Whole exome sequencing of patients with diffuse idiopathic skeletal hyperostosis and calcium pyrophosphate crystal chondrocalcinosis

Parreira B[1,2], Couto AR[1,2], Rocha F[1,2], Sousa M[1,2], Faustino V[1], Power DM[3], Bruges-Armas J[1,2]

ACTA REUMATOL PORT. 2020;45:116-126

ABSTRACT

Objectives: DISH/CC is a poorly understood phenotype characterised by peripheral and axial enthesopathic calcifications, frequently fulfilling the radiological criteria for Diffuse Idiopathic Skeletal Hyperostosis (DISH, MIM 106400), and in some cases associated with Calcium Pyrophosphate Dihydrate (CPPD) Chondrocalcinosis (CC). The concurrence of DISH and CC suggests a shared pathogenic mechanism. In order to identify genetic variants for susceptibility we performed whole exome sequencing in four patients showing this phenotype.

Materials and methods: Exome data were filtered in order to find a variant or a group of variants that could be associated with the DISH/CC phenotype. Variants of interest were subsequently confirmed by Sanger sequencing. Selected variants were screened in a cohort of 65 DISH/CC patients *vs* 118 controls from Azores. The statistical analysis was performed using PLINK V1.07.

Results: We identified 21 genetic variants in 17 genes that were directly or indirectly related to mineralization, several are predicted to have a strong effect at a protein level. Phylogenetic analysis of altered amino acids indicates that these are either highly conserved in vertebrates or conserved in mammals. In case-control analyses, variant rs34473884 in *PPP2R2D* was significantly associated with the DISH/CC phenotype (p=0.028; OR=1.789, 95% CI= 1.060 - 3.021)).

Conclusion: The results of the present and preceding studies with the DISH/CC families suggests that the phenotype has a polygenic basis. The *PPP2R2D* gene could be involved in this phenotype in an as yet unknown way.

Keywords: Rheumatic and musculoskeletal diseases ; Genetic association; Rheumatology.

1. Serviço Especializado de Epidemiologia e Biologia Molecular (SEEBMO) / Hospital de Santo Espírito da Ilha Terceira (HSEIT)
2. Comprehensive Health Research Center, CHRC, Lisbon, Portugal
3. Centre of Marine Science (CCMAR) / University of Algarve

INTRODUCTION

Previous studies undertaken by our group, identified and characterized twelve families affected with Diffuse Idiopathic Skeletal Hyperostosis (DISH, MIM 106400) and/or Calcium Pyrophosphate Dihydrate (CPPD) Chondrocalcinosis (CC), hereafter designated, DISH/CC. DISH/CC is a poorly understood phenotype characterised by peripheral and axial enthesopathic calcifications, frequently fulfilling the radiological criteria for DISH, and in some cases associated with CPPD Chondrocalcinosis. A common pathogenic mechanism, shared by the two conditions, has been suggested[1]. DISH is a common skeletal disorder characterized by progressive calcification and ossification of ligaments and entheses[2-3]. The exact prevalence and incidence of DISH is unknown, however it is well known that it is more frequent in males and its prevalence rapidly increases with age, mainly affecting subjects over the age of 40[4]. The prevalence of DISH in patients over 50 years of age is 25% in males and 15% in females,[5] and this disease is now becoming a serious problem in aging societies. Several lines of evidence suggest that genetic factors might play a part in the etiology of DISH, such as the existence of familial cases with early onset (in the third decade of life)[6] and the higher frequency of DISH in the boxer dogs relative to other dog breeds[7-8]. So far, however, no single gene has been conclusively associated with the disease. Chondrocalcinosis is characterized by deposition of crystals of calcium pyrophosphate dihydrate (CPPD) in articular hyaline and fibro-carti-

lage[9]. For the moment two genes, *ANKH* (CCAL2) and *TNFRSF11B*, are known to be involved in the development of Chondrocalcinosis[10-11]. Previously, we performed genetic studies of DISH/CC using a "Whole genome wide linkage study" and an "Identity by State/Identity by Descent" association study and two genes were identified as good candidate genes for DISH/CC; *RSPO4* on chromosome 20, and *LEMD3* on chromosome 12. Several *RSPO4* gene variants were identified and nucleotide modifications located in the regulatory region were more frequent in control individuals than in the DISH/CC group[12]. Even though the genetic basis of DISH/CC is unknown, it is considered to be a bone forming disease and genes related to calcification and ossification are considered good candidates. In the present study, we took advantage of whole exome sequencing (WES) so that even in the absence of sufficient pedigrees and samples for traditional linkage approaches we could identify rare protein-coding variants that are the cause of the majority of monogenic diseases[13-14]. We performed targeted exome sequencing on four DISH/CC patients, with an apparently autosomal dominant DISH/CC phenotype, in order to capture rare and pathogenic variants expected to have potentially damaging effects on protein function that lead to modified calcification and/or ossification. To our knowledge this is the first report of WES analysis in patients affected with DISH.

MATERIALS AND METHODS

This study was approved by the "Comissão de Ética do Hospital de Santo Espírito da Ilha Terceira". All methods were performed in accordance with the approved guidelines, obtaining informed consent from each subject before conducting the experiments.

EXOME CAPTURE

The selection of patients was made after ruling out mutations in *ANKH* and secondary causes for Chondrocalcinosis. DNA was extracted from peripheral blood samples using a standard salting out procedure. Samples were resequenced, using an ABI-SOLiD platform and Agilent's SureSelect Target Enrichment System for 38 Mb, by "Sistemas Genómicos, S.L." in Valencia, Spain. After standard DNA quality control, SOLiD Fragment libraries were prepared and enriched with SureSelect All Human Enrichment Target Exon. The quality and quantity of the libraries were assessed by analysis with an Agilent 2100 Bioanalyzer and Qubit. Each library went through a process of emulsion PCR for clonal amplification of the fragments, followed by an enrichment process and chemical modification to allow loading into the reaction chamber. The quality and quantity of the beads obtained for each library were estimated taking into account the parameters given by Work Flow Analysis. Then, ligation sequencing was done to obtain sequences of 50nt +35nt (Paired-end) reads using a SOLiD4 sequencer. The data quality was estimated using the parameters provided by the software SETS (SOLiD Experimental Tracking System). Single Nucleotide Variants (SNVs) were classified using Ensembl's nomenclature and grouped using the following scheme: Known and Novel (Coding, Splicing, Others). The coding variants were divided into nonsynonymous and synonymous. The "Others" included intronic, UTR, regulatory region, intergenic, downstream and upstream variants.

WES FILTERING

The segregation analysis of these families indicated that the most likely model for this disease is an autosomal dominant model[1]. However, given that a previous linkage study did not identify a major dominant locus, the data were analyzed assuming both a recessive and a dominant model.

Genetic variants were identified in 815, 917, 872 and 593 genes under a dominant model, for individuals AZ1 to AZ4. Assuming that the genetic cause is the same for the individuals AZ1 to AZ4, identified genes were then filtered to select common, shared genes (supplementary Table I, available in supplementary file). A group of 52 genes were common to the four patients; candidate genes for testing were selected taking into account their involvement in calcification and/or ossification or related conditions that could be associated with the DISH/CC disease. Gene variants shared between the genes common to the four patients were identified and nonsynonymous, splice site, stop lost/gain and frameshift variants were targeted, in anticipation that synonymous and intronic variants would be far less likely to be relevant in the pathogenicity.

CANDIDATE GENES – SELECTION BY FUNCTION

A group of 20 candidate genes was selected and included genes that are reported to be involved in bone metabolism and/or related conditions. These genes were Alkaline Phosphatase (*ALPL/TNAP*), the Calcium

TABLE I. RADIOLOGY RESULTS FROM FOUR SELECTED PATIENTS FOR WES.

Patient	Sex	RADIOLOGY					Age at onset	Other diseases
		C-Spine	T-Spine	L-Spine	Knees	Elbow		
AZ1	M	Normal	Normal	Syndesmophytosis DISH	N/A	Periarticular calcifications	<40	Obesity
AZ2	F	DISH	DISH	DISH	Arthrosis, enthesopathy	Periarticular calcifications	30	Lithiasis, Diabetes mellitus
AZ3	F	N/A	DISH	DISH	Enthesopathy, osteophytosis, arthrosis	Periarticular calcifications	N/A	Obesity
AZ4	M	Normal	Syndesmophytosis	Syndesmophytosis, anterior osteophytosis	Osteophytosis, calcifications in capsule, arthrosis	Periarticular calcifications	<40	Lithiasis, cardiac arrythmia

Definition of Radiological terminology 1) DISH: Continuous ossification along the anterolateral aspect of three or four contiguous vertebral bodies; 2) Syndesmophytosis: Vertical and symmetrical calcification of the lateral margins of the intervertebral disc space; 3) Osteophytosis: presence of spurs, which are outgrows of bone tissue; 4) Enthesopathy: calcification/ossification process at the site of the insertion of ligaments, tendons, fascia or articular capsule into bone; 5) Arthrosis: presence of joint space narrowing, sclerosis and osteophytosis.
Abbreviations: N/A: not available, C-Spine: Cervical Spine, T-Spine: Thoracic spine and L-Spine: Lumbar spine.

Sensing Receptor (*CASR*), Bone Morphogenetic Protein Receptor Type 1B (*BMPR1B*), Osteopontin (*OPN/SPP1*), Integrin Binding Sialoprotein (*IBSP*), Fibroblast Growth Factor 2 (*FGF2*), Inorganic Pyrophosphate Transport Regulator (*ANKH*), Collagen Type XI Alpha 2 (*COL11A2*), Nucleotide Pyrophosphatase 1 (*ENPP1*), Runt-related Transcription Factor (*RUNX2*), Dickkopf WNT Signaling Pathway (*DKK-1*), Insulin Like Growth Factor 1 (*IGF1*), Matrix Gla Protein (*MGP*); Vitamin D (*VDR*), Bone Morphogenetic Protein 4 (*BMP-4*), Collagen Type 1 Alpha 1 (*COL1A1*), Transforming Growth Factor Beta 1 (*TGF* 1), Solute Carrier Family 29 Member 1 (*SLC29A1*), Bone Morphogenetic Protein 2 (*BMP-2*) and Collagen Type VI Alpha 1 (*COL6A1*).

GENETIC VARIANTS – SELECTION BY PATHOGENICITY

In the selection by variant pathogenicity filtering strategy we used two levels; a "variant level" and a "knowledge level" (Figure 1). In the "variant level" we first filtered according to variant type and focused on functional significance (splice sites, frameshift coding, nonsynonymous coding and loss or a gain of a stop). Only the SIFT prediction with Deleterious and Unknown and GERP values equal or higher than 3 were included. We excluded variants with MAF values higher than 0.05 by filtering the SNP_IDs (rs number or chromosome position) against the Ensemble Variant Effect Predictor tool (http://www.ensembl.org/ info/docs/tools/vep/index.html). Then in the selection by "knowledge" we focused only on genes associated with calcification and/or ossification or related conditions. Lastly, we evaluated the functional significance and pathogenic potential of each variant found in the selected genes.

EVALUATION OF GENES AND VALIDATION OF VARIANTS

Information about candidate genes was obtained from several databases, including Ensembl (http://www.ensembl.org/index), National Center for Biotechnology Information (NCBI) (http://www.ncbi.nih.gov), Pubmed (https:// www.ncbi.nlm.nih.gov/pubmedwe), GeneCards (http://www.genecards.org/), Online Mendelian Inheritance in Man (OMIM) (http://www. omim.org/)

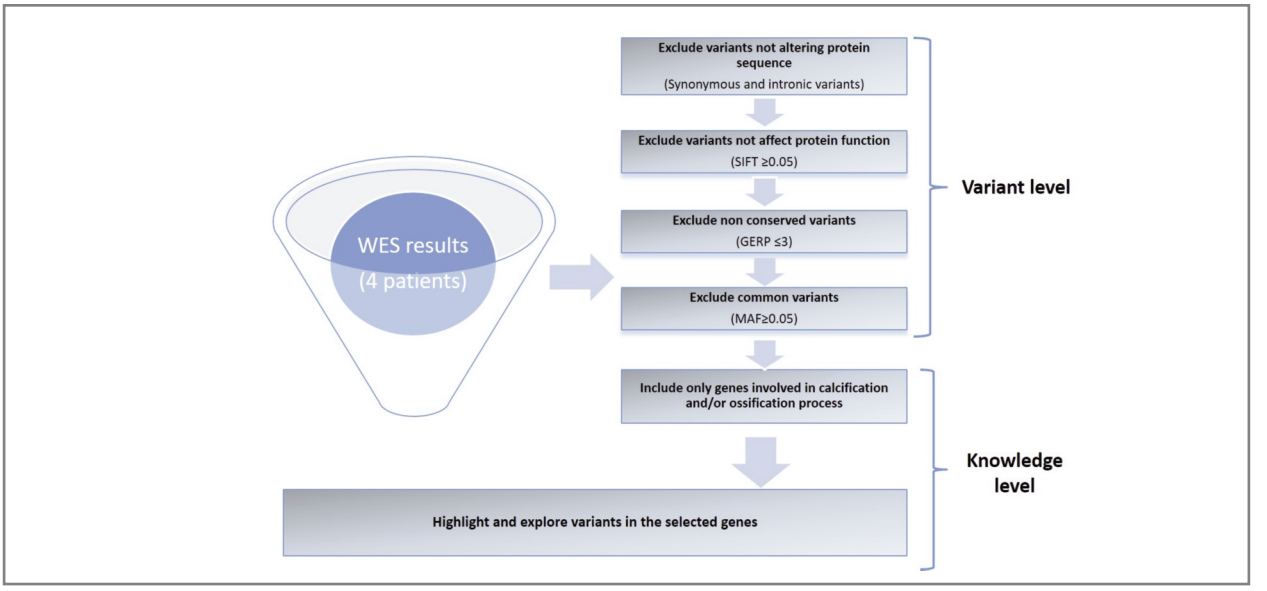

FIGURE 1. The two-level filter approach used to analyze the WES results from 4 patients with DISH/CC disease.
SIFT: Sorting Intolerant From Tolerant, GERP: Genomic Evolutionary Rate Profiling score and MAF: Minor Allele Frequency

and MalaCards (http://www.malacards.org/).

PCR primers were designed using the software Primer3 (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) to amplify and validate variants detected by exome sequencing. Sanger sequencing using standard protocols was performed using an automated DNA sequencer ABI 3130xl (Applied Biosystems©). Genetic variants were screened using Sequencing Analysis and SeqScape (Applied Biosystems©) and using a reference NCBI sequence.

The functional significance and the potential deleterious effect of each variant was explored using the following databases: Ensembl (http://www.ensembl. org/index), Human Gene Mutation Database (HGMD) (http://www.hgmd.cf.ac.uk/ac/index.php) and dbSNP (https://www.ncbi.nlm.nih.gov/projects/SNP/), making use of PolyPhen-2 (Polymorphism Phenotyping v2) (http://genetics.bwh.harvard.edu/pph2/), SIFT and GERP. PolyPhen is classified as, benign [0-0.2], possibly damaging [0.2-0.85], and probably damaging [0.85-1], SIFT as deleterious if less than 0.05 and GERP, ranges from -12.3 to 6.17, with 6.17 being the most conserved[15].

The MAF values of each variant were also analyzed. Protein conservation analysis was performed using ClustalW (http://www.genome.jp/tools/clustalw/) to compare homologous amino acid sequences among multiple vertebrates at the sites where the variations occur (accession numbers of transcripts available in supplementary Table II, available in supplementary file).

ASSOCIATION STUDIES AND STATISTICAL ANALYSIS

In order to verify a possible association between identified genes from WES and the DISH/CC phenotype, selected variants were screened in family members (affected with DISH/CC and unaffected), and the most conserved variants and those that were present in three or four WES patients, were selected for screening in a group of 65 DISH/CC patients and a group of 118 controls. Association testing was performed by chi-squared analysis using PLINK software (V1.07)[16]: p-values <0.05 were considered significant. ORs with a 95% CI were calculated for the minor alleles of each variant. OR >1 indicates a susceptibility allele and OR<1 indicates a protective allele for the disease.

RESULTS

The WES was performed using DNA from four patients from unrelated DISH/CC families (AZ1-AZ4) from Azores. A detailed description of these families is provided in our previous study[1]. The patients who gave informed consent, a blood sample was obtained and standard X-rays were taken from: the knees, axial skeleton, wrists, hands, elbows, and pelvis. The 4 patients were selected because they shared a very evident DISH phenotype. The radiological characterization of patients selected for WES are indicated in Table I.

TABLE II. LIST OF FILTERED CANDIDATE GENETIC VARIANTS CONFIRMED BY SANGER SEQUENCING.

Gene	Chr	SNP_ID	Mutational spectrum	Variant	AA	MAF	SIFT	PolyPhen	Patient			
									AZ1	AZ2	AZ3	AZ4
PPP2R2D	10	rs34473884	HC	c.1072G>A	G358S	0.18	D	B	hm	ht	ht	ht
COL11A2	6	rs9277934	RR	c.826G>A	E276K	0.32	T	Ps D	hm	ht	hm	ht
VDR	12	rs2228570	CM	c.2T>C	M1T	0.33	D	B	hm	hm	ht	ht
CASR	3	rs1801726	RR	c.3031G>C	E1021Q	0.08	T	B		hm	hm	hm
FGF2	4	rs1048201	NA	c.*757C>T	R209Q	0.23	NA	NA	ht	ht		ht
BMP2	20	rs235768	HC	c.570A>T	R190S	0.24	D	Pb D	ht		ht	ht
ENPP1	6	rs1044498	RR	c.517A>C	K173Q	0.34	T	B	ht	ht	ht	
BMP4	14	rs17563	CM	c.455T>C	V152A	0.37	T	B	hm		hm	ht
MGP	12	rs4236	RR	c.304A>G	T102A	0.39	T	B	ht		ht	ht
PLCG2	16	rs138158454	RR	c.2054+7G>A	NA	0.01	NA	NA			ht	ht
ALPL	1	rs3200254	RR	c.787T>C	Y263H	0.27	T	B	ht		ht	
CASR	3	rs1801725	HC	c.2956G>T	A986S	0.09	T	B		ht		ht
COL1A1	17	rs372029024	CM	c.3247G>T	A1083T	<0.01*	T	B	hm			
PLCG2	16	rs374430619	RR	c.3786G>C	K1262N	<0.01*	D	B		ht		
ALPL	1	rs149344982	RR	c.455G>A	R152H	0.01	T	B		ht		
ABCC6	16	rs41278174	CM	c.3190C>T	R1064W	0.01	D	Pb D			ht	
COL11A2	6	rs2229792	RR	c.5165C>T	P1722L	0.01	D	B				ht
TGFβ1	19	rs55659002	RR	c.713-8delC	NA	0.01	NA	NA			ht	
AMER3	2	rs72854996	RR	c.1300C>G	L434V	0.05	T	B	ht			
FLNC	7	rs2291569	HC	c.4700G>A	R1567Q	0.06	D	Pb D		ht		
COL6A1	21	rs1053312	RR	c.2549G>A	R850H	0.27	T	B		hm		

The variants are listed in the table according to occurrence in two or more of the patients AZ1 to AZ4 in the study.
Abbreviations: Chr: chromosome, HC: highly conserved, CM: conserved in mammals, RR: relaxed region, NA: Not Applicable, AA: Aminoacid, MAF: Minor Allele Frequency, hm: homozygous, ht: heterozygous, Blank cells represent wild type, T: Tolerated, D: Deleterious, Pb D for "Probably Damaging", Ps D for "Possibly Damaging" and B for Benign. * Allele frequencies assumed from NHLBI GO Exome Sequencing Project.

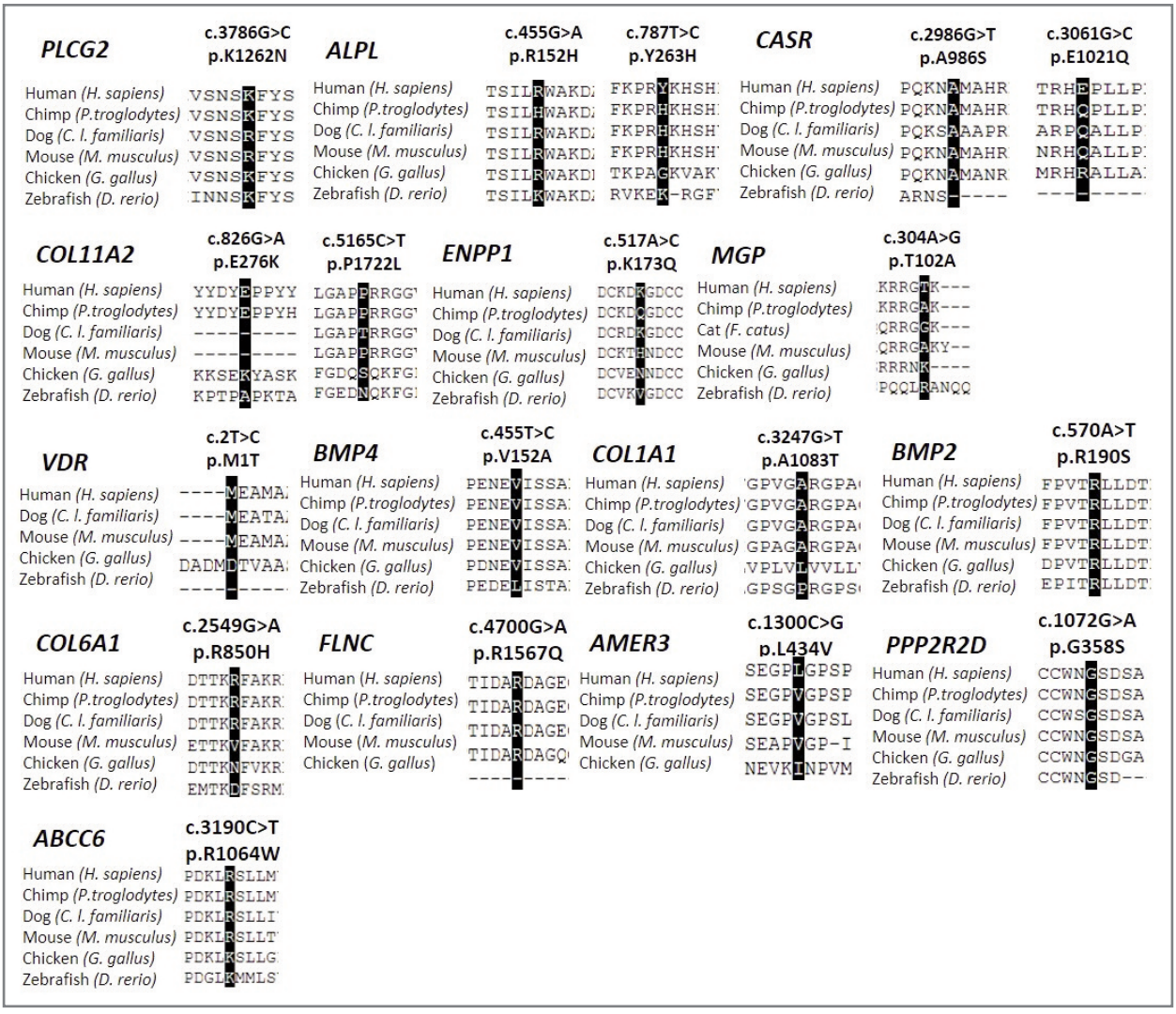

FIGURE 2. Sequence conservation in six vertebrates of the variants identified in candidate genes possibly associated with the DISH/CC phenotype in humans (some genes are not available in all species). The variant for the FGF2 gene is not presented since it is located in the 3' untranslated region.
Abbreviations: PLCG2- Phospholipase C Gamma 2, ALPL- Alkaline Phosphatase, Liver/Bone/Kidney, CASR- calcium-sensing receptor, COL11A2- Collagen Type XI Alpha 2 Chain, ENPP1– ectonucleotide pyrophosphatase/phosphodiesterase 1, MGP- Matrix Gla Protein, VDR- Vitamin D Receptor, BMP4- Bone morphogenetic protein 4, COL1A1- Collagen Type I Alpha 1 Chain, BMP2- Bone morphogenetic protein 2, COL6A1- Collagen Type VI Alpha 1 Chain, FLNC- Filamin C, AMER3- APC Membrane Recruitment Protein 3, PPP2R2D- Protein Phosphatase 2 Regulatory Subunit B delta, ABCC6 – ATP-binding cassete subfamily C, member 6.

FILTERING RESULTS

Approximately 38 Mb of sequence per patient was generated and the capture specificity and sensitivity in all samples was about 55% and 94%, respectively. The results for each sample after SNV calling and indel identification steps are summarized in supplementary Table III.

After filtering, 21 missense, deletion and splice site variants in 17 genes were obtained: *PLCG2*, *ALPL*, *CASR*, *FGF2*, *COL11A2*, *ENPP1*, *MGP*, *VDR*, *BMP-4*, *COL1A1*, *TGF 1*, *BMP-2*, *COL6A1*, *FLNC*, *AMER3*, *PPP2R2D* and *ABCC6*. Ten of the identified variants were present in the HGMD mutation database, linked to phenotypes, other than DISH/CC. The identified variants and available functional information are presented in Table II.

Sequence conservation of the missense variants were

TABLE III. ASSOCIATION STUDY BETWEEN SEVEN VARIANTS FROM SEVEN GENES AND THE DISH/CC PHENOTYPE

Chr	Gene	SNP M/m	Allele MAF	MAF DISH/CC (N=65)	MAF Controls (N=118)	OR (95% CI)	p-value
10	PPP2R2D	rs34473884	G/A	0.262	0.165	1.789 (1.060 - 3.021)	0.028
14	BMP4	rs17563	T/C	0.454	0.390	1.301 (0.843 - 2.006)	0.234
4	FGF2	rs1048201	C/T	0.139	0.131	1.063 (0.569 - 1.985)	0.849
6	ENPP1	rs1044498	A/C	0.200	0.203	0.979 (0.574 - 1.670)	0.938
6	COL11A2	rs9277934	G/A	0.385	0.360	1.110 (0.713 - 1.728)	0.643
20	BMP2	rs235768	A/T	0.292	0.284	1.042 (0.650 - 1.671)	0.865
12	VDR	rs2228570	T/C	0.400	0.373	1.121 (0.723 - 1.739)	0.609

The minor alleles are indicated in bold.
Abbreviations: Chr: Chromossome, SNP: Single nucleotide polymorphism, M/m: major allele/minor allele, MAF: Minor allele frequency and OR (95% CI): Odds Ratio (95% Confidence Interval).

then evaluated in 6 different vertebrates. It is evident in Figure 2 that the variants in the genes *PPP2R2D*, *BMP2*, *FLNC*, and *CASR* are highly conserved in all the vertebrates analyzed. Variants of the genes *VDR*, *BMP4*, *COL1A1* and *ABCC6* are only conserved in mammals. The degree of conservation is not directly linked to lower allele frequency, since several of these variants have high Minor Allele Frequencies (MAF).

ASSOCIATION BETWEEN VARIANTS AND THE DISH/CC PHENOTYPE

Two cohorts, one of 65 unrelated DISH/CC patients (45 males, 20 females; age of onset around 40 years) and another of 118 unrelated controls without any signs of DISH/CC, with a similar ethnic background, were selected after radiological characterization (47 males, 71 females; mean current age, 68 years; range, 60-104) and used for association studies. In cases we selected younger patients around the age of 40, where genetic factors may be associated, however in the controls we selected older individuals to ensure that they have not developed the disease.

All patients had at least 2 highly conserved genetic variants in combination with other variants that are normally conserved in mammals (Table II). Variants with a high degree of conservation, present in 4 or 3 WES patients, irrespective of the MAF, were selected for the association study. The variants rs34473884, rs9277934 and rs2228570 in *PPP2R2D*, *COL11A2* and *VDR* genes, respectively, were present in all four WES patients. The variants rs1048201, rs235768, rs1044498 and rs17563 in *FGF2*, *BMP2*, *ENPP1* and *BMP4* genes, respectively, were present in three of the four patients selected for WES. Seven variants were screened in a group of 65 DISH/CC patients and 118 controls and the results are indicated in Table III. The variant rs34473884 in the *PPP2R2D* gene was the only that gave significant results. It was found to be much more frequent in the DISH/CC group than in the controls (p=0.028; OR=1.789, 95% CI= 1.060 - 3.021).

DISCUSSION

In this study, we used WES as a method to identify candidate genes for DISH/CC aetiology and association studies to investigate some of the identified variants. As expected, thousands of protein coding variants per patient were identified across each exome. Consequently, a number of filtering strategies were used to select potential high risk variants of genes potentially associated with DISH. The filtering strategies employed were based on previous knowledge about gene function. The major limitations of the strategy used to identify candidate genes is that unannotated genes and those with unknown functions were not investigated in the study.

Very few genetic studies have been published about DISH so far. DISH susceptibility genes have previously been investigated, including Human Leukocyte Antigens (HLA),[17-24] Collagen 6A1 gene (*COL6A1*),[25] Fibroblast Growth factor 2 (*FGF2*),[26] Vitamin D (1,25-Dihydroxyvitamin D3) Receptor (*VDR*) and Collagen Type I_1 (*COL1A1*),[27] but only two genes are known to

have a positive association with DISH susceptibility; *COL6A1*[28-29] and *FGF2*.[26] However, the *COL6A1* and *FGF2* variants associated with the disease, are located in non-coding regions and are common variants within the general population, suggesting they have a small genetic effect.

In our study we found 3 genetic variants that were present in all 4 DISH/CC patients and 6 that were present in 3 DISH/CC patients. Minor allele frequencies of these variants were high (mean 0.25), meaning that these were common variants in the general population. Rare nonsynonymous SNPs are two times more likely to be predicted as protein-affecting, when compared to common SNPs. However, in our study three common SNPs, shared by the 4 patients (G358S, E276K and M1T in *PPP2R2D*, *COL11A2* and *VDR* genes, respectively), according to SIFT and/or PolyPhen algorithms were predicted to have a strong effect on the protein. Eight of these genetic variants, in heterozygous and/or homozygous states, were in conserved positions of proteins associated with mineralization; four of which were in regions that were highly conserved across the vertebrates. The four WES selected patients had at least 2 highly conserved genetic variants in combination with other variants that were conserved in mammals. The association study indicated that the SNP rs34473884 in the *PPP2R2D* gene was significantly associated with the DISH/CC phenotype.

In humans, the full *PPP2R2D* (Protein Phosphatase 2 Regulatory Subunit delta) gene structure is composed of 9 exons and spans 79.19 kb in chromosome 10. The gene encodes a crucial serine/threonine protein phosphatase that regulates basal cellular activities by dephosphorylating substrates. Protein Ser/Thr phosphatases are a group of enzymes that catalyze the removal of phosphate groups from serine and/or threonine residues by hydrolysis of phosphoric acid monoesters. They oppose the action of kinases and phosphorylases and are involved in signal transduction. Protein phosphatases have long been postulated to influence TGF- superfamily signaling, which regulates numerous cellular responses[30]. Despite the long-standing suspected influence of protein phosphatases in TGF- signaling, concrete data confirming the interaction only started to emerge recently. In human cell lines, *PPP2R2D* negatively modulates *TGF-* /Activin/ /Nodal signaling by inhibiting the type I receptors ALK4 and ALK5[31].

The variant c.1072G>A (G327S) found in *PPP2R2D* gene is a missense variant which causes substitution of glycine for serine at amino acid 327 in the protein. These two amino acids are hydrophilic but glycine is non-polar and serine is polar. The glycine is normally totally conserved in all vertebrates studied, and the variant has a low, deleterious, SIFT score (0.03) but the Polyphen score (0.33) does not corroborate its harmful effect. The frequency of this variant is high in European populations with a MAF of 0.18 and this variant was identified in all four DISH/CC patients used for WES; AZ1 was homozygous and AZ2, 3 and 4 were heterozygous. As far as we known no phenotype has been associated with this gene.

Rare variants, with a deleterious effect on proteins in the human genome, are the basis for the development of Mendelian diseases. However, common polymorphisms that individually exert small effects might, as a group, also play a substantial genetic role. Despite the lack of association of most of the studied gene variants in the present study, this may be in part linked to the heterogeneous features of the phenotype, and so improved characterization of the disorder under study is essential for the planning of future studies. The results of our study lead us to suggest that DISH/CC is polygenic, and is influenced by the interaction of multiple, small effect gene variants and possibly by unknown environmental factors.

CONCLUSION

Our results underline the polygenic nature of DISH and a number of conserved and sometimes deleterious variants were identified in genes with a role in mineralization. The variant rs34473884 in the *PPP2R2D* gene was significantly associated with the DISH/CC phenotype and we propose it may contribute to the development of this disorder. Further studies will be needed to confirm the association of *PPP2R2D* with the phenotype under study.

ACKNOWLEDGEMENTS

The authors thank all the patients who participated in this research and made it possible. We also thank Isa Dutra (MSc), João Paulo Pinheiro (BsC) and Raquel Meneses (BsC) from Hospital Santo Espírito da Ilha Terceira, EPE, Epidemiology and Molecular Biology Service (SEEBMO), Angra do Heroísmo, Portugal, who performed the DNA extractions and Vânia Machado (MSc), who participated in the elaboration of the pedigrees. BP was supported by "Fundo Regional para a Ciência e Tecnologia (FRCT)" (M3.1.2/F/023/2011).

COMPETING FINANCIAL INTERESTS

The authors declare no competing interests.

CORRESPONDENCE TO
Jácome Bruges Armas
Serviço Especializado de Epidemiologia e Biologia Molecular (SEEBMO)
Hospital de Santo Espírito da Ilha Terceira (HSEIT)
Canada do Briado
9700-049, Angra do Heroísmo
E-mail: brugesarmas@gmail.com

REFERENCES

1. Bruges-Armas J, Couto AR, Timms A, Santos MR, Bettencourt BF, Peixoto MJ, et al. Ectopic calcification among families in the Azores: clinical and radiologic manifestations in families with diffuse idiopathic skeletal hyperostosis and chondrocalcinosis. Arthritis Rheum 2006; 54: 1340-1349.
2. Mader R. Clinical manifestations of diffuse idiopathic skeletal hyperostosis of the cervical spine. Sem Arthritis Rheum 2002; 32: 130-135.
3. Mader R, Verlaan JJ, Buskila D. Diffuse idiopathic skeletal hyperostosis: clinical features and pathogenic mechanisms. Nat Rev Rheumatol 2013; 9: 741-750.
4. Utsinger PD. Diffuse Idiopathic Skeletal Hyperostosis. Clin Rheum Dis 1985; 11: 325-351.
5. Weinfeld RM, Olson PN, Maki DD, Griffiths HJ. The prevalence of diffuse idiopathic skeletal hyperostosis (DISH) in two large American Midwest metropolitan hospital populations. Skeletal Radiol 1997; 26: 222-225.
6. Beardwell A. Familial Ankylosing Vertebral Hyperostosis With Tylosis. Ann Rheum Dis 1969; 28: 518-523.
7. Woodard JC, Poulos PW, Jr., Parker RB, Jackson RI, Jr., Eurell JC. Canine diffuse idiopathic skeletal hyperostosis. Vet Pathol 1985; 22: 317-326.
8. Morgan J, Stavenborn M. Disseminated idiopathic skeletal hyperostosis (DISH) in a dog. Vet Radiol Ultrasound 1991; 32: 65-70.
9. Gutierrez M, Silveri F, Bertolazzi C, Giacchetti G, Tardella M, Di Geso L, et al. [Gitelman syndrome associated with chondrocalcinosis: description of two cases]. Reumatismo 2010; 62: 60-64.
10. Netter P, Bardin T, Bianchi A, Richette P, Loeuille D. The ANKH gene and familial calcium pyrophosphate dihydrate deposition disease. Joint Bone Spine 2004; 71: 365-368.
11. Ramos YF, Bos SD, van der Breggen R, Kloppenburg M, Ye K, Lameijer EW, et al. A gain of function mutation in TNFRSF11B encoding osteoprotegerin causes osteoarthritis with chondrocalcinosis. Ann Rheum Dis 2015; 74: 1756-62.
12. Couto AR, Parreira B, Thomson R, Soares M, Power DM, Stankovich J, et al. Combined approach for finding susceptibility genes in DISH/chondrocalcinosis families: whole-genome-wide linkage and IBS/IBD studies. Hum Genome Var 2017; 4: 17041.
13. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, et al. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet 2010; 42: 30-35.
14. Rabbani B, Tekin M, Mahdieh N. The promise of whole-exome sequencing in medical genetics. J Hum Genet 2014; 59: 5-15.
15. Cooper GM, Stone EA, Asimenos G, Green ED, Batzoglou S, Sidow A. Distribution and intensity of constraint in mammalian genomic sequence. Genome Res 2005; 15: 901-913.
16. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81: 559-575.
17. Rosenthal M, Bahous I, Müller W. Letter: HL-A B27 and modified bone formation. Lancet 1976 1: 543.
18. Rotés Querol J, Ercilla González G. Letter: HLA-B27 and modified bone formation. Lancet 1976 1: 482-3.
19. Brigode M, Francois RJ. Histocompatibility antigens in vertebral ankylosing hyperostosis. J Rheumatol 1977 4: 429-434.
20. Spagnola AM, Bennett PH, Terasaki PI. Vertebral ankylosing hyperostosis (Forestier's disease) and HLA antigens in Pima Indians. Arthritis Rheum 1978 21: 467-72.
21. Perry JD, Wolf H, Festenstein H, Storey GO. Ankylosing hyperostosis: a study of HLA A, B, and C antigens. Ann Rheum Dis 1979 38: 72-73.
22. Dostál C, Ivasková E, Hána I, Sváb V. HLA antigens in vertebral ankylosing hyperostosis (Forestier's disease). J Hyg Epidemiol Microbiol Immunol 1983; 27: 98-102.
23. Rosenthal M, Bahous I, Muller W. Increased frequency of HLA B8 in hyperostotic spondylosis. J Rheumatol Suppl 1977; 3: 94--96.
24. Ercilla González G, Brancós MA, Breisse Y. Histocompatability antigens in Forestier's disease, polyarthrosis and ankylosing spondylitis. HLA and Disease. Paris: INSERM; 1976. p. 28.
25. Tsukahara S, Miyazawa N, Akagawa H, Forejtova S, Pavelka K, Tanaka T, et al. COL6A1, the candidate gene for ossification of the posterior longitudinal ligament, is associated with diffuse idiopathic skeletal hyperostosis in Japanese. Spine 2005 30: 2321--2324.
26. Jun JK, Kim SM. Association study of fibroblast growth factor 2 and fibroblast growth factor receptors gene polymorphism in korean ossification of the posterior longitudinal ligament patients. J Korean Neurosurg Soc 2012; 52: 7-13.
27. Havelka S, Uitterlinden AG, Fang Y, Arp PP, Pavelková A, Veselá M, et al. Collagen type I(alpha 1) and vitamin D receptor polymorphisms in diffuse idiopathic skeletal hyperostosis. Clin Rheumatol 2002 21: 347-348.
28. Tanaka T, Ikari K, Furushima K, Okada A, Tanaka H, Furukawa K, et al. Genomewide linkage and linkage disequilibrium analyses identify COL6A1, on chromosome 21, as the locus for ossification of the posterior longitudinal ligament of the spine. Am J Hum Genet 2003; 73: 812-822.
29. Kong Q, Ma X, Li F, Guo Z, Qi Q, Li W, et al. COL6A1 polymorphisms associated with ossification of the ligamentum flavum and ossification of the posterior longitudinal ligament. Spine (Phila Pa 1976) 2007; 32: 2834-2838.
30. Liu T, Feng XH. Regulation of TGF-beta signalling by protein phosphatases. Biochem J 2010; 430: 191-198.
31. Batut J, Schmierer B, Cao J, Raftery LA, Hill CS, Howell M. Two highly related regulatory subunits of PP2A exert opposite effects on TGF-beta/Activin/Nodal signalling. Development 2008; 135: 2927-2937.

SUPPLEMENTARY FILES

SUPPLEMENTARY TABLE I. NUMBER OF CANDIDATE GENES PER SAMPLE AND NUMBER OF CANDIDATE GENES SHARED BY DISH/CC PATIENTS

Samples	Number of candidate genes	
	Dominant model	Recessive Model
AZ1	815	48
AZ2	917	58
AZ3	872	47
AZ4	593	25
AZ1+AZ2	220	6
AZ1+AZ3	212	4
AZ1+AZ4	139	4
AZ2+AZ3	249	6
AZ2+AZ4	149	4
AZ3+AZ4	162	3
AZ1+AZ2+AZ3	118	2
AZ1+AZ2+AZ4	65	1
AZ2+AZ3+AZ4	78	2
AZ1+AZ3+AZ4	77	1
AZ1+AZ2+AZ3+AZ4	52	1

SUPPLEMENTARY TABLE II. LIST OF GENES AND THEIR ACCESSION TRANSCRIPT NUMBERS USED FOR CONSERVATION ANALYSIS.

GENES	SPECIES					
	Human	Chimpanzee	Dog	Mouse	Chicken	Zebrafish
PLCG2	ENST00000564138.5	ENSPTRT00000015479.3	ENSCAFT00000031815.3	ENSMUST00000081232.8	ENSGALT00000050238.1	ENSDART00000021399.7
ALPL	ENST00000374840.7	ENSPTRT00000000592.2	ENSCAFT00000023578.3	ENSMUST00000030551.10	ENSGALT00000068960.1	ENSDART00000131101.2
CASR	ENST00000498619.3	ENSPTRT00000043996.4	ENSCAFT00000018760.3	ENSMUST00000063597.13	XM_416491.5.1	ENSDART00000010934.8
COL11A2	ENST00000341947.6	ENSPTRT00000048564.4	ENSCAFT00000001409.4	ENSMUST00000087497.10	Ni	ENSDART00000105754.4
ENPP1	ENST00000360971.6	ENSPTRT00000034356.3	ENSCAFT00000000001.3	ENSMUST00000105520.7	ENSGALT00000066498.1	ENSDART00000127350.3
MGP	ENST00000539261.5	ENSPTRT00000008732.4	ENSFCAT00000031714.1 (Cat)	ENSMUST00000032342.2	ENSGALT00000019173.4	ENSDART00000149622.2
VDR	ENST00000229022.7	Ni	ENSCAFT00000043473.2	ENSMUST00000023119.14	ENSGALT00000071682.1	ENSDART00000161892.1
BMP4	ENST00000245451.8	ENSPTRT00000011643.4	ENSCAFT00000023624.3	ENSMUST00000074077.11	ENSGALT00000020316.5	ENSDART00000075150.4
COL1A1	ENST00000225964.9	ENSPTRT00000017231.4	ENSCAFT00000026953.3	ENSMUST00000001547.7	XM_015273228.1.1	ENSDART00000009393.7
BMP2	ENST00000378827.4	ENSPTRT00000024606.3	ENSCAFT00000049690.1	ENSMUST00000028836.6	ENSGALT00000065435.1	ENSDART00000166657.1
COL6A1	ENST00000361866.7	ENSPTRT00000026156.3	ENSCAFT00000018918.3	ENSMUST00000001147.4	ENSGALT00000039669.3	ENSDART00000110608.3
FLNC	ENST00000325888.12	ENSPTRT00000036444.5	ENSCAFT00000002504.3	ENSMUST00000065090.7	NM_204573.1.1	Ni
AMER3	ENST00000423981.1	ENSPTRT00000023124.3	ENSCAFT00000006747.2	ENSMUST00000052670.10	ENSGALT00000055960.1	ENSDART00000149992.2
PPP2R2D	ENST00000455566.5	ENSPTRT00000005829.4	ENSCAFT00000044416.1	ENSMUST00000041097.12	ENSGALT00000081063.1	ENSDART00000172175.1
ABCC6	ENST00000205557.11	ENSPTRT00000014398.4	ENSCAFT00000028908.3	ENSMUST00000002850.7	ENSGALT00000048627.1	ENSDART00000172943.1

Data was retrieved from the Ensembl genome database; accessed on November 2016.
Ni: not identified.

SUPPLEMENTARY TABLE III. KNOWN, NOVEL AND TOTAL NUMBER OF SNVS AND INDELS FOR THE FOUR SAMPLES TESTED (AZ1-4)

	SNVs			Indels				TOTAL/
Sample	Known	Novel	Total	Known	Novel	Complex	Total	sample
AZ1	17281	1619	18900	457	405	51	913	19813
AZ2	18086	1759	19845	440	496	66	1002	20847
AZ3	18327	1650	19977	524	505	73	1102	21079
AZ4	17575	1223	18798	475	424	72	971	19769

SNVs: Single nucleotide variants.