ARTICLE

# A family with ulcerative colitis maps to 7p21.1 and comprises a region with regulatory activity for the aryl hydrocarbon receptor gene

Hans Eiberg[1,7]✉, Josephine B. Olsson[2,3,7], Mads Bak[4], Claus Heiner Bang-Berthelsen[5], Jesper T. Troelsen[2] and Lars Hansen[6]



We have mapped a locus on chromosome 7p22.3-7p15.3 spanning a 22.4 Mb region for ulcerative colitis (UC) by whole genome linkage analyses of a large Danish family. The family represent three generations with UC segregating as an autosomal dominant trait with variable expressivity. The whole-genome scan resulted in a logarithm of odds score (LOD score) of Z = 3.31, and a whole genome sequencing (WGS) of two affected excluded disease-causing mutations in the protein coding genes. Two rare heterozygote variants, rs182281985:G>A and rs541426369:G>A, both with low allele frequencies (MAF A:0.0001, gnomAD ver3.1.2), were found in clusters of ChiP-seq transcription factors binding sites close to the *AHR* (aryl hydrocarbon receptor) gene and the UC associated SNP rs1077773:G>A. Testing the two SNPs in a promoter reporter assay for regulatory activity revealed that rs182281985:G>A influenced the *AHR* promoter. These results suggest a regulatory region that include rs182281985:G>A close to the UC GWAS SNP rs1077773:G>A and further demonstrate evidence that the *AHR* gene on the 7p-tel region is a candidate susceptible gene for UC.

*European Journal of Human Genetics* (2023) 31:1440–1446; https://doi.org/10.1038/s41431-023-01298-9

## INTRODUCTION

Inflammatory Bowel Disease (IBD) is a chronic idiopathic inflammatory disease that includes Crohn's disease (CD) and ulcerative colitis (UC). CD can affect any part of the gastrointestinal tract, but the terminal ileum and colon are the most commonly affected, while UC only affects the colon and rectum [1–3].

The incidence rate of IBD were increasing in the western countries in the second half of the twentieth century and is now stabilizing, whereas the incidence rate in newly industrialized countries is accelerating, and IBD has become a worldwide disease. This emphasizes the need for more research into IBD prevention and treatment [4]. IBD is a multifactorial common disease and the pathogenesis involves a complex interaction between genetic and environmental factors that is not well understood [5, 6]. In up to 25% to 30% of the cases there is a positive family history [7, 8], furthermore first-degree relatives of patients with UC or CD have a tenfold increase in the risk of developing the same disease [9]. A large meta-analysis of monozygotic and dizygotic twin pairs found CD rates of 30.3% versus 3.6% and UC rates of 15.4% versus 3.9%, demonstrating the importance of genetics in both CD and UC risk [10].

Genome-wide association studies (GWAS) have identified more than 240 loci, mostly polymorphic SNPs that show a significant association with UC/IBD [11–15]. In 2014, Ellinghaus et al., discovered that 67.5% of the identified IBD loci were shared between CD and UC, suggesting a heterogeneous and continuous disease spectrum [5], and in OMIM (OMIM 266600) 31 loci have been confirmed or proposed for IBD. The genes involved serve a variety of functions including innate immune response, adaptive immune response activation and regulation [16–21].

In this study we report a large Danish family where UC segregate as an autosomal dominant inherited trait with variable expressivity. A region of 22 Mb on chromosome 7p22.3-7p15.3 was mapped by whole genome linkage analysis, and two affected family members were Whole Genome Sequenced (WGS). GWAS data for UC in the linkage region combined with the WGS data suggested *AHR* (the aryl hydrocarbon receptor gene, OMIM 600253) and a rare SNPs close to the GWAS SNP, rs1077773:G>A, to be involved in UC.

## MATERIALS AND METHODS

### Family data

The study protocols adhered to the tenets of the Declaration of Helsinki and the ARVO statement on human subjects and approved by the Danish National Committee on Health Research Ethics in 2019 (H-19019167). All participants were informed orally and provided written consent. The family was recruited from Copenhagen Family Bank [22] and represent five individuals with diagnosed UC, two undiagnosed individuals with UC symptoms and seven healthy individuals (Fig. 1A). The family was sampled in the period 1982-2022 and clinical data was from their medical doctor and hospital.

[1]RCLINK, Department of Cellular and Molecular Medicine, Panum Institute, University of Copenhagen, Blegdamsvej 3, DK-2200 Copenhagen N, Denmark. [2]Department of Science and Environment, Roskilde University, Roskilde, Denmark. [3]Department of Clinical Immunology, Zealand University Hospital, Naestved, Denmark. [4]Department of Clinical Genetics, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark. [5]Research Group for Microbial Biotechnology and Biorefining, National Food Institute, Technical University of Denmark, Kemitorvet building 202, 2800 Kgs Lyngby, Denmark. [6]Department of Cellular and Molecular Medicine, Panum Institute, University of Copenhagen, Blegdamsvej 3, DK-2200 Copenhagen N, Denmark. [7]These authors contributed equally: Hans Eiberg, Josephine B. Olsson. ✉email: he@sund.ku.dk

Received: 16 August 2022 Revised: 16 November 2022 Accepted: 18 January 2023
Published online: 3 February 2023

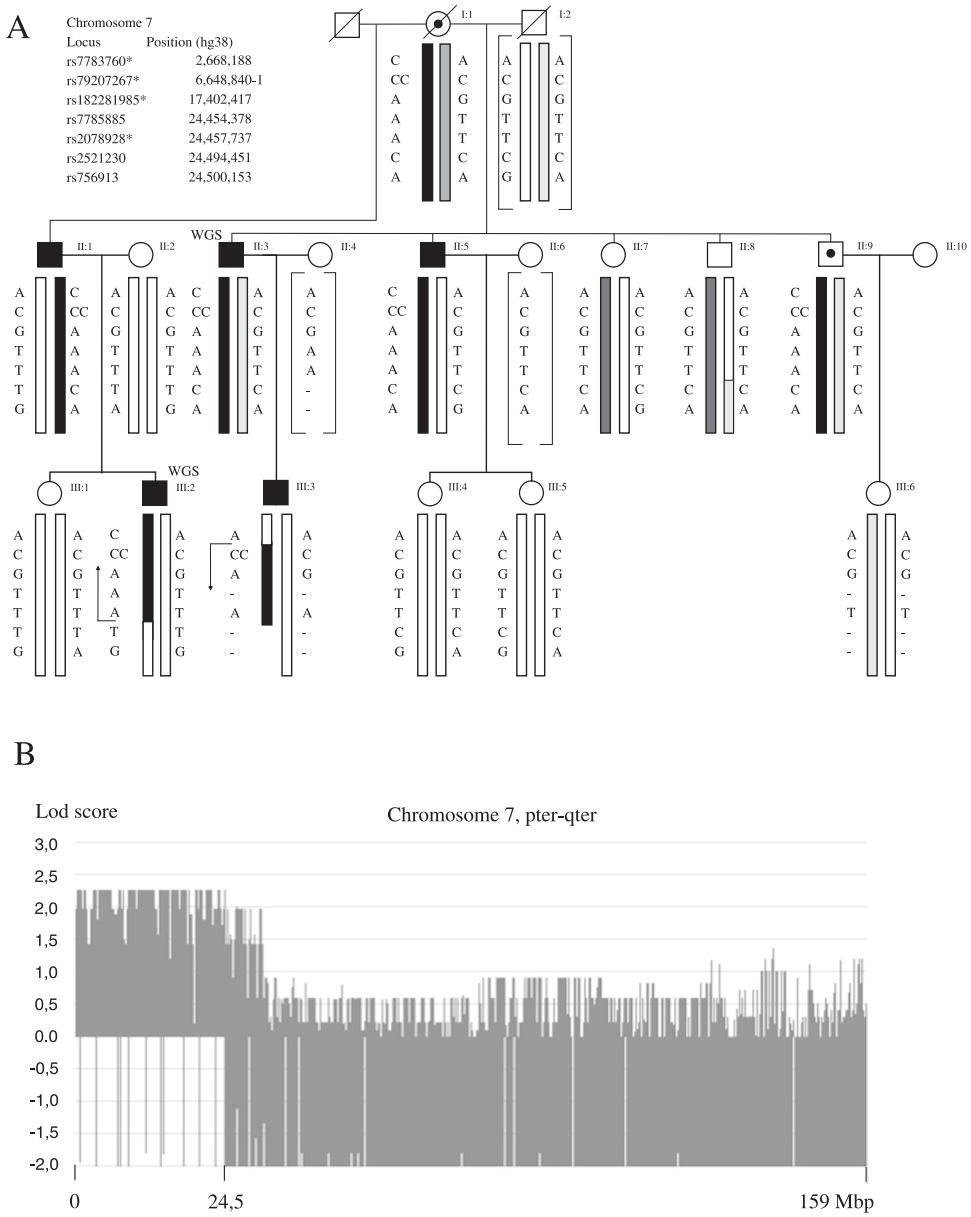


**Fig. 1 The pedigree of the large Danish family and plot of the chromosome 7 LOD scores. A** The pedigree represents the three generation's family and the UC haplotype for the DNA markers in the mapped linkage region. The SNP markers are spanning from 2.7 Mb to 24,5 Mb telomeric on 7p; recombination events are marked by arrows for III:2 and III:3, males are shown by squares, females as circles, healthy individuals as open symbols, affected as filled symbols, and undiagnosed individuals with UC symptoms are shown with a dot in an open symbol. Individuals WG sequenced (II:3 and III:2) are denoted WGS, individuals where the haplotypes are from Sanger sequencing of SNPs are denoted *. The following individuals were genotyped by SNP array: I:1, II:1-3,5,7-9, III:1,2,4,5, four different SNPs (rs7783760:A>C, rs79207267:dup, rs182281985:G>A, rs2078928:A>G) were genotyped by Sanger sequencing for all individuals in the family. **B** Graphic presentation of the LOD scores revealed one segment on chr7p-tel with Z > 2.0 as a continuous region. The plot shows the initial mapping of the linkage region done by SNP array data with a LOD score of Z = 2.26. The negative LOD scores (single lines) seen in the linkage regions are genotyping errors likely due to DNA degradation.

## Whole genome linkage analyses

DNA was extracted using EDTA (ethylenediamine tetra-acetic acid) blood by standard phenol/chloroform extraction protocols and genotyping was carried out by genome-wide SNP microarray (CytoScan™ HD, Affymetrix) for 12 family members (I:1, II:1, II:2, II:3, II:5, II:7, II:8, II:9, III:1, III:2, III:4 and III:5). The CytoScan™ HD genotyping resulted in approximately 800.000 markers. An initial two-point parametric LOD score calculation for the microarray data was done using the linkage program LIPED [23] with an allele frequency of 0.01 for the disease gene and a haplotype frequency of 0.5, and both I:1 and II:9 were set as affected. The Birdseed text files were converted to a single input file and analyzed by LIPED. Calculated LOD scores were sorted by chromosome and position using the SNP6-LINK program package and analyzed by graphic presentation. The resulting 24 chromosome files were analyzed for contiguous regions with LOD scores >2.0 excluding all other regions in the genome (Fig. 1B). Subsequently, the two additional family members (III:3 and III:6) were included in the analysis and additional four informative SNPs (rs7783760:A>C, rs79207267:dup, rs182281985:G>A and rs2078928:A>G) were genotyped in all family members by Sanger sequencing (ABI Big Dye ver1.1 and ABI3100 sequenator, Applied Biosystems) for further fine mapping of the candidate region. The following calculations of LOD score after inclusion of two new family members and four new SNPs was done with a disease haplotype frequency of 0.001 that rs182281985:G>A is rare (MAF = 0.0001, European (non-Finnish), gnomAD v3.1.2).

**Table 1.** Data for the UC family.

| Individual | Phenotype | Age of first symptoms[a] | Age with bloody stool | Age and action of surgery |
|---|---|---|---|---|
| I:2 | UC-symptoms | – | – | – |
| II:1 | UC | 22 | 39 | 54 yrs., hemicolectomy |
| II:3 | UC | 45 | 45 | – |
| II:5 | UC | 13 | 13 | 16 yrs., colectomy with ileostomy |
| II:9 | UC-symptoms | 40 | – | – |
| III:2 | UC | 16 | 23 | – |
| III:3 | UC | 25 | 35 | – |

[a]Frequent and/or urgent bowel movements, abdominal pain, and cramping, but without bloody stool.

### Whole genome sequencing and data analysis

WGS was done by BGI Europe (BGI, Copenhagen Denmark). Briefly a ≤ 800 bp insert normal library was created for the affected individuals II:3 and III:2 and reads were aligned to human reference sequence hg19, GRCh37 using the BWA (0.7.15) aligner [24]. Variant calling was done employing GATK (4.0.11.0) [25], and variant annotation and filtration was done using VarSeq (Golden Helix, USA) and a minimum coverage of 20 reads. The WGS data were analyzed employing the Variant Effect Predictor (VEP, Ensembl) and Provean for Genomic Variants [26], and heterozygous SNPs with MAF (minor allele frequency) values <0.01 were selected and analyzed for affecting coding regions or putative regulatory elements employing GeneHancer [27] and ENCODE Chip-seq data embedded in the UCSC browser [28, 29].

### Cell culture and promoter reporter assays

Caco-2 cells were used for transient transfection analyzing the *AHR* promoter and regulatory activity of two selected SNPs (rs182281985:G>A and rs541426369:G>A). The cells were grown in Dulbecco's modified essential medium (DMEM) with L-Glutamine (Sigma), 10% FBS (HyClone) and 1% PS (Lonza) and sub-cultivated twice a week incubated at 37 °C, 5% CO2. The human *AHR* promoter region was PCR amplified using gDNA from II:1 and cloned into the *Hin*DIII site of the luciferase reporter plasmid pGL4.10 (Promega) using the In-Fusion Cloning strategy [30] generating the plasmid pAHRprom. Subsequently, the putative regulatory regions (reg1) including rs182281985:G>A and (reg2) rs541426369:G>A from II:1 were cloned into the *Bam*HI site of pAHRprom. All clones were bidirectional Sanger sequenced and plamids representing the reference and the alternative alleles were selected. The resulting plamids were named pAHRprom, pAHRprom_reg1(G)/(A) and pAHRprom_reg2(G)/(A). Primer sequences, PCR conditions and genomic positions can be found in Supplementary data.

Caco2 cells were seeded in 24 well plates at a density of $5 \cdot 10^4$ cells per well and transfected with the plasmid constructs and replicated 4 times with a total DNA concentration of 1.2 µg: 0.2 µg AHR promoter construct, 0.1 µg CMV LacZ, and 0.9 µg pSK+ diluted in 150 mM NaCl. Transfection was done using 2 µM PEI in 150 mM NaCl. The luminescence ratio was calculated by dividing the luciferase data by the β-galactosidase data to account for the efficiency of transfection. The promoter reporter assay was performed according to [31], and Graphpad Prism 9 was used to analyze the promoter assay data and statistical significance was determined using One-way ANOVA.

## RESULTS

### The family

The Copenhagen Family Bank [22] included a three-generations Danish family where UC segregates as an autosomal dominant trait over three generations. Inclusion criteria was an UC diagnosis or as undiagnosed but with UC symptoms. Five individuals (II:1, II:3, II:5, III:2, III:3) had an UC diagnosis from their own medical doctors and hospitals and clinical date was achieved be interviewing the family. Two members (I:1 and II:9) were undiagnosed but had UC like symptoms and all were considered as affected in the analysis (Fig. 1A). The grandmother (I:1) was deceased but by the family reported with UC symptoms, II:9 was reported with UC symptoms by his own medical doctor. The disease for all affected starts with frequent and/or urgent bowel movements, abdominal pain, and cramping (UC like symptoms). Both II:1 and II:5 was under hospital treatments, II:1 got a hemicolectomy surgery at age of 54 years, and for II:5 the first sign of UC came at age 13 years followed by colectomy with ileostomy tree years later (Table 1). None of the family members were diagnosed for colon cancer. Based on the clinical data, the family was analyzed as a family with UC segregating as autosomal dominant with variable expressivity.

### The whole genome-scan

The genome-scan included 12 of the family members and additional 2 individuals were later genotyped for informative markers by Sanger sequencing. The initial linkage analysis included all 12 individuals genotyped by SNP microarray assuming I:1 having the UC phenotype and the LOD score calculations were done for all chromosomes with an allele frequency of 0.01 for the disease gene and a allele frequency of 0.5 for the SNPs. The initial linkage analysis resulted in a 24 Mb large telomeric region on chromosome 7p (7p22.3-7p15.3) having a LOD score of $Z = 2.26$ and the 7p-tel locus was the only continuous region with positive LOD scores (Fig. 1B) excluding the remaining part of the genome. Fine-mapping of the region included genotyping of four informative SNPs (rs7783760:A>C, rs79207267:dup, rs182281985:G>A and rs2078928:A>G) in all family members including III:3 and III:6, and resulted a final LOD score of $Z = 3.31$ between the disease and the segregating haplotype.

The final genotyping defined a region between rs7783760:A>C and rs2521230:T>A corresponding to 22.4 Mb having the coordinates chr7:2,668,188-24,494,451 (hg38) due to recombination between rs7783760:A>C and rs79207267:dup in III:3 and between rs2078928:A>G and rs2521230:T>A in III:2 (Fig. 1A).

### GWAS loci and WGS variants

Data mining the GWAS catalog resulted in seven GWAS studies reporting SNPs associated with UC or IBD in the linkage region (Table 2). The SNP positions were at 2.8 Mb (rs798502:A>C and rs1182188:T>C) [11–13, 32, 33] at 6.5 Mb (rs11768365) [13] and in a region at 17-20 Mb (rs1077773:G>A and rs11764116:G>A) [11, 13, 34, 35].

Two individuals, II:3 and III:2, that represent two distant branches in the family were chosen for WG sequencing, and the following analyses revealed more than 45,000 variants in the linkage region. The data were analyzed for heterozygous variants having MAF values <0.01 found in both individuals. The initial analyses excluded all protein coding regions leaving variants in introns and intergenic regions as candidates. The variants were therefor analyzed for position in putative regulatory regions employing ENCODE Chip-seq data, DNaseI hypersensitivity data and GeneHancer data from the UCSC browser. Regions in the proximity of published UC GWAS SNPs were selected as candidate regions for variants and two candidates, rs182281985:G>A and

**Table 2.** Published GWAS SNP's associated with UC in 7p22.3-7p21.1.

| SNP | Trait | II:3 | III:2 | Allele frequency[a] | Haplotype in the family | Reported / mapped genes | Reference (risk allele) |
|---|---|---|---|---|---|---|---|
| rs798502 (A/C) | UC, IBD | A/A | A/C | A:0.7088 | A | *AMZ1, GNA12, CARD11, TTYH3* | [13] (not reported)<br>[32] (A)<br>[33] (A) |
| rs1182188 (A/G) | UC, IBD | A/A | A/G | A:0.7031 | A | *GNA12* | [11] (A)<br>[12] (not reported) |
| rs11768365 (A/G) | UC, IBD | A/A | A/A | A:0.7780 | A | *FLJ20306, DAGLB, KDELR2, GRID2IP* | [13] (A) |
| rs1077773 (G/A) | UC, IBD | G/A | G/A | A:0.5237 | G[b] | *AHR* | [11] (A)<br>[13] (A)<br>[35] (G) |
| rs11764116 (G/T) | UC | G/G | G/G | G:0.7893 | G | *HDAC9* | [34] (A) |

[a]European (non-Finnish) population, gnomAG v3.1.2, common allele.
[b]The G allele for rs1077773 was identified cis to rs182281985 A by sequence analyses of PCR products for regulatory region 1.

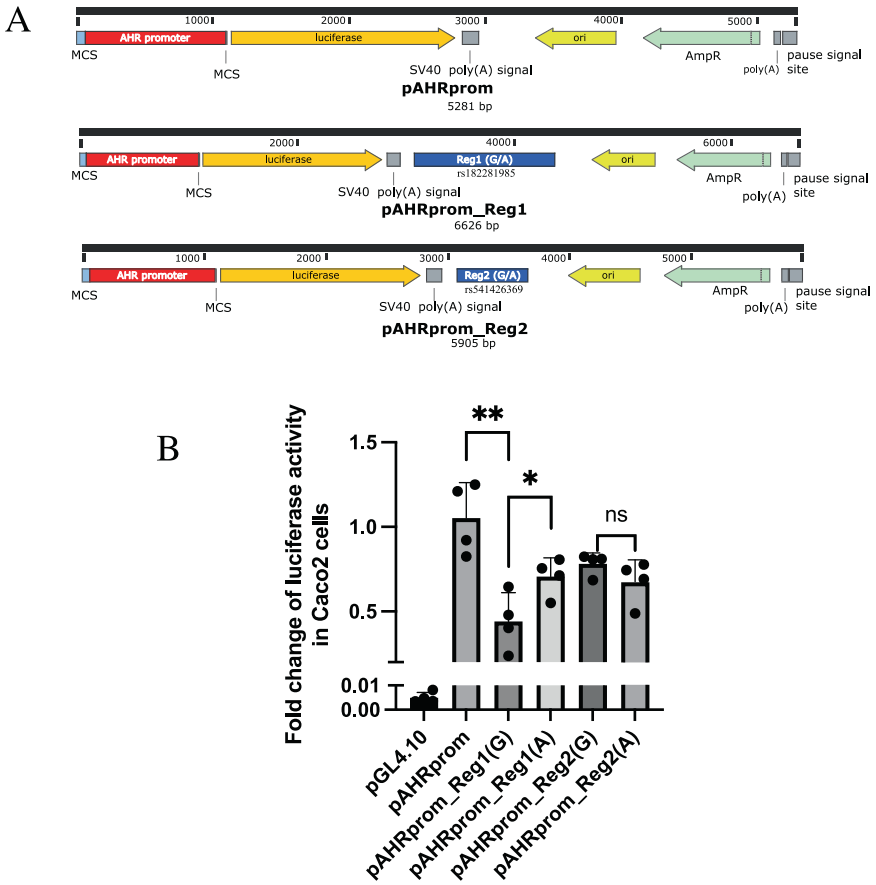


**Fig. 2 Schematic presentation of vector constructs uses for the AHR promoter assays and expression results. A** Plasmid maps of the five pGL4.10 constructs used in the promoter expression assay showing position of the *AHR* promoter and orientation of the two putative regulatory segments. **B** Bar chart of results from promoter reporter assay of *AHR* constructs. The mean ±standard deviation and the data points are shown for the five different constructs and the pGL4.10 vector. pAHRprom_Reg1(G)/(A) is the vector with the AHR promoter and the segment harboring rs182281985:G>A and pAHRprom_Reg2(G)/(A) is the vector with the AHR promoter and the segment with rs541426369:G>A. The data is normalized to pAHRprom. $N = 4$, significance levels: $^*P < 0.05$, $^{**}P < 0.01$ and ns not significant.

rs541426369:G>A, were selected for further study as the only variants in the WGS data that adhered to the filtration criteria and were in the vicinity of a GWAS locus. These two SNPs were located close to the UC SNP rs1077773:G>A, rs182281985:G>A separated by 638 bp and rs541426369:G>A separated by 2,836 bp. Both SNPs were in clusters of transcription factor binding sited predicted by Chip-seq data and in DNaseI hypersensitivity regions, and both had a MAF value of A:0.0001 (European (non-Finnish), gnomAD ver3.1.2) and are approximately 40 kb downstream for the IBD susceptibility candidate gene *AHR* [11, 13, 35]. The WGS did not reveal any candidates close to other UC GWAS SNPs in the region.

One interesting question is, if the risk alleles from the seven GWA studies were found in the mapped linkage region. Combining the WGS data for II;3 and III:2 and the mapped disease

haplotype made it possible to predict the haplotype for the five GWAS SNPs. Risk alleles were reported for all five UC GWAS SNPs represented by an A-allele for rs798502, rs1182188, rs11768365 and rs11764116 [11, 13, 32–34], and both an A and a G-allele for rs1077773 [11, 35]. All risk alleles represent common variants (Table 2) and could be found in the mapped disease haplotype with the exception of rs11764116:A. The significance of finding the risk alleles in the disease haplotype is limited that they mainly represent the common allele.

### AHR promoter constructs and activities

A possible regulatory effect on the *AHR* gene of the two SNPs was analyzed in a promoter expression assay using the luciferase expression vector pGL4.10 and Caco2 human epithelial colon cells. Five vector constructs in total were made comprising a 1,032 bp segment of the *AHR* promoter, two constructs with a 1,338 bp fragment carrying the two alleles of rs182281985:G>A and two constructs with a 618 bp fragment for the two alleles of rs541426369:G>A, respectively, all cloned after the reported gene in the *AHR* promoter construct (Fig. 2A). The promoter segment covered a region of 467 bp upstream for transcription start of *AHR* and 565 bp of the 5'UTR, the segment carrying the two alleles of rs182281985:G>A included rs1077773:G>A. The five constructs and the empty pGL4.10 vector was transfected in Coca2 cells and each assay was replicated 4 times and normalized to the *AHR* promoter construct.

The results from the promoter reporter assays are shown in Fig. 2B. The analyses demonstrated the AHR promoter (pAHR-prom) to be highly active in Caco2 cells. The constructs for rs541426369:G>A (pAHRprom_Reg2(G) and (A)) and the A allele of rs182281985:G>A, (pAHRprom_Reg1(A)) had similar expression corresponding to 75% of the AHR promoter. The construct for the G allele of rs182281985:G>A (pAHRprom_Reg1(G)) had expression corresponding to 50% of the AHR promoter suggesting the G allele to serve a regulatory function on the AHR promoter. It is notable that the G-allele is the major allele of rs182281985:G>A that seems to be involved in regulation of the AHR gene, and the outcome of the expression assay suggests a regulatory element including rs182281985:G>A approximately 40 kb downstream for AHR and close to rs1077773:G>A.

## DISCUSSION

We have in the present study analyzed a Danish family where UC segregates as a Mendelian autosomal dominant trait. The grandmother (I:1) was reported with undiagnosed bowel problems, and the UC phenotype segregates in both branches in two generations originating with the grandmother. The UC trait demonstrate variable expressivity that two individual (I:1 and II:9) are reported with uncharacterized UC symptoms and the disease has different debut age for symptoms, severness and colectomy recognized as early as age 13 and 16 years (II:5 and III:2) and colectomy at age 16 years (II:5) and age 54 years (II:1). The data summerized in Table 1 suggeted inheritable UC segregating in the family.

The genome-wide scan, done by micro-arrays genotype analysis of 12 family members, resulted in a 24 Mb linkage region with a LOD score of $Z = 2.26$ on chromosome 7p21.1 (Fig. 1A, B). Including two additional family members in the analysis by Sanger sequencing four informative SNPs, a LOD score of $Z = 3.31$ was obtained for a defined linkage region of 22.4 Mb. Setting the two undiagnosed individuals I:2 and II:9 as unknown with respect to the disease, the LOD scores is still significant ($Z = 3.01$). WGS of two affected excluded mutations in the gene coding regions and made a mutation in a regulatory element most likely. Two rare SNPs proximate to the UC GWAS SNP rs1077773:G>A were tested for regulatory impact on the AHR gene and suggested a putative regulatory element including the SNP rs182281985:G>A.

UC and IBD is characterized as a multifactorial common disease which is polygenic and involves more than one gene in progression of the disorder [36]. Several studies have found linkage of UC and CD to specific chromosomal regions as reviewed by Mathew and Lewis 2004 [37] and these studies suggested and support that single genetic components are involved development of IBD and UC, but only the NOD2 gene on chromosome 16 has been reported as a susceptibility locus with identified mutations [38–43].

A serie of GWA studies of large cohorts with IBD, UC or CD has expanded the number susceptible candidate loci to more than 200 with different burden. The AHR gene in the 7p-tel linkage region has been suggested as a candidate by Liu et al. in a large study including both UC and IBD study groups and both Liu et al. and de Lange et al. have found association between UC and the SNP rs1077773 close to AHR [11, 13]. The association between the SNP rs1077773:G>A, AHR and UC made us investigate rare variants in the vicinity of the GWAS SNP.

The two SNPs rs182281985:G>A and rs541426369:G>A fulfilled our criteria as possible candidates for further investigation. Both SNPs are rare variants in regions with clusters of transcription factor binding sites approxoimately 40 Kb downstream for AHR. The promoter expression assay done in Coco2 cells revealed a reduction of transcriptional activity for the G-allele of rs182281985:G>A, an effect not observed for the A allele or for alleles of rs541426369:G>A. The result suggested a AHR regulatory element close to the UC associated SNP rs1077773:G>A and the A allele segragating in the family seems to repeal repression of AHR.

As suggested by GWAS data, *AHR* is a candidate susceptible gene for UC. AHR is a cytosolic transcription factor that is ligand-activated and binds a wide range of synthetic and natural molecules [44–46] and is highly expressed on Th17 cells where ligand binding triggers the Th17 cells to produce more cytokines, including IL-22 [11, 47]. It is shown in a clinical trial that activation of the IL-22 pathway through AHR results in effective remission in UC patients [48]. Low *AHR* activity, caused by a deficiency in the receptor or the receptor's ligand, causes the intestine to be in a state of increased immune activation [44] and disrupts intraepithelial lymphocyte homeostasis. Inadequate control of intestinal microbial load and composition, as well as increased immune activation, will result in epithelial damage [11]. Metidji et al. demonstrated in 2018 that *AHR* was a highly important factor in the regeneration of the intestinal epithelial cells [49], furthermore, Benson and Shephard demonstrated that activating *AHR* resulted in the generation of regulatory immune cells, resulting in a decrease in colonic inflammation [44]. Finally, AHR is involved in the regulation of intestinal homeostasis, particularly in terms of immune aspects, according to data from animal models [46] and suggests that *AHR* is a promising target gene for diagnosing and treatment of patients with UC in the future [50].

Combining 7p-tel linkage region with *AHR* and a variant segregating in the family that may be involved in *AHR* regulation further confirm a locus for UC in the region. These findings are in line with several GWA studies that find association with UC in the 7p-tel region and further suggest *AHR* as a candidate gene (Table 2). In addition, all risk alleles reported in the GWAS except for one could be found in the UC haplotype. More studies combining genetic data and the role of *AHR* in development of IBD/UC and treatment of inflammatory bowel are needed to clarify this role of the *AHR*.

In summary, we have mapped a 22.4 Mb region on 7p22.3-7p15.3 in a large family with a LOD score $Z = 3.31$ for an autosomal dominant UC form. Combination of WGS data and promoter expression analysis of the SNP rs182281985:G>A revealed a putative regulatory element for *AHR*. The location of the putative regulatory element close to the UC GWA SNP rs1077773 further emphasize *AHR* as candidate for a UC susceptibility gene.

## Web resources

GWAS catalog, https://www.ebi.ac.uk/gwas/
GnomAD, https://gnomad.broadinstitute.org/
SNP6-LINK package, https://icmm.ku.dk/english/research-groups/eiberg-group/snp6-link/index.html
UCSC Genome Browser, http://genome.ucsc.edu/cgi-bin/hgGateway

## DATA AVAILABILITY

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

## REFERENCES

1. Sairenji T, Collins KL, Evans DV. An update on inflammatory bowel disease. Prim Care. 2017;44:673–92. https://doi.org/10.1016/j.pop.2017.07.010.
2. Torres J, Mehandru S, Colombel JF, Peyrin-Biroulet L. Crohn's disease. Lancet. 2017;389:1741–55. https://doi.org/10.1016/S0140-6736(16)31711-1.
3. Ungaro R, Mehandru S, Allen PB, Peyrin L, Colombel J-F. Ulcerative colitis. Lancet. 2017;389:1756–70. https://doi.org/10.1016/S0140-6736(16)32126-2.
4. Ng SC, Shi HY, Hamidi N, Underwood FE, Tang W, Benchimol EI, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. Lancet. 2017;390:2769–78. https://doi.org/10.1016/S0140-6736(17)32448-0.
5. Ellinghaus D, Bethune J, Petersen BS, Franke A. The genetics of Crohn's disease and ulcerative colitis-status quo and beyond. Scand J Gastroenterol. 2014;50:13–23. https://doi.org/10.3109/00365521.2014.990507.
6. Bager P, Gørtz S, Feenstra B, Andersen NN, Jess T, Frisch M, et al. Increased risk of inflammatory bowel disease in families with tonsillectomy: a Danish national cohort study. Epidemiology. 2019;30:256–62. https://doi.org/10.1097/EDE.0000000000000946.
7. Santos MPC, Gomes C, Torres J. Familial and ethnic risk in inflammatory bowel disease. Ann Gastroenterol. 2018;31:14–23. https://doi.org/10.20524/aog.2017.0208.
8. Younis N, Zarif R, Mahfouz R. Inflammatory bowel disease: between genetics and microbiota. Mol Biol Rep. 2020;47:3053–63. https://doi.org/10.1007/s11033-020-05318-5.
9. Orholm M, Munkholm P, Langholz E, Nielsen OH, Sørensen TIA, Binder V. Familial occurrence of inflammatory bowel disease. N. Engl J Med. 1991;324:84–88. https://doi.org/10.1056/NEJM199101103240203.
10. Brant SR. Update on the heritability of inflammatory bowel disease: the importance of twin studies. Inflamm Bowel Dis. 2011;17:1–5. https://doi.org/10.1002/ibd.21385.
11. Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat Genet. 2015;47:979–86. https://doi.org/10.1038/ng.3359.
12. Ellinghaus D, Jostins L, Spain SL, Cortes A, Bethune J, Han H, et al. Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. Nat Genet. 2016;48:510–8. https://doi.org/10.1038/ng.3528.
13. de Lange KM, Moutsianas L, Lee JC, Lamb CA, Luo Y, Kennedy NA, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. Nat Genet. 2017;49:256–61. https://doi.org/10.1038/ng.3760.
14. Huang H, Fang M, Jostins L, Mirkov MU, Boucher G, Anderson CA, et al. Fine-mapping inflammatory bowel disease loci to single-variant resolution. Nature. 2017;547:173–8. https://doi.org/10.1038/nature22969.
15. Mirkov MU, Verstockt B, Cleynen I. Genetics of inflammatory bowel disease: beyond NOD2. Lancet Gastroenterol Hepatol. 2017;2:224–34. https://doi.org/10.1016/S2468-1253(16)30111-X.
16. Corbaz A, ten Hove T, Herren S, Graber P, Schwartsburd B, Belzer I, et al. IL-18-binding protein expression by endothelial cells and macrophages is up-regulated during active Crohn's disease. J Immunol. 2002;168:3608–16. https://doi.org/10.4049/jimmunol.168.7.3608.
17. Fuss IJ, Heller F, Boirivant M, Leon F, Yoshida M, Fichtner-Feigl S, et al. Non-classical CD1d-restricted NK T cells that produce IL-13 characterize an atypical Th2 response in ulcerative colitis. J Clin Invest. 2004;113:1490–7. https://doi.org/10.1172/JCI19836.
18. van Heel DA, Fisher SA, Kirby A, Daly MJ, Rioux JD, Lewis CM. Inflammatory bowel disease susceptibility loci defined by genome scan meta-analysis of 1952 affected relative pairs. Hum Mol Genet. 2004;13:763–70. https://doi.org/10.1093/hmg/ddh090.
19. Franke A, McGovern DPB, Barrett JC, Wang K, Radford-Smith GLAhmad T, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat Genet. 2010;42:1118–25. https://doi.org/10.1038/ng.717.
20. Yoneno K, Hisamatsu T, Shimamura K, Kamada N, Ichikawa R, Kitazume MT, et al. TGR5 signalling inhibits the production of pro-inflammatory cytokines by in vitro differentiated inflammatory and intestinal macrophages in Crohn's disease. Immunology. 2013;139:19–29. https://doi.org/10.1111/imm.12045.
21. de Lange KM, Barrett JC. Understanding inflammatory bowel disease via immunogenetics. J Autoimmun. 2015;64:91–100. https://doi.org/10.1016/j.jaut.2015.07.013.
22. Eiberg H, Nielsen LS, Klausen J, Dahlén M, Kristensen M, Bisgaard M, et al. Linkage between serum cholinesterase 2 (CHE2) and □-crystalline gene cluster (CRYG): assignment to chromosome 2. Clin Genet. 1989;35:313–321. https://doi.org/10.1111/j.1399-0004.1989.tb02951.x.
23. Ott J. A computer program for linkage analysis of general human pedigrees. Am J Hum Genet. 1976;28:528–9.
24. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60. https://doi.org/10.1093/bioinformatics/btp324.
25. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43:491–8. https://doi.org/10.1038/ng.806.
26. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics. 2015;31:2745–7. https://doi.org/10.1093/bioinformatics/btv195.
27. Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Stein TI, et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards, Database. 2017. https://doi.org/10.1093/database/bax028.
28. Sloan CA, Chan ET, Davidson JM, Malladi VS, Strattan JS, Hitz BC, et al. ENCODE data at the ENCODE portal. Nucleic Acids Res. 2016;44(D1):D726–32. https://doi.org/10.1093/nar/gkv1160.
29. Lee BT, Barber GP, Benet-Pagès A, Casper J, Clawson H, Diekhans M, et al. The UCSC Genome Browser database: 2022 update. Nucleic Acids Res. 2022;50(D1):D1115–D1122. https://doi.org/10.1093/nar/gkab959.
30. Raman M, Martin K. One solution for cloning and mutagenesis: In-Fusion® HD Cloning Plus. Nat Methods. 2014;11:iii–v. https://doi.org/10.1038/nmeth.f.373.
31. Dahlgaard K, Troelsen JT. Identification and functional analysis of gene regulatory sequences interacting with colorectal tumor suppressors. Methods Mol Biol. 2018;1765:57–77. https://doi.org/10.1007/978-1-4939-7765-9_4.
32. Anderson CA, Boucher G, Lees CW, Franke A, D'Amato M, Taylor KD, et al. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. Nat Genet. 2011;43:246–52. https://doi.org/10.1038/ng.764.
33. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 2012;491:119–24. https://doi.org/10.1038/nature11582.
34. Haritunians T, Taylor KD, Targan SR, Dubinsky M, Ippoliti A, Kwon S, et al. Genetic predictors of medically refractory ulcerative colitis. Inflamm Bowel Dis. 2010;16:1830–40. https://doi.org/10.1002/ibd.21293.
35. Burke KE, Khalili H, Garber JJ, Haritunians T, McGovern DPB, Xavier RJ, et al. Genetic markers predict primary nonresponse and durable response to anti-tumor necrosis factor therapy in ulcerative colitis. Inflamm Bowel Dis. 2018;24:1840–8. https://doi.org/10.1093/ibd/izy083.
36. Khor B, Gardet A, Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. Nature. 2011;474:307–17. https://doi.org/10.1038/nature10209.
37. Mathew CG, Lewis CM. Genetics of inflammatory bowel disease: progress and prospects. Hum Mol Genet. 2004;13:R161–8. https://doi.org/10.1093/hmg/ddh079.
38. Hugot J-P, Laurent-Puig P, Gower-Rousseau C, Olson JM, Lee JC, Beaugerie L, et al. Mapping of a susceptibility locus for Crohn's disease on chromosome 16. Nature. 1996;379:821–3. https://doi.org/10.1038/379821a0.
39. Hugot JP, Chamaillard M, Zouali H, Lesage S, Cézard J-P, Belaiche J, et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature. 2001;411:599–603. https://doi.org/10.1038/35079107.
40. Ogura Y, Bonen D, Inohara N, Nicolae D. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature 2001;411:603–6. https://doi.org/10.1038/35079114.
41. Hampe J, Frenzel H, Mirza MM, Croucher PJP, Cuthbert A, Mascheretti S, et al. Evidence for a NOD2-independent susceptibility locus for inflammatory bowel disease on chromosome 16p. Proc Natl Acad Sci USA. 2002;99:321–6. https://doi.org/10.1073/pnas.261567999.
42. Negroni A, Pierdomenico M, Cucchiara S, Stronati L. NOD2 and inflammation: current insights. J Inflamm Res. 2018;11:49–60. https://doi.org/10.2147/JIR.S137606.
43. Philpott DJ, Sorbara MT, Robertson SJ, Croitoru K, Girardin SE. NOD proteins: regulators of inflammation in health and disease. Nat Rev Immunol. 2014;14:9–23. https://doi.org/10.1038/nri3565.

44. Benson JM, Shepherd DM. Aryl hydrocarbon receptor activation by TCDD reduces inflammation associated with Crohn's disease. Toxicol Sci. 2011;120:68–78. https://doi.org/10.1093/toxsci/kfq360.
45. Li Y, Innocentin S, Withers DR, Roberts NA, Gallagher AR, Grigorieva EF, et al. Exogenous stimuli maintain intraepithelial lymphocytes via aryl hydrocarbon receptor activation. Cell. 2011;147:629–40. https://doi.org/10.1016/j.cell.2011.09.025.
46. Lamas B, Natividad JM, Sokol H. Aryl hydrocarbon receptor and intestinal immunity. Mucosal Immunol. 2018;11:1024–38. https://doi.org/10.1038/s41385-018-0019-2.
47. Mizoguchi A, Yano A, Himuro H, Ezaki Y, Sadanaga T, Mizoguchi E. Clinical importance of IL-22 cascade in IBD. J Gastroenterol. 2018;53:465–74. https://doi.org/10.1007/s00535-017-1401-7.
48. Sugimoto S, Naganuma M, Kiyohara H, Arai M, Ono K, Mori K, et al. Clinical efficacy and safety of oral Qing-Dai in patients with ulcerative colitis: a single-center open-label prospective study. Digestion. 2016;93:193–201. https://doi.org/10.1159/000444217.
49. Metidji A, Omenetti S, Crotta S, Li Y, Nye E, Ross E, et al. The environmental sensor AHR protects from inflammatory damage by maintaining intestinal stem cell homeostasis and barrier integrity. Immunity. 2018;49:353–62.e5. https://doi.org/10.1016/j.immuni.2018.07.010.
50. Cannon AS, Nagarkatti PS, Nagarkatti M. Targeting AhR as a novel therapeutic modality against inflammatory diseases. Int J Mol Sci. 2022;23:288. https://doi.org/10.3390/ijms23010288.

## ACKNOWLEDGEMENTS

We thank all family members for their participation in this study and Annemette Friis Mikkelsen and Marianne Lauridsen is thanked for excellent technical assistance and Rigshospitalets Microarray Center for the genotyping. HE was supported by the Danish Medical Research Council (22-02-0208), CBB by the Innovation Fund Denmark (0603-00320B), JTT and JBO received funding from Region Zealand and Roskilde University.

## AUTHOR CONTRIBUTIONS

All authors have contributed to and approved the final manuscript and agreed to the order in which their names are listed. Hans Eiberg: Conceptualization and design of the study, linkage analysis, WGS data analysis and revising the manuscript. Josephine B Olsson: Construction of expression plasmids, cell transfection, promoter analysis and revising the manuscript. Mads Bak: WGS, alignment, and analyses of data. Claus H Bang-Berthelsen: data analysis and associations data. Jesper T Troelsen: Design of promoter analysis, interpretation of data and revising the manuscript. Lars Hansen: Design of the study, data analysis, drafting and writing the manuscript.

## COMPETING INTERESTS

The authors declare no competing interests.

## ETHICS APPROVAL

The study protocols adhered to the tenets of the Declaration of Helsinki and the ARVO statement on human subjects and approved by the Danish National Committee on Health Research Ethics in 2019 (H-19019167).

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41431-023-01298-9.

**Correspondence** and requests for materials should be addressed to Hans Eiberg.

**Reprints and permission information** is available at http://www.nature.com/reprints

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.