MINISTRY OF EDUCATION AND TRAINING
VIETNAM ACADEMY OF SCIENCE AND TECHNOLOGY

GRADUATE UNIVERSITY OF SCIENCE AND TECHNOLOGY

Ma Thi Huyen Thuong

STUDYING OF RELATED VARIANTS TO EPIDERMOLYSIS BULLOSA, ALBINISM AND MICROTIA BY WHOLE EXOME SEQUENCING

SUMMARY OF DISSERTATION IN BIOTECHNOLOGY

Code: 9 42 02 01

HA NOI - 2023

The dissertation is completed at: Graduate University of Science and Technology, Vietnam Academy Science and Technology

Supervisors:

1. Supervisor 1: Assoc. Prof. PhD. Nguyen Dang Ton, Institute of Genome Research
2. Supervisor 2: Assoc. Prof. PhD. Nguyen Hai Ha, Institute of Genome Research

Referee 1: Prof. PhD. Nguyen Van Ba, Vietnam Military Medical University

Referee 2: Prof. PhD. Tran Huy Thinh, Hanoi Medical University

Referee 3: Assoc. Prof. PhD. Dong Van Quyen, Vietnam Institute of Biotechnology

The dissertation will be examined by Examination Board of Graduate University of Science and Technology, Vietnam Academy of Science and Technology at…….. (time, date……)

The dissertation can be found at:

1. Graduate University of Science and Technology Library
2. National Library

INTRODUCTION

1. The necessity of research

In Vietnam, so far, genetic information of three diseases including epidermolysis bullosa, albinism and microtia are very limited. The broad clinical spectrum, sometimes overlapping subtypes within each disease, makes accurate diagnosis more difficult, even unclassifiable in some complex cases. Therefore, genetic tests are necessary, helping to accurately classify disease groups/subtypes. In addition, although potential candidate genes for hypoplasia have been reported, however, no gene variant has been shown to cause this disease. Therefore, it is necessary to continue further studies on genetic polymorphisms of disease-causing genes as well as the mechanisms underlying the three diseases mentioned above.

The results obtained in this study will be the initial foundation to help build a data set of genetic variants in the group of patients with epidermolysis bullosa, albinism and microtia in Vietnam. At the same time, this is also a scientific basis to support doctors/geneticists in genetic counseling as well as the development of further clinical studies towards the goal of treating the diseases mentioned above or prognosis in the future.

2. Research objectives

- Sequencing and analysis of genomic sequences coding on Vietnamese patients with epidermolysis bullosa, albinism and microtia.

- Identification of cause/risk variants, potential genes that may be involved in the underlying mechanisms leading to epidermolysis bullosa, albinism, and microtia.

3. Research contents

1. Exploiting clinical information, collecting blood samples, and total DNA extraction of patients with epidermolysis bullosa, albinism, microtia and family members (if any).

2. Whole exome sequencing of patient’s samples.

3. Analyzing obtained data and comparing with published databases to identify potential disease-related/pathogenic gene variants.

CHAPTER 1. OVERVIEW

1.1. Epidermolysis bullosa

Epidermolysis bullosa (EB) is a rare hereditary skin disorder which is defined by moderate to excessive fragility of epithelial tissues and blister formation. The prevalence of EB is about 11.1/1 million population and 19.6/1 million live births, and it occurs equally commonly in males and females as well as between ethnic groups. Clinical manifestations in patients with EB can range from mild to severe, very severe, or can lead to death. To date, there is no cure for this condition and current treatment only focuses on the symptoms, mainly wound care, infection control, nutritional support, prevention, and treatment of complications.

Over the past three decades, an international consensus group has conducted a total of 5 meetings to revise and update subtypes of EB based on phenotypic characteristics and related genetic information. The four major subtypes of EB include: Epidermolysis bullosa simplex (EBS), Junctional epidermolysis bullosa (JBS), Dystrophic epidermolysis bullosa (DEB) and Kindler syndrome. The different types of EB are characterized by the certain layer in which the blisters form.

To date, variation on at least 21 genes has been described as the cause of EB, which are involved in coding for components of keratin, adhesion contacts, and desmosomes. Mechanically, the impairment at molecular level resulted in aberration of structural and functional integrity of intra-epidermal adhesion or dermo-epidermal anchoring, leading to cell and tissue dehiscence.

1.2. Albinism

Albinism has been described as a rare genetic disease (estimated population prevalence of 1/17,000 to 1/20,000), characterized by concomitant hypopigmentation of the skin, hair and/or eyes, or only impair pigmentation in the eyes. In addition to well-defined hypopigmentation in albinism patients, several visions can occur

including nystagmus, iris transillumination, macular hypoplasia, strabismus, reduces visual acuity, and reduced depth perception.

There are two main types of albinism, oculocutaneous albinism (OCA) and ocular albinism (OA), which are classified based on clinical phenotype, associated with hypopigmentation in the skin, hair, and eyes (OCA), or eyes only (OA). In addition, two rare syndromes of albinism have also been described: Hermansky-Pudlak syndrome (HPS) and Chediak-Higashi syndrome (CHS).

In terms of disease genetics, albinism is classified according to gene variants detected in different disease genes. A total of 20 genes leading to albinism have been reported, of which 19 are associated with OCA cases and one is located on the X chromosome associated with OA.

Like most rare genetic diseases, there is currently no cure for albinism, the current treatment focus on the symptoms, preventing the harmful effects of the sun on the skin and hair, and regular ophthalmological examination.

1.3. Microtia

Microtia is a congenital anomaly, which characterized by a small, abnormally shaped external ears (auricle-pinna). Research based on population performed in Italy, France, Sweden, Finland and United States estimated rates ranging about 0.8-4.2 per 10000 births. This symptom is often accompanied by a narrow, blocked or totally absent ear canal. The completely underdeveloped pinna is referred as anotia. Microtia can be unilateral or bilateral, but the unilateral form is more common, which occurs in 79-93% of affected individuals. Among unilateral microtia, it was often seen the right ear affected in approximately 60% of the cases.

Microtia can occur as independent congenital anomalies or as part of a complex set of malformations or a syndrome associated with abnormalities of the first and second pharyngeal arches during embryonic development. Any effect that leads to abnormalities in the development of the ear during embryonic development can be a risk

factor for microtia. These effects may be environmental factors or may also be related to genetic factors.

As microtia is defined with external ear canal absence, the children with microtia may experience partial or full hearing loss in the affected ear, thereby affecting life quality. If the children with bilateral condition, they may develop speech difficulty as they learn to talk and become self-conscious about their condition as they grow older. For treatment of microtia, the medical ambition is to provide the best shape as well as function of underdevelopment ear.

1.4. Vietnamese research

In Vietnam, there has been no data on prevalence of EB, albinism and microtia. For microtia, the number of newly diagnosed cases is 20-30 per year. Up to 2018, no in-depth genetic study has been conducted on subjects with epidermolysis bullosa, albinism, and microtia in Vietnam. The studies mainly reported on the clinical, subclinical, and complications of the disease, specifically in pediatric patients with bullous epidermolysis bullosa, who were examined and treated at the National Hospital of Pediatrics and Children's Hospital and National Institute of Dermatology in the period 2007-2011.

CHAPTER 2. MATERIALS AND METHODS

2.1. Study subjects

A total of 8 patients with epidermolysis bullosa, 7 patients with albinism and 11 patients with microtia were clinically diagnosed by specialists and samples were collected from Hanoi University Hospital, Hanoi. Peripheral blood samples of patients and family members (except family members of elderly patients with albinism Al005 and microtia group) were collected.

2.2. Inclusion and exclusion criteria

Inclusion criteria were based on clinical criteria along with the patient's and family's consent to participate in the study.

Exclusion criteria were based on failure to meet inclusion criteria and exclude diseases/disorders with similar clinical features according to specific guidelines.

2.3. Ethical considerations

This study was approved by the Institute of Genome Research Review Board, Vietnam Academy of Science and Technology, according to Decision No. 2-2019/NCHG -HĐĐD on April 2, 2019. Written informed consent was obtained from all family members before sample collection.

2.4. Instruments and equipment

All instruments and equipments for this study were provided by Institute of Genome Research, Vietnam Academy of Science and Technology.

2.5. Methodologies

2.5.1. Total DNA extraction

Peripheral blood samples (2 ml per individual) were collected into EDTA K2 tubes and stored at -20°C until use. Received samples were coded and then extracted total DNA using the *ExgeneTM Blood SV mini-Kit* (GeneAll-Korea) following the manufacturer's instructions. Subsequently, total genomic DNA was analyzed on 0.8% agarose gel and the concentration of double-stranded DNA (dsDNA) was measured using *Qubit dsDNA HS Assay Kit* (Life Technologies-USA).

2.5.2. Sequencing the entire coding genome

Whole exome sequencing (WES) was performed on 8 EB patients (except EB008.E), 7 albinism patients and 11 microtia patients. DNA libraries were established using the *Sure Select V6-Post kit* (Agilent Technologies, USA), and sequenced on Illumina's system.

2.5.3. Whole exome sequencing data analysis

The reads were mapped to hg19/GRCh37 human reference genome by BWA.v0.7.12 tool, and Picard was used to marking the duplicates. Genome Analysis Tool Kit (GATK) and Samtools were used to detect single nucleotide variants (SNVs) and short

insertions/deletions (Indels). The remaining variants were filtered from the public databases comprising 1000G, dbSNP and gnomAD. All variants with frequencies greater than 5% will be eliminated. The variants were annotated with the ANNOVAR program. Human Splicing Finder (HSF) and Alamut Visual software were used to splicing prediction.

2.5.4. Sanger Sequencing

The candidate variants indicated after WES analysis were further validated by Sanger sequencing. Sequence regions carrying these variants were PCR using specific primer pairs with sizes ranging from 247 bp to 485 bp, then the PCR product was purified and sequenced using the *BigDye® Terminator v3. 1 Cycle Sequencing Kit* (Applied Biosystems, USA) according to the manufacturer's instructions.

2.5.5. Functional analysis of splice site

Peripheral blood samples (2 ml per individual) of two patients with EB004, EB005 and their parents were collected, preserved in EDTA K2 tubes and stored at 4ºC during the day. After receiving samples were coded, total RNA was extracted using *Monarch® Total RNA Miniprep Kit* (New England Biolab, USA) following the manufacturer's instructions. Reverse transcription (RT) was performed by *ProtoScript® II First Strand cDNA Synthesis Kit* (New England Biolab, USA). The RT-PCR was performed, and PCR product were separated by electrophoresis on a 3% agarose gel.

2.5.6. MLPA

The copy number variation of the *COL7A1* gene in cases EB008 and EB008.E was determined by using *SALSA MLPA P415 COL7A1, KRT5 probemix kit* (MRC-Holland, Amsterdam, Netherlands) following the the manufacturer's protocol.

2.5.7. Protein-protein interaction analysis

In microtia, functional networks of protein-protein interactions were predicted by using STRING (http://string-db.org/).

2.5.8. KEGG enrichment analysis

To identify variants and candidate genes that are potentially risk factors for microtia, enrichment analysis by KEGG (Kyoto Encyclopedia of Genes and Genomes) was performed.

2.5.9. Classification of variant pathogenicity according to the American Society of Molecular Pathology and Medical Genetics

The American Society for Molecular Pathology and Medical Genetics (ACMG) provides guidelines for classifying variants with five levels (5-level system), including: pathogenic, likely pathogenic, uncertain, likely benign and benign. Two sets of criteria have been proposed, one to classify pathogenic or likely pathogenic variants and one to classify benign or likely benign variants.

CHAPTER 3. RESEARCH RESULTS

3.1. Clinical characteristics of patients

3.1.1. Epidermolysis bullosa

Patients present mechano-bulluos lesions of the skin with blistering and scarring. Almost affected individuals were male (7/9 patients) and pediatric and juvenile patients (8/9 patients). In which, three patients including EB002 and brothers EB004, EB005 were severe cases, with widespread blisters, contractures and absent nails. Patients EB007 also had a severe phenotype at birth, but the child's symptoms improved over time as the child grew older.

3.1.2. Albinism

All patients presented with typical clinical features of albinism, including various degrees of eyes, hair, and skin hypopigmentation. Visual effects related to depth vision and refractive errors were observed in 6/7 patients (except patient Al006); nystagmus was identified in 5/7 patients (Al001-Al005); while foveal hypoplasia only appeared in 2/7 patients, two brothers Al001 and Al002. Only patient Al004 (sister of patient Al003) had strabismus and ecchymosis in the eyes.

3.1.3. Microtia

Patients diagnosed with microtia in this study included 6 males (Mi001, Mi002, Mi004, Mi005, Mi007 and Mi011) and 5 females (Mi003, Mi006, Mi008, Mi009 and Mi010). These patients were all diagnosed with at least one ear defect, and all were found to have a right ear affecting. Out of 11 patients, there are 3 people with bilateral form: 2 male patients Mi002, Mi004 and 1 female patient Mi009. In addition, 3 patients including Mi002, Mi009 and Mi011 had manifestations of other malformations in addition to abnormal features in the ear.

3.2. Total DNA extraction

Total genomic DNA was extracted from peripheral blood, evaluated by 0.8% agarose gel, and measured OD at 260 nm and 280 nm to check the quality. The total DNA concentration with high purity (A260/280 > 1.80) ranged from 19.2 to 86.6 ng/µl.

3.3. Raw data of WES

Whole exome sequencing showed that billions of bases were read per sample, with an average number of more than 7 billion bases in each one and an average number of reads were more than 48 million. The accuracy of each nucleotide was assessed using a Phred quality score with a higher Q number indicating higher accuracy (e.g. Q20 equates to less than 1/100 chance of error rate, Q30 equates to likelihood error rate less than 1/1000 occurs). The number of bases with an average Q30 and Q20 was 94.34% (93.19% to 95.63%), and 98, 03% (97.53% to 98.56%), respectively. This result ensured that the data meets enough quality and reliability to conduct further analysis.

3.4. Variants associated with disease

3.4.1. Epidermolysis bullosa

3.4.1.1. Functional prediction of novel variants

Three *in sillico* predictions including SIFT, Polyphen-2 and Mutation Taster were used to predict the function of the encoded protein, that carried the novel missense variant of *COL7A1* c.8279G>A. All three tools indicated that the substitution of glycine by glutamic acid at position 2760 (p.G2760E) on the polypeptide chain was probably damaging/deleterious.

For novel variants located in the intron of the *COL7A1* gene, the Human Splicing Finder (HSF) tool, and Alamut Visual software were used to evaluate the ability to affect the natural splicing pattern of *COL7A1* mRNA. Both tools indicated that the c.5821-2A>G variant would have an impact on mRNA splicing with complete deletion of exon 71 in the polypeptide chain. For the c.4581+2delT variant, there was an opposite in predicting between the two tools. HSF predicted no significant impact on splicing signals, while the effect of skipping exon 43 very likely for c.4518+2delT was predicted by Alamut Visual.

3.4.1.2. Functional analysis of novel splicing COL7A1 c.4581+2delT

RT-PCR was performed to amplify the coding sequence region carrying the *COL7A1 c.4581*+2delT extending from exon 42 to exon 44 in a family with 2 EB children (EB004 and EB005). Next, the Nested-PCR reaction was performed to again amplify this sequence region, the PCR product electrophoresis on 3% agarose gel showed one amplicon of 125 bp in wild type and two amplicons in mutation with an additional shorter amplicon of 89 bp (Figure 3.5). The lack of 36 bp corresponded to the skipping of exon 43, which could be constructed by splice site (c.4518+2delT), this result was similar to the prediction of Alamut Visual software previously.

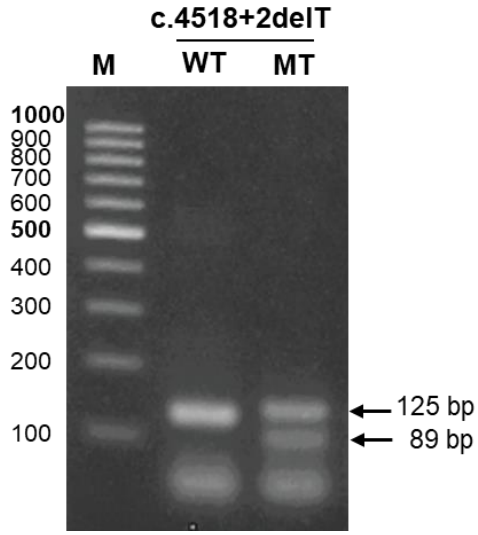

Figure 3.5. Splice defect of c.4518+2delT

Agarose gel electrophoresis of RT-PCR products on 3% agarose gel, the longer band with 125 bp in length presented the wild-type (WT) transcript, the shorter band with 89 bp in length presented the mutated transcript (MT), the bottom bands are primer dimer.

3.4.1.2. DEB cases carrying the COL7A1 variants

Patient EB001

WES analysis of patient EB001 detected homozygous change in *COL7A1* c.8279G>A, leading to the substitution of conserved amino acid glycine by glutamic acid (p.G2760E). This was a novel variant, which has not been found in any database or published before. In addition, this variant was also found in both parents of the patient who were heterozygous for Sanger sequencing. Based on ACMG's standard and guideline, the *COL7A1* c.8279G>A C has been classified as "*pathogenic*" variant with 4 pathogenic criterions including PM2, PM3 and PP1, PP3 - PP5.

Patient EB002

The compound heterozygous frameshift deletion of the *COL7A1* gene *(*c.2858_2859delAG and c.6081delC) was found in patient EB002. Both of these variants were known and resulted in premature termination codon in amino acid at p.E953fs*8 and p.P2029fs*177, respectively, of which the c.6081delC (p.P2029fs*177) was reported as a pathogenic variant on the ClinVar database (VCV000418654). Confirmation by Sanger sequencing also showed that the affected child inherited c.6081delC from the mother and c.2858_2859delAG from the father. The *COL7A1* c.2858_2859delAG (p.E953fs*8) was classified as a "*pathogenic*" variant according to the ACMG classification criteria with one very strong (PVS1), two moderate (PM3, PM4) and two supporting (PP1 and PP4).

Patient EB003 and two brothers of patient EB004, EB005

The combination of nonsense and another variant was detected in patients EB003 and two brothers in patients EB004 and EB005. Patient EB003 carrried a known compound heterozygous of the c.6205C>T (rs121912855) and the c.8233C>T (rs768202310), of which the c.6205C>T (p.R2069C) was described as pathogenic in ClinVar (VCV000017463). Verification by Sanger sequencing indicated that the mother and the father were heterozygous carriers for c.6205C>T and c.8233C>T, respectively.

In patients EB004 and EB005, two heterozygous variants were identified: a deletion point (c.4518+2delT) in intron 43 and a single-

base substitution (c.5047C>T) in exon 54 of *COL7A1* gene. The parent was heterozygous for c.4518+2delT and c.5047C>T, corresponding to the maternal and paternal variants, respectively. In which, the splicing variant of c.4518+2delT was a novel and had not been published in any database. This variant was considered as a "*pathogenic*" variant according to the ACMG classification, which included PVS1, PM2, PP1, PP3 and PP4 evidences. The c.5047C>T was a known pathogenic variant (rs587780490, VCV000130673), generating a terminator at codon 1683 on the amino acid sequence (p.R1683*).

Patient EB006

The conjunction of a splicing variant (c.5821-2A>G) and a missense variant (c.6205C>T) was found in patient EB006. Verification by Sanger sequencing showed that the c.5821-2A>G and c.6205C>T variants were inherited from the father and the mother, respectively. The c.5821-2A>G was a novel substitution variant while the c.6205C>T a known pathogenic variant (rs121912855, VCV000017463) and was also found in case of EB003. Based on the ACMG criteria, the *COL7A1* c.5821-2A>G variant has been classified as a "*pathogenic*" variant with one moderate (PM2) and three supporting (PP1, PP3 and PP4) evidences identified.

Patient EB008

Patient EB008 and his brother (EB008.E) were both diagnosed with EB. WES analysis of EB008 showed that two candidate heterozygous variants located on the *COL7A1* gene including c.1837C>T (p.R613*) and c.3830G>A (p.P1277L) could be the cause of the disease. Both of these variants were known (rs759634066 and rs35761247). In which, the c.3830G>A (p.P1277L) was benign and has been reported on the ClinVar (VCV000255109.16). Sanger sequencing indicated that both variants were found heterozygous in brothers (disease-expression) and maternal (disease-nonexpression), while the father carried neither variant (Figure). Thus, it has not been concluded that these were the two risk variants because the appropriate genetic pattern was not shown (autosomal recessive inheritance).

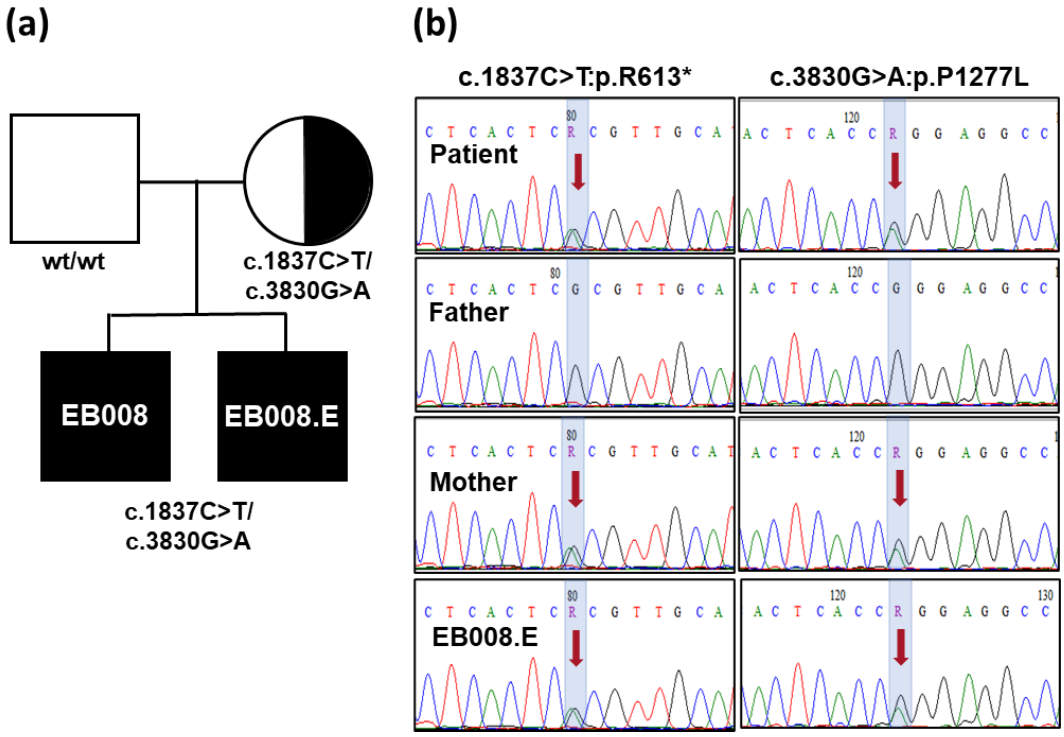

Figure 3.11. Two variants of *COL7A1* c.1837C>T and c.3830G>A were found in the family of patient EB008

(a) Family pedigree chart of patient EB008, black/white represent affected/unaffected individuals. (b) Sanger sequencing detected two heterozygous variants in the patient and maternal samples with no variant found in the father. The nucleotide changes were marked with vertical blue lines and red arrows.

To determine the genetic etiology in the patient's family, MLPA analysis was performed to identify the large deletions of the *COL7A1* gene that were not detected by WES due to limitations in CNVs analysis. However, MLPA results showed no abnormality in copy number in boths EB008 and EB 008.E. Based on the ACMG classification, the *COL7A1* variant *c.1837C*>T (p.R613*) has been considered as an "*uncertain*" with only PVS1 evidence identification.

3.4.1.2. EBS case carrying the KRT5 variant

The only variant located on the *KRT5* gene was the heterozygous variant of c.1429G>A (p.E477K), which found in patient EB007 (3 years old). This was a known variant (rs59190510) and reported as a pathogenic variant on ClinVar database (VCV000021174.4). Sanger sequencing verification indicated that c.1429G>A (p.E477K) was a *de novo* and not found in the parents.

3.4.2. Albinism

3.4.2.1. OCA cases carrying the TYR variants

Patient Al001 and Al002

In brother of patients Al001 (7 years old) and Al002 (5 years old), WES analysis results revealed a compound heterozygous variants located on exon 1 (c.346C>T, p.R116*) and exon 2 (c.926insC, p.R311fs*7) of *TYR* gene (NM_000372.3).

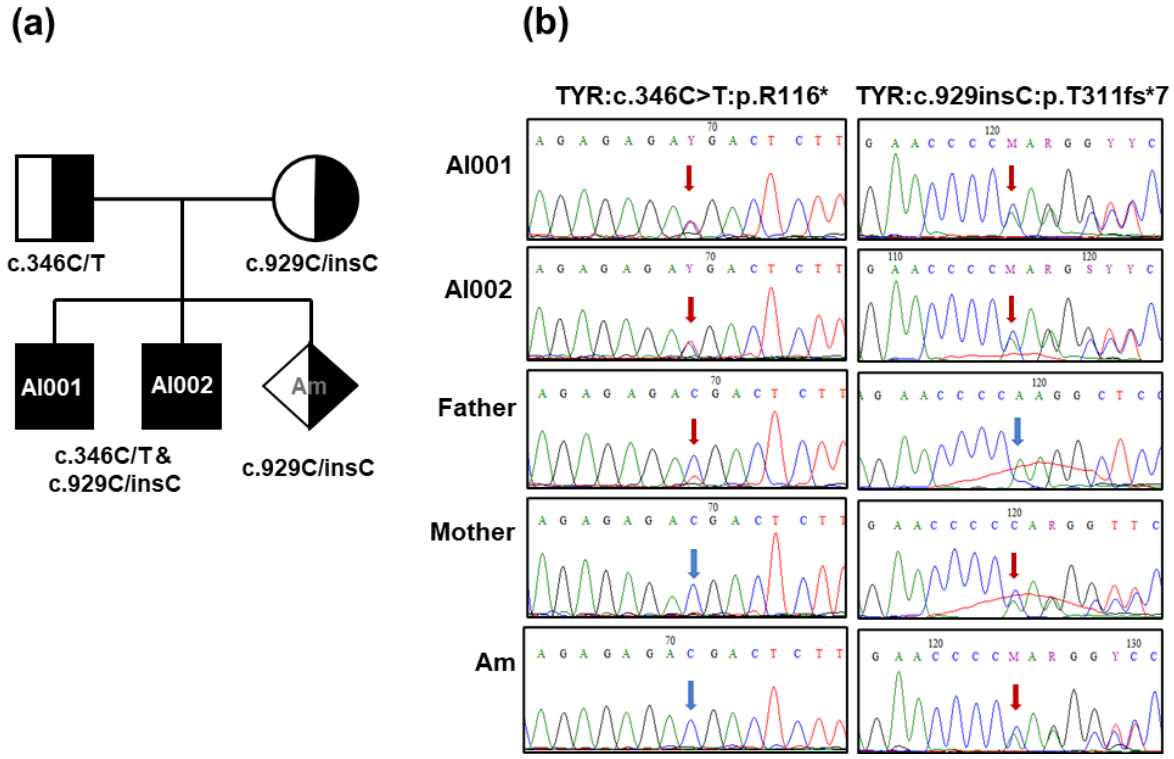

Figure 3.14. A compound heterozygous of *TYR* c.346C>T and c.926insC was detected in a family of patients Al001 and Al002

(a) The pedigree chart of Al001 and Al002 family. (b) Sanger sequencing detected a compound heterozygous of *TYR* variants in the patients, and each variant was carried from the parents. In the amniotic fluid sample (Am), the heterozygous c.346C>T was found. The nucleotide changes were marked with red arrows.

Both variants resulted in two premature termination codons (PTCs) and were reported as two pathogenic variant in ClinVar, and associated with ocularcutaneous albinism type 1 (OCA1). Sanger sequencing confirmed that the affected individuals inherited the c.3416C>T from the father and the c.929insC from the mother. In addition, genetic testing detected c.929insC heterozygous mutation from the amniotic

fluid sample of the third pregnancy of the mother (Figure 3.14). Like the parents, the baby was born without any signs of albinism and is currently developing healthy.

Al003 and Al004

Patients Al003 (25 years old) and Al004 (23 years old) were the first and second child in a family of three children, diagnosed with OCA. WES analysis revealed a homozygous variant of *TYR* c.115T>C (p.W39R). This was reported on the HGMD mutation database (CM100987) and associated with oculocutaneous albinism type 1B (OCA1B). Verification by directed sequencing showed that the heterozygous c.115T>C variant was observed in the parent of P3 and P4 and did not inherit to the third child (with normal phenotype).

Patient Al005

Patient Al005 was a 63-year-old male and was diagnosed with OCA type. WES revealed a homozygous *TYR* c.559_560insCATTATTATGTGTCAAATTATCCCC (c.559_560ins25), which resulted in a PTC (p.G190Cfs*12) on polypeptide chain. This was reported on the HGMD database with code number CN1414183 and related to the OCA1A subtype. Confirmation by Sanger sequencing also showed a similarity to WES analysis. Because Al005 was an elderly patient, the parental sample cannot be obtained, so the exact genetic pattern has not been determined. However, based on the previous findings of Chinese patients carrying the c.559_560ins25 variant and receiving it from their parents. It was possible to predict that both parents of the Al005 patient carry the variant and inherited it to their offspring.

3.4.2.2. OCA cases carrying the OCA2 variant

Patient Al006

Patient Al006 was a 26-years-old woman, the second child in the family. The patient's parents and brother were both phenotypically normal, while the patient was diagnosed with OCA. The results of WES analysis showed that the patient carried the known rare variant of *OCA2* c.2323G>A (p.G775S) (rs774822330) in the homozygous state. Sanger

sequencing verification showed that the patient received both variant alleles from both parents, while the patient's brother received two wildtype alleles in the normal state.

3.4.2.3. OCA case carrying the HPS1 variant

Patient Al007

Patient Al007 was a 5-year-old boy, diagnosed with OCA. WES analysis revealed a homozygous variant of *HPS1* c.972delC (p.M325fs*6). This was a known pathogenic variant, reported in the dbSNPs (rs281865082) and ClinVar (VCV000005280.5) databases and associated with a rare form OCA, Hermansky–Pudlak syndrome. The familial segregation was confirmed for c.972delC with the heterozygous trait in both parents by Sanger sequencing.

3.4.2. Microtia

3.4.3.1. Rare and novel variants founding in patients with microtia

A total of 108 heterozygous variants (MAF<0.05%) were detected on 74 different genes, including 105 missense variants, 2 frameshift variants (FS) and 1 nonsense variation accounts for 97%, 2% and 1%, respectively. Of which, there were 27 novel variants including 2 FS and 25 missense variants. Two novel frameshift variants involving *PCDH15* c.5027delA (p.E676Gfs*128) and *CANT1* c.775dupC (p.H259Pfs*25), were found in patients Mi005 and Mi010, respectively. The *EYA1* c.C922T variant (p.R308*) was the only nonsense variant that was identified in patient Mi002.

To evaluate the influence of variants on protein function, a set of 9 online tools integrated in ANNOVAR was performed. All variants evaluated as damaging/probably damaging on at least 5 predictors as well as FS and nonsense variants were further screened for continuate evaluation. After the screening process, a total of 47 variants met the above criteria located on 45 different genes; of which only 2 variants have been reported to be pathogenic (*EYA1* p.R308*) and likely pathogenic (*ABCC6* p.R1235W) on the ClinVar database. In addition, only *NCF1* p.R90H and *PLEC* p.D3627N variants were identified in

more than one patient of the 47 candidate variants screened, while the others were not.

3.4.3.2. Genetic analysis of single-gene model

A total of 9/11 microtia patients were identified following the monogenetic pattern; in which 5 patients (Mi002, Mi003, Mi006, Mi008 and Mi010) carried heterozygous variants on 5 different genes corresponding to autosomal dominant inheritance (AD) pattern. Candidate gene variants found in these cases were *EYA1* p.R308*, *TRPV4* p.V254L, *ABCC6* p.R1235W, *DNASE1* p.D115Y and *COL3A1* p.A1045T, respectively. In patient Mi004, two heterozygous variants on two different genes following AD patterns were found: *EVC* p.E681K and *LMNA* p.S543L. In patient Mi005, both AD and AR genetic patterns were shown with a total of 5 candidate variants identified. In which, two compound heterozygotes *PCDH15* (p.E1676Gfs*128 and p.E49G) and *PLEC* (p.D3627N and p.Q1737P) were associated with AD pattern. The heterozygous *LRP5* p.V61M was identified according to the AD mode. In patient Mi009, three candidate variants related to the AD were identified, including: *COG4* p.D542Y, *FN1* p.M1693T and *LRP5* p.C967F. In patient Mi011, both AD and AR patterns were shown with a total of 4 candidate variants identified. Two independent heterozygous variants, *COL11A1* p.P1383S and *MITF* p.A111V, correspond to the dominant pattern; while a compound heterozygous of *PLEC* p.R1082C and p.D3627N correspond to AR pattern.

3.4.3.3. Protein-protein interaction network

Protein-protein interaction network analysis (PPI) based on 74 candidate genes carrying rare variants (MAF < 0.5%) screened in 11 microtia patients was performed. The network consisted of 74 nodes corresponding to 74 proteins and 116 edges corresponding to 116 protein-protein interactions. Among them, FN1 and BMP2 were the two central proteins with the most interactions.

3.4.3.4. KEGG enrichment analysis

KEGG enrichment analysis revealed a total of 16 signaling pathways with an FDR < 0.05 (indicating the reliability of the analysis). Among these pathways, six were significantly associated with microtia, including (1) ECM-receptor interaction, (2) Osteoclast differentiation, (3) HPV infectious, (4) MTOR signaling pathway, (5) Protein digestion and absorption, and (6) Amoebiasis.

CHAPTER 4. DISCUSSION

4.1. Genetic diagnosis of epidermolysis bullosa

In this study, one homozygous and four compound heterozygous of *COL7A1* variants were found in 6 EB patients (EB001-EB006). In addition, each parent of patient was a heterozygous carrier and did not present any disease symptoms, thus predicting the AR pattern.

The most severe cases were observed in this study including EB002, brethrens of EB004 and EB005 associated with the premature termination codons (PTCs) variants. Among them, compound heterozygotes PTC variant (*COL7A1* c.2858_2859delAG and c.6081delC) were found in patients EB002 and the combination of a PTC variant (p.R1683*) and a splicing variant (c.4518+2delT) gene *COL7A1* was identified in EB004 and EB005. Patient EB003 was presented with a mild phenotype and carried compound heterozygous consisting of a PTC variant (p.R2745*) and a missense variant (p.R2069C). Other *COL7A1* variants that did not cause PTCs codon usually to produce less severe disease. This could be explained in cases of EB001 (p.G2760E) and EB006 (c.5821-2A>G and c.6205C>T). Due to the genotypes to phenotypes observed from the studied patients, PTC mutations resulted in the most serious cases, it was the same aspect of previously published studies.

The only case of EBS caused by the *KRT5* (c.1429G>A, p.E477K) variant in this study was patient EB007, diagnosed with EB. The p.E477K variant was strongly associated with very severe or lethal of generalized severe EBS (GS-EBS), with at least seven deaths within the first six months of life reported in previous studies. However, the

exceptions were also reported with severe lessions at birth and improving over time. These features were also observed in the case of EB007 with the presence of the heterozygous *KRT5* p.E477K variant.

In this study, only the familial of EB008 was inconclusion of causative variant with finding of two heterozygous *COL7A1* variants (p.R613* and p.P1277L) two affected children (EB008 and EB008.E) and their mother (who did not carry any abnormalities). The MLPA analysis also did not detect any large insertion/deletion variants of *COL7A1*. Therefore, in this case, it could be advisable to extend the discovery in whole genome or microarray analysis in further studies.

The results of the molecular tests would be valuable for appropriate genetic counseling and providing a theoretical prognosis for the progression of the disease.

4.2. Genetic diagnosis of albinism

In this study, all 7 patients were diagnosed with ocularcutaneous albinism (OCA) based on the characteristic clinical manifestations related to hypopigmentation in the skin, hair and eyes. The molecular diagnosis for seven affected individuals with albinism was provided, including identified mutations in *TYR*, *OCA2*, and *HPS1* genes associated with OCA subtype 1 (Al001, Al002, Al003, Al004, and Al005), subtype 2 (Al006), and Hermansky–Pudlak syndrome (HPS) (Al007), respectively. OCA1 and OCA2 were the common subtypes of non-syndromic OCA, while HPS was the rare syndromic OCA.

In five OCA1 cases of our study, three patients (Al001, Al002, and Al005) were classified as OCA1A, carrying PTC mutations (p.R116*; p.R311fs*7; and p.G190Cfs*12) in both alleles, leading to a complete loss of tyrosine function, thereby not producing melanin in the melanocytes. To the knowledge of our knowledge, patient Al005 was the fourth case carrying the c.559_560ins25 (p.G190Cfs*12) variant and the first finding case outside the Chinese. Moreover, genetic counseling was performed in the third pregnancy of the mother's Al001 and Al002, and the baby showed no signs of OCA after birth. This confirms that prenatal genetic screening in risk couples was effective and possible for this disorder. The case of two sick brothers Al003 and

Al004 was classified as OCA1B with the finding of homozygous *TYR* c.115T>C (p.W39R) variant, leading to decreased tyrosine activity and subsequently reduce melanin creation.

An OCA2 case detected in this study was patient Al006, carrying the OCA2 homozygous variant c.2323G>A (p.G775S). This variant was previously found with homozygous in a Vietnamese patient in a study of Markus et al. (University of Regensburg, Germany) in 2011. The authors assumed that this substitution had no harmful consequence on protein due to contradictory of SIFt and Polyphen-2. However, in current work, both tools showed the c.2323G>A (p.G775S) to be potentially damaging the encoded protein. The glycin at c.775 was highly conserved and located within the transmembrane 11 region in total of 12 domains on P polypeptide, encoded by the OCA2 gene. Therefore, the substitutions of G775 may possibly inhibit the folding of P-protein and lead to harmful consequences. Based on these suggestions, we predict that c.2323G>A (p.G775S) could be considered a "*likely pathogenic*" variant according to ACMG classification and should be functionally demonstrated in further studies.

Patient Al007 was the only syndromic OCA case associated with HPS caused by homozygous frameshift *HPS1* c.972delC mutation, making to the appearance of PTC on the polypeptide chain (p.M325fs*6), which was previously reported in causing of failure to the formation or resulting in a protein loss of function after translation. In HPS, the classic clinical features included OCA, a prolonged bleeding due to storage pool-deficient platelets, and development of granulomatous, pulmonary fibrosis, or neutropenia in some cases. However, apart from the main OCA findings, the other features specialty of HPS was not identified in Al007. For this reason, molecular tests could be needed for the specific classification of albinism subtypes.

4.3. Genetic factors in microtia

In this study, WES analysis of 11 microtia patients revealed 108 variants including 81 rare and 27 novel variants on 74 candidate genes. All of these variants were added to the database of variants potentially

associated with microtia. Among them, 2 variants have been reported as pathogenic/likely pathogenic on ClinVar, including the *EYA1* p.R308* (rs121909195, VCV000007929.6) in Mi002 and *ABCC6* p.R1235W (rs63750402, VCV000433323.3) in Mi006. However, based on the genetic pattern as well as the pathology related to the variant, only *EYA1* p.R308* could be the cause of the disease phenotype in case of Mi002. The *EYA1* p.R308* has been reported in cases of Branchio-oto-renal syndrome (BOR) or others than BOR having similar clinical manifestations. BOR was an AD disorder caused by pathogenic variants in *EYA1*, *SIX5* and *SIX1* genes, and characterized by symptoms such as hearing loss, brachial fistulae and cysts, anterior ventral fossa and renal abnormalities. Patient Mi002 was diagnosed with bilateral microtia (grade I in the right ear, grade II in the left ear) and contained other malformations including bilateral atrial rotator cuff fistula, surgical neck fistula, and only one kidney. Thus, it can be seen that the genotype-phenotype correlation in the case of Mi002 with the pathogenic variant of *EYA1* p.R308* was identified. Thereby confirming the accuracy and benefit of WES analysis to detect the etiology leading to hereditary syndrome associated with microtia and may be of value to apply in genetic counseling for risk families.

In addition, the study also identified 18 rare/novel variants that could be related to microtia based on genetic modeling as well as *in silico* analysis with damaging/probably damaging predictors on at least five tools. All of these variants need attention and extensive research on larger samples as well as in further functional studies.

PPI analysis based on 74 candidate genes indicates that BMP2 was one of the two central proteins with the maximum capacity to exchange information with other proteins. Therefore, BMP2 may play some important role in biological pathways associated with microtia.

In KEGG enrichment analysis, 5/6 signaling pathways involved the *COL4A4* gene, including ECM-receptor interaction. In addition, this gene has also been proposed as one of the candidate genes that may be related to microtia in research of 307 genes responsible for deafness in 32 microtia patients. Therefore, *COL4A4* may play some important

roles in the pathogenesis of this disease and should be focused on detailed effects analysis in further studies.

4.4. The role of molecular diagnostics in genetic counseling and rare disease management

Genetic heterogeneity, diversity in clinical manifestations, incomplete penetrance, genetic modifiers, and other environmental factors may further complicate progress toward a specific diagnosis, particularly in cases of rare disorders. This not only causes psychological stress and suffering for patients and their families, but also challenges and burdens health care systems. Therefore, accurate molecular diagnosis is necessary to provide appropriate treatment modalities and effective management of rare diseases. Genetic information is used as an additional source of data, contributing to future family planning-related planning and forming support networks for affected individuals and families. In addition, genetic diagnosis is not only interested in symptomatic subjects but also useful in screening and identifying carriers in asymptomatic groups; therefore it is very significant in the prevention of grade II in both benign and malignant disease groups. In summary, molecular diagnostics should be provided as a comprehensive service that includes pre- and post-test counseling. Highly qualified genetic counselors are key to ensuring that families receive the right support when detecting genetic disorders.

Research limitations

This study was conducted on three groups of rare diseases with very low prevalence in the population; simultaneously, the cost and time limitations make the study still modest in size samples. In particular, microtia is a genetic condition of unknown cause, which can exist independently or as a part of clinical syndromes associated to other congenital anomaly, making it difficult to identify potential or risk variants and need to be continued on a larger sample size in further studies. In addition, the failure to obtain samples of family members and related clinical images in microtia patients is also a limitation. However, any given information or hypothesis can also be a source of data for reference, along with other studies to build large enough evidence to find the etiology as well as the pathogenesis of microtia. Thereby as a premise for further functional studies toward the goal of clinical practice in the future.

CONCLUSIONS AND RECOMMENDATIONS

Conclusion:

1. Whole exome sequencing of 8 epidermolysis bullosa patients (EB), 7 albinism patients and 11 microtia patients were performed with 48 million reads on average and 94% Q30 reached.

2. Pathogenic variants were identified in 7/8 patients with EB, all patients with albinism (7 patients) and 1 risk variant in 1/11 microtia patients were revealed by WES, specifically as follows:

+ Six patients of dystrophic epidermolysis bullosa have been identified novel/rare *COL7A1* variants *(*c.8279G>A, c. 4518+2delT, c.5821-2A>G/c.2858_2859delAG, c.5047C>T, c.6081delC, c.6205C>T and c.8233C>T) and one patient with epidermolysis bullosa simplex carried the *KRT5* c.1429G>A (p.E477K).

+ All albinism patients were diagnosed and classified, including 4 OCA1 patients (3 OCA1A patients and 1 OCA1B patient), 1 OCA2 patient and 1 patient with Hermansky-Pudlak syndrome. The causative variants were also indetified in *TYR* (c.346C>T, c.926insC, c.115T>C and c.559_560ins25), *OCA2* (c.2323G>A) and *HPS1* (c.972delC) genes, respectively. Prenatal diagnosis was made in the third pregnancy in the family with two OCA1A children (Al001 and Al002) and the baby showed health and no signs of OCA after birth.

+ Total of 108 variants were identified in 11 patients, including 81 rare and 27 novel variants on 74 candidate genes that may be related to microtia. In which, *EYA1* p.R308* can be considered as the causative agent in case of Mi002 related to Branchio-oto-renal syndrome (BOR).

Recommendations:

According to data revealed from this study, we would like to give following proposals:

1. Functional analytical studies using appropriate *in vitro*, or animal models should be conducted to assess the true impact of the novel discovered variants in patients with epidermolysis bullosa, from that elucidates the pathogenesis involved.
2. It should be necessary to investigate with a larger number of samples to expand the detection scale of pathogenic variants/risk variants, especially in the group of microtia with a complex molecular mechanism that has not yet been elucidated.

NEW CONTRIBUTIONS OF THE THESIS

This study identified the genetic-cause variants of epidermolysis bullosa, including three novel variants located on the *COL7A1* gene *(*c.8279G>A, c.4518+2delT and c.5281-2A>G).

This study indicated the genetic causes in albinism patients. In which, the *TYR* c.115T>C (p.W39R) which was caused to ocularcutaneous albinism type 1 was found for the first homozygote.

This study identified 27 novel variants on candidate genes that may be associated with microtia.

LIST OF THE PUBLICATIONS RELATED TO THE DISSERTATION

1. **Ma THT**, Luong TLA, Hoang TL, Nguyen TTH, Vu TH, Tran VK, Nguyen DB, Trieu TS, Nguyen HH, Nong VH, Nguyen DT. *Novel and very rare causative variants in the COL7A1 gene of Vietnamese patients with recessive dystrophic epidermolysis bullosa revealed by whole-exome sequencing*. Mol Genet Genomic Med. 2021 Aug;9(8):e1748. doi: 10.1002/mgg3.1748.
2. **Thuong MTH**, Anh LTL, Nhung VP, Ngoc TTB, Lan HT, Phuong DK, Ha NH, Van Hai N, Ton ND. *Genetic analyses of Vietnamese patients with oculocutaneous albinism*. J Clin Lab Anal. 2022 Jul 23:e24625. doi: 10.1002/jcla.24625.
3. **Ma Thi Huyen Thuong**, Dang Tien Truong, Nguyen Hai Ha, Nguyen Dang Ton. *Identification of a de novo mutation in KRT5 gene underlying epidermolysis bullosa simplex by whole exome sequencing in a Vietnamese patient*. Vietnam Journal of Biotechnology 19(2): 223-228, 2021.