IDENTIFICATION AND CLINICAL RELEVANCE OF PHENOTYPE MODIFYING GENETIC FACTORS IN MONOGENIC DISEASES

Ph.D. Thesis

Éva Melinda Pap MD

Supervisors:

Nikoletta Nagy MD, PhD
Department of Medical Genetics
University of Szeged, Szeged, Hungary

Gábor Németh MD, PhD
Department of Gynecology and Obstetrics
University of Szeged, Szeged, Hungary

Doctoral School of Clinical Medicine
University of Szeged, Szeged, Hungary

2020

TABLE OF CONTENTS

LIST OF PUBLICATIONS

Publications providing the basis of the dissertation

I. Pap ÉM, Farkas K, Tóth L, Fábos B, Széll M, Németh G, Nagy N.: Identification of putative genetic modifying factors that influence the development of Papillon-Lefévre or Haim-Munk syndrome phenotypes. **Clin Exp Dermatol**. **2020** Jul;45(5):555-559. **IF: 1.977**

II. Pap ÉM, Farkas K, Széll M, Németh G, Rajan N, Nagy N.: Identification of putative phenotype-modifying genetic factors associated with phenotypic diversity in Brooke-Spiegler syndrome. **Exp Dermatol**. **2020**.07.26. Accepted for publication. **IF: 3.368**

III. Pap ÉM, Széll M, Nagy N, Németh G: Genomikai és fenotípus vizsgálatok Brooke-Spiegler, Papillon-Lefèvre, és Haim-Munk szindrómában. **Magyar Nőorvosok Lapja.** Under review.

1. INTRODUCTION

1.1. General features of monogenic diseases

Monogenic diseases result from modifications in a single gene occurring in all cells of the body. Though they are relatively rare, they can affect millions of patients worldwide. Scientists currently estimate that over 10,000 of human diseases are known to be monogenic. Pure genetic diseases are caused by a single error in a single gene in the human DNA. The nature of the disease depends on the functions of the modified gene. The single-gene or monogenic diseases can basically be inherited in three different manners: dominant, recessive and X-linked.

Dominant diseases are monogenic disorders that involve damage to only one gene copy, recessive diseases are monogenic disorders that occur due to damages in both copies, while X-linked diseases are monogenic disorders that are linked to defective genes on the X chromosome. The X-linked alleles can also be dominant or recessive. These alleles are expressed equally in men and women, more so in men as they carry only one copy of X chromosome (XY) whereas women carry two (XX).

The global prevalence of all single gene diseases at birth is approximately 10/1000 (Orphanet Database, www.orpha.net). Monogenic diseases affect nationwide only a few patients, but altogether they affect a significant portion of the populations since their thousands of different types are known. In general, these disorders are less known than the common, multifactorial ones. They occur rarely in the everyday practice of the medical practitioners and less attention is paid for the research of these diseases (Kelsall *et al*., 2013).

Monogenic diseases can vary greatly in their severity. They can cause mild, severe or very severe symptoms and can be associated with phenotypic diversity like common diseases. Their symptoms can significantly impair the life quality of the patient and they can also result in stigmatization and difficulties in socialization (Kelsall *et al*., 2013).

1.2. Clinical management of monogenic diseases – the importance of genetic counseling

Genetic counseling gives the patient information about how genetic conditions might affect the patient or his/her family. The genetic counselor or other healthcare professional will collect the patient's personal and family health history. They can use this information to determine how likely it is that the patient or his/her family member has a genetic condition. Based on this information, the genetic counselor can help the patient decide whether a genetic test might be right for the patient or his/her relative.

Based on your personal and family health history, your doctor can refer you for genetic counseling. There are different stages in your life when you might be referred for genetic counseling.

Before the patient becomes pregnant genetic counseling can address concerns about monogenic diseases that might affect the baby of the patient during infancy or childhood or the ability of the patient to become pregnant.

Genetic counseling is important while the patient is pregnant, it can address certain tests that may be done during the pregnancy to detect monogenic diseases that might affect the baby of the patient during infancy or childhood.

Genetic counseling can address concerns if a child is showing signs and symptoms of a monogenic disease.

Genetic counseling can be helpful if the patient has symptoms of a monogenic disease or has a family history of a monogenic condition that makes him or her more likely to be affected with that condition.

Following the genetic counseling session, the patient might decide to have genetic testing. Genetic counseling after testing can help the patient better understands the test results and treatment options, help the patient deal with emotional concerns, and refer the patient to other healthcare providers and advocacy and support groups.

1.3. Recent milestones in the background of knowledge explosion in the understanding of monogenic diseases

As a result of the Human Genome Project (HGP) and the development of next-generation sequencing technologies, nowadays we know more and more about

monogenic diseases. In the last 30 years, many disease causing genes and mutations have been identified and the genetic background of many monogenic diseases have been elucidated (Figure 1.). As an example we demonstrate the timeline of the discoveries of the genetic background of Charcot-Marie-Tooth (CMT) disease. However, it can occur that sequencing is unable to answer clinically relevant questions regarding phenotypic diversity and/or disease prognosis (Jarinova *et al.*, 2012; Timmerman *et al.*, 2014; Kiritsi *et al.*, 2015; Smith *et al.*, 2019).

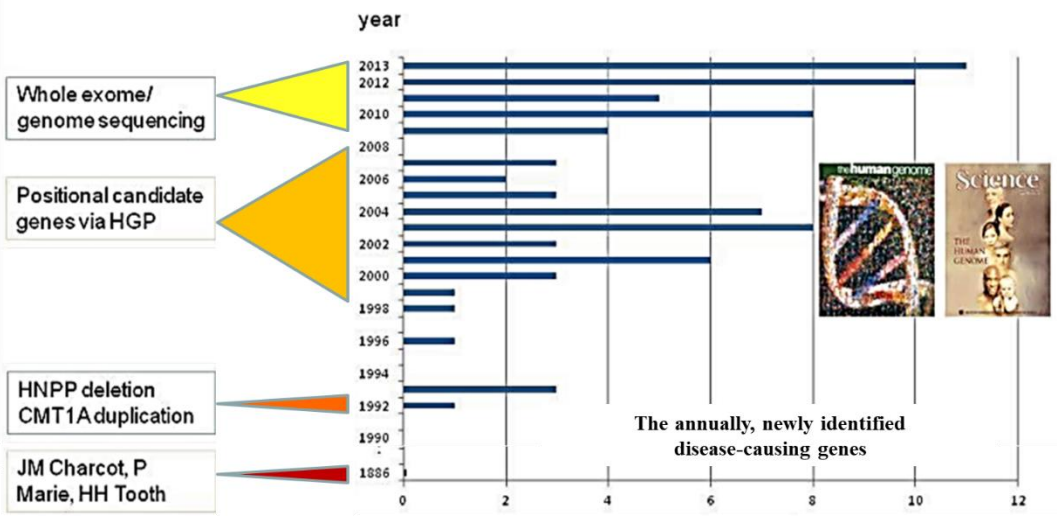

Figure 1. Genetics of Charcot-Marie-Tooth (CMT) Disease within the Frame of the HGP Success. The number of the annually, newly identified causative genes show two peaks: one is explained by the HGP, the other is the consequence of the fast development of sequencing, next-generation sequencing technologies (Timmerman *et al.*, 2014).

1.4. The investigated monogenic diseases

In my thesis, I have summarized the results of my genetic investigations in monogenic diseases: the clinical variants of the cylindromatosis gene (*CYLD*) mutation-caused disease spectrum such as multiple familial trichoepithelioma type 1 (MFT1), familial cylindromatosis (FC) and Brooke-Spiegler syndrome (BSS) and the clinical variants of the cathepsin C (*CTSC*) mutation-caused disease spectrum such as the Papillon-Lefèvre syndrome (PLS) and the Haim-Munk syndrome (HMS). My investigations focused on the identification of putative phenotype modifying genetic factors in these rare monogenic diseases, which are responsible for the observed phenotypic differences among the affected patients carrying the same disease-causing *CYLD* or *CTSC* mutations.

1.4.1. Clinical variants of the *CYLD* mutation-caused spectrum

Brooke-Spiegler syndrome (BSS) is a rare monogenic skin disease characterized by the development of skin appendage tumors such as cylindromas, trichoepitheliomas and spiradenomas (Evans 1954; Bignell *et al.,* 2000). The first symptoms of BSS are small skin-colored papules, which occur in childhood and adolescence (Evans, 1954). These tumors grow slowly in size and continue to appear throughout the lifetime of the patient (Nagy *et al.,* 2015). Expression of the papules exhibits wide variation among and within affected families (Bignell *et al.,* 2000).

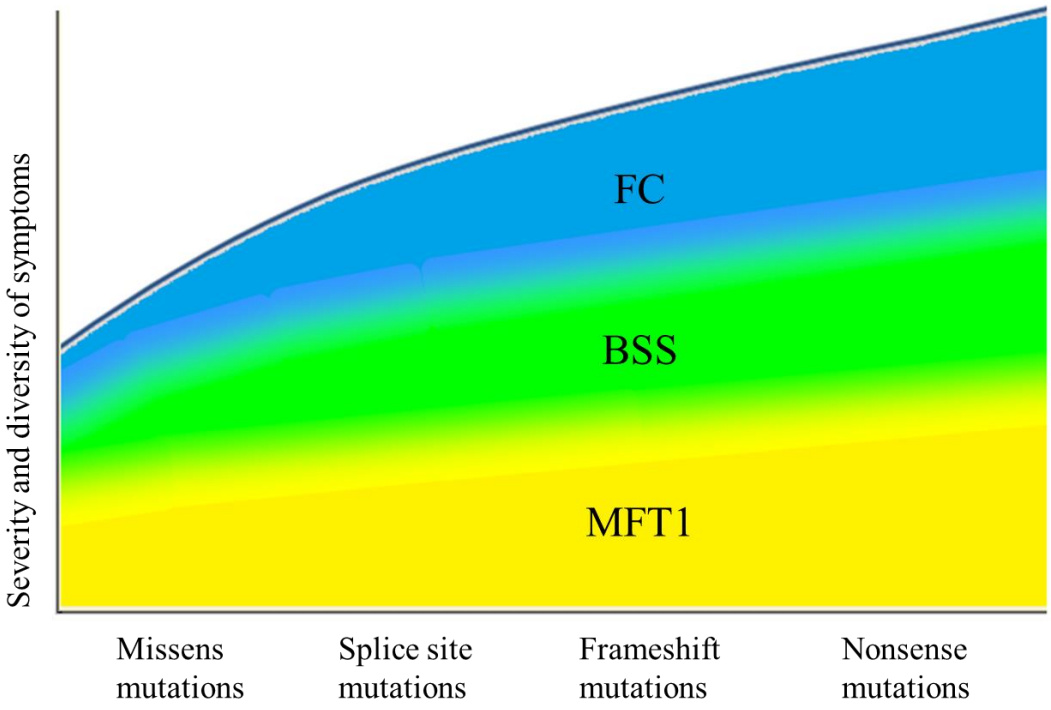

Figure 2. The severity and diversity of the symptoms of the clinical variants of the CYLD mutation-caused spectrum. In general, missense mutations of the *CYLD* gene is associated with the mildest symptoms and in approximately 50% of the cases they results in the development of MFT1. Nonsense mutations of the *CYLD* gene are associated with the most severe symptoms and with the highest diversity of the symptoms of the affected patients, they can lead to the development of FC, MFT1 and BSS with approximately equal frequency. (Nagy *et al.*, 2015).

BSS is transmitted as an autosomal dominant condition affecting males and females equally (Guggenheim and Schnyder, 1961). BSS and its phenotypic variants were independently mapped to chromosome 16q12-q13 by several groups (Fenske *et al.*,

2000; Biggs *et al.*, 1995; Takahashi *et al.*, 2000). Within the mapped region, the cylindromatosis gene *(CYLD)* was identified as the causative gene responsible for the development of the disease (Bignell *et al.*, 2000). The *CYLD* gene (GenBank accession number NM_015247) spans 56 kb and contains 20 exons, the first 3 of which are untranslated, and 19 introns. Of the 17 known splice variants, 13 affect protein coding regions, and the remaining produce non-coding transcripts (http://ensemble.org). The tumor suppressor *CYLD* gene encodes an enzyme with deubiquitinase activity. The CYLD enzyme posttranslationally modifies its target proteins by removing Lys63-linked ubiquitin chains (Kovalenko *et al.*, 2003). The protein interacts with several members of the NF-kB signaling pathway, including the TRAF2, TRAF6, NEMO and BCL3 proteins, acting as a negative regulator (Hutti *et al.*, 2009).

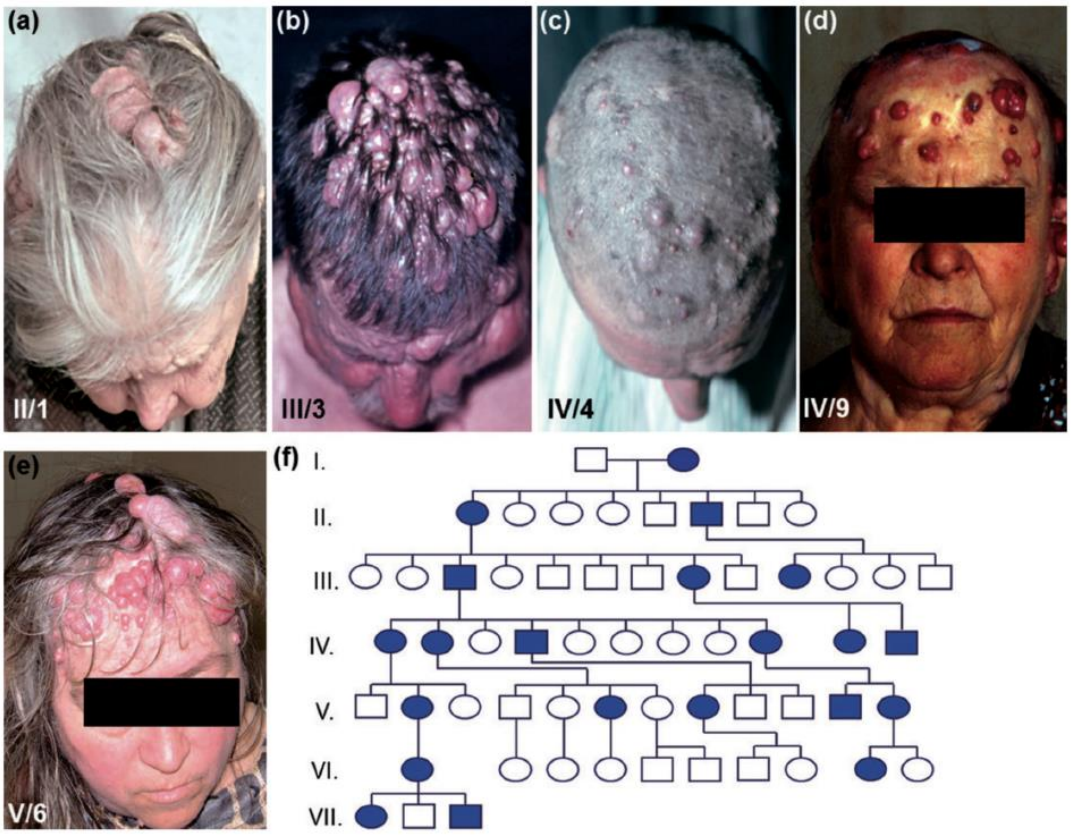

Figure 3. The Hungarian BSS pedigree of Bukovinian (Romania) origin. The clinical pictures of the severe hairy scalp symptoms of the affected individuals are represented from the (a) 2nd, (b) 3rd, (c and d) 4th and (e) 5th generations. (f) The pedigree contains 21 affected family members spanning 7 generations (Nagy *et al.*, 2013).

Concerning the so far reported phenotypes and the associated *CYLD* mutations, it is difficult to establish genotype-phenotype correlations in BSS. However, the elucidation of the genotype-phenotype correlations has significant clinical relevance in promoting the understanding of disease mechanism and contributing to the development of future therapeutic modalities. The picture is even more complex, since mutations of the *CYLD* gene have been identified in patients with phenotypic features of either BSS, FC or MTF1 suggesting that these syndromes are clinical variants of the *CYLD* mutation-caused spectrum (Figure 2., Rajan *et al.*, 2011; Nagy *et al.*, 2015).

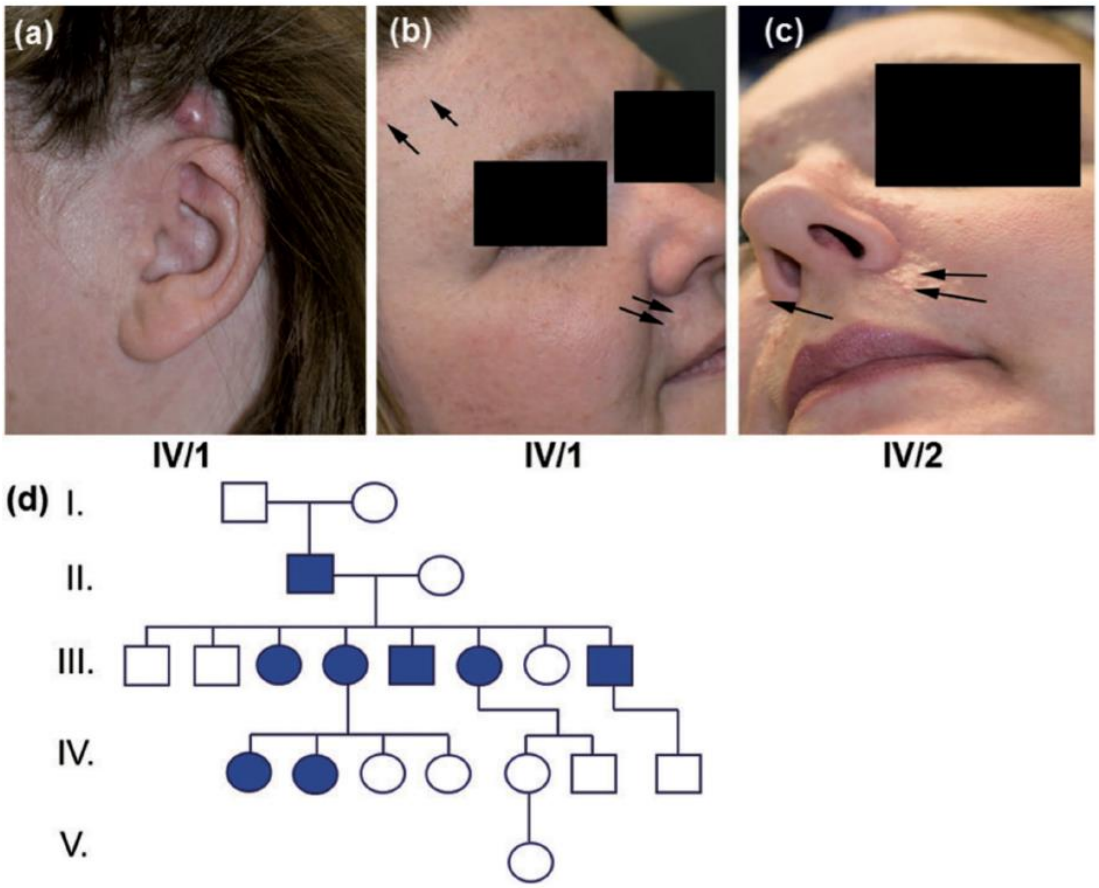

Figure 4. The investigated Anglo-Saxon pedigree from the North of England. Clinical pictures of the affected individuals from the 4th generation show tumors (a) above the ear, (b) on the forehead, and (c) around the nose. (d) The pedigree contains 8 affected family members spanning 5 generations (Nagy *et al.*, 2013).

A Hungarian pedigree from Bukovina (Romania) affected by BSS (Figure 3.) and an English BSS pedigree (Figure 3.) from Northern England were included in this

study. The clinical phenotypes of the affected family members and the pedigrees are reported in detail in a previous paper from our research group (Nagy *et al.*, 2013). The patients of the Hungarian BSS pedigree were presented with severe symptoms: numerous large tumors arising from different skin appendages and developing on the scalp, face and trunk mainly (Nagy *et al.,* 2013). The patients of the Anglo-Saxon pedigree showed moderate severity: few tumors, small size arising from different skin appendages developing primarily on the face (Figure 4., Nagy *et al.,* 2013).

Out of the previously reported patients, two Hungarian and two Anglo-Saxon ones, affected by the different phenotypes of BSS, but carrying the same disease-causing mutation (c.2806C>T, p.Arg936X) in the *CYLD* gene (Nagy *et al.,* 2013) were enrolled to the study.

1.4.2. Clinical variants of the *CTSC* mutation-caused spectrum

Papillon-Lefèvre syndrome (PLS) and Haim-Munk syndrome (HMS) are characterized by overlapping dermatological and dental symptoms, including hyperkeratosis of the palms and soles as well as severe periodontitis (Selvaraju *et al.*, 2003; Nagy *et al.,* 2014). Patients with PLS can also develop mild mental retardation, calcification of the dura mater, hyperhidrosis and increased susceptibility to infections (Gorlin *et al.,* 1964; Haneke *et al.,* 1979; Dalgic *et al.*, 2011). Specific features of HMS include pes planus, arachnodactyly, acroosteolysis and onychogryphosis (Papillon *et al.,* 1924; Haim *et al.*, 1965; Hart *et al.,* 1999).

The prevalence of PLS is approximately four cases per million, and, to date, approximately 300 cases have been reported worldwide. Parental consanguinity has been noted in more than 50% of these cases (Hewitt *et al.,* 2004). The prevalence of HMS is approximately one case per million, and the majority of reported cases are descendants of a few consanguineous families from a religious isolate in Cochin, India. One unrelated Brazilian patient has also been reported. Fewer than 100 HMS cases have been reported in the literature to date (Papillon *et al.,* 1924; Haim *et al.,* 1965; Hart *et al.,* 1999).

The ratio of affected males to females is 1:1 for both syndromes. PLS and HMS are both inherited in an autosomal recessive manner and develop as a consequence of mutations of the *CTSC* gene (Toomes *et al.,* 1999; Adkison *et al.,* 2002).

So far 89 *CTSC* gene mutations have been identified (Sulák *et al.*, 2016). The majority of these mutations have been detected in PLS patients, whereas only 4% have been associated with HMS (Selvaraju *et al.*, 2003; Nagy *et al.*, 2014).

In light of the reported PLS and HMS phenotypes and the associated *CTSC* mutations, it was hypothesized that PLS and HMS are the same entity with different phenotypic appearances (Table I., Sulák *et al.*, 2016). Although it is difficult to establish genotype–phenotype correlations, the elucidation of these correlations is likely to have significant clinical relevance for the development of the different clinical variants (PLS and HMS), the disease mechanism and the development of future therapeutic modalities.

	PLS	HMS
Clinical symptoms	Periodontitis, Palmoplantar hyperkeratosis	Periodontitis, Palmoplantar hyperkeratosis Arachnodactyly, Acroosteolysis, Pes planus, Onychogryposis
CTSC mutations	Any types of mutations	Any types of mutations

Table I. Comparison of the symptoms of the clinical variants of the *CTSC* mutation-caused spectrum.

The clinical phenotypes of the affected patients were reported in detail in a previous paper from our research group (Sulák *et al.*, 2016). A Hungarian woman presented with a typical HMS phenotype (Figure 5.): mild hyperkeratotic plaques were observed symmetrically on her palms and soles, onychogryphosis and arachnodactyly were noted on her fingers and pes planus on her soles.

A Hungarian male patient presented with the classical PLS phenotype (Figure 5.): moderate hyperkeratosis on his palms and soles. Both patients were missing all permanent teeth and using a permanent dental prosthesis. In our previous paper, we also reported haplotype analysis that raise the possibility that these patients are

siblings (Sulák *et al.*, 2016). It was not possible to genotype unaffected relatives (Sulák *et al.*, 2016).

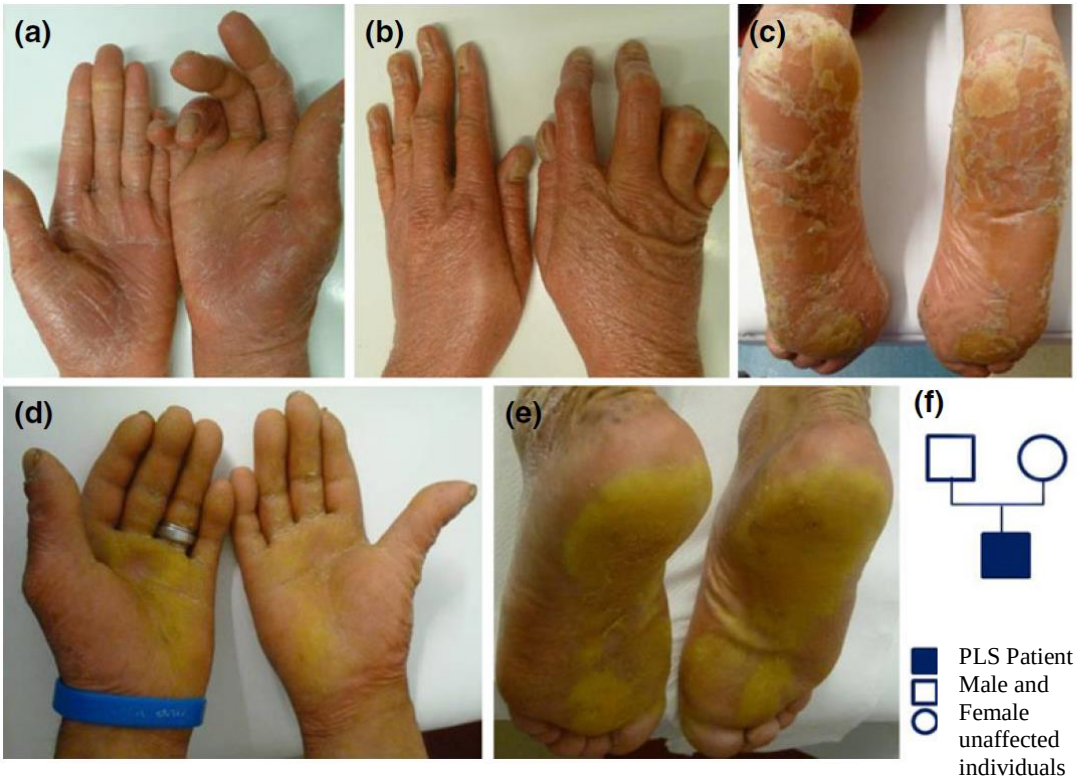

Figure 5. The clinical symptoms of the HMS and PLS Patients. HMS Patient exhibited (a) mild hyperkeratosis on her palms, (b) onychogryphosis and arachnodactly of her fingers and mild hyperkeratosis and (c) pes planus on her soles. PLS Patient was affected by moderate hyperkeratosis of his (d) palms and (f) soles, and exhibited no specific symptoms of HMS. (e) No other affected individuals are known in the family of the PLS Patient, and HMS Patient was brought up in state care without knowing her relatives (Sulák *et al.*, 2016).

The two previously reported Hungarian patients affected by PLS and HMS carrying the same disease-causing mutation (c.748C/T; p.Arg250X) of the *CTSC* gene were investigated (Sulák *et al.*, 2016) in the current study.

2. AIMS

2.1. Aims related to the *CYLD*-mutation caused clinical variants

In order to comprehensively profile coding variants and identify putative phenotype-modifying genetic factors - which could potentially explain the observed clinical differences between the Hungarian and the Anglo-Saxon BSS pedigrees carrying the same causative *CYLD* mutation - we performed whole exome sequencing (WES).

Here we report an investigation of Hungarian and Anglo-Saxon BSS pedigrees, which have been chosen as they both carry the same recurrent germline mutation in the *CYLD* gene (c.2806C>T, p.Arg936X), yet they show striking difference in their phenotypes (Nagy *et al.,* 2013). These two pedigrees provide an excellent opportunity to identify phenotype-modifying genetic factors that are responsible for the phenotypic diversity in BSS.

2.2. Aims related to the *CTSC*-mutation caused clinical variants

In order to identify putative phenotype-modifying genetic factors - which could explain the observed clinical differences between the PLS and HMS patients carrying the same causative *CTSC* mutation – WES was performed.

We have recently investigated two Hungarian patients, one with the symptoms of PLS and one with the symptoms of HMS, who nonetheless carry the same homozygous nonsense mutation (c.748C/T; p.Arg250X) of the *CTSC* gene (Sulák *et al.,* 2016). As there is currently no explanation for why one mutation can lead to these two different clinical variants (PLS and HMS), we became interested in the identification of phenotype-modifying genetic factors that could facilitate the understanding of the phenotypic differences between these patients.

3. PATIENTS AND METHODS

3.1. Patients

The clinical details of the investigated patients are demonstrated in the introduction part of the thesis. Their clinical features have been reported in the previous publications of our research group (Nagy *et al.*, 2013; Sulák *et al.*, 2016).

3.2. Methods

3.2.1. DNA samples

Out of the previously reported patients, two Hungarian and two Anglo-Saxon ones, affected by the different phenotypes of BSS, but carrying the same disease-causing mutation (c.2806C>T, p.Arg936X) in the *CYLD* gene, were investigated (Nagy *et al.*, 2013). DNA samples from the patients (n=4) were used for WES (performed by UD-GenoMed Medical Genomic Technologies Ltd., Debrecen, Hungary; http://www.ud-genomed.hu/).

Two previously reported Hungarian patients affected by PLS and HMS carrying the same disease-causing mutation (c.748C/T; p.Arg250X) of the *CTSC* gene were investigated (Sulák *et al.*, 2016). DNA samples from both patients were subjected to WES, which was performed by UD-GenoMed Medical Genomic Technologies Ltd.

The quality of the DNA samples was evaluated by agarose-gel electrophoresis.

3.2.2. Whole exome sequencing (WES)

Briefly, 4 µg of DNA with a concentration of 100 ng/µL were used for library construction. A liquid chip capture system (Agilent Research Laboratories, Santa Clara, CA, USA) was used to efficiently enrich all human exon regions. High-throughput deep sequencing was subsequently performed on the Illumina (San Diego, CA, USA) platform. An exon kit (SureSelect Human All Exon V6 Kit; Agilent) was used for library construction and capture experiments, and a bioanalyser (Model 2100; Agilent) was subsequently used to verify the library

insert size. The Illumina platform was used for sequencing according to the effective concentration of the library and the data output requirements. High-throughput paired-end sequencing (paired-end 150 bp; PE150) was performed.

3.2.3. Bioinformatics analysis

After WES was completed, bioinformatics analysis was performed, including quality assessment of sequencing data, single-nucleotide polymorphism (SNP) detection and whole exome association analysis.

The sequencing data quality control requirements were as follows: sequencing error rate of each base position < 1%, mean Q20 ratio > 90%, mean Q30 ratio > 80%, mean error rate < 0.1%, alignment rate for sequencing reads ≤ 95% and read depth of the base at one position ≥ 10X.

3.2.4. Single nucleotide polymorphism (SNP) testing

SNP testing was performed as follows: high-quality sequences were aligned with the human reference genome (GRCh37/hg19) to detect sequence variants, and the detected variations were analyzed and annotated. Variants were filtered according to read depth, allele frequency, and prevalence in genomic variant databases such as ExAc (v.0.3), ClinVar and Kaviar.

Variant prioritization tools (PolyPhen2, SIFT, LRT, Mutation Taster, Mutation Assessor) were used to predict the functional impact of the variants. All the identified candidate variants were confirmed by direct sequencing (Delta Bio 2000 Ltd., Szeged, Hungary; http://www.deltabio.hu/).

4. RESULTS

4.1. Identification of three putative genetic variants responsible for phenotypic variants in BSS

Gene	SNP	Polyphen2	SIFT	MutationTaster
ABCA13	rs74790141	Damaging	Deleterious	Disease causing
ARID1B	rs113232635	Unknown	Tolerated	Normal
DECR1	rs550991042	Unknown	Tolerated	Disease causing
EPB41L4A	rs17266567	Damaging	Deleterious	Disease causing
FGFR2	rs765066758	Benign	Tolerated	Disease causing
HLA-A	rs1136741	Damaging	Tolerated	Normal
HLA-A	rs1059563	Damaging	Tolerated	Polymorphism
HLA-A	rs9260179	Probably damaging	Tolerated	Polymorphism
HLA-A	rs9260180	Probably damaging	Tolerated	Polymorphism
KALRN	rs78202770	Damaging	Deleterious	Disease causing
KMT2C	rs74483926	Benign	Tolerated	Polymorphism
LAMB1	Chr10:107638858, ENST00000222399, c.A293G, p.N98S	Damaging	Deleterious	Disease causing
MKI67	rs45438392	Damaging	Tolerated	Normal
NBR1	rs202122812	Probably damaging	Deleterious	Disease causing
POM121L2	rs61736098	Damaging	Deleterious	Normal
SMO	rs111694017	Benign	Tolerated	Damaging
STAT3	rs1053023	Unknown	Unknown	Polymorphism
TNS1	rs140104262	Probably damaging	Deleterious	Disease causing
TRAF3	rs1131877	Benign	Tolerated	Polymorphism
TYW1B	rs181240185	Probably damaging	Deleterious	Unknown

Table II. WES identified 20 variants, which were all present in the Hungarian BSS patients, but not in the Anglo-Saxon ones.

A comparison of the WES data from the Hungarian and Anglo-Saxon BSS patients carrying the same disease-causing mutation (c.2806C>T, p.Arg936X) in the *CYLD* gene identified 20 genetic variants (Table II.), which were all present in the Hungarian patients, but not in the Anglo-Saxon patients. Based on the results of variant prioritization tools and the data of the literature, three of the 20 variants were suggested as putative phenotype-modifying polymorphisms. According to our current knowledge the other 17 variants are not associated with the functions of the CYLD enzyme.

The three putative phenotype-modifying polymorphisms are the followings: the rs1053023 SNP of the *signal transducer and activator of transcription 3 (STAT3)* gene, the rs1131877 SNP of the *tumor necrosis factor receptor-associated factor 3 (TRAF3)* gene and the rs202122812 SNP of the *neighbor of BRCA1 gene 1 (NBR1)* gene. The rs1053023 polymorphism is located in the 3'UTR region of the *STAT3* gene, while the other two polymorphisms (rs1131877 and rs202122812) are common missense variants of the *TRAF3* and *NBR1* genes, respectively. Pathogenicity predictions of the identified phenotype-modifying factors are summarized in Table III.

SNP (gene)	**Location (variant type)**	**SIFT Polyphen2 MutationTaster**	**Clinical associations**	**References**
rs1053023 *(STAT3)*	3' UTR variant	– – Polymorphism	Multiple sclerosis B-NHL BSS	Lu *et al.*, 2005 Butterbach *et al.*, 2018 Pap *et al.*, 2020
rs1131877 *(TRAF3)*	Exonic (Missense)	Tolerated Benign Polymorphism	Postradiotherapy toxicity BSS	De Ruyck *et al.*, 2011 Pap *et al.*, 2020
rs202122812 *(NBR1)*	Exonic (Missense)	Deleterious Probably damaging Disease causing	BSS	Pap *et al.*, 2020

Table III. Pathogenicity predictions and clinical associations of the identified phenotype-modifying factors.

4.2. Identification of two putative genetic variants responsible for phenotypic variants in HMS and PLS

A comparison of the WES data from the PLS and HMS patients carrying the same disease-causing mutation (c.748C/T; p.Arg250X) of the *CTSC* gene identified 35 putative genetic variants (Table IV.), which were present in the HMS patient, but not in the PLS one. The PLS patient did not carry any extra polymorphisms compared to the HMS patient.

Gene	SNP	Polyphen2	SIFT	MutationTaster
HLA- A	rs145046067	Probably damaging	Deleterious	Normal
HLA- A	rs150028516	Probably damaging	Deleterious	Normal
HLA- A	rs9260156	Unknown	Unknown	Disease causing
HLA- B	rs1050683	Benign	Tolerated	Normal
HLA- DQB1	rs41552812	Benign	Tolerated	Normal
HLA- DQB1	rs1071637	Damaging	Deleterious	Normal
HLA- DRB1	rs9269744	Probably damaging	Deleterious	Disease causing
HLA- DRB1	rs3830125	Benign	Deleterious	Unknown
HLA- DRB1	rs1136881	Probably damaging	Deleterious	Disease causing
HLA- DRB1	rs1071752	Benign	Deleterious	Disease causing
HLA- DRB1	rs201614260	Probably damaging	Tolerated	Normal
HLA- DRB1	rs148093782	Benign	Deleterious	Normal
HLA- DRB5	rs41553512	Probably damaging	Tolerated	Disease causing
HLA- DRB5	rs41559420	Benign	Tolerated	Disease causing
HLA- DRB5	rs1064587	Benign	Deleterious	Normal
HLA- DRB5	rs701884	Benign	Deleterious	Disease causing
HLA-DOA	rs2582	Benign	Unknown	Disease causing
HLA-DPB1	rs1042121	Benign	Tolerated	Disease causing
HLA-DQA1	rs12722039	Benign	Tolerated	Disease causing
HLA-DQA1	rs12722042	Damaging	Tolerated	Disease causing

HLA-DQA1	rs35087390	Benign	Deleterious	Disease causing
HLA-DQB2	rs9276570	Benign	Tolerated	Disease causing
HLA-DQB2	rs9276572	Damaging	Deleterious	Disease causing
HLA-DRA1	rs7192	Benign	Tolerated	Unknown
KIR2DL1	rs199644153	Benign	Tolerated	Normal
KIR2DL1	rs200746024	Benign	Tolerated	Normal
KIR2DL1	r538559888	Benign	Tolerated	Normal
KIR2DL2	rs3810343	Benign	Tolerated	Normal
KIR2DL2	rs35719984	Benign	Tolerated	Normal
KIR2DL2	rs78713511	Benign	Tolerated	Normal
KIR2DL2	rs200686594	Benign	Tolerated	Normal
KIR2DL3	rs662386	Benign	Deleterious	Normal
MICA	rs41554412	Damaging	Tolerated	Disease causing
OBP2A	rs55695858	Possibly damaging	Tolerated	Polymorphism
SH2D4A	rs34608771	Benign	Tolerated	Polymorphism

Table IV. WES identified 36 variants, which were all present in the HMS patient, but not in the PLS patient.

Based on the results of variant prioritization tools and the data of the literature, two out of the 35 variants are suggested as putative phenotype-modifying polymorphisms. According to our current knowledge the other 33 variants are not associated with the functions of the CTSC protein.

SNP (gene)	**Location (variant type)**	**SIFT Polyphen2 MutationTaster**	**Clinical associations**	**References**
rs34608771 *(SH2D4A)*	Exonic (missense)	Tolerated Benign Polymorphism	Development of the HMS phenotype	Pap *et al.*, 2020
rs55695858 *(OBP2A)*	Exonic (missense)	Tolerated Possibly damaging Polymorphism	Development of the HMS phenotype	Pap *et al.*, 2020

Table V. Pathogenicity predictions and clinical associations of the identified phenotype-modifying genetic factors.

The two putative phenotype-modifying polymorphisms are the followings: the rs34608771 SNP of the *SH2 domain containing 4A (SH2D4A)* gene and the rs55695858 SNP of the *odorant binding protein 2A (OBP2A)* gene. Both variants are common missense polymorphisms. Pathogenicity predictions for the identified phenotype-modifying factors are summarized in Table V.

5. DISCUSSION

Although the identification of disease-causing mutations is still extremely important for family planning and therapy, direct sequencing is unable to answer clinically relevant questions regarding phenotypic diversity and disease prognosis (Jarinova *et al.,* 2012; Timmerman *et al.,* 2014; Kiritsi *et al.,* 2015; Smith *et al.,* 2019). This limitation of direct sequencing was encountered with the Hungarian and Anglo-Saxon BSS pedigrees previously reported by our workgroup (Nagy *et al.,* 2013). Using haplotype analysis we have demonstrated that the presence of the same *CYLD* mutation in these geographically distant BSS families (the Hungarian one and the Anglo-Saxon one) is the consequence of two independent mutational events (Nagy *et al.,* 2013). These results suggest the position of this recurrent mutation is a mutational hot spot on the *CYLD* gene (Nagy *et al.*, 2013).

Comparing the WES data of the Hungarian and Anglo-Saxon BSS patients, we identified three putative phenotype-modifying genetic variants that potentially explain the striking phenotypic differences among patients carrying the same disease-causing *CYLD* mutation. The genes harbouring these genetic modifying variants code for proteins that are either directly or undirectly functionally linked to CYLD (Figure 6.)

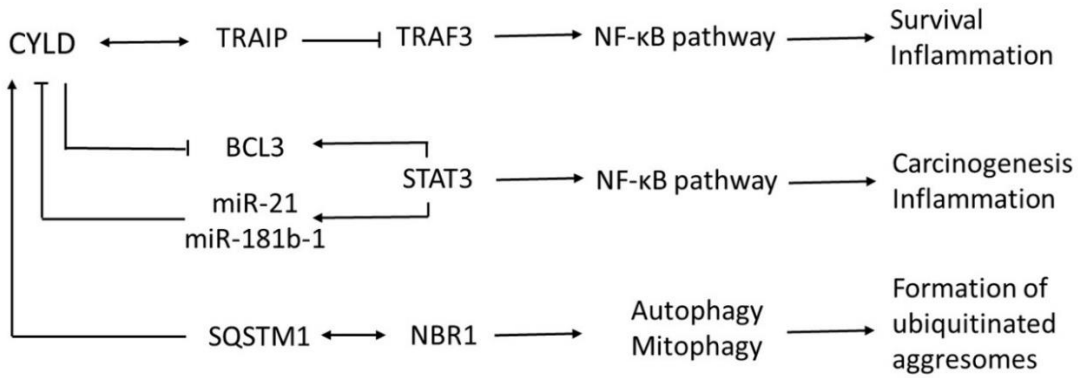

Figure 6. Schematic of the proposed mechanisms of the identified phenotype-modifying factors.

STAT3 is a transcription factor, which is constitutively activated in a variety of human cancers and plays critical roles in cancer cell survival, metastasis and angiogenesis (Yu *et al.,* 20009). STAT3 is activated by interleukin-6 (IL-6) and

directly activates certain microRNAs (miRs), such as miR-21 and miR-181b-1 (Aggarwal *et al.,* 2009). MiR-21 and miR-181b-1 inhibit the enzymatic activity of CYLD, leading to increased nuclear factor-κB (NF-κB) activity. Thus, STAT3 is not only a downstream target of IL-6 but, with miR-21, miR-181b-1 and CYLD, is part of the positive feedback loop that underlies the epigenetic switch that links inflammation to cancer (Iliopoulos *et al.,* 2010). The B-cell CLL/lymphoma 3 (BCL3) protein, which directly interacts with the STAT3 protein, is deubiquitinated by the CYLD enzyme, and abnormal BCL3 ubiquitination has been associated with the development of basal cell carcinomas (Chaudhary *et al.,* 2015). It has been recently reported that *BCL3* serves as an oncogene in cervical cancer and its oncogenic effect is mediated by *STAT3* (Zhao *et al.,* 2016). Interestingly, the rs1053023 SNP of the *STAT3* gene has already been associated with multiple sclerosis (Lu *et al.,* 2005) and reported in B-cell non-Hodgkin lymphoma (Butterbach *et al.,* 2011). However, this is the first study that suggests a potential phenotype-modifying role for *STAT3* in BSS.

TRAF3 is a member of the TRAF family of proteins, which serve as both crucial intracellular adaptors and E3 ubiquitin ligases that mediate signaling after the activation of various receptors. Receptors that signal through TRAF proteins include those involved in inflammation, innate immune responses, cell death and, most notably, interact with the followings: tumor necrosis factor receptors, Toll-like receptors, RIG-1-like receptors and interleukin-1 receptors (Wang *et al.,* 2010; Hacker *et al.,* 2011). The TRAF-interacting protein (TRAIP) interacts with TRAF3, while TRAIP is reported to interact with CYLD (Chapard *et al.,* 2012). TRAIP expression is increased in basal cell carcinomas and in multiple breast epithelial cell lines with oncogenic potentials ranging from non-malignant to highly invasive (Almeida *et al.,* 2011). Mutations in *TRAF3* and *CYLD* leading to constitutive activation of NF-κB have been identified in cancers, including multiple myeloma and solid tumors (Harhaj *et al.,* 2012). The rs1131877 SNP of the *TRAF3* gene is highly predictive for the development grade ≥2 acute esophageal postradiotherapy toxicity (De Ruyck *et al.,* 2011), and here we have demonstrated its association with the phenotypic diversity in BSS.

NBR1 is an autophagic adaptor protein involved in the efficient clearance of damaged mitochondria (Shi *et al.,* 2015). Briefly, upon mitochondrial damage, E3 ubiquitin ligases are recruited from the cytosol to depolarized mitochondria, where

they target damaged mitochondrial proteins for ubiquitination and bulk degradation by autophagy (Shi *et al.*, 2014). NBR1 is a functional homolog of sequestosome 1 (SQSTM1), another autophagic adaptor protein, which is a selective autophagy substrate that also acts as a cargo receptor for the degradation of other substrates (Svenning *et al.*, 2011). The interaction of CYLD with TRAFs is dependent upon SQSTM1, and the absence of SQSTM1 results in the reduction of the activity of the CYLD enzyme (Wooten *et al.*, 2008; Into *et al.*, 2010). The rs202122812 SNP of the *NBR1* gene has not been previously associated with any human diseases: this is the first study to indicate its clinical relevance in the development of the BSS-related phenotypic diversity.

The comparison of the WES data of the HMS and PLS patients identified a putative phenotype-modifying genetic variant (rs34608771 SNP) in the *SH2D4A* gene, which encodes a T-cell-expressed adapter protein that is expressed in T-cells, B-cells, macrophages and dendritic cells (Hashimoto *et al.*, 2000). SH2D4A regulates T-cell receptor (TCR) signal transduction in T-cells, and, in human T-cells, its expression increased in response to T-cell activation (Gonçalves *et al.*, 2018). SH2D4A is linked to cathepsin C via cystatin F. This latter protein is a cysteine-protease inhibitor expressed selectively in immune cells, such as T-cells, NK cells and dendritic cells (Obata-Onai *et al.*, 2002). The rs34608771 polymorphism of the *SH2D4A* gene has not been associated previously with any human diseases: this is the first study which links it to the development of the HMS clinical variant and raises its putative association with the phenotypic differences between PLS and HMS patients.

The other putative phenotype-modifying genetic variant (rs55695858 SNP) is located within the *OBP2A* gene, which encodes an odorant-binding carrier protein that has a known environmental biosensor function (Lacazette *et al.*, 2000). The OBP2A protein is expressed in the nasal structures, salivary and lachrymal glands and lungs, and, thus, has an oral sphere profile (Lapinski *et al.*, 2009 and 2019). OBP2A interacts with the glycosyltransferase 6 domain containing 1 (GLT6D1) protein, encoded by the *GLT6D1* gene, which has been identified as a susceptibility locus for periodontitis by genome-wide association studies and this association has been confirmed by several previous studies (Li *et al.*, 2009; Hasim *et al.*, 2015). Although genetic variants of the *OBP2A* gene have been implicated in influencing

the substrate-binding specificity of the encoded protein, none have previously been associated with the development of a human disease (Halfon *et al.*, 1998; Hamilton *et al.*, 2008). Since periodontitis is a major feature of the PLS and HMS phenotypes, we suggest that the rs55695858 SNP of the *OBP2A* gene might contribute to the phenotypic differences observed between PLS and HMS patients (Figure 7.).

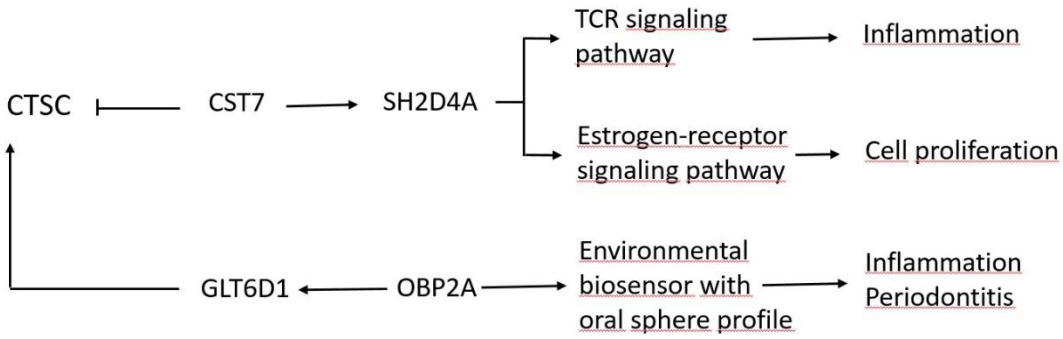

Figure 7. Schematic of the proposed mechanisms of the identified phenotype-modifying factors.

Further functional experiments are currently performed by our research group to unreveal how the modifying effects of these polymorphisms are realized.

6. CONCLUSIONS

Our study aimed to explain the phenotypic differences in BSS patients carrying the same disease-causing *CYLD* mutation by identifying phenotype-modifying genetic polymorphisms (Figure 8.) and to elucidate the phenotypic differences in PLS and HMS patients carrying the same disease-causing *CTSC* mutation by identifying phenotype-modifying genetic polymorphisms (Figure 9.).

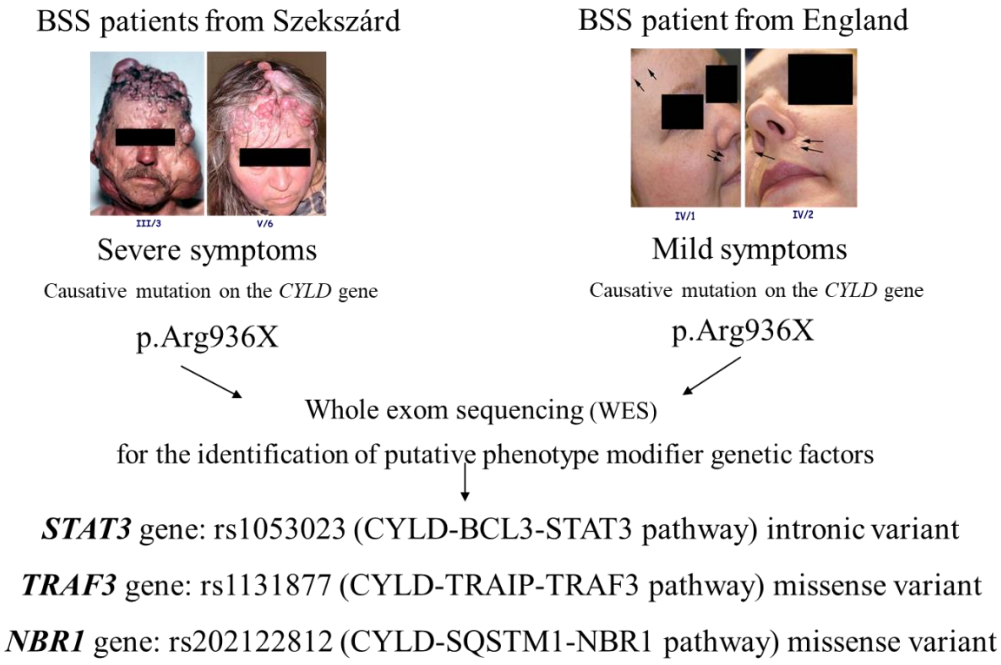

Figure 8. The identified putative phenotype-modifying genetic variants can explain the differences in the severity of the clinical symptoms between BSS patients from Szekszárd and England carrying the same disease-causing *CYLD* mutation (p.Arg936X).

It should be noted that, in addition to genetic factors, environmental or lifestyle factors might also contribute to the observed phenotypic differences between the investigated patients. Further functional studies are needed to prove the clinical relevance of the identified phenotype-modifying genetic factors and to describe the underlying mechanism that explain their phenotype-modifying roles. Our study contributes to the accumulating evidence supporting the clinical importance of phenotype-modifying genetic factors and their potential to facilitate the elucidation of genotype–phenotype correlations, phenotypic diversity and disease prognosis (Lee *et al.*, 2008).

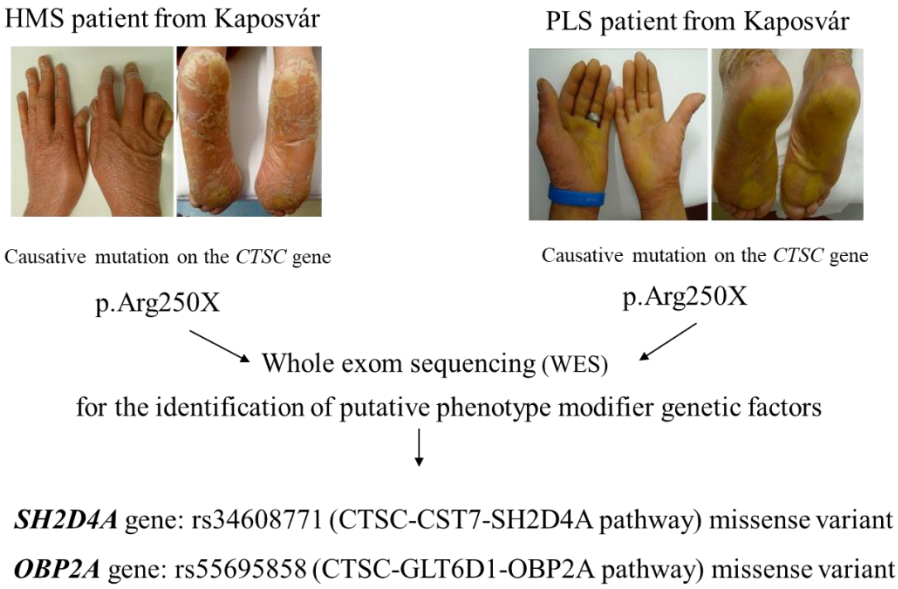

Figure 9. The identified putative phenotype-modifying genetic variants can explain the phenotypic differences between PLS and HMS patients carrying the same disease-causing *CTSC* mutation (p.Arg250X).

Our finding suggests relevant clinical implications. Although the identification of disease-causing mutations remains extremely important for family planning and for the development of novel causative therapeutic modalities, it is not always able to answer clinically relevant questions about phenotype diversity and disease prognosis. According to our view, the next big challenge in human genetics and clinical genetics is to identify these variants. We have joined to this new direction with these investigations demonstrated in my thesis. We are convinced that in the future the results of such investigations will highly contribute to the explanation of the phenotypic differences observed in monogenic diseases and contribute to the deep exploration of the genotype-phenotype relationships. In the clinical practice - primarily in genetic counseling - this knowledge will help us to make a more careful and accurate assessment of the prognosis of monogenic diseases.

.

7. SUMMARY

My investigations focused on the identification of putative phenotype modifying genetic factors in two groups of rare monogenic diseases, which are responsible for the observed phenotypic differences among the affected patients carrying the same disease-causing *CYLD* or *CTSC* mutations.

We recently investigated a Hungarian and an Anglo-Saxon pedigree affected by Brooke-Spiegler syndrome (BSS). Despite carrying the same disease-causing mutation (c.2806C>T, p.Arg936X) of the *CYLD* gene, the affected family members of the two pedigrees exhibit striking differences in their phenotypes. To identify phenotype-modifying genetic factors, WES was performed and the data from the Hungarian and Anglo-Saxon BSS patients were compared.

Three putative phenotype-modifying genetic variants were identified: the rs1053023 SNP of the signal transducer and activator of transcription 3 gene, the rs1131877 SNP of the tumor necrosis factor receptor-associated factor 3 gene and the rs202122812 SNP of the neighbor of BRCA1 gene 1 gene and the functional relevance of the genetic variants were predicted by *in silico* methods.

We investigated two Hungarian patients suffering from different phenotypic variants (PLS and HMS) but carrying the same homozygous nonsense *CTSC* mutation (c.748C/T; p.Arg250X). To gain insights into phenotype modifying associations, WES was performed for both patients and the results were compared to identify potentially relevant genetic modifying factors.

WES revealed two putative phenotype-modifying variants: a missense mutation (rs34608771) of the *SH2 domain containing 4A (SH2D4A)* gene encoding an adaptor protein involved in the intracellular signaling of the cystatin F, a known inhibitor of the cathepsin protein, and a missense variant (rs55695858) of the *odorant binding protein 2A (OBP2A)* gene influencing the function of the cathepsin protein through the glycosyltransferase 6 domain containing 1 (GLT6D1) protein. In silico analysis and literature search results explained how the effected proteins are linked to CTSC and possibly modify its functions.

Our investigations contribute to the accumulating evidence on the clinical importance of the identification of phenotype-modifying genetic factors. In the future it will have high potential in the elucidation of genotype–phenotype correlations and disease prognosis.

8. LIST OF ABBREVIATIONS

BCL3 = B-cell CLL/lymphoma 3

BSS = Brooke-Spiegler syndrome

CTSC = cathepsin C gene

CMT = Charcot-Marie-Tooth disease

CYLD = cylindromatosis gene

HMS = Haim-Munk syndrome

HGP = Human Genome Project

NBR1 = neighbor of BRCA1 gene 1 gene

NF-κB = nuclear factor-κB

PLS = Papillon-Lefèvre syndrome

SNP = single-nucleotide polymorphism

SQSTM1 = sequestosome 1

STAT3 = signal transducer and activator of transcription 3 gene

TRAF3 = tumor necrosis factor receptor-associated factor 3 gene

TRAIP = TRAF-interacting protein

WES = whole exome sequencing

9. REFERENCES

Adkison AM, Raptis SZ, Kelley DG *et al*. Dipeptidyl peptidase I activates neutrophil-derived serine proteases and regulates the development of acute experimental arthritis. J Clin Invest 2002;109:363-71.

Aggarwal BB, Kunnumakkara AB, Harikumar KB et al. Signal transducer and activator of transcription-3, inflammation, and cancer: how intimate is the relationship. Ann N Y Acad Sci. 2009;1171:59-76.

Almeida S, Ryser S, Obarzanek-Fojt M et al. The TRAF-interacting protein (TRIP) is a regulator of keratinocyte proliferation. J Invest Dermatol. 2011;131:349-57.

Bignell GR, Warren W, Seal S et al. Identification of the familial cylindromatosis tumour suppressor gene. Nat Genet. 2000;25:160–5.

Butterbach K, Beckmann L, de Sanjosé S et al. Association of JAK-STAT pathway related genes with lymphoma risk: results of a European case-control study (EpiLymph). Br J Haematol. 2011;153:318-33.

Chapard C, Hohl D, Huber M. The role of the TRAF-interacting protein in proliferation and differentiation. Exp Dermatol. 2012;21:321-6.

Chaudhary SC, Tang X, Arumugam A et al. Shh and p50/Bcl3 signaling crosstalk drives pathogenesis of BCCs in Gorlin syndrome. Oncotarget. 2015;6:36789-814.

Dalgic B, Bukulmez A, Sari S. Eponym: Papillon-Lefevre syndrome. Eur J Pediatr. 2011;170:689-91.

De Ruyck K, Sabbe N, Oberije C et al. Development of a multicomponent prediction model for acute esophagitis in lung cancer patients receiving chemoradiotherapy. Int J Radiat Oncol Biol Phys. 2011;81:537-44.

Evans CD. Turban tumour. Br J Dermatol. 1954;66:434–43.

Gonçalves PF, Harris TH, Elmariah T et al. Genetic polymorphisms and periodontal disease in populations of African descent: A review. J Periodontal Res. 2018;53:164-73.

Gorlin RJ, Sedano H, Anderson VE. The syndrome of palmar-plantar hyperkeratosis and premature periodontal destruction of the teeth: a clinical and genetic analysis of the Papillon-Lefevre syndrome. J Pediat 1964;65:895-908.

Hacker H, Tseng PH, Karin M. Expanding TRAF function: TRAF3 as a tri-faced immune regulator. Nat Rev Immunol. 2011;11:457–68.

Haim S, Munk J. Keratosis palmo-plantaris congenita, with periodontosis, arachnodactyly and a peculiar deformity of the terminal phalanges. Br J Dermatol. 1965;**77:**42-54.

Halfon S, Ford J, Foster J et al. Leukocystatin, a new class II cystatin expressed selectively by hematopoietic cells. J Biol Chem. 1998;273:16400–8.

Hamilton G, Colbert JD, Schuettelkopf AW et al. Cystatin F is a cathepsin C-directed protease inhibitor regulated by proteolysis. EMBO J. 2008;27:499-508.

Haneke E. The Papillon-Lefevre syndrome: keratosis palmoplantaris with periodontopathy: report of a case and review of the cases in the literature. Hum Genet. 1979;51:1-35.

Harhaj EW, Dixit VM. Regulation of NF-kappaB by deubiquitinases. Immunol Rev. 2012;246:107–24.

Hart TC, Hart PS, Bowden DW *et al*. Mutations of the cathepsin C gene are responsible for Papillon-Lefevre syndrome. J Med Genet. 1999;36:881-7.

Hashim NT, Linden GJ, Ibrahim ME et al. Replication of the association of GLT6D1 with aggressive periodontitis in a Sudanese population. J Clin Periodontol. 2015;42:319-24.

Hashimoto SI, Suzuki T, Nagai S et al. Identification of genes specifically expressed in human activated and mature dendritic cells through serial analysis of gene expression. Blood. 2000;96:2206–14.

Hewitt C, McCormick D, Linden G *et al*. The role of cathepsin C in Papillon-Lefevre syndrome, prepubertal periodontitis, and aggressive periodontitis. Hum Mutat. 2004;23:222-8.

Iliopoulos D, Jaeger SA, Hirsch HA et al. STAT3 activation of miR-21 and miR-181b-1 via PTEN and CYLD are part of the epigenetic switch linking inflammation to cancer. Mol Cell. 2010;39:493-506.

Into T, Inomata M, Niida S et al. Regulation of MyD88 aggregation and the MyD88-dependent signaling pathway by sequestosome 1 and histone deacetylase 6. J Biol Chem. 2010;285:35759-69.

Jarinova O, Ekker M. Regulatory variations in the era of next-generation sequencing: implications for clinical molecular diagnostics. Hum Mutat. 2012;33:1021-30.

Kelsall D. With a disease for every day, who will care for the orphans? CMAJ 2013;185:1475.

Kiritsi D, Valari M, Fortugno P et al. Whole-exome sequencing in patients with ichthyosis reveals modifiers associated with increased IgE levels and allergic sensitizations. J Allergy Clin Immunol. 2015;135:280-3.

Lacazette E, Gachon AM, Pitiot G. A novel human odorant-binding protein gene family resulting from genomic duplicons at 9q34: differential expression in the oral and genital spheres. Hum Mol Genet. 2000;9:289-301.

Lapinski PE, Oliver JA, Bodie JN et al. The T-cell-specific adapter protein family: TSAd, ALX, and SH2D4A/SH2D4B. Immunol Rev. 2009;232:240-54.

Lapinski PE, Oliver JA, Kamen LA et al. Genetic analysis of SH2D4A, a novel adapter protein related to T cell-specific adapter and adapter protein in lymphocytes of unknown function, reveals a redundant function in T cells. J Immunol. 2008;181:2019-27.

Lee DS, Park J, Kay KA et al. The implications of human metabolic network topology for disease comorbidity. Proc Natl Acad Sci U S A. 2008;105:9880-5.

Li T, Li W, Lu J et al. SH2D4A regulates cell proliferation via the ERalpha/PLC-gamma/PKC pathway. BMB Rep. 2009;42:516-22.

Lu J, Getz G, Miska EA et al. MicroRNA expression profiles classify human cancers. Nature. 2005;435:834-8.

Nagy N, Farkas K, Kemény L et al. Phenotype-genotype correlations for clinical variants caused by CYLD mutations. Eur J Med Genet. 2015;58:271-8.

Nagy N, Rajan N, Farkas K et al. A Mutational Hotspot in CYLD Causing Cylindromas: A Comparison of Phenotypes Arising in Different Genetic Backgrounds. Acta Derm Venereol. 2013;93:743-5.

Nagy N, Vályi P, Csoma Zs *et al*. CTSC and Papillon-Lefèvre syndrome: detection of recurrent mutations in Hungarian patients, a review of published variants and database update. Mol Gen & Genom Med. 2014;2:217-28.

Obata-Onai A, Hashimoto S, Onai N et al. Comprehensive gene expression analysis of human NK cells and CD8(+) T lymphocytes. Int Immunol. 2002;14:1085–98.

Papillon PH, Lefèvre P. Deuxcas de kératodermiepalmaire et plantairesymétriquefamiliale (maladie de Meleda) chez le frère et la soeur. Coexistence dans les deuxcasd'altérations dentaires graves. Bulletin de la Société française de dermatologie et de vénéorologie, Paris 1924;31:82-7.

Rajan N, Burn J, Langtry J et al. Transition from cylindroma to spiradenoma in CYLD-defective tumours is associated with reduced DKK2 expression. J Pathol. 2011;224:309–321.

Selvaraju V, Markandaya M, Prasad PV *et al*. Mutation analysis of the cathepsin C gene in Indian families with Papillon-Lefèvre syndrome. BMC Med Genet. 2003;4:5.

Shi J, Fung G, Deng H et al. NBR1 is dispensable for PARK2-mediated mitophagy regardless of the presence or absence of SQSTM1. Cell Death Dis. 2015;6:1943.

Shi J, Fung G, Piesik P et al. Dominant-negative function of the C-terminal fragments of NBR1 and SQSTM1 generated during enteroviral infection. Cell Death Differ. 2014;21:1432-41.

Smith DJ, Klein K, Hartel G et al. Mutations in the HFE gene can be associated with increased lung disease severity in cystic fibrosis. Gene. 2019;683:12-17.

Sulák A, Tóth L, Farkas K et al. One mutation, two phenotypes: a single nonsense mutation of the CTSC gene causes two clinically distinct phenotypes. Clin Exp Dermatol. 2016;41:190-5.

Svenning S, Lamark T, Krause K et al. Plant NBR1 is a selective autophagy substrate and a functional hybrid of the mammalian autophagic adapters NBR1 and p62/SQSTM1. Autophagy. 2011;7:993-1010.

Timmerman V, Strickland AV, Züchner S. Genetics of Charcot-Marie-Tooth (CMT) Disease within the Frame of the Human Genome Project Success. Genes (Basel). 2014;5:13-32.

Toomes C, James J, Wood AJ *et al*. Loss-of-function mutations in the cathepsin C gene result in periodontal disease and palmoplantar keratosis. Nat Genet. 1999;23:421-4.

Wang Y, Zhang P, Liu Y et al. TRAF-mediated regulation of immune and inflammatory responses. Sci China Life Sci. 2010;53:159-68.

Wooten MW, Geetha T, Babu JR et al. Essential role of sequestosome 1/p62 in regulating accumulation of Lys63-ubiquitinated proteins. J Biol Chem. 2008;283:6783-9.

Yu H, Pardoll D, Jove R. STATs in cancer inflammation and immunity: a leading role for STAT3. Nat Rev Cancer. 2009;9:798-809.

Zhao H, Wang W, Zhao Q et al. BCL3 exerts an oncogenic function by regulating STAT3 in human cervical cancer. Onco Targets Ther. 2016;9:6619-6629.

10. ACKNOWLEDGEMENT

First and foremost I would like to thank to my supervisors, Dr. Nikoletta Nagy and Prof. Dr. Németh Gábor for their great supervising activity.

My sincere thank also goes to Prof. Dr. Márta Széll, who gave access to perform the genetic investigations of this study at the Department of Medical Genetics, University of Szeged.

Special thanks to all my colleagues for their kind help in the Department of Medical Genetics and in the Department of Obstetrics and Gynecology, University of Szeged.

Last but not the least, I would like to thank to my family for supporting me for everything and encouraging me throughout this experience.

These investigations were supported by the Hungarian EFOP-3.6.1-16-2016-00008.