 doi:10.20944/preprints202009.0543.v1

Case Report

Identification and characterization of a novel *CLCN7* variant associated with osteopetrosis

Dmitrii S. Bug[1], Ildar M. Barkhatov[2], Yana V. Gudozhnikova[3], Artem V. Tishkov[1], Igor B. Zhulin[1,4*] and Natalia V. Petukhova[1*]

[1] Bioinformatics Research Center, Pavlov First Saint Petersburg Medical State University, St. Petersburg, Russian Federation

[2] R.M. Gorbacheva Scientific Research Institute of Pediatric Hematology and Transplantation, Pavlov First Saint Petersburg Medical State University, St. Petersburg, Russian Federation

[3] Pavlov First Saint Petersburg State Medical University Clinic, St. Petersburg, Russian Federation

[4] Department of Microbiology and Translational Data Analytics Institute, The Ohio State University, Columbus OH, USA

* Correspondence: jouline.1@osu.edu, petuhovanv@1spbgmu.ru

Received: date; Accepted: date; Published: date

Abstract: Osteopetrosis is a group of rare inheritable disorders of the skeleton characterized by increased bone density. The disease is remarkably heterogeneous in clinical presentation and often misdiagnosed. Therefore, genetic testing and molecular pathogenicity analysis are essential for precise diagnosis and new targets for preventive pharmacotherapy. Mutations in the *CLCN7* gene give rise to the complete spectrum of osteopetrosis phenotypes and are responsible for about 75% of cases of autosomal dominant osteopetrosis. In this study, we report the identification of a novel variant in the *CLCN7* gene in a patient diagnosed with osteopetrosis and provide evidence for its significance (likely deleterious) based on extensive comparative genomics, protein sequence and structure analysis. A set of automated bioinformatics tools used to predict consequences of this variant identified it as deleterious or pathogenic. Structure analysis revealed that the variant is located at the same "hot spot" as the most common *CLCN7* mutations causing osteopetrosis. Deep phylogenetic reconstruction showed that not only Leu614Arg, but any non-aliphatic substitutions in this position are evolutionarily intolerant, further supporting the deleterious nature of the variant. The present study provides further evidence that reconstructing a precise evolutionary history of a gene helps predicting phenotypical consequences of variants of uncertain significance.

Keywords: genetics, comparative genomics, phylogenetic analysis, osteopetrosis, *CLCN7* gene

1. Introduction

Osteopetrosis ("marble bone disease") is a group of rare inheritable disorders of the skeleton characterized by increased bone density. The estimated prevalence of osteopetrosis is 1 in 100,000 to 500,000 [1, 2]. It exists in two major clinical forms - a benign autosomal dominant form (ADO) and a malignant autosomal recessive form (ARO) with incidents 1 in 20,000 and 1 in 250,000 births, respectively. The autosomal dominant adult (benign) form is associated with few, if any, symptoms, and the autosomal recessive infantile (malignant) form is typically fatal during infancy or early childhood if untreated [1]. General skeletal abnormality can be characterized by increased bone density, diffuse and focal sclerosis, pathological fractures. Osteopetrosis comprises a clinically and genetically heterogeneous group of conditions that share the hallmark of increased bone density on

radiographs. The increase in bone density results from abnormalities in osteoclast differentiation or function. The Nosology Group of the International Skeletal Dysplasia Society classifies increased bone density conditions into several distinct entities based on clinical features, mode of inheritance and underlying molecular and pathogenetic mechanisms [3]. Treatment of osteopetrosis is mostly symptomatic, with some exceptional cases of transplanting hematopoietic stem cells. However, the disease is remarkably heterogeneous in clinical presentation and often misdiagnosed. Therefore, genetic testing and molecular pathogenicity analysis are essential for precise diagnosis and new targets for preventive pharmacotherapy [4].

Osteopetrosis is caused by failure of osteoclast differentiation or function and mutations in at least 10 genes have been identified as causative in humans and collectively account for approximately 80% of patients in the cohorts [4,5]. The genetic basis of this disease is now largely uncovered: mutations in *TCIRG1*, *CLCN7*, *OSTM1*, *SNX10* and *PLEKHM1* lead to osteoclast-rich ARO (in which osteoclasts are abundant but have severely impaired resorptive function), whereas mutations in *TNFSF11* and *TNFRSF11A* lead to osteoclast-poor ARO. In osteoclast-rich ARO, impaired endosomal and lysosomal vesicle trafficking results in defective osteoclast ruffled-border formation and, hence, the inability to resorb bone and mineralized cartilage [2]. Mutations in *TCIRG1* and *CLCN7* together account for nearly 70% of all patients with ARO. Mutations in the *CLCN7* gene are responsible for about 75 % of cases of ADO, 10-15 % of cases of autosomal recessive osteopetrosis, and all known cases of intermediate autosomal osteopetrosis. Mutations in the *CLCN7* gene affect the function of osteoclast-mediated extracellular acidification, resulting in the disturbed dissolution of the bone inorganic matrix and a series of clinical features [6].

The Chloride channel 7 (*CLCN7*) gene is a member of the mammalian CLC gene family. In osteoclasts, the CLCN7 protein resides in the late endocytotic–lysosomal pathway of the ruffled membrane borders and is involved in the acidification of the resorption lacunae [7]. The physiological function of the CLC7 protein was unclear until it was shown that the disruption of the *CLCN7* gene in mice causes severe osteopetrosis, and that the CLC7 protein played an essential role in the acidification of the extracellular resorption lacunae, which is very important for osteoclast-mediated resorption of mineralized bone. In humans, mutations in the *CLCN7* gene give rise to the complete spectrum of osteopetrosis phenotypes [8].

2. Materials and Methods

2.1 Case Report

A 14-yr-old male patient was diagnosed with osteopetrosis in the Pavlov First St. Petersburg Medical State University Clinic. The patient's legal guardian provided written informed consent for genomic sequencing and research. The study was conducted in accordance with the Declaration of Helsinki, all aspects were reviewed and approved by the Ethics committee at the Pavlov First St. Petersburg State Medical University, Russian Federation (protocol 35-2020). The patient's legal guardian provided written informed consent for publication of this case report and any accompanying images.

2.2 DNA sequencing

Patient's whole exome sequencing (WES) was performed on Illumina NextSeq500 by iBinom (Moscow, Russia). Human genome 19 (hg19) build 37 was used as a reference sequence. The following genes associated with osteopetrosis were screened for variants: *TCIRG1*, *CLCN7*, *OSTM1*, *RANKL*, *RANK*, *IKBKG*, *SNX10*, *TNFSFR11A*, *TNFSF11*, *PLEKHM1*, *CA2*. This approach identified a previously unknown heterozygous variant in *CLCN7* gene. No other variants were uncovered that might account for the disease phenotype. Variant was submitted to ClinVar (accession: SCV001190007). Raw sequencing data were deposited in BioProject database (SRA accession: PRJNA613088). Sanger sequencing was performed to verify this variant. PCR products of the gene

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 September 2020 doi:10.20944/preprints202009.0543.v1

locus were obtained by protocol suggested for 20 exone [9] and subjected to sequencing by using BigDye®Terminator v3.1 cycle sequencing kit on an AB3500xl genetic analyzer (Applied Biosystems, U.S.A.).

2.3 Sequence acquisition

Reference sequences of *CLCN7* gene (NG_007567.1), coding nucleotides (NM_001287.6) and amino acids (NP_001278.1) were retrieved from NCBI database.

2.4 Bioinformatics analysis

CLC7 protein sequence domain architecture was analyzed using CDvist [10]. A 3D homology model of the CLCN7 protein was built by the Swiss-Model server [11] using the structure of a eukaryotic CLC protein [12] as a template (PDB accession: 3ORG). BLAST searches of the NCBI RefSeq database were carried out with default parameters using CLCN7 protein sequence (NP_001278.1) as a query. Multiple sequence alignments were constructed using MUSCLE [13] and edited in JalView [14] Neighbor-joining and maximum likelihood phylogenetic trees were built using MEGA 7 [15].

3. Results

3.1 Case representation

A 14-yr-old male patient was diagnosed with osteopetrosis (anamnestic) based on the following symptoms: multiple (more than ten) bone fractures, signs of osteosclerosis, bone marrow failure, hepatosplenomegaly, and congenital anomalies (Figure 1, Table 1, Supplemental Table 1). The patient had a 23-year-old sibling, who was also diagnosed with osteopetrosis based on clinical evaluation, patient history and X-ray imaging. Onset in late childhood/adolescence in both siblings suggested Type II ADO.

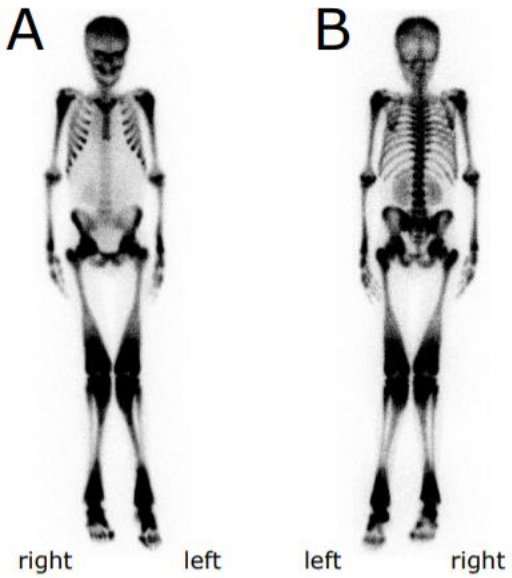

Figure 1. Bone scintigraphy of the proband. Anterior (**A**) and posterior (**B**) views. Metaphyses and diaphyses of humerus, femur, tibia and fibula are widened.

Autosomal Dominant Osteopetrosis Type 2	Proband	Sibling	Relevance/Alternate

 doi:10.20944/preprints202009.0543.v1

Clinical Features*			Explanation
Autosomal dominant inheritance	Yes	Unknown	
Facial nerve palsy	No	Unknown	
Vision loss, severe, beginning in childhood	No	Unknown	
Osteosclerosis, diffuse symmetrical	Yes	Unknown	
Increased long bone fracture rate (75% of patients)	Yes	Yes	
Multiple fractures	Yes	Yes	
Pronounced skull base sclerosis	Unknown	Unknown	
Mandibular osteomyelitis	No	Unknown	
'Rugger-Jersey' spine (vertebral endplate thickening)	Unknown	Unknown	
Endobones (bone within bone)	Unknown	Unknown	
Hip osteoarthritis	No	Unknown	
Facial palsy due to cranial nerve VII compression	No	Unknown	
Bone marrow failure	Yes	No	
Elevated serum acid phosphatase	Unknown	Unknown	
Onset in childhood	Yes	Yes	
Progressive sclerosis with age	Unknown	Unknown	
20-40% patients are asymptomatic	No	No	
Other Clinical Features			
Hepatomegaly	Yes	Unknown	Clinical feature of autosomal recessive osteopetrosis type 4 (OMIM #611490)
Splenomegaly	Yes	Unknown	Same as above
Anemia	Yes	Unknown	Same as above
Reticulocytosis	Yes	Unknown	Same as above
Thrombocytopenia	Yes	Unknown	Same as above
Failure to thrive	Yes	Unknown	Clinical feature of autosomal recessive osteopetrosis type 1 (OMIM #259700)
Hydrocephalus	Yes	Unknown	Same as above
Splayed metaphyses	Yes	Unknown	Same as above
Low serum calcium	Yes	Unknown	Same as above
Elevated alkaline phosphatase	Yes	Unknown	Same as above
Valgus deformity	Yes	Unknown	Clinical feature of autosomal recessive osteopetrosis type 2 (OMIM #259710)
Dental anomalies	Yes	Unknown	Same as above
Elevated serum lactate dehydrogenase	Yes	Unknown	Clinical feature of autosomal recessive osteopetrosis type 5

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 September 2020 doi:10.20944/preprints202009.0543.v1

			(OMIM #259720)
Lymphocytosis	Yes	Unknown	Assumed related
Skeletal effects	Yes	Unknown	Assumed related
Scoliosis	Yes	Unknown	Assumed related
Low hairline	Yes	Unknown	Assumed related
Double xiphoid process	Yes	Unknown	Assumed related

Table 1. Phenotypic features.

3.2 Mutation analysis

Screening of eleven genes associated with osteopetrosis (sequences obtained by low-coverage WES) identified a previously unknown heterozygous variant in CLCN7 gene (Table 2): NM_001287.6:c.1841T>G (NP_001278.1:p.Leu614Arg), consistent with the fact that mutations in the CLCN7 gene are responsible for about 75 % of cases of Type II ADO [7]. The variant was confirmed by Sanger sequencing (Figure 2A). The c.T1769G (p.Leu614Arg) is novel and has not been reported in the 1000 Genomes Project (2,504 samples, accessed 9/18/2019), dbSNP (accessed 03/20/2020), or Genome Aggregation Database (gnomAD v.2.1.1, 125,748 exomes, and gnomAD v3, 71,702 genomes, accessed 03-20-2020). A different variant in the same position, Leu614Pro, was reported in ClinVar database (rs1064794323, reported as "uncertain significance") in a child with severe osteopetrosis, anemia, blindness, neurological impairment and macrocephaly, who also had a deletion in exon 17 of the CLCN7 gene [10]. Another variant in the same position, Leu614Met, was reported in dbSNP database (rs1000353389).

3.3 Comparative protein sequence analysis and modelling

CLCN7 protein domain analysis using CDvist revealed 11 transmembrane helices, N- and C-terminal low complexity regions, and two C-terminal CBS-domains (Supplemental Figure 1). We built a 3D homology model of the CLCN7 protein using the Swiss-Model server and the structure of a eukaryotic CLC protein [12] as a template (PDB accession: 3ORG). Position Leu614 is located at the C-terminal low complexity region between the last transmembrane helix and the first CBS domain (Supplemental Fig. 1). Notably, this position is found at the same location in the protein tertiary structure as the most common mutations causing osteopetrosis - almost all of them are clustered at the intracellular gates of the CLC channel dimer, the CLCN7 mutational "hot spot" (Figure 2B).

Gene	Genomic location	DNA Reference	Protein Reference	Variant Type	Genotype	Origin	Observed Effect
CLCN7	Chr16: 1498724 (GRCh37)	NM_001287.6: c.1841T>G	NP_001278.1: p.Leu614Arg	missense	heterozygous	Unknown	Deleterious

Table 2. Genomic findings.

3.4 Bioinformatics analysis for variant significance confirmation

We used several bioinformatics tools, all of which predicted that the newly identified Leu614Arg variant is damaging (Supplemental Table 2). Still, strictly following ACMG guidelines for interpretation [16], the CLCN7 p.Leu617Arg variant should be classified as "variant of uncertain significance" based on two moderate and one supporting criteria. Therefore, in addition to automated bioinformatics tools, we used a recently developed evolutionary approach based on removing paralogous sequences from analysis [17]. We identified eight paralogs of CLCN7 in the human genome (Supplemental Figure 2). In order to exclude paralogs from further analysis, we

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 September 2020 doi:10.20944/preprints202009.0543.v1

searched for orthologs of nine human proteins (CLCN7 and its paralogs) among the representative Metazoan genomes (Supplemental Figure 3). Species taken in the analysis were evenly distributed over the Tree of Life, with several species per class and approximately one species per order. Then BLAST search for CLCN7 against the NCBI RefSeq database was performed against the representative genomes. From the resulting sequences, a neighbor-joining phylogenetic tree was constructed to reveal CLCN7 orthologs, and exclude all paralogs. Multiple sequence alignment of CLCN7 orthologs showed 100% conservation of Leu614 (Figure 2C).

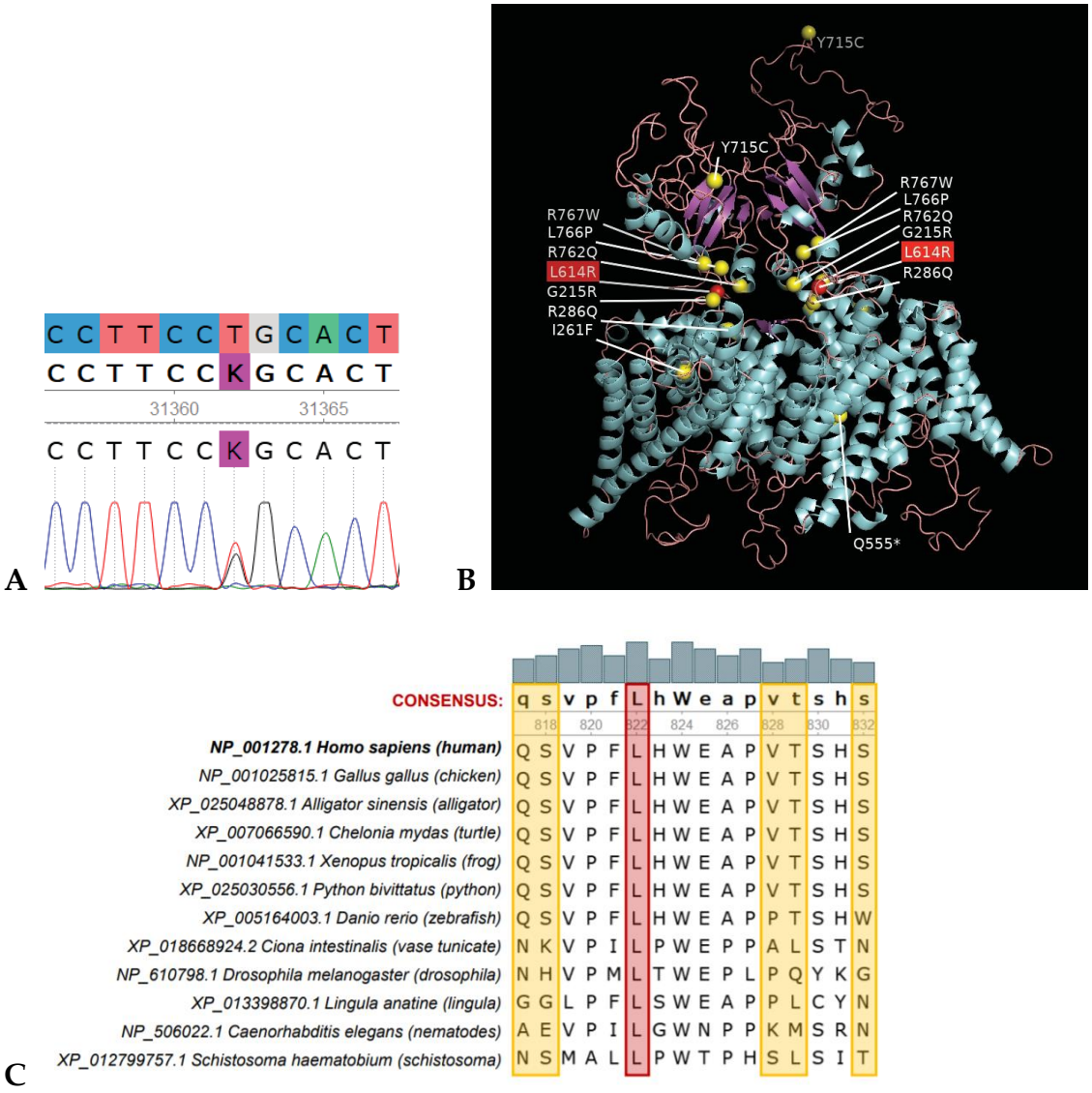

Figure 2. *Leu614Arg substitution occurred in a mutational "hot spot" and it is evolutionarily intolerable.* (**A**) Sanger sequencing showing the heterozygous c.1769T>G, p.Leu614Arg variant. (**B**) CLCN7 structural homology model by Swiss-Model (CmCLC, PDB accession 3ORG, was used as a template). The most common pathogenic variants causing osteopetrosis are shown as yellow spheres; the Leu614Arg variant is shown as a red sphere. (**C**) A fragment of multiple sequence alignment of CLCN7 orthologs from representative metazoan genomes. Position corresponding to Leu614 in the human CLCN7 protein is highlighted in red. Variable positions highlighted in yellow show that there was a significant time for divergence.

In addition, the human CLCN7 protein sequence was used as a query in a BLAST search against all Metazoan genomes. First 1000 resulting sequences were aligned and a neighbor-joining phylogenetic tree was constructed. Using orthologous/paralogous relationships defined with the set of representative genomes, we were able to identify a clade of CLCN7 orthologs, a clade with CLCN6 orthologs and mixed (CLCN6 and CLCN7) wedge-like clade containing only

 doi:10.20944/preprints202009.0543.v1

representatives of invertebrates (Supplemental Figure 4). In the resulting alignment, only four amino acid residues were found in a position corresponding to L614 in the human CLCN7: Leu = 97.0% (970), Met = 1.3% (13), Ile = 0.3% (3), Val = 0.3% (3). These results show that only aliphatic amino acids and functionally related to them methionine are allowed in this position even among close paralogs. A variant Leu614Met reported in dbSNP (rs1000353389) can therefore be considered evolutionarily permitted. In contrast, a substitution Leu614Arg is evolutionarily intolerable and thus should be classified as damaging.

4. Discussion

In this study, we report the identification of a novel variant in the *CLCN7* gene in a patient diagnosed with osteopetrosis and provide evidence for its significance. Although several independent lines of indirect evidence suggested that the newly detected CLCN7 p.Leu617Arg variant is damaging, according to American College of Medical Genetics (ACMG) guidelines for interpretation [16] it should be classified as "variant of uncertain significance". Just one supporting criterion was needed to re-classify it to "likely damaging"; however, because mutations in the *CLCN7* gene account for 75%, not 100% cases, this criterion was not met. On the other hand, we provide a strong argument for this being a likely pathogenic variant by revealing the precise evolutionary history of this gene and showing that all changes to non-aliphatic amino acids in position Leu617 were selected against during hundreds of millions of years. Not a single case of successful substitution to non-aliphatic amino acid in this position can be found among hundreds of CLCN7 orthologs. It is important to stress out that this is not a computational prediction, but a direct observation: mutation Leu617Arg is evolutionarily intolerable and therefore damaging. Taken together with the fact that this variant is found in a gene, mutations in which account for the vast majority of autosomal dominant osteopetrosis cases, and it is located in a mutational "hot spot" of this gene, our evolutionary analysis strongly suggests that this variant as likely pathogenic.

In conclusion, we identified a novel, damaging variant in the *CLCN7* gene. Genomic and protein structure analyses suggest that this variant is located in a mutational "hot spot" and it is evolutionarily intolerable.

Supplementary Material: Figure S1: Predicted domain architecture of the CLCN7 protein. Figure S2: Neighbor-joining phylogenetic tree of CLCN7 paralogs identified in the human genome. Figure S3: Neighbor-joining phylogenetic tree of CLCN7 orthologs from representative metazoan genomes. Figure S4: Separation of CLCN7 orthologs from the closely related paralogs. Table S1: Clinical features of the proband. Table S2: Functional effect of mutation L614R by different SNP predictors.

Author Contributions: Conceptualization, I.B.Z.; Methodology, I.B.Z.; Formal Analysis, A.V.T., N.V.P., D.S.B.; Investigation, A.V.T., N.V.P., D.S.B.; Resources, Y.V.G., I.M.B.; Data Curation, A.V.T., N.V.P., D.S.B.; Writing – Original Draft Preparation, N.V.P.; Writing – Review & Editing, D.S.B. and I.B.Z.; Visualization, D.S.B. and N.V.P.; Supervision, I.B.Z..; Project Administration, I.B.Z.

Funding: This work was supported by Russian Federation Ministry of Health Assignment 056-00016-18-04.

Acknowledgments: We are indebted to the patient and their family for participation in this study. We thank Anna Bilyk for assistance with Sanger sequencing.

Conflicts of Interest: The authors declare that they have no competing interests.

References

1. Stark, Z.; Savarirayan R. Osteopetrosis. *Orphanet. J. Rare. Dis.* **2009**, 4, 5.
2. Sobacchi, C.; Schulz, A.; Coxon, F.P.; Villa, A.; Helfrich, M.H. Osteopetrosis: genetics, treatment and new insights into osteoclast function. *Nat. Rev. Endocrinol.* **2013**, 9, 522–536.
3. Superti-Furga, A.; Unger, S. Nosology and classification of genetic skeletal disorders: 2006 revision. *Am. J. Med. Genet.* **2007**, 143A, 1–18.

4. Askmyr, M.K.; Fasth, A.; Richter, J. Towards a better understanding and new therapeutics of osteopetrosis. *Br. J. Haematol.* **2008**, 140, 597–609.
5. Villa, A.; Guerrini, M.M.; Cassani, B.; Pangrazio, A.; Sobacchi, C. Infantile Malignant, Autosomal Recessive Osteopetrosis: The Rich and The Poor. *Calcif. Tissue. Int.* **2008**, 84, 1–12.
6. Pangrazio, A.; Pusch, M.; Caldana, E.; Frattini, A.; Lanino, E.; Tamhankar, P.M.; Phadke, S.; Lopez, A.G.; Orchard, P.; Mihci, E.; Abinun, M.; Wright, M.; Vettenranta, K.; Bariae, I.; Melis, D.; Tezcan, I.; Baumann, C.; Locatelli, F.; Zecca, M.; Horwitz, E.; Mansour, L.S.; Van Roij, M.; Vezzoni, P.; Villa, A.; Sobacchi, C. Molecular and clinical heterogeneity in CLCN7-dependent osteopetrosis: report of 20 novel mutations. *Hum. Mutat.* **2010**, 31, E1071-E1080.
7. Jentsch, T.J.; Stein, V.; Weinreich, F.; Zdebik, A.A. Molecular structure and physiological function of chloride channels. *Physiol. Rev.* **2002**, 82, 503–568.
8. Kornak, U.; Kasper, D.; Bösl, M.R.; Kaiser, E.; Schweizer, M.; Schulz, A.; Friedrich, W.; Delling, G.; Jentsch, T.J. Loss of the CLC-7 chloride channel leads to osteopetrosis in mice and man. *Cell* **2001**, 104, 205–215.
9. Frattini, A.; Pangrazio, A.; Susani, L.; Sobacchi, C.; Mirolo, M.; Abinun, M.; Andolina, M.; Flanagan, A.; Horwitz, E.M.; Mihci, E.; Notarangelo, L.D.; Ramenghi, U.; Teti, A.; Van Hove, J.; Vujic, D.; Young, T.; Albertini, A.; Orchard, P.J.; Vezzoni, P.; Villa, A. Chloride channel CLCN7 mutations are responsible for severe recessive, dominant, and intermediate osteopetrosis. *J. Bone Miner. Res.* **2003**, 18, 1740–1747.
10. Adebali, O.; Ortega, D.R.; Zhulin, I.B. CDvist: a webserver for identification and visualization of conserved domains in protein sequences. *Bioinformatics* **2014**, 31, 1475–1477.
11. Johansson, M.U.; Zoete, V.; Michielin, O.; Guex, N. Defining and searching for structural motifs using DeepView/Swiss-PDBViewer. *BMC Bioinformatics* **2012**, 13, 173.
12. Feng, L.; Campbell, E.B.; Hsiung, Y.; MacKinnon, R. Structure of a eukaryotic CLC transporter defines an intermediate state in the transport cycle. *Science* **2010**, 330, 635-641.
13. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. **2004**, 32, 1792-1797.
14. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.A.; Clamp, M.; Barton, G.J. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009**, 25,1189-1191.
15. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, 33, 1870-1874.
16. Richards, S.; Aziz, N.; Bale, S.; Bick, D.; Das, S.; Gastier-Foster, J.; Grody, W.W.; Hegde, M.; Lyon, E.; Spector, E.; Voelkerding, K.; Rehm, H.L.; ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **2015**, 17, 405-423.
17. Adebali, O.; Reznik, A.O.; Ory, D.S.; Zhulin, I.B. Establishing the precise evolutionary history of a gene improves prediction of disease-causing missense mutations. *Genet. Med*. **2016**, 18, 1029-1036.