Genes and Immunity (2006) 7, 11–18

www.nature.com/gene

ORIGINAL ARTICLE

Evaluation of AGR2 and AGR3 as candidate genes for inflammatory bowel disease

W Zheng[1,2], P Rosenstiel[1], K Huse[3], C Sina[1], R Valentonyte[1,2], N Mah[1], L Zeitlmann[4], J Grosse[4], N Ruf[5], P Nürnberg[5], CM Costello[1], C Onnie[6], C Mathew[6], M Platzer[3], S Schreiber[1] and J Hampe[1,2]

[1]*Institute of Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany;* [2]*Kiel Center of the German National Genotyping Platform, Christian-Albrechts-University, Kiel, Germany;* [3]*Department of Genome Analysis, Institute of Molecular Biotechnology, Jena, Germany;* [4]*Ingenium Pharmaceuticals AG, Martinsried, Germany;* [5]*Gene Mapping Center, MDC Berlin, Germany and* [6]*Division of Genetics and Molecular Medicine, King's College London, London, UK*

Linkage analyses have implicated chromosome 7p21.3 as a susceptibility region for inflammatory bowel disease (IBD). Recently, the mouse phenotype with diarrhea and goblet cell dysfunction caused by anterior gradient protein 2 dysfunction was reported (European patent WO2004056858). The genes encoding for the human homologues AGR2 and AGR3 are localized on chromosome 7p21.3. The gene structures were verified and mutation detection was performed in 47 IBD patients. A total of 30 single nucleotide polymorphisms (SNPs) were tested for association to ulcerative colitis (UC, N = 317) and Crohn's disease (CD, N = 631) in a German cohort and verified in a UK cohort of 384 CD and 311 UC patients. An association signal was identified in the 5′ region of the AGR2 gene (most significant SNP hcv1702494, nominal $P_{TDT} = 0.011$, $P_{case/control} = 0.0007$, *OR = 1.34, combined cohort). The risk haplotype carried an odds ratio of 1.43 in the German population (P = 0.002). AGR2 was downregulated in UC patients as compared to normal controls (P < 0.001) and a trend toward lower expression was seen in carriers of the risk alleles. Luciferase assays of the AGR2 promoter showed regulation by the goblet cell-specific transcription factors FOXA1 and FOXA2. In summary, AGR2 represents an interesting new avenue into the etiopathophysiology of IBD and the maintenance of epithelial integrity.*
Genes and Immunity (2006) **7,** 11–18. doi:10.1038/sj.gene.6364263; published online 13 October 2005

Keywords: *inflammatory bowel disease; Crohn's disease; ulcerative colitis; candidate gene; human Anterior Gradient 2 (AGR2)*

Introduction

Inflammatory bowel disease (IBD) refers to a group of complex chronically relapsing autoimmune disorders of the gastrointestinal tract. It affects 0.1–0.5% individuals in Western countries, with the median onset age in early adulthood.[1,2] On the basis of clinical and histopathological features, IBD is categorized into two main disease entities: Crohn's disease (CD) and ulcerative colitis (UC).[3,4] Active inflammation is marked by chronic diarrhea, rectal bleeding, abdominal pain, fever, weight loss and extraintestinal manifestations, for example, arthritis, uveitis and skin lesions. IBD is a multifactorial disease caused by the interplay of genetic, environmental and immunological factors.[5,6] Epidemiological studies have consistently shown a familial clustering of IBD[7–9] as well as an increased concordance among monozygotic twins.[10,11] Genetic linkage studies in IBD provide a thorough proof for a genetic background,[12] and the first disease gene (*NOD2* or *CARD15*) has been identified for CD.[13–17] Genome-wide linkage analyses in large patient cohorts of different ethnic backgrounds have defined IBD susceptibility loci located on chromosomes 3, 6, 7, 12 and 16.[12,18–20]

Recently, mutations in the murine Anterior Gradient 2 (*AGR2*) gene have been associated with a spontaneous phenotype characterized by diarrhea and goblet cell dysfunction resembling pathological changes seen in human UC (European patent WO2004056858). Two closely related human homologues (human *AGR2* and *AGR3* genes) of the mouse gene are located on chromosome 7p21.3, a location supported by the sequence of the human genome assembly (NCBI build 35), earlier radiation hybrid mapping and fluorescence *in situ* hybridization.[21] This chromosomal region has been implicated as a susceptibility region for IBD in a previous genome-wide linkage analysis with stronger evidence for linkage in the UC group.[18] Expression of *AGR2* has been reported in mucus-secreting cells and endocrine organs.[22] Thus, we selected the human *AGR2* and *AGR3* (*BCMP11*) genes as functional and positional candidate genes for IBD for an in-depth investigation.

Correspondence: Dr J Hampe, Institute of Clinical Molecular Biology, Christian-Albrechts-Universität Kiel, Schittenhelmstr. 12, 24105 Kiel, Germany.
E-mail: J.Hampe@mucosa.de
Received 12 May 2005; revised 10 August 2005; accepted 12 August 2005; published online 13 October 2005

Subjects and methods

Investigated patient sample

In this study, two groups of participants were investigated. The first group included a sample of 383 cases from families and 565 trios with IBD (631 with CD, 317 with UC) and 537 unrelated healthy control individuals of German extraction. The second group consisted of 604 cases from families and 91 trios (384 with CD and 311 with UC) and 360 unrelated healthy control individuals from the United Kingdom. For each affected individual, the diagnosis of either CD or UC was confirmed by standard diagnostic criteria.[3,23] Ascertainment criteria were determined before the initiation of patient collection. German patients and their family members were recruited at the 1st Department of Medicine at the Hospital Schleswig-Holstein, Campus Kiel (Kiel, Germany), the Charité University Hospital (Berlin, Germany) and other collection centers in Germany. The control individuals were collected through the Department of Transfusion Medicine at the University Hospital Schleswig-Holstein and through the POPGEN population project (www.popgen.de). UK families were sampled through the King's College School of Medicine, Guy's Hospital, and the St Mark's Hospital London (UK). The cohorts have been used in a number of previous studies and the clinical and ethical issues of the recruitment process were reviewed as part of these publications.[19,16,24,15] An overview of the investigated sample is given in Table 1.

cDNA cloning

The public databases and the Celera Discovery Systems were reviewed in order to establish gene models for the *AGR2* and *AGR3* genes. For *AGR2*, two additional 5′ exons were annotated. The presence of the additional 5′ exons was investigated through RT-PCR in the Clontech multiple tissue panels I and II (East Meadow Circle, Palo Alto, USA), using standard supplier's protocols. For the gene structure evaluation of *AGR2*, the following primers were used: (1) for the short form: AGRf5: TCA ACT CTG GCC AGG AAC TC; AGRr5: TAC AGC ACC ATA GTC CAG GG and (2) for the long form: AGRf11: CGA CTC ACA CAA GGC AGG T; AGRr11: GCT GTA TCT GCA GGT TCG T – designed on the basis of the NCBI and Celera gene models.

Mutation detection and genotyping

A total of 21 pairs of primers were designed to cover all exons and the promoter region of the *AGR2* and *AGR3* genes. Amplification was performed with an ABI GeneAmp® PCR System 9700 (Applied Biosystems Inc., Foster City, CA, USA) using the following thermoprofile: 96°C – 10 min, (96°C – 1 min; 59–68.7°C – 1 min (−0.5°C per step); 72°C – 1 min) 16 ×, (96°C – 1 min; 51.2–60.7°C – 1 min; 72°C – 1 min) 25 ×, 72°C – 10 min. The BigDye chemistry (ABI) was used according to the manufacturer's recommendations for the sequencing reactions and analyzed on ABI3700 and ABI3730 automated sequencers. Mutation detection was performed in a set of 47 unrelated German individuals affected with IBD (24 patients with CD, 23 patients with UC). Genotyping was performed on an automated platform using the ABI Taqman technology as described.[25,26]

Statistical analysis

Family-based analyses were performed using the transmission disequilibrium test (TDT)[27] in trios, using TRANSMIT[28] and GENEHUNTER.[29] Haplotype frequency estimates among singletons were obtained using an implementation of the EM algorithm (HAPMAX).[30] Significance testing of haplotype frequency differences was also performed with HAPMAX, making use of the fact that twice the log-likelihood ratio between two nested data models approximately follows a χ^2 distribution with k degrees of freedom, where k is the difference in parameter number between the two models. Significance assessment of associations with or between single locus genotypes was performed using χ^2 or Fisher's exact test for 2×3 contingency tables. All other statistical calculations were performed with SPSS. Throughout the manuscript, nominal P-values are given unless specified differently.

Cell culture and reporter gene constructs

HEK 293 cells were purchased from the German Collection of Microorganisms and Cell Cultures (DSMZ, Braunschweig, Germany). The cells were cultured in DMEM + 10% fetal calf serum. At 1 day before transfection, cells were seeded at a density of 5×10^5 cells/2 ml on 6-well plates. Transfections were performed with Fugene 6 (Roche, Switzerland) according to the manufacturer's manual using 0.08 μg of the target plasmid and 0.02 μg of the pRL-TK reference plasmid (Promega, Mannheim, Germany) for the reporter gene assays. The constructs for the goblet cell transcription factors *FOXA1* and *FOXA2* have been described elsewhere (Rausa, FM 2003 MCB). At 24 h after transfection, the cells were harvested for reporter gene assay. Transfection efficiency was determined by parallel detection of pRL-TK activity in a dual luciferase reporter gene assay (Promega, Madison, WI, USA). Every single transfection experi-

Table 1 Overview of the investigated cohort

Population	*Cases*						*Controls*
	CD			*UC*			
	Trios	*Cases from families*	*Total independent case*	*Trios*	*Cases from families*	*Total independent case*	
Germany	377	254	631	188	129	317	537
UK	56	328	384	35	276	311	360

Non-overlapping categories are given. Single cases were randomly selected from IBD families.

ment was performed in duplicate and was repeated at least three times. The *AGR2* promoter of the short form (NM_006408) from −1542 to −1 was amplified from 100 ng of human genomic DNA by PCR under standard conditions with the following primers (restriction sites underlined): pGL_AGR2_N_sense(*Xho*I): CGC TCG AGA TCT TTA CAG AGG TAA TTA AGT TAA AGT A; pGL_AGR2_N_anti(*Hin*dIII): GCA AGC TTG TTG CTA ACT CAG AAA CGA ACC TTC CTT TCC CCA A. It was then cloned into the pGL3-basic plasmid (Promega). All constructs were sequence-verified with an ABI3700 sequencer (ABI, Foster City, CA, USA) before use.

Dual luciferase reporter gene assay

Luciferase activity was determined with a dual luciferase reporter gene kit (DLR) from Promega according to the manufacturer's manual. The cells lysates were analyzed with a MicroLumatPlus LB96V microplate luminometer (EG&G Berthold, Wellesley, MA, USA) after automatic injection of the necessary substrate solutions. All samples were measured at least in duplicate. The results for firefly luciferase activity were normalized to *Renilla* luciferase activity.

Real-time PCR

RNA transcript levels were measured using quantitative real-time PCR in 138 patient and normal control samples, including 25 normal controls, 56 CD and 57 UC patients. Biopsies were obtained from small and large bowel, with 125 of the 138 samples originating from the sigmoid colon. Patients included in this study consented to the additional research biopsies being taken 24 h before endoscopy. The study protocol was approved by the hospital ethical committee before the start of the study. Total RNA was isolated from snap-frozen biopsies using a commercial kit (Qiagen, Hilden, Germany). A 1 μg portion of total RNA was then reverse-transcribed to cDNA according to the manufacturer's instructions (MultiScribe Reverse Transcriptase, Applied Biosystems, Foster City, CA, USA). The cDNA from each sample was diluted 1:5 and arrayed on 384-well plates for real-time

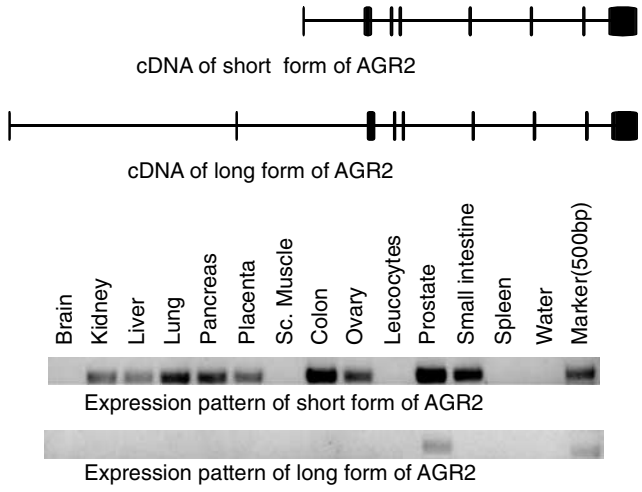

Figure 1 Evaluation of the expression pattern of the two AGR2 splice variants. RT-PCR was performed in the Human Multiple Tissue Panels using the primers given in Subjects and methods.

Table 2 Results of the mutation detection of all exons and the promoters of the *AGR2* and *AGR3* genes

Gene	*Name*	*Position*	*SNP*	*Type*	*Note*
AGR2	07AGR8N1348	1348 in ncbi exon 8	C/T	Intron	Novel
AGR2	07AGR8N1234	1234 in ncbi exon 8	A/G	Intron	Novel
AGR2	07AGR8N104	1044 in ncbi exon 8	A/C	3′UTR	hCV26516309
AGR2	07AGR8N707	707 in ncbi exon 8	A/G	3′UTR	Novel
AGR2	07AGR8N392	392 in ncbi exon 8	A/T	3′UTR	Novel
AGR2	hCV1702536	92 in ncbi exon 7	G/T	Intron	hCV1702536
AGR2	hcv1702535	46 in ncbi exon 7	C/T	Exon	hcv1702535
AGR2	hCV11830196	−60 before ncbi exon 5	A/T	Intron	hCV11830196
AGR2	07AGR1N34	34 in ncbi exon 1	A/G	5′UTR	hCV8302356
AGR2	07AGR1N17	17 in ncbi exon 1	A/G	5′UTR	rs706073
AGR2	07AGRNP53	−53 before ncbi exon 1 (in promoter region)	A/C	Intron	hCV27493993
AGR2	07AGRNP199	−199 before ncbi exon 1 (in promoter region)	C/T	Intron	hcv1170870
AGR2	07AGRNP261	−261 before ncbi exon 1 (in promoter region)	C/T	Intron	rs17136670
AGR2	07AGRC2_122	−122 before celera exon 2	G/T	Intron	Novel
AGR2	hcv9501480	428 in celera exon 1	A/G	Intron	hcv9501480
AGR2	hcv111845	144 in celera exon 1	C/T	5′UTR	hcv111845
AGR2	07AGRCP176	−176 before celera exon 1 (in promoter region)	A/G	Intron	Novel
AGR2	07AGRCP197	−197 before celera exon 1 (in promoter region)	C/T	Intron	Novel
AGR2	07AGRCP207	−207 before celera exon 1 (in promoter region)	C/T	Intron	Novel
AGR2	07AGRCP299	−299 before celera exon 1 (in promoter region)	C/T	Intron	Novel
AGR2	07AGR2C395	−395 before celera exon 1 (in promoter region)	C/T	Intron	rs12674158
AGR2	07AGR2C517	−517 before celera exon 1 (in promoter region)	C/T	Intron	SNP_A-1721592
AGR2	07AGR2CP574	−574 before celera exon 1 (in promoter region)	G/T	Intron	Novel
AGR2	07AGRCP619	−619 before celera exon 1 (in promoter region)	A/G	Intron	Novel
AGR2	07AGRCP626	−626 before celera exon 1 (in promoter region)	A/G	Intron	Novel
AGR3	hCV11170861	58 in ncbi exon 7	A/G	3′UTR	Hcv11170861
AGR3	07AGR3E6_184	−184 before ncbi exon 6	G/C	Intron	Novel
AGR3	07AGR3E2_89	89 in ncbi exon 2	C/T	Intron	Novel
AGR3	rs4472406	−54 before ncbi exon 2	C/T	Intron	rs4472406
AGR3	hCV2571858	290 in ncbi exon 1	A/T	Intron	hCV2571858

PCR quantitation using an Assays-on-Demand Gene Expression Assay for *AGR2* (Hs00180702_m1; context sequence: GTT TGT TGA CCC ATC TCT GAC AGT T) on the ABI Prism 7900HT Sequence Detection System (Applied Biosystems) according to the manufacturer's instructions. Relative transcript levels were determined using the standard curve quantitation method and β-actin as the endogenous control gene.

Results

cDNA cloning

The gene model of *AGR2* was evaluated using RT-PCR in a tissue panel designed on the basis of the NCBI and Celera gene models. The presence of the additional 5′ exon predicted in the Celera database was confirmed. The two different transcripts also show a significantly different expression pattern (Figure 1). The extended form shows a predominant expression in the prostate whereas the shorter form shows a ubiquitous expression pattern. The resultant gene model has been submitted to GenBank.

Mutation detection

The mutation detection experiment identified a total of 30 single nucleotide polymorphisms (SNPs), of which 19 were not previously known. A total of 25 SNPs are located in the *AGR2* gene and five map to the *AGR3* gene. One SNP (hcv111845 – rs4719482) would lead to an amino-acid exchange in the additional N-terminal sequence of the extended splice variant. An overview of all identified SNPs is given in Table 2.

Association analysis

The German study cohort (Table 1) and 537 healthy controls of German descent were genotyped for 30 SNPs of the genes *AGR2* and *AGR3*. All markers were in Hardy–Weinberg equilibrium both in the case and control sample. Case–control and TDT tests of association were performed for IBD and the subphenotypes CD and UC, separately. Results are shown in Table 3. Markers passing a nominal threshold of $P<0.05$ in the German screening sample were further investigated in an independent patient and control sample from the UK. Association statistics in the UK, German and combined cohorts for three diagnostic categories (UC, CD, IBD) are presented in Table 4a–c.

Table 3 Overview of single point association statistics in the German cohort

Gene	SNP	IBD				CD				UC			
		TDT	Case/control			TDT	Case/control			TDT	Case/control		
		P	P	Allele frequencies		P	P	Allele frequencies		P	P	Allele frequencies	
				Cases	Controls			Cases	Controls			Cases	Controls
AGR2	hcv1702558	0.61	**0.024**	0.38	0.33	0.96	**0.046**	0.38	0.33	0.45	0.11	0.38	0.33
AGR2	hcv1702545	0.24	**0.016**	0.37	0.32	0.29	**0.035**	0.37	0.32	0.58	0.08	0.37	0.33
AGR2	07AGR8N392	0.24	0.86	0.02	0.02	0.56	0.25	0.02	0.02	0.16	0.26	0.01	0.02
AGR2	hcv1702537	0.53	0.195	0.22	0.25	0.78	0.26	0.22	0.25	0.48	0.20	0.22	0.25
AGR2	hcv1702535	0.79	0.27	0.49	0.46	0.89	0.25	0.49	0.46	0.49	0.45	0.48	0.46
AGR2	hcv8302351	**0.016**	**0.001**	0.39	0.47	0.19	**0.001**	0.39	0.46	**0.022**	**0.018**	0.40	0.47
AGR2	hcv1702532	0.85	0.66	0.19	0.20	0.42	0.87	0.20	0.20	0.41	0.70	0.19	0.20
AGR2	07AGR1N34	0.2	0.25	0.17	0.19	0.43	0.34	0.17	0.19	0.27	0.24	0.16	0.19
AGR2	07AGR1N17	0.052	0.299	0.17	0.19	0.11	0.63	0.19	0.19	0.27	0.18	0.16	0.19
AGR2	07AGRNP53	0.057	**0.038**	0.21	0.17	0.24	0.11	0.20	0.17	0.11	**0.023**	0.23	0.18
AGR2	07AGRNP199	0.07	0.09	0.05	0.07	**0.03**	0.052	0.05	0.07	0.89	0.38	0.06	0.07
AGR2	07AGRNP261	0.17	0.12	0.21	0.18	0.52	0.27	0.20	0.18	0.15	**0.034**	0.23	0.19
AGR2	hcv1702494	**0.013**	**0.013**	0.43	0.48	0.14	0.06	0.44	0.49	**0.027**	**0.01**	0.41	0.48
AGR2	hcv474914	0.18	0.34	0.31	0.33	0.42	0.67	0.32	0.33	0.23	0.06	0.29	0.33
AGR2	hcv111845	**0.028**	0.11	0.35	0.38	0.23	0.34	0.36	0.38	**0.034**	**0.03**	0.32	0.38
AGR2	07AGRCP176	**0.005**	0.078	0.03	0.05	**0.029**	0.10	0.03	0.05	0.08	0.59	0.04	0.05
AGR2	07AGRCP197	**0.047**	0.42	0.11	0.12	0.24	0.86	0.12	0.12	0.06	0.25	0.09	0.11
AGR2	07AGRCP207	0.21	0.52	0.01	0.001	0.56	0.67	0.001	0.001	0.26	0.14	0.01	0.001
AGR2	07AGRCP299	0.08	0.097	0.22	0.19	0.33	0.21	0.21	0.19	0.1	0.054	0.24	0.19
AGR2	07AGR2C395	0.51	0.82	0.14	0.14	0.46	0.76	0.15	0.15	0.92	0.33	0.13	0.15
AGR2	07AGR2C517	0.56	1	0.14	0.14	0.43	0.54	0.15	0.14	0.92	0.42	0.13	0.14
AGR2	07AGR2CP574	0.76	0.1	0.01	0.001	0.32	0.1	0.01	0.001	0.16	0.52	0.01	0.001
AGR2	07AGRCP619	0.12	0.54	0.14	0.13	0.57	0.18	0.16	0.14	**0.036**	0.66	0.12	0.13
AGR2	07AGRCP626	0.6	0.7	0.14	0.15	0.67	0.80	0.16	0.15	0.77	0.25	0.13	0.15
AGR3	hcv318606	0.84	0.92	0.32	0.32	0.96	0.97	0.32	0.32	0.66	0.74	0.33	0.32
AGR3	hcv11170861	0.68	0.99	0.33	0.33	0.84	0.89	0.32	0.32	0.66	0.84	0.33	0.33
AGR3	hcv2571858	0.94	0.49	0.29	0.28	0.88	0.71	0.29	0.28	0.73	0.41	0.31	0.29
AGR3	hcv2571854	0.29	0.72	0.15	0.15	0.95	0.67	0.16	0.15	0.06	0.5	0.16	0.15
AGR3	hcv2571840	0.42	0.94	0.32	0.32	0.48	0.94	0.33	0.33	0.71	0.96	0.32	0.32
AGR3	hcv2571839	0.93	0.7	0.31	0.30	0.88	0.87	0.31	0.30	0.94	0.69	0.30	0.29

Results that meet nominal *P*-value criterion of 0.05 are highlighted in bold.

Table 4 Replication analysis of significant markers from Table 3 in a UK cohort in the (a) UC subgroup, (b) CD subgroup and (c) joint IBD phenotype

SNP	German				UK				Combined			
	TDT	Case/control			TDT	Case/control			TDT	Case/control		
	P	P	Allele frequencies		P	P	Allele frequencies		P	P	Allele frequencies	
			Cases	Controls			Cases	Controls			Cases	Controls
(a) UC subgroup												
hcv1702558	0.45	0.11	0.38	0.33	0.41	0.39	0.39	0.36	0.29	0.07	0.38	0.34
hcv1702545	0.58	0.08	0.37	0.33	0.11	0.35	0.38	0.34	0.24	**0.04**	0.38	0.33
hcv8302351	**0.022**	**0.018**	0.40	0.47	0.91	0.86	0.40	0.41	**0.04**	**0.04**	0.40	0.44
07AGRNP53	0.11	**0.023**	0.23	0.18	0.39	0.51	0.13	0.21	0.69	**0.013**	0.23	0.19
07AGRNP199	0.89	0.38	0.06	0.07	0.39	0.27	0.05	0.07	0.72	0.21	0.06	0.07
07AGRNP261	0.15	**0.034**	0.23	0.19	0.33	0.497	0.22	0.21	0.84	**0.023**	0.23	0.19
hcv1702494	**0.027**	**0.01**	0.41	0.48	0.17	**0.045**	0.39	0.45	**0.011**	**0.0007**	0.39	0.47
hcv111845	**0.034**	**0.03**	0.32	0.38	0.41	**0.031**	0.33	0.39	**0.029**	**0.005**	0.33	0.38
07AGRCP176	0.08	0.59	0.04	0.05	0.80	0.88	0.05	0.05	0.14	0.63	0.04	0.05
07AGRCP197	0.06	0.25	0.09	0.11	0.22	0.80	0.12	0.12	**0.03**	0.31	0.11	0.12
07AGRCP619	**0.036**	0.66	0.12	0.13	0.15	0.71	0.14	0.13	**0.01**	0.91	0.13	0.13
(b) CD subgroup												
hcv1702558	0.96	0.46	0.38	0.33	0.87	0.21	0.60	0.65	0.38	**0.024**	0.38	0.34
hcv1702545	0.29	**0.035**	0.37	0.32	0.21	0.20	0.39	0.34	0.13	**0.018**	0.38	0.45
hcv8302351	0.19	**0.001**	0.39	0.46	0.10	0.22	0.38	0.42	0.49	**0.0001**	0.38	0.45
07AGRNP53	0.24	0.11	0.20	0.17	0.49	0.17	0.17	0.21	0.22	0.50	0.19	0.18
07AGRNP199	**0.03**	0.052	0.05	0.07	0.24	0.06	0.05	0.07	**0.014**	**0.013**	0.05	0.07
07AGRNP261	0.52	0.27	0.20	0.18	0.66	0.17	0.17	0.21	0.53	0.8	0.19	0.19
hcv1702494	0.14	0.06	0.44	0.49	**0.04**	0.23	0.41	0.45	**0.028**	**0.038**	0.44	0.48
hcv111845	0.23	0.34	0.36	0.38	0.10	0.29	0.35	0.38	0.14	0.18	0.36	0.38
07AGRCP176	**0.029**	0.10	0.03	0.05	0.72	0.7	0.05	0.05	0.11	0.26	0.04	0.05
07AGRCP197	0.24	0.86	0.12	0.12	0.44	0.5	0.11	0.13	0.12	0.65	0.12	0.12
07AGRCP619	0.57	0.18	0.16	0.14	0.22	0.64	0.13	0.12	0.45	0.15	0.15	0.13
(c) Joint IBD phenotype												
hcv1702558	0.61	**0.024**	0.38	0.33	0.06	0.31	0.61	0.64	0.18	**0.03**	0.38	0.34
hcv1702545	0.24	**0.016**	0.37	0.32	0.14	0.25	0.38	0.34	0.054	**0.016**	0.37	0.33
hcv8302351	**0.016**	**0.001**	0.39	0.47	0.17	0.51	0.39	0.41	**0.005**	**0.005**	0.39	0.41
07AGRNP53	0.057	**0.038**	0.21	0.17	0.28	0.7	0.20	0.21	**0.037**	0.13	0.21	0.18
07AGRNP199	0.07	0.09	0.05	0.07	0.77	0.16	0.05	0.07	0.08	0.06	0.05	0.07
07AGRNP261	0.17	0.12	0.21	0.18	0.34	0.73	0.20	0.21	0.12	0.23	0.21	0.19
hcv1702494	**0.013**	**0.013**	0.43	0.48	**0.015**	0.11	0.41	0.45	**0.001**	**0.006**	0.42	0.47
hcv111845	**0.028**	0.11	0.35	0.38	0.26	**0.013**	0.34	0.39	**0.012**	**0.04**	0.35	0.38
07AGRCP176	**0.005**	0.078	0.03	0.05	0.88	0.97	0.05	0.05	**0.031**	0.20	0.04	0.05
07AGRCP197	**0.047**	0.42	0.11	0.12	0.17	0.61	0.12	0.13	**0.013**	0.35	0.11	0.12
07AGRCP619	0.12	0.54	0.14	0.13	0.55	0.77	0.13	0.13	**0.03**	0.5	0.14	0.13

Significant values at the nominal *P*-value criterion of 0.05 are highlighted in bold.

The consistency of the TDT and case–control statistics and the strength of replication between populations were used as a pragmatic guide to judge the trustworthiness of the positional signals. The linkage disequilibrium (LD) structure of the human *AGR2* and *AGR3* gene regions is shown in Figure 2 using the r^2 measure of LD. Taking into account the LD data and the association results (Tables 3 and 4), the association is most consistent in the UC phenotype and localizes to the 5′ region of the *AGR2* gene (Figure 2). The association is most pronounced at hcv1702494 (combined sample: $P_{TDT} = 0.011$, $P_{case/control} = 0.0007$, OR = 1.34) and hcv111845 ($P_{TDT} = 0.029$, $P_{case/control} = 0.005$, OR = 1.2837) for the UC phenotype. A haplotype analysis including the markers 07AGRNP53, 07AGRNP261, hcv1702494 and hcv111845 yielded similar results (German population, UC phenotype, $\chi^2 = 14.8$, df = 3, $P = 0.002$) with an odds ratio for the risk haplotype of 1.43. In the UK cohort UC phenotype, there was no significance ($\chi^2 = 8.18$, df = 4, $P = 0.085$). The association of the CD subphenotype to the marker hcv8302351 is largely limited to the German subpopulation. Further, TDT and case–control test give disparate results for this marker.

Expression analysis

The relative expression of total *AGR2* was quantified by real-time PCR in 25 normal controls, 56 CD and 57 UC patients. Median expression levels (arbitrary normalization units) of 1.07 for normal controls, 0.57 for CD and 0.67 for UC were observed. This expression difference was statistically significant as tested by nonparametric testing ($P = 0.0000001$ for CD versus NC and $P = 0.0000001$ for UC versus NC; Mann–Whitney *U*-test; Figure 3).

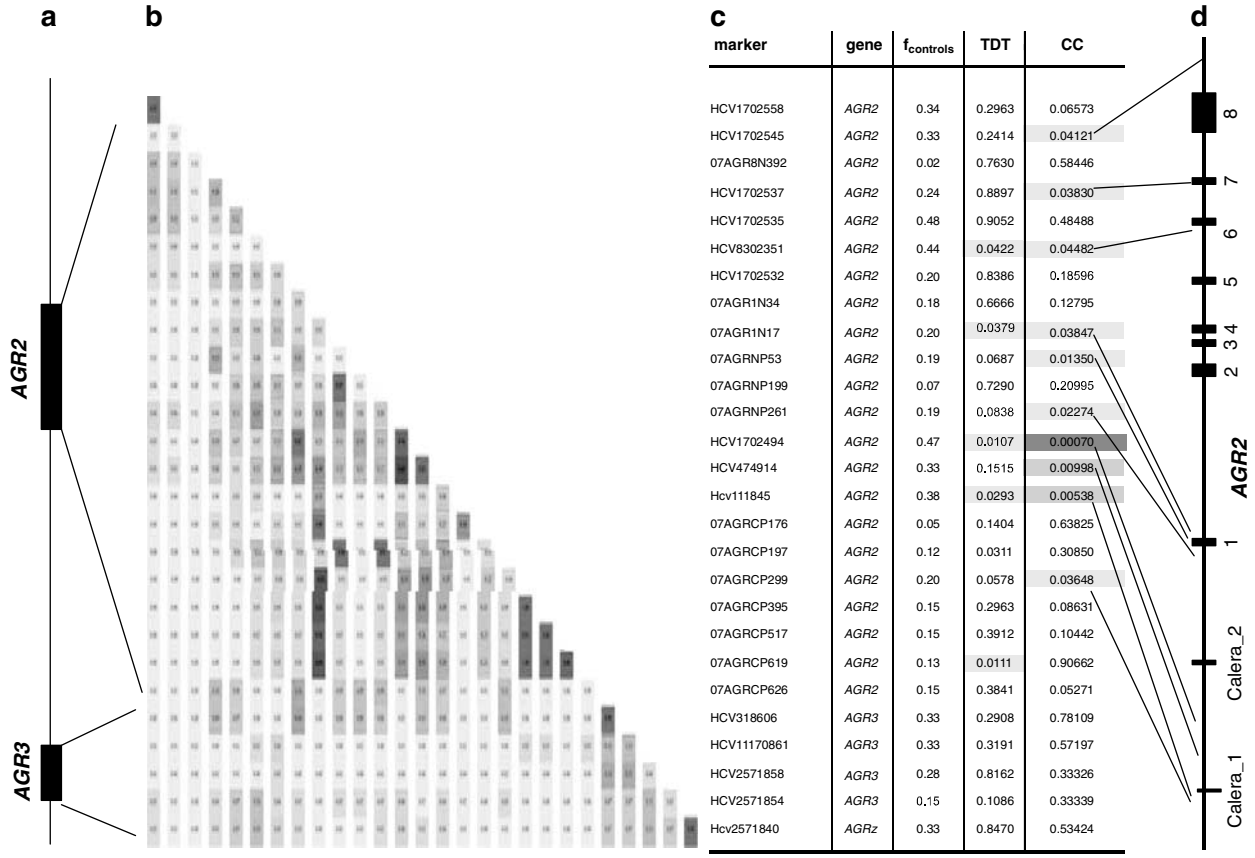

Figure 2 Overview of the LD structure of the human AGR2 and AGR3 region. Panels a and d show the physical map and the exon–intron structure of the AGR2 gene. Panel b shows the r^2 plot for the region. The level of LD is highlighted in gray scale (dark grey $r^2 = 1$, white $r^2 = 0$). The association results for the UC phenotype in the combined German/UK cohort are given in panel c.

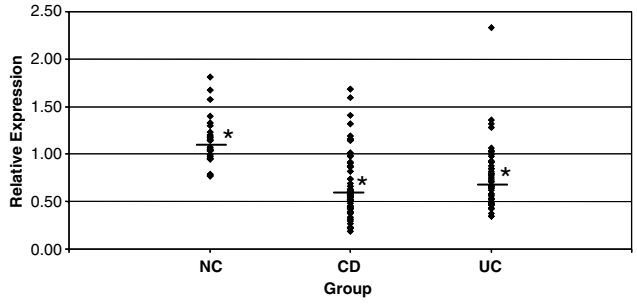

Figure 3 Relative expression of AGR2 in normal and IBD patient samples. The relative expression of AGR2 was quantitated by real-time PCR in 25 normal controls, 56 CD and 57 UC patients. The expression of AGR2 is significantly different between normal controls and either CD or UC samples (P-value <0.001, as shown by Mann–Whitney U-test). *Median.

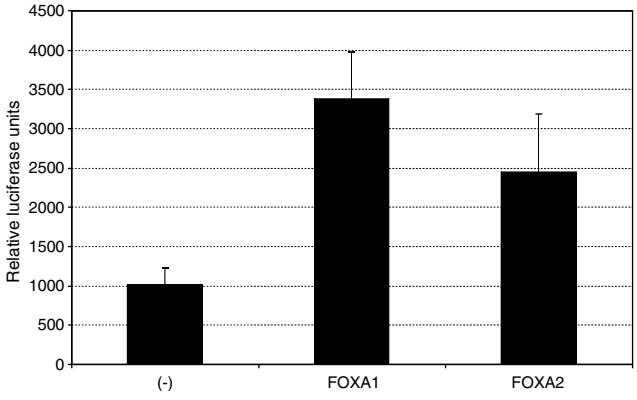

Figure 4 Luciferase activity observed after stimulation with FoxA1 and FoxA2. (−): transfected with pGL3B-AGR2 [−1542] without stimulation; FoxA1: co-transfected with pGL3B-AGR2 [−1542] and FoxA1 gene; FoxA2: transfected with pGL3B-AGR2 [−1542] and FoxA2 gene.

For 91 IBD biopsies, genotype data for the *AGR* markers was available. Individuals who were homozygous for the risk allele at marker hcv111845 showed an overall lower expression relative expression level than the remainder of the samples (relative expression level 0.63 for '22' genotype versus 0.71 for '11' and '12'). However, this difference was not statistically significant owing to the limited sample number ($P > 0.1$). The power to detect this expression difference (0.08) at the $P' \leqslant 0.05$ level in the sample was 48%.

Promoter studies

A luciferase reporter gene construct (pGL3B-AGR2 [−1542]) driven by the *AGR2* promoter was transfected into HEK 293 cells. Co-transfection of the forkhead box transcription factors *FOXA1* and *FOXA2*, which have been implicated in maintaining goblet cell function, led to a significantly increased luciferase activity (Figure 4).

Median luciferase activities (out of seven replicates) of 1023 for the control and 3378 and 2455 for the *FOXA1* and *FOXA2* co-transfection experiments were observed ($P < 0.001$, Mann–Whitney *U*-test).

Discussion

The *AGR2* gene is the human homologue of the *Xenopus laevis* cement gland gene *XAG-1*, mapping to chromosome band 7p21.3.[21] This gene is highly expressed in the trachea, lung stomach, colon, prostate and small intestine.[22] In *X. laevis*, the *XAG* family of genes are expressed in a gradient in the ectoderm during early development of the cement gland, and appear to be important factors during differentiation of this organ.[31] The cement gland arises from the outer layer of embryonic ectoderm and forms a cone of columnar epithelium. *AGR2* has been found to be expressed in tissues that contain mucous-secreting cells and/or function as endocrine organs. Thus, from an evolutionary perspective, the human *AGR* genes may be involved in the epithelial barrier function.

The hypothesis of involvement of the *AGR2* gene in epithelial barrier function is further supported by the regulation of the *AGR2* promoter by transcription factors typical for epithelial goblet cells. Luciferase reporter gene assays show an activation of human *AGR2* promoter by *FOXA1* and *FOXA2*. *FOXA1* contributes to pancreatic beta-cell function[32] and both regulate signaling and transcriptional programs required for morphogenesis and goblet cell differentiation during formation.[33,34] There is a binding site for Hepatic Nuclear Factor 1 (HNF1) in the *AGR2* promoter region at SNP 07AGRNP53. HNF1 and *FOXA1* and *FOXA2* belong to the same family.

The analyses throughout the manuscript present nominal *P*-values. Upon rigorous Bonferroni correction, only the marker hcv8302351 would remain significant in the German screening sample at the 0.05 level. If a conditional analysis was performed, using for instance the TDT results in the German population as a screening test with a $P \leqslant 0.05$ threshold, both hcv1702494 and hcv8302351 would pass this threshold. Overall, we observe association of markers in the 5′ region of *AGR2* primarily with the UC phenotype in two independent cohorts (UK and German extraction). All of those four significant SNPs are located in the 5′ end of the *AGR2* gene (Table 2) and 07AGRNP53 and 07AGRNP261, which are part of the risk haplotype, are located in the promoter region. The expression level of *AGR2* in UC patients was significantly lower than in health controls. Also, a trend toward lower expression of *AGR2* in carriers of the risk alleles was observed. The link between downregulation of the *AGR2* transcript in risk allele carriers and in disease has not yet been fully explored. None of the individual promoter SNPs identified in this study (Table 2) fully defines the disease haplotype, and also the more distant markers hcv1702494 and hcv111845 are needed for the definition of the risk haplotype. This suggests that further, as yet unidentified, private mutations contribute to the downregulation of *AGR2* in disease.

In summary, we demonstrate association of the 5′ region of the *AGR2* gene to the UC phenotype in two independent populations. Functionally, the gene may be involved in the maintenance of epithelial integrity based on the mouse model, phylogenetic background and activation by transcription factors, which are characteristic for epithelial goblet cells. The disease effect is likely to be mediated through downregulation of the *AGR2* transcript in disease, as suggested by association to individual promoter SNPs. The mechanistic risk profile of the risk haplotype is functionally not yet fully explored and possibly includes further private mutations in more distant regulatory elements. The impact of the risk mutations on the overall phenotype is moderate as indicated by allelic odds ratios in the range of 1.3–1.4. Overall, the *AGR2* represents an interesting new avenue into the etiopathophysiology of IBD and warrants further evaluation in additional, independent populations.

Acknowledgements

This study was supported by the German National Genome Research Network (NGFN, DHGP), the Competence Network IBD and the Network 'Fat and metabolism' of the BmBF, the European Commission (FP5 grant), the German Research Council (DFG For423, Ha 3091/1-1) and the POPGEN population project.

References

1 Shivananda S, Lennard-Jones J, Logan R, Fear N, Price A, Carpenter L *et al*. Incidence of inflammatory bowel disease across Europe: is there a difference between north and south? Results of the European Collaborative Study on Inflammatory Bowel Disease (EC-IBD). *Gut* 1996; **39**: 690–697.
2 Probert CS, Jayanthi V, Rampton DS, Mayberry JF. Epidemiology of inflammatory bowel disease in different ethnic and religious groups: limitations and aetiological clues. *Int J Colorectal Dis* 1996; **11**: 25–28.
3 Podolsky DK. Inflammatory bowel disease (1). *N Engl J Med* 1991; **325**: 928–937.
4 Hamilton SR. The differential diagnosis of idiopathic inflammatory disease by colorectal biopsy. *Int J Colorectal Dis* 1987; **2**: 113–117.
5 Fiocchi C. Inflammatory bowel disease: etiology and pathogenesis. *Gastroenterology* 1998; **115**: 182–205.
6 Podolsky DK. Inflammatory bowel disease. *N Engl J Med* 2002; **347**: 417–429.
7 Orholm M, Munkholm P, Langholz E, Nielsen OH, Sorensen IA, Binder V. Familial occurrence of inflammatory bowel disease. *N Engl J Med* 1991; **324**: 84–88.
8 Colombel JF, Grandbastien B, Gower-Rousseau C, Plegat S, Evrard JP, Dupas JL *et al*. Clinical characteristics of Crohn's disease in 72 families. *Gastroenterology* 1996; **111**: 604–607.
9 Satsangi J, Grootscholten C, Holt H, Jewell DP. Clinical patterns of familial inflammatory bowel disease. *Gut* 1996; **38**: 738–741.
10 Tysk C, Lindberg E, Jarnerot G, Floderus-Myrhed B. Ulcerative colitis and Crohn's disease in an unselected population of monozygotic and dizygotic twins. A study of heritability and the influence of smoking. *Gut* 1988; **29**: 990–996.
11 Thompson NP, Driscoll R, Pounder RE, Wakefield AJ. Genetics versus environment in inflammatory bowel disease: results of a British twin study. *BMJ* 1996; **312**: 95–96.
12 Curran ME, Lau KF, Hampe J, Schreiber S, Bridger S, Macpherson AJ *et al*. Genetic analysis of inflammatory bowel disease in a large European cohort supports linkage to chromosomes 12 and 16. *Gastroenterology* 1998; **115**: 1066–1071.

13 Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J *et al.* Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001; **411**: 599–603.
14 Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R *et al.* A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 2001; **411**: 603–606.
15 Hampe J, Cuthbert A, Croucher PJ, Mirza MM, Mascheretti S, Fisher S *et al.* Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. *Lancet* 2001; **357**: 1925–1928.
16 Stoll M, Corneliussen B, Costello CM, Waetzig GH, Mellgard B, Koch WA *et al.* Genetic variation in DLG5 is associated with inflammatory bowel disease. *Nat Genet* 2004; **36**: 476–480.
17 Peltekova VD, Wintle RF, Rubin LA, Amos CI, Huang Q, Gu X *et al.* Functional variants of OCTN cation transporter genes are associated with Crohn disease. *Nat Genet* 2004; **36**: 471–475.
18 Satsangi J, Parkes M, Louis E, Hashimoto L, Kato N, Welsh K *et al.* Two stage genome-wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosomes 3, 7 and 12. *Nat Genet* 1996; **14**: 199–202.
19 Hampe J, Schreiber S, Shaw SH, Lau KF, Bridger S, Macpherson AJ *et al.* A genomewide analysis provides evidence for novel linkages in inflammatory bowel disease in a large European cohort. *Am J Hum Genet* 1999; **64**: 808–816.
20 Hampe J, Shaw SH, Saiz R, Leysens N, Lantermann A, Mascheretti S *et al.* Linkage of inflammatory bowel disease to human chromosome 6p. *Am J Hum Genet* 1999; **65**: 1647–1655.
21 Petek E, Windpassinger C, Egger H, Kroisel PM, Wagner K. Localization of the human anterior gradient-2 gene (AGR2) to chromosome band 7p21.3 by radiation hybrid mapping and fluorescence *in situ* hybridisation. *Cytogenet Cell Genet* 2000; **89**: 141–142.
22 Thompson DA, Weigel RJ. hAG-2, the human homologue of the *Xenopus laevis* cement gland gene XAG-2, is coexpressed with estrogen receptor in breast cancer cell lines. *Biochem Biophys Res Commun* 1998; **251**: 111–116.
23 Lennard-Jones JE. Classification of inflammatory bowel disease. *Scand J Gastroenterol Suppl* 1989; **170**: 2–6; discussion 16–19.
24 Hampe J, Frenzel H, Mirza MM, Croucher PJ, 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**: 321–326. Epub 2001 Dec 18.
25 Hampe J, Wollstein A, Lu T, Frevel HJ, Will M, Manaster C *et al.* An integrated system for high throughput TaqMan based SNP genotyping. *Bioinformatics* 2001; **17**: 654–655.
26 Livak KJ, Marmaro J, Todd JA. Towards fully automated genome-wide polymorphism screening. *Nat Genet* 1995; **9**: 341–342.
27 Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 1993; **52**: 506–516.
28 Clayton D. A generalization of the transmission/disequilibrium test for uncertain-haplotype transmission. *Am J Hum Genet* 1999; **65**: 1170–1177.
29 Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 1996; **58**: 1347–1363.
30 Krawczak M, Konecki DS, Schmidtke J, Duck M, Engel W, Nutzenadel W *et al.* Allelic association of the cystic fibrosis locus and two DNA markers, XV2c and KM19, in 55 German families. *Hum Genet* 1988; **80**: 78–80.
31 Sive HL, Hattori K, Weintraub H. Progressive determination during formation of the anteroposterior axis in *Xenopus laevis*. *Cell* 1989; **58**: 171–180.
32 Lantz KA, Kaestner KH. Winged-helix transcription factors and pancreatic development. *Clin Sci (London)* 2005; **108**: 195–204.
33 Wan H, Dingle S, Xu Y, Besnard V, Kaestner KH, Ang SL *et al.* Compensatory roles of Foxa1 and Foxa2 during lung morphogenesis. *J Biol Chem* 2005; **24**: 24.
34 Wan H, Kaestner KH, Ang SL, Ikegami M, Finkelman FD, Stahlman MT *et al.* Foxa2 regulates alveolarization and goblet cell hyperplasia. *Development* 2004; **131**: 953–964.