Update on SLC26A3 Mutations in Congenital Chloride Diarrhea

Satu Wedenoja, Elina Pekansaari, Pia Hoglund, Siru Mäkelä, Christer Holmberg, Juha Kere

Human Mutation

Update on SLC26A3 Mutations in Congenital Chloride Diarrhea

Journal:	*Human Mutation*
Manuscript ID:	humu-2010-0633.R1
Wiley - Manuscript type:	Mutation Update
Date Submitted by the Author:	27-Feb-2011
Complete List of Authors:	Wedenoja, Satu; University of Helsinki, Department of Medical Genetics Pekansaari, Elina; University of Helsinki, Department of Medical Genetics Hoglund, Pia; Rinnekoti Foundation Mäkelä, Siru; Helsinki University Central Hospital, Department of Oncology Holmberg, Christer; University of Helsinki, Hospital for Children and Adolescents Kere, Juha; Karolinska Institutet, Department of Biosciences and Nutrition, and Clinical Research Centre
Key Words:	SLC26A3, chloride diarrhea, CLD, mutation, anion transport

Update on *SLC26A3* Mutations in Congenital Chloride Diarrhea

Satu Wedenoja[1], Elina Pekansaari[1], Pia Höglund[2], Siru Mäkelä[3], Christer Holmberg[4], Juha Kere[1,5]

[1]Department of Medical Genetics, University of Helsinki, Helsinki, Finland

[2]Rinnekoti Foundation, Espoo, Finland

[3]Department of Oncology, Helsinki University Central Hospital, Helsinki, Finland

[4]Hospital for Children and Adolescents, University of Helsinki, Helsinki, Finland

[5]Department of Biosciences and Nutrition, and Clinical Research Centre, Karolinska Institutet, Huddinge, Sweden

Key words: SLC26A3; chloride diarrhea; CLD; mutation; anion transport

Correspondence to:

Satu Wedenoja, MD, PhD

Department of Medical Genetics

Biomedicum Helsinki, B316a, P.O. Box 63

FI-00014 University of Helsinki

FINLAND

tel. +358-(0)9-19125638, fax +358-(0)9-19125624

e-mail: satu.wedenoja@helsinki.fi

Abstract

Congenital chloride diarrhea (CLD) is an autosomal recessive disorder with around 250 cases reported so far. Life-long secretory diarrhea is caused by mutations in the solute carrier family 26 member 3 (*SLC26A3*) gene disrupting the epithelial Cl^-/HCO_3^- transport in the ileum and colon. Although salt substitution allows favorable outcome, possible manifestations include renal impairment, intestinal inflammation, and male infertility. At least 55 mutations, of which 21 (38%) novel are reported here, cause CLD. Majority of the mutations are single-nucleotide substitutions (n=30; 55%) with 18 missense, 7 nonsense, and 5 splice-site mutations. Additional mutations are minor deletions/insertions or their combinations (n=21; 38%), major deletions (n=3; 5%), and a major insertion (n=1; 2%). Distinct founder mutations appear in Finland, Poland, and Arab Countries, whereas patients from other countries carry rare homozygous or compound heterozygous mutations. None of the studied SLC26A3 mutants shows significant Cl^-/HCO_3^- exchange activity *in vitro*, and accordingly, evidence of genotype-phenotype differencies remain nonexistent. The domain interaction between SLC26A3 and the cystic fibrosis transmembrane conductance regulator (CFTR) raises a possibility of CFTR modulation in the pathogenesis of CLD. This review summarizes the current knowledge of *SLC26A3* mutations and polymorphisms, and their biological and clinical relevance.

Deleted: any

Background

Congenital chloride diarrhea (CLD; OMIM #214700) is a rare diasease with an autosomal recessive inheritance (Norio et al., 1971). It is characterized by chronic secretory diarrhea containing an excess of chloride. Due to the intrauterine onset of diarrhea, CLD pregnancies are complicated by polyhydramnios and infant meconium is lacking (Holmberg, 1986). Children with CLD are often premature and soon after birth, profuse diarrhea leads to dehydration, hypochloremic and hypokalemic metabolic alkalosis, and failure to thrive. CLD diagnosis is based on the clinical picture and measurement of high fecal chloride (>90 mmol/L) after correction of the fluid and salt depletion. While untreated disease is mostly lethal during the first weeks or months of life, oral salt substitution with NaCl and KCl allows normal growth and development, and favorable outcome (Holmberg, 1986; Hihnala et al., 2006a; Wedenoja et al., 2010). Untreated or poorly treated disease is associated with impaired renal function and nephrocalcinosis, and even with end-stage renal disease (Wedenoja et al., 2008b). Associated morbidity involves intestinal inflammation, hyperuricemia, inguinal hernias, spermatoceles, and male subfertility (Hihnala et al., 2006a; Hoglund et al., 2006).

After the first CLD descriptions in two Americans partly of Italian descent, more than 250 cases have been reported worldwide with around one-fifth of the patients originating from Finland (Darrow, 1945; Gamble et al., 1945; Hoglund et al., 1998). For the rarity of CLD, most of its clinical and genetic studies have been performed in the genetically homogeneous series of Finland, where the incidence is in the range of 1 in 30,000 to 40,000 live births. Other countries with a higher than global incidence are Poland (1 in 200,000), Kuwait, and Saudi Arabia. In the countries around the Persian Gulf, and especially in Kuwait, local estimates of incidence total as high as around 1 in 3,200 due to consanguineous marriages

Deleted: -

Deleted: -average

(Hoglund et al., 1998). Single cases with CLD appear worldwide, both in developing and more affluent countries, making diagnostics challenging. If a suspicion of CLD arises, especially in low-incidence countries, mutation analysis may be required to establish the diagnosis (Hoglund et al., 1996; Makela et al., 2002).

CLD is caused by mutations in the solute carrier family 26 member 3 (*SLC26A3)* gene (OMIM #126650) (Kere et al., 1993; Hoglund et al., 1996). Initially, *SLC26A3* was cloned with the name down-regulated in adenoma (*DRA*) as a potential tumor suppressor gene based on its abundant expression in the colon but its absence from colonic adenomas and adenocarcinomas (Schweinfest et al., 1993; Taguchi et al., 1994). The genomic sequence of *SLC26A3* spans around 37.8 kb on the minus strand of chromosome 7q31.1 (www.ncbi.nlm.nih.gov: NG_008046.1). The full-length cDNA, around 3.2 kb on Northern blot (Schweinfest et al., 1993) (www.ncbi.nlm.nih.gov: NM_000111.2), has an open reading frame of 2,295 bp, the 5´untranslated region (UTR) of 211 bp, and the 3´ UTR of 388 bp. The promoter region fragment flanking the area between -398 and -102 bp is responsible for the basal level of *SLC26A3* transcription, and the region between -398 and -688 bp is necessary for high-level transcriptional activation induced by sodium butyrate (Alrefai et al., 2007). Total of 21 exons, of which 20 are coding, range from 55 to 234 bp in size. The translation start codon ATG (Met 1) in exon 2 is followed by the termination codon TAA in exon 21 (Haila et al., 1998).

SLC26A3 encodes for an 85-kDa transmembrane glycoprotein with 764 amino acids forming the predicted 12 to 14 hydrophobic membrane-spanning domains in addition to the cytoplasmic N- and C-terminal domains (Byeon et al., 1996; Bairoch et al., 2005). The SLC26A3 protein is an apical exchanger which transports Cl^- across the apical cell membrane in exchange of HCO_3^- (Hoglund et al., 1996; Moseley et al., 1999) and in conjunction with other epithelial ion transporters (Melvin et al., 1999; Ko et al., 2002; Lamprecht and Seidler,

Deleted: ; Kere et al., 1993

Deleted: Jacob

2006). C-terminal amino acids 525 to 720 form the cytoplasmic sulphate transporter and anti-sigma antagonist (STAS) domain—a binding site for the R domain of cystic fibrosis transmembrane conductance regulator (CFTR) (Aravind and Koonin, 2000; Ko et al., 2002). Four C-terminal amino acids of SLC26A3, glutamic acid-threonine-lysine-phenylalanine (ETKF), form PDZ [postsynaptic density protein (PSD95), the Drosophila homologue Disc-large, and the tight junction protein ZO-1] interaction motif which interacts with several PDZ adapter proteins and couples the activity of SLC26A3 and Na^+/H^+ exchanger 3 (NHE3) (Melvin et al., 1999; Lamprecht and Seidler, 2006). Loss of the SLC26A3-mediated Cl^-/HCO_3^- exchange activity in the ileum and colon, with secondary disruption of the Na^+/H^+ transport, results in intestinal loss of both NaCl and fluid, and to the profuse diarrhea of CLD (Moseley et al., 1999; Melvin et al., 1999; Holmberg, 1978).

Deleted: ate

In both humans and rodents, expression of *SLC26A3* mRNA and protein emerges in the absorptive surface epithelium of the small intestine, especially in the ileum but also in the duodenum, and in the surface epithelium of the colon. In the intestinal crypts, the signal is lacking (Hoglund et al., 1996; Jacob et al., 2002). In the human male reproductive tract, SLC26A3 expression appears in the testis, efferent ducts of the testis and proximal epididymis, and in the epithelium of the seminal vesicle (Hihnala et al., 2006b). Moreover, SLC26A3 is expressed in the epithelium of the human sweat gland (Haila et al., 2000). In cell cultures, expression of SLC26A3 appears in epithelial cells derived from the pancreatic duct (Greeley et al., 2001) and trachea (Wheat et al., 2000). In intestinal adenomas and adenocarcinomas, expression of SLC26A3 is significantly down-regulated or even undetectable (Schweinfest et al., 1993; Antalis et al., 1998). In addition to colon cancer, reduced *SLC26A3* expression may play a role in the pathogenesis of ulcerative colitis (Yang et al., 1998; Asano et al., 2009), and reduction of *SLC26A3* promoter activity by the

inflammatory mediator interferon-gamma (IFN-γ) may contribute to the pathogenesis of intestinal inflammation (Alrefai et al., 2007).

SLC26A3 belongs to he human solute carrier family 26 (SLC26) with 11 structurally homologous anion exchangers (Mount and Romero, 2004). Three of these anion exchangers are associated with rare autosomal recessive diseases: SLC26A3 with CLD (OMIM #214700), SLC26A4 with Pendred syndrome (OMIM #274600) (Everett et al., 1997), and SLC26A2 (alias DTDST) with a continuum of skeletal dysplasia phenotypes including diastrophic dysplasia (OMIM #222600), achondrogenesis type IB (OMIM #600972), atelosteogenesis type II (OMIM #256050), and multiple epiphyseal dysplasia (OMIM #226900) (Hastbacka et al., 1994; Superti-Furga et al., 1996; Hastbacka et al., 1996; Superti-Furga et al., 1999).

Variants of the *SLC26A3* gene

After finding that *SLC26A3* causes CLD (Hoglund et al., 1996), altogether 30 different mutations and 4 polymorphisms have been reported by the Finnish researchers (Makela et al., 2002), and 4 mutations by others (Heinz-Erian et al., 2008; Choi et al., 2009). When the 21 novel mutations of this paper are added to the spectrum of *SLC26A3* mutations, these novel changes represent 38% of the total 55 mutations (Table 1).

The reported *SLC26A3* mutations have been identified by automated sequencing after PCR amplification of genomic DNA with intronic primers to cover all the exons and the complete coding sequence (Haila et al., 1998; Heinz-Erian et al., 2008), except for the recent paper with 3 novel mutations identified by whole-exome sequencing (Choi et al., 2009). The mutation nomenclature is partly different from the previous mutation update (Makela et al., 2002) and based on the current recommendations of the Human Genome Variation Society

(HGVS) checked using the Mutalyzer program (http://www.LOVD.nl/mutalyzer/) (den Dunnen and Antonarakis, 2000; den Dunnen and Paalman, 2003; Wildeman et al., 2008). As for reference, RefSeq sequences NG_008046.1 (genomic) and NM_000111.2 (cDNA) were used with the nucleotide numbering +1 denoting the first nucleotide (A) of the translation start codon (ATG; codon 1) in exon 2 of the *SLC26A3* gene.

Mutation types

Majority of the *SLC26A3* mutations are single-nucleotide substitutions (n=30; 55%). Of these, 13 are transversions with a substitution of a purine for a pyrimidine or vice versa, and 17 are transitions changing a purine to another purine or a pyrimidine to another pyrimidine. This excess of transitional over transversional substitutions, resulting from the relatively high rate of mutation of methylated cytosines to thymine, is consistent with the findings throughout the genome (Arnheim and Calabrese, 2009). At protein level, 18 (60%) of point mutations result in a change of one amino acid (missense), 7 (23%) introduce a stop codon resulting in truncation of the protein product (nonsense), and 5 (17%) are splice site mutations most probably affecting RNA splicing.

Deleted: .

Additional mutations are minor deletions (n=15; 27%) and insertions (n=3; 5%) or their combinations (n=3; 5%). Only three major deletions (n=3; 5%) and one major insertion (n=1; 2%) have been found in association with CLD. All major deletions result in loss of one or several exons and, most likely, absence of the SLC26A3 protein. The largest deletion found in association with CLD is the 8.6 kb deletion reported in this study. This mutation was found in a Swedish patient with Lebanese origin. As for other mutations, direct sequencing was used in mutation analysis. After the failed PCR amplification of exon 1, we selected several SNPs (single nucleotide polymorpshims) in the promoter and intron 1 and identified the borders of the deletion by finding the closest markers to exon 1 present in the patient DNA.

Thereafter, we designed primers around the approximate deletion, spanning from 4.3 kb upstream and 4.2 kb downstream from the exon 1, and PCR amplified and sequenced the exact nucleotides around the breakpoint of the deletion. Similarly, we proved the carriership of this deletion in the parents who were first cousins. As this large deletion removes not only the non-coding exon 1 but also the promoter region essential for the basal level of *SLC26A3* transcription (Alrefai et al., 2007), it is the only mutation that, most likely, disrupts transcription of the gene. The clinical phenotype of this patient is, however, similar to that observed in other patients with different mutations for CLD.

When larger deletions and insertions or their combinations are excluded, all minor deletions and insertions except for 3 introduce a frame-shift with a change of some amino acids, and finally, stop codon terminating the protein synthesis. Those 3 CLD-associated mutations with in-frame deletion/insertion are the Finnish founder mutation c.951_953del (p.V318del) in exon 8, the Polish founder mutation c.2024_2026dup (p.I675dup) in exon 18, and c.1579_1581del (p.Y527del) in exon 14 found in one Polish patient (Table 1).

Deleted: one

Deleted: TAT

Intragenic mutations

Among the all mutations, 82 % (n=45) are located in exons, 13 % (n=7) in or next to intronic splice sites, and 5% (n=3) involve deletion of both exonic and intronic sequences. Although the mutations are scattered across the whole gene, their distribution is uneven (Figure 1A). While 6 mutations (11%) appear both in exon 5 and 12, none of the mutations is located in exons 2, 16, 20, and 21. Most of the mutations are scattered between the exons 3 and 6, and between the exons 12 and 15. At least codon 120, harboring the mutation c.358G>A (p.G120S) in four independent families from different populations, codon 124 with the c.371A>T (p.H124L) mutation reported in 5 patients from 3 different populations, and codons 131 and 454, involving three and two different mutations at the same locus, are

evident mutation hotspots (Table 1). In the case of transitions, such as p.G120S, the likely mechanism for mutations is methylation of cytosine-guanine (CpG) dinucleotides while microdeletions and insertions scattered between the exons 12 and 15 may be associated with short nucleotide repeats (Makela et al., 2002; Arnheim and Calabrese, 2009).

Of those 7 intronic splice site changes, 5 are located at the 3' splice site which terminates the intron with an almost invariant AG sequence. Mutations in this splice acceptor sites are likely to cause splicing error and loss of function of the protein. Although one of the novel splice site mutations is located next to the concensus sequence GU of the 5' splice site (c.2205+3A>G), its heterozygous presence in a patient with another heterozygous mutation (c.951_953del; p.V318del) and typical clinical picture of CLD predict disease-associated disruption of splicing at the splice donor site. In addition, homozygous insertion of an *Alu* Y subfamily sequence (GenBank U14569.1) next to the intron donor site (c.131+2_131+3insU14569.1) is likely to disrupt protein synthesis in a patient with both parents, being first cousins, carrying this insertion in a heterozygous form (Table 1).

Deleted: ,

Founder Mutations

In Finland, genetic studies of CLD have revealed the single ancestral founder mutation c.951_953del (p.V318del) in a homozygous form in all but one patient. Similarly, the founder mutation c.559G>T (p.G187X) accounts for >90% of CLD-causing mutations in Arab Countries, whereas the Polish major mutation c.2024_2026dup (p.I675dup) is involved in almost 50% of the CLD-associated chromosomes (Hoglund et al., 1998). Outside these founder populations, mutation analysis is very likely to show rare, and often previously unreported, *SLC26A3* mutations responsible for CLD (Table 1).

De Novo Mutations

So far, we have been able to verify the carriership of heterozygous *SLC26A3* mutations in all the parental DNA samples studied for the mutation(s) found in their child with CLD. Although *de novo* mutations remain nondetermined so far, the vast spectrum of *SLC26A3* mutations suggests that novel mutations in this gene may be more common than expected based on the CLD incidence (Hoglund et al., 1998).

Polymorphisms

All the SNPs reported in the coding region of *SLC26A3* are presented in Table 1. Of these 17 SNPs, 4 has been identified in association with genetic studies of CLD (Makela et al., 2002) and 16 exist in the dbSNP public database among the total of 311 *SLC26A3* SNPs (Sherry et al., 2001). Only 5 of the coding-area SNPs are silent, while the other SNPs alter the amino acid sequence (Table 1). Functional data confirming the neutral role of an sequence altering variant exist only for the missense change c.921T>G (p.C307W) (Hoglund et al., 1996; Moseley et al., 1999).

Biological Relevance

Animal Model of CLD

The SLC26A3 protein is the colon's most essential apical anion transporter. Evidence for this is not only the life-threatening diarrheal phenotype of CLD but also existence of *SLC26A3*-deficient mice with a diarrheal phenotype closely similar to that of CLD. S*LC26A3*-deficient mice exhibit, however, unique features uncharacteristic of human CLD. These include significant growth retardation, occasionally development of prolapsed rectums, and an

aberrant growth pattern of the colonic mucosa with an expanded proliferative zone of colonic crypts and hyperplasia of the surface epithelium (Schweinfest et al., 2006).

Protein structure

Although the exact structure of the SLC26A3 protein remains undetermined, the UniProt prediction suggests the presence of 12 to 14 transmembranic protein domains (Bairoch et al., 2005). Using the prediction of 12 domains and those 45 mutations (82% of total 55 mutations) associated with the coding sequence only, 2 mutations are located in the cytoplasmic N terminal domain, 28 between the first and last transmembranic domain, and 15 in the cytoplasmic C terminus (Figure 1B).

Deleted: s

STAS domain, the amino acids 525 to 720 (Aravind and Koonin, 2000; Bairoch et al., 2005), involves 10 different mutations representing 18% of all mutations. The regulatory action through the R domain of CFTR and STAS domain of SLC26A3 seems to be essential for epithelial HCO_3^- secretion in the tissues where both these transporters are expressed, meaning at least the small intestine (Jacob et al., 2002; Greger, 2000; Rossmann et al., 2005), pancreatic duct (Greeley et al., 2001; Ko et al., 2004), and tracheal epithelium (Wheat et al., 2000). As this essential physical association between CFTR and SLC26A3 requires intact R and STAS domains, and at least one CLD-causing mutation on the STAS domain of SLC26A3 prevent the activation of CFTR as well, CFTR may also play a role in the pathogenesis of CLD (Ko et al., 2004).

Deleted: ; Aravind and Koonin, 2000

None of the *SLC26A3* mutations is located in the PDZ motif, comprising the four C-terminal amino acids. This motif enables the binding with PDZ adapter proteins and couples the action of SLC26A3 with other transporters such as NHE3 to promote electroneutral NaCl absorption (Melvin et al., 1999; Lamprecht and Seidler, 2006). To underline the importance of this coupled transport, *NHE3* knock-out mice have increased levels of intestinal *SLC26A3*

mRNA, and *SLC26A3*-deficient mice show massive upregulation of intestinal *NHE3* mRNA and protein (Melvin et al., 1999; Schweinfest et al., 2006).

Functional Studies on SLC26A3 mutations

Functional data on mutant forms of SLC26A3 are limited. Regarding the Finnish founder mutation c.951_953del (p.V318del), the anion exchange activity is lacking despite the normal levels of *SLC26A3* mRNA and protein in the CLD colon (Hoglund et al., 1996; Moseley et al., 1999;). An associated feature of this mutation is a substitution of tryptophan for cysteine ten codons upstream c.921T>G (p.C307W), the frequency of which alone is around 14% in the healthy Finnish population. This variant is compatible with normal anion exchange activity, neutral in conjunction with the mutation c.951_953del (p.V318del), and is therefore considered a meaningless *SLC26A3* polymorphism (Hoglund et al., 1996; Moseley et al., 1999).

The mouse p.V318del-homologue mutant (p.V310del) is retained in the endoplasmic reticulum, failing to show any expression on the apical plasma membrane. Similar to the human p.V318del, some other CLD-associated mouse mutants, e.g., p.L489R (human p.L496R) and p.I668-669ins (human p.I675dup), are expressed on the plasma membrane (Ko et al. 2002). The mutations located in the STAS domain are likely to result in loss of the functional protein at the plasma membrane either by disrupting the STAS domain directly (p.I675dup and p.G702Tins9) or by affecting protein folding and/or trafficking pathway (p.Y527del and p.I544N) (Dorwart et al., 2008). Among the functionally studied mutants (those mentioned above plus p.S706AfsX6), only p.Y527del shows some Cl^-/HCO_3^- exchange activity in cell cultures (Ko et al., 2002; Chernova et al., 2003; Dorwart et al., 2008).

Deleted: , as expected,

Deleted: (Ko et al., 2002)

Deleted: None of

Deleted: the studied mutants

Deleted: show,

Deleted: however, any

Deleted:

Formatted: Font color: Auto, English (U.K.)

Clinical and Diagnostic Relevance

Differential Diagnosis

The simple measurement of fecal Cl^- is still sufficient to confirm the diagnosis in most of the cases. Differential diagnosis of CLD includes other inherited diarrheas such as congenital sodium diarrhea (Holmberg and Perheentupa, 1985; Booth et al., 1985), glucose-galactose malabsorption (Martin et al., 1996), congenital sucrase-isomaltase deficiency (Ouwendijk et al., 1996), lysinuric protein intolerance, acrodermatitis enteropathica, and microvillus inclusion disease (Field, 2003). In these diseases and in other secretory diarrheas, alkaline stool pH and metabolic acidosis exclude the possibility of CLD, the hallmarks of which are acidic stools and metabolic alkalosis (Holmberg, 1986; Field, 2003). Importantly, CLD may be confused with other salt-wasting diseases, such as Bartter syndrome (Choi et al., 2009).

Genotype-Phenotype Correlation

Despite the various types of mutations and their wide distribution in different regions of the *SLC26A3* gene, evidence of genotype–phenotype differences in CLD is nonexistent (Makela et al., 2002). Even the identical genetic background of CLD may, however, result in variable clinical course if the diagnosis is delayed or salt substitution insufficient, resulting in severe dehydration, and thereby, reduction in the amount of diarrhea (Holmberg et al., 1977; Hoglund et al., 2001). In addition, the patients with even the same genotype may show different responses to diarrhea-modulating agents such as butyrate, shown to reduce the amount of diarrhea in an Italian patient but not in the Finnish series of CLD (Canani et al., 2004; Wedenoja et al., 2008a). Whether different mutations modulate disease characteristics such as male subfertility, described only in the highly homogenous series of Finland so far (Hoglund et al., 2006), remains to be determined.

Deleted: -

Molecular Diagnostic Strategies

So far, genetic testing for CLD is available by a clinical laboratory (HUSLAB, Helsinki University Central Hospital) only for the Finnish founder mutation c.951_953del (p.V318del). After finding the gene for CLD (Hoglund et al., 1996), more than 160 patients with a clinical suspicion have been tested for *SLC26A3* mutations in our laboratory at research basis. Because CLD is rare and the outcome of treated disease is favorable, prenatal diagnostics or carrier testing are rarely requested or performed. In addition to our laboratory, some other research laboratories have recently reported CLD-associated *SLC26A3* mutations (Heinz-Erian et al., 2008; Choi et al., 2009).

In patients originating from founder populations, a feasible strategy for mutation analysis is to search for the known founder mutations present in 99% of the Finnish, >90% of the Arabic, and around 50% of the Polish CLD-associated chromosomes (Hoglund et al., 1998). In contrast, most patients outside these geographical regions are likely to have rare, and often previously unreported, homozygous or compound heterozygous mutations identified only through sequencing of the whole coding region of *SLC26A3* (Table 1; Figure 1). At least rare homozygous mutations for CLD, such as the large homozygous deletion of 8.6 kb reported here, are likely to arise from areas where consanguineous marriages are relatively common, for instance from the countries around the Persian Gulf (Hoglund et al., 1998). On the other hand, those rare mutations found in single patients from different countries may actually arise from the same population, and therefore, the country of residence may not reveal the roots of the families (Table 1). Because single cases with CLD arise worldwide and the spectrum of mutations is wide, such CLD infrequency makes misdiagnoses highly probable. Especially in low-incidence countries, negative family history in association with difficulties in refining the early clinical diagnosis are likely to delay the diagnosis and result in lethal complications

in early childhood, or in those with the delayed diagnosis, induce retarded growth, and later, development of chronic kidney disease (Wedenoja et al., 2008b).

Our mutation analysis has failed to find CLD-causing mutations in 6 patients so far (Makela et al., 2002). Certainly, any difficulties in assessing phenotypic characteristics may make genetic diagnostics unreliable, or alternatively, mutations may be localized in the areas uncovered by sequencing, such as promoter and non-coding regions outside the exon-intron boundaries. In addition, any major deletions or insertions may be simply missed through direct sequencing. Most probably, these events are relatively rare in association with CLD, supported by the high success rate of our mutation analysis and by the fact that larger deletions or insertions have been found in only 4 of these 165 patients studied so far (Table 1).

When many of the *SLC26A3* mutations found are novel, prediction of their pathogenicity may remain uncertain without functional testing. *SLC26A3* is, however, likely to be sensitive for any coding region mutations because (i) polymorphisms in the coding region are relatively rare (Table 1), (ii) all the studied mutants result in loss of the Cl^-/HCO_3^- activity (Moseley et al., 1999; Ko et al., 2002; Chernova et al., 2003; Dorwart et al., 2008), and (iii) the clinical picture of CLD is invariable independently of genotype. Even up to the 8.6 kb deletion, reported in this study, results in the typical clinical course of CLD despite the fact that it eliminates the promoter region responsible for the basal level *SLC26A3* transcription (Alrefai et al., 2007). These data make assessment of mutations with unknown functional significance relatively conclusive if only the clinical picture is typical for CLD, a highly conserved amino acid is involved, mutations are confirmed on a second PCR product, and parental DNA samples prove the localization of compound heterozygous mutations on separate alleles. Although evidence of *de novo* mutations remain non-existent and we have

found heterozygous changes in all parental samples studied, the spectrum of mutations suggests that novel *SLC26A3* mutations are relatively common.

Future Prospects

Despite the advantages of genetic testing, large-scale mutation analysis is performed only as a part of research programs so far. Measuring high fecal chloride (>90 mmol/L) in a patient with a typical clinical picture and corrected electrolyte and acid-base balance is a reliable sign of CLD, independently of mutation analysis. As measurement of fecal chloride is fast and unexpensive, it remains to be the first line test to confirm/exclude the diagnosis of CLD. It is worth remembering, however, that excessive volume and salt depletion reduces the amount of diarrhea and may result in a low fecal Cl^- of even 40 mmol/L (Holmberg, 1978). Therefore, many patients are likely to die before the diagnosis and treatment, and in those suspected cases of CLD, mutation data may be essential to confirm the unsure diagnosis.

Establishment of local sequencing possibilities for the countries with a higher-than-average incidence of CLD and with known founder mutations would be easy and reliable in the first line diagnostics of CLD. In the future, genetic testing will most probably become easier and faster, allowing genetic diagnostics for CLD and other rare diseases locally. Outside the known founder populations for CLD, a future prospect is that the spectrum of found *SLC26A3* mutations will expand, and hopefully, many of the patients with CLD will get their diagnosis based on both early clinical characterization and mutation analysis. Only better diagnostics could reveal the figures of incidence of CLD which, most probably, are notably higher than expected based on those around 250 cases reported so far. Finally, only the analysis of different genotypes in relation to phenotypes could provide data on genotypic-phenotypic differences—if any—in CLD.

Figure legends

Figure 1. A) Different types of mutations within the *SLC26A3* gene. The exons are shown in scale but introns out of scale. The shaded areas in the first and last exons represent non-coding sequences.

B) Structure of the SLC26A3 protein modified according to the Uniprot prediction by using the model with 12 transmembranic domains (Bairoch et al., 2005). SLC26A3 involves 764 amino acids with 12 (to 14) transmembranic domains, and cytoplasmic N and C terminal ends. The arrows point to the approximate positions of different coding-area mutations from which large deletions and non-coding mutations are excluded. Numbers in the figure denote the amino acid positions. Two codons involve 2 (codon 454) or 3 (codon 131) different mutations at the same locus. The cytoplamic STAS domain involves amino acids 525 to 720 which are shown with the dashed line (Aravind and Koonin, 2000; Bairoch et al., 2005).

Deleted: ; Aravind and Koonin, 2000

References

Alrefai WA, Wen X, Jiang W, Katz JP, Steinbrecher KA, Cohen MB, Williams IR, Dudeja PK, Wu GD. 2007. Molecular cloning and promoter analysis of downregulated in adenoma (DRA). Am J Physiol Gastrointest Liver Physiol 293:G923-34.

Antalis TM, Reeder JA, Gotley DC, Byeon MK, Walsh MD, Henderson KW, Papas TS, Schweinfest CW. 1998. Down-regulation of the down-regulated in adenoma (DRA) gene correlates with colon tumor progression. Clin Cancer Res 4:1857-1863.

Aravind L, Koonin EV. 2000. The STAS domain - a link between anion transporters and antisigma-factor antagonists. Curr Biol 10:R53-5.

Arnheim N, Calabrese P. 2009. Understanding what determines the frequency and pattern of human germline mutations. Nat Rev Genet 10:478-488.

Asano K, Matsushita T, Umeno J, Hosono N, Takahashi A, Kawaguchi T, Matsumoto T, Matsui T, Kakuta Y, Kinouchi Y, Shimosegawa T, Hosokawa M, Arimura Y, Shinomura Y, Kiyohara Y, Tsunoda T, Kamatani N, Iida M, Nakamura Y, Kubo M. 2009. A genome-wide association study identifies three new susceptibility loci for ulcerative colitis in the japanese population. Nat Genet 41:1325-1329.

Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh LS. 2005. The universal protein resource (UniProt). Nucleic Acids Res 33:D154-9.

Booth IW, Stange G, Murer H, Fenton TR, Milla PJ. 1985. Defective jejunal brush-border Na+/H+ exchange: A cause of congenital secretory diarrhoea. Lancet 1:1066-1069.

Byeon MK, Westerman MA, Maroulakou IG, Henderson KW, Suster S, Zhang XK, Papas TS, Vesely J, Willingham MC, Green JE, Schweinfest CW. 1996. The down-regulated in adenoma (DRA) gene encodes an intestine-specific membrane glycoprotein. Oncogene 12:387-396.

Canani RB, Terrin G, Cirillo P, Castaldo G, Salvatore F, Cardillo G, Coruzzo A, Troncone R. 2004. Butyrate as an effective treatment of congenital chloride diarrhea. Gastroenterology 127:630-634.

Chernova MN, Jiang L, Shmukler BE, Schweinfest CW, Blanco P, Freedman SD, Stewart AK, Alper SL. 2003. Acute regulation of the SLC26A3 congenital chloride diarrhoea anion exchanger (DRA) expressed in xenopus oocytes. J Physiol 549:3-19.

Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloglu A, Ozen S, Sanjad S, Nelson-Williams C, Farhi A, Mane S, Lifton RP. 2009. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proc Natl Acad Sci U S A 106:19096-19101.

Darrow DC. 1945. Congenital alkalosis with diarrhea. J Pediatr 26:519-532.

den Dunnen JT, Antonarakis SE. 2000. Mutation nomenclature extensions and suggestions to describe complex mutations: A discussion. Hum Mutat 15:7-12.

den Dunnen JT, Paalman MH. 2003. Standardizing mutation nomenclature: Why bother? Hum Mutat 22:181-182.

Dorwart MR, Shcheynikov N, Baker JM, Forman-Kay JD, Muallem S, Thomas PJ. 2008. Congenital chloride-losing diarrhea causing mutations in the STAS domain result in misfolding and mistrafficking of SLC26A3. J Biol Chem 283:8711-8722.

Everett LA, Glaser B, Beck JC, Idol JR, Buchs A, Heyman M, Adawi F, Hazani E, Nassir E, Baxevanis AD, Sheffield VC, Green ED. 1997. Pendred syndrome is caused by mutations in a putative sulphate transporter gene (PDS). Nat Genet 17:411-422.

Field M. 2003. Intestinal ion transport and the pathophysiology of diarrhea. J Clin Invest 111:931-943.

Gamble JL, Fahey KR, Appleton J, MacLachlan E. 1945. Congenital alkalosis with diarrhea. J Pediatr 26:509-518.

Greeley T, Shumaker H, Wang Z, Schweinfest CW, Soleimani M. 2001. Downregulated in adenoma and putative anion transporter are regulated by CFTR in cultured pancreatic duct cells. Am J Physiol Gastrointest Liver Physiol 281:G1301-8.

Greger R. 2000. Role of CFTR in the colon. Annu Rev Physiol 62:467-491.

Haila S, Hoglund P, Scherer SW, Lee JR, Kristo P, Coyle B, Trembath R, Holmberg C, de la Chapelle A, Kere J. 1998. Genomic structure of the human congenital chloride diarrhea (CLD) gene. Gene 214:87-93.

Haila S, Saarialho-Kere U, Karjalainen-Lindsberg ML, Lohi H, Airola K, Holmberg C, Hastbacka J, Kere J, Hoglund P. 2000. The congenital chloride diarrhea gene is expressed in seminal vesicle, sweat gland, inflammatory colon epithelium, and in some dysplastic colon cells. Histochem Cell Biol 113:279-286.

Hastbacka J, de la Chapelle A, Mahtani MM, Clines G, Reeve-Daly MP, Daly M, Hamilton BA, Kusumi K, Trivedi B, Weaver A. 1994. The diastrophic dysplasia gene encodes a novel sulfate transporter: Positional cloning by fine-structure linkage disequilibrium mapping. Cell 78:1073-1087.

Hastbacka J, Superti-Furga A, Wilcox WR, Rimoin DL, Cohn DH, Lander ES. 1996. Atelosteogenesis type II is caused by mutations in the diastrophic dysplasia sulfate-transporter gene (DTDST): Evidence for a phenotypic series involving three chondrodysplasias. Am J Hum Genet 58:255-262.

Heinz-Erian P, Oberauer M, Neu N, Muller T, Scholl-Burgi S, Hager J, Pittschieler K, Janecke AR. 2008. A novel homozygous SLC26A3 nonsense mutation in a tyrolean girl with congenital chloride diarrhea. J Pediatr Gastroenterol Nutr 47:363-366.

Hihnala S, Hoglund P, Lammi L, Kokkonen J, Ormala T, Holmberg C. 2006a. Long-term clinical outcome in patients with congenital chloride diarrhea. J Pediatr Gastroenterol Nutr 42:369-375.

Hihnala S, Kujala M, Toppari J, Kere J, Holmberg C, Hoglund P. 2006b. Expression of SLC26A3, CFTR and NHE3 in the human male reproductive tract: Role in male subfertility caused by congenital chloride diarrhoea. Mol Hum Reprod 12:107-111.

Hoglund P, Auranen M, Socha J, Popinska K, Nazer H, Rajaram U, Al Sanie A, Al-Ghanim M, Holmberg C, de la Chapelle A, Kere J. 1998. Genetic background of congenital chloride diarrhea in high-incidence populations: Finland, poland, and saudi arabia and kuwait. Am J Hum Genet 63:760-768.

Hoglund P, Haila S, Socha J, Tomaszewski L, Saarialho-Kere U, Karjalainen-Lindsberg ML, Airola K, Holmberg C, de la Chapelle A, Kere J. 1996. Mutations of the down-regulated in adenoma (DRA) gene cause congenital chloride diarrhoea. Nat Genet 14:316-319.

Hoglund P, Hihnala S, Kujala M, Tiitinen A, Dunkel L, Holmberg C. 2006. Disruption of the SLC26A3-mediated anion transport is associated with male subfertility. Fertil Steril 85:232-235.

Hoglund P, Holmberg C, Sherman P, Kere J. 2001. Distinct outcomes of chloride diarrhoea in two siblings with identical genetic background of the disease: Implications for early diagnosis and treatment. Gut 48:724-727.

Holmberg C. 1978. Electrolyte economy and its hormonal regulation in congenital chloride diarrhea. Pediatr Res 12:82-86.

Holmberg C. 1986. Congenital chloride diarrhoea. Clin Gastroenterol 15:583-602.

Holmberg C, Perheentupa J. 1985. Congenital na+ diarrhea: A new type of secretory diarrhea. J Pediatr 106:56-61.

Holmberg C, Perheentupa J, Launiala K, Hallman N. 1977. Congenital chloride diarrhoea. clinical analysis of 21 finnish patients. Arch Dis Child 52:255-267.

Jacob P, Rossmann H, Lamprecht G, Kretz A, Neff C, Lin-Wu E, Gregor M, Groneberg DA, Kere J, Seidler U. 2002. Down-regulated in adenoma mediates apical cl-/HCO3- exchange in rabbit, rat, and human duodenum. Gastroenterology 122:709-724.

Kere J, Sistonen P, Holmberg C, de la Chapelle A. 1993. The gene for congenital chloride diarrhea maps close to but is distinct from the gene for cystic fibrosis transmembrane conductance regulator. Proc Natl Acad Sci U S A 90:10686-10689.

Ko SB, Shcheynikov N, Choi JY, Luo X, Ishibashi K, Thomas PJ, Kim JY, Kim KH, Lee MG, Naruse S, Muallem S. 2002. A molecular mechanism for aberrant CFTR-dependent HCO(3)(-) transport in cystic fibrosis. EMBO J 21:5662-5672.

Ko SB, Zeng W, Dorwart MR, Luo X, Kim KH, Millen L, Goto H, Naruse S, Soyombo A, Thomas PJ, Muallem S. 2004. Gating of CFTR by the STAS domain of SLC26 transporters. Nat Cell Biol 6:343-350.

Lamprecht G, Seidler U. 2006. The emerging role of PDZ adapter proteins for regulation of intestinal ion transport. Am J Physiol Gastrointest Liver Physiol 291:G766-77.

Makela S, Kere J, Holmberg C, Hoglund P. 2002. SLC26A3 mutations in congenital chloride diarrhea. Hum Mutat 20:425-438.

Martin MG, Turk E, Lostao MP, Kerner C, Wright EM. 1996. Defects in Na+/glucose cotransporter (SGLT1) trafficking and function cause glucose-galactose malabsorption. Nat Genet 12:216-220.

Melvin JE, Park K, Richardson L, Schultheis PJ, Shull GE. 1999. Mouse down-regulated in adenoma (DRA) is an intestinal cl(-)/HCO(3)(-) exchanger and is up-regulated in colon of mice lacking the NHE3 na(+)/H(+) exchanger. J Biol Chem 274:22855-22861.

Moseley RH, Hoglund P, Wu GD, Silberg DG, Haila S, de la Chapelle A, Holmberg C, Kere J. 1999. Downregulated in adenoma gene encodes a chloride transporter defective in congenital chloride diarrhea. Am J Physiol 276:G185-92.

Mount DB, Romero MF. 2004. The SLC26 gene family of multifunctional anion exchangers. Pflugers Arch 447:710-721.

Norio R, Perheentupa J, Launiala K, Hallman N. 1971. Congenital chloride diarrhea, an autosomal recessive disease. genetic study of 14 finnish and 12 other families. Clin Genet 2:182-192.

Ouwendijk J, Moolenaar CE, Peters WJ, Hollenberg CP, Ginsel LA, Fransen JA, Naim HY. 1996. Congenital sucrase-isomaltase deficiency. identification of a glutamine to proline substitution that leads to a transport block of sucrase-isomaltase in a pre-golgi compartment. J Clin Invest 97:633-641.

Rossmann H, Jacob P, Baisch S, Hassoun R, Meier J, Natour D, Yahya K, Yun C, Biber J, Lackner KJ, Fiehn W, Gregor M, Seidler U, Lamprecht G. 2005. The CFTR associated protein CAP70 interacts with the apical cl-/HCO3- exchanger DRA in rabbit small intestinal mucosa. Biochemistry 44:4477-4487.

Schweinfest CW, Henderson KW, Suster S, Kondoh N, Papas TS. 1993. Identification of a colon mucosa gene that is down-regulated in colon adenomas and adenocarcinomas. Proc Natl Acad Sci U S A 90:4166-4170.

Schweinfest CW, Spyropoulos DD, Henderson KW, Kim JH, Chapman JM, Barone S, Worrell RT, Wang Z, Soleimani M. 2006. slc26a3 (dra)-deficient mice display chloride-losing diarrhea, enhanced colonic proliferation, and distinct up-regulation of ion transporters in the colon. J Biol Chem 281:37962-37971.

Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. 2001. dbSNP: The NCBI database of genetic variation. Nucleic Acids Res 29:308-311.

Superti-Furga A, Hastbacka J, Wilcox WR, Cohn DH, van der Harten HJ, Rossi A, Blau N, Rimoin DL, Steinmann B, Lander ES, Gitzelmann R. 1996. Achondrogenesis type IB is caused by mutations in the diastrophic dysplasia sulphate transporter gene. Nat Genet 12:100-102.

Superti-Furga A, Neumann L, Riebel T, Eich G, Steinmann B, Spranger J, Kunze J. 1999. Recessively inherited multiple epiphyseal dysplasia with normal stature, club foot, and double layered patella caused by a DTDST mutation. J Med Genet 36:621-624.

Taguchi T, Testa JR, Papas TS, Schweinfest C. 1994. Localization of a candidate colon tumor-suppressor gene (DRA) to 7q22-q31.1 by fluorescence in situ hybridization. Genomics 20:146-147.

Wedenoja S, Hoglund P, Holmberg C. 2010. Review article: The clinical management of congenital chloride diarrhoea. Aliment Pharmacol Ther 31:477-485.

Wedenoja S, Holmberg C, Hoglund P. 2008a. Oral butyrate in treatment of congenital chloride diarrhea. Am J Gastroenterol 103:252-254.

Wedenoja S, Ormala T, Berg UB, Halling SF, Jalanko H, Karikoski R, Kere J, Holmberg C, Hoglund P. 2008b. The impact of sodium chloride and volume depletion in the chronic kidney disease of congenital chloride diarrhea. Kidney Int 74:1085-1093.

Wheat VJ, Shumaker H, Burnham C, Shull GE, Yankaskas JR, Soleimani M. 2000. CFTR induces the expression of DRA along with cl(-)/HCO(3)(-) exchange activity in tracheal epithelial cells. Am J Physiol Cell Physiol 279:C62-71.

Yang H, Jiang W, Furth EE, Wen X, Katz JP, Sellon RK, Silberg DG, Antalis TM, Schweinfest CW, Wu GD. 1998. Intestinal inflammation reduces expression of DRA, a transporter responsible for congenital chloride diarrhea. Am J Physiol 275:G1445-53.

Wildeman M, van Ophuizen E, den Dunnen JT, Taschner PE. 2008. Improving sequence variant descriptions in mutation databases and literature using the mutalyzer sequence variation nomenclature checker. Hum Mutat 29:6-13.

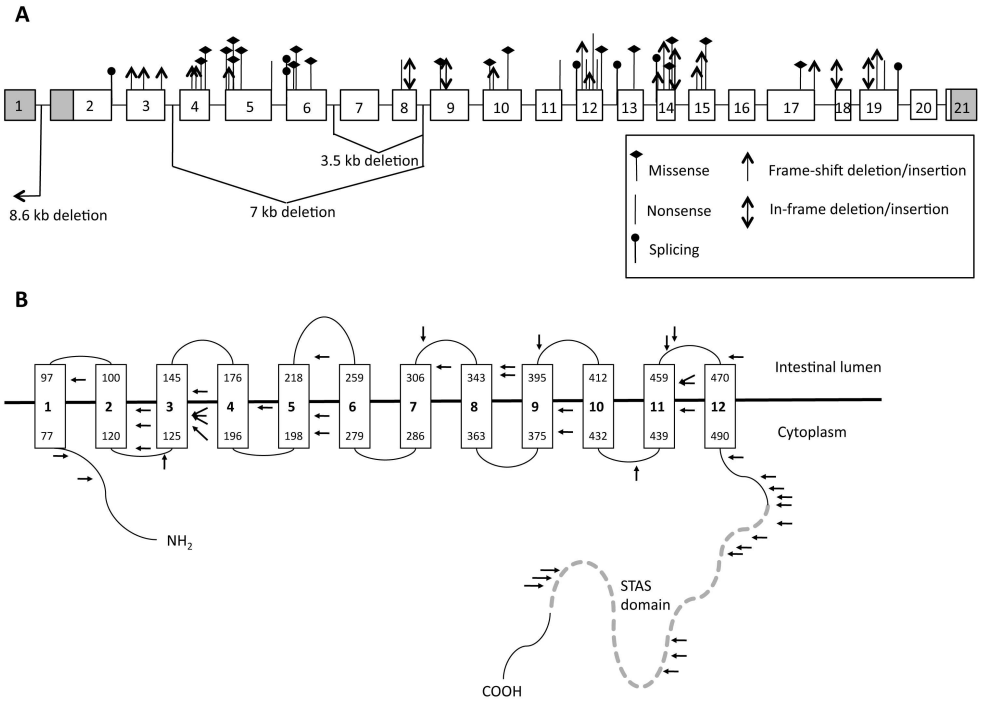

Figure 1. A) Different types of mutations within the SLC26A3 gene. The exons are shown in scale but introns out of scale. The shaded areas in the first and last exons represent non-coding sequences.

B) Structure of the SLC26A3 protein modified according to the Uniprot prediction by using the model with 12 transmembranic domains (Bairoch et al., 2005). SLC26A3 involves 764 amino acids with 12 (to 14) transmembranic domains, and cytoplasmic N and C terminal ends. The arrows point to the approximate positions of different coding-area mutations from which large deletions and non-coding mutations are excluded. Numbers in the figure denote the amino acid positions. Two codons involve 2 (codon 454) or 3 (codon 131) different mutations at the same locus. The cytoplamic STAS domain involves amino acids 525 to 720 which are shown with the dashed line (Aravind and Koonin, 2000; Bairoch et al., 2005).

260x190mm (300 x 300 DPI)

Table 1. *SLC26A3* mutations and coding-region single-nucleotide polymorphisms (SNPs).

Exon/Intron	Mutation Type	DNA Variant	Predicted/Demonstrated Effect on Protein	Patients (n)	Country of Origin	Reference
Promoter to Intron 1	Deletion	8.6 kb deletion	Exon 1 deletion	1	Sweden	This study
Intron 2	Insertion/splice-site	c.131+2_131+3insU14569.1*	Intron donor site change	1	Qatar	This study
Exon 3	Deletion	c.145_157del	p.K49LfsX8	2	Belgium	1
Exon 3	Insertion	c.177dup	p.I60HfsX11	1	USA	1
Exon 3	Insertion	c.269_270insAA	p.G91KfsX3	3	Hong Kong, China	1
Intron 3 to 8	Deletion	7 kb deletion	Exon 4 to 8 deletion	1	Austria	This study
Exon 4	Deletion	c.332del	p.F111SfsX4	1	Sweden	This study
Exon 4	Deletion	c.344del	p.I115TfsX19	2	Poland	1
Exon 4	Missense	c.358G>A	p.G120S	6	Poland, Sweden, Norway, China	1
Exon 4	Missense	c.371A>T	p.H124L	5	Poland, Sweden, France	1
Exon 5	Missense	c.386C>T	p.P129L	1	Austria	This study
Exon 5	Deletion	c.392del	p.P131RfsX3	1	Austria	This study
Exon 5	Missense	c.392C>G	p.P131R	3	USA	1
Exon 5	Missense	c.392C>T	p.P131L	1	USA	This study
Exon 5	Missense	c.408G>A	p.M136I	1	UK	This study
Exon 5	**Nonsense**	**c.559G>T**	**p.G187X**	**25**	**Saudi-Arabia, Kuwait, UK**	**1**
Intron 5	Splice-site change	c.571-2A>G	Intron acceptor site AG loss	1	Canada	1
Intron 5	Splice-site change	c.571-1G>T	Intron acceptor site AG loss	1	USA	1
Exon 6	Missense	c.610T>G	p.Y204D	1	Spain	This study
Exon 6	Missense	c.616T>C	p.S206P	1	Netherlands	1
Exon 6	Missense	c.659A>C	p.H220P	1	Spain	This study
Intron 6 to 8	Deletion	3.5 kb deletion	Exon 7 to 8 deletion	2	Japan	1
Exon 8	Nonsense	c.915C>A	p.Y305X	2	Poland	1
Exon 8	**Deletion**	**c.951_953del**	**p.V318del**	**45**	**Finland, Sweden**	**1**
Exon 9	Missense	c.1028G>A	p.C343Y	1	France	This study
Exon 9	Deletion/insertion	c.1030_1047delinsGATGCC	p.F344_V349delinsDA	1	Poland	This study
Exon 10	Missense	c.1136G>C	p.G379A	1	UK	This study
Exon 10	Deletion	c.1148_1149del	p.I383SfsX74	1	Spain	This study
Exon 10	Missense	c.1193C>T	p.S398F	1	Austria	This study
Exon 11	Nonsense	c.1306C>T	p.Q436X	1	Netherlands	1
Intron 11	Splice-site change	c.1312-1G>A	Intron acceptor site AG loss	1	Poland	1
Exon 12	Deletion	c.1342_1343del	p.L448KfsX9	1	Japan	1
Exon 12	Nonsense	c.1360C>T	p.Q454X	1	Canada	This study
Exon 12	Deletion	c.1362del	p.Q454HfsX5	3	Belgium, Turkey	2, this study
Exon 12	Nonsense	c.1386G>A	p.W462X	1	UK	1
Exon 12	Nonsense	c.1387C>T	p.R463X	1	France	This study
Exon 12	Missense	c.1403A>T	p.D468V	1	Poland	1
Intron 12	Splice-site change	c.1408-1G>A	Intron acceptor site AG loss	2	Germany	1
Exon 13	Missense	c.1487T>G	p.L496R	3	Hong Kong	1

Intron 13	Splice-site change	c.1515-2del	Intron acceptor site AG loss	1	Kuwait	1
Exon 14	Deletion	c.1517del	p.P506QfsX30	1	Poland	1
Exon 14	Deletion	c.1551_1554del	p.N518SfsX17	1	Poland	1
Exon 14	Missense	c.1559A>G	p.Y520C	1	Turkey	2
Exon 14	Missense	c.1563G>C	p.K521N	2	Sweden	This study
Exon 14	Deletion	c.1579_1581del	p.Y527del	1	Poland	1
Exon 15	Deletion	c.1609del	p.I537FfsX39	1	Canada	1
Exon 15	Deletion/insertion	c.1624_1626delinsC	p.S542PfsX11	1	Turkey	This study
Exon 15	Missense	c.1631T>A	p.I544N	2	Vietnam	1
Exon 17	Missense	c.1954G>A	p.D652N	4	Germany, Turkey	2, this study
Exon 17	Deletion	c.1990del	p.V664X	1	Sweden	This study
Exon 18	**Insertion**	**c.2024_2026dup**	**p.I675dup**	**15**	**Poland, Turkey**	**1**
Exon 19	Deletion/insertion	c.2104_2105delins29	p.G702Tins9	2	Norway	1
Exon 19	Deletion	c.2116del	p.S706AfsX6	1	Finland	1
Exon 19	Nonsense	c.2132T>G	p.L711X	1	Italy	3
Intron 19	Splice-site change	c.2205+3A>G	Intron donor site change	1	Sweden	This study
Coding SNPs					rs number	
Exon 2		c.119A>G	p.K40R		rs78133952	4
Exon 3		c.203G>A	p.R68Q		rs10280704	4
Exon 4		c.332T>C	p.F111S		rs75733585	4
Exon 4		c.357C>T	No change		rs73419912	4
Exon 5		c.514G>A	p.E172K		rs71566741	4
Exon 8		c.921T>G	p.C307W		rs34407351	1,4
Exon 8		c.923A>G	p.D308G		rs80222394	4
Exon 8		c.950T>G	p.V317G		rs78983942	4
Exon 9		c.996C>T	No change		rs35576676	4
Exon 11		c.1299G>A	No change		rs3735605	1,4
Exon 12		c.1314C>T	No change		-	1
Exon 14		c.1529C>T	p.T510M		rs60147601	4
Exon 15		c.1661G>A	p.R554Q		rs2301635	1,4
Exon 17		c.1802T>C	p.I601T		rs35776303	4
Exon 17		c.1953T>C	No change		rs41669	4
Exon 18		c.2038dup	p.D680GfsX8		rs35617203	4
Exon 20		c.2258A>G	p.N753S		rs35342296	4

Founder mutations are shown in bold. References: 1) Makela et al., 2002. 2) Choi et al., 2009. 3) Heinz-Erian et al., 2008. 4) http://www.ncbi.nlm.nih.gov/SNP/ (Sherry et al., 2001). *U14569.1 indicates the GenBank accession of human Alu-Sb1 (Alu Y) subfamily consensus sequence (HSU14569). Mutation data are based on the RefSeq sequences NG_008046.1 (genomic) and NM_000111.2 (cDNA).