....51

Table 7 Clinical features of mutated enzyme gene detected patients 52

Table 8 List of significant variants detected in *UGT1A1* gene ...56

Table 9 Primer sets for all significant variants in membrane protein 58

Table 10 SDS-PAGE results of RBC membrane protein in HS patients 69

Table 11 Comparison of Osmotic Fragility Test , PBS and Genetic Test results in HS patients.....72

Table 12 Baseline characteristics of 11 patients, who had no significant variant in RBC membrane genes 76

LIST OF FIGURES

Figure 1 Schematic diagram of research workflow.....13

Figure 2 Summary of data processing pipeline algorithms 24

Figure 3 Sample of base quality score for forward direction and reverse direction.....28

Figure 4 Average depth of coverage per gene and sample.....29

Figure 5 Number of observed significant variants per gene and samples. 30

Figure 6 Variants profile for each patient.....34

Figure 7 Characteristics of significant variants for membrane protein genes.....41

Figure 8 3-D molecular structure of membrane proteins.....45

Figure 9 Characteristics of significant variants for enzyme genes.....49

Figure 10 3-D molecular structure of enzymes 53

Figure 11 Significant variants diagrams for *UGT1A1* gene.....55

Figure 12 Comparison of gene study result between USA and Korean HS patients 70

Figure 13 Comparison of SDS-PAGE results between Japanese and Korean HS patients.....71
Figure 14 Comparison of mean hemoglobin concentration according to gene mutation.....73
Figure 15 Comparison of mean hemoglobin concentration according to pathogenicity of variants 74
Figure 16 Comparison of mean hemoglobin concentration according to number of variants.....75
Figure 17 Diagnostic work flow for HS suggested by international working parties and this study.....78

LIST OF ABBREVIATIONS

CBC	complete blood count
EMA	eosin-5´-maleimide
HHA	hereditary hemolytic anemia
HS	hereditary spherocytosis
LDH	lactate dehydrogenase
NGS	next generation sequencing
OFT	osmotic fragility test
PBS	peripheral blood smear
PCR	polymerase chain reaction
RBC	red blood cell
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SNP	single nucleotide polymorphism
TIBC	total iron-binding capacity

1. INTRODUCTION

1. 1. Epidemiology

Hereditary spherocytosis (HS), characterized by the presence of spherocytes in a peripheral blood smear, is the most common cause of hereditary hemolytic anemia[1, 2]. It occurs in 1 in 2000 Caucasians and has been known to be less common in Asians[1, 3, 4]. Although the exact incidence in Korea is not known, few studies have reported that HS is diagnosed in 1 in every 5000 births[5]. 75% cases of HS are inherited as autosomal dominant, while the rest are of autosomal recessive or de-novo inheritance.

1. 2. Clinical manifestations

A major clinical manifestation of HS is hemolytic anemia. Spherocytes are selectively captured in the spleen and destroyed, leading to hemolytic anemia in a wide range of clinical manifestations, from asymptomatic to life-threatening anemia requiring regular transfusion. Other clinical symptoms include splenomegaly, jaundice, and gallstone, depending on the severity of the disease.

1. 3. Pathophysiology

HS occurs due to the deficiency or a defect in red blood cell (RBC) membrane proteins and some studies have reported mutations of the genes that encode these proteins. Defective membrane protein disrupts the vertical linkage between the RBC membrane skeleton and phospholipid bilayer,

causing the RBC to lose its biconcave characteristics and become spherical in shape[3, 4, 6]. Various membrane proteins including ankyrin, band 3, protein 4.2, and spectrin are known to be associated with HS. Spectrin forms a cytoskeleton on the inside of the RBC membrane and linkage proteins such as ankyrin and protein 4.2 connect spectrin to band 3, a transmembrane protein. Among these RBC membrane proteins, spectrin is most frequently reported to be involved in HS in infants, while band 3 is mostly associated with adult HS.

1. 4. Diagnostic criteria

HS is generally diagnosed based on clinical history and routine laboratory tests. According to the guideline suggested by Bolton-Maggs et al.[7], patients with family history of HS, clinical features and typical laboratory tests results (presence of spherocytes on the blood film, raised mean corpuscular hemoglobin concentration(MCHC), increased reticulocyte count) do not require any additional tests. If the diagnosis is equivocal, they recommend special tests such as, cryohemolysis test and eosin-5′ -maleimide (EMA) binding test. They also recommend gel electrophoresis analysis of RBC membrane in atypical cases. Park et al.[8] used their own diagnostic criteria for HS, including anemia (patients aged 2-12 years : <11.5 g/dL, male patients >12 years old: <13.0 g/dL, female patients >12 years old: <12.0 g/dL), hyperbilirubinemia(>1.2 mg/dL for total bilirubin), splenomegaly, reticulocytosis(patients aged 2-6 or 12-15 years: ≥1.5%, patients aged 6-12 years: ≥1.9%, patients ≥15 years: ≥1.8%), and spherocytosis(≥2/high-

power field). Recently, International Council for Standardization in hematology (ICSH) suggested guideline for diagnosis of HS[9]. In diagnostic flow chart suggested by ICSH, if patient shows evidence of hemolysis and typical manifestations (jaundice, gallstones, hepato-splenomegaly), family history should be checked firstly. Next, laboratory tests including direct antiglobulin test (DAT) test, peripheral blood smear (PBS), bilirubin, and haptoglobin should be performed. If diagnosis is not clear, then screening tests (osmotic fragility test (OFT), acid glycerol lysis time (AGLT) test, EMA binding test) may be conducted. If results of screening tests are normal, further studies (sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), molecular analysis) or other diagnosis (enzymopathy, hemoglobinopathy) should be considered.

1. 5. Laboratory tests: Indirect methods

HS cause chronic extravascular hemolysis, so markers of hemolysis are important criteria for diagnosing HS. Complete blood count (CBC) including RBC indices is basic test in hematology. Hemoglobin concentration may be presented variously according to disease severity. Trait HS patients show normal hemoglobin concentration, while mild, moderate and severe HS patients show 11.0-15.0 g/dL, 8.0-12.0g/dL, and 6.0-8.0 g/dL, respectively[7]. Mean corpuscular volume (MCV) is normal and MCHC is usually increased in HS, which reflect a decrease in cell surface. Reticulocytes are immature RBC, and they present about 0.5-1.5% in adult, 2.5-6.5% in neonate. Reticulocytes usually increase in response to compensate for hemolysis in HS

patients. Haptoglobin is a protein present in α_2-globulin fraction. Haptoglobin binds free hemoglobin with high affinity, so serum haptoglobin is decrease when hemoglobin is released from destructed RBC. HS patients usually show decreased haptoglobin level. Bilirubin is a component of bile acid. It is produced by the metabolism of hemoglobin, and increased in hemolysis. Indirect bilirubin is predominantly increased in HS patients. Lactate dehydrogenase(LDH) is an enzyme, which catalyze the pyruvate to lactate conversion reaction and reverse reaction. LDH is released from destructed RBC, so serum LDH level is increase in hemolysis.

The DAT is a method for detecting autoimmune-related hemolytic anemia. RBC is washed by saline and 2-5% cell suspension is made. It is subsequently centrifuged with antihuman globulin added and then observed for aggregation formation in the tube. If the aggregation occurs, it is interpreted as positive which implies the RBC is attached to antibodies. Consequently, DAT test is negative for HS patient.

Peripheral blood smear is essential test for diagnosing HS. PBS is for blood cell morphology, and spherocytes are graded according to the mean percentage of observed spherocytes on 10 high power field (x1,000). Spherocytes are presented in three-graded format: slight or 1+ (1-5%), moderate or 2+ (6-15%), marked or 3+ (>15%)[10].

Osmotic fragility test is a common test for diagnosing HS indirectly by using the character of abnormal RBC breaking down more readily in hypotonic solution. The first step is to suspend the RBC in a tube containing 0.0%-0.9% hypotonic NaCl solution and incubate it for 30 minutes in room temperature.

After centrifugation, percentage of RBC hemolysis in the supernatant is recorded by NaCl concentration to graph it. If RBC is spherical, it is lysed in higher concentration due to decreased surface/volume ratio compared to normal RBC, thus having increased osmotic fragility in HS patients. In most HS patient, freshly drawn blood has increased osmotic fragility; however, it may be normal in mild HS patient. In this case, another step of 24h incubation in 37℃ can help to detect them. Increased OFT result is one character of HS patient although it is not specific. Recently, osmotic fragility test utilizing flow cytometry is introduced[13, 14]. In flow cytometry, deionized water which is hemolysis inducing agent spikes into RBC suspension during acquisition period and it is analyzed simultaneously. Residual RBC is applied for early lysed RBC and is expressed as '%residual red cell' for measuring osmotic fragility. OFT employing flow cytometry technique is more convenient, quantitative, objective and cost-effective compared to conventional OFT.

The AGLT test is used as screening test for diagnosing HS[11]. It uses the principle that glycerol retards entering water into RBC. AGLT test add 0.0053M sodium phosphate to improve sensitivity and specificity of the conventional glycerol lysis time test. Normally lysis time measured >900s, but 23s-45s in HS patients[9].

EMA binding test is an indirect method for diagnosing HS by detecting the band 3 protein defect[15]. Band 3 protein is a RBC membrane component which has EMA binding site. RBC is stained with EMA and then analyzed by flow cytometry. Here, band 3 protein defect can be measured by fluorescence

intensity of EMA which is bounded band 3 protein. HS is suspected when EMA fluorescence intensity is lower than 86.9% compared to normal control group[8].

The principle of cryohemolysis test is applying the character of abnormal RBC being more susceptible in temperature change when suspended in hypertonic solution compared to normal cells[12]. RBC is suspended in warm buffered sucrose and incubated in 37℃ for 10 minutes followed by 10 minute ice-cold bath incubation. Supernatant is collected after centrifugation and then measured for absorbance. Hemolysis is more progressed in HS patient compared to normal population. Cryohemolysis test is inexpensive, fast and convenient.

1. 6. Laboratory tests: Direct methods

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) is a direct method to detect protein defect of RBC membrane in HS patient. RBC membrane proteins are isolated and each fraction is separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis for analysis. The Fiarbanks[16] and Laemmli[17] methods are most commonly used, with 3.5% to 17% exponential gradient and 4% to 17% linear gradient gels, respectively. In these methods, proteins are fractionated according to their moving velocity and the result is interpreted by assigning the most-lagging strand as band 1, followed by band 2, and so on. In this way, proteins constituting RBC membrane can be distinguished and deficient proteins can be identified. However, SDS-PAGE is limited by its low accuracy as the results are greatly

affected by the processing of materials and also by the number of detectable abnormalities as only partial information regarding RBC membrane proteins can be obtained by this method. Meanwhile, it has been reported that detecting defective ankyrin by SDS-PAGE is difficult, resulting in low reliability of the method as well[1]. Moreover, the test takes approximately a week, which is a significant amount of time, as well as requiring constant monitoring, which makes the process laborious. Furthermore, interpreting the result is complicated and can often be subjective. Therefore, it cannot be routinely employed as a diagnostic tool in the general laboratory.

Analyzing the sequence and detecting mutations in the genes encoding RBC membrane proteins is another direct method to identify defects in membrane proteins. In the early 1990s, researchers performed sequence analysis of genes encoding several RBC membrane proteins from HS patient samples and identified a large number of deleterious mutations[14-19]. However, most studies were performed in USA and Europe, while a few have been reported from Asian countries, especially Japan[20, 21]. Moreover, most reports involved genetic mutations in SPTA1, SPTB, ANK1, and SLC4A1, which encode spectrin α, spectrin β, ankyrin-1, and band 3 protein, respectively. In addition, performing sequence analysis for all the genes encoding for RBC membrane protein is inefficient in terms of cost and time effectiveness.

Mass spectrometry measures the mass of a material under charged status and has been used mainly as a tool to analyze small molecular-weight molecules such as drugs and metabolites. Of late, the usage of mass analysis has gradually extended to protein profiling[13]. RBCs are lysed, the membrane

protein is isolated, and the sample is sequentially subjected to denaturation, reduction, and alkylation, followed by digestion into peptides by trypsin. Subsequently, molecular weight of the peptide and sequence of amino acids are matched with those of proteins registered in the database, which enables identification of the protein. Although the cost for the actual test is low, the mass spectrometer is quite expensive, costing several billion Wons; therefore, it is not affordable for relatively small laboratory. Additionally, the test method is complicated and requires skill, while interpretation is quite subjective. Therefore, it is not suitable to be used routinely as a diagnostic tool.

1. 7. SOP suggested by Korean society of hematology

Recently, the Hereditary Hemolytic Anemia Working Party of Korean society of hematology suggested standard operating procedure (SOP) for the diagnosis of hereditary hemolytic anemia[5]. It is similar to ICSH guideline except for excluding AGTL method in screening test and including mass spectrometry method in confirm test. According to the SOP, diagnosis of HS requires family history, physical examination, and laboratory tests. Family history of anemia, splenectomy, gallstone, or cholecystectomy is noted while taking patient history, while anemia, jaundice, and splenomegaly are determined during physical examination. Laboratory tests include complete blood count (CBC) with RBC index, reticulocyte count (corrected reti%), PBS, iron study (serum iron, TIBC, ferritin), serum LDH, serum haptoglobin, total and direct bilirubin, Coombs test, and hemoglobin EP. When HS is suspected

from history, physical examination, and laboratory tests, while OFT, autohemolysis test, flow cytometry (OF and EMA binding test), and protein analysis using mass spectrometry can be additionally tested[8, 15]. If these tests indicate normal membrane proteins, RBC enzyme test using mass spectrometry or genetic test may be considered.

1. 8. Next Generation Sequencing

Recently, sequence analysis methods termed as next-generation sequencing have been developed that enable the generation of data several thousand to million times faster than the previous, Sanger method[18]. The naming convention employs a comparison to the Sanger method, which is considered a conventional sequencing method for a long time. More precisely, the Sanger method is called 1st generation sequencing, where the chromosome is cleaved into small fragments, followed by amplification. A process where a large number of sequences are read in parallel is called 2nd generation sequencing, whereas reading the sequence from a single cell from end to end, without fragmentation is called 3rd generation sequencing. Currently, next-generation sequencing refers to 2nd generation sequencing. There are various types of platforms for 2nd generation sequencing, where Illumina and Ion Proton platforms are most frequently employed[19]. Illumina analyzes nucleotide sequence by taking images of the fluorescent signal generated upon DNA synthesis, and Ion Proton analyzes the sequence by measuring the amount of hydrogen ion released upon DNA synthesis using semi-conductors. Next-generation sequencing reduces the amount of time required from several

months to hours. Additionally, the Illumina platform is known to generate maximum of 750 GB data within 4 days and this is a significantly rapid progress compared to a maximum of several KB (kilo bases) in a day using the Sanger method. Depending on the subject of analysis, next-generation can be classified into whole-genome sequencing, whole-exome sequencing, and targeted sequencing. For sequencing genomes of human and specific animal and plant species, whole-genome sequencing is employed, and whole-exome sequencing is performed to detect causal genes for specific diseases. Targeted sequencing selectively amplifies exons or introns of specific genes and is used to analyze multiple genes simultaneously. For targeted sequencing, a specific region of interest in the genome is captured and amplified using microdroplet polymerase chain reaction (PCR), solid-phase capture, and solution-phase capture[20]. Since targeted sequencing enables simultaneous analysis of multiple genes, it is very useful to detect mutations in specific genes or genes of specific proteins, which are previously known or expected to be involved with diseases.

1. 9. Aims of this study

Considering the limitations of existing direct methods, developing a simple, easy, and direct method to measure RBC membrane protein abnormalities to diagnose HS is required. To the best of our knowledge, large-scale research on gene mutations in Korean patients with HS has not yet been conducted. In this study, we tried to establish genetic method to measure RBC membrane protein abnormalities to diagnose HS. To investigate the patterns of genetic variation

associated with HS, we analyzed for mutations in genes encoding RBC membrane proteins, using next-generation targeted sequencing, in Korean patients with HS. To rule out enzymopathy, we have analyzed the gene responsible for encoding RBC enzyme. We also analyzed genes involved in thalassemia and Gilbert syndrome, known to be inherited along with HS, as well as genes involved in paroxysmal nocturnal hemoglobinuria and congenital dyserythropoietic anemia, which manifest similar clinical symptoms as HS and therefore need to be distinguished.

2. MATERIALS AND METHODS

2. 1. Overall study design

The flow chart of this study is shown in figure 1. The whole blood of HS patient was collected, medical record was reviewed and DNA was extracted accordingly. Targeted sequencing for determined gene panel was performed and significant variant was selected through data analysis. Each variant was expressed as 3-D graphic representation by using molecular visualization tool. Primers were determined based on selected variants which could subsequently establish the gene panel for diagnosing hereditary spherocytosis.

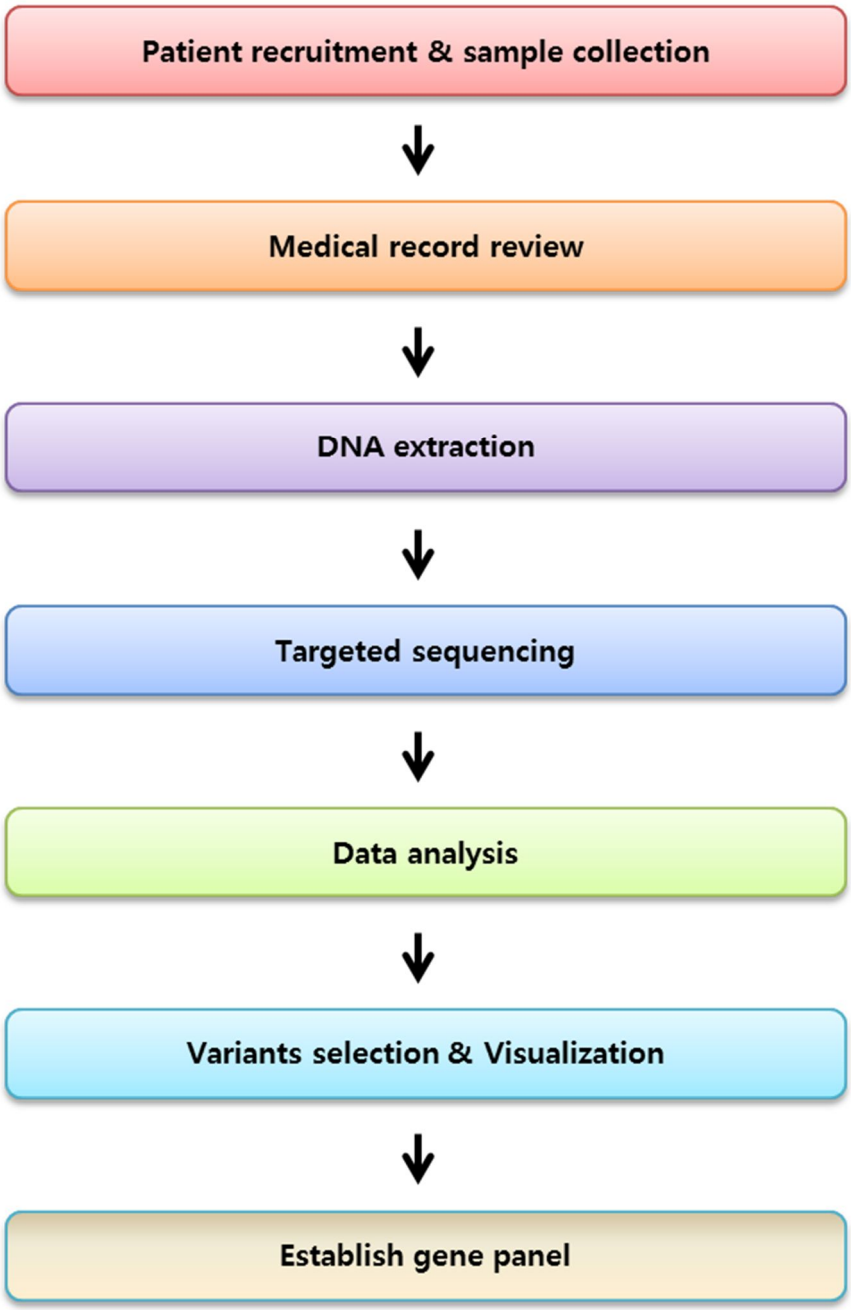


Figure 1. Schematic diagram of research workflow

2. 2. Patients

A total of 61 patients with HS, including 32 males and 29 females with a mean age of 10 years, were registered between July 2013 and July 2014 from pediatrics and internal medicine departments at 25 institutions nationwide. After reviewing their medical records, clinical data including age, sex, symptoms, family history, and results of laboratory test including CBC with RBC index, reticulocyte, total bilirubin concentration, indirect bilirubin concentration, lactate dehydrogenase, iron, total iron-binding capacity (TIBC), ferritin, peripheral blood smear, osmotic fragility test, and hereditary hemolytic anemia-related specialized test were collected (Table 1). A total of 20 patients (33%) had family history of HS, whereas symptoms of splenomegaly, jaundice, and hepatomegaly were exhibited by 39 (64%), 33 (54%), and 10 (16%) patients, respectively. Mean values for laboratory tests were as follows: hemoglobin concentration 8.3 g/dL, corpuscular volume 80.7 fL, corpuscular hemoglobin concentration 35.3 g/dL, reticulocyte count indicating hemolysis 7.3%, total bilirubin/indirect bilirubin 4.0/3.7 mg/dL, LDH 498 IU/L, and parameters representing iron profile included iron 100.8 μg/dL, TIBC 266.4 μg/dL, and ferritin concentration 346.6 ng/mL. Peripheral blood smear was rated for spherocytes on a 4-point scale from 0 to 3+ and the number of smears that showed 0, 1+, 2+ and 3+ were 7(11%), 15(25%), 21(34%), and 17(28%), respectively. Osmotic fragility test results were obtained from 46 (75%) patients, of which 39 showed increased osmotic fragility. Some laboratory tests contained missing data: 1 in PBS and MCHC, 2 in Reticulocyte count and Total Bilirubin, 13 in LDH, 14 in Indirect

Bilirubin, 15 in OFT, 21 in Iron, 22 in TIBC and Ferritin, respectively. Blood samples were collected from each patient after written consent and this study was approved by an Institutional Review Board (IRB) of each respective institution (IRB No. 1308-006-507).

Table 1. Baseline characteristics of 61 patients

Characteristics	HS Patients
Sex (%)	
Male	32 (52)
Female	29 (48)
Age (years)	
Mean	10
Range	1 ~ 81
Family History (%)	
Positive	20 (33)
Negative	41 (67)
Clinical symptoms (%)	
Splenomegaly	39 (64)
Jaundice	33 (54)
Hepatomegaly	10 (16)
Hematologic parameters (range)	
Hemoglobin (g/dL)	8.3 (3.6-13.5)
MCV (fL)	80.7 (62.0-107.0)
MCHC (g/dL) *	35.3 (30.8-38.2)
Markers of hemolysis (range)	
Reticulocyte count (%)*	7.3 (0.3-24.8)
Total Bilirubin (mg/dL)*	4.0 (0.4-19.1)
Indirect Bilirubin (mg/dL) *	3.7 (0.4-17.2)
LDH (IU/L) *	498 (187-1557)
Iron status parameters (range)	
Iron (ug/dL) *	100.8 (26.0-245.0)

TIBC (ug/dL) *	266.4 (108.0-486.0)
Ferritin (ng/mL) *	346.6 (31.6-4671.2)
Grading of Peripheral Spherocytes	Number of patients(%)
0	7 (11)
1+	15 (25)
2+	21 (34)
3+	17 (28)
NA	1 (2)
Osmotic fragility tests	Number of patients(%)
Positive	39 (64)
Negative	7 (11)
NA	15 (25)

*: calculated except missing data

2.3. DNA extraction

Genomic DNA was extracted from whole blood of all patients using QIAamp DNA Blood Mini Kit (Qiagen Biosciences, Germantown, MD, USA) according to the manufacturer instructions. Quality of purified DNA was evaluated via 260/280 absorbance ratio using ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA).

2.4. Targeted sequencing

Nucleotide sequences were captured using a kit manufactured based on a 43-gene panel (Table 2). This panel included all exon regions from 17 genes encoding the following membrane proteins: *SPTA1, SPTB, ANK1, SLC4A1, EPB41, EPB42, MPP1, ADD2, GYPA, GYPB, GYPC, STOM, AQP1, DMTN, ACTB, TMOD1*, and *TPM3*. Among these, *SPTA1, ANK1*, and *EPB42* were analyzed with their intron regions, since mutations involved in HS were reported in their intron region. In addition, in order to distinguish hereditary hemolytic anemia caused by enzyme defect, 20 genes including *HK1*, *GPI*, *ENO1*, *LDHB*, *GSR*, *ACHE*, *GAPDH*, *PFKM*, *PGK1*, *PKRL*, *G6PD*, *AK1*, *ADA*, *CMPK1*, *NT5C3A*, *ALDOA*, *ALDOB*, *TPI1*, *PGD*, and *GPX1* were added to the panel. Furthermore, *HBA1*, *HBA2*, *HBB*, and *UGT1A1* were included to determine the coinheritance of thalassemia and Gilbert syndrome, known to be co-inherited with HS. *SEC23B* was included to diagnose congenital dyserythropoietic anemia and *PIGA* was included to rule out paroxysmal nocturnal hemoglobinuria. Targeting the regions ranging a total of 407 Kbp, nucleotide sequence analysis using Illumina Hiseq 2500 platform was performed in rapid-run mode employing a paired-end, 150-bp probe. The complete process, including manufacturing a standard library, fragmenting genomic DNA, hybridization, and sequencing were performed in Celemics (Celemics Inc., Korea).

Table 2. Gene panel for targeted sequencing

	Protein	Gene	Location
Membrane	spectrin alpha	*SPTA1*	1q21
protein	spectrin beta	*SPTB*	14q23-q24.2
	ankyrin	*ANK1*	8p11.1
	band 3	*SLC4A1*	17q21.31
	protein 4.1	*EPB41*	1p33-p32
	protein 4.2	*EPB42*	15q15-q21
	p55	*MPP1*	Xq28
	beta-adducin	*ADD2*	2p13.3
	Glycophorin A	*GYPA*	4q31.21
	Glycophorin B	*GYPB*	4q31.21
	Glycophrins C/D	*GYPC*	2q14-q21
	Stomatin	*STOM*	9q34.1
	Aquaporin	*AQP1*	7p14
	Dematin	*DMTN*	8p21.1
	beta-actin	*ACTB*	7p22
	Tropomodulin 1	*TMOD1*	9p22.3
	Tropomyosin 3	*TPM3*	1q21.2

Enzyme	Hexokinase	*HK1*	10q22
	Glucose phosphate isomerase	*GPI*	19q13.1
	Enolase	*ENO1*	1p36.2
	Lactate dehydrogenase	*LDHB*	12p12.2-p12.1
	Glutathione reductase	*GSR*	8p21.1
	Acetylcholinesterase	*ACHE*	7q22
	Glyceraldehyde-3-phosphate dehydrogenase	*GAPDH*	12p13
	Phosphofructokinase	*PFKM*	12q13.3
	Phosphoglycerate kinase	*PGK1*	Xq13.3
	Pyruvate kinase	*PKRL*	1q21
	Glucose-6-phosphate dehydrogenase	*G6PD*	Xq28
	Adenylate kinase	*AK1*	9q34.1
	Adenosine deaminase	*ADA*	20q13.12
	Pyrimidine 5'-nucleotidase(CMP)	*CMPK1*	1p32
	Pyrimidine 5'-nucleotidase(UMP)	*NT5C3A*	7p14.3
	Aldolase A	*ALDOA*	16p11.2
	Aldolase B	*ALDOB*	9q21.3-q22.2
	Triosephosphate isomerase	*TPI1*	12p13

	6-phosphogluconic dehydrogenase	*PGD*	1p36.22
	Glutathione peroxidase	*GPX1*	3p21.3
Thalassemia	hemoglobin alpha1	*HBA1*	16p13.3
	hemoglobin alpha2	*HBA2*	16p13.3
	hemoglobin beta	*HBB*	11p15.5
CDAII	Sec23 homolog B	*SEC23B*	20p11.23
Gilbert syndrome	UDP glucuronosyltransferase 1 family	*UGT1A1*	2q37
PNH	Phosphatidylinositol Glycan Anchor Biosynthesis, Class A	*PIGA*	Xp22.1

2.5. Data analysis

FastQC program was employed to analyze the quality of raw data and following analyses were performed according to the DNAseq variant analysis workflow recommended by Best Practices of the Genome Analysis Toolkit (GATK)[21, 22]. Figure 2 represents the data analysis pipeline and schematic diagram of annotation and selection of the detected mutation. Raw data was mapped onto human reference genome hg19 from National Centre for Biotechnology information using the Burrow-Whleer Aligner (BWA v0.7.2) program[23]. Mapped data was aligned by the Picard (v1.109) program, a read group name was added, and the duplicated read group was deleted by MarkDuplicate.jar script from the same program. Next, realignment around indels was performed using the GATK (v3.1-1) program and base quality score recalibration was executed using BaseRecalibrator script. UnifiedGenotyper script from GATK was used for variants calling and snpEff program for annotation[24]. Additionally, dbNSFP script from the SnpSift program was used for more accurate annotation[25, 26]. Mutations were selected from the identified variants by running algorithms including SIFT, Polyphen2, MutationTaster, MutationAssessor, FATHMM, GERP++, and phyloP, and by referring to dbSNP138, 1000Gp, ESP6500[27-30]. In addition, an in-house Korean single nucleotide polymorphism(SNP) database (N=273) was used to filter polymorphisms out. All analyses were performed by a Linux-based computer using non-commercial version of the programs.

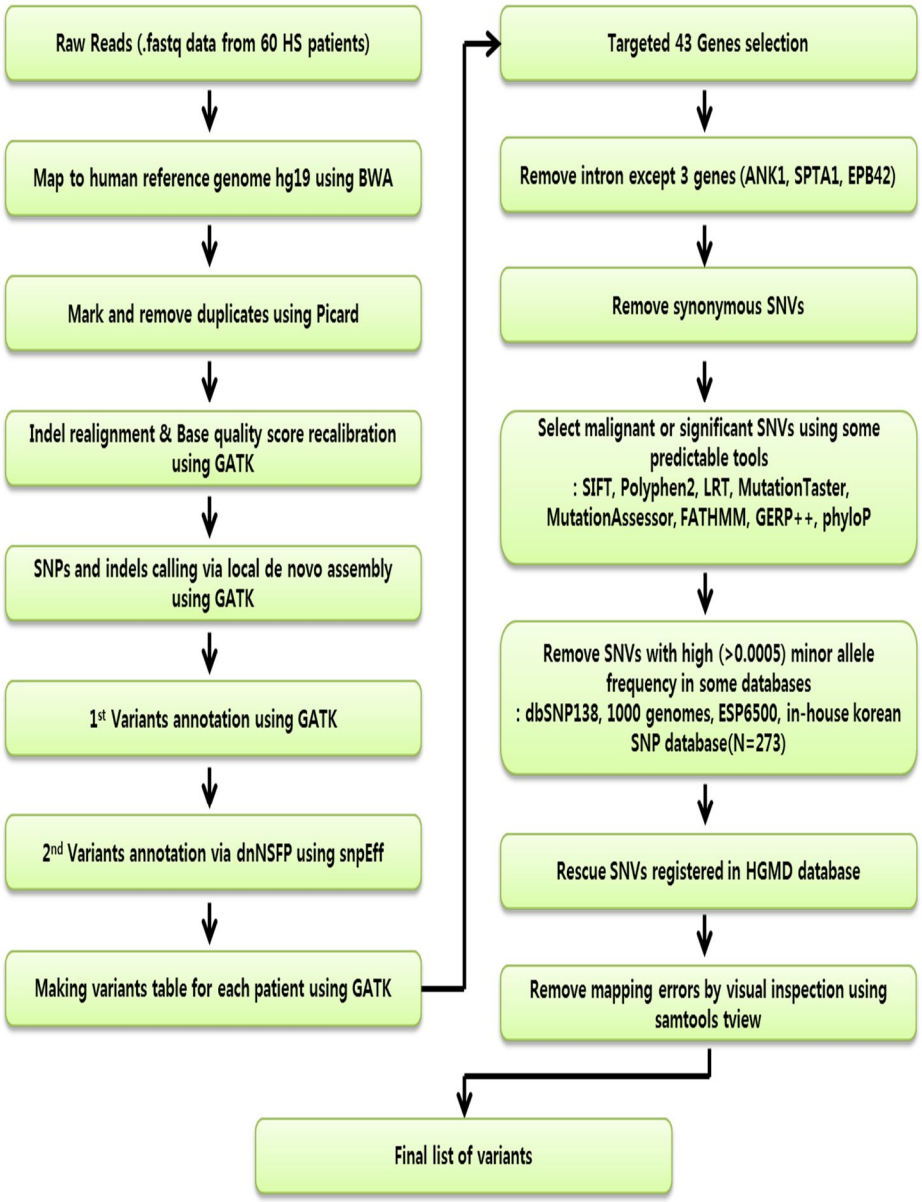


Figure 2. Summary of data-processing pipeline algorithm

2.6. Simulation of mutated gene effect on the protein structure

In order to predict how mutations affect the protein structure, we visualized three-dimensional (3-D) spatial protein structure. We selected and took protein template as protein data bank (PDB) format from web site ,'The Protein Model Portal (http://www.proteinmodelportal.org/)' (Table 3). We visualized 3-D graphic representation using PyMOL (The PyMOL Molecular Graphics System), stand-alone molecular visualization tool, and modified protein structure based on genetic mutation profile from NGS results. We represented protein structure as cartoon graphic style, and the region affected by frame-shift mutation was represented as simple line style. Dots were used to indicate one amino acid change by missense mutation. Protein structures followed by nonsense mutation were removed.

Table 3. List of protein simulation templates

Protein	UnitProt number	Template	Start position	End position	Length
Spectrin beta chain	P11277	Model 8 from SWISSMODEL	55	600	546
		Model 32 from SWISSMODEL	652	951	300
		Experimental structure 1S35	1064	1275	212
		Model 25 from SWISSMODEL	1277	1580	304
		Experimental structure 3KBU	1582	1906	325
		Experimental structure 3LBX	1901	2084	184
Ankyrin-1	P16157	Model 5 from SWISSMODEL	9	789	781
		Experimental structure 1N11	405	812	408
		Experimental structure 3UD1	912	1233	322
		Experimental structure 2YQF	1394	1497	104
Band 3	P02730	Experimental structure 1BH7	803	835	33
Spectrin alpha chain	P02549	Model 4 from SWISSMODEL	1606	1816	211
Protein 4.1	P11171	Model 1 from SWISSMODEL	208	519	312
Protein 4.2	P16452	Model 2 from: SWISSMODEL	5	689	685
Aldolase B, fructose-bisphosphate	P05062	Experimental structure 1QO5	2	364	363
Glucose-6-phosphate dehydrogenase	P11413	Experimental structure 1QKI	2	515	514
Glyceraldehyde-3-phosphate dehydrogenase	P04406	Experimental structure 4WNC	1	335	335
Glutathione reductase	P00390	Experimental structure 3DJG	45	522	478

3. RESULTS

3.1. Summary of targeted sequencing

All 61 samples were sequenced in both directions on the Illumina HiSeq 2500 platform. The base-quality scores of all read positions in all samples were adequate for analysis (Figure 3). In figure3, base-quality phred score was separated by colors: green for over 28, amber for 20-28, pink for less than 20. Although, there may be some differences of quality score between forward and reverse direction due to library composition, all the mean value of the base-quality phred score are located in the green zone. The mean sequencing depth for the 43 target genes was 548× (range, 177–1280×), and for the 61 samples was 984× (range 534–1171×) (Figure 4). Target genes and samples showed read depths of greater than 100×, which is considered good coverage for sequencing analyses. Significant variants were observed in 11 of the 43 total genes. A total of 63 variants were detected, and most (28 variants) were observed in the *SPTB* gene. Nineteen variants were observed in *ANK1*, and 5 were detected in *UGT1A1*. The *SLC4A1* and *SPTA1* genes contained 3 and 2 variants, respectively. One variant in each of *EPB41, EPB42, GAPDH, GSR, G6PD*, and *ALDOB* was observed (Figure 5A).

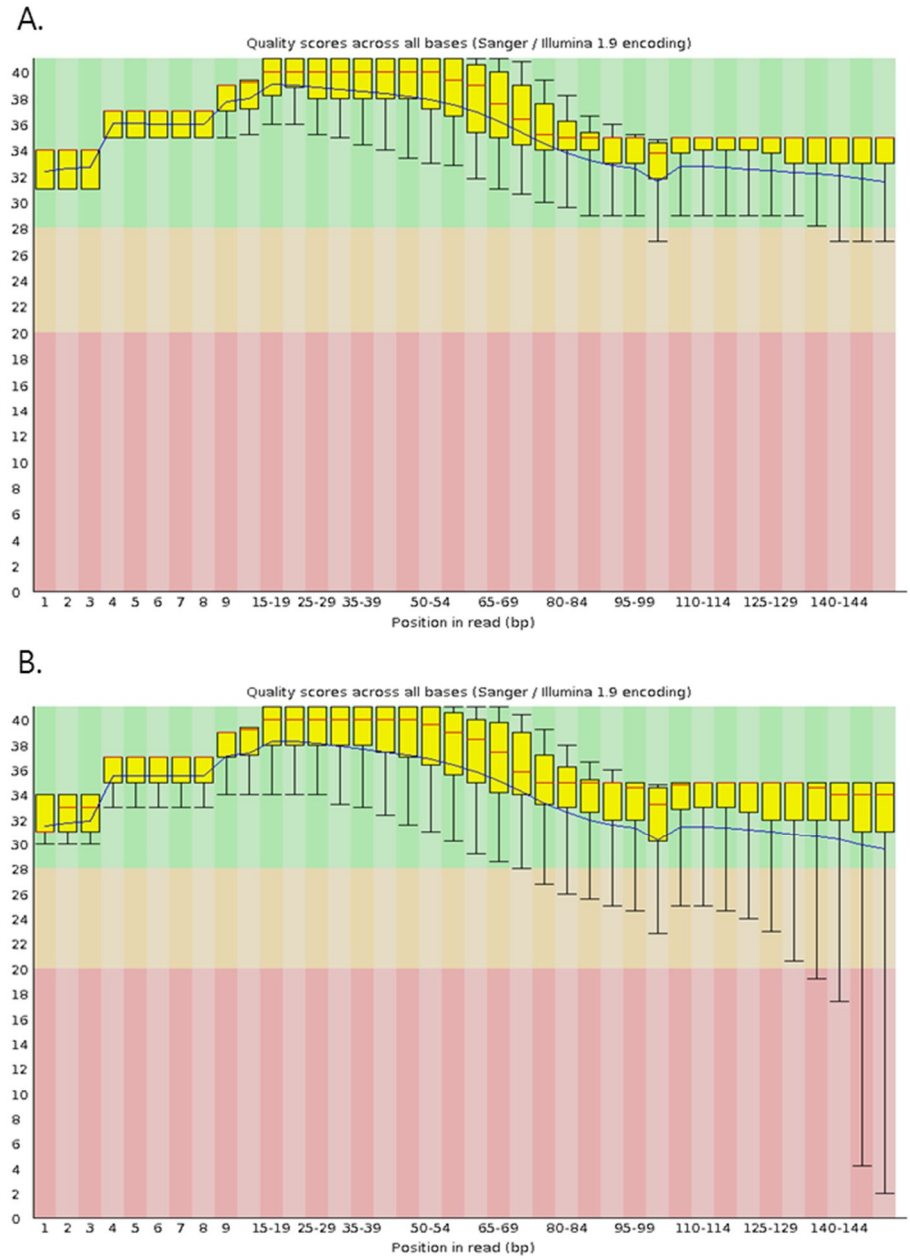


Figure 3. Sample of base quality score for forward direction (A) and reverse direction(B)

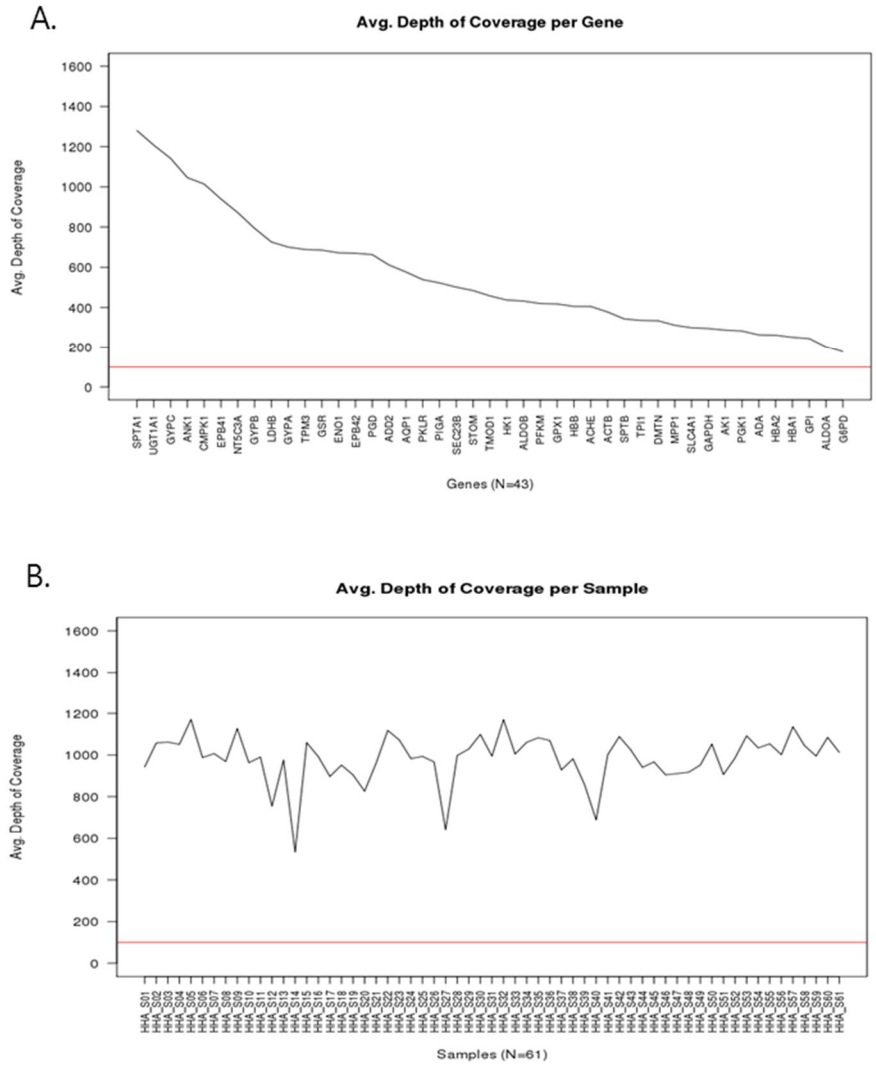


Figure 4. Average depth of coverage per gene(A) and sample(B)

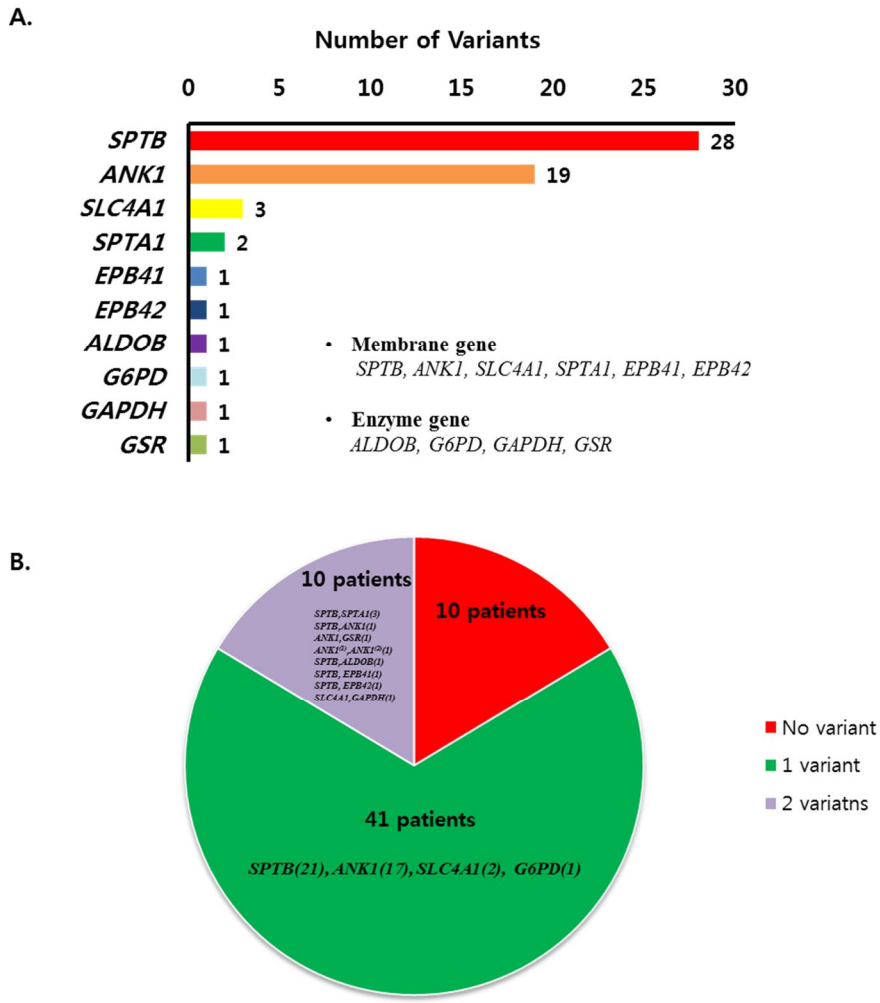


Figure 5. Number of observed significant variants per gene (A) and samples (B)

3.2. Variant profile characteristics

More than half of patients (n = 41, 67%) had a single variant, 10 patients had 2 variants (16%), and more than 3 variants were not observed in any patient (Figure 5B). Of the 61 HS patients, 50 (82%) had at least one variant in a membrane protein gene (Table 4). Among patients with a single variant, the most frequently mutated genes were *SPTB* followed by *ANK1*, each of which was mutated in 21 and 17 patients, respectively. Two patients had a *SCL4A1* mutation, and one patient had a *G6PD* mutation. Among patients with 2 variants, one patient had 2 mutations in a single gene, *ANK1*. Three patients had mutations in both *SPTB* and *SPTA1*. Combinations of variants on *SLC4A1* and *GAPDH*, *ANK1* and *GSR*, *SPTB* and *ANK1*, *SPTB* and *ALDOB, SPTB* and *EPB41,* and *SPTB* and *EPB42* were detected in one patient each. Ten patients had no variant on the membrane protein or enzyme genes (Figure 6).

Table 4. List of significant variants detected in patients

Patient ID	Gene	Chrom	SNP ID	Variant(nucleotide)	Variant(Amino acid)	Effect	ACMG Classification
1	SPTB	chr14	.	c.647+1G>C		SPLICE_SITE_DONOR	Uncertain significance
	EPB41	chr1	rs189183599	c.848C>T	p.Thr283Ile	NON_SYNONYMOUS_CODING	Uncertain significance
2	ANK1	chr8	.	c.4095delC		FRAME_SHIFT	Likely pathogenic
3	SPTB	chr14	.	c.4515delC		FRAME_SHIFT	Likely pathogenic
4	SPTB	chr14	.	c.4842G>C	p.Lys1614Asn	NON_SYNONYMOUS_CODING	Uncertain significance
6	ANK1	chr8	.	c.2029C>T	p.Gln677*	STOP_GAINED	Likely pathogenic
7	SPTB	chr14	.	c.5015delA		FRAME_SHIFT	Likely pathogenic
8	SPTB	chr14	.	c.639_640insG		FRAME_SHIFT	Likely pathogenic
	SPTA1	chr1	rs182430449	c.5093T>C	p.Val1698Ala	NON_SYNONYMOUS_CODING	pathogenic
9	SPTB	chr14	.	c.187T>C	p.Trp63Arg	NON_SYNONYMOUS_CODING	Uncertain significance
	SPTA1	chr1	rs182430449	c.5093T>C	p.Val1698Ala	NON_SYNONYMOUS_CODING	pathogenic
11	SPTB	chr14	.	c.5153_5154insG		FRAME_SHIFT	Likely pathogenic
12	ANK1	chr8	.	c.511A>T	p.Lys171*	STOP_GAINED	Likely pathogenic
13	SPTB	chr14	.	c.643_644insC		FRAME_SHIFT	Likely pathogenic
15	ANK1	chr8	.	c.830A>G	p.His277Arg	NON_SYNONYMOUS_CODING	pathogenic
16	ANK1	chr8	.	c.3754C>T	p.Arg1252*	STOP_GAINED	pathogenic
18	ANK1	chr8	.	c.4387_4390delAACA		FRAME_SHIFT	pathogenic
19	ANK1	chr8	.	c.2227C>T	p.Gln743*	STOP_GAINED	pathogenic
20	SPTB	chr14	.	c.5798+1G>A		SPLICE_SITE_DONOR	Uncertain significance
	SPTA1	chr1	.	c.2187C>G	p.His729Gln	NON_SYNONYMOUS_CODING	Uncertain significance
21	SLC4A1	chr17	.	c.2423G>A	p.Arg808His	NON_SYNONYMOUS_CODING	pathogenic
23	SPTB	chr14	.	c.4577_4586del		FRAME_SHIFT	pathogenic
25	ANK1	chr8	.	c.2390_2393delTAGT		FRAME_SHIFT	pathogenic
26	ANK1	chr8	.	c.2227C>T	p.Gln743*	STOP_GAINED	pathogenic
27	ANK1	chr8	.	c.2739delT		FRAME_SHIFT	Likely pathogenic
28	SPTB	chr14	.	c.3489_3490delCC		FRAME_SHIFT	Likely pathogenic
29	ANK1	chr8	.	c.3157C>T	p.Arg1053*	STOP_GAINED	pathogenic
	GSR	chr8	rs149225584	c.1520C>T	p.Thr507Ile	NON_SYNONYMOUS_CODING	Uncertain significance
30	SPTB	chr14	.	c.3466_3467insA		FRAME_SHIFT	Likely pathogenic

	ALDOB	chr9	rs145252200	c.401G>A	p.Arg134His	NON_SYNONYMOUS_CODING	Uncertain significance
31	SPTB	chr14	.	c.3241C>T	p.Gln1081*	STOP_GAINED	Likely pathogenic
32	SLC4A1	chr17	rs373916826	c.2278C>T	p.Arg760Trp	NON_SYNONYMOUS_CODING	pathogenic
33	SPTB	chr14	.	c.208C>T	p.Arg70*	STOP_GAINED	Likely pathogenic
34	SPTB	chr14	.	c.4987A>T	p.Arg1663*	STOP_GAINED	Likely pathogenic
35	SPTB	chr14	rs121918651	c.1A>G	p.Met1Val	START_LOST	pathogenic
	EPB42	chr15	rs368756068	c.835G>A	p.Val279Met	NON_SYNONYMOUS_CODING	Uncertain significance
36	SPTB	chr14	.	c.5938-2A>G		SPLICE_SITE_ACCEPTOR	Uncertain significance
37	ANK1	chr8	.	c.3235G>T	p.Glu1079*	STOP_GAINED	Likely pathogenic
39	ANK1	chr8	.	c.5066_5067delTG		FRAME_SHIFT	Likely pathogenic
40	SPTB	chr14	.	c.376C>T	p.Gln126*	STOP_GAINED	Likely pathogenic
41	ANK1	chr8	.	c.856C>T	p.Arg286*	STOP_GAINED	Likely pathogenic
42	ANK1	chr8	.	c.2803C>T	p.Arg935*	STOP_GAINED	Likely pathogenic
43	ANK1	chr8	.	c.4103A>T	p.Lys1368Met	NON_SYNONYMOUS_CODING	Uncertain significance
	ANK1	chr8	.	c.2106C>A	p.Tyr702*	STOP_GAINED	Likely pathogenic
44	SPTB	chr14	.	c.360_361insTC		FRAME_SHIFT	Likely pathogenic
	ANK1	chr8	.	c.593A>G	p.Asn198Ser	NON_SYNONYMOUS_CODING	Uncertain significance
45	ANK1	chr8	.	c.1924delG		FRAME_SHIFT	Likely pathogenic
46	SPTB	chr14	.	c.4267C>T	p.Arg1423*	STOP_GAINED	Likely pathogenic
47	SPTB	chr14	.	c.4321delC		FRAME_SHIFT	Likely pathogenic
48	G6PD	chrX	.	c.1142T>G	p.Phe381Cys	NON_SYNONYMOUS_CODING	pathogenic
49	SPTB	chr14	.	c.764-1G>T		SPLICE_SITE_ACCEPTOR	Uncertain significance
50	SPTB	chr14	.	c.2533C>T	p.Gln845*	STOP_GAINED	pathogenic
51	ANK1	chr8	.	c.2095_2096insT		FRAME_SHIFT	Likely pathogenic
52	SPTB	chr14	.	c.2014delG		FRAME_SHIFT	Likely pathogenic
54	ANK1	chr8	.	c.4387_4390delAACA		FRAME_SHIFT	pathogenic
56	SPTB	chr14	.	c.4291C>T	p.Arg1431*	STOP_GAINED	Likely pathogenic
57	SLC4A1	chr17	.	c.1526G>T	p.Gly509Val	NON_SYNONYMOUS_CODING	Uncertain significance
	GAPDH	chr12	.	c.164A>G	p.Lys55Arg	NON_SYNONYMOUS_CODING	Uncertain significance
58	SPTB	chr14	.	c.472C>T	p.Gln158*	STOP_GAINED	Likely pathogenic
59	SPTB	chr14	.	c.6022+1G>A		SPLICE_SITE_DONOR	Uncertain significance
60	SPTB	chr14	.	c.4873C>T	p.Arg1625*	STOP_GAINED	Likely pathogenic

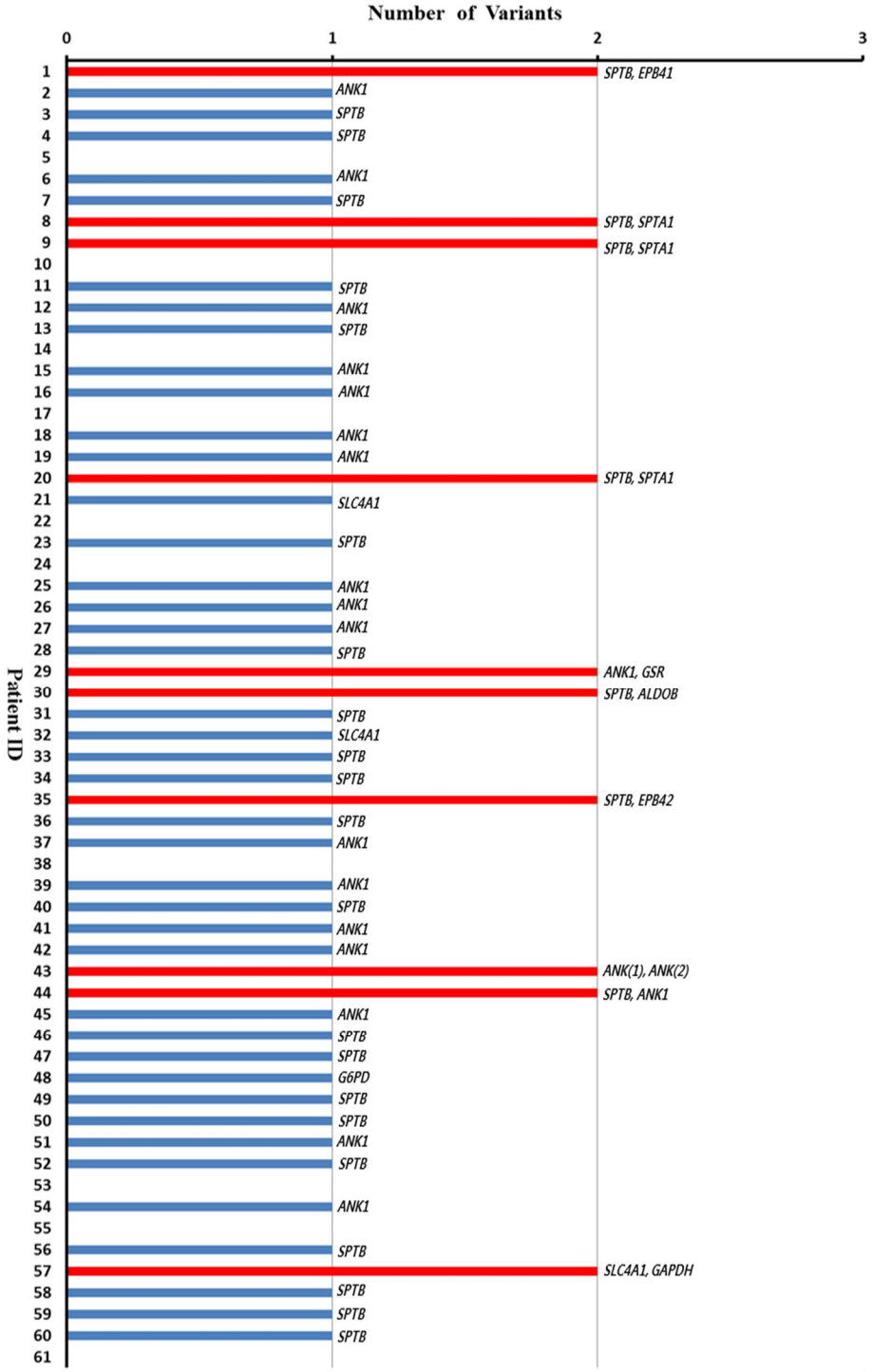


Figure 6. Variants profile for each patient

3.3. Variants profile characteristics for membrane protein genes

Among 17 membrane protein genes examined, significant disease-related variants were observed in 6 genes, *SPTB, ANK1, SLC4A1, SPTA1, EPB41*, and *EPB42* (Figure 7). A total of 54 significant variants were observed, of which 8 were previously reported as pathogenic variants in HS patients and 46 variants were novel variants (Table 5)[31-38]. According to the American College of Medical Genetics and Genomics (ACMG) guidelines[39], 12 were pathogenic variants (including 8 previously reported variants) and 29 were likely pathogenic variants. Thirteen variants were classified as uncertain significance category.

The *SPTB* gene, which encodes the spectrin beta protein, had the most variants (n = 28), two of which were previously reported as pathogenic variants in HS patients[31, 32]. Three variants were classified as pathogenic, 18 and 7 variants were likely pathogenic and uncertain significance, respectively, in ACMG guidelines. One variant was a missense mutation at codon 1, c.1A>G, resulting in the replacement of methionine with valine and the loss of the translation start signal. The other variant was a nonsense mutation at codon 845, c.2533C>T, resulting in the replacement of glutamine with a stop codon. All other variants were novel. Eleven were frameshift mutations and have been designated as follows, according to HGVS recommendations: c.360_361insTC, c.639_640insG, c.643_644insC, c.2014delG, c.3466_3467insA, c.3489_3490delCC, c.4321delC, c.4515delC, c.4577_4586del, c.5015delA, c.5153_5154insG. Eight novel nonsense

mutations were detected as follows: c.208C>T, p.Arg70*; c.376C>T, p.Gln126*; c.472C>T, p.Gln158*; c.3241C>T, p.Gln1081*; c.4267C>T, p.Arg1423*; c.4291C>T, p.Arg1431*; c.4873C>T, p.Arg1625*; and c.4987A>T, p.Arg1663*. The following two missense mutations were detected: c.187T>C, p.Trp63Arg and c.4842G>C, p.Lys1614Asn. At the splice sites near the exonic regions, 5 significant variants were detected: c.647+1G>C, c.764-1G>T, c.5798+1G>A, c.5938-2A>G, and c.6022+1G>A. All variants were heterozygous.

Nineteen significant variants were observed in the *ANK1* gene, which encodes the ankyrin protein. Six variants were classified as pathogenic, 11 and 2 variants were likely pathogenic and uncertain significance, respectively, in ACMG guidelines. Four variants that were previously reported as pathogenic in HS patients [33-36], were detected, including two nonsense mutations (c.3157C>T, p.Arg1053* and c.3754C>T, p.Arg1252*), one frameshift mutation (c.2390_2393delTAGT), and one missense mutation (c.830A>G, p.His277Arg). The others were novel variants. Seven were nonsense mutations: c.511A>T, p.Lys171*; c.856C>T, p.Arg286*; c.2029C>T, p.Gln677*; c.2106C>A, p.Tyr702*; c.2227C>T, p.Gln743*; c.2803C>T, p.Arg935*; and c.3235G>T, p.Glu1079*. Six were frameshift mutations: c.1924delG, c.2095_2096insT, c.2739delT, c.4095delC, c.4387_4390delAACA, and c.5066_5067delTG. Two were missense mutations: c.593A>G, p.Asn198Ser and c.4103A>T, p.Lys1368Met. No significant variants were detected in the intronic region. All variants were heterozygous.

The *SLC4A1* gene encodes the band 3 protein. It contained 2 variants that were previously reported as pathogenic in HS patients [37, 38]. Both were missense mutations: c.2278C>T, p.Arg760Trp and c.2423G>A, p.Arg808His. It also contained a third, novel variant that was a missense mutation: c.1526G>T, p.Gly509Val. This novel variant was classified as uncertain significance. All variants detected in *SLC4A1* gene were heterozygous.

Two significant variants were detected in the *SPTA1* gene, which encodes the spectrin alpha protein. One variant was a novel missense mutation: c.2187C>G, p.His729Gln, which classified as uncertain significance. The other was previously reported as single nucleotide polymorphism (SNP, rs373916826), but we regarded it as a significant variant and classified as pathogenic in ACMG guidelines. It was missense mutation: c.5093T>C, p.Val1698Ala.

EPB41 encodes the band 4.1 protein. One variant detected in *EPB41* was a previously report SNP (rs189183599), which was classified as uncertain significance: c.848C>T, p.Thr283Ile.

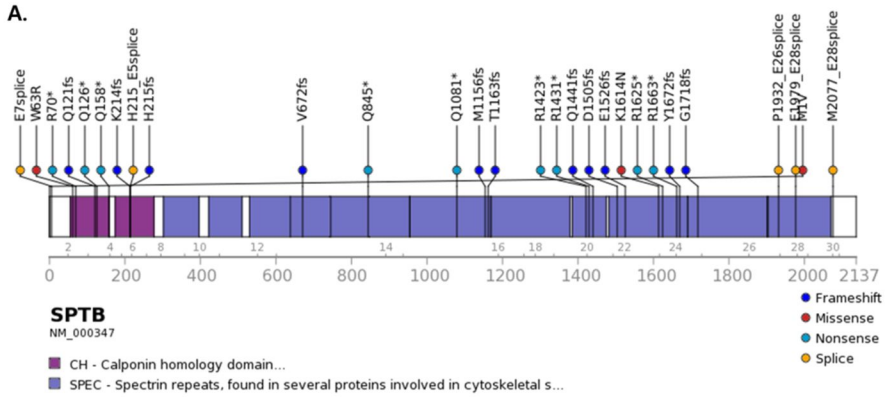
One variant was detected in *EPB42*, which encodes the band 4.2 protein. It was a missense mutation: c.835G>A, p.Val279Met, and was previously reported as a SNP (rs368756068). This variants was also classified as uncertain significance in ACMG guildelines.

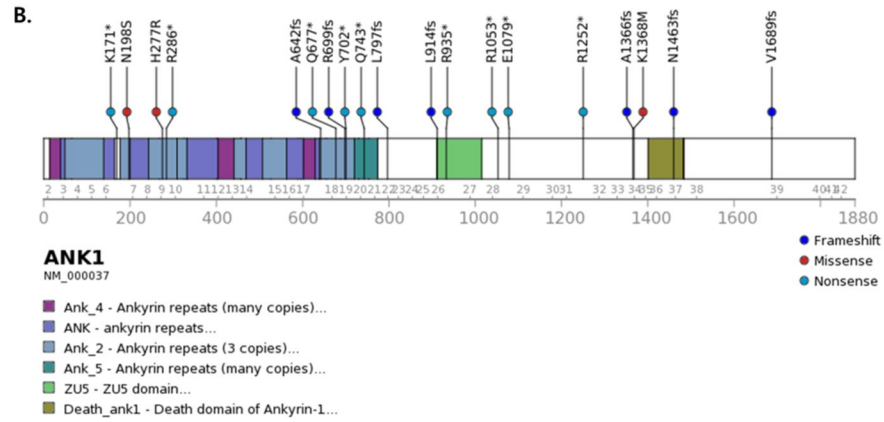
Each membrane protein was visualized 3-D spatial structure and effects of gene mutations were reflected in adequate graphic style. Spectrin beta chain structure was partially constructed by six different prediction model templates. All gene mutations, except c.1A>G, p.Met1Val, were reflected in models.

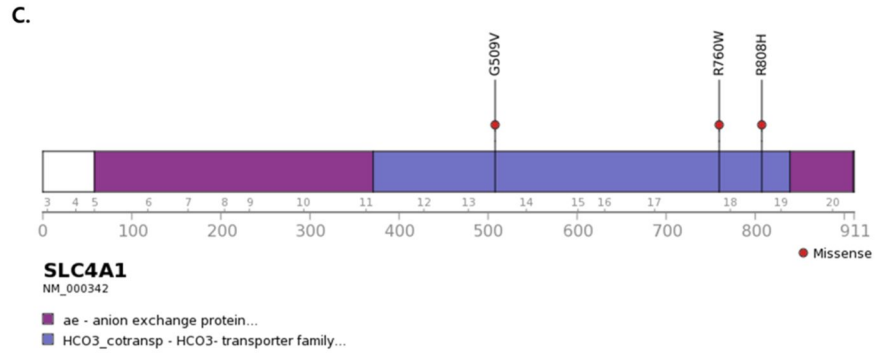
First template was Model 8 from SWISSMODEL based on 4d1eA, which comes from the crystal structure of human muscle alpha-actinin-2(Figure 8A1). It contained amino acid sequence of spectrin beta chain from 56 to 600. Nine mutations were reflected in eight separated frames(Figure 8A2-B2). Second template was Model 32 from SWISSMODEL based on 1u4qB, which comes from the crystal structure of repeats 15,16 and 17 of chicken brain alpha spectrin(Figure 8B3). It contained amino acid sequence of spectrin beta chain from 652 to 951. Two mutations were reflected in two separated frames(Figure 8B4,B5). Third template was experimental structure 1S35, which comes from crystal structure of repeat 8 and 9 of human erythroid spectrin. It contained amino acid sequence of spectrin beta chain from 1064 to 1275(Figure 8B6). Three mutations were reflected in three separated frames(Figure 8B7-C2). Fourth template was Model 25 from SWISSMODEL based on 1u4qB, same as second template(Figure 8C3). It contained amino acid sequence of spectrin beta chain from 1277 to 1580. Five mutations were reflected in five separated frames(Figure 8C4-D1). Fifth template was experimental structure 3KBU, which comes from Crystal structure of the ankyrin binding domain of human erythroid beta spectrin (repeats 13-15) in complex with the spectrin binding domain of human erythroid ankyrin (ZU5-ANK), EMTS derivative(Figure 8D2). It contained amino acid sequence of spectrin beta chain from 1582 to 1906. Five mutations were reflected in six separated frames(Figure 8D3-E17). Sixth template was experimental structure 3LBX, which comes from crystal structure of the erythrocyte spectrin tetramerization domain complex(Figure 8E2). It contained amino acid

sequence of spectrin beta chain from 1901 to 2084. Three mutations were reflected in three separated frames(Figure 8E3-E5). Ankyrin-1 was partially constructed by four different prediction model templates. Among 19 mutations, four mutations (c.3754C>T, c.4095delC, c.4103A>T, c.5066_5067delTG) could not be reflected because absence of templates. First template was Model 5 from SWISSMODEL based on 4rlvA, which comes from crystal structure of AnkB 24 ankyrin repeats in Complex with AnkR autoinhibition segment(Figure 8E6). It contained amino acid sequence of Ankyrin-1 chain from 9 to 789. Four mutations were reflected in six separated frames(Figure 8E7-F5). Second template was experimental structure 1N11, which comes from D34 region of human ankyrin-r and linker(Figure 8E6). It contained amino acid sequence of Ankyrin-1 from 405 to 812. Six mutations were reflected in six separated frames(Figure 8E7-G5). Third template was experimental structure 3UD1, which comes from crystal structure of ZU5A-ZU5B domains of human erythrocyte ankyrin(Figure 8G6). It contained amino acid sequence of Ankyrin-1 from 912 to 1233. Four mutations were reflected in four separated frames(Figure 8G7-H3). Fourth template was experimental structure 2YQF, which comes from Solution structure of the death domain of Ankyrin-1(Figure 8H4). It contained amino acid sequence of Ankyrin-1 from 1394 to 1497. One mutation were reflected in one separated frames(Figure 8H5). Band 3 was partially constructed by one prediction model template. Among three mutations, two mutations (c.1526G>T, c.2278C>T) could not be reflected because absence of templates. The used template was experimental structure 1BH7, which comes from a low energy

structure for the final cytoplasmic loop of Band 3, NMR, minimized average structure(Figure 8H6). It contained amino acid sequence of Band 3 from 803 to 835. One mutation was reflected in one separated frame(Figure 8H7). Spectrin alpha chain was partially constructed by one prediction model template. Among two mutations, one mutation (c.2187C>G) could not be reflected because absence of template. The used template was Model 4 from SWISSMODEL based on 1u5pA, which comes from crystal structure of repeats 15 and 16 of chicken brain alpha spectrin(Figure 8I1). It contained amino acid sequence of spectrin alpha chain from 1606 to 1816. One mutation was reflected in one separated frame(Figure 8I2). Protein 4.1 was partially constructed by one prediction model template. The used template was Model 1 from SWISSMODEL based on 2i1jA, which comes from moesin from spodoptera frugiperda at 2.1 angstroms resolution(Figure 8I3). It contained amino acid sequence of protein 4.1 from 208 to 519. One mutation was reflected in one separated frame(Figure 8I4). Protein 4.2 was almost completely constructed by one prediction model template. The used template was Model 2 from SWISSMODEL based on 4pygB, which comes from crystal structure of transglutaminase 2 with GTP complex and amino acid sequence evidence of evolution of GTP binding site(Figure 8I5). It contained amino acid sequence of protein 4.2 from 5 to 689. One mutation was reflected in two separated frames(Figure 8I6,I7).


A.
E7splice
W63R
R70*
Q121fs
Q126*
Q158*
K214fs
H215_E5splice
H215fs
V672fs
Q845*
Q1081*
M1156fs
T1163fs
R1423*
R1431*
Q1441fs
D1505fs
E1526fs
K1614N
R1625*
R1663*
Y1672fs
G1718fs
P1932_E26splice
E1979_E28splice
M2077_E28splice
0 200 400 600 800 1000 1200 1400 1600 1800 2000 2137
SPTB
NM_000347
Frameshift
Missense
Nonsense
Splice
CH - Calponin homology domain...
SPEC - Spectrin repeats, found in several proteins involved in cytoskeletal s...


B.
K171*
N198S
H277R
R286*
A642fs
Q677*
R699fs
Y702*
Q743*
L797fs
L914fs
R935*
R1053*
E1079*
R1252*
A1366fs
K1368M
N1463fs
V1689fs
0 200 400 600 800 1000 1200 1400 1600 1880
ANK1
NM_000037
Frameshift
Missense
Nonsense
Ank_4 - Ankyrin repeats (many copies)...
ANK - ankyrin repeats...
Ank_2 - Ankyrin repeats (3 copies)...
Ank_5 - Ankyrin repeats (many copies)...
ZU5 - ZU5 domain...
Death_ank1 - Death domain of Ankyrin-1...


C.
G509V
R760W
R808H
0 100 200 300 400 500 600 700 800 911
SLC4A1
NM_000342
Missense
ae - anion exchange protein...
HCO3_cotransp - HCO3- transporter family...

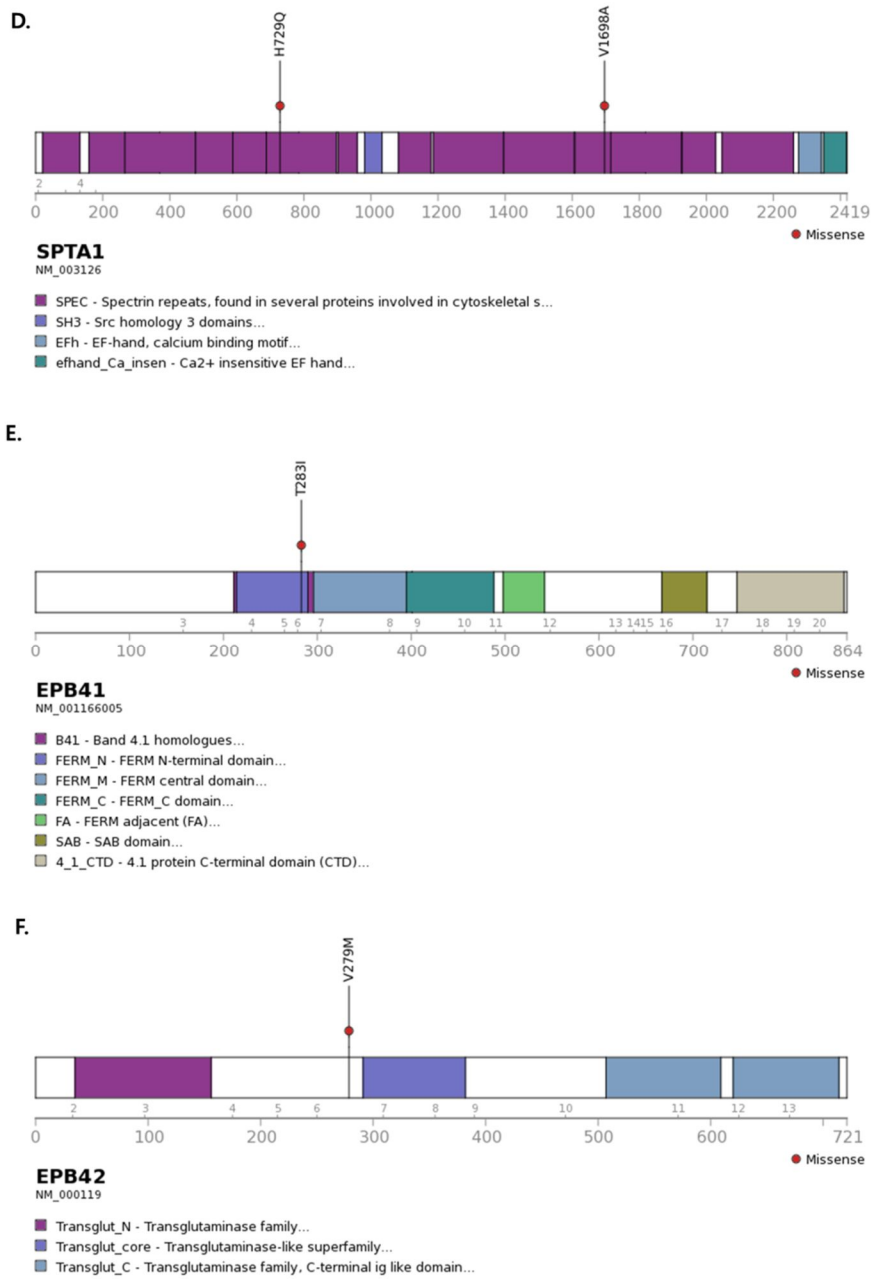


Figure 7. Characteristics of significant variants for membrane protein genes (A) *SPTB*, (B) *ANK1*, (C) *SLC4A1*, (D) *SPTA1*, (E) *EPB41*, and (F) *EPB42*.

Table 5. List of significant variants detected in membrane protein genes

Gene	Chrom	Pos	SNP ID	Variant(nucleotide)	Variant(Amino acid)	Zygosity	ACMG Classification	Patient ID
SPTB	chr14	65289812	rs121918651	c.1A>G†	p.Met1Val	Heterozygous	pathogenic	35
		65271770	.	c.187T>C	p.Trp63Arg	Heterozygous	Uncertain significance	9
		65271749	.	c.208C>T	p.Arg70*	Heterozygous	Likely pathogenic	33
		65270438	.	c.360_361insTC		Heterozygous	Likely pathogenic	44
		65270423	.	c.376C>T	p.Gln126*	Heterozygous	Likely pathogenic	40
		65270327	.	c.472C>T	p.Gln158*	Heterozygous	Likely pathogenic	58
		65268479	.	c.639_640insG		Heterozygous	Likely pathogenic	8
		65268475	.	c.643_644insC		Heterozygous	Likely pathogenic	13
		65268471	.	c.647+1G>C		Heterozygous	Uncertain significance	1
		65267587	.	c.764-1G>T		Heterozygous	Uncertain significance	49
		65260366	.	c.2014delG		Heterozygous	Likely pathogenic	52
		65259848	.	c.2533C>T‡	p.Gln845*	Heterozygous	pathogenic	50
		65253442	.	c.3241C>T	p.Gln1081*	Heterozygous	Likely pathogenic	31
		65253216	.	c.3466_3467insA		Heterozygous	Likely pathogenic	30
		65253192	.	c.3489_3490delCC		Heterozygous	Likely pathogenic	28
		65246649	.	c.4267C>T	p.Arg1423*	Heterozygous	Likely pathogenic	46
		65246625	.	c.4291C>T	p.Arg1431*	Heterozygous	Likely pathogenic	56
		65246594	.	c.4321delC		Heterozygous	Likely pathogenic	47
		65245922	.	c.4515delC		Heterozygous	Likely pathogenic	3
		65242097	.	c.4577_4586del		Heterozygous	pathogenic	23
		65241843	.	c.4842G>C	p.Lys1614Asn	Heterozygous	Uncertain significance	4
		65241215	.	c.4873C>T	p.Arg1625*	Heterozygous	Likely pathogenic	60
		65240129	.	c.4987A>T	p.Arg1663*	Heterozygous	Likely pathogenic	34
		65240100	.	c.5015delA		Heterozygous	Likely pathogenic	7
		65239958	.	c.5153_5154insG		Heterozygous	Likely pathogenic	11
		65237602	.	c.5798+1G>A		Heterozygous	Uncertain significance	20
		65235838	.	c.5938-2A>G		Heterozygous	Uncertain significance	36
		65235751	.	c.6022+1G>A		Heterozygous	Uncertain significance	59
ANK1	chr8	41583380	.	c.511A>T	p.Lys171*	Heterozygous	Likely pathogenic	12

		41583298	.	c.593A>G	p.Asn198Ser	Heterozygous	Uncertain significance	44
		41580722	.	c.830A>G†	p.His277Arg	Heterozygous	pathogenic	15
		41580696	.	c.856C>T	p.Arg286*	Heterozygous	Likely pathogenic	41
		41566369	.	c.1924delG		Heterozygous	Likely pathogenic	45
		41563729	.	c.2029C>T	p.Gln677*	Heterozygous	Likely pathogenic	6
		41563662	.	c.2095_2096insT		Heterozygous	Likely pathogenic	51
		41561982	.	c.2106C>A	p.Tyr702*	Heterozygous	Likely pathogenic	43
		41561627	.	c.2227C>T	p.Gln743*	Heterozygous	pathogenic	19,26
		41559135	.	c.2390_2393delTAGT†		Heterozygous	pathogenic	25
		41554101	.	c.2739delT		Heterozygous	Likely pathogenic	27
		41554038	.	c.2803C>T	p.Arg935*	Heterozygous	Likely pathogenic	42
		41552280	.	c.3157C>T†	p.Arg1053*	Heterozygous	pathogenic	29
		41552202	.	c.3235G>T	p.Glu1079*	Heterozygous	Likely pathogenic	37
		41550270	.	c.3754C>T†	p.Arg1252*	Heterozygous	pathogenic	16
		41547753	.	c.4095delC		Heterozygous	Likely pathogenic	2
		41547746	.	c.4103A>T	p.Lys1368Met	Heterozygous	Uncertain significance	43
		41543669	.	c.4387_4390delAACA		Heterozygous	pathogenic	18,54
		41529900	.	c.5066_5067delTG		Heterozygous	Likely pathogenic	39
SLC4A1	chr17	42334818	.	c.1526G>T	p.Gly509Val	Heterozygous	Uncertain significance	57
		42330519	rs373916826	c.2278C>T†	p.Arg760Trp	Heterozygous	pathogenic	32
		42328845	.	c.2423G>A†	p.Arg808His	Heterozygous	pathogenic	21
SPTA1	chr1	158636139	.	c.2187C>G	p.His729Gln	Heterozygous	Uncertain significance	20
		158607919	rs182430449	c.5093T>C	p.Val1698Ala	Heterozygous	pathogenic	8,9
EPB41	chr1	29342222	rs189183599	c.848C>T	p.Thr283Ile	Heterozygous	Uncertain significance	1
EPB42	chr15	43501559	rs368756068	c.835G>A	p.Val279Met	Heterozygous	Uncertain significance	35

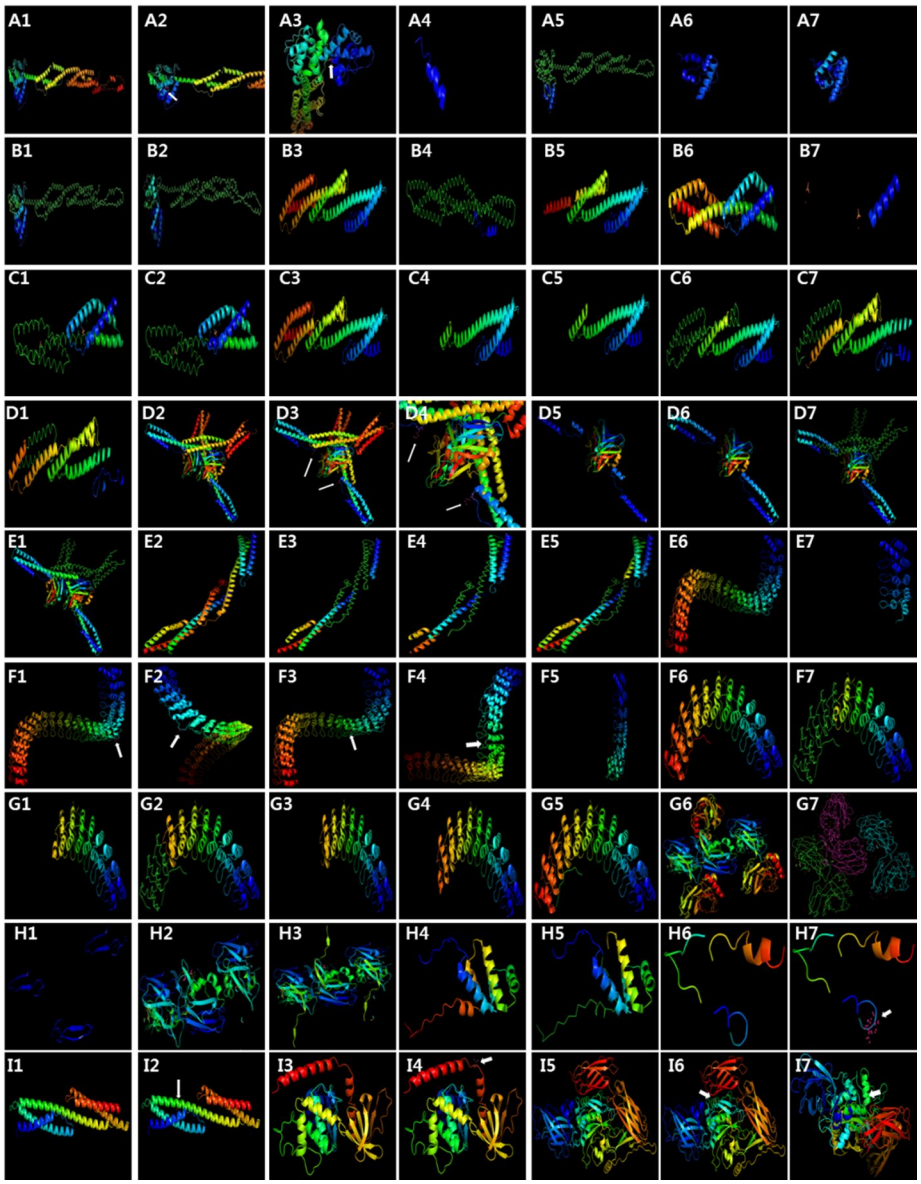


Figure 8. 3-D molecular structure of membrane proteins

3.4. Variant profile characteristics for enzyme genes

In total, 20 genes encoding enzymes were examined, 4 of which contained significant variants, *ALDOB*, *G6PD*, *GAPDH*, and *GSR* (Figure 9). One variant was observed for each gene (Table 6).

The *ALDOB* gene encodes fructose-1,6-bisphosphate aldolase. It contained a heterozygous variant that was previously reported as a SNP (rs145252200), which we further classified as a missense mutation: c.401G>A, p.Arg134His. According to ACMG guidelines, this variant was uncertain significance. One mutation was observed in *G6PD*, which encodes glucose-6-phosphate dehydrogenase. It was a novel missense mutation: c.1142T>G, p.Phe381Cys. The variant was classifies as pathogenic in ACMG guidelines, and only hemizygous variant in our study. *GAPDH* encodes glyceraldehyde-3-phosphate dehydrogenase. It contained a novel missense mutation in the HS patients: c.164A>G, p.Lys55Arg, which belongs uncertain significance category. One mutation was observed in the *GSR* gene, which encodes glutathione reductase. The variant was previously reported as a SNP (rs149225584) and we further characterized it as a missense mutation: c.1520C>T, p.Thr507Ile.

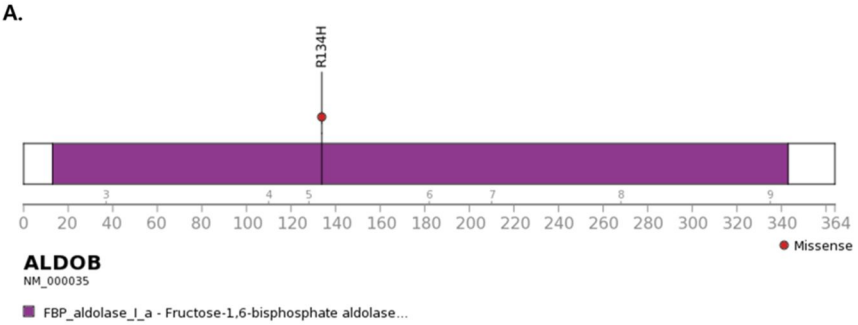
These four mutated genes were found in different four patients (Table 7). The mutated *GSR* gene was detected in a 3-year-old male patient (ID=29), who had mutated *ANK1* gene. The variant of *ANK1* gene was a known mutation in HS patients. He had no family history and no splenomegaly. Spherocytes were observed as grade 2+ in PBS, and osmotic fragility was

increased. The mutated *ALDOB* gene was found in a 18-year-old male patient (ID=30), who had mutated *SPTB* gene. The variant of *SPTB* gene was a novel frame-shift mutation. He had no family history, and presented splenomgaly. Spherocytes were observed as grade 3+ in PBS, and osmotic fragility was increased. The mutated *G6PD* gene was detected in a 3-year-old male patient (ID=48), who had no other mutated gene. He had equivocal family history of HS and presented splenomegaly. Spherocytes were not observed in PBS, and osmotic fragility was increased. The mutated *GAPDH* gene was found in a 13-year-old male patient (ID=57), who had mutated *SLC4A1* and *UGT1A1* gene. The variant of *SLC4A1* gene was a novel missense mutation, and the variant of *UGT1A1* gene was a known mutation in hyperbilirubinemia patients. He had no family history, and presented splenomgaly. Spherocytes were observed as grade 1+ in PBS, and osmotic fragility was increased.

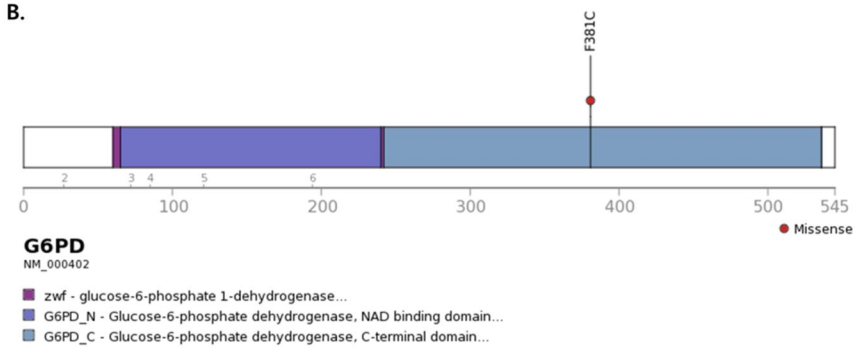
Each enzyme was visualized 3-D spatial structure and effects of gene mutations were reflected in adequate graphic style. Aldolase B, fructose-bisphosphate structure was completely constructed by one experimental template. The template was experimental structure 1QO5, which comes from fructose 1,6-bisphosphate aldolase from human liver tissue(Figure 10A1,A2). It contained amino acid sequence of Aldolase B, fructose-bisphosphate from 2 to 364. One mutation was reflected in two separated frames (Figure 10A3,A4). Glucose-6-phosphate dehydrogenase structure was completely constructed by one experimental template. The template was experimental structure 1QKI, which comes from the crystal structure reveals a structural NADP(+) molecule and provides insights into enzyme deficiency(Figure 10B1,B2). It

contained amino acid sequence of Glucose-6-phosphate dehydrogenase from 2 to 515. One mutation was reflected in two separated frames(Figure 10B3,B4). Glutathione reductase structure was partially constructed by one experimental template. The template was experimental structure 3DJG, which comes from catalytic cycle of human glutathione reductase near 1 A resolution(Figure 10C1,C2). It contained amino acid sequence of Glutathione reductase from 45 to 522. One mutation was reflected in two separated frames(Figure 10C3,C4). Glyceraldehyde-3-phosphate dehydrogenase structure was completely constructed by one experimental template. The template was experimental structure 4WNC, which comes from crystal structure of human wild-type GAPDH at 1.99 angstroms resolution(Figure 10D1). It contained amino acid sequence of Glyceraldehyde-3-phosphate dehydrogenase from 1 to 335. One mutation was reflected in one separated frame(Figure 10D2).

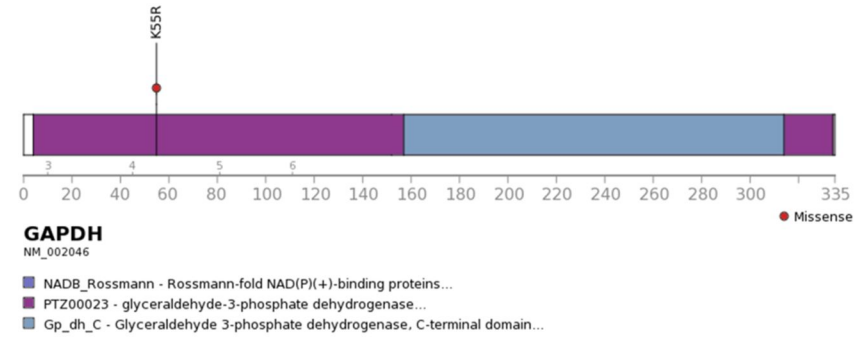
A.


R134H
3
4
5
6
7
8
9
0
20
40
60
80
100
120
140
160
180
200
220
240
260
280
300
320
340
364
Missense
ALDOB
NM_000035
FBP_aldolase_I_a - Fructose-1,6-bisphosphate aldolase...

B.


F381C
2
3
4
5
6
0
100
200
300
400
500
545
Missense
G6PD
NM_000402
zwf - glucose-6-phosphate 1-dehydrogenase...
G6PD_N - Glucose-6-phosphate dehydrogenase, NAD binding domain...
G6PD_C - Glucose-6-phosphate dehydrogenase, C-terminal domain...

C.


K55R
3
4
5
6
0
20
40
60
80
100
120
140
160
180
200
220
240
260
280
300
335
Missense
GAPDH
NM_002046
NADB_Rossmann - Rossmann-fold NAD(P)(+)-binding proteins...
PTZ00023 - glyceraldehyde-3-phosphate dehydrogenase...
Gp_dh_C - Glyceraldehyde 3-phosphate dehydrogenase, C-terminal domain...

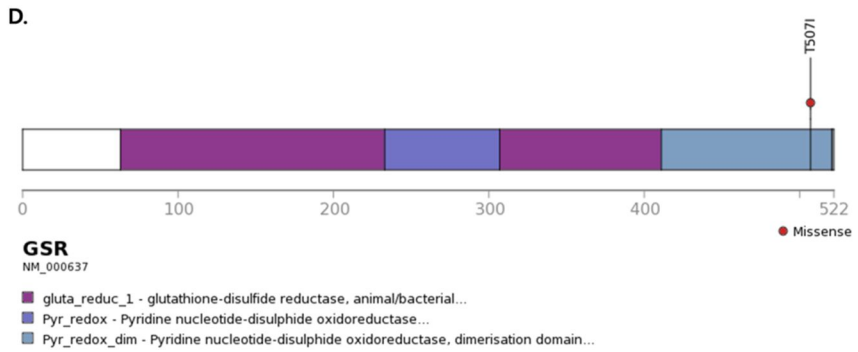


Figure 9. Characteristics of significant variants for enzyme genes (A) *ALDOB*, (B) *G6PD*, (C) *GAPDH*, and (D) G*SR*.

Table 6. List of significant variants detected in enzyme genes

Gene	Chrom	Pos	SNP ID	Variant(nucleotide)	Variant(Amino acid)	Zygosity	ACMG Classification	Patient ID
ALDOB	chr9	104189903	rs145252200	c.401G>A	p.Arg134His	Heterozygous	Uncertain significance	30
G6PD	chrX	153760927	.	c.1142T>G	p.Phe381Cys	Hemizygous	pathogenic	48
GAPDH	chr12	6645884	.	c.164A>G	p.Lys55Arg	Heterozygous	Uncertain significance	57
GSR	chr8	30537086	rs149225584	c.1520C>T	p.Thr507Ile	Heterozygous	Uncertain significance	29

Table 7. Clinical features of mutated enzyme gene detected patients

	Patient ID			
	29	**30**	**48**	**57**
Mutated Gene	ANK1, GSR	SPTB, ALDOB	G6PD	SLC4A1, GAPDH, UGT1A1
Sex	M	M	M	M
Age	3	18	3	14
Familiy history	abscent	abscent	present	abscent
Splenomegaly	abscent	present	present	present
Hg	10	11.4	6.4	12.8
MCHC	33.6	NA	31.5	37.4
Reticulocyte	4.3	6.92	4.1	3.09
T.bil	1.3	7.5	1.2	3.8
Ind.bil	0.94	6.8	0.8	NA
PBS spherocytes	2+	3+	0	1+
OFT	positive	positive	positive	positive

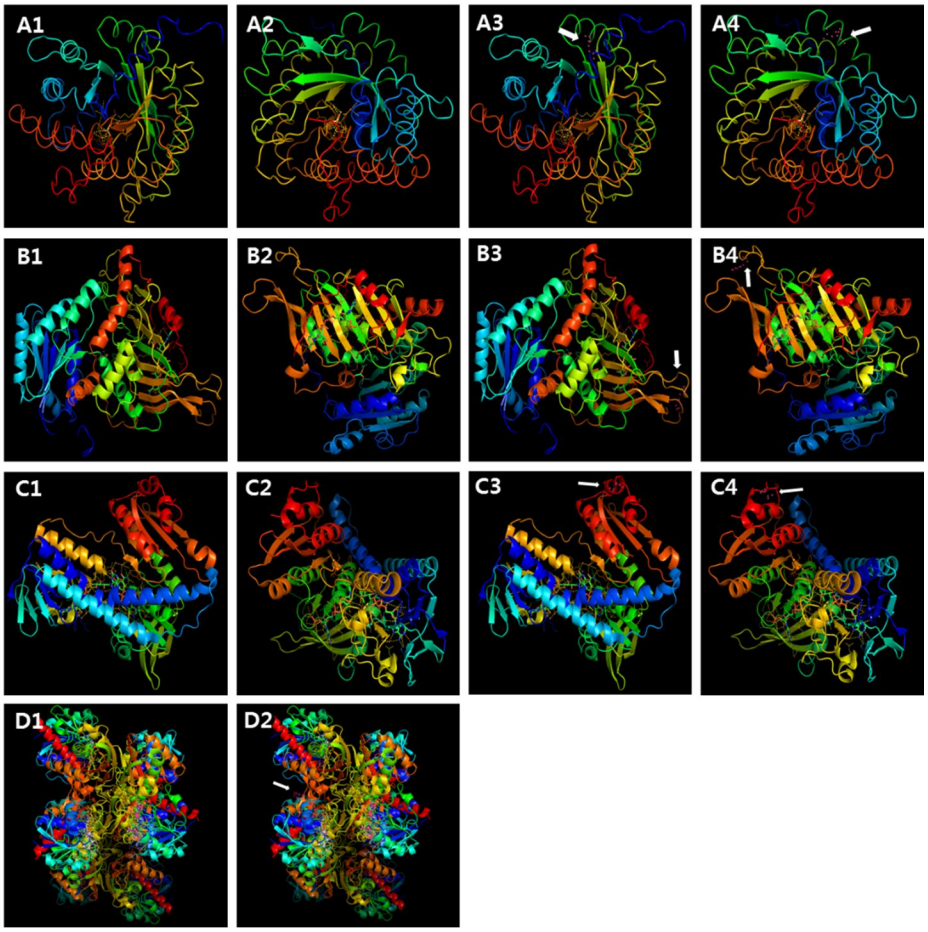


Figure 10. 3-D molecular structure of enzymes

3.5. Variants profile characteristics for other candidate genes

Six additional genes were examined, *HBA1*, *HBA2*, and *HBB* (associated with thalassemia), *SEC23B* (associated with congenital dyserythropoietic anemia), *PIGA* (associated with paroxysmal nocturnal hemoglobinuria), and *UGT1A1* (associated with Gilbert's syndrome). Significant variants were observed only in *UGT1A1* (Figure 11).

UGT1A1 encodes UDP glucuronosyltransferase 1 family, polypeptide A1. Five significant *UGT1A1* variants were detected, and all were pathogenic variants according to ACMG guidelines and previously reported variants in diseases such as hyperbilirubinemia, Gilbert's syndrome, and Crigler-Najjar syndrome (Table 8)[30, 40-48]. Two variants were previously reported in Gilbert's syndrome, both of which were missense mutations: c.686C>A, p.Pro229Gln and c.1091C>T, p.Pro364Leu. Two other variants were previously reported in hyperbilirubinemia patients, both of which were missense mutations: c.211G>A, p.Gly71Arg and c.1352C>T, p.Pro451Leu. One variant was previously reported in Crigler-Najjar syndrome. It was a missense mutation: c.1456T>G, p.Tyr486Asp.

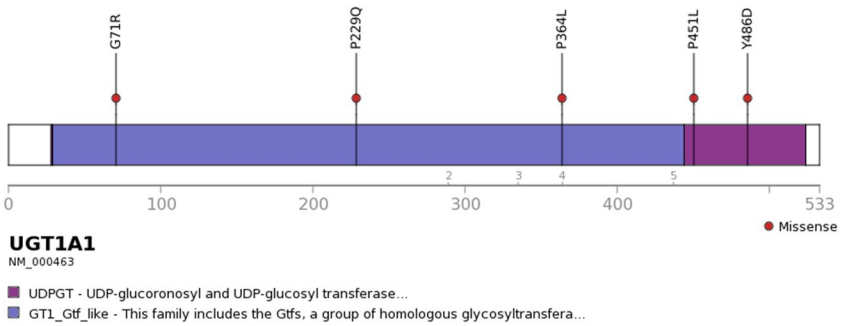


Figure 11. Significant variants diagrams for *UGT1A1* gene

Table 8. List of significant variants detected in *UGT1A1* gene

Gene	Chrom	Pos	SNP ID	Variant(nucleotide)	Variant(Amino acid)	ACMG Classification	Patient ID
UGT1A1	Chr2	234669144	rs4148323	c.211G>A	p.Gly71Arg	pathogenic	1,4,11,13,16, 17,19,20,21, 32,33,35,38, 39,43,44,45, 50,52,57,59
		234669619	rs35350960	c.686C>A	p.Pro229Gln	pathogenic	22
		234676872	rs34946978	c.1091C>T	p.Pro364Leu	pathogenic	24,32,46
		234680955	rs114982090	c.1352C>T	p.Pro451Leu	pathogenic	19
		234681059	rs34993780	c.1456T>G	p.Tyr486Asp	pathogenic	13

3.6. The development of primer sets for the diagnosis of HS by using the mutational profile

Thirty-five primer sets are necessary to assess all significant variants detected in membrane protein genes (Table 9). The total product size is 16,836 bp, and the average product size is 481 bp. The average melting temperature (Tm) is 59.9°C (range 59.0–60.4°C). For coverage of the 28 variants detected in the *SPTB* gene, 16 primer sets are needed. Twelve primer sets can cover the 19 variants detected in *ANK1*, and 3, 2, 1, and 1 primer sets are needed for variants detected in *SLC4A1*, *SPTA1*, *EPB41*, and *EPB42*, respectively.

Table 9. Primer sets for all significant variants in membrane protein

Gene	Forward	Reverse	Tm	Product size
SPTB	AACAAACGACCCAGAGACCC	TGTGAGAATGCCCTCCAACC	60	436
	CCTCTGTTCTGGTGGCATGT	GAGAGACCAGTCAGCCACAC	60	517
	GGAGGGTGGGTGCTTATGAC	GGGGGAGGAGGCTAAAAAGC	60.2	570
	GCCTCTGTGTGTGGTTAGCT	TTGTACAGAGGCAGGGAGGT	60.1	511
	ACACTCTTCCATTCTTGCTCTGT	GACATGGGTCTCAGGTGGAC	59.8	507
	TGAGGAGCTGAGCAACATGG	CTTGTGCTTTTTCCCCAGGG	59.6	502
	GGCTGCTCTCTGGTGAAGAT	AACAAAGGACATCCCAGGGC	59.9	502
	ACCCTGAGCAGAAGGAGGAT	GCCCTGAATGCCTCTCTACC	60	507
	ACCCCTTTTGCAGTGGGTAG	TAAGGGGTGAGGTGACCAGT	59.9	632
	CCCACTGTCTGACTAGGAGC	TGGAGACCCCAAAGCTACCA	59.5	452
	CCTGAGGGTGCAAGAATGGT	CCAGGTAGTACTGCTGTGCC	60	438
	GATCGACTGCCAGGACCTTG	AGTGGGTAAACAGGAGGCAC	60	434
	TCCAGAGAGGAAGGTAGGGC	ATCTGAAGGGACGTTGGCTG	60	448
	GCATATGTTGGCTCCCTGGA	GGGGTTACAGGTGTCACCAG	60	618
	GCAGAGGCCATCCAGAACAA	CCTGACTCTGGAGCGATTGG	60.2	448
	GTGCACACACTTTGGAGCTG	GCACGCTGGCTAGTTAGAGT	60	578
ANK1	TTTGGTGGAGGTAGGTCCCT	AAGAGCACGTTTCTGGGCAT	60	518
	GTGCAGTGGGAGGCAATAGT	GTGGCTGCTTTTCTGCTACG	60	561
	CAAGCCTGAGGAGCTGTCTC	TGAAATGTACGCAGGGCTGT	60	533
	CAGGAAGAGCTTTGCCTGGA	GCTTTTGTGTGGATCTGGGG	59.6	554
	TCATGTGATGCCCAGCTCAG	GGCCCTAGACACGTGCATTA	60	589
	CAGCAGTGGAGTGGATGGG	AGTCAATGTTCCCTCACCTAAGT	59.6	502
	ACGTGGCACCTGGTTTCTAG	GCAAGCTCACCTCAGGAACT	60	532
	TGTGTGAGAGGATGCTCAGC	GCACCAGACAAAAGTGTGGG	59.7	536
	CTCCCGTCTGGATGGAAAGG	CCTCGTCTACAGAGGGAGGA	59.7	404
	TTCCCTAAACCAGACAGCCG	CCAGCACTGCACTGCAAATT	59.9	412
	CTCAGAGCTCCCTGTGCTC	GTGGGGACTACAGGCATGAG	59.7	530
	TTGTGGAGGACGACACAGTG	CATACCTGGTGGGGTCTTGG	59.8	452
SLC4A1	GGTAGTTCTGTCCTTGCCCC	TCCTGATCTCGGGTGATCCA	59.9	444
	AAACCTGAGCGCAAGATGGT	CCTTTCCTCCAGGATCCCCT	60.3	413
	GCTCCATATGGTGCCTGTGT	GATGATCTGGATGCCCGTGA	59.8	409

SPTA1	ATGTCTTCTCCCAGGGACCC	ACCAAAGACACTGAGCACCC	60.4	507
	TCTTTGAGCCAGACCAGCTG	AGCTGAGGCAAACATGAAGT	59	431
EPB41	AATCGCTTGAACCTGGGAGG	ACCAGAAACTTTCAATCCTAGGCT	60	450
EPB42	TGATGAGCCCTCCTTACCCA	GCTGCCTGGGACAACTACTT	60	407

4. DISCUSSION

HS is the most common and well-studied disease in hereditary hemolytic anemia. Various tests are used to diagnose HS, but most are indirect methods to detect RBC membrane protein defects. Gene mutation analysis is simple and direct method, while it has not been widely studied. In this study, we analyzed 17 membrane protein encoded genes in 61 Korean HS patients. Among 61 patients, 50 (82%) had at least one significant variant in a membrane protein gene. In 6 membrane protein genes (*SPTB, ANK1, SPTA1, SLC4A1, EPB41*, and *EPB42*), total 54 variants were detected. Eight variants were previously reported as pathogenic mutation in HS, and 46 variants were novel. Novel variants were considered to be related HS through many predictable tools. The most variants (28) were found in *SPTB*, and followed by *ANK1* (19), *SLC4A1* (3), *SPTA1* (2), and *EPB41* (1), *EPB42* (1). This result has both similarities and differences compared to previous USA result[49]. USA study examined only 4 genes, *SPTA, SPTB, ANK1* and *SLC4A1*. Spectrin encoding gene contained the most variants in both studies. *SPTB* and *SPTA* gene mutation rate were 51% and 5%, respectively in Korea, and each was same as 45% in USA (figure 12). *ANK1* gene mutation rate was 33% in Korea, and 10% in USA. However, *SLC4A1* gene mutation was not reported in USA study.

These differences were detected in previous studies using SDS-PAGE methods[50-55]. RBC membrane protein defects in Korean patients with HS exhibited different characteristics than those observed in European, American,

and Japanese patients examined in previous studies (Table 10). In Italy, single defects in band 3 and spectrin are the primary variants reported, and a combined defect in spectrin/ankyrin is frequently detected in patients in the USA and Spain. The incidence of HS among individuals in Asian countries is highest in Japan, and the defect in the 4.2 protein is more frequent in Japan than in the USA and Europe. A Korean report from 2000 showed a similar pattern to that observed in Japan, where 4.2 protein defects were detected at a higher frequency than that of band 3 in the USA and Europe (Figure 13). However, a single ankyrin defect is most frequently observed in Korean patients, unlike in Japan and other countries[50-55].

On the other hand, a sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis showed largely different results from other reports for the same country, but the frequency of cases for which a defect was not observed was also high. This may be due to complexity of SDS-PAGE methods and the subjective, complicated interpretation of results. In contrast, the nucleotide sequence analysis that we performed was straightforward and the interpretation of results was objective. Therefore, it can be considered a more reliable method than SDS-PAGE. In this study, we performed a sequence analysis for almost all genes encoding RBC membrane proteins among other candidate genes to differentially diagnose HS; such an analysis has not been performed in other countries to date.

In previous studies, membrane protein defects have mainly been observed in the spectrin, ankyrin, and band 3 proteins[14-19]. In this study, we detected *SPTB*, *ANK1*, and *SLC4A1* variants, which encode the spectrin, ankyrin, and

band 3 proteins, respectively. However, we found some differences with the results of previous studies. In an analysis of RBC membrane proteins using SDS-PAGE in Korean patients in 2000, most variants were detected in the gene encoding ankyrin, which most variants in our study were in the gene encoding spectrin. In addition, band 3 and band 4.2 protein defects were detected in 16% and 22% of cases, respectively, in the SDS-PAGE study, but in only 5% and 2% of cases, respectively, in our analysis of the genes encoding those proteins. Additionally, band 4.1 protein defects were not detected in the SDS-PAGE analysis, while variants of *EPB41* gene, which encodes the band 4.1 protein, was detected in 4% of cases in our study. Despite the different patients groups between studies, it is clear that the results previous SDS-PAGE analysis and our genetic analysis had both similarities and differences. The distribution of variants in our study was similar to that of a study conducted in the USA, which showed that the majority of variants are in the genes encoding spectrin, followed by genes encoding ankyrin. However, the mutation frequencies were different and variants were not detected in *SLC4A1* in the USA study.

Inherited pattern of HS is different depending on the gene. In *SPTB*, HS inherited in autosomal dominant or de novo pattern. In *ANK1*, HS in known to be inherited in all three patterns, autosomal dominant, autosomal recessive, and de novo. Autosomal dominant or autosomal recessive pattern is reported in *SPTA1*, and only autosomal dominant pattern was reported in *SLC4A1*. In *EPB42*, HS inherited in autosomal recessive pattern. And inherited pattern is not clearly revealed in *EPB41*. All significant variants detected in membrane

protein coding genes were heterozygous. So we can predict that significant variants in *SPTB, ANK1, SPTA1*, and *SLC4A1* gene are inherited as autosomal dominant manner and are likely cause of HS. However, Variants detected in *EPB41* and *EPB42* gene have low potential in direct cause of HS. Actually, each patients who has variant in *EPB41* and *EPB42* gene, has another variant in SPTB gene.

Variants detected in enzyme coding genes were also heterozygous except in *G6PD* gene. Inherited pattern of fructose intolerance which caused by *ALDOB* gene defect is autosomal recessive, and other enzyme deficiency caused by *GSR, GAPDH* and *G6PD* gene is not clearly revealed. Therefore, heterozygous variants detected in *ALDOB, GSR*, and *GAPDH* were thought to be low potential for each enzyme deficiency. Variant detected in *G6PD* gene was hemizygous, and this might be cause glucose phosphate dehydrogenase deficiency. Although the patient, who has *G6PD* gene variant, has no variant in membrane coding gene, HS cannot be ruled out HS because of other clinical and laboratory results including family history and OFT.

It is the limitation of this study that we couldn't investigate protein change through direct methods. Instead, we visualized 3-D spatial structure and expect the effects of gene mutations for both membrane protein and enzyme encoding genes (Figure 8, 10). Although exact changes of protein structure cannot be known by 3-D spatial structure, large scale modification of the protein due to frame shift or nonsense mutation can be visualized and subsequent functional changes can be expected from these images.

To investigate clinical utility, we compared the result of the gene test to OFT and PBS tests (Table 11). Among 46 patients who have OFT test results, 31 patients(68%) showed positivity in both OFT and gene test, while 2 patients (4%) showed negative results in both OFT and gene test. Eight patients (17%) had positive OFT result with no gene mutation, and 5 patients (11%) had gene mutation with negative OFT result. In that, 13 patients (29%) showed mutually exclusive result. Therefore, complementary use of these 2 tests can give higher sensitivity and specificity for diagnosing HS. In the comparison of gene test and PBS test, 12 patients (20%) showed mutually exclusive result, thus, these 2 tests can be used as complement method.

Meanwhile, to find out the correlation between gene test results and disease severity, mean hemoglobin concentrations were compared according to 4 genes (*SPTB, ANK1, SPTA1, SLC4A1*) (Figure 14). The patients with *SPTB* variant had mean hemoglobin concentration of 7.9 g/dL, which is lower than that of the patients with other gene variant (8.9 g/dL). However, there was no statistical significance (*P* value was 0.125). The patients with or without *ANK1* variant showed mean hemoglobin concentration of 8.6, 8.1 g/dL, respectively, and with or without *SPTA1* mutation showed 8.4, 7.2 g/dL respectively, which showed no statistical significance (*P* value was 0.493, 0.416, respectively). There was statistical significance in *SLC4A1* gene (*P* value was 0.022), but the patients with the mutation had hemoglobin concentration of 11.3 g/dL which is higher than that of the patients without the (8.1 g/dL), therefore, it had no clinical significance.

To investigate the differences in disease severity according to pathogenicity, we compared mean hemoglobin concentrations (Figure 15). Patients with pathogenic variants showed higher mean hemoglobin concentrations (8.8 g/dL) than patients with likely pathogenic or uncertain significant variants (8.1 g/dL), and had no statistical significance (*P* value was 0.4). Mean hemoglobin concentrations of patients with pathogenic or likely pathogenic variants was 8.2 g/dL, while patients who have only uncertain significant variants showed 8.7 g/dL, which was also not statistically significant (*P* value was 0.63).

To find out if there are differences in disease severity according to number of variants, we compared mean hemoglobin concentrations (Figure 16). Mean hemoglobin concentration of patients who had no variant was 8.9 g/dL, and those of patients who had 1 variant and 2 variants were 8.1 and 8.7 g/dL, respectively. However, there was no statistical significance.

Of the 61 patients examined in this study, 23 (38%) had significant *UGT1A1* gene variants. Variants in the *UGT1A1* gene, which is causally related to Gilbert's syndrome, are observed in approximately 3–10% of the general population, and some studies have shown that mutations at this locus are more frequent in HS patients than in the general population[56-58]. In our study, the frequency of *UGT1A1* mutations was much higher than that of previous studies. We found that total bilirubin, indirect bilirubin concentration, and the incidence of jaundice were significantly higher in patients with *UGT1A1* mutations. Therefore, if *UGT1A1* is included in the diagnostic gene panel for HS and UDP-glucuronosyltransferase enzyme is used as a supplementary treatment for patients with the mutation, the incidence of jaundice and

gallstones is expected to be reduced. However, most previous studies have reported variants in the *UGT1A1* promoter region, and all variants detected in our study were exonic. Accordingly, further studies are required to explore this difference.

Among 61 HS patients, 11 (18%) patients had no significant variant in RBC membrane genes. These patients show similar baseline characteristics in most aspects compared to those of total 61 patients (Table 12). Although, sex ratio was inverted, it had no statistical significance. Mean age was 8 years, and proportion of family history, clinical symptoms, Grading of peripheral spherocytes, and Osmotic fragility tests were not significantly different those of total patients group. Other laboratory test results including hematologic parameters, markers of hemolysis, and iron status parameters showed no statistical significance. The reason of detecting no significant variant is probably because significant variant of these patients exist in intron, not exon. Another possible cause we can think is that causative variants are on other genes. We examined 17 membrane protein coding genes. If other membrane protein related genes were defected, we could not find them. Further study including whole exome sequencing or whole genome sequencing might be find significant variants in these 10 HS patients.

As described above, mutations in RBC membrane genes were observed in 50 out of 61 patients (82%) diagnosed with HS. This finding indicates a close correlation between clinical diagnosis and gene mutations. Additionally, tests for genetic mutations are typically more sensitive than OFT or EMA binding tests, which are currently used for screening. However, some HS patients

show mutually exclusive results in OFT and genetic test. In addition, some HS patients, who had no spherocytes in PBS, harbor membrane gene mutation (Table 11). Therefore, mutation analyses are thought to be complementary method for diagnosing HS. Also, gene mutation profile in Korean HS patients showed difference in that of other countries. So, gene panel kit for diagnosing Korean HS patients can be manufactured using 54 variants defined in this study and several known mutations from previous studies. This could provide a cheaper, faster, and more convenient method than current strategies for diagnosis. Therefore, in addition to the diagnostic guidelines suggested by international working parties, we suggest that gene test should be conducted with other special tests including OFT and EMA-binding tests (Figure 17).

However, the variants identified in this study and their causal link with HS should be confirmed. Segregation analysis based on family study is one of the widely used methods to confirm pathogenicity. So, further study including family members of HS patients should be conducted to validate the significance of variants. On the other hands, to prove relationship between genetic mutation and protein defect, researchers should consider knockout gene study or other prospective studies. Moreover, to determine clinical utilities of genetic tests in HS patients, studies for the relationship between genetic tests and disease severity including splenectomy results should be performed in many countries.

In conclusion, this was the first large-scaled genetic study of Korean HS patients, and we detected 54 significant variants including 46 novel variants. We demonstrated that gene test is sensitive and might be used as

complementary method for diagnosing HS. Consequently, we determined gene panel with 35 primer sets and suggested that gene test should be conducted with other special tests.

Table 10. SDS-PAGE results of RBC membrane protein in HS patients

Membrane protein	Italy1[51] (n=300)	Italy2[54] (n=87)	USA1[52] (n=166)	USA2[6] (n=55)	Spain[53] (n=62)	Japan1[59] (n=47)	Japan2[55] (n=60)	Korea[50] (n=27)
Band 3	53	26	23	18	0	27	12	11
Spectrin only	32	43	0	13	30	8	0	7
Ankyrin only	0	0	0	0	6	4	0	30
Spectrin/ankyrin	4	19	60	11	58	0	4	4
4.2 protein	1	6	2	0	0	14	27	15
Undetected	10	6	15	58	5	47	17	33

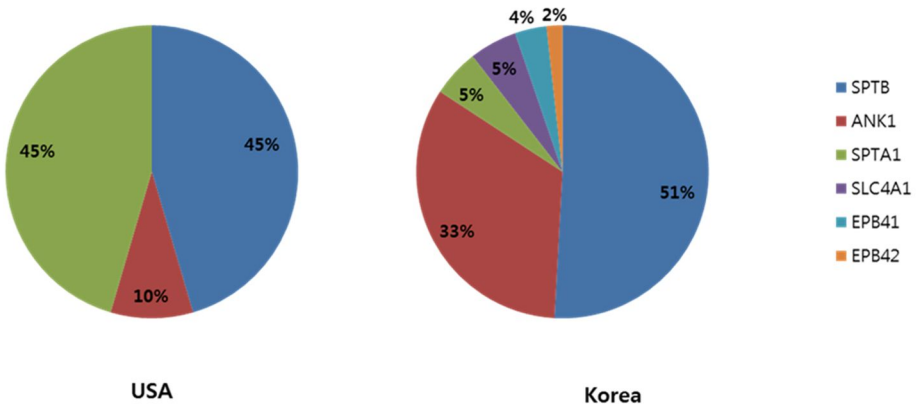


Figure 12. Comparison of gene study result between USA and Korean HS patients

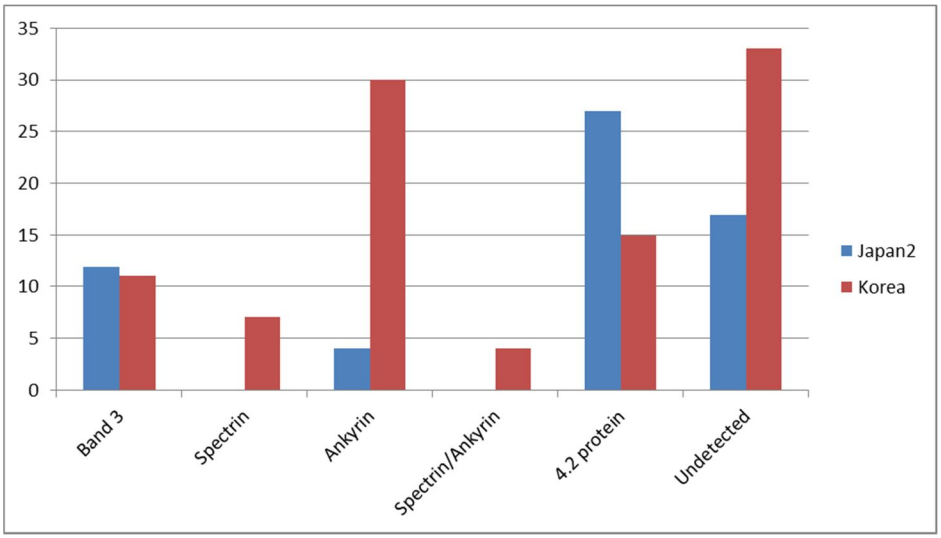


Figure 13. Comparison of SDS-PAGE results between Japanese and Korean HS patients

Table 11. Comparison of Osmotic Fragility Test , PBS and Genetic Test results in HS patients

		Membrane protein gene	
		Mutated	Non-mutated
Osmotic Fragility Test (N = 46)	Positive	31 (68%)	8 (17%)
	Negative	5 (11%)	2 (4%)
Peripheral Blood Smear (N = 60)	Positive	46 (77%)	8 (13%)
	Negative	4 (7%)	2 (3%)

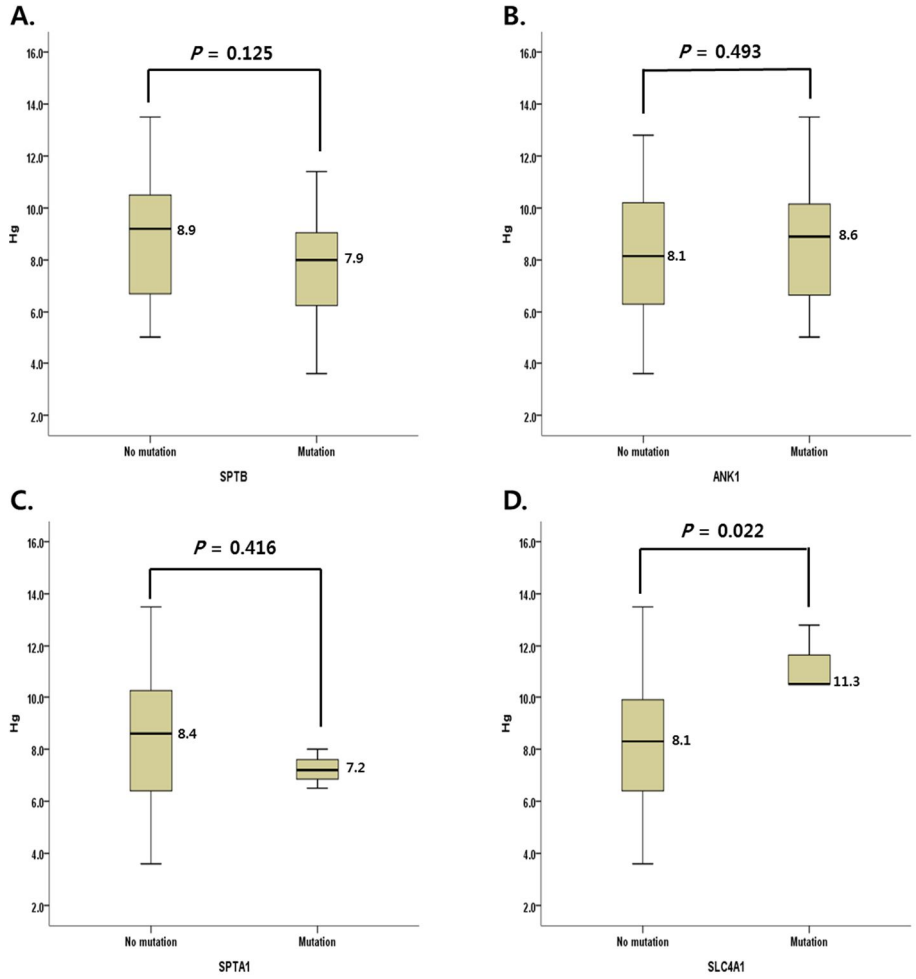


Figure 14. Comparison of mean hemoglobin concentration according to gene mutation. (A) *SPTB*, (B) *ANK1*, (C) *SPTA1*, and (D) *SCL4A1* gene

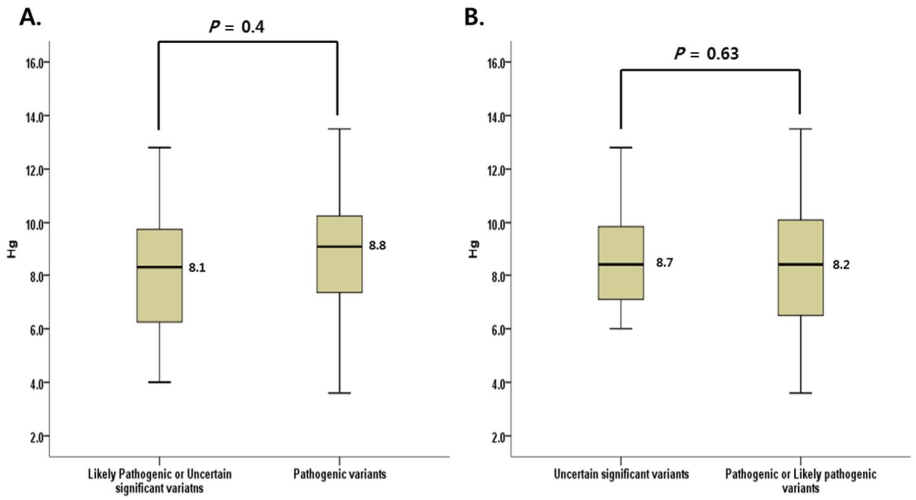


Figure 15. Comparison of mean hemoglobin concentration according to pathogenicity of variants. (A) Pathogenic vs likely pathogenic or uncertain significant variants, (B) Pathogenic or likely pathogenic vs uncertain significant variants

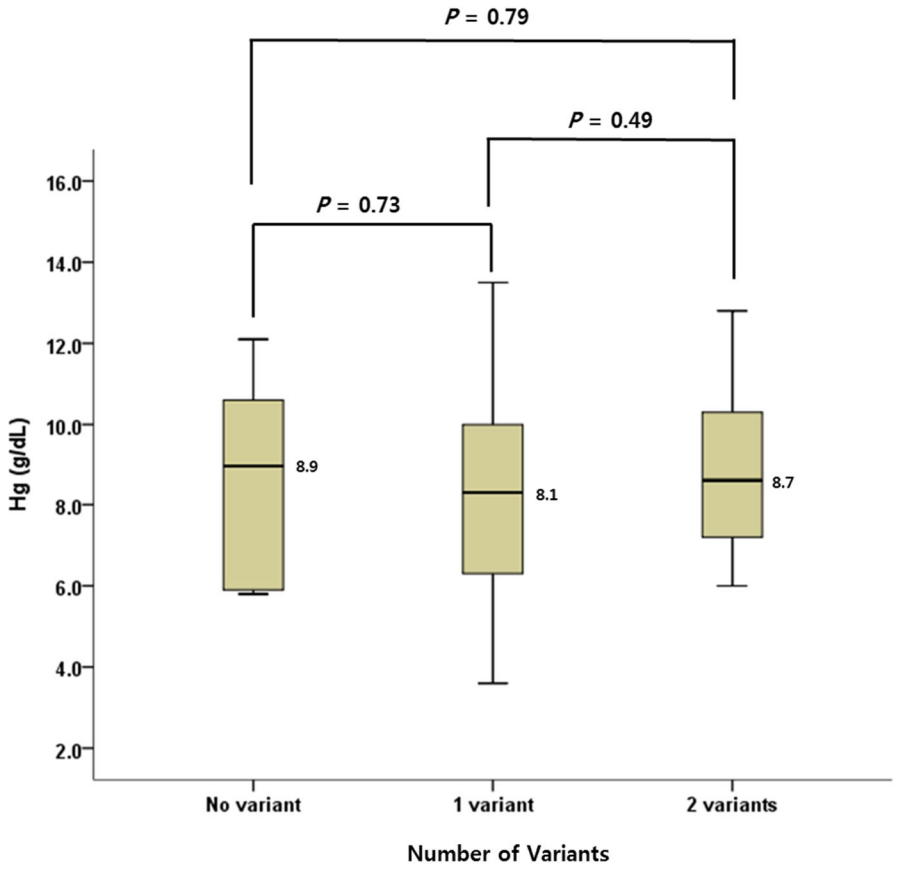


Figure 16. Comparison of mean hemoglobin concentration according to number of variants.

Table 12. Baseline characteristics of 11 patients, who had no significant variant in RBC membrane genes

Characteristics	HS Patients
Sex (%)	
Male	5 (45)
Female	6 (55)
Age (years)	
Mean	8
Range	2 ~ 17
Family History (%)	
Positive	5 (45)
Negative	6 (55)
Clinical symptoms (%)	
Splenomegaly	8 (73)
Jaundice	7 (64)
Hepatomegaly	1 (9)
Hematologic parameters (range)	
Hemoglobin (g/dL)	8.5 (5.8-12.1)
MCV (fL)	83.2 (62.0-107.0)
MCHC (g/dL)	35.0 (31.5-37.9)
Markers of hemolysis (range)	
Reticulocyte count (%)*	6.4 (0.3-13.3)
Total Bilirubin (mg/dL)*	3.9 (0.4-15.3)
Indirect Bilirubin (mg/dL)*	4.0 (0.7-14.6)
LDH (IU/L)*	363 (198-730)

Iron status parameters (range)	
Iron (ug/dL) *	114.3 (51.0-245.0)
TIBC (ug/dL) *	252.4 (207.0-280.0)
Ferritin (ng/mL) *	291.5 (55.1-918.0)
Grading of Peripheral Spherocytes	Number of patients(%)
0	3 (27)
1+	2 (19)
2+	5 (45)
3+	1 (9)
NA	0 (0)
Osmotic fragility tests	Number of patients(%)
Positive	8 (72)
Negative	2 (19)
NA	1 (9)

*: calculated except missing data

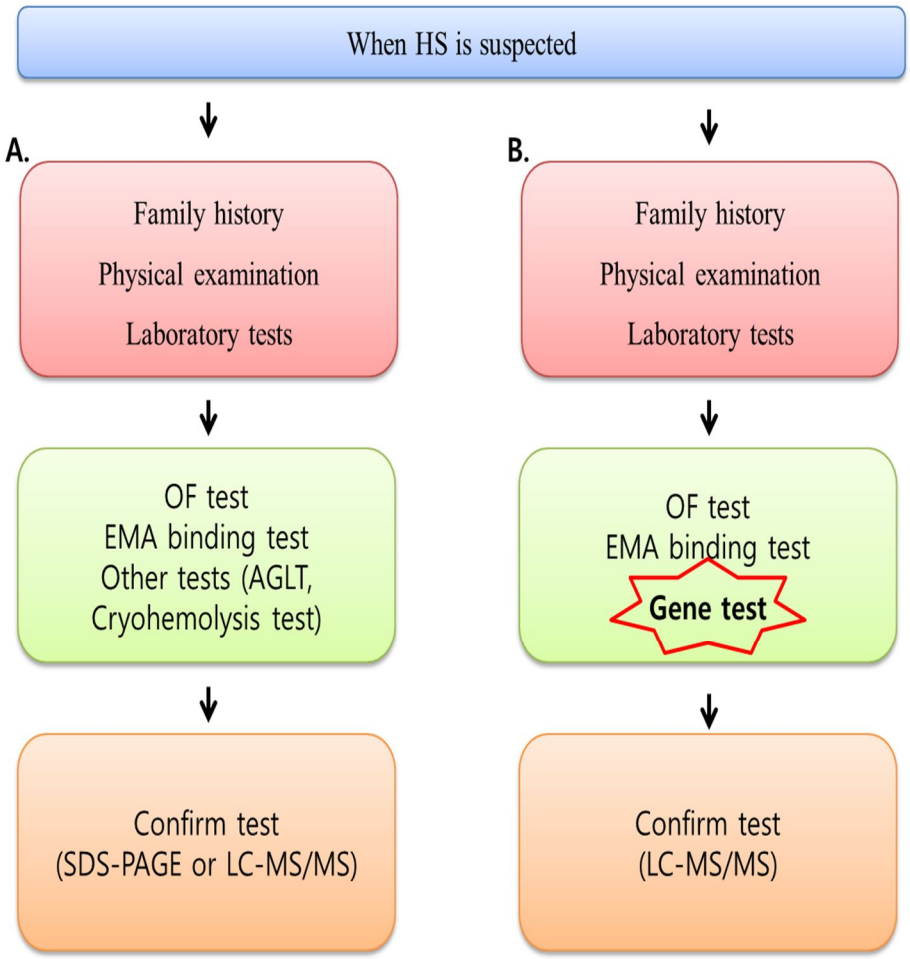


Figure 17. Diagnostic work flow for HS suggested by (A) international working parties and (B) this study

REFERENCES

1. Perrotta S, Gallagher PG, Mohandas N: **Hereditary spherocytosis**. *Lancet* 2008, **372**(9647):1411-1426.
2. Iolascon A, Miraglia del Giudice E, Perrotta S, Alloisio N, Morle L, Delaunay J: **Hereditary spherocytosis: from clinical to molecular defects**. *Haematologica* 1998, **83**(3):240-257.
3. Da Costa L, Galimand J, Fenneteau O, Mohandas N: **Hereditary spherocytosis, elliptocytosis, and other red cell membrane disorders**. *Blood reviews* 2013, **27**(4):167-178.
4. Barcellini W, Bianchi P, Fermo E, Imperiali FG, Marcello AP, Vercellati C, Zaninoni A, Zanella A: **Hereditary red cell membrane defects: diagnostic and clinical aspects**. *Blood transfusion = Trasfusione del sangue* 2011, **9**(3):274-277.
5. Jung HL: **A new paradigm in the diagnosis of hereditary hemolytic anemia**. *Blood research* 2013, **48**(4):237-239.
6. Cynober T, Mohandas N, Tchernia G: **Red cell abnormalities in hereditary spherocytosis: relevance to diagnosis and understanding of the variable expression of clinical severity**. *The Journal of laboratory and clinical medicine* 1996, **128**(3):259-269.
7. Bolton-Maggs PH, Langer JC, Iolascon A, Tittensor P, King MJ, General Haematology Task Force of the British Committee for Standards in H: **Guidelines for the diagnosis and management of hereditary spherocytosis--2011 update**. *British journal of haematology* 2012, **156**(1):37-49.
8. Park SH, Park CJ, Lee BR, Cho YU, Jang S, Kim N, Koh KN, Im HJ, Seo JJ, Park ES *et al*: **Comparison study of the eosin-5'-maleimide binding test, flow cytometric osmotic fragility test, and cryohemolysis test in the diagnosis of hereditary spherocytosis**. *American journal of clinical pathology* 2014, **142**(4):474-484.
9. King MJ, Garcon L, Hoyer JD, Iolascon A, Picard V, Stewart G, Bianchi P, Lee SH, Zanella A, International Council for

Standardization in H: **ICSH guidelines for the laboratory diagnosis of nonimmune hereditary red cell membrane disorders.** *International journal of laboratory hematology* 2015, **37**(3):304-325.

10. Constantino BT: **Reporting and grading of abnormal red blood cell morphology.** *International journal of laboratory hematology* 2015, **37**(1):1-7.
11. Zanella A, Izzo C, Rebulla P, Zanuso F, Perroni L, Sirchia G: **Acidified glycerol lysis test: a screening test for spherocytosis.** *British journal of haematology* 1980, **45**(3):481-486.
12. Streichman S, Gescheidt Y: **Cryohemolysis for the detection of hereditary spherocytosis: correlation studies with osmotic fragility and autohemolysis.** *American journal of hematology* 1998, **58**(3):206-212.
13. Shim YJ, Won DI: **Flow cytometric osmotic fragility testing does reflect the clinical severity of hereditary spherocytosis.** *Cytometry Part B, Clinical cytometry* 2014, **86**(6):436-443.
14. Won DI, Suh JS: **Flow cytometric detection of erythrocyte osmotic fragility.** *Cytometry Part B, Clinical cytometry* 2009, **76**(2):135-141.
15. Kar R, Mishra P, Pati HP: **Evaluation of eosin-5-maleimide flow cytometric test in diagnosis of hereditary spherocytosis.** *International journal of laboratory hematology* 2010, **32**(1 Pt 2):8-16.
16. Fairbanks G, Steck TL, Wallach DF: **Electrophoretic analysis of the major polypeptides of the human erythrocyte membrane.** *Biochemistry* 1971, **10**(13):2606-2617.
17. Laemmli UK: **Cleavage of structural proteins during the assembly of the head of bacteriophage T4.** *Nature* 1970, **227**(5259):680-685.
18. Metzker ML: **Sequencing technologies - the next generation.** *Nature reviews Genetics* 2010, **11**(1):31-46.
19. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR,

Bertoni A, Swerdlow HP, Gu Y: **A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers**. *BMC genomics* 2012, **13**:341.

20. Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, Howard E, Shendure J, Turner DJ: **Target-enrichment strategies for next-generation sequencing**. *Nature methods* 2010, **7**(2):111-118.

21. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J *et al*: **From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline**. *Current protocols in bioinformatics / editoral board, Andreas D Baxevanis [et al]* 2013, **11**(1110):11 10 11-11 10 33.

22. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M *et al*: **The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data**. *Genome research* 2010, **20**(9):1297-1303.

23. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform**. *Bioinformatics* 2009, **25**(14):1754-1760.

24. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM: **A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3**. *Fly* 2012, **6**(2):80-92.

25. Liu X, Jian X, Boerwinkle E: **dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations**. *Human mutation* 2013, **34**(9):E2393-2402.

26. Liu X, Jian X, Boerwinkle E: **dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions**. *Human mutation* 2011, **32**(8):894-899.

27. Reva B, Antipin Y, Sander C: **Predicting the functional impact of protein mutations: application to cancer genomics**. *Nucleic acids research* 2011, **39**(17):e118.

28. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR: **A method and server for predicting damaging missense mutations**. *Nature methods* 2010, **7**(4):248-249.

29. Kumar P, Henikoff S, Ng PC: **Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm**. *Nature protocols* 2009, **4**(7):1073-1081.

30. Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA: **A map of human genome variation from population-scale sequencing**. *Nature* 2010, **467**(7319):1061-1073.

31. Basseres DS, Vicentim DL, Costa FF, Saad ST, Hassoun H: **Beta-spectrin Promiss-ao: a translation initiation codon mutation of the beta-spectrin gene (ATG --> GTG) associated with hereditary spherocytosis and spectrin deficiency in a Brazilian family**. *Blood* 1998, **91**(1):368-369.

32. Maillet P, Alloisio N, Morle L, Delaunay J: **Spectrin mutations in hereditary elliptocytosis and hereditary spherocytosis**. *Human mutation* 1996, **8**(2):97-107.

33. Ozcan R, Jarolim P, Lux SE, Ungewickell E, Eber SW: **Simultaneous (AC)n microsatellite polymorphism analysis and single-stranded conformation polymorphism screening is an efficient strategy for detecting ankyrin-1 mutations in dominant hereditary spherocytosis**. *British journal of haematology* 2003, **122**(4):669-677.

34. Nakanishi H, Kanzaki A, Yawata A, Yamada O, Yawata Y: **Ankyrin gene mutations in japanese patients with hereditary spherocytosis**. *International journal of hematology* 2001, **73**(1):54-63.

35. Leite RC, Basseres DS, Ferreira JS, Alberto FL, Costa FF, Saad

ST: **Low frequency of ankyrin mutations in hereditary spherocytosis: identification of three novel mutations**. *Human mutation* 2000, **16**(6):529.

36. Eber SW, Gonzalez JM, Lux ML, Scarpa AL, Tse WT, Dornwell M, Herbers J, Kugler W, Ozcan R, Pekrun A *et al*: **Ankyrin-1 mutations are a major cause of dominant and recessive hereditary spherocytosis**. *Nature genetics* 1996, **13**(2):214-218.

37. Yawata A, Kanzaki A, Gilsanz F, Delaunay J, Yawata Y: **A markedly disrupted skeletal network with abnormally distributed intramembrane particles in complete protein 4.1-deficient red blood cells (allele 4.1 Madrid): implications regarding a critical role of protein 4.1 in maintenance of the integrity of the red blood cell membrane**. *Blood* 1997, **90**(6):2471-2481.

38. Jarolim P, Rubin HL, Brabec V, Chrobak L, Zolotarev AS, Alper SL, Brugnara C, Wichterle H, Palek J: **Mutations of conserved arginines in the membrane domain of erythroid band 3 lead to a decrease in membrane-associated band 3 and to the phenotype of hereditary spherocytosis**. *Blood* 1995, **85**(3):634-640.

39. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E *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**. *Genetics in medicine : official journal of the American College of Medical Genetics* 2015, **17**(5):405-424.

40. Chen Z, Su D, Ai L, Jiang X, Wu C, Xu Q, Wang X, Fan Z: **UGT1A1 sequence variants associated with risk of adult hyperbilirubinemia: a quantitative analysis**. *Gene* 2014, **552**(1):32-38.

41. Mimura Y, Maruo Y, Ohta Y, Sato H, Takeuchi Y: **Effect of common exon variant (p.P364L) on drug glucuronidation by the human UDP-glucuronosyltransferase 1 family**. *Basic & clinical pharmacology & toxicology* 2011, **109**(6):486-493.

42. Maruo Y, Ozgenc F, Mimura Y, Ota Y, Matsui K, Takahashi H, Mori A, Taga T, Takano T, Sato H *et al*: **Compound heterozygote of a novel missense mutation (p.K402T) and a double missense mutation (p.[G71R;Y486D]) in type II Crigler-Najjar syndrome**. *Journal of pediatric gastroenterology and nutrition* 2011, **52**(3):362-365.

43. Teng HC, Huang MJ, Tang KS, Yang SS, Tseng CS, Huang CS: **Combined UGT1A1 and UGT1A7 variant alleles are associated with increased risk of Gilbert's syndrome in Taiwanese adults**. *Clinical genetics* 2007, **72**(4):321-328.

44. Kaniwa N, Kurose K, Jinno H, Tanaka-Kagawa T, Saito Y, Saeki M, Sawada J, Tohkin M, Hasegawa R: **Racial variability in haplotype frequencies of UGT1A1 and glucuronidation activity of a novel single nucleotide polymorphism 686C> T (P229L) found in an African-American**. *Drug metabolism and disposition: the biological fate of chemicals* 2005, **33**(3):458-465.

45. Fu WP, Liu Y: **[Role of genetic factors in occurrence of neonatal jaundice in Guangxi region]**. *Zhonghua er ke za zhi Chinese journal of pediatrics* 2005, **43**(10):743-747.

46. Takeuchi K, Kobayashi Y, Tamaki S, Ishihara T, Maruo Y, Araki J, Mifuji R, Itani T, Kuroda M, Sato H *et al*: **Genetic polymorphisms of bilirubin uridine diphosphate-glucuronosyltransferase gene in Japanese patients with Crigler-Najjar syndrome or Gilbert's syndrome as well as in healthy Japanese subjects**. *Journal of gastroenterology and hepatology* 2004, **19**(9):1023-1028.

47. Gagne JF, Montminy V, Belanger P, Journault K, Gaucher G, Guillemette C: **Common human UGT1A polymorphisms and the altered metabolism of irinotecan active metabolite 7-ethyl-10-hydroxycamptothecin (SN-38)**. *Molecular pharmacology* 2002, **62**(3):608-617.

48. Aono S, Yamada Y, Keino H, Hanada N, Nakagawa T, Sasaoka Y, Yazawa T, Sato H, Koiwai O: **Identification of defect in the genes**

for bilirubin UDP-glucuronosyl-transferase in a patient with Crigler-Najjar syndrome type II. *Biochemical and biophysical research communications* 1993, **197**(3):1239-1244.

49. Pei Hui M, PhD, Vincent Shultz, PhD, Michael Krauthammer, PhD, Jon Morrow, MD, PhD.: **Yale Blood Cell Disease Reference Laboratory Programs: Rapid Mutation Scanning by Next Generation Sequencing**. *Poster presentation* 2012.

50. Lee YK, Cho HI, Park SS, Lee YJ, Ra E, Chang YH, Hur M, Shin HY, Ahn HS: **Abnormalities of erythrocyte membrane proteins in Korean patients with hereditary spherocytosis**. *Journal of Korean medical science* 2000, **15**(3):284-288.

51. Mariani M, Barcellini W, Vercellati C, Marcello AP, Fermo E, Pedotti P, Boschetti C, Zanella A: **Clinical and hematologic features of 300 patients affected by hereditary spherocytosis grouped according to the type of the membrane protein defect**. *Haematologica* 2008, **93**(9):1310-1317.

52. Jarolim P, Murray JL, Rubin HL, Taylor WM, Prchal JT, Ballas SK, Snyder LM, Chrobak L, Melrose WD, Brabec V *et al*: **Characterization of 13 novel band 3 gene defects in hereditary spherocytosis with band 3 deficiency**. *Blood* 1996, **88**(11):4366-4374.

53. Ricard MP, Gilsanz F, Millan I: **Erythroid membrane protein defects in hereditary spherocytosis. A study of 62 Spanish cases**. *Haematologica* 2000, **85**(9):994-995.

54. Miraglia del Giudice E, Iolascon A, Pinto L, Nobili B, Perrotta S: **Erythrocyte membrane protein alterations underlying clinical heterogeneity in hereditary spherocytosis**. *British journal of haematology* 1994, **88**(1):52-55.

55. Yawata Y, Kanzaki A, Yawata A, Doerfler W, Ozcan R, Eber SW: **Characteristic features of the genotype and phenotype of hereditary spherocytosis in the Japanese population**. *International journal of hematology* 2000, **71**(2):118-135.

56. Lee JH, Moon KR: **Coexistence of gilbert syndrome and**

hereditary spherocytosis in a child presenting with extreme jaundice. *Pediatric gastroenterology, hepatology & nutrition* 2014, **17**(4):266-269.

57. Rawa K, Adamowicz-Salach A, Matysiak M, Trzemecka A, Burzynska B: **Coexistence of Gilbert syndrome with hereditary haemolytic anaemias.** *Journal of clinical pathology* 2012, **65**(7):663-665.

58. Rocha S, Costa E, Ferreira F, Cleto E, Barbot J, Rocha-Pereira P, Quintanilha A, Belo L, Santos-Silva A: **Hereditary spherocytosis and the (TA)nTAA polymorphism of UGT1A1 gene promoter region--a comparison of the bilirubin plasmatic levels in the different clinical forms.** *Blood cells, molecules & diseases* 2010, **44**(2):117-119.

59. Inoue T, Kanzaki A, Yawata A, Wada H, Okamoto N, Takahashi M, Sugihara T, Yamada O, Yawata Y: **Uniquely higher incidence of isolated or combined deficiency of band 3 and/or band 4.2 as the pathogenesis of autosomal dominantly inherited hereditary spherocytosis in the Japanese population.** *International journal of hematology* 1994, **60**(4):227-238.

국문 초록

서론: 유전구형적혈구증은 유전용혈적혈구질환 중 가장 흔한 질환으로, 적혈구 막 단백의 이상으로 인해 구형이 된 적혈구가 비장에서 파괴되어 용혈 빈혈이 발생한다. 유전구형적혈구증의 진단을 위해 다양한 검사가 이용되고 있으나, 대부분 용혈을 발견하거나 간접적인 검사법이며, 직접적으로 막 단백의 이상을 검사하는 방법들은 매우 복잡하고 검사가 어렵다. 최근 차세대 염기서열 분석법이 개발되어 대규모 유전자 돌연변이 검사가 가능하게 되었다. 본 연구에서는 한국인 유전구형적혈구증 환자를 대상으로 차세대 염기서열 분석법을 이용하여 적혈구 막 단백 관련 유전자 돌연변이를 분석함으로써, 유전구형적혈구증의 원인 유전적 변이 양상을 규명하고, 진단을 위한 분자 기전 연구의 토대를 마련하고자 하였다.

방법: 유전구형적혈구증으로 진단받은 환자 61 명의 환자들이 이 연구에 포함되었다. 43 개 유전자(막 단백 유전자 17 개, 효소 유전자 20 개, 동반질환 및 감별질환 유전자 6 개)에 대해 타겟 염기서열 분석법으로 돌연변이를 검색하고 결과를 분석하였다.

결과: 총 61 명의 환자 중 82%에 해당하는 50 명의 환자에서 적혈구 막 단백 유전자의 유의미한 변이가 관찰되었다. 17 개 막 단백 관련 유전자 중, 유의한 변이가 발견된 유전자는 *SPTB, ANK1, SPTA1,*

SLC4A1, EPB41, EPB42 등 6 개였고, 총 54 개 변이(기보고 있는 변이 8 개, 새로운 변이 46 개)가 관찰 되었다. *SPTB* 유전자에서 가장 많은 28 개의 변이가 관찰되었다. 20 개의 효소 관련 유전자 중, 유의한 변이가 발견된 유전자는 *ALDOB, G6PD, GAPDH, GSR* 등 4 개 유전자였고, 총 4 개의 기보고 있는 변이가 관찰 되었다. 6 개 관련 질환 유전자 중, *UGT1A1* 유전자에서 5 개의 기보고 된 돌연변이가 관찰 되었다. 관찰된 돌연변이 프로파일을 기반으로 하여, 유전구형적혈구증 진단을 위한 35 쌍의 시발체가 제안되었다.

결론: 본 연구는 한국인 유전구형적혈구증 환자를 대상으로 시행한 첫 번째 대규모 유전자 연구이다. 본 연구를 통해, 한국인 유전구형적혈구증 환자들의 원인 유전적 변이 양상을 규명하였으며, 이 결과를 토대로 하여 진단을 위한 유전자 패널을 구성하여 적용한다면, 보다 쉽고 빠르게 유전구형적혈구증 진단할 수 있을 것으로 기대한다.

주요어: 유전구형적혈구증, 차세대염기서열분석, 유전자 돌연변이

학번: 2012-30790