OPEN ACCESS

Original research

A disorder clinically resembling cystic fibrosis caused by biallelic variants in the *AGR2* gene

Aida Bertoli-Avella ,[1] Ronja Hotakainen,[1] Maryam Al Shehhi,[2] Alice Urzi,[1] Catarina Pareira,[1] Anett Marais,[1] Khoula Al Shidhani,[3] Sumaya Aloraimi,[2] Galina Morales-Torres,[1] Steffen Fisher,[1] Laura Demuth,[1] Laila Abdel Moteleb Selim,[4] Nihal Al Menabawy,[4] Maryam Busehail,[5] Mohammed AlShaikh,[5] Naser Gilani,[6] Dler Nooruldeen Chalabi,[7] Nasser S Alharbi,[8] Majid Alfadhel ,[9,10] Mohammed Abdelrahman,[11] Hanka Venselaar,[12] Nadeem Anjum,[13] Anjum Saeed,[13] Malak Ali Alghamdi,[14] Hamad Aljaedi,[15] Hisham Arabi,[16] Vasiliki Karageorgou,[1] Suliman Khan,[1] Zahra Hajjari,[1] Mandy Radefeldt,[1] Ruslan Al-Ali,[1] Kornelia Tripolszki,[1] Amer Jamhawi,[1] Omid Paknia ,[1] Claudia Cozma,[1] Huma Cheema,[13] Najim Ameziane,[1] Saleh Al-Muhsen,[11] Peter Bauer[1]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/jmedgenet-2021-108150).

For numbered affiliations see end of article.

Correspondence to
Dr Aida Bertoli-Avella, Centogene GmbH, Rostock, Germany; Aida.Bertoli-Avella@centogene.com

AB-A and RH contributed equally.

Received 11 August 2021
Accepted 10 December 2021
Published Online First 24 December 2021

To cite: Bertoli-Avella A, Hotakainen R, Al Shehhi M, *et al*. *J Med Genet* 2022;**59**:993–1001.

ABSTRACT

Purpose We sought to describe a disorder clinically mimicking cystic fibrosis (CF) and to elucidate its genetic cause.

Methods Exome/genome sequencing and human phenotype ontology data of nearly 40 000 patients from our Bio/Databank were analysed. RNA sequencing of samples from the nasal mucosa from patients, carriers and controls followed by transcriptome analysis was performed.

Results We identified 13 patients from 9 families with a CF-like phenotype consisting of recurrent lower respiratory infections (13/13), failure to thrive (13/13) and chronic diarrhoea (8/13), with high morbidity and mortality. All patients had biallelic variants in *AGR2*, (1) two splice-site variants, (2) gene deletion and (3) three missense variants. We confirmed aberrant *AGR2* transcripts caused by an intronic variant and complete absence of *AGR2* transcripts caused by the large gene deletion, resulting in loss of function (LoF). Furthermore, transcriptome analysis identified significant downregulation of components of the mucociliary machinery (intraciliary transport, cilium organisation), as well as upregulation of immune processes.

Conclusion We describe a previously unrecognised autosomal recessive disorder caused by *AGR2* variants. *AGR2*-related disease should be considered as a differential diagnosis in patients presenting a CF-like phenotype. This has implications for the molecular diagnosis and management of these patients. AGR2 LoF is likely the disease mechanism, with consequent impairment of the mucociliary defence machinery. Future studies should aim to establish a better understanding of the disease pathophysiology and to identify potential drug targets.

INTRODUCTION

Cystic fibrosis (CF (OMIM 219700)) is characterised by a triad of chronic obstructive pulmonary disease, exocrine pancreatic insufficiency and elevation of sodium and chloride concentration in sweat caused by biallelic pathogenic variants in the *CFTR* gene.[1 2] The disorder was first described in 1938[3] and was originally known as mucoviscidosis, given the observation that the patients presented abnormally thick mucus (reviewed in a previous work[4]).

Mucus is an essential polymer secreted mainly by specialised cells in the respiratory and digestive tracts. It plays a vital role in the protection against infectious and toxic agents by clearing debris and pathogens through mucociliary clearance. Furthermore, mucus also protects the sensitive epithelial surfaces in the airways and intestine.[5] Mucins are the major macromolecular constituents of epithelia mucus and play a relevant role in health and disease.[6] The secreted or gel-forming mucins are responsible for the viscoelasticity of mucus, with MUC5B and MUC5AC being the major gel-forming mucins present in the airways, where they have different airway clearance functionalities.[7 8] In the small intestine and colon, MUC2 forms the layer of mucus and is responsible for the protection of the gut barrier, the regulation of microbiome homeostasis and the prevention of diseases.[9 10]

Within this study, we present 13 patients from 9 unrelated families suffering from a previously undescribed genetic disorder characterised by recurrent lower respiratory infections, chronic diarrhoea and failure to thrive—a phenotype clinically resembling cystic fibrosis (CF). By performing exome/genome sequencing (ES/GS) and an analysis of our rare disease-centric Bio/Databank, we identified six different *AGR2* biallelic variants as disease-causing for these patients. In mice, Agr2 is relevant for the normal functioning of mucins from the respiratory and gastrointestinal tract,[11–13] suggesting that *AGR2*-related disease might be a novel mucus disorder.

METHODS

This project has been conducted within a diagnostic setting, and as a second step, used deidentified data and samples. Thus, this did not require institutional review board (IRB) approval in our jurisdiction. Written informed consent was obtained from all nine families for genetic studies as well as for scientific publication of anonymised clinical data and clinical photographs. Additionally, the consent declaration included information regarding storage of the data and further processing for research purposes. The informed consent form is available in English and several other languages (https://wwwcentogenecom/downloadshtml).

Exome sequencing (ES)

DNA was extracted using standard methods from dried blood spots (DBS) submitted on filter cards (CentoCard®). Details of the laboratory procedures, bioinformatics analysis and evaluation of the exome data are provided in the Supplemental data. *AGR2* exon numbering was based on transcript NM_006408.3 (8 exons). Variant nomenclature followed standard Human Genome Variation Society recommendations.

Analysis of proprietary Bio/Databank

Our Bio/Databank[14] contains data from 65 005 individuals with ES and/or GS data, as well as corresponding clinical information, which is registered in the Bio/Databank as human phenotype ontology (HPO) terms. A total of 39 756 of these individuals are patients with at least one HPO term used for phenotype description. After identification of *AGR2* as a candidate gene in the first family, the Bio/Databank was investigated for rare biallelic variants in the *AGR2* gene (ExAC/gnomAD<1%). Variants with high or moderate predicted impact on protein structure or function (CADD raw score ≥4) were prioritised. For the identified cases, HPOs, all available clinical information, and test results were reviewed to elucidate the associated phenotype. The ES and/or GS data were reanalysed to investigate whether other variants could contribute to the phenotype of the patients. Referring clinicians were then recontacted (for cases with consent provided).

RNA sequencing and data analysis

Nasal swabs were taken for the probands (families 2, 3, 4, 6, 7 and 8) using ORE-100 RNA collection kits (Steinbrenner Laborsysteme GmbH), and RNA was extracted with the Quick-RNA micro prep kit (Zymogen) following manufacturer's instructions. The TruSeq stranded mRNA kit (Illumina) was used to generate next-generation sequencing (NGS) barcoded libraries. After pooling, the libraries were sequenced on a NextSeq 500 system using the 75 bp paired-end protocol. An average of 34 052 907 reads (>Q30) were obtained for each sample. RNA-seq reads were aligned using two-pass mode with STAR V.2.7.6a[15] against human genome GRCh37/Release 38 (www.gencodegenes.org). The read groups were fixed, and the duplicates were marked using Picard tools V.2.23.8. Counting the reads was performed by featureCounts/subread V.2.0.1.[16] Initial quality control and differential expression analysis was performed with DESeq2_1.32.0,[17] and pathway analysis was performed with ToppGene.[18]

Total RNA was converted to cDNA using reverse transcriptase Superscript IV (Invitrogen). Primers were designed to amplify *AGR2* fragments from exons 1–8, 2–6 and 4–7 (primer sequences available on request). After PCR, aliquots were electrophoresed on 1% agarose gels at 90 V for 90 min, stained with SYBR safe (Invitrogen).

Protein structural analysis

Experimentally solved structures for both the monomeric and dimeric conformations of AGR2 can be found in the Protein Data Bank. These files, PDB-codes 2LNT and 2LNS, respectively, contain the C-terminal residues 36–175. We used the YASARA[19] and WHAT IF software[20] to study these structures.

Ceramide26 quantification in DBS

C26 Ceramide species were quantified in DBS extracts using a method previously described,[21] as well as multiple reaction monitoring mass spectrometry.

A detailed methods description can be found in the Supplemental data.

RESULTS

Clinical description of the affected individuals

Thirteen patients from nine families presented with a similar CF-like phenotype consisting of recurrent respiratory infections, chronic diarrhoea and failure to thrive (online supplemental table 1). The patients' ages ranged from 10 months to 10 years old. Symptoms started usually around the neonatal period. The children suffered from recurrent coughing, wheezy episodes, pneumonia, interstitial lung disease and bronchiectasis. Further episodes of vomiting and chronic diarrhoea led to poor weight gain. Additionally, four patients presented hepatosplenomegaly, and two had cardiovascular abnormalities (mitral valve insufficiency and right heart failure with severe pulmonary hypertension). Most patients had appropriate neurodevelopment; only two cases presented developmental issues. Respiratory and gastrointestinal complaints caused frequent and prolonged hospital admissions. Family history was positive in four families (out of eight) with several similarly affected relatives. Sweat chloride tests and pancreatic elastase results were normal (when performed). Two patients had cultures positive for *Pseudomonas*. Patient III-1 (family 6) had a bronchoalveolar lavage cytology that showed cellular fluid composed of bronchial epithelial cells and alveolar macrophages with strands of thick mucus. Nasal ciliary brush study was done in two patients, with motile cilia seen under light microscopy with 9+2 normal configuration (family 6 and family 7). However, ultrastructural electron microscopy analysis of a nasal brush sample (family 8) detected ciliary abnormalities in 34% of the 189 examined transverse cilia sections. The abnormalities were related to missing central doubles, triplets instead of central doubles with missing dynein arms (inner and/or outer) at peripheral doubles, and duplication of central doublets with missing dynein arms (inner and/or outer) at peripheral doublets (online supplemental figure 1). This result was considered as inconclusive since in patients with a primary ciliary defect, most cilia would be expected to be abnormal.[22]

Additional clinical data can be found in the supplementary data (online supplemental table 1 and Clinical summaries).

Exome and Bio/Databank analyses

In family 1, exome analysis focused on 'diagnostic' genes which did not detect any relevant variant. We then performed an extended exome analysis focusing on genes not yet linked to a human phenotype. Given the positive family history and parental consanguinity, we prioritised homozygous variants. Eight variants were identified (online supplemental table 2), from which the novel missense variant in the *AGR2* gene (NM_006408.3:c.211C>A, p.Pro71Thr) was selected as the best

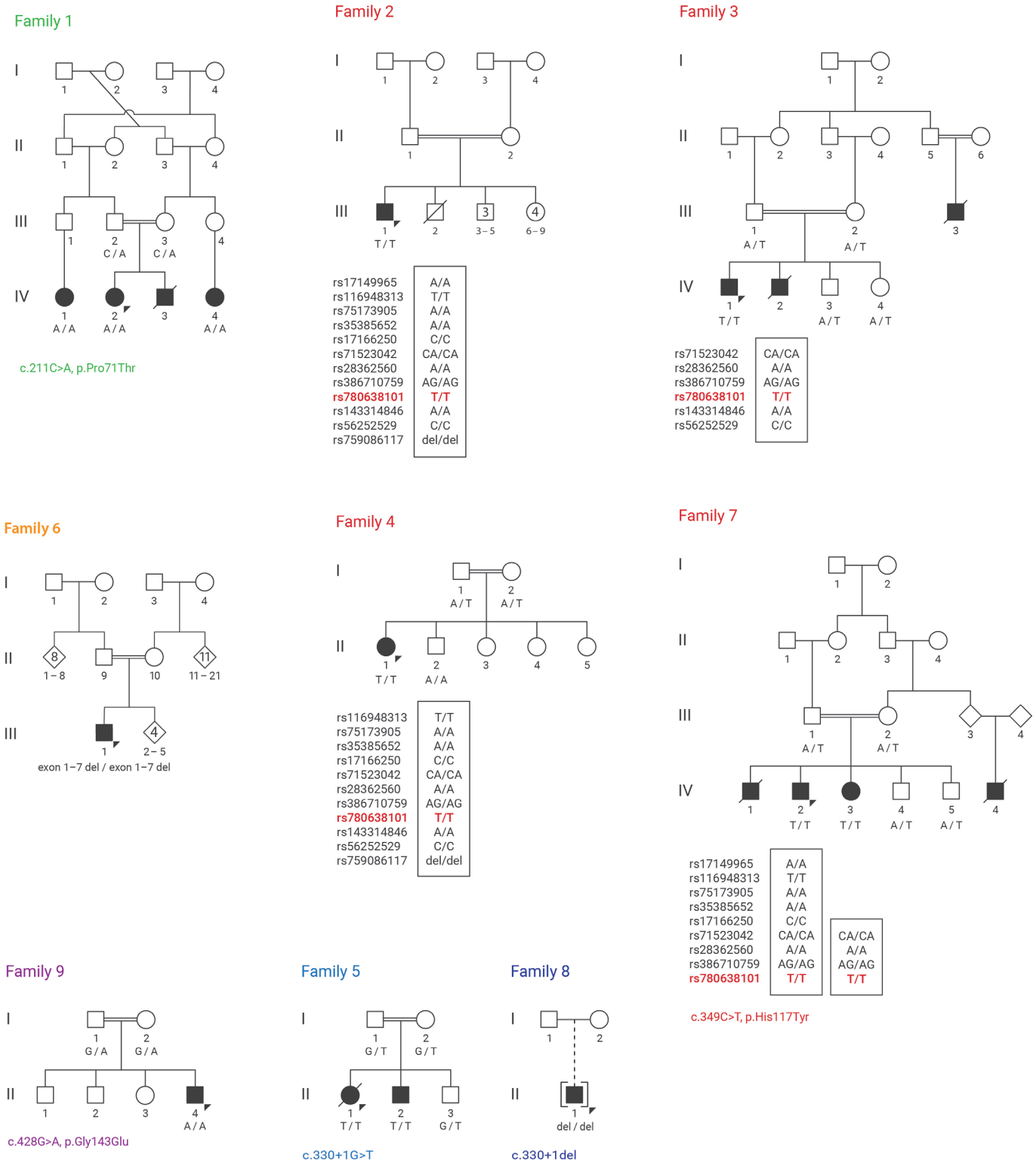


Figure 1 Summarised family trees of the nine families and the identified *AGR2* variants. Variants are colour-coded, the founder missense variant is shown in red font, with the corresponding haplotypes (families 2, 3, 4 and 7). Genotypes are shown below available individuals. *AGR2* genotypes show full co-segregation with the phenotype.

candidate given its rarity, high conservation, and known function and protein localization.[12 13 23] Targeted testing confirmed that the parents were heterozygote carriers, and both affected cousins were homozygotes for the same variant, fully co-segregating with the disease (figure 1).

This finding prompted us to mine our Bio/Databank to identify additional cases with overlapping phenotypes, no genetic diagnosis established, and rare biallelic variants in the *AGR2* gene. Exome/genome data from 39 756 patients was investigated. We prioritised *AGR2* rare, homozygous and compound

heterozygous variants, with predicted impact on the protein. Additionally, we searched for copy number variants (CNVs) affecting the *AGR2* gene.

This resulted in the identification of a total of 13 patients from nine families (figure 1 and online supplemental table 1). All patients presented with very similar phenotypic features, mainly including recurrent lower respiratory tract infections, and no evident immunological abnormalities. Relevant rare homozygous variants are shown in online supplemental table 2. The missense variant in exon 6 of *AGR2* (NM_006408.3:c.349C>T, p.His117Tyr) was present in patients from families 2, 3, 4 and 7. Analysis of the genomic region around *AGR2* indicated a shared haplotype for these families (from rs71523042 to rs780638101, 2.8 Mb), suggesting a common ancestor. Families 2, 4 and 7, from Syrian origin, had a larger shared chromosomal region of approximately 8.2 Mb (figure 1). An additional missense variant was identified in family 9 (c.428G>A, p.Gly143Glu).

We also detected two splice site variants (NM_006408.3:c.330+1G>T and c.330+1del) in families 5 and 8, respectively. The variant was confirmed as heterozygous in the parents and unaffected sibling and homozygous in the similarly affected brother in family 5 (figure 1). The index from family 8 is adopted, and no biological relatives could be tested. The variants affect the canonical splicing site and are predicted to abolish normal exon 5 splicing.

Furthermore, we identified a large homozygous deletion including exon 1 to exon 7 of the *AGR2* gene and affecting the neighbouring gene *AGR3* in family 6 (full gene deletion, genomic coordinates chr7:16,834,229–16,936,407, figure 2A). This prompted us to query our Bio/Databank for potential causative variants in *AGR3*. We identified several individuals with homozygous missense, splicing and nonsense variants in *AGR3*. However, these variants were present in unaffected adults (parents) and patients with variable phenotypes not overlapping with the clinical features of the patients described in this study. Therefore, these data do not support a causative effect of *AGR3* variants.

We also searched other data repositories for variants, such as gnomAD (v2.1.1) and Decipher for *AGR2* variants (SNVs and CNVs). Genes related to autosomal recessive diseases are relatively unconstrained.[24] However, loss of function (LoF) variants in *AGR2* are ultrarare, with not a single individual reported as homozygote in these data repositories. The variants detected in our patients are novel or ultrarare in gnomAD, with high conservation and CADD scores supporting adverse consequences (online supplemental table 3).

RNA sequencing analysis

The AGR2 protein is mainly detected in mucus-secreting organs from the gastrointestinal tract, the respiratory tract and the reproductive system.[23] To assess the effect of the splicing variant and the large deletion and to learn about the putative affected pathways, we performed RNA sequencing using RNA isolated from the nasal mucosa from 6 patients, 4 heterozygote carriers and 11 controls. RNA sequencing confirmed that the c.330+1del variant causes aberrant splicing of *AGR2*, disturbing exon 5 splicing with retention of intronic regions and altering the reading frame, finally leading to LoF (figure 2C and online supplemental figure 2AC). Abnormal splicing was confirmed by targeted *AGR2* RT-PCR (online supplemental figure 2B). We also confirmed that the large deletion detected in the index patient from family 6 leads to a complete loss of *AGR2* transcripts (online supplemental figure 2C). Both patients had very low to nearly no *AGR2* expression (adjusted p value=0.002), confirming that both variants are leading to LoF. For the founder missense variant c.349C>T, there is no evident splicing effect (figure 2C and online supplemental figure 2A-C). Additionally, differential gene expression analysis detected biological processes that were significantly dysregulated in the patients compared with control and carrier samples. Processes such as cell/leucocyte activation and others related to the immune system were transcriptionally upregulated, whereas processes such as microtubule-based movement, process, transport and cilium organisation were significantly downregulated (online supplemental file 1). As an exploratory analysis, we evaluated differential gene expression in the two patients with proven LoF variants (large gene deletion and c.330+1del) compared with the controls. Two relevant mucins (*MUC2* and *MUC5AC*) and *CLCA2* (from the calcium-dependent chloride channel family) are at the top of the downregulated genes (online supplemental file 1).

Protein structural analysis (missense variants)

Based on the published AGR2 protein structure,[25] we investigated the possible effects of the missense variants p.Pro71Thr, p.His117Tyr and p.Gly143Glu. We first examined whether the variants could directly affect dimer formation. The Pro71 residue is semi-buried in the core of the protein, whereas the His117 and Gly143 occur in surface loops but are not immediately at the dimerisation face (figure 3A and B). For Pro71Thr, the proline side chain is slightly larger than threonine; the main differences between these two residues lie in the tendency of the proline residue to make rigid turns that stabilise the protein structure (figure 3C). A substitution of histidine by tyrosine would result in loss of an amino acid that can potentially store electrons (figure 3D). Lastly, Gly143Glu is a clear example of the introduction of a larger side chain that will no longer fit at that position. Glutamic acid with its charged γ-carboxyl side chain would be disruptive and large compared with the glycine with its single hydrogen side chain (figure 3E). This change will affect the local structure simply by restructuring the surrounding residues to remove steric clashes, and may affect interactions with other proteins by changing the surface of AGR2.

Whole blood ceramides (Cer26) analysis

Lipid metabolism imbalances have been consistently reported in patients with CF, as measured in plasma,[26] human primary bronchial cells[27] and mesenchymal stem cells.[28] Among sphingolipids, ceramide is emerging as one of the players of the pulmonary dysfunction in inflammatory lung diseases; enhanced sphingolipid metabolism leads to an increased ceramide content, which in turn contributes to maintaining the chronic inflammatory status.[28] We measured the levels of ceramide26 (Cer26) in DBS from patients with homozygous biallelic variants in *AGR2* (n=5), in molecularly confirmed patients with CF (n=11) and healthy controls (n=10). Patients with CF had slightly lower Cer26 levels, but this difference was not significant. All four *AGR2* patients with severe pulmonary disease consistently showed pathologically elevated levels of Cer26cis, Cer26trans and Cer26total isomers, while the patient with mild respiratory symptoms showed normal levels (family 3). Taken as a whole, the *AGR2* patient group had significantly elevated levels of all measured Cer26 isomers when compared with patients with CF and healthy controls (Cer26total F=10.94, p<0.001; online supplemental figure 3). These findings suggest an altered Cer26 metabolism in patients with *AGR2*-related disease.

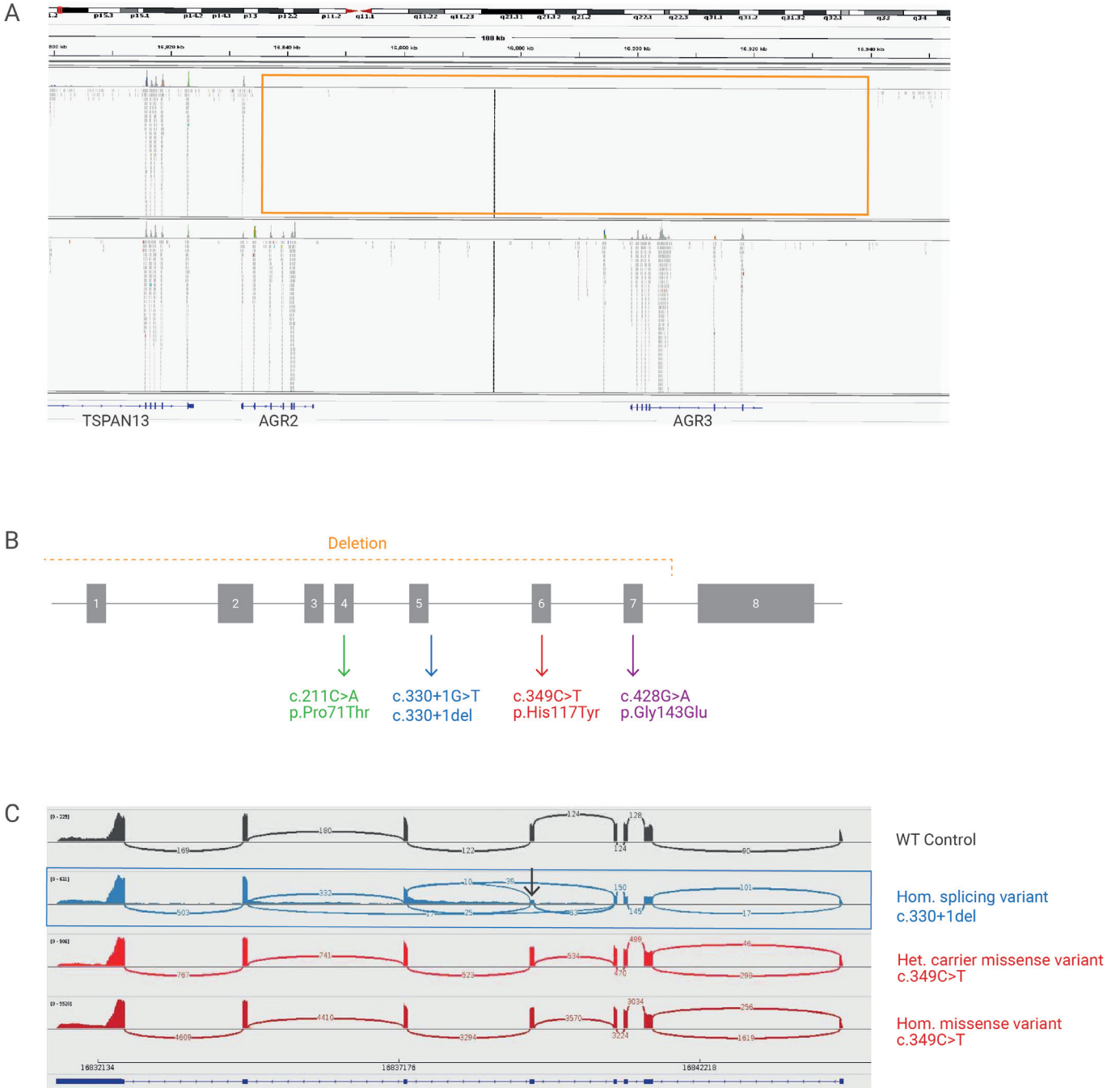


Figure 2 *AGR2* variants identified in the patients and abnormal splicing caused by an intronic variant. (A) The deletion region of *AGR2/3* is shown in the Integrative Genome Viewer (IGV). Reads for the exonic regions of the *AGR2/3* genes can be seen in the control (lower panel), whereas no reads are seen in the index sample III-1 (deleted region is boxed, chr7:16834456–16918247). This deletion was confirmed by qPCR. (B) Schematic representation of *AGR2* gene, with the detected variants shown (font colours match the respective family). (C) Sashimi plots from IGV, illustrating *AGR2* splicing junctions. Arcs represents splice junctions and connect the exons, the number of reads split is displayed across the junction. The variant c.330+1del causes aberrant splicing, note the junctions skipping exon 5 (arrow). See also online supplemental figure 2A–C.

DISCUSSION

By combining ES and Bio/Databank analyses, we identified 13 patients from 9 families with rare homozygous variants in *AGR2* (figures 1 and 2). Three of these variants are very likely leading to a LoF (affecting canonical splicing site and large gene deletion), suggesting that loss of AGR2 is likely causing the phenotype in at least a subset of these patients. Affected individuals presented very early in life with recurrent coughing, wheezing, low tract respiratory infections, chronic diarrhoea and failure to thrive which resembled CF (online supplemental table 1). However, these patients presented normal sweat/elastase tests. Although four of them presented hepatomegaly with undetermined cause, meconium ileus, pancreatic insufficiency, steatorrhea or pancreatitis were not reported in our patients. Thus, on a closer look, CF can be clinically excluded in these patients. This was later confirmed by genetic testing with no relevant variants detected in the *CFTR* gene. From a clinical perspective, it is important to consider AGR2-related disease as a differential diagnosis of patients presenting a CF-like phenotype (and normal sweat/elastase tests).

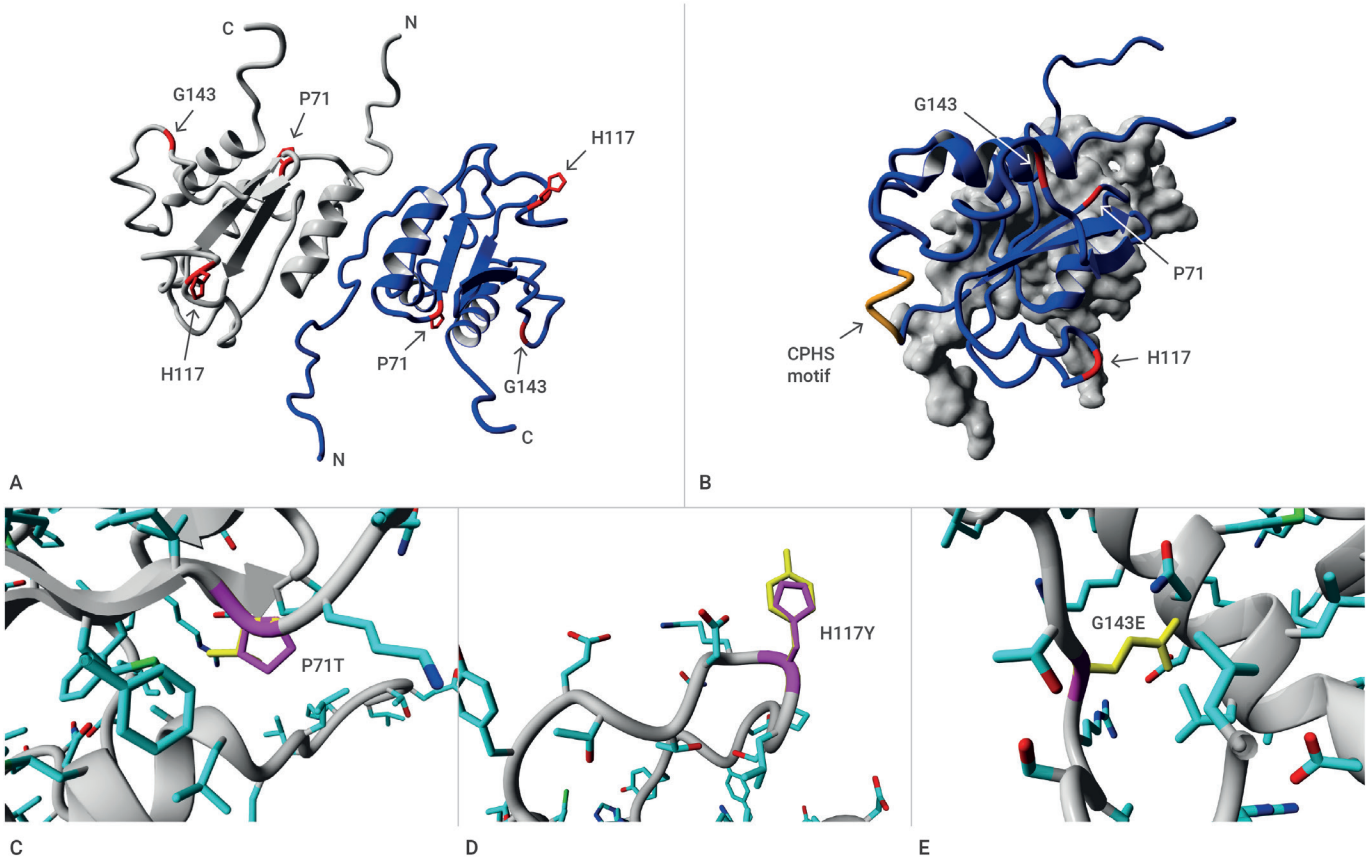


Figure 3 Structural protein analysis of the missense variants. (A) Dimer of the AGR2 residues 36–175. Monomers are individually coloured in grey or blue. Side chains of the residues are not shown, except for the mutated P71, H117 and G143 in red. (B) Overview of AGR2 as seen from the side; one monomer is shown as grey surface only. This view shows the distance between the mutated residues (red side chains) and the putative active site of the protein CPHS-motif (orange). (C) Variant P71T: The proline side chain is shown in magenta; note the attachment of the side chain to its own backbone; the threonine side chain is shown in yellow. Side chains of the protein are coloured by atom type (carbon=cyan, oxygen=red, nitrogen=blue, sulfur=green). The proline side chain is slightly larger than threonine, but the main differences between these two residues lie in the shape of the side chain and proline tendency to make rigid turns that stabilise the protein structure. (D) Variant H117Y: The histidine side chain is shown in magenta, whereas the tyrosine side chain is yellow. Other atoms are coloured as described. The change from histidine to tyrosine indicates a small difference in size, and a different potential for interactions since histidine's side chain can be used for electron storage. (E) Variant G143E: The side chain of the mutant residue glutamic acid will not fit in the same space (note that wild-type glycine does not have a side chain). The change in charge and side chain size will affect the local structure and may affect interactions with other proteins.

Interestingly, the index case from family 8 presented cilia abnormalities in the nasal epithelium (online supplemental figure 1), although these could be secondary cilia changes, as described also in CF and chronic bronchitis.[29] Together with our results from the RNA differential expression and pathway analysis, cilia abnormalities may occur as part of the *AGR2*-related phenotype; however, more patients would need to be examined.

The AGR2 protein is detected at high levels in tissues that secrete mucus or function as endocrine organs, including the respiratory tract, stomach, colon, prostate and small intestine (reviewed in a previous work[23]). At the cellular level, AGR2 is expressed in Paneth and goblet cells (intestine/colon), ciliated cells (airways) and glandular cells (pancreas and prostate, among others)[12 30 31] (Human Protein Atlas, http://www.proteinatlas.org). Subcellularly, AGR2 localises to the lumen of the endoplasmic reticulum (ER), indirectly associates with ER membrane-bound ribosomes, and it is involved in the maintenance of ER homeostasis. Knockdown of AGR2 significantly alters the expression of components of the ER-associated degradation machinery and reduces the ability of cells to cope with acute ER stress.[32]

AGR2 is required for adequate production of intestinal mucin MUC2 and airway mucins MUC5B and MUC5AC.[11–13 33] Mouse Muc5b is required for mucociliary clearance, for controlling infections in the airways and middle ear, and for maintaining immune homeostasis in mouse lungs, whereas Muc5ac is dispensable.[34] On the other hand, MUC2 is the major intestinal mucin,[11 35] and it has been implicated in inflammatory bowel disease and colorectal cancer.[36 37] *Agr2* knockout mice are born healthy but are unable to produce intestinal mucin and are highly susceptible to experimentally induced colitis, with profound weight loss and intestinal bleeding suggesting a role of Agr2 in protection from disease. With ageing, *Agr2* knockout mice develop rectal prolapse, a feature observed in mouse models with colitis.[11] Extensive spontaneous ileitis and colitis were also described in mice lacking Agr2.[12] Schroeder *et al* described a considerable reduction of Muc5ac and Muc5b in the airways of allergen challenged *Agr2*-deficient mice, with abnormal allergen response compared with wild-type controls. This is likely due to impaired mucin transit through the ER, where these mucins were found to accumulate.[13]

Thus, the evidence presented in the studies of the Agr2-deficient mice points to an important role of Agr2 in mucin/mucus production, as well as homeostasis of the respiratory and intestinal tract. These findings are also compatible with the phenotype observed in the patients described in this study and support the hypothesis that the detected variants are likely acting via a LoF mechanism.

Interestingly, one patient presented a large homozygous deletion affecting seven out of eight exons of *AGR2* and the complete *AGR3* gene. The Agr3 protein is detected in ciliated cells in the airway epithelium, and unlike Agr2, it is not induced by ER stress. Mice lacking Agr3 are viable and develop ciliated cells with normal-appearing cilia, which have reduced ciliary beat frequency in the airways, associated with impaired mucociliary clearance in Agr3-deficient animals.[38] No differences in phenotype were observed between the patient with the homozygous *AGR2-AGR3* deletion and the rest of the patients. Cases with *AGR3* biallelic variants and overlapping phenotypes have not been described, to our knowledge.

Pathway transcriptome analysis of nasal samples from patients, carriers and healthy controls detected upregulated biological processes such as immune response, leucocyte activation and immune effector processes which could be related to the recurrent airway infections suffered by the patients. Further, downregulation of cilia-related processes (intraciliary transport, microtubule-based transport, cilium organisation) could also reflect a defective cilia function in the patients. This could be secondary to a primary mucus abnormality.

Interestingly, a recent article reports two siblings with severe congenital enteropathy, but also recurrent respiratory infections and wheezing—a phenotype that overlaps with the clinical features of our patients. The siblings had the same homozygous missense founder variant (c.349C>T, p.His117Tyr) reported in five of our patients. The authors detected very low levels of MUC2 protein in the intestinal wall of the patients.[30] They found high levels of mislocalised AGR2 protein in the epithelial surfaces of gastric and bowel sections of the patients compared with controls. In our transcriptome analysis, no significant differences in mRNA *AGR2* expression were detected when comparing average *AGR2* expression in patients, with carriers, or controls. Combining this information, this suggests that the mutant AGR2 (His117Tyr) protein is produced but has impaired functionality, probably leading to accumulation and mislocalisation at the affected epithelia. Since AGR2 is essential for MUC2/mucus production,[11] loss of AGR2 functioning could explain the low levels of MUC2 protein detected intestinal mucosa of the patients compared with controls.[30] This also aligns with our transcriptome differential gene expression analysis focused on the patients with proven LoF variants, which showed significantly reduced levels of *AGR2*, but also *MUC2* and *MUC2AC*. Interestingly, this analysis also showed a significant reduction of *CLCA2* levels. CLCA2 belongs to the calcium-dependent chloride channel family which are involved in the regulation of electrolytic fluxes and modulate secretion, absorption, cell volume and membrane potential, predominantly expressed in the digestive tract and trachea.[39]

Based on the AGR2 structure published by Patel *et al*,[25] our structural analysis suggests that the missense variants detected in the patients could affect proper AGR2 interactions with other proteins (His117Thr) or might affect the protein structure (Pro71Thr and Gly143Glu). Importantly, His117Thr occurs roughly on the same side of the protein as Cys81. Change of Cys81 into serine is described as causing loss of interaction with Muc2.[11] This is also in line with a recent cellular model that found reduced binding of mutant AGR2 (His117Thr) to MUC2.[30]

Sphingolipid metabolism and ceramide content is altered in airway epithelial cells from patients with CF.[27 28] Ceramides are implicated in inflammation and their accumulation in CF cells was previously demonstrated.[40] In Cftr-deficient mice, ceramide accumulation leads to constitutive age-dependent pulmonary inflammation, death of respiratory epithelial cells, deposits of DNA in bronchi and high susceptibility to severe *Pseudomonas* infections.[40] We directly measured DBS extracts as validated by us previously.[21] In four out of the five *AGR2* patients, ceramide isomers (Cer26) were significantly higher than in healthy controls and patients with CF. Ceramides have been found consistently elevated in the airways of patients with CF and CF animal models, and its accumulation significantly contributes to sustained inflammation and inability to fight lung infections. Conversely, low plasma ceramides have been found in patients with CF, which has been attributed to the abnormal lipid metabolism and malabsorption (reviewed in a previous work[41]). We also detected lower Cer26 levels in patients with CF, although this difference was not significant. Our findings in *AGR2* patients point to a role of ceramides, specifically Cer26 in the AGR2-disease pathophysiology. Whether this is part of a specific disease mechanism or a reflection of a systemic inflammation process, will require further investigation. Cer26 determination could potentially be used as a rapid screening method for *AGR2*-related disease.

In conclusion, we describe a previously unrecognised autosomal recessive disease which is caused by biallelic variants in the *AGR2* gene, likely acting via a LoF mechanism. Paediatric patients presenting a CF-like phenotype should be tested for *AGR2*. Our findings are relevant for the early genetic diagnosis and timely clinical management of the patients—acting as the first step to unravelling the pathophysiology of this disease.

Author affiliations
[1]Medical Reporting & Genomic Research, Centogene GmbH, Rostock, Germany
[2]Child Health Department, The Royal Hospital, Muscat, Oman
[3]Department of Pediatric Pulmonology, The Royal Hospital, Muscat, Oman
[4]Pediatric Neurology and Metabolic division, Cairo University Childrens Hospital, Cairo, Egypt
[5]Department of Pediatrics, Salmaniya Medical Complex, Manama, Bahrain
[6]Farabi Medical Laboratory, Erbil, Iraq
[7]Hawler Medical University College of Medicine, Erbil, Kurdistan, Iraq
[8]Pulmonology Unit, Department of Pediatrics, College of Medicine, King Saud University, Riyadh, Saudi Arabia
[9]Medical Genomic Research department, King Saud Bin Abdulaziz University for Health Sciences (KSAU-HS), King Abdullah International Medical Research Center, Ministry of National Guard Health Affairs (MNG-HA), Riyadh, Saudi Arabia
[10]Genetics and Precision Medicine department (GPM), King Abdullah Specialized Children's Hospital (KASCH), King Abdulaziz Medical City, Ministry of National Guard Health Affairs (MNG-HA), Riyadh, Saudi Arabia
[11]Immunology Research laboratory, Department of Pediatrics, College of Medicine and King Saud University Medical City, King Saud University, Riyadh, Saudi Arabia
[12]Centre for Molecular and Biomolecular Informatics, Radboudumc, Radboud Institute for Molecular Life Sciences, Nijmegen, Gelderland, Netherlands
[13]Department of Pediatric Gastroenterology, Children's Hospital of Lahore, Lahore, Pakistan
[14]Medical Genetics Department, College of Medicine, King Saud University, Riyadh, Saudi Arabia
[15]Department of Pathology, College of Medicine, King Saud University Medical City, King Saud University, Riyadh, Saudi Arabia
[16]Department of Pediatrics, King Abdullah Specialized Children's Hospital (KASCH), King Abdulaziz Medical City, Ministry of National Guard Health Affairs (MNG-HA), Riyadh, Saudi Arabia

Acknowledgements We thank the patients and families for their entusiastic participation and CENTOGENE's colleagues and staff that efficiently supported this project.

Contributors MA-S, KAS, SA, LAMS, NAM, MB, MAS, NG, DNC, NA, MA, MA, NA, AS, MAA, HA, HC and SA referred the patients, and acquired and provided clinical data. AM, AU, SK, CP, OP, KT, RA, MR, ZH, AJ and VK processed the samples, analysed the NGS, Sanger or RNA-seq data. GMT, SF, LD and CC performed experiments and analysed the biochemistry data. HV performed the structural protein analysis. RH and AMB-A wrote the manuscript. AMB-A and PB designed and supervised the research project and the manuscript (both authors are guarantors of this publication).

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests AB-A, RH, AU, CP, AM, GM-T, SF, LD, VK, SK, ZH, MR, RA-A, KT, AJ, OP, CC, NA and PB are employees of CENTOGENE GmbH. None of the other authors declared a potential conflict of interest.

Patient consent for publication Not applicable.

Ethics approval The current project has been conducted within a diagnostic setting and in a second step, used deidentified data and samples, and thus did not require IRB approval in our jurisdiction. Informed consents were obtained, including specific consents for scientific publications. The form contains a section for consent for genetic testing related to the disease(s) of the patient and consent for research (related to the main concern, but implicating genes not yet associated to human diseases). Additionally, the consent declaration included information regarding storage of the data and further processing for research purposes. The informed consent form is available in English and several other languages at https://www.centogene.com/downloads. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Data are available on request.

ORCID iDs
Aida Bertoli-Avella http://orcid.org/0000-0001-9544-1877
Majid Alfadhel http://orcid.org/0000-0002-9427-7240
Omid Paknia http://orcid.org/0000-0003-1173-5363

REFERENCES

1 Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M, Tsui LC. Identification of the cystic fibrosis gene: genetic analysis. *Science* 1989;245:1073–80. doi:10.1126/science.2570460
2 Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, Zielenski J, Lok S, Plavsic N, Chou JL, Drumm M, Iannuzzi M, Collins F, Tsui LC. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* 1989;245:1066–73. doi:10.1126/science.2475911
3 ANDERSEN DH. Cystic fibrosis of the pancreas and its relation to celiac disease: a clinical and pathologic study. *Am J Dis Child* 1938;56:344–99. doi:10.1001/archpedi.1938.01980140114013
4 Asay LD. Cystic fibrosis. *Calif Med* 1965;102:292–300.
5 Hansson GC. Mucus and mucins in diseases of the intestinal and respiratory tracts. *J Intern Med* 2019;285:479–90. doi:10.1111/joim.12910
6 Rose MC, Voynow JA. Respiratory tract mucin genes and mucin glycoproteins in health and disease. *Physiol Rev* 2006;86:245–78. doi:10.1152/physrev.00010.2005
7 Ermund A, Meiss LN, Rodriguez-Pineiro AM, Bähr A, Nilsson HE, Trillo-Muyo S, Ridley C, Thornton DJ, Wine JJ, Hebert H, Klymiuk N, Hansson GC. The normal trachea is cleaned by MUC5B mucin bundles from the submucosal glands coated with the MUC5AC mucin. *Biochem Biophys Res Commun* 2017;492:331–7. doi:10.1016/j.bbrc.2017.08.113
8 Ostedgaard LS, Moninger TO, McMenimen JD, Sawin NM, Parker CP, Thornell IM, Powers LS, Gansemer ND, Bouzek DC, Cook DP, Meyerholz DK, Abou Alaiwa MH, Stoltz DA, Welsh MJ. Gel-Forming mucins form distinct morphologic structures in airways. *Proc Natl Acad Sci U S A* 2017;114:201703228. doi:10.1073/pnas.1703228114
9 Johansson MEV, Sjövall H, Hansson GC. The gastrointestinal mucus system in health and disease. *Nat Rev Gastroenterol Hepatol* 2013;10:352–61. doi:10.1038/nrgastro.2013.35
10 Liu Y, Yu X, Zhao J, Zhang H, Zhai Q, Chen W. The role of MUC2 mucin in intestinal homeostasis and the impact of dietary components on MUC2 expression. *Int J Biol Macromol* 2020;164:884–91. doi:10.1016/j.ijbiomac.2020.07.191
11 Park S-W, Zhen G, Verhaeghe C, Nakagami Y, Nguyenvu LT, Barczak AJ, Killeen N, Erle DJ. The protein disulfide isomerase AGR2 is essential for production of intestinal mucus. *Proc Natl Acad Sci U S A* 2009;106:6950–5. doi:10.1073/pnas.0808722106
12 Zhao F, Edwards R, Dizon D, Afrasiabi K, Mastroianni JR, Geyfman M, Ouellette AJ, Andersen B, Lipkin SM. Disruption of Paneth and goblet cell homeostasis and increased endoplasmic reticulum stress in Agr2-/- mice. *Dev Biol* 2010;338:270–9. doi:10.1016/j.ydbio.2009.12.008
13 Schroeder BW, Verhaeghe C, Park S-W, Nguyenvu LT, Huang X, Zhen G, Erle DJ. Agr2 is induced in asthma and promotes allergen-induced mucin overproduction. *Am J Respir Cell Mol Biol* 2012;47:178–85. doi:10.1165/rcmb.2011-0421OC
14 Trujillano D, Oprea G-E, Schmitz Y, Bertoli-Avella AM, Abou Jamra R, Rolfs A. A comprehensive global genotype-phenotype database for rare diseases. *Mol Genet Genomic Med* 2017;5:66–75. doi:10.1002/mgg3.262
15 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. Star: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29:15–21. doi:10.1093/bioinformatics/bts635
16 Liao Y, Smyth GK, Shi W. featureCounts: an efficient General purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014;30:923–30. doi:10.1093/bioinformatics/btt656
17 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol* 2014;15:550. doi:10.1186/s13059-014-0550-8
18 Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res* 2009;37:W305–11. doi:10.1093/nar/gkp427
19 Krieger E, Koraimann G, Vriend G. Increasing the precision of comparative models with YASARA NOVA-a self-parameterizing force field. *Proteins* 2002;47:393–402. doi:10.1002/prot.10104
20 Vriend G. What if: a molecular modeling and drug design program. *J Mol Graph* 1990;8:52–6. doi:10.1016/0263-7855(90)80070-V
21 Cozma C, Iurașcu M-I, Eichler S, Hovakimyan M, Brandau O, Zielke S, Böttcher T, Giese A-K, Lukas J, Rolfs A. C26-Ceramide as highly sensitive biomarker for the diagnosis of Farber disease. *Sci Rep* 2017;7:6149. doi:10.1038/s41598-017-06604-2
22 Papon JF, Coste A, Roudot-Thoraval F, Boucherat M, Roger G, Tamalet A, Vojtek AM, Amselem S, Escudier E. A 20-year experience of electron microscopy in the diagnosis of primary ciliary dyskinesia. *Eur Respir J* 2010;35:1057–63. doi:10.1183/09031936.00046209
23 Delom F, Mohtar MA, Hupp T, Fessart D. The anterior gradient-2 interactome. *Am J Physiol Cell Physiol* 2020;318:C40–7. doi:10.1152/ajpcell.00532.2018
24 Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, England EM, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B, Lek M, Ware JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferriera S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S, Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K, Tolonen C, Wade G, Talkowski ME, Neale BM, Daly MJ, MacArthur DG, Genome Aggregation Database Consortium. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020;581:434–43. doi:10.1038/s41586-020-2308-7
25 Patel P, Clarke C, Barraclough DL, Jowitt TA, Rudland PS, Barraclough R, Lian L-Y. Metastasis-Promoting anterior gradient 2 protein has a dimeric thioredoxin fold structure and a role in cell adhesion. *J Mol Biol* 2013;425:929–43. doi:10.1016/j.jmb.2012.12.009
26 Guilbault C, Wojewodka G, Saeed Z, Hajduch M, Matouk E, De Sanctis JB, Radzioch D. Cystic fibrosis fatty acid imbalance is linked to ceramide deficiency and corrected by fenretinide. *Am J Respir Cell Mol Biol* 2009;41:100–6. doi:10.1165/rcmb.2008-0279OC
27 Loberto N, Mancini G, Bassi R, Carsana EV, Tamanini A, Pedemonte N, Dechecchi MC, Sonnino S, Aureli M. Sphingolipids and plasma membrane hydrolases in human primary bronchial cells during differentiation and their altered patterns in cystic fibrosis. *Glycoconj J* 2020;37:623–33. doi:10.1007/s10719-020-09935-x
28 Zulueta A, Peli V, Dei Cas M, Colombo M, Paroni R, Falleni M, Baisi A, Bollati V, Chiaramonte R, Del Favero E, Ghidoni R, Caretti A. Inflammatory role of extracellular sphingolipids in cystic fibrosis. *Int J Biochem Cell Biol* 2019;116. doi:10.1016/j.biocel.2019.105622
29 Tilley AE, Walters MS, Shaykhiev R, Crystal RG. Cilia dysfunction in lung disease. *Annu Rev Physiol* 2015;77:379–406. doi:10.1146/annurev-physiol-021014-071931
30 Al-Shaibi AA, Abdel-Motal UM, Hubrack SZ, Bullock AN, Al-Marri AA, Agrebi N, Al-Subaiey AA, Ibrahim NA, Charles AK, Elawad M, Uhlig HH, Lo B, COLORS in IBD-Qatar Study Group. Human AGR2 deficiency causes mucus barrier dysfunction and infantile

inflammatory bowel disease. *Cell Mol Gastroenterol Hepatol* 2021;12:1809–30. doi:10.1016/j.jcmgh.2021.07.001
31 Thul PJ, Åkesson L, Wiking M, Mahdessian D, Geladaki A, Ait Blal H, Alm T, Asplund A, Björk L, Breckels LM, Bäckström A, Danielsson F, Fagerberg L, Fall J, Gatto L, Gnann C, Hober S, Hjelmare M, Johansson F, Lee S, Lindskog C, Mulder J, Mulvey CM, Nilsson P, Oksvold P, Rockberg J, Schutten R, Schwenk JM, Sivertsson Åsa, Sjöstedt E, Skogs M, Stadler C, Sullivan DP, Tegel H, Winsnes C, Zhang C, Zwahlen M, Mardinoglu A, Pontén F, von Feilitzen K, Lilley KS, Uhlén M, Lundberg E. A subcellular map of the human proteome. *Science* 2017;356. doi:10.1126/science.aal3321. [Epub ahead of print: 26 05 2017].
32 Higa A, Mulot A, Delom F, Bouchecareilh M, Nguyên DT, Boismenu D, Wise MJ, Chevet E. Role of pro-oncogenic protein disulfide isomerase (PDI) family member anterior gradient 2 (AGR2) in the control of endoplasmic reticulum homeostasis. *J Biol Chem* 2011;286:44855–68. doi:10.1074/jbc.M111.275529
33 Fahy JV, Dickey BF. Airway mucus function and dysfunction. *N Engl J Med Overseas Ed* 2010;363:2233–47. doi:10.1056/NEJMra0910061
34 Roy MG, Livraghi-Butrico A, Fletcher AA, McElwee MM, Evans SE, Boerner RM, Alexander SN, Bellinghausen LK, Song AS, Petrova YM, Tuvim MJ, Adachi R, Romo I, Bordt AS, Bowden MG, Sisson JH, Woodruff PG, Thornton DJ, Rousseau K, De la Garza MM, Moghaddam SJ, Karmouty-Quintana H, Blackburn MR, Drouin SM, Davis CW, Terrell KA, Grubb BR, O'Neal WK, Flores SC, Cota-Gomez A, Lozupone CA, Donnelly JM, Watson AM, Hennessy CE, Keith RC, Yang IV, Barthel L, Henson PM, Janssen WJ, Schwartz DA, Boucher RC, Dickey BF, Evans CM. Muc5B is required for airway defence. *Nature* 2014;505:412–6. doi:10.1038/nature12807
35 Maldonado-Contreras AL, McCormick BA. Intestinal epithelial cells and their role in innate mucosal immunity. *Cell Tissue Res* 2011;343:5–12. doi:10.1007/s00441-010-1082-5
36 Miner-Williams WM, Moughan PJ. Intestinal barrier dysfunction: implications for chronic inflammatory conditions of the bowel. *Nutr Res Rev* 2016;29:40–59. doi:10.1017/S0954422416000019
37 Velcich A, Yang W, Heyer J, Fragale A, Nicholas C, Viani S, Kucherlapati R, Lipkin M, Yang K, Augenlicht L. Colorectal cancer in mice genetically deficient in the mucin MUC2. *Science* 2002;295:1726–9. doi:10.1126/science.1069094
38 Bonser LR, Schroeder BW, Ostrin LA, Baumlin N, Olson JL, Salathe M, Erle DJ. The endoplasmic reticulum resident protein AGR3. required for regulation of ciliary beat frequency in the airway. *Am J Respir Cell Mol Biol* 2015;53:536–43. doi:10.1165/rcmb.2014-0318OC
39 Agnel M, Vermat T, Culouscou JM. Identification of three novel members of the calcium-dependent chloride channel (CACC) family predominantly expressed in the digestive tract and trachea. *FEBS Lett* 1999;455:295–301. doi:10.1016/S0014-5793(99)00891-1
40 Teichgräber V, Ulrich M, Endlich N, Riethmüller J, Wilker B, De Oliveira-Munding CC, van Heeckeren AM, Barr ML, von Kürthy G, Schmid KW, Weller M, Tümmler B, Lang F, Grassme H, Döring G, Gulbins E. Ceramide accumulation mediates inflammation, cell death and infection susceptibility in cystic fibrosis. *Nat Med* 2008;14:382–91. doi:10.1038/nm1748
41 Ghidoni R, Caretti A, Signorelli P. Role of sphingolipids in the pathobiology of lung inflammation. *Mediators Inflamm* 2015;2015:1–19. doi:10.1155/2015/487508

Supplementary data

A disorder clinically resembling cystic fibrosis caused by biallelic variants in the *AGR2* gene

Methods

Exome Sequencing (ES)

ES was performed as previously described [1]. In short, Twist Human Core Exome Plus, the Nextera Rapid Capture Exome Kit (Illumina, San Diego, CA) or the SureSelect Human All Exon kit (Agilent, Santa Clara, CA) were used for the enrichment, and a HiSeq4000 (Illumina) instrument for the sequencing with the 150 paired-end protocol to yield at least 20x coverage of depth for >98% of the target region. An in-house bioinformatics pipeline, including read alignment to human genome reference (hg19), variant calling (single nucleotide and small deletion/insertion variants) and variant annotation with publicly available databases, was used [1].

All provided clinical data, family history, consanguinity, disease onset/course, and available test results were considered. The type of variant and frequency in public databases, such as gnomAD, ExAc, as well as disease-centered databases (HGMD and CentoMD®), were considered.

Sanger Validation and Co-Segregation Analysis

The *AGR2* exons containing the variants were amplified (primers available upon request) and Sanger-sequenced in both forward and reverse direction on a 3730xl sequencer (Thermo Fisher Scientific, Waltham, MA). The copy number variant (deletion) was confirmed by quantitative PCR assays (qPCR), targeting several exons within the copy number variant and

1-2 additional fragments outside the deletion. Products were run in a LightCycler 480 II (Roche).

Ceramide26 Quantification in Dried Blood Spots

C26 Ceramide species were quantified in dried blood spots extract using a method previously described[2], using Multiple reaction monitoring- mass spectrometry.

Perforations of 3.2 mm in diameter were cut using a DBS puncher (Perkin Elmer LAS, Germany) and placed in deep well plate (Thermo Scientific, Germany). 50 µL extraction solution (DMSO: water, 1:1) and 100 µL internal standards solution in ethanol were added on top. Plate was sealed and placed in an incubator (Heidolph, Germany) for 30 minutes at 37 °C under agitation at 700 rpm. After incubation, the plate was sonicated for 10 minutes at maximum power and then the liquid was transferred to an AcroPrep Filter Plate with PTFE membrane (PALL, Germany) placed on a 96 well V-shape bottom plate (VWR, Germany). The samples were filtrated by centrifugation for 5 minutes at 3500 rpm in a Hermle Z300 plate centrifuge (Hermle Labortehnik, Germany). The clear extract was measured using LC/MRM-MS on a Waters Acquity UPLC (Waters, UK) coupled with an ABSciex 5500 TripleQuad mass spectrometer (ABSciex, Germany). Chromatographic run was performed on a C8, 3 µm, Column, 50 × 2.1 mm (ACE, ACE, Germany) using a flow rate of 0.9 mL/min preheated at 60 °C. The analytes were eluted using a gradient with type 6 curve from 40% A (50 mM formic acid in water) to 100% B (50 mM formic acid in acetone: acetonitrile vol. 1:1). Multiple reaction monitoring- mass spectrometry (MRM-MS) analyses were performed in positive ion mode using the following parameters: CUR gas 10 psi, IS voltage 5 kV, CAD 8 psi, cone temperature 200 °C, GS1 45 psi, GS2 60 psi, EP 10 V.

Clinical Description of the Affected Individuals

Family 1, individuals IV-2, IV-1 and IV-4

The index patient (IV-2) is a female born to healthy consanguineous parents from Oman (Figure 1A). The family history is positive, with three relatives having a similar phenotype of recurrent respiratory infections and failure to thrive. One affected sibling deceased with a similar phenotype. The index has presented recurrent lower respiratory tract infections, wheezing, and failure to thrive since infancy, which caused regular hospital admissions. A chest X-ray showed hyperinflated lungs and mild peribronchial wall thickening. A high-resolution computed tomography (HRCT) showed mild bilateral bronchiectasis and mosaic pattern. Repeated thorax CT scan showed diffuse mosaic pattern, bronchiectasis, and hilar lymphadenopathy. Given the clinical suspicions of cystic fibrosis, a sweat chloride test was performed, which was normal. Exome sequencing, indicated for diagnostic purposes, did not identify any relevant variant among known disease genes. Then, a second analysis was performed including genes not yet associated to a human phenotype.

Individual IV-1 (cousin). She has been suffering from recurrent respiratory infections and wheezy episodes since the neonatal period, with multiple hospital admissions. Between these episodes she has been having daily wet coughing, as well as poor weight gain. Immunological work up and cystic fibrosis investigations were all normal. A HRCT of the thorax at the age of nine months showed features suggestive of mild basal bronchiectasis and of bronchiolitis obliterans. A bronchoscopy, echocardiography, and a barium swallow test were normal. A tuberculosis workup was negative. A lung biopsy at the age of two years showed mild peri bronchial lymphocytic inflammation with no evidence of lung fibrosis. Currently, she has poor weight gain, and suffers from chronic coughing, and exertional dyspnea with intermittent wheezing. The most recent HRCT showed a diffuse mosaic pattern with bronchial wall thickening, mild bronchiectasis, fibrotic bands, mild collapse, few pulmonary nodules, and hilar lymphadenopathy. Detailed studies: Microbiology surveillance (sputum and throat swap) were negatives (no culture growth). Bronchoalveolar lavage: sample of whitish color, semi-mucoid, by microscopy groups of reactive respiratory epithelial cells were observed, mucus

with admixture of neutrophil and lymphocytes, no fungal elements are demonstrated by GMS, the results are consistent with an inflammatory process. No bacterial growth. Lung biopsy: The sections show four profiles of lung covered by pleura. The pleura is edematous and shows prominent vessels. There are some adhesions. The lung parenchyma appears normal with slight over distension of the air spaces, there is no subpleural fibrosis. However, at the resection margin the parenchyma is collapsed. No cartilage containing bronchi are present within the specimen. There is a focal lymphocytic and histiocytic infiltrate around bronchioles that is nodular. There is no diffuse inflammatory cell infiltrate in the interstitial or air spaces. PAS staining is negative. There are collections of foamy cells in the lumen of one of the bronchioles. There is no alveolar proteinosis and no type 2 pneumocyte hyperplasia is seen. The appearances are not specific in these sections. There is some mild peri bronchial lymphocytic infiltration and the presence of foamy cells within the lumina suggest possible mild bronchial obstruction. The appearances do not suggest surfactant protein deficiency and immunostaining for surfactant protein B is normal. The vasculature appears unremarkable on elastic staining. Conclusion: Mild peri bronchial lymphocytic inflammation.

Individual IV-4 (cousin). Since the neonatal period, she has presented with persistent wet coughing, recurrent wheezing episodes, and dyspnea. She had multiple admissions due to persistent lower respiratory infections. A physical examination revealed no dysmorphism, no clubbing, bilateral crackles at chest auscultation, and delayed motor development with right hemiplegia. A head CT scan displayed a left basal ganglia and left thalamus smaller than the right side, indicating atrophic changes (Wallerian degeneration) probably due to an old ischemic insult. Currently, she suffers from intermittent productive coughing. After completing an extensive rehabilitation plan for the hemiplegia, she can walk and run relatively normal. Her HRCT has showed segmental areas of mosaic perfusion with bronchial wall thickening, mild bronchial dilatation, and tiny nodules with tree-in-bud appearance. Mediastinal and hilar lymphadenopathy were seen. The bronchoscopy was normal with no

airway anomalies noted, but a clear secretion was observed all over the airways. Bronchoalveolar lavage: three smears of bronchial lavage examined, they show numerous macrophages and mixed inflammatory cell infiltrate. Few reactive bronchial epithelial cells were seen, and no malignant cells noted. Microbiology: no growth. Microbiology surveillance by sputum culture and throat swap was negative. Laboratory investigations including targeted genetic analysis excluded primary immunodeficiency, cystic fibrosis, primary ciliary dyskinesia, aspiration syndrome and mitochondrial cytopathy.

Family 2, individual III-1

The index is a male born to consanguineous parents, who are from Syrian origin. His birth weight was 3.2 kg. The patient had history of passing frequent loose stools since birth, for which his formula milk has been changed several times with no improvement. He was admitted to the hospital for acute gastroenteritis and pneumonia during the neonatal period. The patient was again readmitted after three weeks from discharge with history of still passing frequent loose watery stools that was greenish in color. In addition, the patient had also history of chronic cough that was productive in nature, associated with vomiting whitish sputum. To rule out cystic fibrosis, sweat chloride test was performed that came normal (27 mmol/L) and repeated (33 mmol/L).

The patient was again readmitted a few months later with the impression of pneumonia and receive a course of antibiotics, upper GI study was requested and was found to have severe GERD. The patient had frequent follow ups with the respiratory and GI team and was been managed as a case of hyperactive airway disease, cow's milk allergy and severe GERD.

Upon examination patient was noticed to have subtle dysmorphic features, prominent forehead, upslanting palpebral fissures, and thin upper lips.

Family 3, individual IV-1

The index is a male born to consanguineous asymptomatic parents from Iraq. His brother died during infancy after a long-term hospitalization due to chronic diarrhea, vomiting, and renal failure. He has a similarly affected cousin (deceased, Figure 1A). The index presented soon after birth with severe vomiting and diarrhea leading to admission to the intensive care unit.

Family 4, individual II-1

The index is a female born to healthy consanguineous parents from Syrian origin. She was born at full term with normal birth weight (3 kg). It was noticed that she presented with persistent coughing and she was admitted with the impression of severe pneumonia and treated with a course of antibiotics. During early childhood, she was examined for speech delay and was found to have sensorineural hearing impairment. Currently, the patient is still suffering from recurrent respiratory infections and a chronic productive cough, which regularly requires hospital admission. Exome sequencing detected a homozygous likely pathogenic variant in the *SLC26A4* gene with the diagnosis of autosomal recessive deafness type 4, which explains the hearing impairment. However, this finding did not clarify the cause of the respiratory symptoms.

According to the parents, the patient is still having recurrent lower respiratory infections requiring hospital admissions and chronic cough that is productive in nature despite being on prophylactic antibiotics weekly.

Family 5, individual II-1

The index is a deceased female. Her consanguineous parents are from Egypt. The index's younger brother is similarly affected (Figure 1). The index presented with recurrent lower respiratory tract infection, interstitial lung disease, hepatosplenomegaly, hypotonia, and global developmental delay. The recurrent lower respiratory tract infections started at the age of eight months, leading to prolonged hospitalizations in the intensive care unit. She also had

right-sided heart failure, which was thought to be secondary to the respiratory condition. A brain MRI and EMG (upper and lower limbs) were normal. The thorax CT scan showed bilateral patchy ground glass haze and right lower lobe patchy consolidation with trans bronchial spread and intervening areas of hyperventilation. In addition, progressive diffuse reticulo - nodular infiltrates and bilateral atelectatic bands were detected. An echocardiography showed right ventricular and right atrial dilatation, with tricuspid regurgitation and severe pulmonary hypertension. An ultrasound of the abdomen revealed an enlarged liver with homogenous echogenic parenchymal texture. The immunological profile was normal, apart from a slightly low percentage of CD4+ T-cells. A sweat chloride test was normal. Her brother has presented with recurrent lower respiratory infections that led to multiple hospitalizations. He has had normal neurodevelopment with slightly delayed motor milestones. A physical examination revealed pectus carinatum and hepatomegaly of approximately 3 cm. A CT scan of the thorax showed well defined patch areas of ground glass appearance and scattered consolidations in both lungs.

Family 6, individual III-1

The index is a male born to asymptomatic consanguineous parents, who are residents of Oman, with no family history of similar clinical picture. He presented with recurrent lower respiratory tract infections since infancy. Later, he had recurrent episodes of ear infection and otorrhea, which did not respond well to antibiotic therapy. He also had recurrent upper tract respiratory infections and coughing. A nasal ciliary brush study was done with motile cilia seen under light microscopy. Unfortunately, the sample was not adequate for electron microscopy. His bronchoscopy test was completely normal. Bronchoalveolar lavage (BAL) culture was positive for pseudomonas. The BAL cytology showed cellular fluid composed of bronchial epithelial cells and alveolar macrophages. Strands of thick mucus were seen in a

background containing many neutrophils. No alveolar cast, micro-organisms or atypical cells were detected. However, scattered lipid laden macrophages were observed.

Family 7, individuals IV-2 and IV-3

The index is a male from Syria, born to consanguineous parents. He was born after an uneventful pregnancy and delivery (at term). The index presented soon after birth with chronic diarrhea, poor weight gain, and mild hepatomegaly. During early childhood, he had recurrent otorrhea and middle ear infections, which did not respond well to antibiotic therapy. He also had upper tract respiratory infections and dried cough. He was evaluated by the immunology team (upon clinical suspicion of primary immunodeficiency); however, all lab tests were normal. A chest CT scan showed randomly distributed bilateral airspace consolidations, some are nodular with no cavitation, and mild bronchial wall thickening. Multiple enlarged mediastinal and hilar lymph nodes, mild bilateral pleural effusion with no pericardial effusion or pneumothorax. Microbiology cultures detected pseudomonas in both airway secretion and ear discharge. Nasal brush examination revealed rare epithelial cells with cilia, with 9+2 normal configuration. Ultrastructural electron microcopy examination was not possible due to a suboptimal quality of the sample. The liver shows diffuse low density could be due to fatty infiltration. Abdominal ultrasound showed a mildly enlarged liver with homogeneous parenchyma and no focal lesion. Screening for several infectious diseases had negative results as well (tuberculosis, CMV, EBV, HIV). Other test performed included sweat chloride, pancreatic elastase, Alpha 1 antitrypsin, and nasal brush test (also normal). Laboratory testing resulted normal excluded intestinal parasite infections. Multiple blood, urine and stool cultures were negative as well.

There are other similarly affected relatives. A male sibling deceased during infancy with a clinical picture of chronic diarrhea and progressive respiratory disease. A cousin deceased

with a progressive respiratory disease. A female sibling (IV-3) is affected with chronic diarrhea and recurrent lower tract respiratory infections (Figure 1A).

Family 8, individual II-1

The index is a male, presenting since early childhood with recurrent low tract respiratory infections and persistent rhinorrhea. He is adopted and history of his biological relatives is not available. He had persistent vomiting and dysphagia with hard food. Esophagogastroduodenoscopy with histopathology study, and echo were normal. CT chest showed bronchiectatic changes and persistent segmental collapse in the left lower lobe. Also, persistent direct hyperbilirubinemia with mild hepatomegaly, were detected with other liver function tests within normal limits. Furthermore, he is followed by a sleep therapist for obstructive sleep apnea likely due to upper airway obstruction.

Immunological workup revealed normal results including immunoglobulin, lymphocyte subset analysis and oxidative burst test. Ciliary abnormalities were detected in 34% of the examined cilia, the abnormalities were related to missing central doubles, triplets instead of central doubles with missing dynein arms (inner and/or outer) at peripheral doubles, and duplication of central doublets with missing dynein arms (inner and/or outer) at peripheral doublets (Supplementary figure). Although suggestive of primary ciliary dyskinesia, in most patients with a true ciliary defect most cilia are abnormal [3]. Additional testing included throat swab cultures (Escherichia coli in 3 different occasions).

Family 9, individual II-4

The index is a male, born to consanguineous parents from Pakistan. During the neonatal period, he presented respiratory distress. A few weeks later, he developed respiratory distress followed by inter costal and sub costal recessions and lethargy. He was admitted to the hospital for more than a month with pneumonia. Later, he presented loose stools and frequent

episodes of dehydration. He persisted to have course of respiratory symptoms and marked weight loss due to diarrheal episodes. His neurodevelopment is normal.

Supplementary Table 1. Clinical characteristics of patients identified with ***AGR2*** homozygous variants (NM_006408.3)

	Fam. 1, IV-1 (2427168)	**Fam. 1, IV-2** (2427168)	**Fam.1, IV-4** (2427168)	**Fam. 2, III-1** (2438720)	**Fam. 3, IV-1** (2399903)	**Fam. 4, II-1** (2337168)	**Fam. 5, II-1** (2151518)	**Fam. 5, II-2** (2151518)	**Fam. 6, III-1** (2451078)	**Family 7** (2518771)	**Family 7** (2518771)	**Family 8** (2508357)	**Family 9** (2534592)
AGR2 variant - NM_006408.3	c.211C>A p.Pro71Thr exon 4	c.211C>A p.Pro71Thr exon 4	c.211C>A p.Pro71Thr exon 4	c.349C>T p.His117Tyr exon 6	c.349C>T p.His117Tyr exon 6	c.349C>T p.His117Tyr exon 6	c.330+1G>T intron 5	c.330+1G>T intron 5	Large deletion (exon 1- 7 chr7:16834456-16918247)	c.349C>T p.His117Tyr exon 6	c.349C>T p.His117Tyr exon 6	c.330+1del intron 5	c.428G>A p.Gly143Glu (exon 7)
Current life stage	Childhood	Childhood	Childhood	Childhood	Infancy	Childhood	Deceased	Childhood	Childhood	Childhood	Early childhood	Childhood	Early childhood
Sex	Female	Female	Female	Male	Male	Female	Female	Male	Male	Male	Female	Male	Male
Consanguinity	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unknown	Yes
Geographical region	Oman	Oman	Oman	Bahrain (of Syrian origin)	Iraq	Bahrain (of Syrian origin)	Egypt	Egypt	Oman	Saudi Arabia	Saudi Arabia	Saudi Arabia	Pakistan
Family history	Yes (cousins)	Yes (cousins and sibling)	Yes (cousins)	No	Yes (deceased brother and cousin)	No	Yes (brother)	Yes (deceased sister)	No	Yes (siblings and deceased cousin)	Yes (siblings and deceased cousin)	N/A (adopted)	No
Age at onset	2 weeks	6 months	1 week	At birth	2 days	1 year	8 months	10 days	6 months	At birth	At birth	2 years	3 days
Failure to thrive	Yes, weight below 5th percentile	Yes, weight below 5th percentile, height at 10th percentile	Yes, weight below 5th percentile	Yes	Yes, weight, height and OFC below 5th percentile	Yes, weight 5th percentile	Yes	Yes	Yes	Low weight (weight <3rd percentile, height 25th percentile)	Low weight (weight <3rd percentile, height 10th-25th percentile)	Yes	Yes, height and weight below 5th percentile
Dysmorphism	None	None	None	Prominent forehead, Upslanting palpebral fissures, Thin upper lips	None	None	None	None	None	None	None	None	None
Motor development	Appropriate for age	Appropriate for age	Delayed motor development with right hemiplegia	Appropriate for age	Mild motor delay	Appropriate for age	Appropriate for age	Mild motor delay	Appropriate for age	Appropriate for age	Appropriate for age	Appropriate for age	Appropriate for age

Mental development	Appropriate for age	Appropriate for age	Appropriate for age	Appropriate for age	Appropriate for age	Appropriate for age	Speech delay	Appropriate for age	Appropriate for age	Appropriate for age	Appropriate for age	Appropriate for age	Appropriate for age
Neurological abnormalities	None reported	None reported	Hemiparesis, Paucity in the movement of the right side of the body	None reported	None reported	None reported	Global developmental delay, Hypotonia	None reported	None reported	None reported	None reported	None reported	None reported
Recurrent lower respiratory tract infections	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pulmonary abnormalities	Chronic coughing, Exertional dyspnea, Basal crackles, Bronchial wall thickening, Hilar lymphadenopathy Mild bronchiectasis, fibrotic bands	Chronic coughing, Bilateral crackles, Mild bronchiectasis Hilar lymphadenopathy	Chronic coughing, Recurrent wheezing episodes, Dyspnea, Bilateral crackles, Bronchial wall thickening, Mediastinal and hilar lymphadenopathy	Chronic coughing, Pneumonia, Hyperactive airway disease	Mild respiratory tract infections	Chronic coughing, Severe pneumonia	Interstitial lung disease	Recurrent wheezing episodes, Patch areas of ground glass appearance and scattered consolidations in both lungs	Bronchiectasis, Chronic coughing	Chronic cough, Pleural effusion, hilar lymphadenopathy Bronchiectasis	Chronic cough, hilar lymphadenopathy	Bronchiectasis, Persistent segmental collapse in the left lower lobe, Chronic productive cough	Collapse/consolidation in segments of both lungs. Subsegmental atelectasis. Small bilateral axillary lymph nodes
Immunological abnormalities	None reported	None reported	None reported	None reported	None reported (see test results)	None reported	Slightly low percentage of CD4+ T-cells	None reported	None reported	Leukocytosis, Lymphocytosis	None reported	None reported	Leucocytosis
Gastroenteric abnormalities	None	None	None	Acute gastroenteritis, Vomiting, Severe gastroesophageal reflux, Chronic diarrhea	Chronic diarrhea, Episodic vomiting, lethargy	None	Hepatomegaly	Choking, vomiting and chronic diarrhea, Hepatomegaly	None	Chronic diarrhea (improved after 2 y), hepatomegaly	Chronic diarrhea	Persistent vomiting, hepatomegaly and persistent cholestasis	Chronic diarrhea, abdominal distention with prominent veins, no visceromegaly
Cardiovascular abnormalities	None, Echocardiogram - normal	Mitral valve prolapse, Mitral regurgitation	None	None	None	None	Right sided heart failure, Right ventricular and right atrial dilatation,	None	None	None	None	None	None, Echocardiogram - normal

							Tricuspid regurgitation, Severe pulmonary hypertension						
Other tests results	Sweat chloride test, Bronchoscopy, Imunological profile – all normal	Heterozygous VUS in CFTR, c.4091C>T, p.Ala1364Val Sweat chloride test - normal	Bronchoscopy - normal	Sweat chloride test, Immunological profile - normal	Decreased T-cell count with low CD4+/CD8+ ratio, low B-cell count and slightly increased NK-cell count	Sweat chloride test, Immunological profile – normal Hom LP SLC26A4, NM_000441.1:c.1339_1340delinsTCT	Sweat chloride test - normal	Liver function test - normal	Bronchoalveolar lavage culture - positive for pseudomonas	Sweat chloride test, Pancreatic elastase, Nasal brush test (light microscopy) – all normal Het pathogenic variant *GAA* NM_000152.3:c.-32-13T>G	Hom pathogenic variant *GAA* NM_000152.3:c.-32-13T>G	Lymphocyte subset analysis, Immunoglobulins, and oxidative burst test – all normal. EM nasal brush - ciliary abnormalities in 34% of examined cilia	Sweat chloride test, Pancreatic elastase – normal
Clinical suspicion	Cystic fibrosis, Primary ciliary dyskinesia	Cystic fibrosis, Primary ciliary dyskinesia, Primary Immunodeficiency	Primary immunodeficiency, Cystic fibrosis, Primary ciliary dyskinesia, Aspiration syndrome and Mitochondrial cytopathy	Cystic fibrosis	Type 1 distal, renal tubular acidosis, congenital enteropathies, chloride losing diarrhea, Primary Immunodeficiency	Cystic fibrosis	Cystic fibrosis, Niemann-Pick disease type 2	Cystic fibrosis	Cystic fibrosis, Primary ciliary dyskinesia	Primary ciliary dyskinesia, Cystic fibrosis Primary immunodeficiency	Cystic fibrosis, Primary immunodeficiency, Malabsorption	Primary ciliary dyskinesia	Cystic fibrosis, Primary immunodeficiency
Other				Cow's milk allergy		Otitis media. Sensorineural hearing impairment (cochlear implant)			Rhinorrhea	Chronic suppurative otitis media, Mediastinal lymphadenopathy	Recurrent otitis media	Sino-nasal polyposis by CT, obstructive sleep apnoea, Rhinitis, Recurrent otitis media	

Footnote: Infancy: < 1 year of age, early childhood: >1-5 years, childhood >5-14 years. OFC: Occipitofrontal circumference

Supplementary table 2. Rare homozygous coding variants remaining as candidates in family 1, 2, 3, 4, 5, 7, 8, and 9. Homozygous variants detected in other samples in our or external data repositories (healthy individuals) were excluded.

Chr	Genomic coordinate	*Gene*	Reference seq: nucleotide change	Reference seq.: protein change	Variant type	dbSNP	OMIM	PyloP	Cadd raw	PopFreq Max
Family 1 - 2427168										
chr10	96117097	*NOC3L*	NM_022451.10:c.351-10delT		Splice region variant & intron variant	rs753480410	610769			0.0002
chr12	66707785	*HELB*	NM_033647.4:c.1700C>T	NM_033647.4:p.Ser567Leu	Missense	rs149157869	614539	0.388	0.34089	0.0053
chr2	85661241	*SH2D6*	NM_201594.2:c.41+7G>A		Splice region variant & intron variant					
chr2	97820417	*ANKRD36*	NM_001164315.1:c.1199T>C	NM_001164315.1:p.Leu400Pro	Missense	-	-	-0.113	0.77015	
chr2	234394541	*USP40*	NM_018218.2:c.3313G>A	NM_018218.2:p.Ala1105Thr	Missense	rs374106216	610570	0.548	-1.37204	0.0003
chr20	7895021	*HAO1*	NM_017545.2:c.335C>A	NM_017545.2:p.Thr112Asn	Missense	rs377526496	605023	9.513	4.90385	0.001
chr7	**16840820**	***AGR2***	**NM_006408.4:c.211C>A**	**NM_006408.3:p.Pro71Thr**	**Missense**	**-**	**606358**	**9.276**	**6.14466**	
chr9	141015116	*CACNA1B*	NM_000718.3:c.6272C>T	NM_000718.3:p.Pro2091Leu	Missense	rs746163681	601012 [Neurodevelopmental disorder with seizures and nonepileptic hyperkinetic movements]	0.285	1.59689	0.0015
Family 2 -2438720										
chr10	82298271	*SH2D4B*	NM_207372.2:c.184G>A	NM_207372.2:p.Ala62Thr	Missense	rs749601744		4.15	1.70092	0.0003
chr15	23686239	*GOLGA6L2*	NM_001304388.1:c.1382_1383insCGAGGAGGAGAAGATGCGGGA	NM_001304388.1:p.Arg460_Glu461insAspGluGluGluLysMetArg	Disruptive inframe insertion					
chr17	6928019	*BCL6B*	ENST00000293805.5:c.720_731delCAGCAGCAGCAG	ENST00000293805.5:p.Ser241_Ser244del	Disruptive inframe deletion		608992			
chr17	7918378	*GUCY2D*	NM_000180.3:c.2769+9T>G		Splice region variant& intron variant	rs771741738	600179 [?Choroidal dystrophy, central areolar 1,Cone-rod dystrophy 6,Leber congenital amaurosis 1,Night blindness, congenital stationary, type 1I]			0
chr17	48504265	*ACSF2*	NM_001288968.1:c.129-1G>C		splice_acceptor_variant&intron_variant	rs189245546	610465	0.048	0.032011	0.0086
chr17	48629001	*SPATA20*	NM_022827.3:c.1706G>A	NM_022827.3:p.Arg569Gln	Missense	rs144320831	613939; n/a	5.669	7.69479	0.001
chr19	52497739	*ZNF615*	NM_001321323.1:c.638C>T	NM_001321323.1:p.Thr213Ile	Missense	rs369585230		-2.753	-1.16086	0.0001
chr19	54561565	*VSTM1*	NM_198481.3:c.349delG	NM_198481.3:p.Val117fs	Frameshift	rs745734767	616804			0.0001
chr2	197537074	*CCDC150*	NM_001080539.1:c.942_943delGCinsTT	NM_001080539.1:p.LeuGln314*	Stop gain					

chr2	210881322	*RPE*	NM_001318926.1:c.488C>T	NM_001318926.1:p.Pro163Leu	Missense	rs370757730	180480; 613833	4.96	7.33988	0.0001
chr4	83424050	*TMEM150C*	NM_001080506.1:c.165A>C	NM_001080506.1:p.Ile55Ile	Splice region variant & synonymous variant	rs377209343	617292			0.0004
chr6	32489844	*HLA-DRB5*	NM_002125.3:c.206_208delTCGinsACA	NM_002125.3:p.PheAsp69TyrAsn	Missense		604776			
chr6	100838957	*SIM1*	NM_005068.2:c.1581T>A	NM_005068.2:p.His527Gln	Missense		603128	0.644	3.55834	
chr7	**16837300**	***AGR2***	**NM_006408.4:c.349C>T**	**NM_006408.4:p.His117Tyr**	**Missense**	**rs780638101**	**606358**	**3.236**	**6.07564**	**0.0001**
chr7	20824043	*SP8*	NM_182700.5:c.1387_1392delGGCGGC	NM_182700.5:p.Gly463_Gly464del	Conservative inframe deletion	rs759086117	608306			0.0057
chrX	39932647	*BCOR*	NM_001123385.1:c.1952T>C	NM_001123385.1:p.Ile651Thr	Missense	rs746064364	300485 [Microphthalmia, syndromic 2]	8.942	3.8841	0
Family 3 - 2399903										
chr1	3417196	*MEGF6*	NM_001409.3:c.2707+1G>A		Splice donor & intron	rs546771819	604266	5.977	4.01321	0
chr1	3732033	*CEP104*	NM_014704.3:c.2711G>T	NM_014704.3:p.Gly904Val	Missense		616690 [Joubert syndrome 25]	4.719	5.62154	
chr14	94517551	*DDX24*	NM_020414.3:c.2566A>G	NM_020414.3:p.Thr856Ala	Missense		606181; 608338	-0.329	1.18224	
chr3	165491198	*BCHE*	NM_000055.3:c.1781G>T	NM_000055.3:p.Ser594Ile	Missense	rs142859898	177400 [Apnea, postanesthetic, susceptibility to, due to BCHE deficiency,Butyrylcholinesterase deficiency]	3.448	5.89742	0.0009
chr7	**16837300**	***AGR2***	**NM_006408.4:c.349C>T**	**NM_006408.4:p.His117Tyr**	**Missense**	**rs780638101**	**606358**	**3.236**	**6.07564**	**0.0001**
chrX	36053879	*CFAP47*	NM_001304548.1:c.3719A>G	NM_001304548.1:p.Lys1240Arg	Missense			0.686		
Family 4 - 2337168										
chr1	79116314	*IFI44*	NM_006417.4:c.434A>T	NM_006417.4:p.Asp145Val	Missense	rs146103588	610468; 613975	3.127	4.4882	0.0019
chr14	88945502	*PTPN21*	NM_007039.3:c.2273T>C	NM_007039.3:p.Leu758Pro	Missense		603271	1.421	0.730167	
chr16	461495	*DECR2*	NM_020664.3:c.796G>C	NM_020664.3:p.Val266Leu	Missense		615839	6.656	3.47079	
chr16	720513	*RHOT2*	NM_138769.2:c.496G>A	NM_138769.2:p.Val166Ile	Missense	rs146373820	613889; 618290	9.856	4.17403	0
chr16	1498755	*CLCN7*	NM_001287.5:c.1810A>G	NM_001287.5:p.Met604Val	Missense		602727 [Hypopigmentation, organomegaly, and delayed myelination and development,Osteopetrosis, autosomal dominant 2,Osteopetrosis, autosomal recessive 4]; 618740	2.031	1.60287	
chr16	16355487	*NOMO3*	NM_001004067.3:c.1349A>G	NM_001004067.3:p.His450Arg	Missense	rs750513109	609159	3.811	-0.22054	0
chr17	6928019	*BCL6B*	ENST00000293805.5:c.720_731delCAGCAGCAGCAG	ENST00000293805.5:p.Ser241_Ser244del	Disruptive inframe deletion		608992			

chr17	7918378	*GUCY2D*	NM_000180.3:c.2769+9T>G		Splice region & intron	rs771741738	600179 [?Choroidal dystrophy, central areolar 1,Cone-rod dystrophy 6,Leber congenital amaurosis 1,Night blindness, congenital stationary, type 1I]			0
chr17	56429356	*TSPOAP1-AS1*	NR_038410.1:n.755-4A>G		Splice region & intron		n/a; 603555; 612482 [Sessile serrated polyposis cancer syndrome]			
chr17	76502807	*DNAH17*	NM_173628.3:c.4798G>T	NM_173628.3:p.Val1600Leu	Missense	rs76449350	610063 [Spermatogenic failure 39]; n/a	0.71	3.12798	0.0091
chr18	10800427	*PIEZO2*	NM_022068.2:c.1286T>C	NM_022068.2:p.Ile429Thr	Missense		613629 [?Marden-Walker syndrome,Arthrogryposis, distal, type 3,Arthrogryposis, distal, type 5,Arthrogryposis, distal, with impaired proprioception and touch]	6.467	2.96644	
chr19	6047495	*RFX2*	NM_000635.3:c.13G>A	NM_000635.3:p.Glu5Lys	Missense	rs142338131	142765	7.211	7.44236	0.0045
chr19	8400050	*KANK3*	NM_198471.2:c.661A>G	NM_198471.2:p.Lys221Glu	Missense		614611	4.529	4.76978	
chr19	8577979	*ZNF414*	NM_001146175.1:c.250G>A	NM_001146175.1:p.Gly84Ser	Missense	rs772627384		-0.205	1.02072	0
chr19	9061720	*MUC16*	NM_024690.2:c.25726T>C	NM_024690.2:p.Phe8576Leu	Missense	rs776428876	606154	-0.753	0.126963	0.0001
chr19	9204530	*OR1M1*	NM_001004456.1:c.610G>A	NM_001004456.1:p.Gly204Arg	Missense	rs199800381		-3.28	3.31593	0.0011
chr19	14236925	*ASF1B*	NM_018154.2:c.225+9G>A		Splice region & intron	rs543610465	609190			
chr19	18391795	*JUND*	NM_005354.5:c.491_499dupCCGCCGCCG	NM_005354.5:p.Ala164_Ala166dup	Conservative inframe insertion	rs529130306	165162; n/a			0.021
chr20	5935314	*MCM8*	NM_001281521.1:c.314G>A	NM_001281521.1:p.Arg105Lys	Missense		608187 [?Premature ovarian failure 10]; n/a	0.392	0.145186	
chr20	9440301	*PLCB4*	NM_000933.3:c.3056A>G	NM_000933.3:p.Gln1019Arg	Missense	rs377707845	600810 [Auriculocondylar syndrome 2]	6.823	2.71346	0.0024
chr20	39990473	*EMILIN3*	NM_052846.1:c.1736C>T	NM_052846.1:p.Ser579Leu	Missense	rs772266071	608929; 605520	7.181	5.3479	0
chr3	130134482	*COL6A5*	NM_001278298.1:c.4762-7A>G		Splice region & intron	rs575983094	611916			0.0072
chr5	66478938	*CD180*	NM_005582.2:c.1733C>T	NM_005582.2:p.Pro578Leu	Missense	rs185244476	602226	5.621	5.62279	0.0011
chr5	72875903	*UTP15*	NM_032175.3:c.1541A>C	NM_032175.3:p.Lys514Thr	Missense	rs142841898	616194	0.367	1.87684	0.0008
chr7	20824043	*SP8*	NM_182700.5:c.1387_1392delGGCGGC	NM_182700.5:p.Gly463_Gly464del	Conservative inframe deletion	rs759086117	608306			0.0057
chr7	103835705	*ORC5*	NM_002553.3:c.442-3C>T		Splice region & intron	rs747497110	602331			0.0003
chr7	107334923	*SLC26A4*	NM_000441.1:c.1339_1340delAAinsTCT	NM_000441.1:p.Lys447fs	Frameshift & missense & splice region		605646 [Deafness, autosomal recessive 4, with enlarged vestibular aqueduct, Pendred syndrome]			
chr7	107334924	*SLC26A4*	NM_000441.1:c.1340_1341insTCT	NM_000441.1:p.Lys447delinsAsnLeu	Splice region & disruptive inframe insertion		605646 [Deafness, autosomal recessive 4, with enlarged vestibular aqueduct,Pendred syndrome]			

chr7	116199040	*CAV1*	NM_001753.4:c.236A>G	NM_001753.4:p.His79Arg	Missense	rs376004565	601047 [?Lipodystrophy, congenital generalized, type 3,Lipodystrophy, familial partial, type 7,Pulmonary hypertension, primary, 3]	8.947	3.18652	0.0005
chr7	121678816	*PTPRZ1*	NM_002851.2:c.5375A>G	NM_002851.2:p.Asn1792Ser	Missense		176891	6.197	3.58104	
chr7	**16837300**	***AGR2***	**NM_006408.4:c.349C>T**	**NM_006408.4:p.His117Tyr**	**Missense**	**rs780638101**	**606358**	**3.236**	**6.07564**	**0.0001**
chr8	21955123	*FAM160B2*	NM_022749.5:c.394C>T	NM_022749.5:p.Pro132Ser	Missense	rs371809549		6.142	5.48401	0.0003
chr9	139298499	*ENTR1*	NM_001039707.1:c.1208+8C>T		Splice region & intron		618289			
chr9	140267963	*EXD3*	NM_017820.4:c.209G>A	NM_017820.4:p.Arg70Gln	Missense	rs200718958		-0.133	1.33609	0.0036
Family 5 - 2151518										
chr7	24745875	*GSDME*	NM_001127453.1:c.1111T>G	NM_001127453.1:p.Cys371Gly	Missense	rs138301435	608798 [Deafness, autosomal dominant 5]	-2.439	-3.47024	0.0023
chr7	**16839367**	***AGR2***	**NM_006408.4:c.330+1G>T**		**Splice donor & intron**	**rs1483660993**	**606358**	**6.2**	**5.1938**	
Family 7 - 2518771 – index and sister										
chr10	31137766	*ZNF438*	NM_001143766.1:c.1568G>A	NM_001143766.1:p.Arg523Gln	Missense	rs745473764		0.854	1.66277	0.0001
chr10	32580137	*EPC1*	NM_025209.3:c.929T>G	NM_025209.3:p.Phe310Cys	Missense	rs868826577	610999	3.411	4.08598	
chr10	37482114	*ANKRD30A*	NM_052997.2:c.2374G>T	NM_052997.2:p.Ala792Ser	Missense	rs189204441	610856	-0.491	-1.68588	0.015
chr10	45473142	*DEPP1*	NM_007021.3:c.337C>T	NM_007021.3:p.Gln113*	stop_gained	rs867293581	611309	3.966	11.5352	
chr10	45959681	*MARCHF8*	NM_001282866.1:c.242+6A>G		splice_region_variant&intron_variant	rs115355800	613335			0.022
chr10	50819325	*SLC18A3*	NM_003055.2:c.539C>G	NM_003055.2:p.Ala180Gly	Missense	rs771402838	600336 [Myasthenic syndrome, congenital, 21, presynaptic]; 118490 [Myasthenic syndrome, congenital, 6, presynaptic]	3.537	3.50895	0.0002
chr10	50943403	*OGDHL*	NM_018245.2:c.2910-6C>T		splice_region_variant&intron_variant		617513			
chr10	52569761	*A1CF*	NM_001198819.1:c.1550A>C	NM_001198819.1:p.Glu517Ala	Missense		618199	7.674	6.41814	
chr10	52610477	*A1CF*	NM_001198819.1:c.71A>G	NM_001198819.1:p.Lys24Arg	Missense		618199	-1.819	-1.58506	
chr12	14599904	*ATF7IP*	NM_181352.1:c.1954-8_1954-4dupTTTTT		splice_region_variant&intron_variant		613644			
chr14	35739656	*PRORP*	NM_014672.3:c.1474C>A	NM_014672.3:p.His492Asn	Missense		609947	3.207	6.351	
chr14	105417623	*AHNAK2*	NM_138420.2:c.4165G>C	NM_138420.2:p.Ala1389Pro	Missense		608570	-0.362	2.02195	
chr17	27620989	*NUFIP2*	NM_020772.2:c.86_88dupAGC	NM_020772.2:p.Gln29dup	conservative_inframe_insertion	rs577779578	609356			0.003
chr17	28791746	*CPD*	NM_001304.4:c.4057A>G	NM_001304.4:p.Thr1353Ala	Missense	rs115003383	603102	8.469	4.03671	0.011

chr17	29685568	*NF1*	NM_001042492.2:c.8041A>G	NM_001042492.2:p.Ile2681Val	Missense	rs146315101	613113 [Leukemia, juvenile myelomonocytic,Neurofibromatosis-Noonan syndrome,Neurofibromatosis, familial spinal,Neurofibromatosis, type 1,Watson syndrome]	3.221	-0.019102	0.0011
chr17	33904517	*PEX12*	NM_000286.2:c.220T>C	NM_000286.2:p.Tyr74His	Missense		601758 [Peroxisome biogenesis disorder 3A (Zellweger),Peroxisome biogenesis disorder 3B]	8.495	5.48378	
chr17	36485313	*GPR179*	NM_001004334.3:c.4139C>T	NM_001004334.3:p.Pro1380Leu	Missense	rs764207680	614515 [Night blindness, congenital stationary (complete), 1E, autosomal recessive]	0.455	0.154142	0.0001
chr18	14852348	*ANKRD30B*	NM_001145029.1:c.4048G>A	NM_001145029.1:p.Glu1350Lys	Missense		616565	4.374	3.9738	
chr18	21660713	*TTC39C*	NM_001135993.1:c.625G>A	NM_001135993.1:p.Glu209Lys	Missense			3.464	2.01257	
chr4	128625370	*INTU*	NM_015693.3:c.1504-8delT		splice_region_variant&intron_variant	rs772920778	610621 [?Orofaciodigital syndrome XVII,?Short-rib thoracic dysplasia 20 with polydactyly]			0.0059
chr7	132070014	*PLXNA4*	NM_001105543.1:c.1412C>T	NM_001105543.1:p.Thr471Met	Missense	rs183271681	604280	-0.687	-0.047923	0.0074
chr7	**16837300**	***AGR2***	**NM_006408.3:c.349C>T**	**NM_006408.3:p.His117Tyr**	**Missense**	**rs780638101**	**606358**	**3.236**	**6.07564**	**0.0001**
chr9	19346501	*DENND4C*	NM_017925.6:c.3587A>G	NM_017925.6:p.Asp1196Gly	Missense	rs149094194		5.795	2.19924	0.0011
chrX	65252335	*VSIG4*	NM_007268.2:c.669C>A	NM_007268.2:p.Ser223Arg	Missense	rs749453785	300353	-1.006	2.20312	0
chrX	153880610	*CTAG2*	NM_020994.4:c.565G>T	NM_020994.4:p.Glu189*	stop_gained		300396	-1.488	6.97086	
Family 8 - 2508357										
chr1	69369	*OR4F5*	NM_001005484.1:c.279G>T	NM_001005484.1:p.Gln93His	Missense			0.502	3.75252	
chr1	180464666	*ACBD6*	NM_032360.3:c.223-6A>G		Splice region & intron	rs151129855	616352			0.0038
chr1	197871814	*C1orf53*	NM_001024594.2:c.35C>T	NM_001024594.2:p.Ala12Val	Missense	rs374493997		0.447	0.697828	0.0092
chr1	197887095	*LHX9*	NM_020204.2:c.142G>A	NM_020204.2:p.Ala48Thr	Missense	rs113693840	606066	6.277	2.96391	0.011
chr1	200584666	*KIF14*	NM_014875.2:c.1184C>T	NM_014875.2:p.Thr395Met	Missense	rs138621008	611279 [?Meckel syndrome 12,Microcephaly 20, primary, autosomal recessive]	3.493	2.47981	0.011
chr1	201190604	*IGFN1*	NM_001164586.1:c.9931A>G	NM_001164586.1:p.Thr3311Ala	Missense	rs370519814	617309	4.578	3.80368	0.0008
chr1	203743568	*LAX1*	NM_017773.3:c.956G>C	NM_017773.3:p.Gly319Ala	Missense	rs755267095		0.304	-0.857094	0.0002
chr1	207642169	*CR2*	NM_001006658.2:c.659G>A	NM_001006658.2:p.Arg220Gln	Missense	rs147633291	120650 [Systemic lupus erythematosus, susceptibility to, 9,Immunodeficiency, common variable, 7]	-0.236	-0.85175	0.0002

chr1	207651374	*CR2*	NM_001006658.2:c.3047C>T	NM_001006658.2:p.Ser1016Leu	Missense	rs138062179	120650 [Systemic lupus erythematosus, susceptibility to, 9,Immunodeficiency, common variable, 7]	0.006	1.26183	0.0038
chr1	214638055	*PTPN14*	NM_005401.4:c.92A>G	NM_005401.4:p.Asn31Ser	Missense	rs151121546	603155 [Choanal atresia and lymphedema]	2.19	0.070051	0.0008
chr1	219352740	*LYPLAL1*	NM_001350628.1:c.192-2A>G		Splice acceptor & intron	rs530475818	616548			0.0008
chr1	220300171	*IARS2*	NM_018060.3:c.1823C>G	NM_018060.3:p.Ser608Cys	Missense		612801 [?Cataracts, growth hormone deficiency, sensory neuropathy, sensorineural hearing loss, and skeletal dysplasia]	2.424	5.65751	
chr1	220324939	*RAB3GAP2*	NM_012414.3:c.4026+9A>G		Splice region & intron		609275 [Martsolf syndrome,Warburg micro syndrome 2]		0.18342	
chr1	228494751	*OBSCN*	NM_001271223.2:c.14947A>G	NM_001271223.2:p.Met4983Val	Missense		608616	3.952	3.31168	
chr1	228560723	*OBSCN*	NM_001271223.2:c.25115G>A	NM_001271223.2:p.Arg8372His	Missense	rs772771183	608616	0.06	6.83572	0.0001
chr10	115391287	*NRAP*	NM_001261463.1:c.1823T>C	NM_001261463.1:p.Ile608Thr	Missense	rs867292545	602873	6.616	5.40354	
chr10	118618610	*ENO4*	NM_001242699.1:c.595C>T	NM_001242699.1:p.Pro199Ser	Missense	rs150721071	131375	1.127	2.55975	0.0061
chr11	556970	*LRRC56*	NM_173573.2:c.841G>A	NM_173573.2:p.Ala281Thr	Missense	rs200530408	618227 [Ciliary dyskinesia, primary, 39]	-3.775	0.407053	0.0015
chr11	615268	*IRF7*	NM_001572.3:c.21-9C>T		Splice region & intron	rs775307916	605047 [?Immunodeficiency 39]		4.43317	0
chr11	637408	*DRD4*	NM_000797.3:c.104C>T	NM_000797.3:p.Ala35Val	Missense	rs767779176	126452 [Attention deficit-hyperactivity disorder,Autonomic nervous system dysfunction]	-0.436		0.0009
chr11	640063	*DRD4*	NM_000797.3:c.814G>C	NM_000797.3:p.Gly272Arg	Missense		126452 [Attention deficit-hyperactivity disorder,Autonomic nervous system dysfunction]	-0.327	-1.0774	
chr11	640167	*DRD4*	NM_000797.3:c.918_925delCGGCTCCAinsTGGCCCCG	NM_000797.3:p.CysGlySerAsn306CysGlyProAsp	Missense		126452 [Attention deficit-hyperactivity disorder,Autonomic nervous system dysfunction]		3.0393	
chr11	704536	*EPS8L2*	NM_001276274.1:c.742G>A	NM_001276274.1:p.Ala248Thr	Missense	rs140750324	614988 [Deafness autosomal recessive 106]			0.0053
chr11	798222	*SLC25A22*	NM_001293167.1:c.610_612dupCGC	NM_001293167.1:p.Arg204dup	Conservativeinframe inser tion	rs572464433	609302 [Epileptic encephalopathy, early infantile, 3]		1.3418	0.003
chr11	1028702	*MUC6*	NM_005961.2:c.1535G>A	NM_005961.2:p.Arg512His	Missense	rs748763819	158374	-1.055	2.60293	0.0001
chr11	1093705	*MUC2*	ENST00000441003.2:c.5524G>T	ENST00000441003.2:p.Ala1842Ser	Missense	rs546395242	158370			0.0008
chr11	1103869	*MUC2*	ENST00000441003.2:c.8168C>T	ENST00000441003.2:p.Ser2723Leu	Missense	rs770253933	158370		-0.95932	0.0013
chr11	1156634	*MUC5AC*	XM_003403450.4:c.1482G>A	XM_003403450.4:p.Met494Ile	Missense	rs747862882	158373	-0.554		0
chr11	1157570	*MUC5AC*	XM_003403450.4:c.1582C>A	XM_003403450.4:p.Pro528Thr	Missense	rs145633450	158373	2.484		0.0008

chr11	13729560	*FAR1*	NM_032228.5:c.479G>A	NM_032228.5:p.Arg160His	Missense	rs139416149	616107 [Peroxisomal fatty acyl-CoA reductase 1 disorder]	5.818	1.02607	0.001
chr11	28135086	*METTL15*	NM_001113528.1:c.205G>A	NM_001113528.1:p.Ala69Thr	Missense	rs150513835	618711	-0.107	0.403354	0.012
chr11	44331153	*ALX4*	NM_021926.3:c.460T>A	NM_021926.3:p.Cys154Ser	Missense	rs182274454	605420 [Craniosynostosis 5, susceptibility to,Frontonasal dysplasia 2,Parietal foramina 2]	6.759		0.0079
chr11	57077855	*TNKS1BP1*	NM_033396.2:c.2330A>C	NM_033396.2:p.Lys777Thr	Missense	rs150510361	607104	-0.909	-1.54772	0.0045
chr11	64132915	*RPS6KA4*	NM_003942.2:c.1049C>T	NM_003942.2:p.Pro350Leu	Missense	rs141809902	603606	4.182	0.186073	0.0064
chr11	70277319	*CTTN*	NM_001184740.1:c.1088A>G	NM_001184740.1:p.Gln363Arg	Missense	rs141534651	164765	1.569		0.0024
chr11	117052513	*SIDT2*	NM_001040455.1:c.306-10C>T		Splice region & intron	rs183950705	617551		-1.20542	0.011
chr12	2027813	*CACNA2D4*	NM_172364.4:c.-174C>T		5 prime UTR premature start codon gain	rs374494829	608171 [Retinal cone dystrophy 4]		5.84564	0.0083
chr12	2775846	*CACNA1C*	NM_199460.2:c.4671-6T>A		Splice region & intron		114205 [Brugada syndrome 3,Long QT syndrome 8,Timothy syndrome]		6.05791	
chr12	2997396	*RHNO1*	NM_001252499.2:c.488C>T	NM_001252499.2:p.Ser163Leu	Missense	rs145733432	614085	1.743	2.7842	0.0015
chr12	120794697	*MSI1*	NM_002442.3:c.652+8C>T		Splice region & intron	rs145402971	603328		-0.392542	0.0084
chr12	124297973	*DNAH10*	NM_207437.3:c.3053G>A	NM_207437.3:p.Cys1018Tyr	Missense	rs138151312	605884	6.743	2.54644	0.0048
chr12	124364296	*DNAH10*	NM_207437.3:c.8228C>T	NM_207437.3:p.Pro2743Leu	Missense	rs755673190	605884	2.659	2.82796	0.0001
chr14	89044465	*ZC3H14*	NM_024824.4:c.1260T>G	NM_024824.4:p.Asp420Glu	Missense	rs201108116	613279 [Mental retardation, autosomal recessive 56]	0.159	3.26121	0.0003
chr16	11370095	*PRM2*	NM_001286356.1:c.133G>A	NM_001286356.1:p.Glu45Lys	Missense	rs768731173	182890	0.205	-0.25231	0.0001
chr17	28511782	*NSRP1*	NM_032141.3:c.767C>T	NM_032141.3:p.Ala256Val	Missense	rs148657875	616173	4.478	-0.748862	0.0014
chr17	48632896	*SPATA20*	NM_022827.3:c.2282G>C	NM_022827.3:p.Arg761Pro	Missense	rs373370910	613939	3.363		0
chr19	362273	*THEG*	NM_016585.4:c.1067C>G	NM_016585.4:p.Pro356Arg	Missense	rs780542408	609503	0.29	6.78287	0
chr19	507696	*MADCAM1*	NM_033513.2:c.190A>G	NM_033513.2:p.Ile64Val	Missense	rs760844827	102670	2.559	4.54354	0.0006
chr19	871258	*MED16*	NM_005481.2:c.2099-7_2099-6delTC		Splice region & intron		604062			
chr19	1061796	*ABCA7*	NM_019112.3:c.5479G>C	NM_019112.3:p.Gly1827Arg	Missense		605414 [Alzheimer disease 9, susceptibility to]	9.312	7.29201	
chr19	1487830	*REEP6*	NM_017573.4:c.547T>G	NM_017573.4:p.Tyr183Asp	Missense	rs202179680	609346 [Retinitis pigmentosa 77]	-0.13	0.852528	0.0033
chr19	1556218	*MEX3D*	NM_001174118.1:c.1300T>C	NM_001174118.1:p.Phe434Leu	Missense	rs746484985	611009	3.222	4.57011	0
chr19	2763724	*SGTA*	NM_003021.3:c.424G>A	NM_003021.3:p.Ala142Thr	Missense	rs374479958	603419	0.338	4.70748	0.0008

chr19	2934729	*ZNF77*	NM_021217.2:c.396C>A	NM_021217.2:p.His132Gln	Missense	rs34184381	194551	-2.339	2.49175	0.011
chr19	55870332	*COX6B2*	NM_001145402.1:c.1904C>T	NM_001145402.1:p.Ala635Val	Missense	rs148112323	618127	-1.76	-1.57011	0.014
chr19	55870797	*COX6B2*	NM_001145402.1:c.1439G>A	NM_001145402.1:p.Arg480Gln	Missense	rs140308319	618127	-1.188		0.03
chr2	85778964	*GGCX*	NM_000821.6:c.1580C>T	NM_000821.6:p.Thr527Ile	Missense	rs78504541	137167 [Pseudoxanthoma elasticum-like disorder with multiple coagulation factor deficiency,Vitamin K-dependent clotting factors, combined deficiency of, 1]	7.311		0.0095
chr2	87072063	*CD8B*	NM_172213.3:c.602G>A	NM_172213.3:p.Arg201Gln	Missense		186730	1.112		
chr2	99012373	*CNGA3*	NM_001298.2:c.740C>T	NM_001298.2:p.Thr247Met	Missense	rs148616345	600053 [Achromatopsia 2]	0.357	2.63198	0.005
chr2	105956062	*C2orf49*	NM_024093.2:c.122A>T	NM_024093.2:p.Asp41Val	Missense	rs147396516		3.837		0.0009
chr2	109365537	*RANBP2*	NM_006267.4:c.1225A>G	NM_006267.4:p.Ile409Val	Missense	rs201087513	601181 [Encephalopathy, acute, infection-induced, 3, susceptibility to]	-0.064		0.0011
chr2	116572446	*DPP10*	NM_001321905.1:c.1829A>G	NM_001321905.1:p.Asn610Ser	Missense	rs373895432	608209	2.901	-0.908639	0.0002
chr2	118587033	*DDX18*	NM_006773.3:c.1861G>C	NM_006773.3:p.Val621Leu	Missense		606355	9.184	4.64068	
chr2	132021201	*POTEE*	NM_001083538.1:c.2173G>A	NM_001083538.1:p.Asp725Asn	Missense		608914	5.197	0.55247	
chr2	238977928	*SCLY*	NR_037904.1:n.880-10A>G		Splice region & Intron		611056		0.255702	
chr21	46913070	*COL18A1*	ENST00000359759.8:c.3460-7A>T		Splice region & intron	rs576172127	120328 [Glaucoma, primary closed-angle,Knobloch syndrome, type 1]		2.29856	0.001
chr3	9959601	*IL17RC*	NM_153461.3:c.341-6A>G		Splice region & Intron		610925 [Candidiasis, familial, 9]			
chr3	10420100	*ATP2B2*	NM_001001331.2:c.1043-6C>G		Splice region & intron	rs111358898	108733 [Deafness, autosomal recessive 12, modifier of]		4.79197	0.011
chr3	12641745	*RAF1*	NM_002880.3:c.896A>G	NM_002880.3:p.Asn299Ser	Missense	rs866428774	164760 [Cardiomyopathy, dilated, 1NN, LEOPARD syndrome 2,Noonan syndrome 5]	2.879	0.850035	
chr3	15115639	*RBSN*	NM_001302378.1:c.2005G>A	NM_001302378.1:p.Glu669Lys	Missense	rs368892679	609511	4.257		0.0001
chr4	37446261	*NWD2*	NM_001144990.1:c.2651C>T	NM_001144990.1:p.Ser884Leu	Missense	rs371806771		5.809		0.0029
chr4	38051428	*TBC1D1*	NM_015173.3:c.1819C>T	NM_015173.3:p.Pro607Ser	Missense	rs766339856	609850	0.018		0.0002
chr4	77138780	*SCARB2*	NM_001242936.1:c.20G>A	NM_001242936.1:p.Cys7Tyr	Missense	rs576816603	602257 [Epilepsy, progressive myoclonic 4, with or without renal failure]	0.365		0.0008
chr4	79308538	*FRAS1*	NM_025074.6:c.3658G>A	NM_025074.6:p.Val1220Met	Missense	rs367770853	607830 [Fraser syndrome 1]	2.875		0.0001
chr5	32263226	*MTMR12*	NM_001040446.2:c.706T>A	NM_001040446.2:p.Cys236Ser	Missense	rs199989524	606501	2.52		0.0029

chr5	45695991	*HCN1*	NM_021072.3:c.205G>A	NM_021072.3:p.Gly69Ser	Missense		602780 [Epileptic encephalopathy, early infantile, 24,Generalized epilepsy with febrile seizures plus, type 10]	0.064		
chr5	63496630	*RNF180*	NM_001113561.2:c.1-5C>T		Splice region & intron	rs145598166	616015			0.0023
chr6	137323457	*IL20RA*	NM_014432.3:c.900C>A	NM_014432.3:p.Phe300Leu	Missense		605620	1.674		
chr6	147136238	*ADGB*	NM_024694.3:c.4889C>T	NM_024694.3:p.Ala1630Val	Missense		614630	2.326		
chr6	149699835	*TAB2*	NM_001292034.2:c.784T>C	NM_001292034.2:p.Ser262Pro	Missense		605101 [Congenital heart defects, nonsyndromic, 2]	2.941		
chr6	159653708	*FNDC1*	NM_032532.2:c.2164C>T	NM_032532.2:p.Pro722Ser	Missense	rs767489430	609991	-0.98		0.0005
chr6	168439229	*KIF25*	NM_030615.2:c.318-4G>A		Splice region & intron	rs376908446	603815			0.001
chr7	1522175	*INTS1*	NM_001080453.2:c.3703+4_3703+7delCGGTinsTGGC		Splice region & intron		611345 [Neurodevelopmental disorder with cataracts, poor growth, and dysmorphic facies]			
chr7	1532637	*INTS1*	NM_001080453.2:c.2165+9A>G		Splice region & intron	rs202080075	611345 [Neurodevelopmental disorder with cataracts, poor growth, and dysmorphic facies]			0.0008
chr7	2695729	*TTYH3*	NM_025250.2:c.1024C>T	NM_025250.2:p.Pro342Ser	Missense	rs202147266	608919	2.048		
chr7	4824638	*AP5Z1*	NM_014855.2:c.890C>G	NM_014855.2:p.Ala297Gly	Missense		613653 [Spastic paraplegia 48, autosomal recessive]	5.732		
chr7	4830809	*AP5Z1*	NM_014855.2:c.2217C>G	NM_014855.2:p.His739Gln	Missense		613653 [Spastic paraplegia 48, autosomal recessive]	0.228		
chr7	12675734	*SCIN*	NM_001112706.2:c.1384C>T	NM_001112706.2:p.Arg462Trp	Missense	rs771548886	613416	1.186		0.0002
chr7	**16839366**	***AGR2***	**NM_006408.3:c.330+1delG**		**Splice donor & intron**		**606358**			
chr8	37706129	*BRF2*	NM_018310.3:c.199C>T	NM_018310.3:p.Arg67Cys	Missense	rs140395188	607013	4.524		0.0014
chr8	38853939	*ADAM9*	NM_078473.2:c.20C>T	NM_078473.2:p.Pro7Leu	Missense	rs750181557	602713 [Cone-rod dystrophy 9]	6.11		0.0003
chrX	48123331	*SSX1*	NM_001278691.1:c.445_449delGAGAAinsAAGAG	NM_001278691.1:p.GluLys149LysArg	Missense		312820 [?Sarcoma, synovial]			
chrX	48213442	*SSX3*	NM_021014.3:c.272G>T	NM_021014.3:p.Gly91Val	Missense		300325	-1.038		
Family 9 – 2534592										
chr7	**16834610**	***AGR2***	**NM_006408.3:c.428G>A**	**NM_006408.3:p.Gly143Glu**	**Missense**	**rs923936131**	**606358**	**7.316**	**7.1477**	
chr17	55962604	*CUEDC1*	NM_001271875.1:c.322A>G	NM_001271875.1:p.Ser108Gly	Missense	rs765302468		7.559	5.34241	0.0001

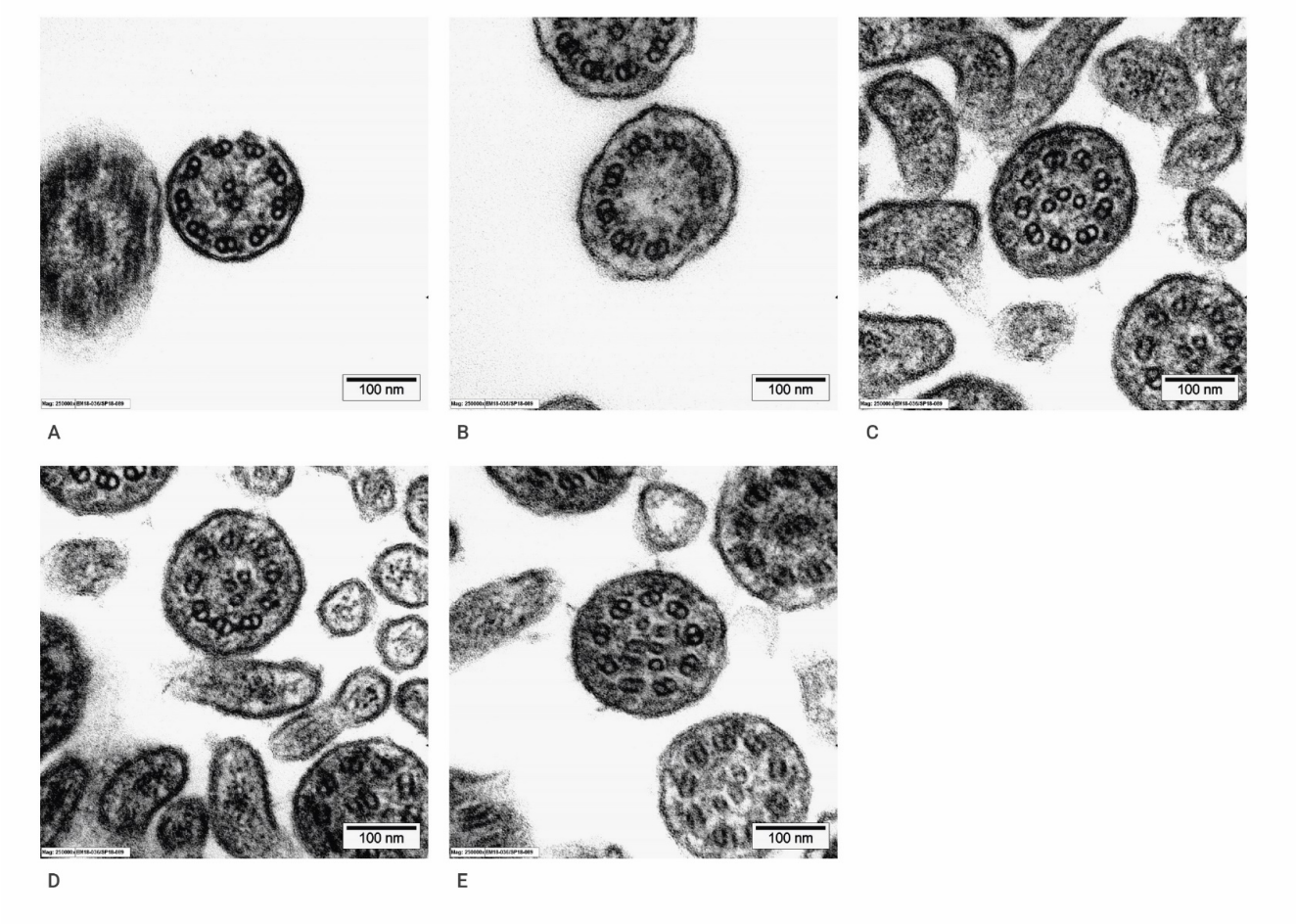
PopMaxFreq (Population maximum frequency - indicates the highest frequency of the variant observed in databases gnomAD, ESP and 1000G). PhyloP scores indicate evolutionary conserved positions (high positive). CADD (Combined Annotation-Dependent Depletion) ranks genetic variants, including single

nucleotide variants (SNVs) and short inserts and deletions (InDels), throughout the human genome reference assembly. [4]RAW score above 4 are considered as likely damaging [4].

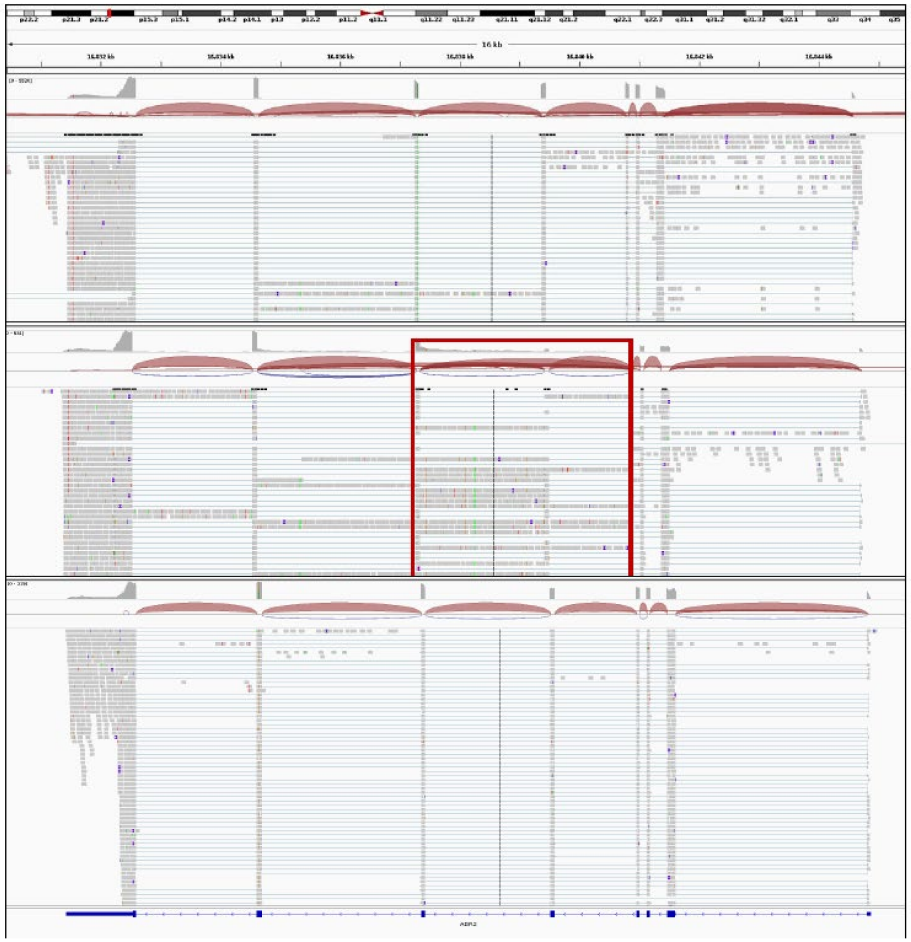
Supplementary table 3: AGR2 variants detected in this study are novel or ultra-rare (gnomAD v2.1.1), with high CADD and conservation scores.

AGR2 Variant (NM_006408.3)	Allele number	Number of homozygotes	Cadd raw	PHRED_CADD	phylop100way_ vertebrate
c.330+1G>T	1/249340	0	5	33	6
c.330+1del	0	0	NA	NA	NA
c.211C>A, p.Pro71Thr	0	0	4	27	9
c.349C>T, p.His117Tyr	1/250930	0	4	27	3
c.428G>A, p.Gly143Glu	0	0	5	32	7

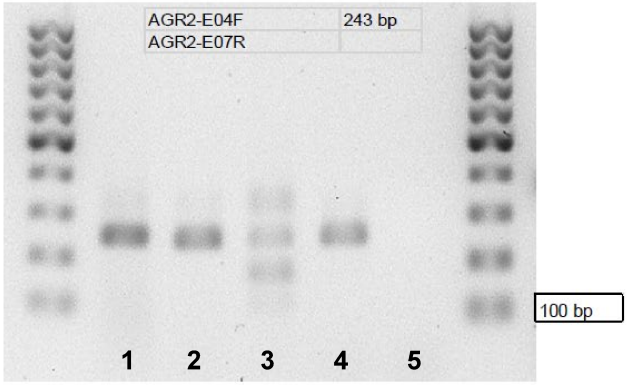
Supplementary Figure 1. Electron microscopy findings.in patient II-1 (family 8). A. Control cilia showing nine peripheral and two central pairs of microtubules with outer and inner dynein arms. B. Missing central microtubular doublets. C and D. Triples instead of central doublets with missing dynein arms t peripheral doublets. E. Duplication of central doublets with missing dynein arms at peripheral doublets (250000x).

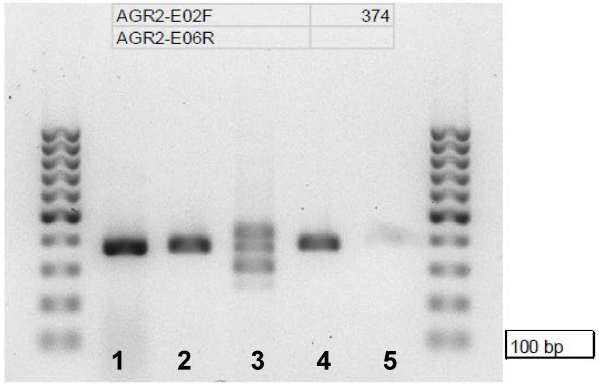


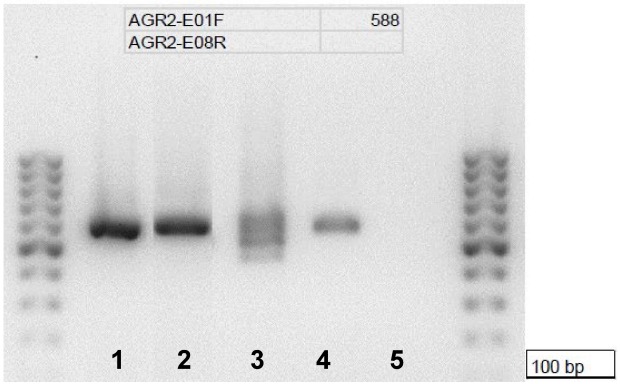
Supplementary Figure 2A: Splice junction track from the Integrative Genomic Viewer (IGV), showing *AGR2* RNA sequencing data. Note the aberrant splicing in the sample with the homozygous c.330+1del variant, producing a disturbance in the splicing around intron 5 (red box).



Supplementary Figure 2B: The variant c.330+1del lead to abnormal *AGR2* splicing. Agarose gel with bands of expected sizes obtained for samples with the missense variant (exon 4-7, exon2-6, exon 1-8). However, for the patient with the homozygous splicing variant several additional bands are detected, corresponding to abnormal *AGR2* transcripts. No *AGR2* PCR products were detected in blood derived RNA.







Lane 1: Hom. c.349C>T, p.His117Tyr

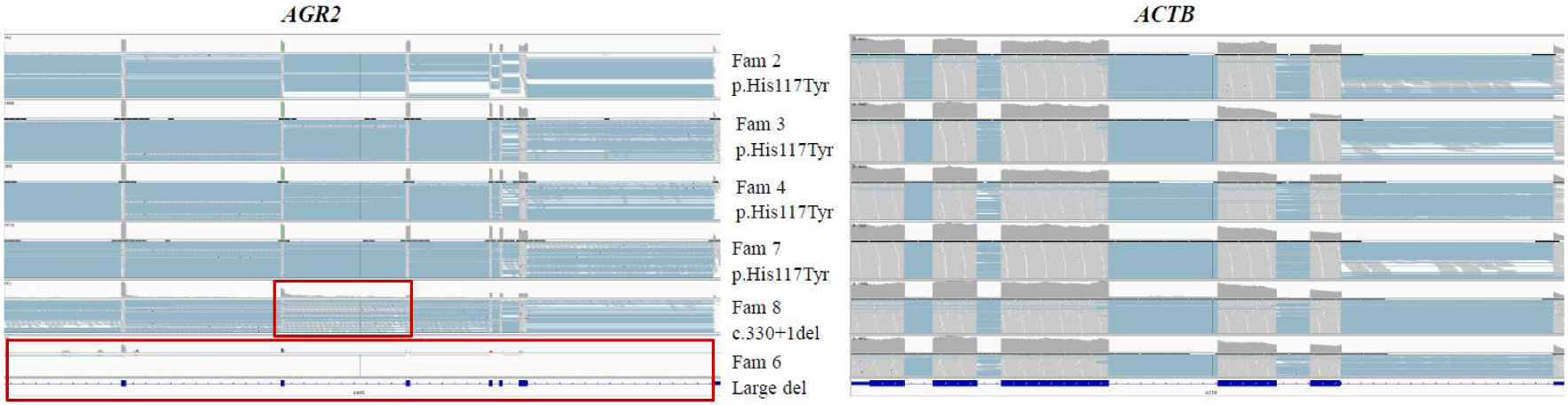
Lane 2: Het. c.349C>T, p.His117Tyr

Lane 3: Hom. c.330+1del

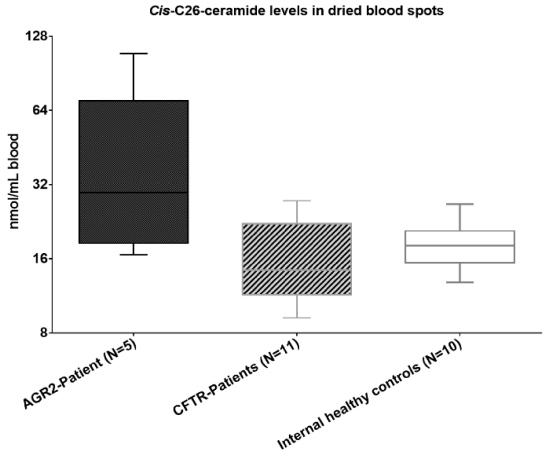
Lane 4: Control

Lane 5: RNA isolated from blood

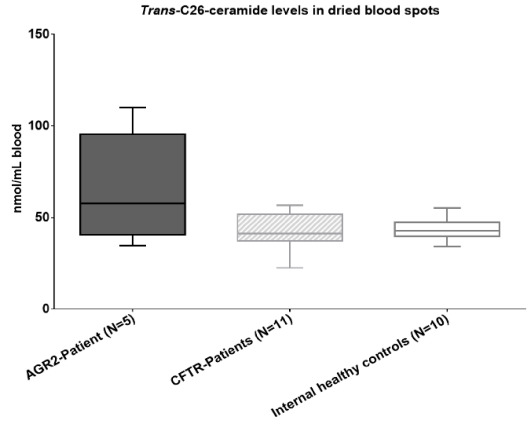
Supplementary Figure 2C: IGV tracks showing *AGR2* RNA sequencing data. Left panel is showing the aligned RNAseq data for *AGR2*. Right panel is showing aligned data for the housekeeping gene *ACTB* (as control gene). Note abnormal reads corresponding to intron 5. This is consistent with the retention of the intron caused by the c.330+1del variant in the index case from Family 8 (red box). Note absence of *AGR2* transcripts caused by the large homozygous deletion in *AGR2* in the index case of Family 6 (red box).



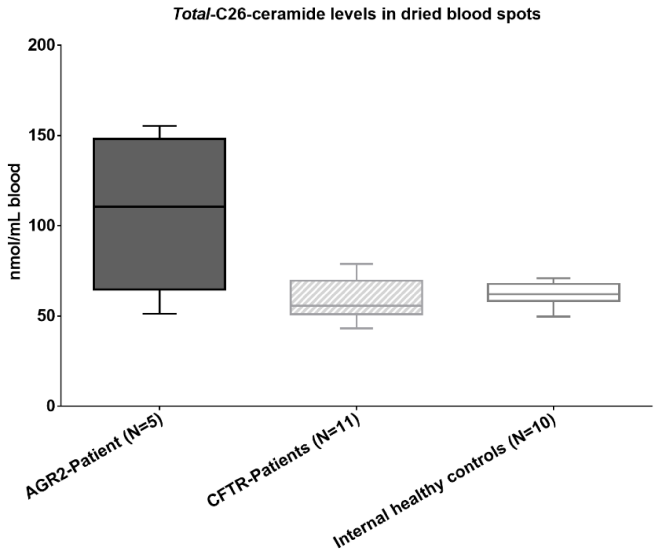
Supplementary Figure 3: Blood Cis, Trans and Total-Cer26 are significantly elevated in *AGR2* patients compared to healthy controls and *CFTR*-patients. ANOVA test Cis-Cer26 F=4.13, P=0.03; Trans-Cer26=4.75, P=0.02; Trans-Cer26 F=10.94, P<0.001.



	AGR2-Patients (N=5)	CFTR-Patients (N=11)	Healthy controls (N=10)
Number of values	5	11	10
Minimum	16.60	9.210	12.80
25% Percentile	18.50	11.40	15.43
Median	29.70	14.40	18.10
75% Percentile	69.95	22.20	20.78
Maximum	109.0	27.50	26.70
Mean	41.32	16.77	18.46
Std. Deviation	38.32	6.010	4.043
Std. Error of Mean	17.14	1.812	1.278
Lower 95% CI of mean	-6.257	12.74	15.57
Upper 95% CI of mean	88.90	20.81	21.35



	AGR2-Patient (N=5)	CFTR-Patients (N=11)	Healthy controls (N=10)
Number of values	5	11	10
Minimum	34.70	22.40	34.20
25% Percentile	40.50	37.20	39.90
Median	57.70	41.30	42.80
75% Percentile	95.45	51.80	47.40
Maximum	110.0	56.70	55.30
Mean	65.92	42.77	43.74
Std. Deviation	29.98	10.62	5.939
Std. Error of Mean	13.41	3.203	1.878
Lower 95% CI of mean	28.70	35.64	39.49
Upper 95% CI of mean	103.1	49.91	47.99



	AGR2-Patient (N=5)	CFTR-Patients (N=11)	Internal healthy controls (N=10)
Number of values	5	11	10
Minimum	51.30	43.10	49.80
25% Percentile	64.70	50.90	58.20
Median	110.6	55.70	62.05
75% Percentile	148.1	69.60	67.65
Maximum	155.3	78.90	71.00
Mean	107.2	59.55	62.20
Std. Deviation	43.11	11.76	6.442
Std. Error of Mean	19.28	3.546	2.037
Lower 95% CI of mean	53.72	51.64	57.59
Upper 95% CI of mean	160.8	67.45	66.81

References

1. Trujillano D, Bertoli-Avella AM, Kumar Kandaswamy K, et al. Clinical exome sequencing: results from 2819 samples reflecting 1000 families. Eur J Hum Genet 2017;25:176-82.
2. Cozma C, Iurașcu MI, Eichler S, et al. C26-Ceramide as highly sensitive biomarker for the diagnosis of Farber Disease. Sci Rep 2017;7:6149.
3. Papon JF, Coste A, Roudot-Thoraval F, et al. A 20-year experience of electron microscopy in the diagnosis of primary ciliary dyskinesia. Eur Respir J 2010;35:1057-63.
4. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res 2019;47:D886-d94.