

RESEARCH ARTICLE

Rare functional genetic variants in *COL7A1*, *COL6A5*, *COL1A2* and *COL5A2* frequently occur in Chiari Malformation Type 1

Aintzane Urbizu^{1,2}, Melanie E. Garrett¹, Karen Soldano¹, Oliver Drechsel^{3,4}, Dorothy Loth⁵, Anna Marcé-Grau², Olga Mestres i Soler⁶, Maria A. Poca^{6,7}, Stephan Ossowski^{3,4}, Alfons Macaya², Francis Loth^{8,9}, Rick Labuda¹⁰, Allison Ashley-Koch^{1*}



1 Duke Molecular Physiology Institute, Duke University Medical Center, Durham, NC, United States of America, **2** Pediatric Neurology Research Group, Vall d'Hebron Research Institute, Barcelona, Spain, **3** Genomic and Epigenomic Variation in Disease Group, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain, **4** Universitat Pompeu Fabra, Barcelona, Spain, **5** Department of Psychology, Conquer Chiari Research Center, University of Akron, Akron, OH, United States of America, **6** Neurotraumatology and Neurosurgery Research Unit, Vall d'Hebron Research Institute, Universitat Autònoma de Barcelona, Barcelona, Spain, **7** Department of Neurosurgery, Vall d'Hebron University Hospital, Universitat Autònoma de Barcelona, Barcelona, Spain, **8** Department of Biomedical Engineering, Conquer Chiari Research Center, University of Akron, Akron, OH, United States of America, **9** Department of Mechanical Engineering, Conquer Chiari Research Center, University of Akron, Akron, OH, United States of America, **10** Conquer Chiari, Wexford, PA, United States of America

OPEN ACCESS

Citation: Urbizu A, Garrett ME, Soldano K, Drechsel O, Loth D, Marcé-Grau A, et al. (2021) Rare functional genetic variants in *COL7A1*, *COL6A5*, *COL1A2* and *COL5A2* frequently occur in Chiari Malformation Type 1. PLoS ONE 16(5): e0251289. <https://doi.org/10.1371/journal.pone.0251289>

Editor: Obul Reddy Bandapalli, German Cancer Research Center (DKFZ), GERMANY

Received: February 13, 2020

Accepted: April 24, 2021

Published: May 11, 2021

Copyright: © 2021 Urbizu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Although the authors cannot make their study's data publicly available at the time of publication, all authors commit to make the data underlying the findings described in this study fully available without restriction to those who request the data, in compliance with the PLOS Data Availability policy. For data sets involving personally identifiable information or other sensitive data, data sharing is contingent on the data being handled appropriately by the data requester and in accordance with all applicable local requirements. Appropriate requests of the

* allison.ashleykoch@duke.edu

Abstract

Chiari Malformation Type 1 (CM-1) is characterized by herniation of the cerebellar tonsils below the foramen magnum and the presence of headaches and other neurologic symptoms. Cranial bone constriction is suspected to be the most common biologic mechanism leading to CM-1. However, other mechanisms may also contribute, particularly in the presence of connective tissue disorders (CTDs), such as Ehlers Danlos Syndrome (EDS). Accumulating data suggest CM-1 with connective tissue disorders (CTD+) may have a different patho-mechanism and different genetic risk factors than CM-1 without CTDs (CTD-). To identify CM-1 genetic risk variants, we performed whole exome sequencing on a single large, multiplex family from Spain and targeted sequencing on a cohort of 186 unrelated adult, Caucasian females with CM-1. Targeted sequencing captured the coding regions of 21 CM-1 and EDS candidate genes, including two genes identified in the Spanish family. Using gene burden analysis, we compared the frequency of rare, functional variants detected in CM-1 cases versus publically available ethnically-matched controls from gnomAD. A secondary analysis compared the presence of rare variants in these genes between CTD+ and CTD- CM-1 cases. In the Spanish family, rare variants co-segregated with CM-1 in *COL6A5*, *ADGRB3* and *DST*. A variant in *COL7A1* was present in affected and unaffected family members. In the targeted sequencing analysis, rare variants in six genes (*COL7A1*, *COL5A2*, *COL6A5*, *COL1A2*, *VEGFB*, *FLT1*) were significantly more frequent in CM-1 cases compared to public controls. In total, 47% of CM-1 cases presented with rare variants in at least one of the four significant collagen genes and 10% of cases harbored variants in multiple significant collagen genes. Moreover, 26% of CM-1 cases presented with

data will require an existing IRB at the recipient institution, as well as data use/transfer agreements established between the recipient institution and the institutions owning the data. Data requests can be sent to the following authors: Allison Ashley-Koch--allison.ashleykoch@duke.edu Aintzane Urbizu--aintzaubi@gmail.com Dorothy Loth--dloth@uakron.edu.

Funding: This work was supported by a grant from Conquer Chiari to AAK. Collection of the Chiari1000 study participants utilized in this study was supported by a grant from Conquer Chiari to FL at University of Akron. Collection of the Duke study participants utilized in this study was supported by a grant from the National Institutes of Health (NS063273). A.U. was the recipient of a Postdoctoral Fellowship from Fundación Ramón Areces (Spain). RL is the Executive Director of Conquer Chiari which provided some of the funding for this work. For the manuscript, he assisted with revising and editing the manuscript. The funders did have a role in study design, but had no role in data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: As noted in the previous section, RL is the Executive Director of Conquer Chiari, an organization which provided some funding for the work presented. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

rare variants in the *COL6A5* gene. We also identified two genes (*COL7A1*, *COL3A1*) for which the burden of rare variants differed significantly between CTD+ and CTD- CM-1 cases. A higher percentage of CTD+ patients had variants in *COL7A1* compared to CTD- patients, while CTD+ patients had fewer rare variants in *COL3A1* than did CTD- patients. In summary, rare variants in several collagen genes are particularly frequent in CM-1 cases and those in *COL6A5* co-segregated with CM-1 in a Spanish multiplex family. *COL6A5* has been previously associated with musculoskeletal phenotypes, but this is the first association with CM-1. Our findings underscore the contribution of rare genetic variants in collagen genes to CM-1, and suggest that CM-1 in the presence and absence of CTD symptoms is driven by different genes.

Introduction

Chiari Malformations (CMs) are a group of malformations characterized by a downward herniation of the caudal part of the cerebellum and/or medulla oblongata into the spinal canal. Chiari Malformation Type 1 (CM-1) is the most common form of CMs, and is typically diagnosed as cerebellum tonsillar herniation (TH) of 3mm or more below the foramen magnum using magnetic resonance imaging (MRI) [1]. Clinically, CM-1 frequently presents with headache, ocular disturbances, vertigo, sleep apnea, lower cranial nerve signs, and motor and sensory symptoms. The symptoms are thought to result from a direct compression of neural tissue at the craniocervical junction and can also include cerebrospinal fluid disturbances, such as syringomyelia. Importantly, CM-1 individuals meeting diagnostic criteria on imaging can also be asymptomatic. TH in CM-1 is typically attributed to a cranial constriction mechanism caused by a small posterior cranial fossa (PCF) and concomitant shorter occipital bone. This mechanism may be caused by an insufficient paraxial mesoderm (classical CM-1) [2], premature synchondrosis closure, premature fusion of the skull bones (craniosynostosis) or a consequence of other conditions such as acromegaly or Paget's disease [3]. However, additional mechanisms can also lead to TH, such as space-occupying intracranial lesions, cranial settling or tethered cord syndrome [3]. The other forms of CMs are generally less common and it remains unclear whether the biologic mechanisms leading to other CMs are related to those underlying CM-1.

Even within CM-1, there is clinical complexity. CM-1 is often comorbid with other conditions, especially connective tissue disorders (CTDs). In particular, CM-1 frequently co-occurs with Ehlers-Danlos Syndrome (EDS), a heterogeneous group of heritable CTDs characterized mainly by skin hyperextensibility, abnormal wound healing, and joint hypermobility [4, 5]. There are thirteen EDS subtypes caused by mutations in collagen-encoding genes or in genes encoding collagen-modifying enzymes [6]. CM-1 can also be comorbid with other CTD genetic syndromes, such as Marfan syndrome and Klippel Feil syndrome [7, 8]. Importantly, comorbidity with CTDs is also associated with differences in cranial morphology [5], suggesting that the comorbidity represents distinct clinical subtypes of CM-1.

The co-segregation of CM-1 with other known genetic conditions also suggests that genetic factors contribute to risk for CM-1. However, even in the absence of other comorbid conditions, several lines of evidence support a genetic basis for CM-1 risk. Between 12–20% of individuals with CM-1 report having a relative who has also been diagnosed with CM-1 or has similar clinical symptoms [9]. Several families comprised of multiple individuals with CM-1 (multiplex families) have been described, some presenting with complete or incompletely

penetrant autosomal dominant inheritance and others with autosomal recessive inheritance [9–14]. Monozygotic twins who are concordant for CM-1 have also been observed [7]. Together, these findings strongly argue for a genetic contribution to CM-1 pathogenicity in at least a subset of cases. Despite this, the precise genetic variants contributing to CM-1 remain elusive in most cases.

Identifying the genetic risk factors for CM-1 can improve the understanding of the mechanistic processes that lead to the condition, but also provide the opportunity for genetic testing to families and individuals and ultimately, the possibility of therapeutic intervention. Several genetic approaches (linkage analysis, genetic association, expression analysis, sequencing studies) have been applied to CM-1. These studies have implicated genes involved in a multitude of biologic processes, including bone development and differentiation, angiogenesis and vasculogenesis, components of the extracellular matrix (ECM) and connective tissue, cell proliferation and differentiation, and ribosomal proteins [8, 11, 12, 15–18]. Most of the findings from these studies have been limited to single cohorts, with little or no evidence for replication. In short, a wide array of cellular and developmental processes may play a role in risk for CM-1, making the task of finding the relevant genes all the more challenging. Reducing phenotypic heterogeneity, and purportedly, the underlying genetic heterogeneity, can improve the ability to detect associated genetic factors. For example, stratifying families and patient cohorts on the presence of CTD-related symptoms or MRI-morphometric traits has successfully uncovered genetic variants that were not identified when analyzing all families or patients together [12, 15–17].

In recent years, genetic and genomic technologies have shifted towards next generation sequencing (NGS) approaches. Application of these approaches to CM-1 could improve the identification of causal variants in family and cohort studies. Exome sequencing has previously been applied to two Italian families with CM-1 resulting in three potential candidates, including *DKK1*, *LRP4* and *BMP1* [19]. The mutations in those genes co-segregated in the two families with CM-1, but screening 65 unrelated additional CM-1 probands only identified one additional missense mutation in *DKK1*. Another study performed exome sequencing on seven large multiplex Russian CM-1 families and linkage analysis suggested two chromosomal regions on 1q43-44 and 12q23-24.11, but the actual gene(s) and variants responsible were not definitively identified [14]. These findings highlight the potential power of NGS in identifying CM-1 genes, but also underscore the extensive genetic heterogeneity and the challenges for replicating genetic findings for this condition.

Here, we have utilized NGS technology to identify rare, putative causal genetic variants, and explored genetic heterogeneity as a function of comorbid symptoms consistent with CTD. We first performed whole exome sequencing in a Spanish family with multiple CM-1 affected individuals in an attempt to increase the likelihood of identifying at least one CM-1 gene. Next, we performed targeted sequencing in a US Caucasian cohort of the most promising candidate genes identified in the Spanish family, as well as several other genes previously associated with CM-1 and EDS. Our findings implicate several extracellular matrix and bone development genes in risk for CM-1, and suggest that rare variants in collagen genes are commonly present in CM-1.

Results

WES analysis of Spanish multiplex family

Variant detection from the whole exome sequencing data performed in four CM-1 affected members (see Fig 1, individuals II.1, II.4, III.1 and III.4) of a Spanish multiplex family yielded 974,936 total variants. After filtering as described in Methods, 2516 candidate variants remained (exonic and UTR). Among those, 1580 variants were annotated as affecting the

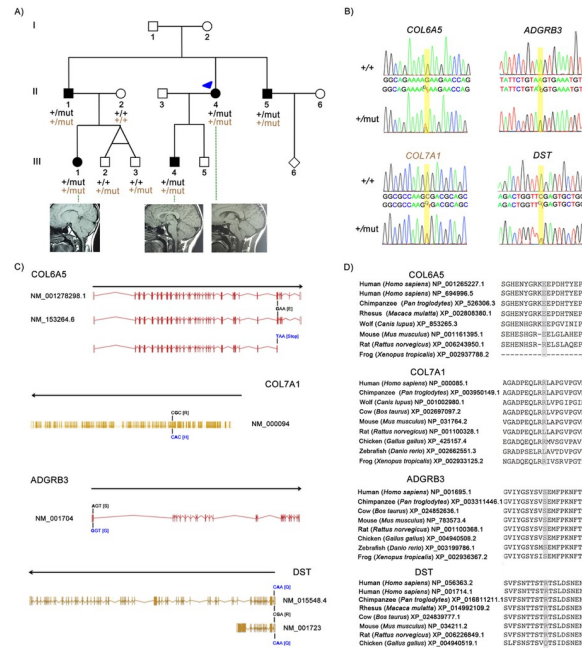


Fig 1. Mutations identified in the multiplex family with CM-1. (A) Pedigree of the Spanish family including five individuals with CM-1 carrying the heterozygous mutations E2272* in *COL6A5*, S45G in *ADGRB3*, R22Q in *DST*, in the top row; and R1202H in *COL7A1* in the bottom row. The arrow head indicates the index case (individual II.4). Healthy individuals are denoted in white and CM-1 affected in black. (B) Electropherograms showing the altered nucleotides (+/mut) and the respective wild type (+/+) sequences. (C) Location and consequence of the alteration in the human transcripts. (D) Evolutionary conservation of the E2272*, R1202H, S45G and R22Q residues in different species demonstrated by protein sequence alignment generated by MUSCLE version 3.6 (using option:-maxiters 2).

<https://doi.org/10.1371/journal.pone.0251289.g001>

protein sequence (nonsynonymous or splicing). We excluded variants with a MAF>0.01 (n = 1495), those where the functional prediction was neutral, tolerant or benign in five or more of the six prediction programs considered (n = 43), or those not exhibiting conservation across 20 different mammalian species (n = 18). Among the remaining 24 variants, seven were in pathways involved in angiogenesis (*ADGRB3*, *ADGRA2*) or the extracellular matrix (*COL6A5*, *COL7A1*, *DST*, *COL15A1*, *ITIH5*) (Table 1) and were confirmed by Sanger sequencing. Variants located in *COL6A5*, *ADGRB3* and *DST* were also present in the remaining affected sibling (individual II.5), but not in other family members who were not diagnosed with CM-1 (II.2, III.2 and III.3), suggesting cosegregation with the disease in the family (Fig 1A and 1B). The variant located in *COL7A1*, however, was present not only in the affected

Table 1. Single nucleotide variations identified in the family with five CM-1 affected members when considering a dominant inheritance model.

Position (GRCh37/hg19)	Gene	Ref	Alt	dbSNP.b141	Impact variant	PP2	Shift	Mutation Taster	ExAC ENF	1000G EUR	GoESP6500SI
chr3:48623625	<i>COL7A1</i>	C	T	rs149011081	NS	D	D	D	0.0029	0.0030	0.0030
chr3:130187662	<i>COL6A5</i>	G	T	rs115375867	Stop gained	-	-	-	0.0112	0.0010	0.0070
chr6:69348700	<i>ADGRB3</i>	A	G	rs138295002	NS	B	T	D	0.0005	0.0010	0.0010
chr6:56507522	<i>DST</i>	C	T	rs749722200	NS	D	D	-	0.0001	-	-
chr8:37693279	<i>ADGRA2</i>	G	A	rs77369926	NS	D	T	D	0.0058	0.0050	0.0030
chr9:101816909	<i>COL15A1</i>	A	G	rs35544077	NS	T	D	D	0.0088	0.0050	0.0080
chr10:7628008	<i>ITIH5</i>	G	A	-	NS	D	D	D	-	-	-

NS (nonsynonymous), D (deleterious), B (benign), T (tolerant)

<https://doi.org/10.1371/journal.pone.0251289.t001>

members but also in the twins who were not diagnosed with CM-1 (II.5, III.2, and III.3; [Fig 1A and 1B](#)), suggesting this variant could be incompletely penetrant. Variants located in *ADGRA2*, *COL15A1*, *ITIH5* were not present in the other affected sibling (individual II.5; data not shown), and thus were removed from consideration.

The variant located in *COL6A5* (chr3:130468818) is present in two isoforms (NM_001278298.1: c.6814G>T for transcript 1 and NM_153264.6: c.6814G>T for transcript 2) and leads to a stopgain: p.Glu2272* in both precursors (NP_001265227.1 and NP_694996.5) ([Fig 1C](#)). The affected amino acid position 2272 in the *COL6A5* protein is conserved across different mammalian species, but not in amphibians (frog) where the protein is shorter (2296 amino acids) ([Fig 1D](#)). The variant located in *COL7A1* (chr3:48586192): NM_000094:c.G3605A leads to an amino acid change: p.Arg1202His (NP_000085.1) ([Fig 1C](#)), producing a hydrophilicity change from hydrophilic to moderate, but not a change in charge. The affected amino acid position 1202 in the *COL7A1* protein is conserved in most mammals (except in rat), birds (such as chicken) and amphibians (frog) ([Fig 1D](#)). The missense variant in *ADGRB3* (chr6: 68638808): NM_001704:c.A133G leads to an amino acid change, p.Ser45Gly (NP_001695.1) ([Fig 1C](#)), and the hydrophilicity changes from polar to hydrophobic. This amino acid is conserved in mammals, birds (such as chicken), fish (zebrafish) and amphibians (frog) ([Fig 1D](#)). However, PolyPhen2 and Shift predicted the variant to be benign and tolerant, respectively ([Table 1](#)). The variant located in *DST* (chr 6: 56642724) impacts multiple transcripts of the gene: NM_001723:c.65G>A for transcript 1e and NM_015548.4:c.65G>A for transcript variant 1eA ([Fig 1C](#)). The variant leads to a missense mutation p.Arg22Gln in both isoforms of the protein (NP_001714.1 and NP_056363.2, respectively) and produces a loss of charge. This amino acid is conserved in some isoforms of human, chimpanzee and mouse, but not in birds (such as chicken), fish (zebrafish) or amphibians (frog) ([Fig 1D](#)). Based on these data, the functional consequence of the *COL6A5* variant (stopgain) appeared to be the most significant, but we were unable to conclusively rule out any of the other variants. Thus, we sought to evaluate these and other, related genes in an independent cohort.

Network protein-protein interaction

Since variants in two collagens (*COL6A5* and *COL7A1*) were detected in the Spanish family, we hypothesized that a common pathway of collagens (including EDS genes) could be involved in CM-1 pathogenesis. Thus, in the targeted sequencing analysis, we considered *COL6A5* and *COL7A1*, as well as *COL1A2*, *COL5A1*, *COL5A2*, *COL1A1*, *ADAMTS2*, *TNXB* and *PLOD1*. We also considered five genes that were previously described to be associated with CM-1: *ALDH1A2*, *CDX1*, *FLT1*, *MSGN1*, and *RDH10*. Those genes were input into the STRING protein network database to identify a potential protein-protein interaction (PPI) network for CM-1. Seven other genes were ultimately included (*COL3A1*, *DSE*, *HIF1A*, *NRP1*, *PGF*, *VEGFA* and *VEGFB*) to create a fully connected network. The PPI enrichment was statistically significant ($p < 1.0e-16$), meaning that these proteins have more interactions among themselves than what would be expected for a random set of proteins of similar size. Such enrichment suggests that the proteins are biologically connected. Through this STRING analysis, we determined that *ADGRB3* and *DST* were not connected to this network, even after extending the interaction connections five times. Thus, we omitted those two genes from the targeted sequencing analysis and only examined the remaining 21 ECM genes belonging to the collagen/EDS network ([Fig 2](#)).

Targeted sequencing

A total of 1833 variants were observed across the samples subjected to targeted sequencing. Two samples had low coverage (<10X) and were removed. After zeroing out variants called

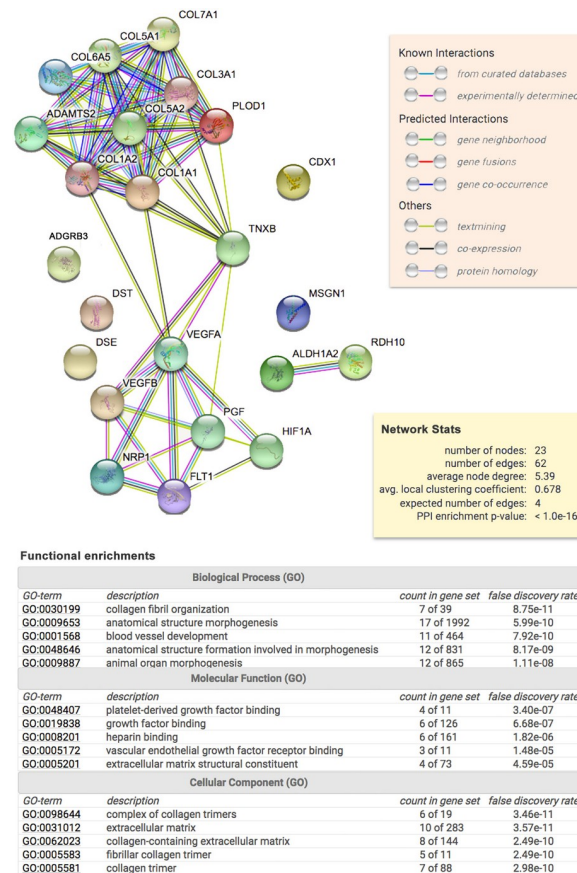


Fig 2. Protein network and enrichment analysis obtained from STRING. In the upper part, the diagram shows the result obtained upon entering a set of 23 proteins suspected to be involved in CM-1. Each node represents all the proteins produced by a single, protein-coding gene locus. Colored nodes indicate query proteins and first set of interactions. Filled node content indicates that protein 3D structure is known/predicted. Edges represent specific and meaningful protein-protein associations. In the middle, the yellow rectangular shows the reported enrichment of functional connections among the set of proteins. At the bottom, the table shows the statistical enrichment detected in functional subsystems.

<https://doi.org/10.1371/journal.pone.0251289.g002>

from fewer than 3 reads, three additional samples had missing rates > 50%, and therefore were removed from the analysis. Two samples were removed because the average number of variants detected was > 3 standard deviations from the mean, likely indicating a sample quality issue. Finally, one sample had over half of the total variants called as homozygous alternate, again, indicating a sample quality issue and thus, it was also removed. After all quality control measures, 178 CM-1 cases (average coverage of 143X) and 1345 variants (777 of which were exonic and 489 of which were functional) were included in the subsequent analyses. Of note, three of the candidate genes (*ALDH1A2*, *CDX1* and *RDH10*) were also excluded from analysis at this point because these genes did not contain variants that passed the quality control filters.

To compare rare, functional variant allele frequencies in CM-1 cases and controls (publicly available from the gnomAD database), we restricted the analysis to only missense variants with $MAF < 0.05$. We identified six genes (*COL7A1*, *COL5A2*, *COL6A5*, *COL1A2*, *VEGFB* and *FLT1*) with a higher burden of rare, functional variants in CM-1 cases compared to controls (p 's < 0.01; Tables 2 and S3) with odds ratios ranging from 1.8 (*COL5A2*) to 8.3 (*COL1A2*) (Table 2). STRING (PPI) enrichment showed that these proteins are at least partially biologically connected ($p = 3.49e-09$). All of them present extracellular regions (Fig 3A), functioning

Table 2. Candidate gene burden tests comparing the number of variants in CM-1 cases with public controls from gnomAD.

Candidate Gene	Number of SNPs passing filtering	Fisher exact 1-sided p-value	Odds Ratio (OR)	Lower Confidence Interval of OR	Upper Confidence Interval of OR
<i>PLOD1</i>	2	0.8149	0.794	0.265	1.785
<i>MSGN1</i>	1	+			
<i>COL3A1</i>	6	0.2935	1.362	0.633	2.519
<i>COL5A2</i>	12	0.0001	1.814	1.327	2.426
<i>COL7A1</i>	14	0.0006	2.424	1.501	3.69
<i>COL6A5</i>	23	<0.0001	1.99	1.463	2.648
<i>CDX1</i>	0	*			
<i>ADAMTS2</i>	5	0.2719	1.198	0.751	1.807
<i>TNXB</i>	48	0.9999	0.228	0.167	0.306
<i>VEGFA</i>	1	+			
<i>DSE</i>	4	0.8806	0.757	0.383	1.326
<i>COL1A2</i>	3	0.0095	8.273	2.303	20.678
<i>RDH10</i>	0	*			
<i>COL5A1</i>	7	0.0934	1.399	0.895	2.078
<i>NRP1</i>	1	+			
<i>VEGFB</i>	3	0.0036	7.436	2.467	16.884
<i>FLT1</i>	8	0.001	3.354	1.752	5.766
<i>HIF1A</i>	4	0.3037	1.509	0.568	3.178
<i>PGF</i>	1	+			
<i>ALDH1A2</i>	0	*			
<i>COL1A1</i>	4	0.0709	2.532	0.953	5.339

Significant p-values are indicated in bold.

+These genes only contained a single variant that surpassed our quality control filters and thus, were dropped from the burden analysis.

*These genes did not contain variants that surpassed our quality control filters and thus, were dropped from analysis.

<https://doi.org/10.1371/journal.pone.0251289.t002>

in growth factor binding (*FLT1* and *COL1A2*), SMAD binding (*COL5A2* and *COL1A2*) or extracellular matrix structural constituent (*COL6A5*, *COL7A1* and *COL1A2*) (**Fig 3B**). The

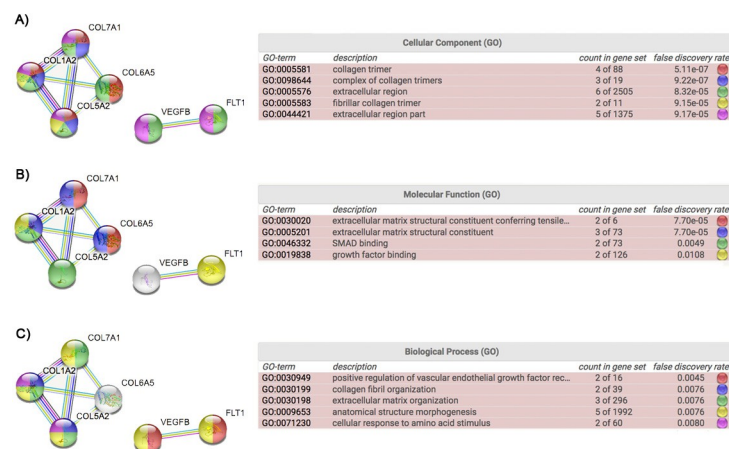


Fig 3. Protein enrichment analysis obtained from STRING for the six genes with significantly more rare variants in CM-1 compared with controls: *COL5A2*, *COL6A5*, *COL1A2*, *COL7A1*, *VEGFB* and *FLT1*. (A), (B) and (C) show the results obtained from the cellular component, molecular function and biological process enrichment, respectively.

<https://doi.org/10.1371/journal.pone.0251289.g003>

gene ontology (GO) biological process enrichment indicated that these proteins play a role in the vasculature (FLT1 and VEGFB), collagen fibril organization (COL1A2 and COL5A2), extracellular matrix organization (COL1A2, COL5A2 and COL7A1) and anatomical structure morphogenesis (FLT1, VEGFB, COL1A2, COL5A2 and COL7A1) (Fig 3C). Interestingly, 46.63% of CM-1 cases presented with rare variants in at least one of the four significant collagen genes and 9.55% harbor variants in multiple significant collagen genes. Moreover, 25.8% of CM-1 cases presented rare variants in *COL6A5* (S5 Table). These data suggest that the impact of these specific collagen genes on risk for CM-1 may be more universal compared to other genes.

When testing for an excess burden of functional variants between CTD+ compared to CTD- CM-1 cases, we identified two genes (*COL7A1*, *COL3A1*) where the burden of rare variants was significantly different between the CM-1 cases as a function of CTD status (*COL7A1* $p = 0.0081$ *COL3A1* $p = 0.0193$; Tables 3 and S3). CTD+ CM-1 cases had a significantly higher percentage of rare, functional variants in *COL7A1* and a significantly lower percentage of rare, functional variants in *COL3A1*. These associations remained significant when CM-1 patients who had a formal EDS diagnosis were excluded from the analysis (*COL7A1* $p = 0.0025$; *COL3A1* $p = 0.0385$; Table 3).

Discussion

In the present study, we provide evidence for an important role of collagens and other members of the extracellular matrix (ECM) in the pathogenesis of CM-1. We identified four

Table 3. Gene burden tests comparing CTD+ versus CTD- individuals.

Candidate Gene	N SNPs (N SNPs after removing EDS cases)	SKAT p-value	SKAT p-value excluding EDS cases
<i>PLOD1</i>	1 (1)	0.6325	0.5892
<i>MSGN1</i>	5 (5)	0.6314	0.6173
<i>COL3A1</i>	6 (6)	0.0193	0.0385
<i>COL5A2</i>	11 (10)	0.5262	0.5998
<i>COL7A1</i>	13 (13)	0.0081	0.0010
<i>COL6A5</i>	22 (21)	0.9536	0.9572
<i>CDX1</i>	0	*	*
<i>ADAMTS2</i>	5 (4)	0.6949	0.7651
<i>TNXB</i>	40 (38)	0.2784	0.2964
<i>VEGFA</i>	1 (1)	0.7303	0.7147
<i>DSE</i>	3 (3)	0.4226	0.4187
<i>COL1A2</i>	9 (9)	0.6671	0.7511
<i>RDH10</i>	0	*	*
<i>COL5A1</i>	11 (11)	0.4796	0.4058
<i>NRP1</i>	3 (3)	0.7976	0.7886
<i>VEGFB</i>	6 (4)	0.2104	0.2553
<i>FLT1</i>	9 (9)	0.1223	0.1337
<i>HIF1A</i>	4 (4)	0.5227	0.4971
<i>PGF</i>	1 (1)	0.2303	0.2147
<i>ALDH1A2</i>	0	*	*
<i>COL1A1</i>	13 (13)	0.4138	0.4937

Significant p-values are indicated in bold.

*These genes did not contain variants that surpassed our quality control filters and thus, were dropped from analysis.

<https://doi.org/10.1371/journal.pone.0251289.t003>

variants in genes involved in the ECM (*COL6A5*, *COL7A1*), angiogenesis (*ADGRB3*) and adhesion junctions (*DST*) that cosegregated with CM-1 in a large Spanish family. Follow-up analysis of ECM genes in an independent cohort (186 CM-1 patients) identified an enrichment of rare functional variants in *COL7A1*, *COL5A2*, *COL6A5*, *COL1A2*, *VEGFB* and *FLT1* in CM-1 cases compared with controls. In our CM-1 cohorts, nearly half of unrelated CM-1 cases harbored rare variants in at least one of the significant collagen genes (*COL7A1*, *COL5A2*, *COL6A5*, *COL1A2*), suggesting that these genes may be particularly important for risk for CM-1. We also observed association between the presence of CTD symptoms among CM-1 patients with the occurrence of rare variants in *COL7A1* and *COL3A1*.

A role for genetic variation in collagens, and more broadly components of the ECM, in risk for CM-1 is perhaps not unexpected. Bone and ligaments are derived from connective tissue which is composed of the ECM (consisting of protein fibers such as collagens) and a variety of support cells. In particular, ligaments are composed of collagen (75% of the dry weight) and non-collagenous proteins (such as proteoglycans, glycoproteins and elastin) that are synthesized by fibroblasts. In bones, the connective tissue is highly vascularized and contains a specialized ECM mainly composed of collagen types I, III, and V that are synthesized by osteoblasts. The importance of collagens and other ECM genes in risk for EDS, a group of CTDs frequently comorbid with CM-1, is well known [6]. But there is also evidence supporting a role in CM-1 pathogenesis more directly. *COL2A1*, although not examined in this study, was found to be significantly upregulated in CM-1 patients with specific cranial morphometrics [16]. Moreover, the relationship between collagen and retinoic acid exposure has been observed in chondrocyte and osteoblast differentiation [20, 21] and over-expression of vitamin A can reduce bone mineral density [22] and even produce a Chiari-like phenotype [2]. Two SNPs (rs2899611 and rs6493979) of the *ALDH1A2* gene (an enzyme that catalyzes the synthesis of retinoic acid) have been previously associated with risk for CM-1 [15]. Thus, there is supportive evidence implicating collagens and associated pathways in CM-1. However, to our knowledge, this is the first report associating rare genetic variants in collagens and other ECM genes with the occurrence of CM-1.

Many of the individual genes associated with CM-1 in our study have known biologic roles that could impact CM-1 risk. Genetic variation in *COL6A5* influences bone mineral density in both mouse and human [23]. *COL7A1* is typically known for its association with epidermolysis bullosa, a blistering skin condition caused by abnormal collagen organization [24, 25]. However, *COL7A1* is also downregulated in articular cartilage and subcondral bone in the setting of osteoarthritis [26]. *COL5A2* and *COL1A2* can participate in binding SMAD, a protein that is the main signal transducer for receptors of TGF- β . While SMAD has not been directly associated with CM-1, *TGFBR2* and *GDF6* are two other members of the TGF- β gene family that have been previously associated with CM-1 [12, 17]. Mutations in *COL3A1* have been associated with a vascular form of EDS [6], primarily characterized by risk for early death due to the rupture of major arteries [27]. *VEGFB* and *FLT1* are vascular genes, and it is known that vascularization and collagen is important for the bone mineralization process [28–30]. Vasculature is thought to serve as a guide for deposition of the mineralized bone matrix (osteoclast-secreted collagen type I), which helps define bone shape during embryonic development [29]. Alteration in the interactions between blood vessels and bone cells can lead to skeletal diseases such as craniofacial dysmorphism or idiopathic osteonecrosis [31].

Collagens in co-occurring diseases

Classical CM-1 occurs because of a smaller and shallower posterior cranial fossa (PCF), as a consequence of a change in the length and slope of the basilar part of the occipital bone (basiocciput) [5, 32, 33]. However, other mechanisms can also lead to a small PCF, such as craniosynostosis

(Crouzon's syndrome or Apert syndrome). Craniosynostosis syndromes are frequently associated with mutations in fibroblast growth factor receptors (*FGFR2* or *FGFR3*) which play an important role in bone growth, particularly during embryonic development and are complexed with ECM components. Other conditions that can be comorbid with CM-1, such as osteogenesis imperfecta (OI) or scoliosis, are associated with alterations in the ECM [8]. OI is a genetic disorder predominantly affecting bone, and is characterized by bone fragility and abnormalities in skeletal, teeth, skin, and soft tissues, as well as joint hyperlaxity. OI is caused by alterations in collagen I, either as a consequence of haploinsufficiency of the genes (*COL1A1* or *COL1A2*) or due to genetic mutations in other genes related to collagen I biosynthesis and organization. Scoliosis, frequently associated with CM-1 [34, 35], is an abnormal curving of the spine due to connective tissue alterations and some forms of scoliosis are associated with abnormalities in collagens [36, 37].

Since each collagen is generally expressed in several different tissues, and is tightly associated with other ECM components, alterations in the collagen genes can result in widely overlapping clinical features [38]. For example, mutations in *COL1A2* can give rise to OI, EDS type VIIB, recessive EDS Classical type, idiopathic osteoporosis, and atypical Marfan syndrome. Mutations in *COL2A1* are associated with a wide range of clinical conditions including severe Ulrich congenital muscular dystrophy, hypochondrogenesis, and different forms of EDS to name a few [28]. Additional work will be needed to fully understand why mutations in the same gene result in different clinical presentations, but we hypothesize it is caused by allelic heterogeneity or by an incomplete penetrance resulting from compound inheritance of both, rare and common variants, at single or multiple loci as has been demonstrated for non-syndromic craniosynostosis and congenital scoliosis [39, 40].

Study limitations and future work

CM-1 is highly clinically heterogeneous. In an attempt to reduce that heterogeneity, we took two different approaches. The first was to restrict to a more homogeneous demographic group of patients. The second was to subsequently stratify that group on the presence or absence of CTDs. However, there are many other ways to address the clinical heterogeneity in this patient population [8] and those may have yielded different findings than what we identified here.

With respect to the demographics of this population, we focused primarily on adult females of European descent. The Spanish family used in the exome sequencing study included pediatric and male subjects, but the unrelated CM-1 patients examined in the targeted sequencing analysis were exclusively Caucasian, adult females. CM-1 is more frequent in females, and thus, this restriction allowed us to achieve a reasonable sample size without concern for potential gender differences. The fact that we observed rare functional variants in *COL6A5* and *COL7A1* in both the Spanish family and the unrelated CM-1 cases suggests that these loci may not be unique to females with CM-1. It is possible that other mechanisms may be driving the gender differences. Recently Houston *et al.* observed a shorter PCF in males compared to females [41], underscoring the effect of gender on CM-1 traits. More work will be needed to disentangle the mechanisms leading to the gender differences. We also restricted the unrelated CM-1 cohort to adults, since it has been reported that the age of onset of CM-1 symptoms (pediatric versus adult) can produce different forms of the disease [42–44]. It is also important to note that the unrelated CM-1 cases utilized in the targeted sequencing were from the US, while the whole exome sequencing family was from Spain. However, both data sets have European origin, and variants in *COL6A5* and *COL7A1* were identified in both.

Although the evaluation of rare variants as a function of CTD status had considerably less statistical power than did the analysis of all the CM-1 cases versus public controls, we did observe evidence for some genes having a higher prevalence of rare variants as a function of CTD status.

The association of *COL7A1* variants in the CTD+ subset is consistent with a role for collagens in connective tissue as described above. The association of *COL3A1* in the CTD- subset is more surprising. However, mutations in *COL3A1* have been typically associated with the vascular subtype of EDS [6]. Thus, it is possible that these individuals may not have received a diagnosis of EDS as a result of a more predominant vascular phenotype rather than joint involvement. More investigation will be needed to fully evaluate this association, including examination of a larger cohort of CM-1 patients. Nonetheless, the observation of different genetic etiologies as a result of CTD status among CM-1 patients is consistent with previous findings [12].

Our analytic approach for this study was to focus on rare variants in the coding regions of the genes, based on the assumption that rare, highly penetrant changes in these regions would produce an altered protein, resulting in CM-1. However, alterations in introns, regulatory regions, and long noncoding RNAs can also cause clinical presentations since these types of variants often impact gene expression. Future genetic studies in CM-1 should ideally focus on whole genome sequencing to ensure that these other types of mutations are fully evaluated in CM-1 etiology. We also focused the targeted sequencing analysis on genes belonging to a common biologic pathway. However, it is important to note that while *DST*, which was identified in the Spanish multiplex family, does not belong to the biologic pathway, there is recent work implicating this gene in a patient presenting with syringomyelia [45], a condition often comorbid with CM-1. Similarly, *ADGRB3*, also does not belong to the pathway that was sequenced in the US cohort, but it has been associated with intellectual disability and cerebellar atrophy [46]. Thus, we cannot exclude the possibility that the *DST* and *ADGRB3* variants contribute to CM-1 pathogenesis in the Spanish family. The finding of multiple variants in reasonable candidate genes in a single multiplex family is consistent with CM-1 being a complex genetic disorder, with possibly several variants contributing to the phenotype even within the same family. Taking a multi-omics approach to understanding CM-1 pathology may help disentangle the multiple candidate genes. Previous work has demonstrated evidence for genomic differences contributing to CM-1 subtypes [15, 16]. Thus, a more comprehensive analysis of genomic variation in CM-1 patients could shed light on relevant candidate genes and mechanistic processes.

The clinical data for the unrelated CM-1 patients in the targeted sequencing analysis was obtained from a variety of sources, including some self-report data. For the diagnosis of CM-1, medical record data were obtained, and MRIs, in the majority of cases, were available to confirm the diagnosis. Thus, misdiagnosis of CM-1 is not likely in this analysis. In contrast, the assignment of CTD status was rarely accompanied by a physician diagnosis and was primarily generated by self-report. As a result, the CTD classification used in this study has less precision. Nonetheless, that limitation would only affect the CTD status analysis, not the primary comparison of CM-1 cases to controls.

Future work should evaluate the effect of genetic variation on cranial morphometrics. Several studies have previously shown that CM-1 patients have unique cranial morphology [5], and also that genetic variation is associated with cranial morphology [15–17]. Cranial morphology is a quantitative phenotype which should yield more power to detect associations with genetic variants, and may also help distinguish different CM-1 subtypes. In addition, while this study has provided statistical evidence of association between rare functional variants and risk for CM-1, careful molecular functional analysis will be needed to fully define the manner in which the variants in these genes impact risk for CM-1.

Conclusions

Rare functional variants within collagens and other related ECM genes are a risk factor for CM-1. CTD status may impact which collagen genes are involved, and thus may be a hallmark

not only for clinical heterogeneity in CM-1 but also genetic heterogeneity. Follow-up studies in other CM-1 patients is warranted, but these findings may ultimately provide an opportunity for improved diagnosis and clinical management strategies in CM-1 patients.

Methods

Ethics statement

The present investigation was conducted according to the principles expressed in the Declaration of Helsinki and was approved by the Ethics Committee of the Vall d'Hebron University Hospital (Spain) and the institutional review boards of the University of Akron (OH, US) and Duke University School of Medicine (NC, US). All participants provided informed consent prior to their participation in the study.

Subjects

Spanish multiplex family. A female, 30 years old Caucasian patient (index case, individual II.4) visited the Vall d'Hebron University Hospital (Spain). Due to the patient's symptoms (cough headaches, neck pain, paresthesia in the upper and lower limbs, sensory loss, and impaired temperature sensation in the thorax, right arm and in both lower limbs), a cranial and complete spinal cord MRI was performed, and the neurosurgeon (MAP) diagnosed her with CM-1. The patient had 10mm of tonsillar herniation, retrocurved odontoid and cervical syringomyelia. Posterior fossa reconstruction was performed (extensive suboccipital craniectomy with a wide opening of the foramen magnum and C1 laminectomy; the dura mater was opened and closed watertight using a wide dural lyophilized allograft [Lyodura-S, Braun-Melsungen AG, Germany], wide enough to allow a spacious posterior fossa to be reconstructed with good results (new artificial cisterna magna formed with upward migration of the cerebellum and reduction of the syrinx [47]). Clinical interview and cranial MRI demonstrated CM-1 in four additional family members (Fig 1). The main clinical characteristics of the other affected members of the family included:

- **Individual II.1:** male, 38 years old. CM-1 (tonsillar herniation below the foramen magnum of 3 mm) without syringomyelia. Headaches.
- **Individual II.5:** male, 27 years old. Retrocerebellar arachnoid cyst, initially asymptomatic. Two years later, the patient reported symptoms that increased with stress: hemicranial headaches, visual alterations, and abnormal oral perceptions, which were attributed to migraine.
- **Individual III.1:** female, 9 years old, daughter of individual II.1. Frequent headaches. CM-1 with tonsillar herniation of 6 mm without syringomyelia.
- **Individual III.4:** male, 8 years old, son of individual II.4. CM-1 (tonsillar herniation below the foramen magnum of 4 mm) without syringomyelia.

All individuals studied had a normal ventricular size (Evans Index < 0.30) [48].

US caucasian cohort. Targeted candidate gene sequencing was performed in an independent cohort comprised of 186 unrelated adult American Caucasian females. Participants were recruited at two different sites: Ninety-four were recruited through the Chiari1000 project (<https://chiari1000.uakron.edu>) and ninety-two were selected from a larger family study of CM-1 ascertained at Duke University Medical Center [12]. Chiari1000 project participants shared pre-surgery MRI data, and those who had indicated interest in participating in future studies were contacted by email and invited to participate in the genetic study. Saliva samples were collected through the mail and extracted at Duke University Medical Center. Participants

in Chiari1000 who ultimately provided a saliva sample were almost exclusively female and all Caucasian. As a result, only unrelated Caucasian females were selected from the Duke cohort in an effort to avoid potential confounding of racial, ethnic, and gender differences by site. Participants from the Duke family study were ascertained as part of a larger study to investigate the genetic basis of CM-1. Like the Chiari1000 subjects, Duke subjects were recruited primarily through self-referral across the US and provided medical records, including MRIs, which were used to confirm the diagnosis. Because Duke subjects were recruited primarily for the purpose of a genetic study, the majority had a positive family history of CM-1.

Clinical features of the Chiari1000 and Duke participants are described in Table 4. Among the 186 participants, 176 were diagnosed with CM-1 and ten were diagnosed with CM-1 and EDS. All 186 women presented with a minimum of 3mm cerebellar tonsillar herniation on MRI. Exclusion criteria included the following: 1) individuals thought to have a secondary form of CM-1 (e.g. due to a brain tumor, tethered cord syndrome, hydrocephalus, atlantoaxial assimilation, craniosynostosis or pseudotumor cerebri), 2) participants with another known genetic syndrome other than EDS (e.g. Marfan syndrome, Klippel-Feil syndrome, Neurofibromatosis, Growth hormone deficiency, abnormal pituitary, hyperostosis), and/or 3) participants that were not of non-Hispanic Caucasian ethnicity.

Previous work has demonstrated that, even in the absence of a formal CTD diagnosis, stratifying study participants on the presence (CTD+) or absence (CTD-) of CTD associated symptoms can improve the ability to detect a genetic signal [12]. Thus, we assigned CTD status based on a positive family history of CTD associated symptoms including self-reported hypermobility, a Beighton hypermobility score of 6 or greater [49], kyphosis, aneurysm, mitral valve prolapse, pectus excavatum, scoliosis, orthostatic hypotension, supraventricular tachycardia, heart valve disease, and/or heart murmur [12]. Study participants with a family history of at least one of these characteristics were classified as CTD+; all other participants were classified as CTD-.

Brain Magnetic Resonance Imaging (MRI) protocol

MRI data for the Spanish family were acquired at Vall d'Hebron University Hospital using a 1.5 T scanner (MAGNETOM Symphony or MAGNETOM Vision, Siemens, Erlangen, Germany) equipped with a circular polarized receiver head array coil. Sagittal conventional spin-echo T1-weighted sequences were obtained (TR 450–600 msec, TE 12–20 msec, acquisitions 2). All sequences were obtained with 4–5-mm section thickness and 0.1–0.3 interslice gap, with 144–256 × 256–384 matrix, and 196 × 230 mm field of view. MRI images were voluntarily provided by the participants recruited through the Chiari1000 project and the Duke family study, details about which have been described previously [12, 41]. In this analysis, the MRI's were used to confirm the CM-1 diagnosis.

Table 4. Clinical characteristics of Chiari1000 and Duke subjects.

	Chiari 1000 (N = 94)		Duke (N = 92)	
	Number	%	Number	%
CTD+	28	29.79	53	57.61
CTD-	66	70.21	27	29.35
Unknown CTD			12	13.04
EDS-	89	94.68	87	94.57
EDS+	5	5.32	5	5.43
Syringomyelia	14	14.89	22	23.91
No syringomyelia	80	85.11	70	76.09

<https://doi.org/10.1371/journal.pone.0251289.t004>

DNA isolation and quantification

Genomic DNA from the Spanish family members was isolated from peripheral blood lymphocytes by the salting-out method [50]. Genomic DNA from the Chiari1000 participants was isolated from saliva using the Oragene DNA Self-Collection kit according to the manufacturer's recommendations (DNA Genotek Inc. Ottawa, Ontario, Canada). For patients from the Duke cohort, genomic DNA was extracted from peripheral blood samples using Puregene reagents (Qiagen, Valencia, California). DNA quality was assessed by running a small amount of DNA on a 0.8% agarose gel. DNA concentrations were determined on a NanoDrop spectrophotometer (NanoDrop Technologies, LLC, Wilmington, DE) or using Quant-iT PicoGreen dsDNA Reagent (ThermoFisher Scientific, Foster City, CA).

Design, library preparation and next generation sequencing

Whole exome sequencing of Spanish multiplex family. A Covaris ultrasonicator (Covaris, Woburn, MA, USA) was used to shear 1 μ g of genomic DNA from individuals II.1, II.4, III.1 and III.4 resulting in fragments 300-400bp in size. Library generation was performed using the TruSeq DNA kit (Illumina, San Diego, CA, USA) following manufacturer's instructions. Pools of six samples were subjected to whole exome capture applying Nimblegen Seq-Cap EZ Human Exome v3.0 (Roche NimbleGen, Madison, WI, USA) followed by sequencing on a HiSeq 2500 using the 100bp paired end protocol. The sequencing was performed at the Centre for Genomic Regulation in Barcelona (Spain). We obtained a coverage of at least 30X in > 80% of the enriched regions per sample.

Targeted sequencing of CM-1 cases. Amplicon targets were designed using DesignStudio, a web-based software tool (Illumina, San Diego, CA, USA). Initially, 23 genes were prioritized: those identified from whole exome sequencing analysis of the Spanish family, genes previously associated with CM-1 [15] or Ehlers Danlos Syndrome (EDS) (<https://ghr.nlm.nih.gov/condition/ehlers-danlos-syndrome#genes>) and genes in pathways connected to those genes as identified using the STRING network database [51]. STRING was used to identify a potential protein-protein interaction (PPI) network among the candidate genes for the purpose of focusing our analysis on a defined molecular pathway. In total, 1144 amplicons were designed to capture the exonic regions of 21 genes. Libraries were prepared using the TruSeq Custom Amplicon Library Prep kit manufactured by Illumina, following the manufacturer's recommendations. Sequencing of the libraries was performed on an Illumina MiSeq v2 using 150bp paired end reads by the Duke Center for Genomic and Computational Biology (GCB) at Duke University Medical Center.

Whole exome sequencing

Alignment and variant detection. Sequencing data preparation and read alignment were performed according to Genome Analysis Toolkit (GATK) Best Practice recommendations provided by Broad Institute. In brief, sequencing reads were aligned to the hg19 reference genome using BWA-MEM [52]. Variants were called using GATK HaplotypeCaller (version 3.3) [53]. Initial quality control was performed using default thresholds in GATK and subsequently variants failing to meet those criteria were removed. Additionally, calls based on fewer than three reads (low coverage) and heterozygous calls with less than 30% of reads attributable to each allele (allele balance) were zeroed out. Only biallelic variants were considered for variant annotation.

Variant annotation and prioritization. Variants were annotated, filtered and prioritized using both functional annotations (population allele frequency, impact on protein function and conserved elements) and pedigree information. ANNOVAR [54] was used to annotate the

variants, including: i) information on population allele frequencies based on: 1000 genomes 2015aug (Eur) [55], ExAC version 0.3 (NFE) [56] and esp6500siv2 (Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA; URL: <http://evs.gs.washington.edu/EVS/>; date accessed: February, 2017)); ii) dbSNP version 147 with left normalization [Database of Single Nucleotide Polymorphisms (dbSNP). Bethesda (MD): National Center for Biotechnology Information, National Library of Medicine. (dbSNP Build ID: build 147)]; iii) dataset functional prediction dbnsfp30a which includes PolyPhen-2 [57], SIFT [58], Mutation-Taster [59], LTR [60], Mutation Assessor [59] and FATHMM [61]; and iv) information about evolutionarily conserved elements in mammalian species (phastCons20way mammalian) [62].

As the inheritance model in the pedigree appeared to be dominant, we performed variant prioritization based on this mode of inheritance and focused on exonic (non-synonymous and splicing) variants. In order to reduce the candidate variant list, we prioritized variants that had not been described previously (ie. were novel), presented with a minor allele frequency (MAF) <0.01, were predicted to be deleterious in at least two of the six functional prediction programs, and had a conservation site score over 0.9. Lastly, we checked the genotypes and the proportion of the reads for each individual using the Integrative Genomics Viewer (IGV) [63, 64] for the candidate variants in order to remove potential false positives, such as variants where the allele with less coverage had less than 30% of the reads at that locus. To ensure there were no sequencing errors, we confirmed the variants passing these filters using Sanger sequencing.

Targeted sequencing

Raw sequencing reads from 21 targeted genes were processed identically to those from the Spanish multiplex family. We applied the same GATK quality filters, zeroed out low coverage calls and heterozygous calls with allele balance < 30%, however, we did not exclude variants based on MAF, predicted function, or conservation score. Although not used as a filter, variants were annotated using ANNOVAR to classify variants based on MAF thresholds and predicted function. Genotypes for rare (MAF<1%), functional exonic (nonsynonymous or splicing) variants were confirmed via Sanger sequencing.

Sanger confirmation

Direct sequencing was used to confirm high priority candidate variants identified through WES and targeted sequencing. WES candidate variants were sequenced for all family members in order to evaluate co-segregation with disease. For the targeted sequencing, we confirmed variants in the genes that were significant in the burden tests, described below, in an effort to remove any false positives. Primers used were designed with the identified variant genomic target coordinates using Primer3 [65] generating amplicons of 200-600bp in length. The sequences are detailed in S1 and S2 Tables. Amplicons were amplified by PCR (S3 and S4 Tables show the reaction conditions) and purified and sequenced either by Genewiz (Research Triangle Park, NC) for the targeted sequencing or the High Technology Unit (UAT) at Vall d'Hebron Institut de Recerca for the Spanish family. All sequences were analyzed using Sequencher v 5.4.5 (GeneTools, Ann Arbor, MI).

Statistical analysis of targeted sequencing

Because we only sequenced CM-1 cases, we utilized controls from the Genome Aggregation Database (gnomAD; <https://gnomad.broadinstitute.org/>), restricting the gnomAD controls to those of non-Finnish European descent to more closely match the ethnic background of our CM-1 cases. Allele counts were compiled for CM-1 cases and non-Finnish European controls

for each variant and were summed by gene to obtain a gene-based burden of risk alleles. Importantly, we required that the variant be present in the gnomAD database to be considered for further analysis. This was to avoid potential false positive associations that could be driven by variants detected only in the CM-1 cases. Chi-square tests were used to test for an increased burden of risk alleles (1-tailed test) in the CM-1 cases compared to controls. If cell counts were less than five, a Fisher's exact test was employed. The gene-burden tests were performed with PROC FREQ in SAS version 9.4 (SAS Institute, Cary, NC) and were restricted to functional variants with $MAF < 5\%$. To test for gene-based differences in allele frequencies between CTD+ and CTD- CM-1 cases, we used SKAT-O [66], which combines a standard gene-burden test with the sequence kernel association test (SKAT) to maximize power. Comparisons by CTD status were based on variant functional annotation (only functional variants) and allele frequency threshold ($MAF < 5\%$) to be consistent with the case-control analysis.

Supporting information

S1 Table. Primers details used to confirm candidate variants identified in WES.

(DOCX)

S2 Table. Primers details used to confirm candidate variants identified in and targeted sequencing.

(DOCX)

S3 Table. PCR conditions used to validate variants identified in whole exome sequencing of Spanish family.

(DOCX)

S4 Table. PCR conditions used to validate variants identified in targeted sequencing of Duke and Chiari1000 cohorts.

(DOCX)

S5 Table. SNPs identified from targeted sequencing that contributed to gene burden tests.

The listed variants were either a) significantly different between CM-1 cases and public controls or b) significantly different between CTD+ and CTD- CM-1 cases.

(DOCX)

Acknowledgments

We would like to thank all patients and their relatives for participating in this study.

Author Contributions

Conceptualization: Aintzane Urbizu, Melanie E. Garrett, Olga Mestres i Soler, Allison Ashley-Koch.

Data curation: Aintzane Urbizu, Melanie E. Garrett.

Formal analysis: Aintzane Urbizu, Melanie E. Garrett.

Funding acquisition: Aintzane Urbizu, Francis Loth, Allison Ashley-Koch.

Investigation: Aintzane Urbizu, Melanie E. Garrett, Karen Soldano, Oliver Drechsel, Allison Ashley-Koch.

Methodology: Aintzane Urbizu, Melanie E. Garrett, Karen Soldano, Stephan Ossowski, Alfons Macaya, Allison Ashley-Koch.

Project administration: Aintzane Urbizu, Allison Ashley-Koch.

Resources: Dorothy Loth, Maria A. Poca, Stephan Ossowski, Alfons Macaya, Francis Loth, Allison Ashley-Koch.

Software: Melanie E. Garrett.

Supervision: Allison Ashley-Koch.

Validation: Aintzane Urbizu, Melanie E. Garrett, Karen Soldano, Anna Marcé-Grau.

Visualization: Aintzane Urbizu, Melanie E. Garrett, Karen Soldano.

Writing – original draft: Aintzane Urbizu, Melanie E. Garrett, Allison Ashley-Koch.

Writing – review & editing: Aintzane Urbizu, Melanie E. Garrett, Karen Soldano, Oliver Drechsel, Dorothy Loth, Anna Marcé-Grau, Maria A. Poca, Stephan Ossowski, Alfons Macaya, Francis Loth, Rick Labuda, Allison Ashley-Koch.

References

1. Barkovich A. J., Wippold F. J., Sherman J. L. & Citrin C. M. Significance of cerebellar tonsillar position on MR. *AJNR Am J Neuroradiol* 7, 795–799 (1986). PMID: [3096099](#)
2. Marin-Padilla M. & Marin-Padilla T. M. Morphogenesis of experimentally induced Arnold—Chiari malformation. *J Neurol Sci* 50, 29–55 (1981). [https://doi.org/10.1016/0022-510x\(81\)90040-x](https://doi.org/10.1016/0022-510x(81)90040-x) PMID: [7229658](#)
3. Milhorat T. H., Nishikawa M., Kula R. W. & Dlugacz Y. D. Mechanisms of cerebellar tonsil herniation in patients with Chiari malformations as guide to clinical management. *Acta Neurochir (Wien)* 152, 1117–1127, <https://doi.org/10.1007/s00701-010-0636-3> (2010). PMID: [20440631](#)
4. Milhorat T. H., Bolognese P. A., Nishikawa M., McDonnell N. B. & Francomano C. A. Syndrome of occipitoatlantoaxial hypermobility, cranial settling, and chiari malformation type I in patients with hereditary disorders of connective tissue. *J Neurosurg Spine* 7, 601–609, <https://doi.org/10.3171/SPI-07/12/601> (2007). PMID: [18074684](#)
5. Eppelheimer M. S. et al. A Retrospective 2D Morphometric Analysis of Adult Female Chiari Type I Patients with Commonly Reported and Related Conditions. *Front Neuroanat* 12, 2, <https://doi.org/10.3389/fnana.2018.00002> (2018). PMID: [29403363](#)
6. Malfait F. et al. The 2017 international classification of the Ehlers-Danlos syndromes. *Am J Med Genet C Semin Med Genet* 175, 8–26, <https://doi.org/10.1002/ajmg.c.31552> (2017). PMID: [28306229](#)
7. Speer M. C. et al. Chiari Type I Malformation With or Without Syringomyelia: Prevalence and Genetics. *Journal of Genetic Counseling* 12, 297–311 (2003). <https://doi.org/10.1023/A:1023948921381> PMID: [26141174](#)
8. Urbizu A., Khan T. N. & Ashley-Koch A. E. Genetic dissection of Chiari malformation type 1 using endophenotypes and stratification. *Journal of Rare Disease Research and Treatment* 2, 35–42 (2017).
9. Milhorat T. H. et al. Chiari I malformation redefined: clinical and radiographic findings for 364 symptomatic patients. *Neurosurgery* 44, 1005–1017 (1999). <https://doi.org/10.1097/00006123-199905000-00042> PMID: [10232534](#)
10. Speer M. C. et al. A genetic hypothesis for Chiari I malformation with or without syringomyelia. *Neurosurgical focus* 8, E12 (2000). <https://doi.org/10.3171/foc.2000.8.3.12> PMID: [16676924](#)
11. Boyles A. L. et al. Phenotypic definition of Chiari type I malformation coupled with high-density SNP genome screen shows significant evidence for linkage to regions on chromosomes 9 and 15. *Am J Med Genet A* 140, 2776–2785, <https://doi.org/10.1002/ajmg.a.31546> (2006). PMID: [17103432](#)
12. Markunas C. A. et al. Stratified whole genome linkage analysis of Chiari type I malformation implicates known Klippel-Feil syndrome genes as putative disease candidates. *PLoS One* 8, e61521, <https://doi.org/10.1371/journal.pone.0061521> (2013). PMID: [23620759](#)
13. Abbott D., Brockmeyer D., Neklason D. W., Teerlink C. & Cannon-Albright L. A. Population-based description of familial clustering of Chiari malformation Type I. *J Neurosurg* 128, 460–465, <https://doi.org/10.3171/2016.9.JNS161274> (2018). PMID: [28156254](#)
14. Musolf A. M. et al. Small posterior fossa in Chiari I malformation affected families is significantly linked to 1q43-44 and 12q23-24.11 using whole exome sequencing. *Eur J Hum Genet*, <https://doi.org/10.1038/s41431-019-0457-7> (2019). PMID: [31227808](#)

15. Urbizu A. et al. Chiari malformation type I: a case-control association study of 58 developmental genes. *PLoS One* 8, e57241, <https://doi.org/10.1371/journal.pone.0057241> (2013). PMID: 23437350
16. Markunas C. A. et al. Identification of Chiari Type I Malformation subtypes using whole genome expression profiles and cranial base morphometrics. *BMC Med Genomics* 7, 39, <https://doi.org/10.1186/1755-8794-7-39> (2014). PMID: 24962150
17. Markunas C. A. et al. Genetic evaluation and application of posterior cranial fossa traits as endophenotypes for Chiari type I malformation. *Ann Hum Genet* 78, 1–12, <https://doi.org/10.1111/ahg.12041> (2014). PMID: 24359474
18. Lock E. F. et al. Joint eQTL assessment of whole blood and dura mater tissue from individuals with Chiari type I malformation. *BMC Genomics* 16, 11, <https://doi.org/10.1186/s12864-014-1211-8> (2015). PMID: 25609184
19. Merello E. et al. Exome sequencing of two Italian pedigrees with non-isolated Chiari malformation type I reveals candidate genes for cranio-facial development. *Eur J Hum Genet* 25, 952–959, <https://doi.org/10.1038/ejhg.2017.71> (2017). PMID: 28513615
20. Li X. et al. Retinoic acid stimulates chondrocyte differentiation and enhances bone morphogenetic protein effects through induction of Smad1 and Smad5. *Endocrinology* 144, 2514–2523, <https://doi.org/10.1210/en.2002-220969> (2003). PMID: 12746314
21. Jeradi S. & Hammerschmidt M. Retinoic acid-induced premature osteoblast-to-preosteocyte transitioning has multiple effects on calvarial development. *Development* 143, 1205–1216, <https://doi.org/10.1242/dev.129189> (2016). PMID: 26903503
22. Conaway H. H., Henning P. & Lerner U. H. Vitamin a metabolism, action, and role in skeletal homeostasis. *Endocr Rev* 34, 766–797, <https://doi.org/10.1210/er.2012-1071> (2013). PMID: 23720297
23. Wang X. et al. Joint mouse-human phenome-wide association to test gene function and disease risk. *Nat Commun* 7, 10464, <https://doi.org/10.1038/ncomms10464> (2016). PMID: 26833085
24. Ryyanen J. et al. Type VII collagen gene expression by cultured human cells and in fetal skin. Abundant mRNA and protein levels in epidermal keratinocytes. *J Clin Invest* 89, 163–168, <https://doi.org/10.1172/JCI115557> (1992). PMID: 1729268
25. Lapiere J. C. et al. Epitope mapping of type VII collagen. Identification of discrete peptide sequences recognized by sera from patients with acquired epidermolysis bullosa. *J Clin Invest* 92, 1831–1839, <https://doi.org/10.1172/JCI116774> (1993). PMID: 7691888
26. Chou C. H. et al. Direct assessment of articular cartilage and underlying subchondral bone reveals a progressive gene expression change in human osteoarthritic knees. *Osteoarthritis Cartilage* 21, 450–461, <https://doi.org/10.1016/j.joca.2012.11.016> (2013). PMID: 23220557
27. Kuivaniemi H. & Tromp G. Type III collagen (COL3A1): Gene and protein structure, tissue distribution, and associated diseases. *Gene* 707, 151–171, <https://doi.org/10.1016/j.gene.2019.05.003> (2019). PMID: 31075413
28. Theocharis A. D., Manou D. & Karamanos N. K. The extracellular matrix as a multitasking player in disease. *FEBS J* 286, 2830–2869, <https://doi.org/10.1111/febs.14818> (2019). PMID: 30908868
29. Ben Shoham A. et al. Deposition of collagen type I onto skeletal endothelium reveals a new role for blood vessels in regulating bone morphology. *Development* 143, 3933–3943, <https://doi.org/10.1242/dev.139253> (2016). PMID: 27621060
30. Duan X., Bradbury S. R., Olsen B. R. & Berendsen A. D. VEGF stimulates intramembranous bone formation during craniofacial skeletal development. *Matrix Biol* 52–54, 127–140, <https://doi.org/10.1016/j.matbio.2016.02.005> (2016). PMID: 26899202
31. Filipowska J., Tomaszewski K. A., Niedzwiedzki L., Walocha J. A. & Niedzwiedzki T. The role of vasculature in bone development, regeneration and proper systemic functioning. *Angiogenesis* 20, 291–302, <https://doi.org/10.1007/s10456-017-9541-1> (2017). PMID: 28194536
32. Urbizu A. et al. Machine learning applied to neuroimaging for diagnosis of adult classic Chiari malformation: role of the basion as a key morphometric indicator. *J Neurosurg* 129, 779–791, <https://doi.org/10.3171/2017.3.JNS162479> (2018). PMID: 29053075
33. Eppelheimer M. S. et al. Quantification of changes in brain morphology following posterior fossa decompression surgery in women treated for Chiari malformation type 1. *Neuroradiology* 61, 1011–1022, <https://doi.org/10.1007/s00234-019-02206-z> (2019). PMID: 31119343
34. Strahle J. et al. The association between Chiari malformation Type I, spinal syrinx, and scoliosis. *J Neurosurg Pediatr* 15, 607–611, <https://doi.org/10.3171/2014.11.PEDS14135> (2015). PMID: 26030330
35. Kelly M. P., Guillaume T. J. & Lenke L. G. Spinal Deformity Associated with Chiari Malformation. *Neurosurg Clin N Am* 26, 579–585, <https://doi.org/10.1016/j.nec.2015.06.005> (2015). PMID: 26408068
36. Bushell G. R., Ghosh P. & Taylor T. K. Collagen defect in idiopathic scoliosis. *Lancet* 2, 94–95, [https://doi.org/10.1016/s0140-6736\(78\)91400-9](https://doi.org/10.1016/s0140-6736(78)91400-9) (1978). PMID: 78312

37. Uden A., Nilsson I. M. & Willner S. Collagen changes in congenital and idiopathic scoliosis. *Acta Orthop Scand* 51, 271–274, <https://doi.org/10.3109/17453678008990798> (1980). PMID: 7435186
38. Arseni L., Lombardi A. & Orioli D. From Structure to Phenotype: Impact of Collagen Alterations on Human Health. *Int J Mol Sci* 19, <https://doi.org/10.3390/ijms19051407> (2018). PMID: 29738498
39. Timberlake A. T. et al. Two locus inheritance of non-syndromic midline craniosynostosis via rare SMAD6 and common BMP2 alleles. *Elife* 5, <https://doi.org/10.7554/eLife.20125> (2016). PMID: 27606499
40. Wu N. et al. TBX6 null variants and a common hypomorphic allele in congenital scoliosis. *N Engl J Med* 372, 341–350, <https://doi.org/10.1056/NEJMoa1406829> (2015). PMID: 25564734
41. Houston J. R. et al. A morphometric assessment of type I Chiari malformation above the McRae line: A retrospective case-control study in 302 adult female subjects. *J Neuroradiol* 45, 23–31, <https://doi.org/10.1016/j.neurad.2017.06.006> (2018). PMID: 28826656
42. Aitken L. A. et al. Chiari type I malformation in a pediatric population. *Pediatr Neurol* 40, 449–454, <https://doi.org/10.1016/j.pediatrneurol.2009.01.003> (2009). PMID: 19433279
43. Tubbs R. S., Lyerly M. J., Loukas M., Shoja M. M. & Oakes W. J. The pediatric Chiari I malformation: a review. *Childs Nerv Syst* 23, 1239–1250, <https://doi.org/10.1007/s00381-007-0428-0> (2007). PMID: 17639419
44. Greenlee J. D., Donovan K. A., Hasan D. M. & Menezes A. H. Chiari I malformation in the very young child: the spectrum of presentations and experience in 31 children under age 6 years. *Pediatrics* 110, 1212–1219, <https://doi.org/10.1542/peds.110.6.1212> (2002). PMID: 12456921
45. Cappuccio G. et al. Expanding the phenotype of DST-related disorder: A case report suggesting a genotype/phenotype correlation. *Am J Med Genet A* 173, 2743–2746, <https://doi.org/10.1002/ajmg.a.38367> (2017). PMID: 28767192
46. Scuderi C. et al. Biallelic intragenic duplication in ADGRB3 (BAI3) gene associated with intellectual disability, cerebellar atrophy, and behavioral disorder. *Eur J Hum Genet* 27, 594–602, <https://doi.org/10.1038/s41431-018-0321-1> (2019). PMID: 30659260
47. Sahuquillo J. et al. Posterior fossa reconstruction: a surgical technique for the treatment of Chiari I malformation and Chiari I/syringomyelia complex—preliminary results and magnetic resonance imaging quantitative assessment of hindbrain migration. *Neurosurgery* 35, 874–884; discussion 884–875, <https://doi.org/10.1227/00006123-199411000-00011> (1994). PMID: 7838336
48. Evans J. W. A. An encephalographic ratio for estimating ventricular enlargement and cerebral atrophy. *Archives of Neurology and Psychiatry* 47, 931–937 (1942).
49. Beighton P. & Horan F. Orthopaedic aspects of the Ehlers-Danlos syndrome. *J Bone Joint Surg Br* 51, 444–453 (1969). PMID: 5820785
50. Miller S. A., Dykes D. D. & Polesky H. F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16, 1215 (1988). <https://doi.org/10.1093/nar/16.3.1215> PMID: 3344216
51. Szklarczyk D. et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* 45, D362–D368, <https://doi.org/10.1093/nar/gkw937> (2017). PMID: 27924014
52. Li, H. Vol. 00 1–3 (Elsevier, arXiv, 2013).
53. Poplin, R. et al. (bioRxiv, 2017).
54. Wang K., Li M. & Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38, e164, <https://doi.org/10.1093/nar/gkq603> (2010). PMID: 20601685
55. Auton A. et al. A global reference for human genetic variation. *Nature* 526, 68–74, <https://doi.org/10.1038/nature15393> (2015). PMID: 26432245
56. Lek M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285–291, <https://doi.org/10.1038/nature19057> (2016). PMID: 27535533
57. Adzhubei I. A. et al. A method and server for predicting damaging missense mutations. *Nat Methods* 7, 248–249, <https://doi.org/10.1038/nmeth0410-248> (2010). PMID: 20354512
58. Ng P. C. & Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* 31, 3812–3814 (2003). <https://doi.org/10.1093/nar/gkg509> PMID: 12824425
59. Flicek P. et al. Ensembl 2012. *Nucleic Acids Res* 40, D84–90, <https://doi.org/10.1093/nar/gkr991> (2012). PMID: 22086963
60. Chun S. & Fay J. C. Identification of deleterious mutations within three human genomes. *Genome Res* 19, 1553–1561, <https://doi.org/10.1101/gr.092619.109> (2009). PMID: 19602639

61. Shihab H. A. et al. Ranking non-synonymous single nucleotide polymorphisms based on disease concepts. *Hum Genomics* 8, 11, <https://doi.org/10.1186/1479-7364-8-11> (2014). PMID: [24980617](https://pubmed.ncbi.nlm.nih.gov/24980617/)
62. Pollard K. S., Hubisz M. J., Rosenbloom K. R. & Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* 20, 110–121, <https://doi.org/10.1101/gr.097857.109> (2010). PMID: [19858363](https://pubmed.ncbi.nlm.nih.gov/19858363/)
63. Robinson J. T. et al. Integrative genomics viewer. *Nat Biotechnol* 29, 24–26, <https://doi.org/10.1038/nbt.1754> (2011). PMID: [21221095](https://pubmed.ncbi.nlm.nih.gov/21221095/)
64. Thorvaldsdóttir H., Robinson J. T. & Mesirov J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14, 178–192, <https://doi.org/10.1093/bib/bbs017> (2013). PMID: [22517427](https://pubmed.ncbi.nlm.nih.gov/22517427/)
65. Tsai M. F. et al. PrimerZ: streamlined primer design for promoters, exons and human SNPs. *Nucleic Acids Res* 35, W63–65, <https://doi.org/10.1093/nar/gkm383> (2007). PMID: [17537812](https://pubmed.ncbi.nlm.nih.gov/17537812/)
66. Lee S. et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet* 91, 224–237, <https://doi.org/10.1016/j.ajhg.2012.06.007> (2012). PMID: [22863193](https://pubmed.ncbi.nlm.nih.gov/22863193/)