



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 Pereira,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 Bauer¹

► 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



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

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 1938³ and was originally known as mucoviscidosis, given the observation that the patients presented abnormally thick mucus (reviewed in a previous work⁴).

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://www.centogenecom/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/Databank¹⁴ 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 34052907 reads ($>Q30$) were obtained for each sample. RNA-seq reads were aligned using two-pass mode with STAR V.2.7.6a¹⁵ against human genome GRCh37/Release 38 (www.encodegenes.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 YASARA¹⁹ and WHAT IF software²⁰ 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

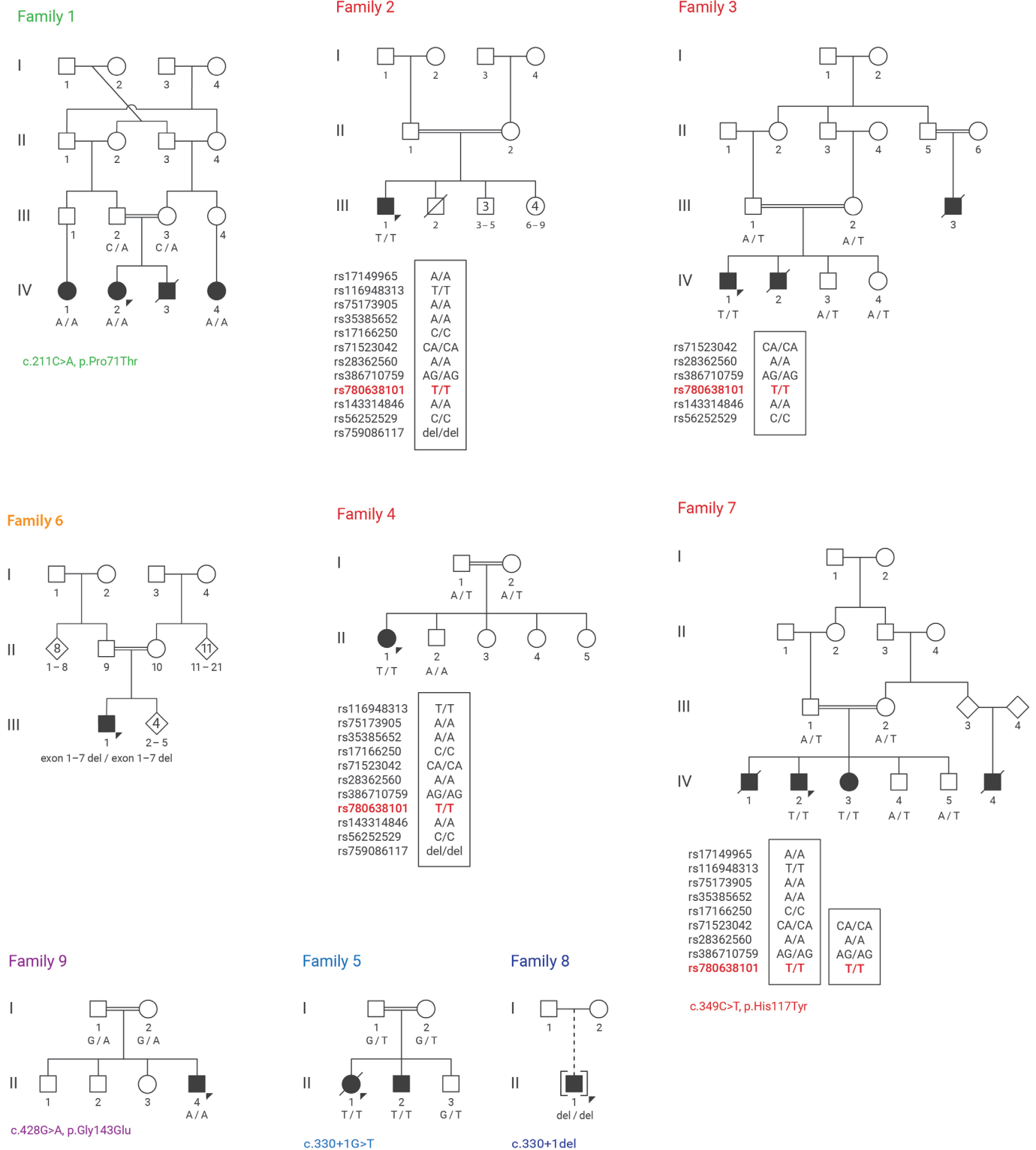


Figure 1 Summarised family trees of the nine families and the identified *AGR2* variants. Variants are colour-coded, the founder missense variant is shown in red font, with the corresponding haplotypes (families 2, 3, 4 and 7). Genotypes are shown below available individuals. *AGR2* genotypes show full 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 39756 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 down-regulated 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 cells²⁷ 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.

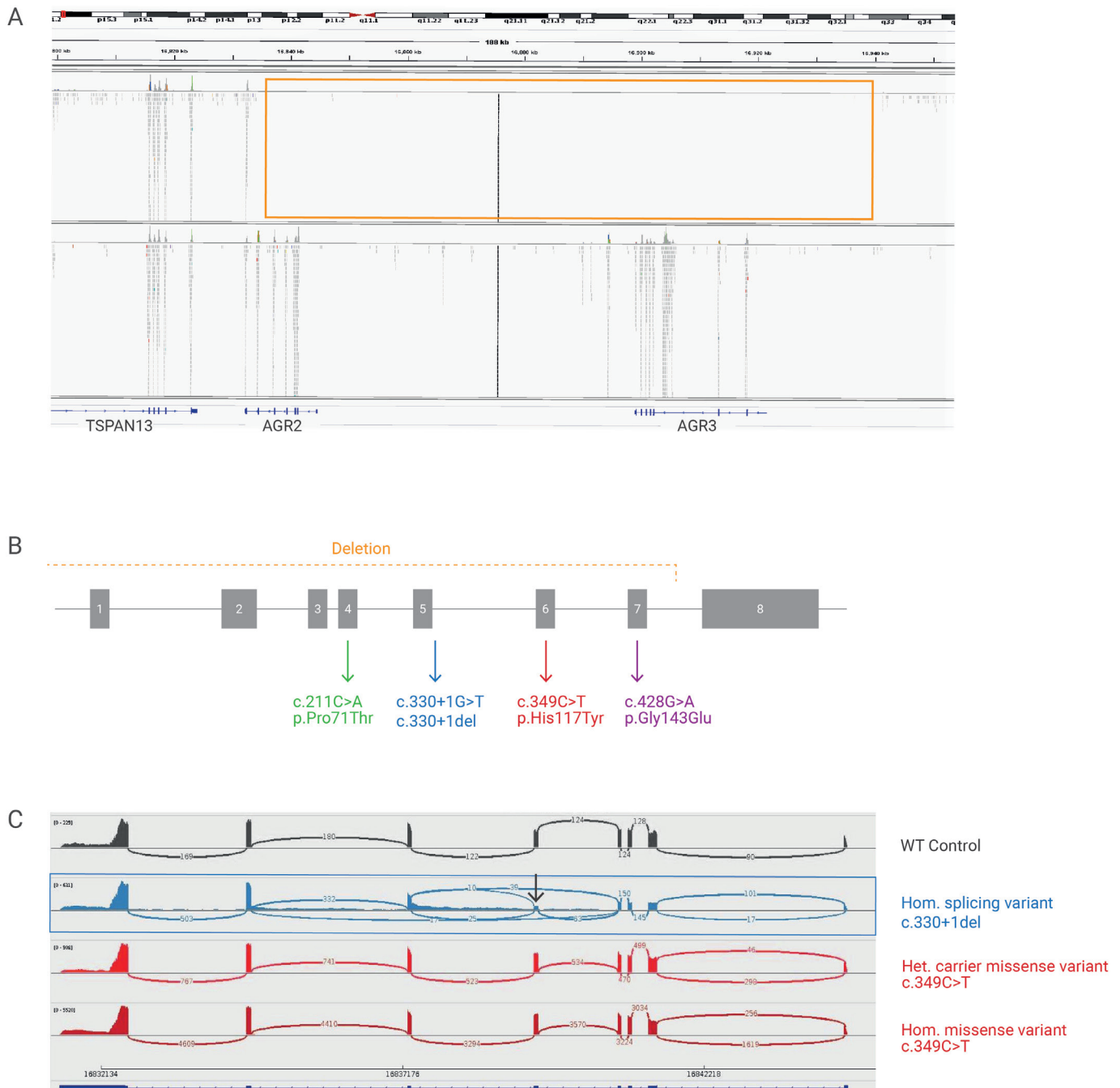


Figure 2 *AGR2* variants identified in the patients and abnormal splicing caused by an intronic variant. (A) The deletion region of *AGR2/3* is shown in the Integrative Genome Viewer (IGV). Reads for the exonic regions of the *AGR2/3* genes can be seen in the control (lower panel), whereas no reads are 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 represent 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/elasticase 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/elasticase tests).

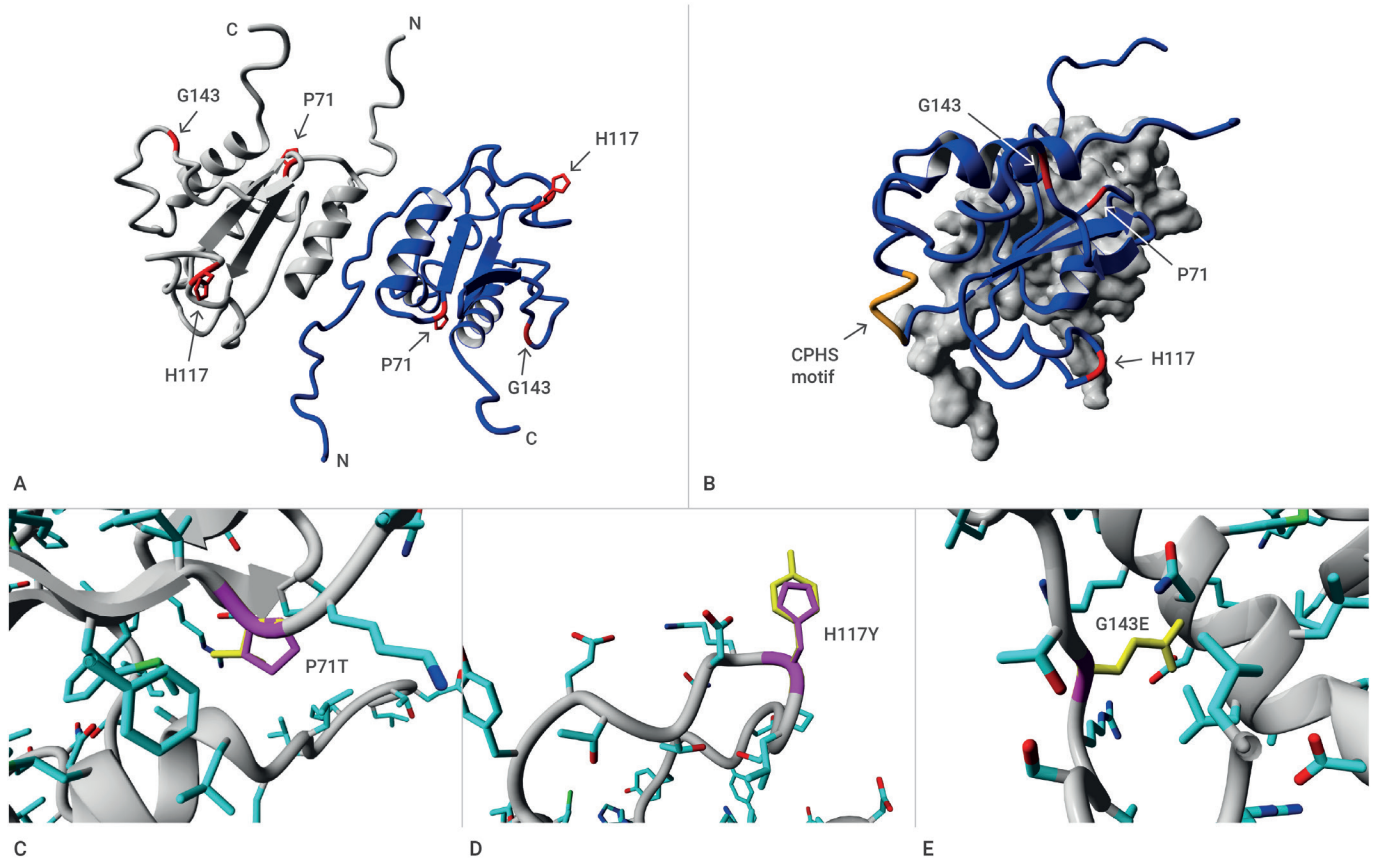


Figure 3 Structural protein analysis of the missense variants. (A) Dimer of the AGR2 residues 36–175. Monomers are individually coloured in grey or blue. Side chains of the residues are not shown, except for the mutated P71, H117 and G143 in red. (B) Overview of AGR2 as seen from the side; one monomer 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 work²³). 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 work⁴¹). 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

- ¹Medical Reporting & Genomic Research, Centogene GmbH, Rostock, Germany
- ²Child Health Department, The Royal Hospital, Muscat, Oman
- ³Department of Pediatric Pulmonology, The Royal Hospital, Muscat, Oman
- ⁴Pediatric Neurology and Metabolic division, Cairo University Childrens Hospital, Cairo, Egypt
- ⁵Department of Pediatrics, Salmaniya Medical Complex, Manama, Bahrain
- ⁶Farabi Medical Laboratory, Erbil, Iraq
- ⁷Hawler Medical University College of Medicine, Erbil, Kurdistan, Iraq
- ⁸Pulmonology Unit, Department of Pediatrics, College of Medicine, King Saud University, Riyadh, Saudi Arabia
- ⁹Medical 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
- ¹⁰Genetics 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
- ¹¹Immunology Research laboratory, Department of Pediatrics, College of Medicine and King Saud University Medical City, King Saud University, Riyadh, Saudi Arabia
- ¹²Centre for Molecular and Biomolecular Informatics, Radboudumc, Radboud Institute for Molecular Life Sciences, Nijmegen, Gelderland, Netherlands
- ¹³Department of Pediatric Gastroenterology, Children's Hospital of Lahore, Lahore, Pakistan
- ¹⁴Medical Genetics Department, College of Medicine, King Saud University, Riyadh, Saudi Arabia
- ¹⁵Department of Pathology, College of Medicine, King Saud University Medical City, King Saud University, Riyadh, Saudi Arabia
- ¹⁶Department 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 enthusiastic 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.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

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

- 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
- 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
- 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
- Asay LD. Cystic fibrosis. *Calif Med* 1965;102:292–300.
- 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
- Rose MC, Vovnow JA. Respiratory tract mucin genes and mucin glycoproteins in health and disease. *Physiol Rev* 2006;86:245–78. doi:10.1152/physrev.00010.2005
- 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
- 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
- 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
- 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
- Park S-W, Zhen G, Verhaeghe C, Nakagami Y, Nguyen 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
- 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
- Schroeder BW, Verhaeghe C, Park S-W, Nguyen 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
- 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
- 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
- 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
- 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
- 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
- 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
- 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
- Cozma C, Iurascu 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
- 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
- Delom F, Mohtar MA, Hupp T, Fessard D. The anterior gradient-2 interactome. *Am J Physiol Cell Physiol* 2020;318:C40–7. doi:10.1152/ajpcell.00532.2018
- Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi 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, Ferreira 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
- Patel P, Clarke C, Barraclough DL, Jowitz 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
- Guilbault C, Wojewodka G, Saeed Z, Hajdich 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
- Loberio N, Mancini G, Bassi R, Carsana EV, Tamanini A, Pedemonte N, Dececchi 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
- 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
- Tilley AE, Walters MS, Shaykhiiev R, Crystal RG. Cilia dysfunction in lung disease. *Annu Rev Physiol* 2015;77:379–406. doi:10.1146/annurev-physiol-021014-071931
- Al-Shaibi AA, Abdel-Motal UM, Hubrack SZ, Bullock AN, Al-Marri AA, Agrebi N, Al-Subaiey AA, Ibrahim NA, Charles AK, Elawad M, Uhligh 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, Bouche-careilh M, Nguyễn 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.2014-03180C
- 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