Human Molecular Genetics, 2010, Vol. 19, No. 17 **3468–3476**
doi:10.1093/hmg/ddq248
Advance Access published on June 22, 2010

Fucosyltransferase 2 (FUT2) non-secretor status is associated with Crohn's disease

Dermot P.B. McGovern[1,2,*], Michelle R. Jones[3], Kent D. Taylor[2], Kristin Marciante[4], Xiaofei Yan[2], Marla Dubinsky[1], Andy Ippoliti[1], Eric Vasiliauskas[1], Dror Berel[1], Carrie Derkowski[1], Deb Dutridge[2], International IBD Genetics Consortium, Phil Fleshner[1], David Q. Shih[1], Gil Melmed[1], Emebet Mengesha[2], Lily King[2], Sheila Pressman[2], Talin Haritunians[2], Xiuqing Guo[2], Stephan R. Targan[1] and Jerome I. Rotter[2]

[1]Inflammatory Bowel and Immunobiology Research Institute, [2]Medical Genetics Institute and [3]Endocrinology, Diabetes & Metabolism, Cedars-Sinai Medical Center, Los Angeles, CA, USA and [4]Cardiovascular Health Research Unit, Department of Internal Medicine, University of Washington, Seattle, WA, USA

Received January 19, 2010; Revised June 3, 2010; Accepted June 10, 2010

Genetic variation in both innate and adaptive immune systems is associated with Crohn's disease (CD) susceptibility, but much of the heritability to CD remains unknown. We performed a genome-wide association study (GWAS) in 896 CD cases and 3204 healthy controls all of Caucasian origin as defined by multidimensional scaling. We found supportive evidence for 21 out of 40 CD loci identified in a recent CD GWAS meta-analysis, including two loci which had only nominally achieved replication (rs4807569, 19p13; rs991804, *CCL2/CCL7*). In addition, we identified associations with genes involved in tight junctions/epithelial integrity (*ASHL, ARPC1A*), innate immunity (*EXOC2*), dendritic cell biology [*CADM1* (*IGSF4*)], macrophage development (*MMD2*), TGF-β signaling *(MAP3K7IP1)* and *FUT2* (a physiological trait that regulates gastrointestinal mucosal expression of blood group A and B antigens) (rs602662, $P = 3.4 \times 10^{-5}$). Twenty percent of Caucasians are 'non-secretors' who do not express ABO antigens in saliva as a result of the *FUT2 W134X* allele. We demonstrated replication in an independent cohort of 1174 CD cases and 357 controls between the four primary *FUT2* single nucleotide polymorphisms (SNPs) and CD (rs602662, combined *P*-value 4.90×10^{-8}) and also association with *FUT2 W143X* ($P = 2.6 \times 10^{-5}$). Further evidence of the relevance of this locus to CD pathogenesis was demonstrated by the association of the original four SNPs and CD in the recently published CD GWAS meta-analysis (rs602662, $P = 0.001$). These findings strongly implicate this locus in CD susceptibility and highlight the role of the mucus layer in the development of CD.

INTRODUCTION

Crohn's disease (CD), one of the major forms of the inflammatory bowel diseases (IBDs), is a chronic, debilitating disease characterized by recurrent gastrointestinal (GI) inflammation, postulated to occur as a result of an abnormal immune reaction to commensal flora in genetically susceptible individuals. The role of commensal flora in precipitating chronic GI mucosal inflammation is substantiated by data from established rodent models of IBD such as the $Il10^{-/-}$ mouse and the *Hla-B27* transgenic rat that are disease free when housed in germ-free environments, but develop inflammation when raised under pathogen-free conditions (1,2). Furthermore, in both of these models, the bacterial load and the nature of the commensal flora can influence both the site and the degree of GI inflammation (1,3,4). In addition, in human disease, both antibiotic and probiotic therapy can be effective in modifying some of the manifestations of IBD (5,6), and our group and others have had a long-standing interest in serological responses to commensal flora and their association with CD (7).

Through utilizing genome-wide association studies (GWASs), in addition to linkage informed positional candidate gene

---

[*] To whom correspondence should be addressed at: Inflammatory Bowel and Immunobiology Research Institute, Cedars-Sinai Medical Center, 8730 Alden Drive, Los Angeles, CA 90048, USA. Tel: +1 3104234100; Fax: +1 3104230146; Email: mcgovernd@cshs.org

Table 1. Replication ($P < 0.05$) of confirmed and 'nominally associated' CD susceptibility loci, or a proxy marker for these loci, from CD GWAS meta-analysis by Barrett *et al.* (15)

Known CD SNP	Chr.#	Position	Minor allele	OR (95%CI)	*P*-value	Proxy marker for known CD SNP (D')	Gene(s) of interest
rs11465804	1p31	67 475 114	G	0.53 (0.41–0.68)	1.02×10^{-06}	Same SNP	*IL23R*
rs9286879	1q24	171 128 857	G	1.21 (1.07–1.37)	0.002	Same SNP	*Gene desert*
rs11584383	1q32	199 202 489	C	0.88 (0.78–0.99)	0.047	rs3767498 (0.96)	*CACNA1S, KIF21B*
rs917997	2q11	102 437 000	T	1.19 (1.05–1.35)	0.001	rs13015714 (1.0)	*IL18RAP*
rs2241880	2q37	233 848 107	A	0.76 (0.68–0.85)	3.29×10^{-6}	Same SNP	*ATG16L1*
rs3197999	3p21	49 696 536	A	1.25 (1.11–1.40)	1.8×10^{-04}	rs4625 (1.0)	*MST1*
rs4613763	5p13	40 428 485	C	1.31 (1.11–1.54)	0.001	Same SNP	*Gene desert (PTGER4)*[a]
rs2188962	5q31	131 798 704	T	1.36 (1.21–1.52)	1.37×10^{-07}	Same SNP	*IBD5 locus*
rs10045431	5q33	158 747 111	A	0.69 (0.61–0.79)	7.12×10^{-08}	Same SNP	*IL12B*
rs6908425	6p22	20 836 710	T	0.81 (0.70–0.93)	0.004	Same SNP	*CDKAL1*
rs2301436	6q27	167 357 978	T	1.37 (1.22–1.53)	5.92×10^{-08}	Same SNP	*CCR6, FGFR10P, RNASE2*
rs10758669	9p24	4 971 602	C	1.154 (1.03–1.29)	0.016	Same SNP	*JAK2*
rs11190140	10q24	101 281 583	G	0.79 (0.06–0.71)	5.3×10^{-05}	Same SNP	*NKX2-3*
rs7927894	11q13	75 978 964	T	1.20 (1.07–1.34)	0.0012	rs7130588 (1.0)	*C11orf30*
rs2066847	16q12	49 321 279	T	1.62 (1.42–1.86)	2.15×10^{-15}	rs5743289[b]	*NOD2*[b]
rs991804	17q12	29 611 838	A	0.78 (0.68–0.89)	0.002	Same SNP	*CCL2, CCL7*
rs2872507	17q21	35 294 289	A	1.17 (1.04–1.31)	0.006	rs907092 (1.0)	*ORMDL3*
rs744166	17q21	37 767 727	T	0.86 (0.76–0.96)	0.010	Same SNP	*STAT3*
rs2542151	18p11	12 769 947	G	1.22 (1.05–1.43)	0.009	rs1893217 (1.0)	*PTPN2*
rs4807569	19p13	1 074 378	C	1.25 (1.09–1.43)	0.001	rs2024092 (1.0)	*STK11, SBNO2, 729119*
rs1736135	21q21	15 727 091	C	0.84 (0.75–0.94)	0.003	rs1736148 (1.0)	—

Chr. #, chromosome number. Chromosomal position according to db129 (Build 36). *P*-value is from the current GWAS.
[a] *PTGER4* is outside the region but has been implicated by eQTL analysis.
[b] No proxy found for specific *NOD2* SNP cited in Barrett *et al.*—alternative SNP used in view of widespread replication of *NOD2*/CD association.

approaches, there has been considerable success in identifying CD susceptibility loci in populations of Northern European origin (8–14). To date, more than 30 loci are definitively known to be associated with CD, although these loci account only for a minority of the genetic variance to CD in this population (15). A number of the CD susceptibility genes encode important components of the innate immune system genes such as *NOD2/CARD15* (12,13), the Toll-like receptors (16,17) and the autophagy genes *ATG16L1(9)* and *IRGM (18)*, further emphasizing the importance of the microbial–host interaction in the development of CD. Our group and others have identified antibodies to bacterial antigens that define certain sub-groups of CD patients reinforcing the essential role that bacteria play in 'driving' CD (19).

We report herein the findings of a CD GWAS identifying a number of putative associations with CD. Given our group's long-standing interest in the host–microbial interaction, we were particularly interested in the CD association with four *Fucosyltransferase 2 (FUT2)* single nucleotide polymorphisms (SNPs), particularly as genetic variation in *FUT2* has been implicated in susceptibility to infections including *Helicobacter pylori* (20), Norovirus (Norwalk virus) (21–23) and progression of HIV (24). *FUT2* alleles have also been associated with circulating serum vitamin B12 levels (25). FUT2 is a physiological trait that regulates the expression of the H antigen, a precursor of the blood group A and B antigens, on the GI mucosa. Approximately 20% of Caucasians are non-secretors (Se-) who do not express ABO antigens in saliva as they are homozygous for *FUT2* null alleles (26). In addition to the genetic associations mentioned above, non-secretion of ABO blood group antigens into body fluids has been shown to be associated with duodenal ulceration (27), the development of oral candidiasis (28,29), rheumatic fever (30), recurrent urinary tract infection (31), cholera (32) and infection with meningococcus (33), pneumococcus (33) and haemophilus influenzae (34). These data taken together implicate *FUT2* as an obvious gene of interest in IBD pathogenesis. Furthermore, our genome-wide association scan reported herein demonstrated four *FUT2* SNPs with association with CD ($P > 4.0 \times 10^{-5}$), including a non-synonymous (*Ser258Gly*) polymorphism. The data presented herein indicate an association between the non-secretor status associated *FUT2* genotype and CD.

RESULTS

A CD GWAS meta-analysis previously identified or confirmed association with 30 loci and demonstrated nominal association with a further 10 loci (15). We were able to confirm association (uncorrected $P < 0.05$ and association with the previously identified risk allele or appropriate proxy with association in the same direction) with 21 of these loci in our GWAS (Table 1) and these loci served as an internal control for our data set and as indicators of the relative power of the discovery phase of the study (see Materials and Methods). Three of these loci were from the nominally replicated list of SNPs (rs4807569, 19p13; rs991804, 17q12, *CCL2, CCL7*; rs917997, 2q11, *IL18RAP*) from the CD meta-analysis, and the data presented in Table 1 therefore provide further evidence of their relevance in CD susceptibility. The *IL18RAP* association has previously been confirmed by others (35). In addition to these loci, we also demonstrated the association between another locus implicated in both CD and UC namely *CARD9* (35,36) [a total of five associated

SNPs $\leq 1 \times 10^{-4}$, including a non-synonymous N12S SNP ($P = 1.4 \times 10^{-6}$)].

We did not demonstrate association ($P < 0.05$) with CD and the other 19 loci identified in the GWAS meta-analysis, including 5q33 (*IRGM*) (no proxy), 9q32 (*TNFSF15*) (no proxy), 10p11 [proxy SNP (D' 1.0) $P = 0.07$], 10q21 (*ZNF365*) (no proxy), 12q12 (*SLC2A13, LRRK2*) (no proxy), 13q14 (rs3764147) (no proxy), 1p13 (*PTPN22*) ($P = 0.43$), 6q21 (*PRDM1*) (no proxy), 8q24 (no proxy), 1q23 (*ITLN1, CD24*) (no proxy), 6p25(*LYRM4*) (no proxy), 2p16 (*PUS10*) (no proxy), 6p25 (*SLC22A23*) ($P = 0.051$), 6q25 ($P = 0.97$), 2p23 (*GCKR*) ($P = 0.25$), 7p12 ($P = 0.40$), 21q22 (*ICOSLG*) (no proxy), 18q11 (rs8098673) ($P = 0.07$) and 6p21 (*BTNL2, DRA, DQA, DRB*) [proxy SNP (D' 0.93) $P = 0.15$].

In addition, we identified associations between CD and a number of putative loci (see Supplementary Material, Table S1). These include genes involved in tight junctions/epithelial integrity [*ASH1L* on 1q22 (4.3×10^{-5})], Wnt and JNK1 signaling [*RHOU* on 1q42 (4.2×10^{-5})], substance P signaling [*TACR3* on 4q25 (5.4×10^{-5})], innate immunity [*EXOC2* on 6p25 (5.6×10^{-5})] (37), dendritic cell biology [*CADM1* (*IGSF4*) on 11q23 (2.3×10^{-6})], macrophage development [*MMD2* on 7p22 (2.80×10^{-5})], asthma susceptibility [*DENND1B* on 1q31 (4.75×10^{-5})] (38,39), integrin regulation [*ACER2* on 9p22 (6.5×10^{-5})], and TGF-β signaling [*MAP3K7IP1* on 22q13 (6.0×10^{-6})]. We also identified two CD-associated loci specifically involved in the host–microbial interaction, namely *SPG20* on 13q13 (3.92×10^{-6}) (endosomal trafficking) and *FUT2* on 19p13 (3.44×10^{-5}). An expression study previously utilized to implicate *PTGER4* in CD pathogenesis (14,40) indicated a strong *cis* effect in three of the putative loci that we identified, including *FDPS* (effect of rs11264359*A LOD score 6.6, $P = 3.3 \times 10^{-8}$), *SPG20* (effect of rs912927*A LOD score 7.2, $P = 7.6 \times 10^{-9}$) and *SMHT1* (effect of rs8080966*C LOD score 4.2, $P = 1.1 \times 10^{-5}$) (Supplementary Material, Table S1). The overall GWA results are summarized in the form of a Manhattan plot (Supplementary Material, Fig. S3).

Independent replication of *FUT2*

From our candidate associations, we chose *FUT2* as the leading gene for independent replication, given our group's interest in the host–microbial interaction in CD pathogenesis and *FUT2*'s known association with a number of infective processes and GI diseases. Furthermore, *FUT2* is located under a previously identified peak of linkage for CD on chromosome 19 (41) and there were a total of four SNPs (one of which is non-synonymous) with strong association to CD in our GWAS (see Table 2 and Fig. 1). In addition to these four SNPs (rs504963: 3′ UTR; rs676388: 3′ UTR; rs485186: synonymous exon 2 SNP; rs602662: S258G) from the GWAS, we also genotyped rs492602 (synonymous exon 2) and rs601338 (W143X), the common null allele in Caucasians associated with the ABO non-secretory phenotype, in the independent confirmatory cohort. We were able to replicate the initial association with the four SNPs from the discovery cohort, as well as demonstrate association with the additional two SNPs, including the allele for non-secretor status. Further analysis of the *FUT2* association seen in the replication

Table 2. Summary of association between *FUT2* and CD in index GWAS, confirmatory cohort of 1174 cases and 357 controls and the CD GWAS meta-analysis from Barrett *et al.* (15)

SNP	SNP location and effect	Assoc. allele	GWAS *P*-value for association in GWAS	Replication cohorts Case/control allele frequencies	Genotype frequencies in cases	H–W in cases	Genotype frequencies in controls	H–W in controls	Replication *P*-value[a]	p TS[b]	p TSM[b]	Combined *P*-value[c]	Barrett *et al.* (15) *P*-value
rs492602	Exon 2, synon.	G	n/a	0.503, 0.407	27.4/46.4/26.2	0.017	16.8/48.7/34.9	0.99	5.27×10^{-5}	1.0×10^{-4}	9.0×10^{-5}	n/a	n/a
rs601338	W143X	A	n/a	0.505, 0.412	27.7/46.2/26.2	0.0092	16.1/50.2/33.7	0.84	8.56×10^{-5}	3.0×10^{-5}	2.6×10^{-5}	n/a	n/a
rs602662	Ser258Gly	A	3.4×10^{-5}	0.520, 0.433	29.1/46.3/24.7	0.015	18.1/50.4/33.7	0.91	2.74×10^{-4}	7.8×10^{-5}	6.9×10^{-5}	4.90×10^{-08}	0.001
rs485186	Exon 2, synon.	G	3.8×10^{-5}	0.566, 0.479	24.6/46.0/29.4	0.0098	18.6/49.6/31.8	0.99	2.98×10^{-4}	3.6×10^{-5}	3.3×10^{-5}	2.70×10^{-08}	0.001
rs504963	3′ UTR	A	4.0×10^{-5}	0.520, 0.431	29.1/46.0/24.9	0.0087	17.9/50.4/31.7	0.91	2.69×10^{-4}	2.8×10^{-5}	2.6×10^{-5}	2.26×10^{-08}	0.0007
rs676388	3′ UTR	C	3.5×10^{-5}	0.515, 0.444	28.6/45.7/25.7	0.0051	20.1/48.6/31.3	0.96	2.80×10^{-3}	3.8×10^{-5}	4.1×10^{-5}	3.07×10^{-08}	0.0006

H–W, Hardy–Weinberg.

[a] *P*-value calculated using logistic regression.

[b] Association analysis results from CSIG: combining test of logistic regression and Hardy–Weinberg proportion test for genetic association studies (see Materials and Methods).

[c] Combined *P*-value calculated for *P*-value in index GWAS, and replication cohort (p TSM).

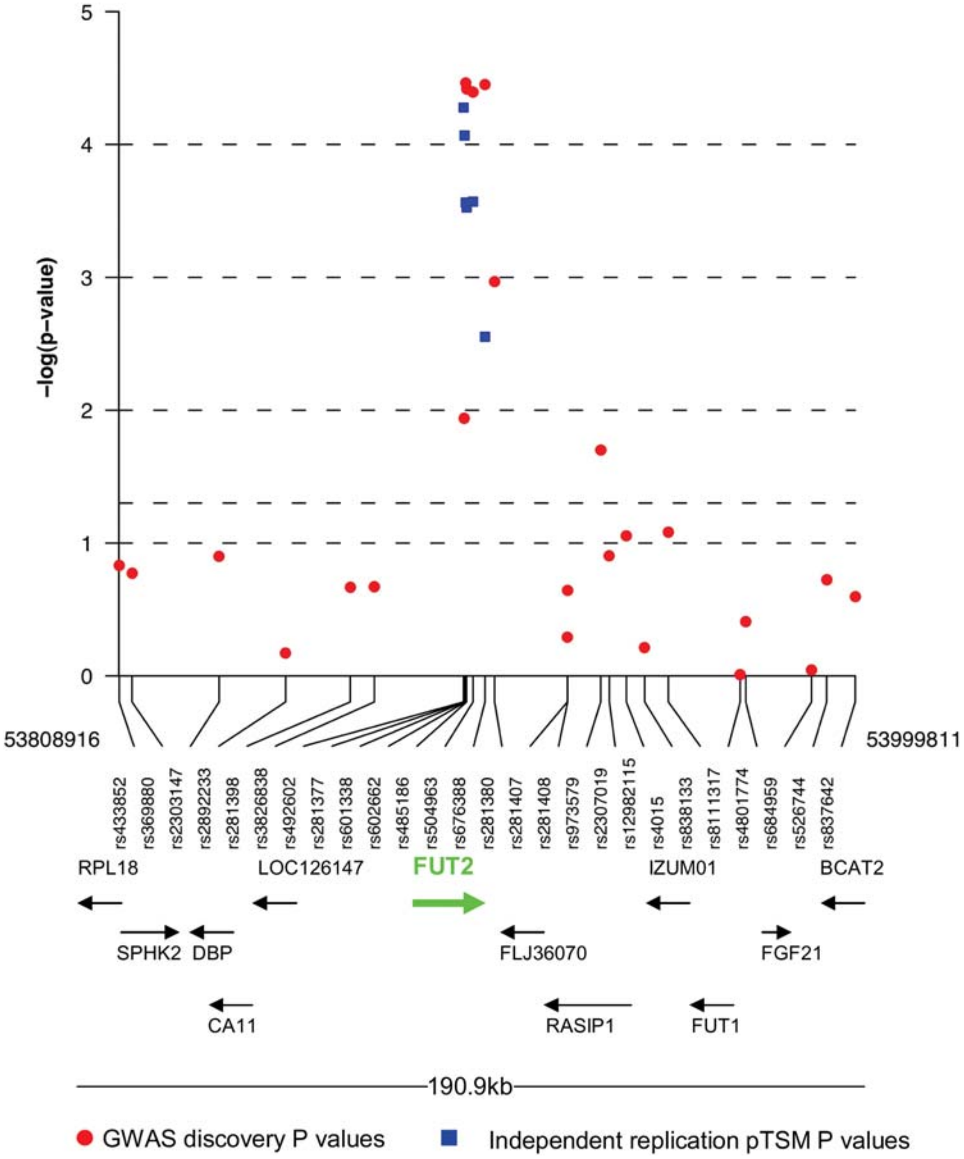

Figure 1. Regional association plot for the *FUT2* locus and CD. A 190 kb region of chromosome 19q13.33 showing position of SNPs included in both the discovery GWAS (red circles) and the replication study (blue squares) within the *FUT2* region (x-axis) and the –log (P-value) for association with CD (y-axis).

cohort reveals deviation from the Hardy–Weinberg equilibrium (HWE) in the cases ($P < 0.05$) but not the controls ($P > 0.84$) (Table 2). The raw genotype frequencies suggest an excess of homozygotes (A allele in rs601338) in cases [325 out of 1174 cases (27.7%)] compared with controls [58 out of 357 (16.2%)] (Table 2). There is no excess of heterozygotes in CD compared with controls and would also be in keeping with the proposed hypothesis that the association seen at this locus is 'driven' by an association between non-secretor status and *FUT2*. In addition, we demonstrated no association between *FUT2* and ulcerative colitis (UC), and there was no evidence of deviation from the HWE in the UC cases (data not shown).

Further evidence for the association between this locus and CD susceptibility is provided by the CD meta-analysis published by Barrett *et al*. (15) in which all four of the *FUT2* SNPs highlighted in the GWAS presented here are also associated with CD (Table 2). We combined the *FUT2* association signals from the study index GWAS and the replication cohort in a meta-analysis comprising in total of our two studies consisting of a total of 2270 CD cases and 4337 controls, achieving the stringent criteria for genome-wide statistical significance ($P > 5 \times 10^{-8}$) for all four SNPs (Table 2 and Fig. 1).

The six SNPs included in the replication study are in strong linkage disequilibrium (Supplementary Material, Fig. S4).

DISCUSSION

In this study, we have confirmed the association with a number of known CD loci, provided further evidence for association to CD with two other loci previously only nominally associated with disease (*19p13* and *17q12*), and identified a number of candidate loci. The region on *19p13* contains *SBNO2* and

GPX4 (glutathione peroxidase 4). Little is known about *SBNO2*, while *GPX4* is known to protect cells against oxidative damage and may have a regulatory role in leukotriene biosynthesis (42). The *17q12* locus is located in a cytokine gene cluster, containing *CCL2, CCL8, CCL11* and *CCL7* genes. These genes encode Cys–Cys cytokine genes that are involved in immunoregulatory and inflammatory processes and are therefore attractive candidate genes for CD susceptibility. This locus has previously also been implicated in susceptibility to asthma (43), mycobacterial infection (44) as well as with HIV progression (45).

We identified a number of putative loci associated with CD in our population. However, our most intriguing result is the evidence of association with *FUT2*. We present independent confirmation for association between *FUT2* and CD in a distinct replication cohort, and also see association with these four SNPs in the meta-analysis published by Barrett *et al.* (15). The combined analysis at this locus of both our index study and the replication cohort demonstrate association attaining stringent criteria of genome-wide statistical significance ($P < 5.0 \times 10^{-8}$). We were particularly interested in this gene given our interest in the host–bacterial interface and the previously documented associations between this gene and infective processes. Furthermore, the association identified herein potentially extends our knowledge regarding the scope of the host–microbial interaction in CD as previously identified genetic associations with CD have highlighted the role of innate (12,13,16,17) and adaptive immune systems (46,47). The data presented here extend this interaction to the mucus layer of the GI tract. *FUT2* encodes the secretor type α (1,2) fucosyltransferase (also known as the Se enzyme) that is responsible for regulating the secretion of the ABO antigens in both the digestive mucosa and secretory glands. Approximately 20% of individuals are non-secretors who fail to express ABO antigens in both the GI tract and saliva as a result of being homozygous for non-secretor alleles (26). The prevalence of the non-secretor status (Se-) is similar between populations (48), although the point mutations that lead to Se- differ. The dominant non-secretor polymorphism in Caucasians is the *Trp143Ter* (*W143X*) (*26*) and our detailed analysis lead us to conclude that this polymorphism is the most likely causative SNP at this locus.

Pathogens utilize host cell surface molecules, including oligosaccharides (synthesized by glycosyltransferases), for invasion. It is likely that the high prevalence of non-secretor phenotypes in the population occurs as the absence of particular carbohydrate molecules in the mucosa may have conferred some historical protection to infection as demonstrated with non-secretor status and protection from *Helicobacter pylori* infection and GI ulceration (20,27). Lactobacilli, a known commensal bacteria, bind to the precursor glycolipid GA1, implying a role of the GI mucosal glycolipid profile in the adherence of commensal and ‘beneficial’ bacteria, in addition to pathogenic organisms (49). Furthermore, *Lactobacilli* can also displace pathogens such as *Clostridium* from mucus (50) and also inhibit the *Shigella*–host interaction (51). Commensal bacteria probably induce glycolipid expression, as the fucosylglycolipid FGA1 is found in the small bowel of conventionally bred mice but not in germ-free mice (52). Furthermore, FGA1 expression is induced by administration of microbes, and FUT2 transcripts in the ileum were induced in germ-free mice 48 h after administration of feces from conventionally bred mice (53). *Fut2*-null mice do not express the fucosylglycolipid FGA1 in the cecum and colon, whereas normal mice do (52). In the mammalian gut, blocking the CRK and JNK pathways inhibits the ability of bacterial colonization to induce fucosyltransferase activity and FUT2 mRNA expression, both of which are hallmarks of the adult mammalian colon (54). Commensal bacteria and probiotics may exert their protective effects through preventing adherence or even displacing pathogenic bacteria, thus emphasizing the potential role of *FUT2* and non-secretor status on GI bacterial profile (55). Se- individuals may thus have a disrupted immunogenic/ homeostatic equilibrium that makes them more susceptible to the development of chronic mucosal inflammation. Furthermore, changes in the microflora of IBD patients have been well documented (56). In addition, Fut2 null mice display an increased susceptibility to experimental yeast vaginitis and cervical mucins containing Fut2 are partly protected from induced vaginal candidiasis (57). Mucin 2 (muc-2), the predominant secreted mucin in the colon, plays a key barrier role in intercepting and excluding bacteria from the mucosal cell surface, thereby reducing host susceptibility to colitis and inflammation-associated neoplasia (58–60). Recent genetic studies have clarified the importance *o*-glycan structures of muc-2 protein in these biologic roles (61,62). Both the core 1- and core 3-derived *o*-glycans of mucin core proteins are terminally fucosylated, which serves as a binding structure for bacterial interception (63). Accordingly, the present findings with *FUT2* may represent human genetic evidence linking IBD susceptibility to the functional state of intestinal mucin.

Although *FUT2* is a strong candidate gene for CD susceptibility given its tissue expression, its influence on the GI bacterial profile and the mode of inheritance we have observed at this locus, the associations identified herein may reflect association with other genetic variants at this locus in linkage disequilibrium with the described *FUT2* SNPs (see Fig. 1). We therefore explored the LD pattern at this locus using the latest version of HapMap (64) and identified that LD (defined as $D' > 0.80$) extends into neighboring genes, including interesting candidate genes that are also potentially involved in the host–bacterial interaction, such as *FUT1* [alpha-1-2-fucosyltransferase 1—*FUT*, genetic variation in pigs is associated with alterations in *Escherichia coli* adherence (65)] and *RASIP1* (RAS interacting protein 1—an RAS effector localized to the Golgi membranes), as well as *DBP* (D-site of albumin promoter-binding protein) and *FGF21* [fibroblast growth factor 21—involved in insulin sensitivity, adipocyte function and growth hormone signaling (66,67)]. While we believe *FUT2* is the most attractive candidate gene at this locus, and we have demonstrated the association with a variant with a known consequence on gene expression, further work will be needed to fully map this locus. We have also identified some candidate loci for further investigation, including genes involved in tight junctions, Substance P signaling, macrophage development, dendritic cell function and NK T-cell function. Further work on these and other loci listed in Supplementary Material, Table S1 will be necessary.

In summary, the data presented here provide strong evidence that non-secretor status increases CD susceptibility. The non-secretor variants from other ethnic groups have been well documented, and studies of these variants within the relevant IBD populations will help elucidate the exact role of *FUT2* in CD susceptibility. Studies on the effect of *FUT2* on clinical and serological phenotype, and in particular its role on the microbiome of non-secretor individuals, may help investigators understand further the role of commensal bacteria in CD susceptibility, and also further determine those CD patients who might most benefit from probiotic- or antibiotic-based therapies for prevention and treatment of CD.

MATERIALS AND METHODS

Study subjects

Discovery cohort. The discovery cohort used in the GWAS comprised 1096 CD subjects and 3970 healthy population controls. Cases were recruited from the Cedars-Sinai IBD and Pediatric IBD Centers and were diagnosed with CD according to standard clinical, radiological, endoscopic and histological criteria. This population consists of a pediatric cohort (39% of the sample, with a mean age of onset of 12.8 years) and adult CD subjects (61% of the sample, with a mean age of onset 37.7 years). Controls were obtained from the Cardiovascular Health Study (CHS), a population-based longitudinal study of risk factors for cardiovascular disease and stroke in adults 65 years of age or older, recruited at four field centers (68). A total of 5201 predominantly Caucasian individuals were recruited in 1989–1990 from random samples of Medicare eligibility lists, followed by an additional 687 African-Americans recruited in 1992–1993 (total $n = 5888$).

Replication cohort. The replication cohort used in Taqman genotyping of the *FUT2* locus consisted of 1174 CD cases and 357 healthy controls. All subjects in the replication cohort were of Northern European origin and independent of the cohort in the GWAS. Cases were recruited at the Cedars-Sinai IBD and Pediatric IBD Centers and diagnosed with the same criteria as those included in the discovery cohort. As with the discovery cohort, this replication sample consisted of both pediatric (9.4% of the total sample, mean age of onset of 11.2 years) and adult cases (90.6% of the total sample, mean age of onset of 31.1 years). Controls were recruited through the Cedars-Sinai IBD center as unrelated acquaintances and spouses of cases (who were not included in the current analysis set) with no personal or family history of IBD or autoimmune disease. All cases and controls provided informed consent prior to study participation and following approval of participating centers' institutional review boards. All cases and controls were independent of the cases and controls included in the published CD meta-analysis.

Genotyping

Discovery GWAS. All genotyping was performed at the Medical Genetics Institute at Cedars-Sinai Medical Center using Illumina Infinium whole-genome genotyping technology, following the manufacturer's protocol (Illumina, San Diego, CA, USA) (69,70). All cases were genotyped with the Illumina Human 610Quad platform. Controls were genotyped with the Illumina 370Duo platform. Samples with genotyping rates >98% were retained in the analysis. In addition, case and control cohorts were both investigated using Identity-By-Descent using the -- genome command in PLINK. Identity-By-Descent is estimated within PLINK utilizing Identity-By-State distance clustering which estimates relatedness within the sample as Pi $\text{hat} = P(\text{IBD} = 2) + 0.5 \times P(\text{IBD} = 1)$ and identifies cryptic relatedness. Pairwise subjects comparisons with Pi hat scores >0.5 had one subject from the pair being analyzed for relatedness removed from the downstream analysis. Eleven CD subjects and 259 controls were removed based on cryptic relatedness. SNPs were excluded based on the following criteria: test of HWE ($P < 0.001$); SNP failure rate >3%; minor allele frequency <5%; and SNPs not found in dbSNP Build 129. SNPs were also examined in order to exclude case/control disparity in missingness, and SNPs with a missingness P-value <0.001 were excluded from the results [PLINK (71)]. A total of 304 825 SNPs passed our QC criteria were available in all data sets and included in the logistic regression association analysis.

Replication of the FUT2 locus. The six SNPs tested in the replication cohort were genotyped using the TaqMan$^{\text{TM}}$ Minor Groove Binder chemistry utilizing Assays on Demand according to the manufacturer's instructions (Applied Biosystems, Foster City, CA, USA). Genotyping concordance among duplicate samples was 100% and the genotyping rate was 96.0% across all SNPs (94.8–97.6% for individual SNPs). All SNPs were in HWE in the controls.

Statistical analysis

GWAS analysis. Population structure was detected using multidimensional scaling [PLINK (71)]. In total, 10 principal components (PCs) were calculated and plotted for graphical representation of population substructure within the cohort. Self-reported ethnicity data were used to confirm the identification of ethnicity based on cluster plots. Subjects lying above 0.025 on the y-axis (PC 1) of the population structure plot (Supplementary Material, Fig. S1) were identified as subjects of African-American ancestry. To reduce false-positive discovery due to population substructure and the predominantly Caucasian make-up of the cases, the African-American subjects were excluded from downstream analysis. In total, 896 CD and 3204 control subjects were carried forward for association testing with the CD phenotype using a logistic regression in PLINK using all 10 PCs as covariates in the model (71). We calculated that the cohorts from this study give us power ranging from 0.36 [mean allele frequency (MAF) = 0.1] to 0.73 (MAF = 0.4) assuming an alpha of 0.05 and an effect size of 1.15 (approximate effect size seen in the 40 meta-analysis loci) to reproduce the associations demonstrated by Barrett *et al*. (15). Genomic control was calculated using the -- adjust function within PLINK simultaneously with a logistic regression with the appropriate quality control filters and revealed a genomic inflation (λ_{GC}) of 1.09. The Q–Q plot is shown in Supplementary Material, Figure S2.

Confirmation study. The association of the *FUT2* SNPs with CD was tested in an independent confirmation cohort. While a logistic regression model is typically used for testing association between a single SNP and disease status, testing deviation from Hardy–Weinberg proportion among cases has been proposed as another approach for genetic association studies (72) (73). Combining information from HWE tests in association studies has been proposed and demonstrated to be effective in improving study power while type I error can be effectively controlled (74–76). We utilized the mean-based tail-strength (TS) measure for association as proposed by Taylor and Tibshirani (75) and the extended median-based measure (TSM) for association as proposed by Wang and Shete (76) to combine the evidence for association from both resources to improve our study power. The program CSIG which can be used to implement both TS and TSM was used to perform the confirmation association analysis. An additive genetic model was assumed in the association analysis.

Meta-analysis. To improve study power and combine the evidence for association from our GWAS, we performed a meta-analysis for the four overlapping SNPs genotyped in both the initial GWAS and replication genotyping of the *FUT2* locus. Given that a consistent directionality between our GWAS sample and the confirmation sample was observed for each of the four SNPs, we utilized Fisher's method for combining *P*-values in our meta-analysis, consisting of a total of 2270 CD cases and 4337 controls.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

Conflict of Interest statement. None declared.

FUNDING

This study was supported in part by NCRR grant M01-RR00425 to the Cedars-Sinai General Research Center Genotyping core; NIH/NIDDK grant P01-DK046763; the Diabetes Endocrinology Research Center grant, DK 063491; Cedars-Sinai Medical Center Inflammatory Bowel Disease Research Funds; The Feintech Family Chair in IBD (S.R.T.); The Abe and Claire Levine Chair in Pediatric IBD (M.D.) and The Cedars-Sinai Board of Governors' Chair in Medical Genetics (J.I.R.). Additional funding through grants DK76984 (M.D.), DK062413 (D.P.B.M.) and DK084554 (M.D. and D.P.B.M.). CHS research reported in this article was supported by contract numbers N01-HC-85079 through N01-HC-85086, N01-HC-35129, N01 HC-15103, N01 HC-55222, N01-HC-75150, N01-HC-45133; grant numbers U01 HL080295 and R01 HL087652 from the NHLBI, with additional contribution from the National Institute of Neurological Disorders and Stroke. A full list of principal CHS investigators and institutions can be found at http://www.chs-nhlbi.org/pi.htm.

REFERENCES

1. Kim, S.C., Tonkonogy, S.L., Albright, C.A., Tsang, J., Balish, E.J., Braun, J., Huycke, M.M. and Sartor, R.B. (2005) Variable phenotypes of enterocolitis in interleukin 10-deficient mice monoassociated with two different commensal bacteria. *Gastroenterology*, **128**, 891–906.
2. Rath, H.C., Herfarth, H.H., Ikeda, J.S., Grenther, W.B., Hamm, T.E. Jr, Balish, E., Taurog, J.D., Hammer, R.E., Wilson, K.H. and Sartor, R.B. (1996) Normal luminal bacteria, especially Bacteroides species, mediate chronic colitis, gastritis, and arthritis in HLA-B27/human beta2 microglobulin transgenic rats. *J. Clin. Invest.*, **98**, 945–953.
3. Rath, H.C., Wilson, K.H. and Sartor, R.B. (1999) Differential induction of colitis and gastritis in HLA-B27 transgenic rats selectively colonized with Bacteroides vulgatus or *Escherichia coli*. *Infect. Immun.*, **67**, 2969–2974.
4. Rath, H.C., Ikeda, J.S., Linde, H.J., Scholmerich, J., Wilson, K.H. and Sartor, R.B. (1999) Varying cecal bacterial loads influences colitis and gastritis in HLA-B27 transgenic rats. *Gastroenterology*, **116**, 310–319.
5. Gionchetti, P., Rizzello, F., Venturi, A., Brigidi, P., Matteuzzi, D., Bazzocchi, G., Poggioli, G., Miglioli, M. and Campieri, M. (2000) Oral bacteriotherapy as maintenance treatment in patients with chronic pouchitis: a double-blind, placebo-controlled trial. *Gastroenterology*, **119**, 305–309.
6. Rutgeerts, P., Hiele, M., Geboes, K., Peeters, M., Penninckx, F., Aerts, R. and Kerremans, R. (1995) Controlled trial of metronidazole treatment for prevention of Crohn's recurrence after ileal resection. *Gastroenterology*, **108**, 1617–1621.
7. Landers, C.J., Cohavy, O., Misra, R., Yang, H., Lin, Y.C., Braun, J. and Targan, S.R. (2002) Selected loss of tolerance evidenced by Crohn's disease-associated immune responses to auto- and microbial antigens. *Gastroenterology*, **123**, 689–699.
8. Duerr, R.H., Taylor, K.D., Brant, S.R., Rioux, J.D., Silverberg, M.S., Daly, M.J., Steinhart, A.H., Abraham, C., Regueiro, M., Griffiths, A. *et al.* (2006) A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science*, **314**, 1461–1463.
9. Hampe, J., Franke, A., Rosenstiel, P., Till, A., Teuber, M., Huse, K., Albrecht, M., Mayr, G., De La Vega, F.M., Briggs, J. *et al.* (2007) A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat. Genet.*, **39**, 207–211.
10. Rioux, J.D., Xavier, R.J., Taylor, K.D., Silverberg, M.S., Goyette, P., Huett, A., Green, T., Kuballa, P., Barmada, M.M., Datta, L.W. *et al.* (2007) Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat. Genet.*, **39**, 596–604.
11. Yamazaki, K., McGovern, D., Ragoussis, J., Paolucci, M., Butler, H., Jewell, D., Cardon, L., Takazoe, M., Tanaka, T., Ichimori, T. *et al.* (2005) Single nucleotide polymorphisms in TNFSF15 confer susceptibility to Crohn's disease. *Hum. Mol. Genet.*, **14**, 3499–3506.
12. Hugot, J.P., Chamaillard, M., Zouali, H., Lesage, S., Cezard, J.P., Belaiche, J., Almer, S., Tysk, C., O'Morain, C.A., Gassull, M. *et al.* (2001) Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature*, **411**, 599–603.
13. Ogura, Y., Bonen, D.K., Inohara, N., Nicolae, D.L., Chen, F.F., Ramos, R., Britton, H., Moran, T., Karaliuskas, R., Duerr, R.H. *et al.* (2001) A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature*, **411**, 603–606.
14. Libioulle, C., Louis, E., Hansoul, S., Sandor, C., Farnir, F., Franchimont, D., Vermeire, S., Dewit, O., de Vos, M., Dixon, A. *et al.* (2007) Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. *PLoS Genet.*, **3**, e58.
15. Barrett, J.C., Hansoul, S., Nicolae, D.L., Cho, J.H., Duerr, R.H., Rioux, J.D., Brant, S.R., Silverberg, M.S., Taylor, K.D., Barmada, M.M. *et al.* (2008) Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat. Genet.*, **40**, 955–962.
16. De Jager, P.L., Franchimont, D., Waliszewska, A., Bitton, A., Cohen, A., Langelier, D., Belaiche, J., Vermeire, S., Farwell, L., Goris, A. *et al.* (2007) The role of the Toll receptor pathway in susceptibility to inflammatory bowel diseases. *Genes Immun.*, **8**, 387–397.
17. Saruta, M., Targan, S.R., Mei, L., Ippoliti, A.F., Taylor, K.D. and Rotter, J.I. (2009) High-frequency haplotypes in the X chromosome locus TLR8 are associated with both CD and UC in females. *Inflamm. Bowel. Dis.*, **15**, 321–327.
18. Parkes, M., Barrett, J.C., Prescott, N.J., Tremelling, M., Anderson, C.A., Fisher, S.A., Roberts, R.G., Nimmo, E.R., Cummings, F.R., Soars, D. *et al.* (2007) Sequence variants in the autophagy gene IRGM and multiple

other replicating loci contribute to Crohn's disease susceptibility. *Nat. Genet.*, **39**, 830–832.

19. Mow, W.S., Vasiliauskas, E.A., Lin, Y.C., Fleshner, P.R., Papadakis, K.A., Taylor, K.D., Landers, C.J., Abreu-Martin, M.T., Rotter, J.I., Yang, H. *et al.* (2004) Association of antibody responses to microbial antigens and complications of small bowel Crohn's disease. *Gastroenterology*, **126**, 414–424.
20. Ikehara, Y., Nishihara, S., Yasutomi, H., Kitamura, T., Matsuo, K., Shimizu, N., Inada, K., Kodera, Y., Yamamura, Y., Narimatsu, H. *et al.* (2001) Polymorphisms of two fucosyltransferase genes (Lewis and Secretor genes) involving type I Lewis antigens are associated with the presence of anti-*Helicobacter pylori* IgG antibody. *Cancer Epidemiol. Biomarkers Prev.*, **10**, 971–977.
21. Marionneau, S., Airaud, F., Bovin, N.V., Le Pendu, J. and Ruvoen-Clouet, N. (2005) Influence of the combined ABO, FUT2 and FUT3 polymorphism on susceptibility to Norwalk virus attachment. *J. Infect. Dis.*, **192**, 1071–1077.
22. Thorven, M., Grahn, A., Hedlund, K.O., Johansson, H., Wahlfrid, C., Larson, G. and Svensson, L. (2005) A homozygous nonsense mutation (428G−>A) in the human secretor (FUT2) gene provides resistance to symptomatic norovirus (GGII) infections. *J. Virol.*, **79**, 15351–15355.
23. Carlsson, B., Kindberg, E., Buesa, J., Rydell, G.E., Lidon, M.F., Montava, R., Abu Mallouh, R., Grahn, A., Rodriguez-Diaz, J., Bellido, J. *et al.* (2009) The G428A nonsense mutation in FUT2 provides strong but not absolute protection against symptomatic GII.4 Norovirus infection. *PLoS One*, **4**, e5593.
24. Kindberg, E., Hejdeman, B., Bratt, G., Wahren, B., Lindblom, B., Hinkula, J. and Svensson, L. (2006) A nonsense mutation (428G−>A) in the fucosyltransferase FUT2 gene affects the progression of HIV-1 infection. *AIDS*, **20**, 685–689.
25. Hazra, A., Kraft, P., Selhub, J., Giovannucci, E.L., Thomas, G., Hoover, R.N., Chanock, S.J. and Hunter, D.J. (2008) Common variants of FUT2 are associated with plasma vitamin B12 levels. *Nat. Genet.*, **40**, 1160–1162.
26. Kelly, R.J., Rouquier, S., Giorgi, D., Lennon, G.G. and Lowe, J.B. (1995) Sequence and expression of a candidate for the human Secretor blood group alpha(1,2)fucosyltransferase gene (FUT2). Homozygosity for an enzyme-inactivating nonsense mutation commonly correlates with the non-secretor phenotype. *J. Biol. Chem.*, **270**, 4640–4649.
27. Evans, D.A., Horwich, L., McConnell, R.B. and Bullen, M.F. (1968) Influence of the ABO blood groups and secretor status on bleeding and on perforation of duodenal ulcer. *Gut*, **9**, 319–322.
28. Thom, S.M., Blackwell, C.C., MacCallum, C.J., Weir, D.M., Brettle, R.P., Kinane, D.F. and Wray, D. (1989) Non-secretion of blood group antigens and susceptibility to infection by Candida species. *FEMS Microbiol. Immunol.*, **1**, 401–405.
29. Aly, F.Z., Blackwell, C.C., MacKenzie, D.A., Weir, D.M., Elton, R.A., Cumming, C.G., Sofaer, J.A. and Clarke, B.F. (1991) Chronic atrophic oral candidiasis among patients with diabetes mellitus—role of secretor status. *Epidemiol. Infect.*, **106**, 355–363.
30. Haverkorn, M.J. and Goslings, W.R. (1969) Streptococci, ABO blood groups, and secretor status. *Am. J. Hum. Genet.*, **21**, 360–375.
31. Kinane, D.F., Blackwell, C.C., Brettle, R.P., Weir, D.M., Winstanley, F.P. and Elton, R.A. (1982) ABO blood group, secretor state, and susceptibility to recurrent urinary tract infection in women. *Br. Med. J. (Clin Res Ed)*, **285**, 7–9.
32. Chaudhuri, A. and DasAdhikary, C.R. (1978) Possible role of blood-group secretory substances in the aetiology of cholera. *Trans. R. Soc. Trop. Med. Hyg.*, **72**, 664–665.
33. Blackwell, C.C., Jonsdottir, K., Hanson, M., Todd, W.T., Chaudhuri, A.K., Mathew, B., Brettle, R.P. and Weir, D.M. (1986) Non-secretion of ABO antigens predisposing to infection by *Neisseria meningitidis* and *Streptococcus pneumoniae*. *Lancet*, **2**, 284–285.
34. Blackwell, C.C., Jonsdottir, K., Hanson, M.F. and Weir, D.M. (1986) Non-secretion of ABO blood group antigens predisposing to infection by *Haemophilus influenzae*. *Lancet*, **2**, 687.
35. Zhernakova, A., Festen, E.M., Franke, L., Trynka, G., van Diemen, C.C., Monsuur, A.J., Bevova, M., Nijmeijer, R.M., van 't Slot, R., Heijmans, R. *et al.* (2008) Genetic analysis of innate immunity in Crohn's disease and ulcerative colitis identifies two susceptibility loci harboring CARD9 and IL18RAP. *Am. J. Hum. Genet.*, **82**, 1202–1210.
36. McGovern, D.P., Gardet, A., Torkvist, L., Goyette, P., Essers, J., Taylor, K.D., Neale, B.M., Ong, R.T., Lagace, C., Li, C. *et al.* Genome-wide association identifies multiple ulcerative colitis susceptibility loci. *Nat. Genet.*, **42**, 332–337.
37. Ishikawa, H., Ma, Z. and Barber, G.N. (2009) STING regulates intracellular DNA-mediated, type I interferon-dependent innate immunity. *Nature*, **461**, 788–792.
38. Laitinen, T., Polvi, A., Rydman, P., Vendelin, J., Pulkkinen, V., Salmikangas, P., Makela, S., Rehn, M., Pirskanen, A., Rautanen, A. *et al.* (2004) Characterization of a common susceptibility locus for asthma-related traits. *Science*, **304**, 300–304.
39. Sleiman, P.M., Flory, J., Imielinski, M., Bradfield, J.P., Annaiah, K., Willis-Owen, S.A., Wang, K., Rafaels, N.M., Michel, S., Bonnelykke, K. *et al.* Variants of DENND1B associated with asthma in children. *N. Engl. J. Med.*, **362**, 36–44.
40. Dixon, A.L., Liang, L., Moffatt, M.F., Chen, W., Heath, S., Wong, K.C., Taylor, J., Burnett, E., Gut, I., Farrall, M. *et al.* (2007) A genome-wide association study of global gene expression. *Nat. Genet.*, **39**, 1202–1207.
41. van Heel, D.A., Fisher, S.A., Kirby, A., Daly, M.J., Rioux, J.D. and Lewis, C.M. (2004) Inflammatory bowel disease susceptibility loci defined by genome scan meta-analysis of 1952 affected relative pairs. *Hum. Mol. Genet.*, **13**, 763–770.
42. Villette, S., Kyle, J.A., Brown, K.M., Pickard, K., Milne, J.S., Nicol, F., Arthur, J.R. and Hesketh, J.E. (2002) A novel single nucleotide polymorphism in the 3' untranslated region of human glutathione peroxidase 4 influences lipoxygenase metabolism. *Blood Cells Mol. Dis.*, **29**, 174–178.
43. Batra, J., Rajpoot, R., Ahluwalia, J., Devarapu, S.K., Sharma, S.K., Dinda, A.K. and Ghosh, B. (2007) A hexanucleotide repeat upstream of eotaxin gene promoter is associated with asthma, serum total IgE and plasma eotaxin levels. *J. Med. Genet.*, **44**, 397–403.
44. Thye, T., Nejentsev, S., Intemann, C.D., Browne, E.N., Chinbuah, M.A., Gyapong, J., Osei, I., Owusu-Dabo, E., Zeitels, L.R., Herb, F. *et al.* (2009) MCP-1 promoter variant −362C associated with protection from pulmonary tuberculosis in Ghana, West Africa. *Hum. Mol. Genet.*, **18**, 381–388.
45. Modi, W.S., Goedert, J.J., Strathdee, S., Buchbinder, S., Detels, R., Donfield, S., O'Brien, S.J. and Winkler, C. (2003) MCP-1-MCP-3-Eotaxin gene cluster influences HIV-1 transmission. *AIDS*, **17**, 2357–2365.
46. Shen, C., Landers, C.J., Derkowski, C., Elson, C.O. and Targan, S.R. (2008) Enhanced CBir1-specific innate and adaptive immune responses in Crohn's disease. *Inflamm. Bowel. Dis.*, **14**, 1641–1651.
47. Duchmann, R., Schmitt, E., Knolle, P., Meyer zum Buschenfelde, K.H. and Neurath, M. (1996) Tolerance towards resident intestinal flora in mice is abrogated in experimental colitis and restored by treatment with interleukin-10 or antibodies to interleukin-12. *Eur. J. Immunol.*, **26**, 934–938.
48. Pang, H., Koda, Y., Soejima, M., Fujitani, N., Ogaki, T., Saito, A., Kawasaki, T. and Kimura, H. (2001) Polymorphism of the human ABO-Secretor locus (FUT2) in four populations in Asia: indication of distinct Asian subpopulations. *Ann. Hum. Genet.*, **65**, 429–437.
49. Yamamoto, K., Miwa, T., Taniguchi, H., Nagano, T., Shimamura, K., Tanaka, T. and Kumagai, H. (1996) Binding specificity of *Lactobacillus* to glycolipids. *Biochem. Biophys. Res. Commun.*, **228**, 148–152.
50. Lee, Y.J., Yu, W.K. and Heo, T.R. (2003) Identification and screening for antimicrobial activity against *Clostridium difficile* of *Bifidobacterium* and *Lactobacillus* species isolated from healthy infant faeces. *Int. J. Antimicrob. Agents*, **21**, 340–346.
51. Moorthy, G., Murali, M.R. and Devaraj, S.N. (2009) Lactobacilli facilitate maintenance of intestinal membrane integrity during Shigella dysenteriae 1 infection in rats. *Nutrition*, **25**, 350–358.
52. Iwamori, M. and Domino, S.E. (2004) Tissue-specific loss of fucosylated glycolipids in mice with targeted deletion of alpha(1,2)fucosyltransferase genes. *Biochem. J.*, **380**, 75–81.
53. Lin, B., Saito, M., Sakakibara, Y., Hayashi, Y., Yanagisawa, M. and Iwamori, M. (2001) Characterization of three members of murine alpha1,2-fucosyltransferases: change in the expression of the Se gene in the intestine of mice after administration of microbes. *Arch. Biochem. Biophys.*, **388**, 207–215.
54. Meng, D., Newburg, D.S., Young, C., Baker, A., Tonkonogy, S.L., Sartor, R.B., Walker, W.A. and Nanthakumar, N.N. (2007) Bacterial symbionts induce a FUT2-dependent fucosylated niche on colonic epithelium via ERK and JNK signaling. *Am. J. Physiol. Gastrointest. Liver Physiol.*, **293**, G780–G787.

55. Collado, M.C., Meriluoto, J. and Salminen, S. (2007) Role of commercial probiotic strains against human pathogen adhesion to intestinal mucus. *Lett. Appl. Microbiol.*, **45**, 454–460.
56. Swidsinski, A., Loening-Baucke, V., Vaneechoutte, M. and Doerffel, Y. (2008) Active Crohn's disease and ulcerative colitis can be specifically diagnosed and monitored based on the biostructure of the fecal flora. *Inflamm. Bowel Dis.*, **14**, 147–161.
57. Hurd, E.A. and Domino, S.E. (2004) Increased susceptibility of secretor factor gene Fut2-null mice to experimental vaginal candidiasis. *Infect. Immun.*, **72**, 4279–4281.
58. Park, S.W., Zhen, G., Verhaeghe, C., Nakagami, Y., Nguyenvu, L.T., Barczak, A.J., Killeen, N. and Erle, D.J. (2009) The protein disulfide isomerase AGR2 is essential for production of intestinal mucus. *Proc. Natl Acad. Sci. USA*, **106**, 6950–6955.
59. Johansson, M.E., Phillipson, M., Petersson, J., Velcich, A., Holm, L. and Hansson, G.C. (2008) The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. *Proc. Natl Acad. Sci. USA*, **105**, 15064–15069.
60. van der Sluis, M., Bouma, J., Vincent, A., Velcich, A., Carraway, K.L., Buller, H.A., Einerhand, A.W., van Goudoever, J.B., Van Seuningen, I. and Renes, I.B. (2008) Combined defects in epithelial and immunoregulatory factors exacerbate the pathogenesis of inflammation: mucin 2-interleukin 10-deficient mice. *Lab. Invest.*, **88**, 634–642.
61. Heazlewood, C.K., Cook, M.C., Eri, R., Price, G.R., Tauro, S.B., Taupin, D., Thornton, D.J., Png, C.W., Crockford, T.L., Cornall, R.J. *et al.* (2008) Aberrant mucin assembly in mice causes endoplasmic reticulum stress and spontaneous inflammation resembling ulcerative colitis. *PLoS Med.*, **5**, e54.
62. An, G., Wei, B., Xia, B., McDaniel, J.M., Ju, T., Cummings, R.D., Braun, J. and Xia, L. (2007) Increased susceptibility to colitis and colorectal tumors in mice lacking core 3-derived *O*-glycans. *J. Exp. Med.*, **204**, 1417–1429.
63. Linden, S.K., Sutton, P., Karlsson, N.G., Korolik, V. and McGuckin, M.A. (2008) Mucins in the mucosal barrier to infection. *Mucosal. Immunol.*, **1**, 183–197.
64. Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M. *et al.* (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature*, **449**, 851–861.
65. Meijerink, E., Neuenschwander, S., Fries, R., Dinter, A., Bertschinger, H.U., Stranzinger, G. and Vogeli, P. (2000) A DNA polymorphism influencing alpha(1,2)fucosyltransferase activity of the pig FUT1 enzyme determines susceptibility of small intestinal epithelium to *Escherichia coli* F18 adhesion. *Immunogenetics*, **52**, 129–136.
66. Berglund, E.D., Li, C.Y., Bina, H.A., Lynes, S.G., Michael, M.D., Shanafelt, A.B., Kharitonenkov, A. and Wasserman, D.H. (2009) Fibroblast growth factor 21 (FGF21) controls glycemia via regulation of hepatic glucose flux and insulin sensitivity. *Endocrinology*, **150**, 4084–4093.
67. Inagaki, T., Lin, V.Y., Goetz, R., Mohammadi, M., Mangelsdorf, D.J. and Kliewer, S.A. (2008) Inhibition of growth hormone signaling by the fasting-induced hormone FGF21. *Cell Metab.*, **8**, 77–83.
68. Fried, L.P., Borhani, N.O., Enright, P., Furberg, C.D., Gardin, J.M., Kronmal, R.A., Kuller, L.H., Manolio, T.A., Mittelmark, M.B., Newman, A. *et al.* (1991) The Cardiovascular Health Study: design and rationale. *Ann. Epidemiol.*, **1**, 263–276.
69. Gunderson, K.L., Steemers, F.J., Lee, G., Mendoza, L.G. and Chee, M.S. (2005) A genome-wide scalable SNP genotyping assay using microarray technology. *Nat. Genet.*, **37**, 549–554.
70. Gunderson, K.L., Steemers, F.J., Ren, H., Ng, P., Zhou, L., Tsan, C., Chang, W., Bullis, D., Musmacker, J., King, C. *et al.* (2006) Whole-genome genotyping. *Methods Enzymol.*, **410**, 359–376.
71. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
72. Nielsen, D.M., Ehm, M.G. and Weir, B.S. (1998) Detecting marker–disease association by testing for Hardy–Weinberg disequilibrium at a marker locus. *Am. J. Hum. Genet.*, **63**, 1531–1540.
73. Wittke-Thompson, J.K., Pluzhnikov, A. and Cox, N.J. (2005) Rational inferences about departures from Hardy–Weinberg equilibrium. *Am. J. Hum. Genet.*, **76**, 967–986.
74. Hoh, J., Wille, A. and Ott, J. (2001) Trimming, weighting, and grouping SNPs in human case–control association studies. *Genome Res.*, **11**, 2115–2119.
75. Taylor, J. and Tibshirani, R. (2006) A tail strength measure for assessing the overall univariate significance in a dataset. *Biostatistics*, **7**, 167–181.
76. Wang, J. and Shete, S. (2008) A test for genetic association that incorporates information about deviation from Hardy–Weinberg proportions in cases. *Am. J. Hum. Genet.*, **83**, 53–63.