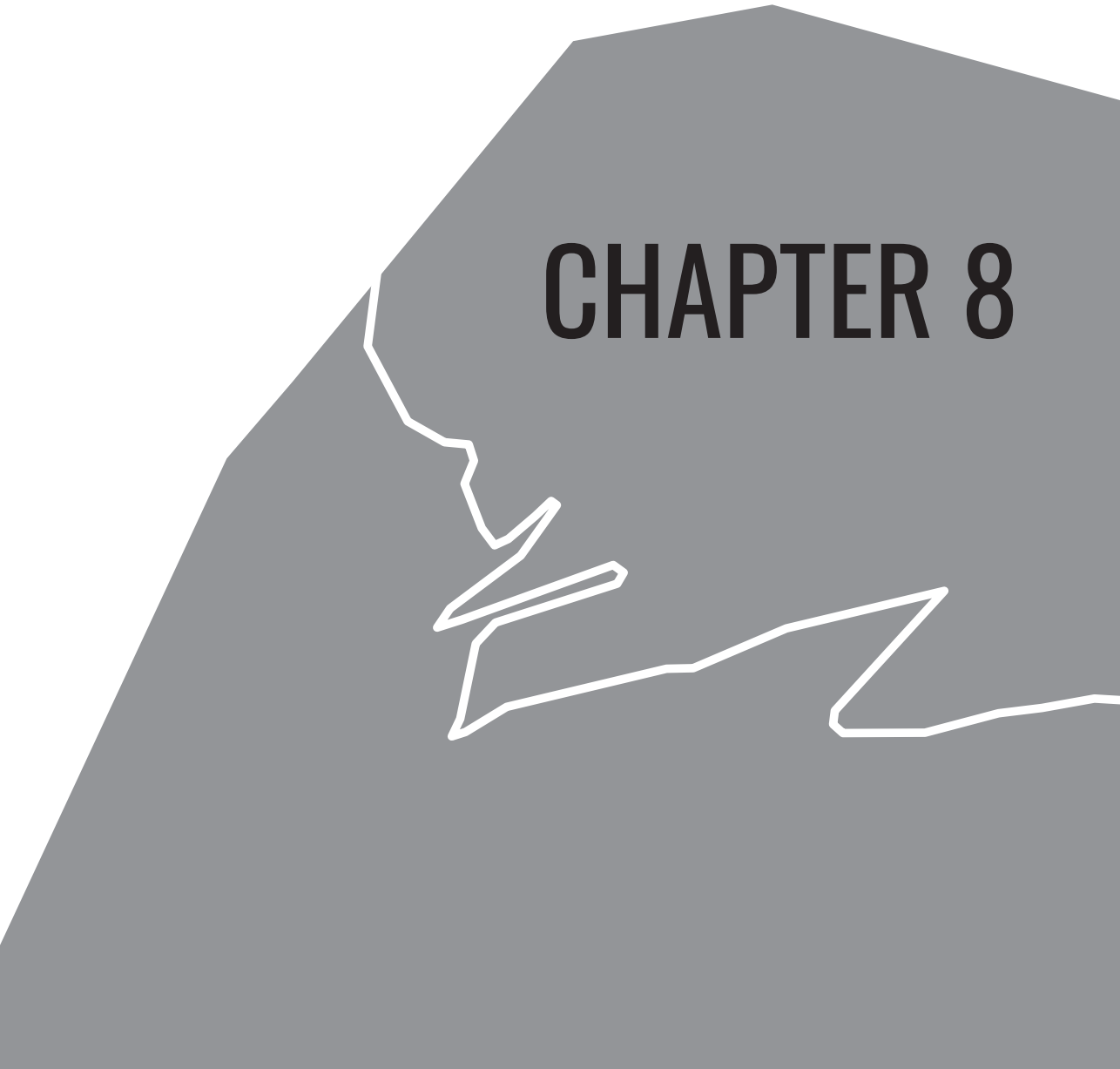
CHAPTER 8



EXOME SEQUENCING IN PATIENT-PARENT TRIOS REVEALS NEW CANDIDATE GENES FOR EARLY-ONSET PRIMARY SCLEROSING CHOLANGITIS

Sjoukje-Marije Haisma
Rinse K Weersma
Maria E Joosse
Barbara AE de Koning
Tim de Meij
Bart GP Koot
Victorien Wolters
Obbe Norbruis
Mark J Daly
Christine Stevens
Ramnik J Xavier
Jukka Koskela
Manuel A Rivas
Marijn C Visschedijk
Henkjan J Verkade
Ruggero Barbieri
Dianne BH Jansen
Eleonora AM Festen
Patrick F van Rheenen
Cleo C van Diemen

Manuscript submitted

ABSTRACT

Background & aims: Primary sclerosing cholangitis (PSC) is a rare bile duct disease strongly associated with inflammatory bowel disease (IBD). While several rare mutations are known to contribute to very early onset IBD, similar variants have not been identified for early-onset PSC. We therefore performed whole-exome sequencing (WES) in patients diagnosed with PSC before the age of 13 to investigate.

Methods: In this multicenter study, WES was performed on 95 DNA samples from 29 index patients with early-onset PSC and their biological parents and eight single early-onset PSC patients. 81% of patients had IBD at the time of PSC diagnosis. We performed patient-parent trio analyses and selected rare (minor allele frequency <0.1%) coding and splice-site variants that matched recessive (homozygous and compound heterozygous variants) and dominant (de novo) inheritance in the index patients. Variant pathogenicity was predicted by an in-house developed algorithm (GAVIN), and PSC-relevant variants were selected using gene expression data and gene function, amongst other factors.

Results: In two separate trios we identified compound heterozygous variants in the genes ABCB6, DACT1 and JMJDC1. In eight other trios we identified 10 de novo variants in 10 genes with predicted pathogenic effects on protein function. The genes identified have roles in bile salt homeostasis, adaptive and innate immunity, and epithelial barrier function.

Conclusion: For 10 out of 29 families, we identified rare protein-altering genetic variants in 13 genes that may explain a substantial part of the etiology of PSC. The functional consequences of these newly discovered variants, and the associated susceptibility to PSC, will require further verification using replication studies and functional testing.

INTRODUCTION

Primary sclerosing cholangitis (PSC) is a rare chronic cholestatic disease characterized by progressive inflammation and obliterative fibrosis of the intra- and extrahepatic bile ducts.[1] There is a strong relation between PSC and inflammatory bowel disease (IBD). Patients who initially present with isolated PSC may go on to develop IBD years later.[2,3] In adult-onset disease, approximately two-thirds of patients with PSC have concurrent IBD.[1] The co-occurrence of PSC and IBD is higher in children than in adults, varying from 76% to 97%.[4-7]

Disease progression is inevitable in the majority of PSC patients, with the development of biliary cirrhosis and portal hypertension requiring repeated endoscopic procedures. Liver transplantation is the only curative treatment option, but the disease recurs in 20-25% of transplanted patients.[4,6] Cholangiocarcinoma and colorectal cancer are feared complications in PSC and the most common causes of death.[8]

The pathogenesis of PSC is largely unknown. Genome wide association studies (GWAS) in adult-onset PSC carried out by the International PSC Study Group recently identified 31 risk loci, but the associated genetic variants so far explain <10% of disease susceptibility.[9] It has been speculated that rare variants with large effect size may play a role in the onset of complex disorders, but these variants are so rare in allele frequency (many of them private mutations) that their genetic signals are not detected by GWAS. In contrast, whole exome sequencing (WES) in patients with extreme phenotypes, such as young age of disease onset, has led to the identification of potentially causative genetic variants in IBD, chronic obstructive pulmonary disease and diabetes type 1.[10-13] Likewise, in a subset of patients with early-onset PSC, we expect to find rare genetic variants resembling a monogenic or oligogenic inheritance pattern. We therefore performed WES in a Dutch cohort of patients with early-onset PSC and their parents.

METHODS

Study design, participants and setting

In this multicenter parent-offspring study we collected DNA from PSC patients with disease-onset prior to their 13th birthday and from their biological parents. PSC diagnosis was confirmed by cholangiography (presence of multifocal strictures, focal dilatation, or beading of the biliary tree) or liver histology (presence of bile duct damage, onion-skinned peri-ductal fibrosis, inflammation, portal edema or fibrosis, ductopenia, ductular proliferation, or cholestasis), or both. Patients with sclerosing cholangitis due to secondary causes such as surgery, trauma, cancer or infection were excluded from participation.

Patients were recruited in five tertiary care hospitals in the Netherlands – University Medical Center Groningen (UMCG, a referral pediatric liver transplant center), Erasmus University Medical Center–Sophia Children's Hospital, VU University Medical Center, Amsterdam University Medical Center–Emma Children's Hospital, University Medical Center Utrecht–Wilhelmina Children's Hospital – and one large general teaching hospital, the Isala Hospital. Eligible patients were those regularly attending the (pediatric) gastroenterology and hepatology clinics as part of standard care. After informed consent was given, the following information was obtained from the local patient records and entered in an online clinical registry using Castor Electronic Data Capture (Amsterdam, the Netherlands): age at PSC diagnosis, findings on cholangiography and/or histology, and follow-up data on medication use and appearance of biliary cirrhosis, portal hypertension or malignancies. If applicable, age at IBD diagnosis, IBD type and location based on the Paris Classification[14] were also entered. Between January 2017 and July 2017 blood was collected from patients and volunteering parents for genomic DNA extraction according to standard protocols.

Ethical considerations

The Medical Ethical Committee of the UMCG approved the study protocol (METC 2016/289), and secondary approval was obtained from all participating centers. All participating parents and teenagers 12-19 years old gave informed consent prior study inclusion.

Whole-exome sequencing
Libraries were prepared using the Illumina Nextera prep kit and hybrid capture (Illumina Rapid Capture Enrichment – 37 Mb target), and sequencing was performed using the Illumina HiSeq 2500 at the Broad Institute of MIT and Harvard. All raw data underwent quality control steps (https://hub.docker.com/r/broadinstitute/gatk/) without any noticeable negative features to achieve 86.06 million high quality reads per sample with 98.85% of reads aligned, on average, resulting in a coverage of 81% of the target region with a read depth of >30X. Sequence reads were aligned to the human reference genome using Novoalign (http://www.novocraft.com). Next, the Genome Analysis Toolkit of the Broad Institute [15] was used for calling single-nucleotide polymorphism and insertions/deletions.

Variant annotation
Variants were annotated with SNPEff [16], using publicly available data from Ensembl and Refseq, and with GAVIN, an annotation tool with an algorithm that scores the likely pathogenicity of the variants.[17] Additional annotations at the variant-, exon- and gene-level were obtained from the 1000 Genomes Project (http://www.1000genomes.org); National Heart, Lung and Blood Institute GO Exome Sequencing Project Exome Variant Server (http://evs.gs.washington.edu/EVS); PolyPhen2[18] and the Exome Aggregation Consortium (ExAC, http://exac.broadinstitute.org).

Variant filtering
We used variants with a sequence coverage of five or greater. We used a Genomics Data Management System (Alissa Interpret – Agilent technologies) to create a filtering tree specifically designed for this study (see figure 1).

We performed patient-parent trio analyses. On the variant-level, we selected variants matching recessive (homozygous and compound heterozygous variants) and dominant (de novo) inheritance in the children. HLA-variants were excluded from this analysis. We filtered for variants with a minor allele frequency <0.1% in data from the 1000 Genomes Project (http://www.1000genomes.org), EVS (http://evs.gs.washington.edu/EVS) and ExAC (http://exac.broadinstitute.org). Variants were then selected based upon whether they were [1] deemed to be coding (missense- and nonsense mutations, frameshift insertions and deletions) or to have an effect on splicing and [2] predicted to be

(likely) pathogenic according to GAVIN [17].

Further variant prioritization was based on evidence from the literature for the relevance of the gene to the disease recovered from multiple databases including Genecards (www.genecards.com), Reactome (www.reactome.org) and OMIM (Online Mendelian Inheritance in Man; www.omim.org). Genes were selected when they were expressed on the mRNA-level in the liver, gallbladder or intestines. We then selected genes based on one of the following criteria:

1. the gene function was already known to be associated with the PSC phenotype or a similar phenotype (including immunological, inflammatory or bile salt pathways),
2. the gene was not (or rarely) reported in the literature and therefore cannot be excluded from having a potential role in the disease pathogenesis,
3. the gene function is well known but not directly associated to the disease phenotype (i.e. gene function in DNA replication) and can therefore also not be excluded from having a potential role in disease pathogenesis.

Genes were excluded when there was substantial literature evidence that the gene codes for a function not related to the PSC-phenotype (i.e. olfactory receptor genes (TAARs), motile cilia function genes (CCDC40), or keratin associated protein genes (KRTAP5-6)).

Variant verification and validation

De novo variants were manually checked for quality in the BAM files. If there was doubt about the validity of the variant, confirmatory Sanger sequencing was performed. To find further supporting evidence for new candidate genes, the exomes of the patients whose parental DNA was not obtained (D1–D8) were checked for possible disease-causing variants within the candidate genes from the trio-analyses.

RESULTS

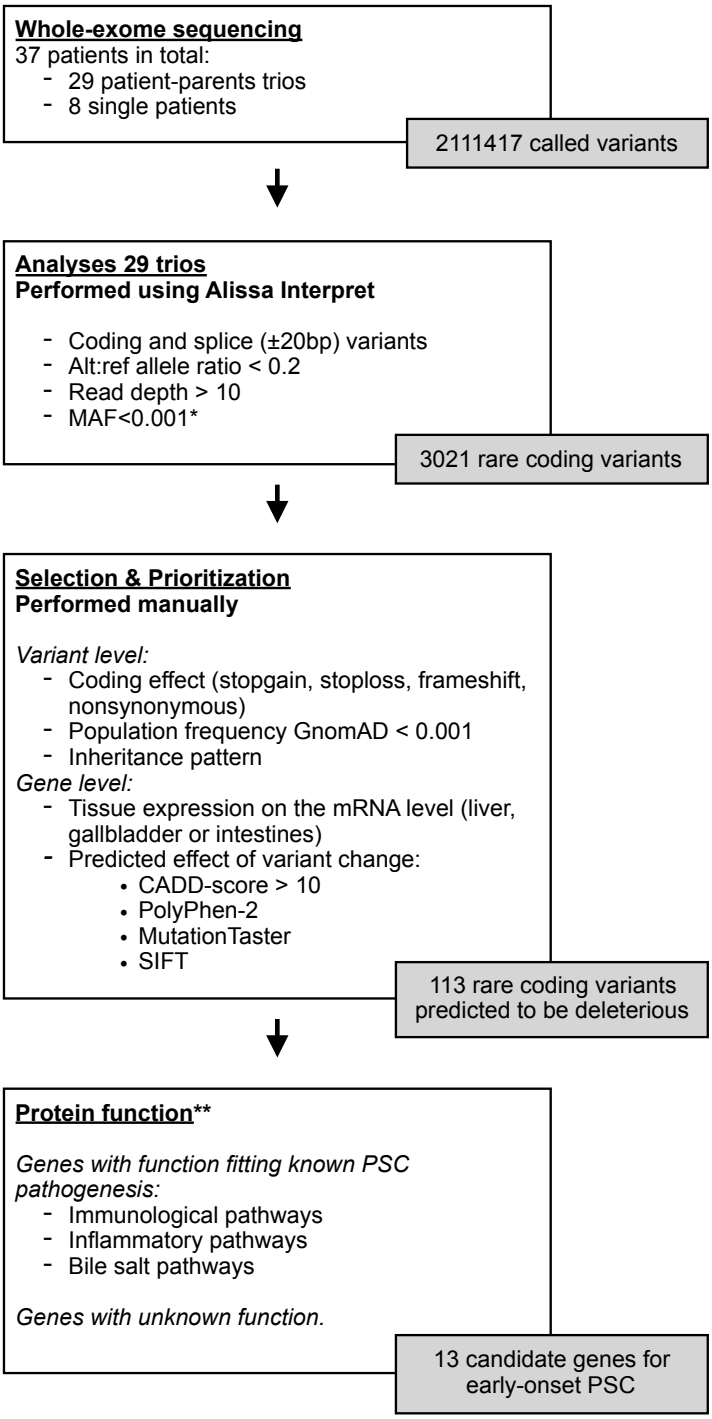
A total of 37 patients with early-onset PSC were enrolled in this study (see figure 1). WES was performed on all 95 DNA samples (29 patient-parent trios and 8 single patients). Table 1 shows the patient characteristics. Patients were diagnosed with PSC at a median age of 10.3 years (range: 2.5–12.8) and were predominantly male (70%). 81% of patients (n=30) had concurrent IBD, with ulcerative colitis significantly more prevalent than Crohn's disease (73% vs. 27%). Other autoimmune disorders included celiac disease (n=1), idiopathic thrombocytopenic purpura (n=1) and vitiligo (n=1). None of the parents had liver disease, but three had IBD.

Table 1. Patient characteristics

(n = 37)	
Median age at PSC diagnosis, yrs (range)	110.3 (2.5-12.8)
Male gender	70%
Autoimmune hepatitis overlap syndrome	38%
Inflammatory bowel disease (IBD)	81%
Median age at IBD diagnosis, yrs (range)	10.3 (2.5-16.4)
Type of IBD	
Ulcerative colitis	73% (56-86)
Crohn's disease	27% (14-44)
IBD in first-degree relatives	8%
Liver disease in first-degree relatives	0%

Values are percentages (95% confidence interval) unless otherwise stated.

The median time between PSC diagnosis and inclusion in this study was 5.0 years. Biliary complications including cholangitis or bile duct obstruction had occurred in three patients (8%), and cirrhosis had occurred in eight (22%). Two patients (5%) underwent a liver transplantation after a disease duration of 10 and 11 years, respectively, and two other patients were on the waiting list for liver transplantation. One of the cirrhotic patients had experienced bleeding of esophageal varices and required a transjugular intrahepatic portosystemic shunt procedure. Thirty-four patients (92%) were prescribed ursodeoxycholic acid.



← Figure 1. Variant selection & prioritization
* Population databases used: ExAC, gnoMAD, 1000 Genomes project.
** Information from multiple databases including Genecards, Reactome and OMIM.
Abbreviations: MAF, Minor allele frequency; CADD-score, Combined Annotation-Dependent Depletion score; PolyPhen2, Polymorphism Phenotyping 2; SIFT, Sorting Intolerant From Tolerant.

Patient-parent trio analyses of WES data

Figure 1 provides an overview of WES variant selection and prioritization. We identified 13 candidate genes matching a recessive or dominant inheritance pattern that are also known to be expressed on the mRNA-level in the liver, gallbladder or intestines and predicted to be deleterious by at least one pathogenicity prediction tool (see Table 2). We identified compound heterozygous variants in the *ABCB6* gene in trio 3 and compound heterozygous variants in the *DACT1* and *JMJDC1* gene in trio 21. These variants disrupt highly conserved regions of the proteins and were therefore considered to be protein-altering variants. In eight other trios, we identified 10 de novo protein-altering variants.

Figure 2 provides an overview of the possible pathogenic mechanisms of our findings. The genes *MARCH1* and *PTX4* encode for proteins that have roles associated with the immune system. The genes *TRDN*, *SLC9B1* and *ABCB6* are related to the 'Transport of glucose and other sugars, bile salts and organic acids, metal ions and amine compounds' pathway (www.reactome.org). We identified one de novo intronic variant positioned exactly at a splice-donor consensus sequence site in *CDHR2*, disrupting splicing of the transcript. De novo nonsynonymous variants were found in the genes *WISP1, CHST11, PLXDC1, CALCRL* and *SMCHD1*, but we could not directly link the known gene functions to the disease.

We also sequenced the DNA of eight early-onset PSC patients whose parental DNA was not obtained. We checked the WES data of these eight patients for possible disease-causing variants in the 13 candidate genes from our trio-analyses and identified a nonsynonymous variant in gene *PLXDC1* (Chr17: 37263667:T/A; p.Tyr235Phe) in patient D3.

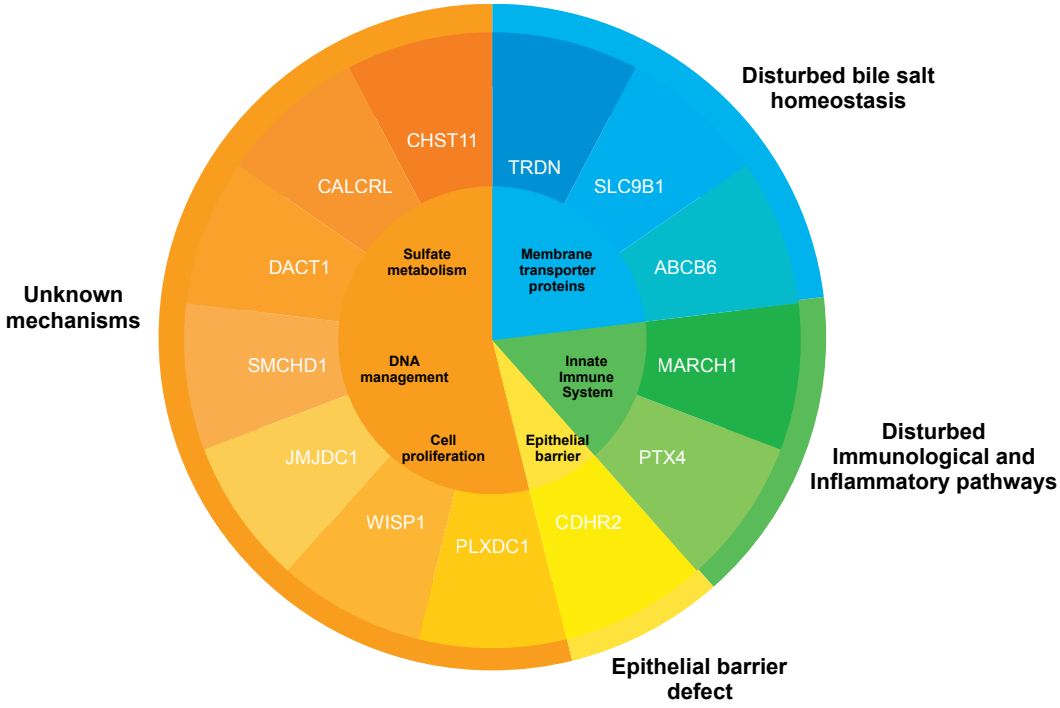


Figure 2. Possible protein-altering genetic variants associated with susceptibility to early-onset PSC (and IBD). The core indicates the protein functions, the inner ring shows the candidate genes and the outer ring represents the possible pathogenic mechanism. Information on genetic functions and pathologic mechanisms comes from multiple databases including Genecards (www.genecards.com), Reactome (www.reactome.org) and OMIM (Online Mendelian Inheritance in Man; www.omim.org).

Table 2. Candidate risk genes for early-onset PSC resulting from the 29 patient-parent trio analyses.

Trio	Chr: position:alleles	rs number	Candidate risk gene	Inheritance mode (parental allele)	GnomAD allele count = Population frequency
3	2:220075521:C/T	rs148211042	*ABCB6*	Compound heterozygous (mother)	217 = 0.00077
	2:220078006:C/T	rs145526996	*ABCB6*	Compound heterozygous (father)	1236 = 0.0043
4	4:164775272:C/T	Unknown	*MARCH1*	De novo	Unknown
	16:1537911:C/T	rs775407157	*PTX4*	De novo	3= 0.000012
	17:37234300:G/A	Unknown	*PLXDC1*	De novo	Unknown
10	8:48866910:./T	Unknown	*WISP1*	De novo	Unknown
17	15:176002840:G/A	rs780769740	*CDHR2*	De novo	5 = 0.00003
18	2:188228104:G/A	Unknown	*CALCRL*	De novo	Unknown
21	10:64927837:C/T	rs71508957	*JMJD1C*	Compound heterozygous (father)	1162 =0.0041
	10:64974807:C/G	rs200016210	*JMJD1C*	Compound heterozygous (mother)	132 = 0.00047

Amino Acid change	CADD-score	Protein function
p.R723Q	35.0	Binds heme and porphyrins and functions in their ATP-dependent uptake into the mitochondria. Mutations in this gene underlie familial pseudohyperkalemia (OMIM 609153) and dyschromatosis universalis hereditarian (OMIM 615402). [19-21]
p.G588S	32.0	
p.W4*	38.0	Downregulates surface expression of major histocompatibility complex (MHC) class II molecules and other glycoproteins by directing them to the late endosomal/lysosomal compartment.[22,23]
p.V63M	12.5	Pentraxins are part of the humoral arm of innate immunity and behave as functional ancestors of antibodies by mediating agglutination, complement activation and opsonization.[24]
p.A351V	23.8	Plays a critical role in endothelial cell capillary morphogenesis.[25,26]
p.C79Y	27.2	Mediates diverse developmental processes, such as control of cell proliferation, adhesion, cell polarity and establishment of cell fates.[27,28]
p.= SPLICE_ SITE_ DONOR	16.9	Intermicrovillar adhesion molecule that controls the packing of microvilli at the apical membrane of epithelial cells.[29,30]
p.P209L	29.6	Receptor for calcitonin-gene-related peptide (CGRP) and adrenomedullin.[31,32]
p.E2531K	26.5	A candidate histone demethylase thought to be a coactivator for key transcription factors. Plays a role in the DNA-damage response pathway.[33,34]
p.D374H	26.2	

Table 2 continues on next page

Continuation of previous page

	14:59104943:C/T	Unknown	*DACT1*	Compound heterozygous (mother)	5 = 0
	14:59113376:T/C	rs200977826	*DACT1*	Compound heterozygous (father)	166 = 0.001
23	4:103832611:G/A	rs75599926	*SLC9B1*	De novo	2 = 0.000011
24	6:123786033:./A	rs201431159	*TRDN*	De novo	Unknown
26	18:2722603:G/A	Unknown	*SMCHD1*	De novo	Unknown
28	12: 105151159:G/A	Unknown	*CHST11*	De novo	Unknown

Abbreviations: CADD-score, Combined Annotation-Dependent Depletion score;

p.T8M	23.6	Interacts with, and positively regulates, dishevelled-mediated signaling pathways during development. [35,36] Associated with Townes-Brocks syndrome-2 (OMIM 617466) [37] Sodium/hydrogen exchanger and transmembrane protein. Among its related pathways are Transport of glucose and other sugars, bile salts and organic acids, metal ions and amine compounds and Ion channel transport.[52,53]
p.W679R	25.5	
p.R305*	36.	Sodium/hydrogen exchanger and transmembrane protein.[38,39]
p.S297Ffs*32	n.a.	Contributes to regulation of luminal Ca2+ release via the sarcoplasmic reticulum calcium release channels.[40] Associated to ventricular tachycardia (OMIM 615441)[41]
p.D849N	31	Involved in DNA management and plays an essential role in X chromosome inactivation.[42]
p.G213S	32.	Catalyzes the transfer of sulfate in chondroitin.[43] Diseases associated with CHST11 include Mucinoses and Costello Syndrome (OMIM 618167).

DISCUSSION

Key results

In this family-based study we examined the exomes of patients with early-onset PSC and their biological parents. The trio-analysis approach revealed 13 candidate disease-causing variants with large effects on protein function in several genes, many of them involved in immunological or bile salt pathways. Our findings strengthen the hypothesis that rare variants can contribute to the development of PSC in patients with extreme PSC phenotypes such as early-onset disease.

Interpretation

Previous studies aiming to identify rare variants in PSC

Low-frequency and rare genetic variants often have larger effect-sizes on protein function than common variants.[44,45] Previous studies have shown that these rare variants may also contribute to the development of complex disorders.[10,11,46] Proper analyses of these variants, revealed by WES, requires massive numbers of cases and controls. Therefore, in 2013, the BROAD Institute partnered with researchers worldwide to develop a collaborative exome sequencing network in IBD, and this initiative is currently ongoing. A similar project is now up and running in PSC with the aim to meta-analyze the exomes of more than 1000 patients of European ancestry.

An alternate method for studying low-frequency and rare variants is to focus on extreme PSC-phenotypes because they are more likely to be caused by rare variants.[45] We therefore decided to select only patients with early-onset disease in this study. WES of all protein-coding genes results in many variants of uncertain clinical significance. A trio-analysis design, i.e. the inclusion of patients and parents in the analysis of rare genetic variants, helps to immediately ascertain whether a variant is inherited or de novo. Population stratification is a major concern with rare variants because they tend to be more geographically clustered than common variants. However, in contrast to case-control studies, trio association studies are less sensitive to population stratification.

Potential role of the identified new risk genes in PSC pathogenesis

Previous studies on the genetics of PSC have confirmed the auto-immune origins of the disease, with the predominant genetic findings localized within the human leukocyte antigen (HLA) complex on chromosome 6 and most non-HLA loci associated with other immune-mediated or auto-immune pathways.[1,9] Furthermore, liver biopsies from patients with PSC showed mainly T cells and, to a lesser degree, macrophages and neutrophils in the infiltrates.[1] In our search for potential disease-causing variants, we therefore prioritized genes that related to immunological or inflammatory pathways.

Based on previous studies, the gene *MARCH1* perfectly segregates with an auto-immune phenotype. A de novo stop-gain variant located at the very beginning of this gene (transcript position 4) was found in a boy of 8 years of age with PSC–autoimmune hepatitis overlap syndrome, also called autoimmune sclerosing cholangitis. Functional studies of *MARCH1* have confirmed that this gene mediates the immunosuppressive effect of the anti-inflammatory cytokine interleukin 10 (IL10) on antigen presentation in monocytes via ubiquitination and degradation of major histocompatibility complex (MHC) class II molecules.[22,23] Knockdown of *MARCH1* strongly inhibited IL-10–dependent down-regulation of cell surface HLA-DR.[22] The exact contribution of the *MARCH1* gene regulation to immunopathology remains to be explored.

Some of the previously identified PSC susceptibility loci harbor genes that are potentially involved in bile acid homeostasis.[1] In our cohort, we identified multiple variants in genes related to the 'Transport of glucose and other sugars, bile salts and organic acids, metal ions and amine compounds' pathway (www.reactome.org). A girl with disease-onset of both PSC and IBD at age 7, had compound heterozygous *ABCB6* variants. The gene encodes a member of the ATP-binding cassette (ABC) transporter superfamily and is known to bind heme and porphyrins and function in their ATP-dependent uptake in the mitochondria.[21,47] Both variants were predicted to be damaging and to disrupt the highly conserved ABC transport and ABC transmembrane regions of the protein, respectively. Interestingly, mutation of another ABC-transporter gene (*ABCB4*) leads to progressive familial

intrahepatic cholestasis (PFIC) type 3, a Mendelian cholestatic syndrome with many similarities to PSC.

Although we can link several of the new candidate genes we found with the known pathogenesis of PSC, this is more difficult for other candidate genes as their protein function remains unknown. Rare variants with large coding effects in genes of unknown function were therefore also included in our list of new candidate genes.

Strengths & limitations

This is the first family-based WES study performed in a relatively large subgroup of PSC patients with an extreme phenotype, namely young age of disease-onset. Previous studies have used this method to identify disease-causing rare variants in isolated cases, and then performed targeted-sequencing in patients with similar phenotypes.[46,48] Our findings suggest that, within this young subgroup of PSC patients with a severe phenotype, rare variants largely affect the onset of their disease, resembling a more monogenic or oligogenic inheritance pattern. We have not yet performed functional tests to confirm this presumption, and this paper should therefore be seen as hypothesis-generating, providing a starting point for further studies. Uncovering the functional consequences of the newly discovered genetic variants and the mechanisms involved in the onset of PSC will require detailed functional experiments involving different functional read-outs, given the broad nature of the identified genes, and further verification of our findings in independent cohorts.

Implications for clinical practice

The lack of understanding of PSC pathogenesis hampers the development of effective therapies. Investigating the genetic basis of a disease can help reveal mechanisms of disease pathology and guide the selection of new targets for drug discovery. Each genetic risk locus can be seen as a potential drug target and the starting point of new treatment opportunities. This has successfully been demonstrated in the field of IBD, in which small-molecule inhibitors were used to recapitulate the anti-inflammatory function of *CARD9* variants associated with protection from IBD.[49] Scientists now recognize that genes with evidence for causality in disease are more promising for identification of new drug targets, and this has led to an increased interest in disease-associated genes with variants that reduce gene function, such as nonsense,

frameshift or essential splice-site variants.[50] With our study we have shown that combining a trio design with WES can reveal loss-of-function variants. Uncovering genetic variants that provide causal evidence will provide valuable insights into disease biology, but this will also, most importantly, help to design new targeted therapies.

Conclusion

We identified 13 rare protein-altering genetic variants in 10 out of 29 patient-parent trios, including variants in genes involved in immunological, epithelial barrier and bile salt pathways. The functional consequences of these variants and the associated susceptibility to PSC will require further verification, but we have shown the trio-design to be a successful method for discovering candidate disease-causing genes in rare diseases.

Acknowledgements

We thank Krista van Dijk en Kristin Abbott for assistance with variant filtering and interpretation. We thank Kate Mc Intyre for editorial assistance.

REFERENCES

1. Karlsen TH, Folseraas T, Thorburn D, Vesterhus M. Primary sclerosing cholangitis - a comprehensive review. J Hepatol 2017;67:1298-1323.
2. European Association for the Study of the Liver. EASL Clinical Practice Guidelines: management of cholestatic liver diseases. J Hepatol 2009;51:237-267.
3. Saich R, Chapman R. Primary sclerosing cholangitis, autoimmune hepatitis and overlap syndromes in inflammatory bowel disease. World J Gastroenterol 2008;14:331-337.
4. Deneau MR, Mack C, Abdou R, Amin M, Amir A, Auth M, et al. Gamma Glutamyltransferase Reduction Is Associated With Favorable Outcomes in Pediatric Primary Sclerosing Cholangitis. Hepatol Commun 2018;2:1369-1378.
5. Deneau M, Jensen MK, Holmen J, Williams MS, Book LS, Guthery SL. Primary sclerosing cholangitis, autoimmune hepatitis, and overlap in Utah children: epidemiology and natural history. Hepatology 2013;58:1392-1400.
6. Deneau MR, El-Matary W, Valentino PL, Abdou R, Alqoaer K, Amin M, et al. The natural history of primary sclerosing cholangitis in 781 children: A multicenter, international collaboration. Hepatology 2017;66:518-527.
7. Valentino PL, Wiggins S, Harney S, Raza R, Lee CK, Jonas MM. The Natural History of Primary Sclerosing Cholangitis in Children: A Large Single-Center Longitudinal Cohort Study. J Pediatr Gastroenterol Nutr 2016;63:603-609.
8. Boonstra K, Weersma RK, van Erpecum KJ, Rauws EA, Spanier BW, Poen AC, et al. Population-based epidemiology, malignancy risk, and outcome of primary sclerosing cholangitis. Hepatology 2013;58:2045-2055.
9. Jiang X, Karlsen TH. Genetics of primary sclerosing cholangitis and pathophysiological implications. Nat Rev Gastroenterol Hepatol 2017;14:279-295.
10. Qiao D, Lange C, Beaty TH, Crapo JD, Barnes KC, Bamshad M, et al. Exome Sequencing Analysis in Severe, Early-Onset Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med 2016;193:1353-1363.
11. Pazmandi J, Kalinichenko A, Ardy RC, Boztug K. Early-onset inflammatory bowel disease as a model disease to identify key regulators of immune homeostasis mechanisms. Immunol Rev 2019;287:162-185.
12. Kelsen JR, Baldassano RN. The role of monogenic disease in

children with very early onset inflammatory bowel disease. Curr Opin Pediatr 2017;29:566-571.
13. Bansal V, Gassenhuber J, Phillips T, Oliveira G, Harbaugh R, Villarasa N, et al. Spectrum of mutations in monogenic diabetes genes identified from high-throughput DNA sequencing of 6888 individuals. BMC Med 2017;15:3.
14. Levine A, Griffiths A, Markowitz J, Wilson DC, Turner D, Russell RK, et al. Pediatric modification of the Montreal classification for inflammatory bowel disease: the Paris classification. Inflamm Bowel Dis 2011;17:1314-1321.
15. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297-1303.
16. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 2012;6:80-92.
17. van der Velde, K J, de Boer EN, van Diemen CC, Sikkema-Raddatz B, Abbott KM, Knopperts A, et al. GAVIN: Gene-Aware Variant INterpretation for medical sequencing. Genome Biol 2017;18:7.
18. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet 2013;Chapter 7:Unit7.20.
19. Bawazir WM, Flatt JF, Wallis JP, Rendon A, Cardigan RA, New HV, et al. Familial pseudohyperkalemia in blood donors: a novel mutation with implications for transfusion practice. Transfusion 2014;54:3043-3050.
20. Helias V, Saison C, Ballif BA, Peyrard T, Takahashi J, Takahashi H, et al. ABCB6 is dispensable for erythropoiesis and specifies the new blood group system Langereis. Nat Genet 2012;44:170-173.
21. Mitsuhashi N, Miki T, Senbongi H, Yokoi N, Yano H, Miyazaki M, et al. MTABC3, a novel mitochondrial ATP-binding cassette protein involved in iron homeostasis. J Biol Chem 2000;275:17536-17540.
22. Thibodeau J, Bourgeois-Daigneault MC, Huppe G, Tremblay J, Aumont A, Houde M, et al. Interleukin-10-induced MARCH1 mediates intracellular sequestration of MHC class II in monocytes. Eur J Immunol 2008;38:1225-1230.
23. De Gassart A, Camosseto V, Thibodeau J, Ceppi M, Catalan N, Pierre P, et al. MHC class II stabilization at the surface of human

dendritic cells is the result of maturation-dependent MARCH I down-regulation. Proc Natl Acad Sci U S A 2008;105:3491-3496.
24. Martinez de la Torre, Y, Fabbri M, Jaillon S, Bastone A, Nebuloni M, Vecchi A, et al. Evolution of the pentraxin family: the new entry PTX4. J Immunol 2010;184:5055-5064.
25. Carson-Walter EB, Watkins DN, Nanda A, Vogelstein B, Kinzler KW, St Croix B. Cell surface tumor endothelial markers are conserved in mice and humans. Cancer Res 2001;61:6649-6655.
26. St Croix B, Rago C, Velculescu V, Traverso G, Romans KE, Montgomery E, et al. Genes expressed in human tumor endothelium. Science 2000;289:1197-1202.
27. Pennica D, Swanson TA, Welsh JW, Roy MA, Lawrence DA, Lee J, et al. WISP genes are members of the connective tissue growth factor family that are up-regulated in wnt-1-transformed cells and aberrantly expressed in human colon tumors. Proc Natl Acad Sci U S A 1998;95:14717-14722.
28. Xu L, Corcoran RB, Welsh JW, Pennica D, Levine AJ. WISP-1 is a Wnt-1- and beta-catenin-responsive oncogene. Genes Dev 2000;14:585-595.
29. Pinette JA, Mao S, Millis BA, Krystofiak ES, Faust JJ, Tyska MJ. Brush border protocadherin CDHR2 promotes the elongation and maximized packing of microvilli in vivo. Mol Biol Cell 2019;30:108-118.
30. Crawley SW, Shifrin DA, Grega-Larson NE, McConnell RE, Benesh AE, Mao S, et al. Intestinal brush border assembly driven by protocadherin-based intermicrovillar adhesion. Cell 2014 ;157:433-446.
31. Foord SM, Craig RK. Isolation and characterisation of a human calcitonin-gene-related-peptide receptor. Eur J Biochem 1987;170:373-379.
32. McLatchie LM, Fraser NJ, Main MJ, Wise A, Brown J, Thompson N, et al. RAMPs regulate the transport and ligand specificity of the calcitonin-receptor-like receptor. Nature 1998;393:333-339.
33. Lee JW, Choi HS, Gyuris J, Brent R, Moore DD. Two classes of proteins dependent on either the presence or absence of thyroid hormone for interaction with the thyroid hormone receptor. Mol Endocrinol 1995;9:243-254.
34. Saez MA, Fernandez-Rodriguez J, Moutinho C, Sanchez-Mut JV, Gomez A, Vidal E, et al. Mutations in JMJD1C are involved in Rett syndrome and intellectual disability. Genet Med 2016;18:378-385.
35. Shi Y, Ding Y, Lei YP, Yang XY, Xie GM, Wen J, et al. Identification of novel rare mutations of DACT1 in human neural tube defects.

Hum Mutat 2012;33:1450-1455.
36. Katoh M, Katoh M. Identification and characterization of human DAPPER1 and DAPPER2 genes in silico. Int J Oncol 2003;22:907-913.
37. Cheyette BN, Waxman JS, Miller JR, Takemaru K, Sheldahl LC, Khlebtsova N, et al. Dapper, a Dishevelled-associated antagonist of beta-catenin and JNK signaling, is required for notochord formation. Dev Cell 2002;2:449-461.
38. Ye G, Chen C, Han D, Xiong X, Kong Y, Wan B, et al. Cloning of a novel human NHEDC1 (Na+/H+ exchanger like domain containing 1) gene expressed specifically in testis. Mol Biol Rep 2006;33:175-180.
39. Holmes RS, Spradling-Reeves KD, Cox LA. Evolution of Vertebrate Solute Carrier Family 9B Genes and Proteins (SLC9B): Evidence for a Marsupial Origin for Testis Specific SLC9B1 from an Ancestral Vertebrate SLC9B2 Gene. J Phylogenetics Evol Biol 2016;4:9002.1000167. Epub 2016 Jun 10.
40. Oddoux S, Brocard J, Schweitzer A, Szentesi P, Giannesini B, Brocard J, et al. Triadin deletion induces impaired skeletal muscle function. J Biol Chem 2009;284:34918-34929.
41. Chopra N, Yang T, Asghari P, Moore ED, Huke S, Akin B, et al. Ablation of triadin causes loss of cardiac Ca2+ release units, impaired excitation-contraction coupling, and cardiac arrhythmias. Proc Natl Acad Sci U S A 2009;106:7636-7641.
42. Blewitt ME, Gendrel AV, Pang Z, Sparrow DB, Whitelaw N, Craig JM, et al. SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. Nat Genet 2008;40:663-669.
43. Yamauchi S, Mita S, Matsubara T, Fukuta M, Habuchi H, Kimata K, et al. Molecular cloning and expression of chondroitin 4-sulfotransferase. J Biol Chem 2000;275:8975-8981.
44. Bomba L, Walter K, Soranzo N. The impact of rare and low-frequency genetic variants in common disease. Genome Biol 2017;18:4.
45. Rivas MA, Beaudoin M, Gardet A, Stevens C, Sharma Y, Zhang CK, et al. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. Nat Genet 2011;43:1066-1073.
46. Kelsen JR, Dawany N, Martinez A, Grochowski CM, Maurer K, Rappaport E, et al. A de novo whole gene deletion of XIAP detected by exome sequencing analysis in very early onset inflammatory bowel disease: a case report. BMC Gastroenterol 2015;15.

47. Krishnamurthy PC, Du G, Fukuda Y, Sun D, Sampath J, Mercer KE, et al. Identification of a mammalian mitochondrial porphyrin transporter. Nature 2006;443:586-589.
48. Li Q, Lee CH, Peters LA, Mastropaolo LA, Thoeni C, Elkadri A, et al. Variants in TRIM22 That Affect NOD2 Signaling Are Associated With Very-Early-Onset Inflammatory Bowel Disease. Gastroenterology 2016;150:1196-1207.
49. Leshchiner ES, Rush JS, Durney MA, Cao Z, Dancik V, Chittick B, et al. Small-molecule inhibitors directly target CARD9 and mimic its protective variant in inflammatory bowel disease. Proc Natl Acad Sci U S A 2017;114:11392-11397.
50. Minikel Eric Vallabh, Karczewski Konrad J, Martin Hilary C, Cummings Beryl B, Whiffin Nicola, Alföldi Jessica, Trembath Richard C, van Heel David A, Daly Mark J, Genome Aggregation Database Production Team, Genome Aggregation Database Consortium, Schreiber Stuart L, MacArthur Daniel G. Evaluating potential drug targets through human loss-of- function genetic variation. bioRxiv preprint first posted online Jan 28, 2019; doi: http://dx.doi.org/10.1101/530881.

Addendum