ACADEMIA JOURNAL OF BIOLOGY 2022, 44(4): 111–122
DOI: 10.15625/2615-9023/17470

GENETIC VARIATIONS OF FILAGGRIN ENCODING GENE (*FLG*) IN THE VIETNAMESE POPULATION REVEALED FROM WHOLE-EXOME SEQUENCING

Vu Phuong Nhung[1], Nguyen Huong Giang[3], Nguyen Thi Hong Nhung[1], Nguyen Dang Ton[1,2], Nguyen Hai Ha[1,2,*]

[1]Institute of Genome Research, VAST, Vietnam
[2]Graduate University of Science and Technology, VAST, Vietnam
[3]University of Science and Technology of Hanoi, VAST, Vietnam

Received 30 August 2022; accepted 9 December 2022

ABSTRACT

Filaggrin is a key protein that facilitates terminal differentiation of the epidermis and maintains skin barrier function. Mutations in the gene encoding filaggrin (*FLG*) have been identified to cause ichthyosis vulgaris, increase the risk of atopic dermatitis and other skin diseases. In this study, we established the database of *FLG* gene obtained by whole-exome sequencing (WES) of 244 Vietnamese. We also estimated allele and genotype frequencies of the *FLG* gene in this Vietnamese population and predicted the impact of novel variants on protein function using *in silico* analysis tools. The detected variants included 126 nonsynonymous, six nonsense mutations, six frameshift insertions/deletions, and one non-frameshift deletion, mostly located in exon 3. Of which, there were 11 novel variants have been identified and four of them were predicted as detrimental for encoding protein. Remarkable pathogenic variants were mostly nonsense variants, showing the main genetic factor underlying the pathology of diseases caused by *FLG*. The results obtained in this study would provide essential information about the genetic characteristics of *FLG* in the Vietnamese population and the risk of occurrence of diseases related to this gene, which facilitates the development of new specific and accurate diagnosis, treatment, and prevention options for *FLG*-related diseases, particularly in Vietnamese.

Keywords: *FLG*, whole-exome sequencing, genetic variants, ichthyosis vulgaris, atopic dermatitis.

Citation: Vu Phuong Nhung, Nguyen Huong Giang, Nguyen Thi Hong Nhung, Nguyen Dang Ton, Nguyen Hai Ha, 2022. Genetic variations of filaggrin encoding gene (*FLG*) in the Vietnamese population revealed from whole-exome sequencing. *Academia Journal of Biology*, 44(4): 111–122. https://doi.org/10.15625/2615-9023/17470

[*]*Corresponding author email*: nguyenhaiha@igr.ac.vn

INTRODUCTION

The filaggrin protein is particularly important in the formation of the skin barrier, both for its fundamental role in terminal epidermal differentiation and its implication in some of the most common dermatological diseases. It gathers the structural proteins in the outermost skin cells to form tight bundles, flattening and strengthening the cells to create a strong barrier (FLG gene: MedlinePlus Genetics). In addition, the processing of filaggrin proteins leads to the production of molecules that play a role in the hydration of the skin. These molecules also maintain the correct acidity (pH) of the skin, which is another crucial aspect of the barrier (FLG gene: MedlinePlus Genetics).

The filaggrin protein is synthesized as a giant precursor protein called profilaggrin, which is the main component of the keratohyalin granules in the stratum granulosum of the epidermis. Profilaggrin is encoded by the *FLG* gene, which is located in the epidermal differentiation complex on chromosome 1 (locus 1q21), which comprises three exons and two introns. Exon 1 consists only of a 5' untranslated (UTR5) sequence. Transcription starts at exon 2 which contains the translation initiation codon. The majority of the profilaggrin protein is encoded by exon 3, one of the largest exons in the genome. The profilaggrin protein is polymorphic, which contains between 10 and 12 repetitions of filaggrin flanked by N- and C-terminus domains. Mutations in *FLG* have recently been identified as the cause of common genetic skin disorders. To date, approximately 40 loss-of-function *FLG* mutations have been identified in ichthyosis vulgaris and/or atopic eczema (Global Variome shared LOVD).

The human monogenic skin disease ichthyosis vulgaris is characterized by dry, flaky skin, extra lines on the palms and soles (hyperlinearity), and a strong association with atopic disorders. The disorder is believed to be caused by mutations to the gene encoding profilaggrin. Sequencing of exon 3 of the *FLG* gene (which encodes almost the entire profilaggrin protein) in individuals with ichthyosis vulgaris revealed two null mutations: nonsense mutation (R501X; mutation of arginine codon 501 to a stop codon) and a frameshift mutation (2282del4; deletion of a 4-bp sequence at position 2282 in the filaggrin-coding DNA sequence). These mutations lead to premature stop codons in the first filaggrin repeat, thereby preventing all filaggrin synthesis from these alleles (Smith et al., 2006).

Atopic dermatitis (also known as eczema) is inflammation of the skin, typically characterized by itchiness, redness, and a rash. Mutations in *FLG* have been identified as the underlying cause of ichthyosis vulgaris and have also been shown to predispose patients to atopic eczema. Eczema is a very common, highly heritable, associated with asthma in around 30% of cases, as well as with a range of other allergies. Subsequent screening of an Irish childhood eczema case series and a Scottish asthma case series disclosed a strong association with the two prevalent loss-of-function *FLG* mutations R501X and 2282del4 (Palmer et al., 2006). These associations have been strongly replicated in more than 25 studies in European populations (Irvine, 2007).

Next-generation sequencing (NGS) technologies are increasingly used in many fields. Their power consists of the possibility to obtain a huge amount of data and discover novel and essential information about the human genome. NGS is a powerful approach for genotyping that enables the sequencing of the entire human genome within a single day. Due to the huge size, polymorphic variations in the number of filaggrin repeats, and highly repetitive nature, sequencing of the entire gene with the Sanger method is difficult (Osawa et al., 2011). So, using WES is effective for the investigation of the *FLG* gene. We have analyzed the WES data of 244 Vietnamese individuals to identify the spectrum of *FLG* genetic variants, including the pathogentic variants associated with several disorders. For the novel variants, we predicted their impact on protein function by *in silico* analysis. Therefore, this research would contribute to the fundamental

knowledge base of Vietnamese *FLG* genetic variants, which provide a scientific background for further studies with the goal of treating *FLG*-related diseases.

MATERIALS AND METHODS

Study population

The subjects of this study were 244 Vietnamese people, including healthy people or those suffering from one of the following diseases: Parkinson's, xeroderma pigmentosum, myasthenia gravis, cerebral palsy, etc. All study subjects had no related reports of diseases caused by the *FLG* gene. The peripheral blood samples were collected and stored by Genome Analysis Laboratory, Institute of Genome Research, Viet Nam Academy of Science and Technology.

Genomic DNA isolation

Peripheral blood samples from 244 study subjects (5 mL each) were collected and stored in EDTA tubes at -20 ^{o}C until used. Genomic DNA samples were isolated from blood using Exgene™ Blood SV mini Kit (Geneall Biotechnology, Korea) according to the manufacturer's protocol. Finally, Eppendorf tubes with DNA-containing suspension were stored at -20 ^{o}C.

Whole-exome sequencing (WES)

The DNA library was carried out using Sure Select V6-Post from Agilent Technologies, Santa Clara, California, USA, according to the protocol of the manufacturer. The exome sequencing process consists of three steps: random fragmentation of DNA, adapter ligation and enrichment of DNA fragments. Enriched library quantification was performed by using The Qubit™ dsDNA HS Assay Kit from Thermo Fisher Scientific, Waltham, Massachusetts, USA. The Bioanalyzer using High sensitivity DNA chip from Agilent Technologies, Santa Clara, California, USA, was used to check the library size distribution, with an expected size range from 200 bp to 400 bp. The sequencing was carried out by using an Illumina NovaSeq 6000 platform from Illumina, San Diego, CA, USA, with paired reads of 150 bp.

Data analysis

We used hg19/GRCh37 human reference genome to map the reads by using the BWA.v0.7.12 tool. Picard performs the marking of duplicates. The detection of single nucleotide variants (SNVs) and short Indels was accomplished using GATK and Sam tools. In addition, both variants with depth read lower than 20x and short Indels in repeat regions and within the 10 bp range from the start and end of the read were dismissed for excluding the false positive. The data on the *FLG* gene of each individual was filtered out from the WES data and aggerated into an Excel file using the bioinformatics tool developed at the Genome Analysis Department. Variant classification analyzes were performed using the Microsoft Excel tool. Finally, the remaining variants were filtered from the public database, including 1000 Genome Project (1000G) and Exome Aggregation Consortium (ExAC).

Genetic variant annotation and functional prediction

ANNOVAR program was used for the annotation of all variants. To predict the possible impact of an amino acid substitution on the protein function of nonsynonymous variants, the *in silico* analysis was carried out by using SIFT and Polyphen-2. In Polyphen-2 prediction models, a combination of two pairs of datasets, including HumVar and HumDiv, was used to test. Variants with scores of 0.0 are considered benign. Values closer to 1.0 are more confidently predicted to be damaging. SIFT scores use the same range with Polyphen-2, 0.0 to 1.0, but with opposite meanings. Variants with scores closer to 0.0 are predicted to be deleterious. Variants with scores very close to 1.0 are more confidently predicted to be tolerated.

RESULTS

In current study, people who have Parkinson's disease accounted for 40% of the research scope. Besides, other diseases such as xerodermal pigmentosa, cerebral palsy, myasthenia gravis po 5% for each. Healthy people also constituted 5% of total samples.

The remaining 35% of the sample set were other diseases, only possessed small proportions of the population (about 1 patient each disease) (Fig. 1).

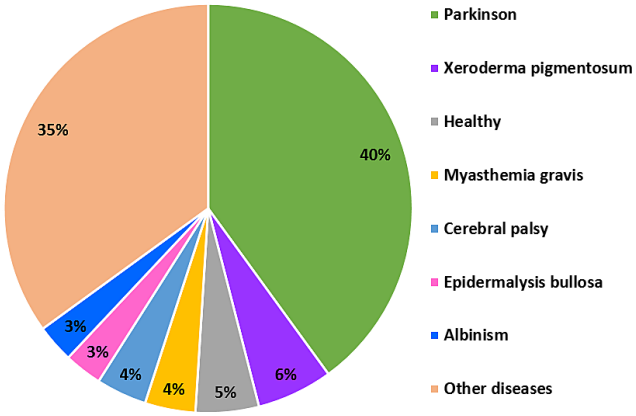

Figure 1. Distribution of samples based on the health conditions among investigated population. Studied subjects included of 244 individuals whose suffering from one of the following diseases: Parkinson's, xeroderma pigmentosum, myasthenia gravis, cerebral palsy, etc. None of enrolled subjects gave report of diseases related to *FLG* gene

Nonsense variants detected by WES

There were six nonsense mutations detected from 244 Vietnamese's WES data, which were c.A12064T, p.(K4022*); c.G7264T, p.(E2422*); c.C6117G, p.(Y2039*); c.C3905A, p.(S1302*); c.6950_6957del, p.(S2317*); c.C10225T, p.(R3409*). Searching through different databases including ExAC_EAS, 1000G-KHV, we found that the frequencies of these alleles were quite low (< 0.01). The number of detected subjects of each mutation in 244 samples was recorded in Table 1. It was clear that c.A12064T, p.(K4022*) and c.G7264T, p.(E2422*) were most frequently found (7 and 6 subjects, respectively). Additionally, all variants were found at the heterozygous stage.

Table 1. Nonsense variants observed in Kinh Vietnamese

Nucleotide change	Amino acid change	Reference SNP ID	ExAC_EAS	KHV	No	Het	Hom
c.A12064T	p.K4022*	rs146466242	0.0193	0.0152	7	7	0
c.G7264T	p.E2422*	rs374588791	0.0028	0.0101	6	6	0
c.C6117G	p.Y2039*	rs774850661	0.0003	N/A	1	1	0
c.C3905A	p.S1302*	rs754812742	0.0013	N/A	1	1	0
c.6950_6957del	p.S2317*	rs578184315	0.0031	0.0051	1	1	0
c.C10225T	p.R3409*	rs201356558	0.0006	0.0000	1	1	0

Note: N/A: Not available; No: Number of detected subjects; Het: Heterozygote; Hom: Homozygote.

Insertion and deletion variants identified by WES

In total, we detected two frameshift insertion mutations of *FLG*, including a known mutation: c.2758_2759insGG, p.(A920Gfs*203) and a novel mutation: c.8593_8594insGG, p.(A2865Gfs*28). Additionally, three reported frameshift deletion mutations were also identified including c.2755_2756del, p.(H919Cfs*5); c.3321delA, p.(G1109Efs*13); c.7945delA, p.(S2649Vfs*94) and a novel one c.8590_8591del, p.(H2864Cfs*5). Only one

novel non-frameshift deletion mutation was defined, which was c.12180_12181ins TACTAT, p.(Y4060_E4061insYY). These novel genetic variants have not been reported in any database such as ExAC and 1000G. Based on ExAC_EAS and 1000G-KHV, the allele frequencies of known variants have been shown to be low (< 0.01) among Asian populations as well as Kinh Vietnamese. All individuals were found to be heterozygous for detected variants (Table 2). The frameshift indels variants are remarkable because they can produce short or no functional protein products, one of the leading causes of diseases.

Table 2. Frameshift insertion, frameshift deletion, non-frameshift deletion identified by WES

Nucleotide change	Amino acid change	Reference SNP ID	ExAC_ EAS	KHV	No	Het	Hom
Frameshift insertion							
c.2758_2759insGG	p.A920Gfs*203	rs770547637	0,0003	N/A	7	7	0
c.8593_8594insGG	p.A2865Gfs*28	Novel	N/A	N/A	7	7	0
c.2755_2756del	p.H919Cfs*5	rs759142251	0,0003	N/A	6	6	0
Frameshift deletion							
c.3321delA	p.G1109Efs*13	rs200519781	0,0095	N/A	4	4	0
c.8590_8591del	p.H2864Cfs*5	Novel	N/A	N/A	7	7	0
c.7945delA	p.S2649Vfs*94	rs538406713	0,0035	0,0101	1	1	0
None-frameshift							
c.12180_12181insTACTAT	p.Y4060_E4061insYY	Novel	N/A	N/A	1	1	0

Note: N/A: Not available; No: Number of detected subjects; Het: Heterozygote; Hom: Homozygote.

Pathogenic variants discovered in Kinh Vietnamese

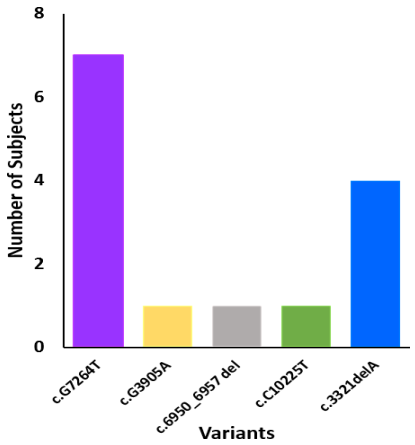

Figure 2. Pathogenic variants distribution among Kinh subjects. Five pathogenic variants of *FLG*, including c.G7264T, p.(E2422*); c.C3905A, p.(S1302*); c.6950_6957del, p.(S2317*); c.C10225T, p.(R3409*); which were reported to be pathogenic in ClinVar database. The c.G7264T, p.(E2422*) variant accounted for highest frequency

We restricted the data set to the status of pathogenic without any conflicting interpretations. We found four nonsense variants of *FLG*, including c.G7264T, p.(E2422*); c.C3905A, p.(S1302*); c.6950_6957del, p.(S2317*); c.C10225T, p.(R3409*); which were reported to be pathogenic in ClinVar database. The heterozygous frameshift deletion mutation c.3321delA, p.(G1109Efs*13) was also detected in our Vietnamese samples. The list of clinically annotated variants from ClinVar was summarized in Figure 2. As shown in the result, the c.G7264T, p.(E2422*) variant was detected in the study's population with the highest frequency (7 subjects).

Nonsynonymous variants

According to data from our 244 samples, the nonsynonymous mutations are the majority, with 126 variants found. Additionally, these variants were mainly detected in heterozygous form. According to the result of SIFT and Polyphen-2, 10 variants were predicted to be damaging (Table 3).

Many of those have a high allele frequency from Kinh in the Ho Chi Minh city population and also high frequency from our dataset: c.A10559C, p.(Q3520P) (20 subjects); c.A8660C, p.(Q2887P) (44 subjects); c.C5355G, p.(S1785R) (51 subjects); c.C5095T, p.(R1699C) (200 subjects). Eight novel nonsynonymous variants were defined, of which two were predicted to be damaging by both SIFT and Polyphen-2: c.C11315A, p.(S3772Y); c.C9829G, p.(R3277G) with one sample has been found for each variant.

Table 3. Nonsynonymous variants in investigated population

Nucleotide change	Amino acid change	Reference SNP ID	1000g2015aug	KHV	Polyphen-2	SIFT	Het	Hom
c.G3263A	p.G1088D	rs568796343	0.00059904	0.0101	D	B	2	0
c.T2407C	p.S803P	rs571501863	0.00059904	0.0101	D	B	2	0
c.G2212A	p.G738R	rs571540987	0.00059904	0.0101	T	B	2	0
c.C4220T	p.T1407I	rs754660081	N/A	N/A	T	B	1	0
c.C7694G	p.S2565C	rs182787187	0.00219649	0.0051	D	D	2	0
c.G4421A	p.R1474Q	rs200551704	0.00039936	0	T	B	1	0
c.C10277T	p.A3426V	rs79212820	0.00059904	0	T	B	1	0
c.T7747C	p.W2583R	rs775486975	N/A	N/A	T	B	1	0
c.G8692C	p.E2898Q	rs150427834	N/A	N/A	T	P	3	0
c.A10871G	p.E3624G	rs549118701	0.00019968	0.0051	T	B	1	0
c.T9997G	p.W3333G	Novel	N/A	N/A	D	B	1	0
c.A2974G	p.R992G	rs761429483	N/A	N/A	D	B	1	0
c.C5377T	p.R1793C	rs138288591	N/A	N/A	T	D	1	0
c.G3410T	p.S1137I	rs546382754	0.00039936	0	T	B	3	0
c.C7271G	p.S2424C	rs370240442	N/A	N/A	D	D	1	0
c.C9829G	p.R3277G	Novel	N/A	N/A	D	P	1	0
c.G10807A	p.A3603T	rs200249011	N/A	N/A	T	B	1	0
c.G6101A	p.S2034N	rs201132811	0.0009984	0	T	B	1	0
c.G5989A	p.A1997T	rs147125842	0.00119808	0	D	B	1	0
c.T5591C	p.V1864A	Novel	N/A	N/A	T	B	1	0
c.C2174T	p.T725I	rs3120655	0.11222	0	T	B	1	0
c.A8420C	p.D2807A	rs747864780	N/A	N/A	D	B	1	0
c.G10564A	p.G3522S	rs114892044	0.00539137	0	T	D	1	0
c.G8290C	p.G2764R	rs536445756	0.00039936	0.0051	D	B	1	0
c.G6416A	p.R2139H	rs200136954	N/A	N/A	T	B	2	0
c.A11146G	p.T3716A	Novel	N/A	N/A	T	P	1	0
c.T8600G	p.I2867S	rs562496957	0.0139776	0	T	B	8	0
c.A8587T	p.T2863S	Novel	N/A	N/A	T	B	4	0
c.T7442C	p.L2481S	rs55650366	0.342452	0.5556	T	B	3	2
c.G5051A	p.R1684H	rs12407807	0.296725	0.5505	T	D	3	2
c.A4126G	p.R1376G	rs11581433	0.343051	0.5606	T	B	3	2
c.G4079A	p.R1360H	rs11586631	0.28774	0.5505	T	B	3	2
c.G9658C	p.D3220H	rs200240824	0.00299521	0.0303	D	D	6	0
c.C2921T	p.A974V	rs143643121	0.00179712	0.0152	T	B	6	0
c.C7739G	p.S2580C	rs543657891	0.00019968	0.0051	D	P	1	0
c.G2509A	p.G837S	rs76413899	0.00499201	0.0051	D	B	7	0
c.C1342A	p.Q448K	rs371857058	0.00079872	0.0051	T	P	4	0
c.C1555A	p.H519N	rs12036682	0.0245607	0.096	D	P	1	0
c.G8548A	p.G2850S	rs2184952	N/A	N/A	T	B	23	0
c.G4406A	p.R1469H	rs145675213	0.0009984	0.0101	T	B	3	0
c.G3858C	p.L1286F	rs367739212	N/A	N/A	T	B	1	0

Nucleotide change	Amino acid change	Reference SNP ID	1000g2015aug	KHV	Polyphen-2	SIFT	Het	Hom
c.G9629A	p.R3210H	rs201718564	0.00079872	0.0051	T	B	1	0
c.T9700A	p.S3234T	rs546555106	0.00519169	0	T	P	1	0
c.G8581A	p.A2861T	rs746839571	N/A	N/A	T	B	1	0
c.A4381C	p.S1461R	rs762458364	N/A	N/A	D	B	1	0
c.G5378A	p.R1793H	rs3126080	0.00079872	0.0051	T	B	3	0
c.A6116C	p.Y2039S	rs746728852	N/A	N/A	T	B	1	0
c.C4867T	p.R1623W	rs78982044	0.00079872	0.0051	T	B	1	0
c.A9478C	p.S3160R	rs145597576	N/A	N/A	T	B	1	0
c.G1738A	p.G580S	rs144209313	0.00019968	0	T	D	1	0
c.G4944C	p.E1648D	rs533186028	0.00059904	0.0101	T	B	2	0
c.G6891C	p.E2297D	rs78179835			T	B	118	0
c.G2986A	p.G996R	rs149106390	0.00599042	0.0152	T	B	6	0
c.C8326A	p.H2776N	Novel	N/A	N/A	D	B	1	0
c.A8802C	p.Q2934H	rs368566122	N/A	N/A	T	B	2	0
c.G8396T	p.G2799V	rs371578635	N/A	N/A	D	B	1	0
c.A278T	p.N93I	rs199918968	0.00039936	0	T	P	3	0
c.C11909T	p.S3970L	rs3814299	0.0249601	0.0505	D	B	31	0
c.G6919A	p.A2307T	rs201076371	N/A	N/A	T	B	7	0
c.G11554A	p.G3852R	rs867113991	N/A	N/A	D	P	1	0
c.A7767T	p.R2589S	rs780916893	N/A	N/A	T	B	2	0
c.A9520G	p.S3174G	rs74925349	0.0329473	0.0051	T	B	1	0
c.G4366A	p.G1456R	rs554065456	0.00019968	0.0051	D	B	2	0
c.G11722A	p.D3908N	rs3814300	0.0105831	0.0707	T	B	29	0
c.A10559C	p.Q3520P	rs80088153	0.00978435	0.0455	D	D	20	0
c.G10492A	p.G3498S	rs146680739	0.00219649	0.0152	T	P	7	0
c.C8480T	p.A2827V	rs751639020	N/A	N/A	T	B	1	0
c.T11317C	p.Y3773H	rs200306807	N/A	N/A	T	B	1	0
c.C11315A	p.S3772Y	Novel	N/A	N/A	D	D	1	0
c.G5962A	p.G1988R	rs576649932	0.00019968	0	T	B	3	0
c.G9748A	p.G3250S	rs143078977	0.00379393	0.0051	T	D	5	0
c.G4279A	p.A1427T	rs148844389	N/A	N/A	T	B	1	0
c.G4996C	p.A1666P	rs555631824	N/A	N/A	D	D	1	0
c.C2579G	p.S860W	rs201661720	0.00219649	0.0101	D	P	1	0
c.G10289A	p.G3430D	Novel	N/A	N/A	T	D	1	0
c.G5960A	p.R1987H	rs150720370	N/A	N/A	T	B	2	0
c.C3500G	p.A1167G	rs58001094	0.534145	0.6717	T	B	96	114
c.G3338A	p.R1113H	rs140923774	0.0107827	0.101	D	B	47	4
c.C2225A	p.S742Y	rs3120654	0.121006	0.101	T	D	47	4
c.A1360G	p.T454A	rs2011331	0.464058	0.6717	T	B	102	124
c.A6626G	p.H2209R	rs66977240	N/A	N/A	T	B	182	0
c.T6058G	p.S2020A	rs7512857	N/A	N/A	T	D	186	0
c.G11213A	p.R3738H	rs77422831	0.286941	0.5505	T	B	110	91
c.G10590T	p.R3530S	rs72697000	0.290535	0.5505	T	D	83	91
c.A8807G	p.D2936G	rs80221306	N/A	N/A	T	B	108	0
c.A8506C	p.S2836R	rs11582087	N/A	N/A	T	B	143	0
c.A7330G	p.K2444E	rs71625200	0.343051	0.5556	T	B	108	91
c.G7192C	p.E2398Q	rs71625201	0.330471	0.5354	T	B	66	47
c.G5672A	p.R1891Q	rs12407748	0.28754	0.5505	T	B	110	91
c.C5414T	p.A1805V	rs12405241	0.28754	0.5505	T	D	110	91

Nucleotide change	Amino acid change	Reference SNP ID	1000g2015aug	KHV	Polyphen-2	SIFT	Het	Hom
c.C5095T	p.R1699C	rs12405278	0.28774	0.5505	D	D	109	91
c.C4445A	p.S1482Y	rs11204978	0.287141	0.5505	D	B	109	91
c.G2263A	p.E755K	rs74129461	0.327676	0.5606	T	B	110	92
c.C1432T	p.P478S	rs11584340	0.344649	0.5606	T	B	110	91
c.G1330A	p.G444R	rs11588170	0.287141	0.5505	T	P	110	91
c.G995T	p.G332V	rs41267154	0.340655	0.5606	T	D	110	92
c.G9809A	p.R3270H	rs147429418	0.00459265	0.0455	D	P	26	0
c.T9536G	p.V3179G	rs2065957	0.39357	0.5859	T	B	19	18
c.G1816A	p.G606R	rs749271421	N/A	N/A	D	D	2	0
c.G10903A	p.D3635N	rs75448155	N/A	N/A	D	D	13	0
c.G7015A	p.D2339N	rs139476473	N/A	N/A	D	D	14	0
c.G10779C	p.E3593D	rs12083389	0.361422	0.4646	B	T	54	12
c.C10736G	p.T3579R	rs3126075	0.477835	0.6717	B	T	23	15
c.G10691A	p.R3564H	rs7518080	0.25639	0.3788	B	T	8	0
c.T10663C	p.W3555R	rs12728605	N/A	N/A	B	T	2	0
c.G10307C	p.G3436A	rs2065955	0.514177	0.6616	P	T	97	122
c.A8660C	p.Q2887P	rs143280202	0.00978435	0.096	P	D	41	3
c.G7633A	p.G2545R	rs3126072	0.47484	0.6566	B	T	98	123
c.C7521G	p.H2507Q	rs3126074	0.46226	0.6566	B	T	7	2
c.G6604A	p.G2202S	rs146891517	0.0119808	0.101	B	T	22	4
c.T6580C	p.Y2194H	rs2184953	0.539736	0.6717	B	T	7	2
c.A6574C	p.K2192Q	rs66954353	N/A	N/A	B	T	7	0
c.A6462C	p.Q2154H	rs74129452	0.34385	0.4444	B	T	9	0
c.C6455A	p.S2152Y	rs77249082	N/A	N/A	P	T	4	0
c.C6430T	p.P2144S	rs78771632	0.0107827	0.0909	B	T	40	0
c.T6355C	p.Y2119H	rs7512553	0.327676	0.4394	B	T	222	0
c.C6323T	p.A2108V	rs7522925	N/A	N/A	B	T	224	0
c.A6190C	p.K2064Q	rs74129455	N/A	N/A	B	T	51	0
c.G6134C	p.S2045T	rs7546186	N/A	N/A	P	T	202	0
c.G5896A	p.G1966S	rs111661820	N/A	N/A	B	T	4	0
c.C5883A	p.H1961Q	rs3126079	0.538139	0.6717	B	T	7	2
c.T5839G	p.W1947G	rs80059102	N/A	N/A	B	T	9	0
c.T5828A	p.L1943H	rs113544881	0.313498	0.4394	B	T	8	0
c.C5355G	p.S1785R	rs72477394	0.0107827	0.101	P	D	47	4
c.G4240A	p.G1414R	rs145962910	0.0107827	0.101	B	T	47	4
c.C3551T	p.S1184L	rs3120649	0.120607	0.101	B	D	4	0

Note: N/A: Not available; Het: Heterozygote; Hom: Homozygote; B: Benign; T: Tolerate; D: Damaging.

Functional prediction of novel nonsynonymous variants using SIFT and PolyPhen-2

To evaluate the possible effects of nonsynonymous mutations, SIFT and Polyphen-2 were performed (Fig. 3). The functional prediction results of these new variants c.T5591C, p.(V1864A); c.A8587T, p.(T2863S) show that these amino acid changes had no serious impact on the function of the protein. These variations: c.C8326A, p.(H2776N); c.(T9997G), p.(W3333G); c.G10289A, p.(G3430D); c.A11146G, p.(T3716A) were predicted to be possibly damaging by one of the two tools SIFT or Polyphen-2. Therefore, it is not possible to draw precise conclusions about the effect on protein function caused by these substitutions. The two novel variants c.C11315A, p.(S3772Y) and c.C9829G, p. (R3277G) were predicted to be damaging by both SIFT and Polyphen-2.

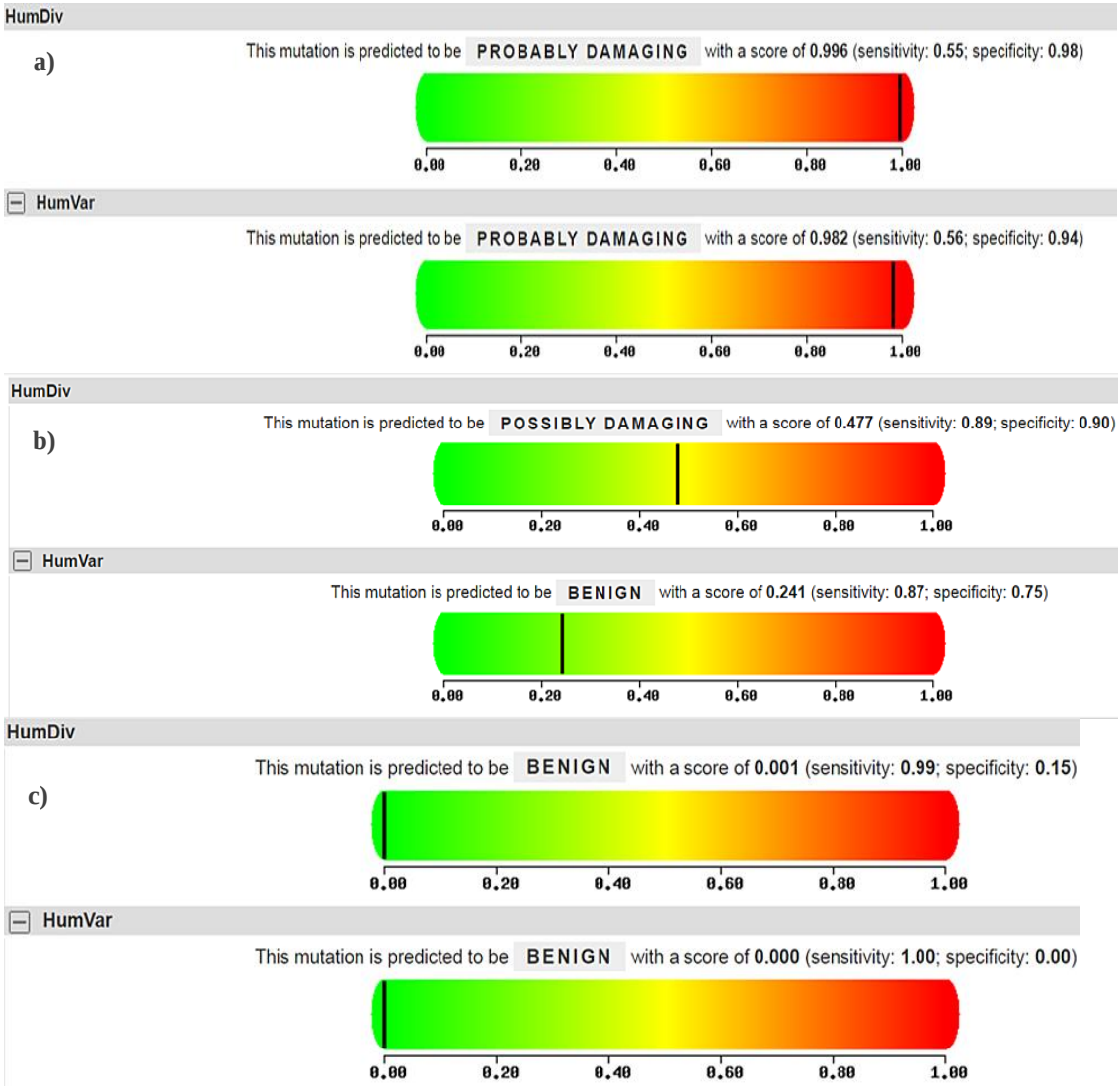

Figure 3. Polyphen-2 prediction of amino acid substitution caused by nonsynonymous mutations. a) Variant c.C11315A, p.(S3772Y) was predicted to be damaged; b) Variant c.A8587T, p.(T2863S) was predicted to be damamaged by HumDiv but benign by HumVar; c) Variant c.T5591C, p.(V1864A) was predicted to be benign

DISCUSSION

Filaggrin plays a key role in forming the skin barrier, not only for its fundamental role in epidermal differentiation but also for its association with some of the most common human dermatological diseases, such as atopic dermatitis and ichthyosis vulgaris. Homozygotes with a mutation in that gene manifest an acute form of atopy. Heterozygotes may be its asymptomatic carriers or manifest a benign or moderate form of atopy (Brown & Irvine, 2008; Brown et al., 2008). As previously reported, different populations have unique distribution patterns of *FLG* genetic variants (Zhang et al., 2011).

The pathogenic variants of our Kinh Vietnamese samples were summarized from the ClinVar database. The nonsense c.G7264T, p.(E2422*) variant was found in patients with albinism, myasthenia gravis, and

xeroderma pigmentosum. This mutation was considered to be related to ichthyosis vulgaris, which was found in Singapore and China populations (Chen et al., 2011). In addition, the heterozygous frameshift deletion mutation c.3321delA, p.(G1109Efs*13) was found in patients suffering from Cornelia de Lange syndrome, cerebral palsy and also in healthy people. This mutation has been reported in previous studies in patients with ichthyosis vulgaris and atopic dermatitis (Hamada et al., 2008; Kang et al., 2009). It can be seen that all pathogenic variants identified in this study were in a state that leads to a premature stop codon, producing a truncated protein, which lacked function. Furthermore, such individuals who carry heterozygous mutations, would be likely non-pathogenic or exhibit mild symptoms of skin diseases.

The pathogenic variants were detected in healthy individuals, cerebral palsy patients, and patients diagnosed with Cornelia de Lange syndrome. However, we lack information on the patient's skin condition, so it is not possible to draw precise conclusions about the causative factor of pathogenic variants on the alteration of filaggrin function. Nevertheless, given the pathogenic status of these variants, it is still noteworthy to give them advice to carefully examine whether they have any skin disorder. In case they exhibit symptoms of skin related disease, this evidence would better interpret their genotype-phenotype correlation.

In this study, we also found six nonsense and six frameshift indels variants which were not categorized as pathogenic in any populations. Both nonsense/indel variants will ultimately create premature stop codons, leading to harmful effects on encoding protein. Since these variants have not been reported in other populations nor the ClinVar database (possibly due to methodological limitations of previous studies that did not examine the entire exome), the re-examination of information on phenotypes of diseases related to *FLG* mutations may be conducted to assess the influence of this mutation in Vietnamese. Particularly for the non-frameshift deletion variant c.12180_12181insTACTAT, p.(Y4060_E4061insYY) that loses three nucleotides and does not change the open reading frame, the effect of this variant on functional protein is still unclear because there is no further information on the person carries this mutation, so it cannot be evaluated yet.

From the nonsynonymous dataset, 18 mutations were predicted to be damaging or possibly damaging by both SIFT and Polyphen-2. Many of those have a high allele frequency from Kinh in the Ho Chi Minh city population and also high frequency from our dataset: c.A10559C, p.(Q3520P) (20 subjects); c.A8660C, p.(Q2887P) (44 subjects); c.C5355G, p.(S1785R) (51 subjects); c.C5095T, p.(R1699C) (200 subjects). These mutations may also deserve closer examination to study any possible effects on disease penetrance.

The symptoms of dermatological diseases are quite disruptive to the daily activities of life. Indeed, atopic dermatitis is often called the itch that rashes since itch is a predominant feature of this disease. Though treatments are available to help the rash resolve and to help with the itch, this disease will continue to appear when treatments are stopped and often become infected. To date, both atopic dermatitis and ichthyosis vulgaris have no cure, the goal of treatment is to manage the condition. On the other hand, several researchers have studied *FLG*-related skin diseases in different populations, such as in Japanese and some European countries (Nomura et al., 2007; Sandilands et al., 2007). However, there is currently a lack of *FLG* genetic variant information in the Vietnamese population. Therefore, genetic screening for *FLG* in Kinh Vietnamese is necessary. The discovery of filaggrin function in the Vietnamese population has enabled us to better understand the pathogenesis of various disorders associated with alterations of the skin barrier. In addition, the study of its functions and the consequences of its deficit may have important therapeutic implications in the future, which facilitate the development

of new specific and accurate diagnosis, treatment, and prevention options in which this gene is altered, particularly in Vietnam. According to the current data of *FLG* genetic variants, we assume that genetic screening of nonsense and frameshift variants should be put in priority due to the destructive impact of these mutations on filaggrin protein.

This study recorded eight novel nonsynonymous mutations, which were located in exon 3 of the *FLG* gene. The new variants c.T5591C, p.(V1864A); c.A8587T, p.(T2863S), which related to a valine-to-alanine (V to A) substitution at position 1864 and a threonine-to-serine substitution at position 2863, respectively. The results of functional prediction using specialized software such as SIFT and Polyphen-2 show that these amino acid changes had no serious impact on the function of the protein.

These variations: c.C8326A, p.(H2776N); c.(T9997G), p.(W3333G); c.G10289A, p.(G3430D); c.A11146G, p.(T3716A) were caused by nucleotide replacement at positions 2776, 9997, 3333, 3430, 3716, respectively. These variants are predicted to be possibly damaged by one of the two tools SIFT or Polyphen-2. Therefore, it is not possible to draw precise conclusions about the effect on protein function caused by these substitutions.

The other nonsynonymous variant c.C11315A, p.(S3772Y) results in the substitution of a serine residue for tyrosine (S to Y) at position 3772. The variant c.C9829G, p. (R3277G) is a mutation that changes the amino acid at position 3277 from arginine to glycine (R to G). These two mutations were predicted to be damaging by both SIFT and Polyphen-2. This leads to the presumption that these novel amino acid substitutions might exert an influence on protein function.

This study only performed functional prediction based on bioinformatics tools. Therefore, in further studies, it is necessary to substantiate the predictions with suitable *in vitro* or *in vivo* models in order to draw accurate conclusions about the ability of these novel variants to influence protein function. Furthermore, monitoring patients carrying that variant can also assess some aspects, such as if they have any symptoms of skin disease.

CONCLUSION

In this study, we have identified the variants of the *FLG* gene from the WES data of 244 Vietnamese. The detected variants included 126 nonsynonymous, six nonsense mutations, six frameshift insertion/deletion, and one non-frameshift deletion, mostly located in exon 3. Of these, 11 novel variants have been identified and four of them were predicted as detrimental for encoding protein. The remarkable pathogenic variants were mostly nonsense variants, showing the main genetic factor underlying the pathology of the *FLG* related skin disorder. The genetic landscape of *FLG* collected in the current study would provide essential information for functional analysis of this protein as well as evaluate its effect on phenotype in future work.

Acknowledgements: This work was completed with the support of the Institute of Genome Research-Vietnam Academy of Science and Technology.

REFERENCES

Brown S. J. , Irvine A. D., 2008. Atopic eczema and the filaggrin story. *Semin Cutan Med Surg*, 27(2): 128–137. https://doi.org/10.1016/j.sder.2008.04.001

Brown S. J., Relton C. L., Liao H., Zhao Y., Sandilands A., Wilson I. J., Burn J., Reynolds N. J., McLean W. H., Cordell H. J., 2008. Filaggrin null mutations and childhood atopic eczema: a population-based case-control study. *J Allergy Clin Immunol*, 121(4): 940–946. https://doi.org/10.1016/j.jaci.2008.01.013

Chen H., Common J. E., Haines R. L., Balakrishnan A., Brown S. J., Goh C. S., Cordell H. J., Sandilands A., Campbell L. E., Kroboth K., Irvine A. D., Goh D. L., Tang M. B., van Bever H. P., Giam Y. C., McLean W. H., Lane E. B., 2011. Wide spectrum of filaggrin-null mutations in atopic dermatitis highlights differences

between Singaporean Chinese and European populations. *Br J Dermatol* 165(1): 106–114. https://doi.org/10.1111/j.1365-2133.2011.10331.x

FLG gene: MedlinePlus Genetics. Available from: https://medlineplus.gov/genetics/gene/flg/

Global Variome shared LOVD. Available from: https://databases.lovd.nl/shared/variants/FLG/unique

Hamada T., Sandilands A., Fukuda S., Sakaguchi S., Ohyama B., Yasumoto S., McLean W. H., Hashimoto T., 2008. De novo occurrence of the filaggrin mutation p.R501X with prevalent mutation c.3321delA in a Japanese family with ichthyosis vulgaris complicated by atopic dermatitis. *J Invest Dermatol*, 128(5): 1323–1325. https://doi.org/10.1038/sj.jid.5701164

Irvine A. D., 2007. Fleshing out filaggrin phenotypes. *J Invest Dermatol*, 127(3): 504–507. https://doi.org/10.1038/sj.jid.5700695

Kang T. W., Lee J. S., Oh S. W., Kim S. C., 2009. Filaggrin mutation c.3321delA in a Korean patient with ichthyosis vulgaris and atopic dermatitis. *Dermatology*, 218(2): 186–187. https://doi.org/10.1159/000163083

Nomura T., Sandilands A., Akiyama M., Liao H., Evans A. T., Sakai K., Ota M., Sugiura H., Yamamoto K., Sato H., Palmer C. N., Smith F. J., McLean W. H., Shimizu H., 2007. Unique mutations in the filaggrin gene in Japanese patients with ichthyosis vulgaris and atopic dermatitis. *J Allergy Clin Immunol*, 119(2): 434–440. https://doi.org/10.1016/j.jaci.2006.12.646

Osawa R., Akiyama M., Shimizu H., 2011. Filaggrin gene defects and the risk of developing allergic disorders. *Allergol Int*, 60(1): 1–9. https://doi.org/10.2332/allergolint.10-RAI-0270

Palmer C. N., Irvine A. D., Terron-Kwiatkowski A., Zhao Y., Liao H., Lee S. P., Goudie D. R., Sandilands A., Campbell L. E., Smith F. J., O'Regan G. M., Watson R. M., Cecil J. E., Bale S. J., Compton J. G., DiGiovanna J. J., Fleckman P., Lewis-Jones S., Arseculeratne G., Sergeant A., Munro C. S., El Houate B., McElreavey K., Halkjaer L. B., Bisgaard H., Mukhopadhyay S., McLean W. H., 2006. Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. *Nat Genet* 38(4): 441–446. https://doi.org/10.1038/ng1767

Sandilands A., Terron-Kwiatkowski A., Hull P. R., O'Regan G. M., Clayton T. H., Watson R. M., Carrick T., Evans A. T., Liao H., Zhao Y., Campbell L. E., Schmuth M., Gruber R., Janecke A. R., Elias P. M., van Steensel M. A., Nagtzaam I., van Geel M., Steijlen P. M., Munro C. S., Bradley D. G., Palmer C. N., Smith F. J., McLean W. H., Irvine A. D., 2007. Comprehensive analysis of the gene encoding filaggrin uncovers prevalent and rare mutations in ichthyosis vulgaris and atopic eczema. *Nat Genet*, 39(5): 650–654. https://doi.org/10.1038/ng2020

Smith F. J., Irvine A. D., Terron-Kwiatkowski A., Sandilands A., Campbell L. E., Zhao Y., Liao H., Evans A. T., Goudie D. R., Lewis-Jones S., Arseculeratne G., Munro C. S., Sergeant A., O'Regan G., Bale S. J., Compton J. G., DiGiovanna J. J., Presland R. B., Fleckman P., McLean W. H., 2006. Loss-of-function mutations in the gene encoding filaggrin cause ichthyosis vulgaris. *Nat Genet*, 38(3): 337–342. https://doi.org/10.1038/ng1743

Zhang H., Guo Y., Wang W., Shi M., Chen X., Yao Z., 2011. Mutations in the filaggrin gene in Han Chinese patients with atopic dermatitis. *Allergy*, 66(3): 420–427. https://doi.org/10.1111/j.1398-9995.2010.02493.x