Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER

Linkage analysis identifies novel genetic modifiers of microbiome traits in families with inflammatory bowel disease

Arunabh Sharma [ID][a], Silke Szymczak [ID][a#], Malte Rühlemann [ID][b], Sandra Freitag-Wolf [ID][a], Carolin Knecht[a], Janna Enderle [ID][c], Stefan Schreiber [ID][b,d], Andre Franke [ID][b], Wolfgang Lieb [ID][c], Michael Krawczak [ID][a], and Astrid Dempfle [ID][a]

aInstitute of Medical Informatics and Statistics, Kiel University, Kiel, Germany; bInstitute of Clinical Molecular Biology, Kiel University, Kiel, Germany; cInstitute of Epidemiology, Kiel University, Kiel, Germany; dDepartment of Internal Medicine I, University Hospital Schleswig-Holstein, Kiel, Germany

ABSTRACT

Dysbiosis of the gut microbiome is a hallmark of inflammatory bowel disease (IBD) and both, IBD risk and microbiome composition, have been found to be associated with genetic variation. Using data from families of IBD patients, we examined the association between genetic and microbiome similarity in a specific IBD context, followed by a genome-wide quantitative trait locus (QTL) linkage analysis of various microbiome traits using the same data. SNP genotypes as well as gut microbiome and phenotype data were obtained from the Kiel IBD family cohort (IBD-KC). The IBD-KC is an ongoing prospective study in Germany currently comprising 256 families with 455 IBD patients and 575 first- and second-degree relatives. Initially focusing upon known IBD risk loci, we noted a statistically significant (FDR<0.05) association between genetic similarity at SNP rs11741861 and overall microbiome dissimilarity among pairs of relatives discordant for IBD. In a genome-wide QTL analysis, 12 chromosomal regions were found to be linked to the abundance of one of seven microbial genera, namely *Barnesiella* (chromosome 4, region spanning 10.34 cM), *Clostridium_XIVa* (chr4, 3.86 cM; chr14, two regions spanning 7.05 and 13.02 cM respectively), *Pseudoflavonifractor* (chr7, 12.80 cM) *Parasutterella* (chr14, 8.26 cM), *Ruminococcus* (chr16, two overlapping regions spanning 8.01 and 16.87 cM, respectively), *Roseburia* (chr19, 7.99 cM), and *Odoribacter* (chr22, three regions spanning 0.89, 5.57 and 1.71 cM, respectively), as well as the Shannon index of α diversity (chr3, 1.47 cM). Our study thus shows that, in families of IBD patients, pairwise genetic similarity for at least one IBD risk locus is associated with overall microbiome dissimilarity among discordant pairs of relatives, and that hitherto unknown genetic modifiers of microbiome traits are located in at least 12 human genomic regions.

Introduction

Inflammatory bowel disease (IBD) is a chronic relapsing condition of the intestine[1] that affects more than 6.8 million people worldwide.[2] IBD comprises two main sub-entities, namely, Crohn's disease (CD) and ulcerative colitis (UC), both of which have a complex etiology involving genetic, microbial and environmental factors as well as their interactions.[3] More than 200 genetic variants associated with an increased IBD risk have been identified so far.[4] In addition, dysbiosis of the gut microbiome has been shown to be a hallmark of IBD, presenting itself as a reduction in overall gut microbial biodiversity.[4–6] Whether the latter is a cause or a consequence of the disease, however,

is not well understood. In any case, some bacterial species clearly exert a protective effect against IBD whereas others have a predisposing role,[7] as is exemplified by a decrease of beneficial bacteria, such as *Faecalibacterium praunitzii* and *Roseburia intestinalis*, and an increase of harmful pathogens[6] in IBD patients.

The gut microbiome compositions of members of one and the same family tend to be more similar than those of unrelated individuals,[8,9] a phenomenon usually ascribed to the fact that family members often live in the same environment and have similar dietary habits.[10] At the same time, however, blood relatives also share genetic factors which means that familial microbiome similarity may also reflect genetic

similarity.[10] Indeed, heritability analyses of the gut microbiomes of twins revealed that the abundance of a number of taxa, including *Christensenellaceae*, *Ruminococcaceae*, and *Lachnospiraceae*, is under some kind of genetic controls.[10,11] In addition, genome-wide association studies (GWAS) showed evidence for various links between host genetic variation and gut microbiome composition.[12–14] In the particular context of IBD, statistical associations were detected between DNA sequence variation at the *NOD2* gene locus and the abundance of *Enterobacteriaceae*.[15] In addition, an IBD genetic risk score based upon 11 disease-associated genetic variants was found to be negatively correlated with the abundance of *Roseburia*.[6] These results notwithstanding, however, the fact that monozygotic twins discordant for UC tend to have different microbiomes, with a notable reduction of diversity seen in the affected sib,[16] suggests that the joint role of genome and microbiome in IBD etiology is complex.

Most previous studies of the connection between gut microbiome and genetic variation employed a GWAS approach. However, GWAS are usually unable to discern the impact of rare variants upon the microbiome and, particularly when carried out in population-representative samples, cannot take the specificities of a certain disease context into account. This prompted us to adopt a family-based design to study the link between host genetics and gut microbiome in IBD etiology. Family-based studies are capable of localizing genetic effects irrespective of the frequencies of the variants involved as long as the effects of interest segregate in families.[17,18] We therefore used genetic and microbiome data from the IBD kindred cohort (IBD-KC), a nationwide prospective study in Germany of IBD patients and their families, to examine the heritability of the abundance of individual microbial genera and the diversity of whole microbiomes, followed by a genome-wide quantitative trait locus (QTL) linkage analysis of the abundances and α diversity at genus level. Finally, we performed association analyses within the identified linkage regions to identify individual SNPs associated with one or more of these microbiome traits.

Results

Cohort characteristics

Our study data were derived from the Kiel IBD kindred cohort (IBD-KC, see Methods), which at the time comprised 1030 individuals from 256 families, including 455 IBD patients. Genotype, phenotype and microbiome data were available for 703 IBD-KC participants (Table 1). None of the potential confounders of a possible genotype-phenotype relationship (e.g. age, sex) differed notably between affected and non-affected family members. A total of 22 bacterial genera yielded a count ≥10 in at least 50% of the samples (Supplementary Table 1), which qualified them as 'highly prominent' for the purpose of the subsequent genome-wide linkage analysis. Alpha and β diversity plots of the microbiome data are provided in Supplementary Fig. 1.

Heritability analysis

For the 22 highly prominent genera in our study (i.e. count ≥10 in ≥50% of samples), the covariate-adjusted, pedigree-based estimates of the heritability (h^2) ranged from 0 to 27.6% (Figure 1, Supplementary Table 2).

Pairwise genetic and microbiome similarity

Our study of the relationship between pairwise genetic and microbiome similarity was focussed upon genomic regions implicated in IBD before through GWAS. We analyzed genetic and microbiome similarity separately for pairs of blood relatives who were concordant or discordant for IBD. For the overall genetic similarity at IBD risk loci, measured by the sum of locus-specific excess identity-by-descent sharing values ($\Delta\rho$) as defined in the 'Methods' section, no significant association with β diversity was observed in linear

Table 1. Characteristics (median [1st quartile, 3rd quartile] or number [%]) of study samples according to IBD affection status.

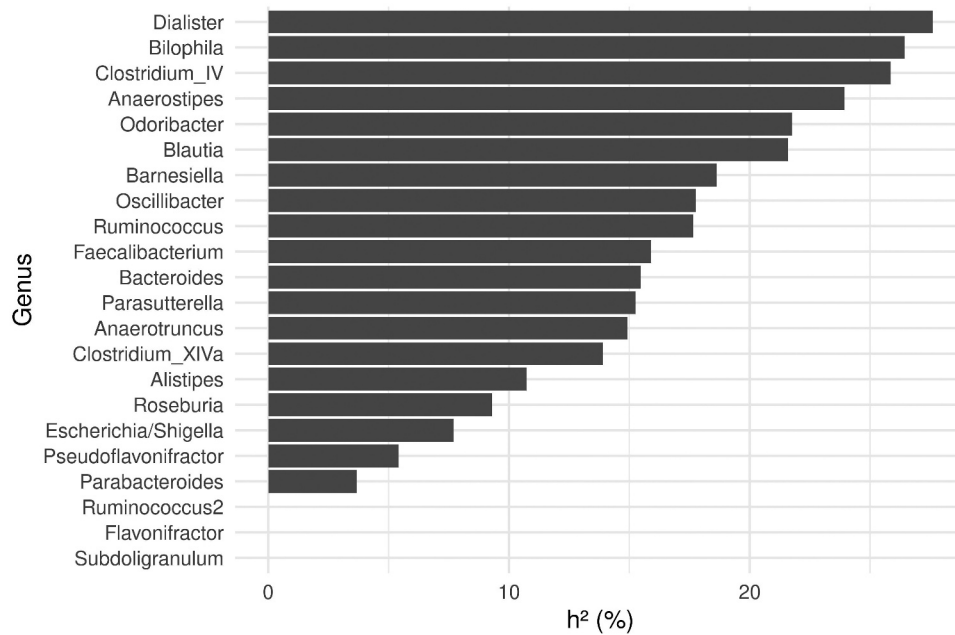|	IBD	Non-IBD
Total	294	409
Age	41.9 [30.2, 53.0]	42.6 [26.0, 58.2]
Body mass index (BMI)	23.9 [20.9, 26.4]	24.8 [21.4, 27.3]
Female [%]	182 [61.9]	231 [56.5]
Ever smoker [%]	29 [9.2]	50 [12.2]

Figure 1. Heritability (h^2) estimates for 22 microbial genera.

regression analyses after multiple testing correction, irrespective of whether β diversity was calculated from all amplicon sequence variants (ASVs) or from all genera. However, when Δρ was considered for individual SNPs, and when β diversity was based upon all genera, one significant association between the two similarity measures emerged among IBD-discordant pairs after multiple testing correction (FDR ≤0.05), namely for SNP rs11741861 (p = 2.8×10^{-5}, q = 0.005). Results for the 20 SNPs with the lowest p values are listed in Supplementary Table 3. Marker rs11741861, which is located on chromosome (chr) 5 at position 150277909, is an intronic variant of both the *IRGM* (immunity-related GTPase M) and the *ZNF300* (zinc finger protein 300) genes, according to the GWAS Catalog.[19] The linear regression coefficient relating Δρ to β was estimated as 0.08 (95% CI: [0.04, 0.13]), indicating that IBD-discordant pairs of relatives who were genetically more similar were less similar in terms of their microbiome composition (Figure 2). No particular phenotype such as IBD subentity or age at diagnosis was found to be related to Δρ.

Genome-wide QTL linkage analysis

Evidence for linkage at genome-wide significance level (i.e. logarithm of odds [LOD] score >3) was obtained between 12 different chromosomal regions

and the abundance of one of seven microbial genera, namely *Barnesiella* (maximum LOD score 3.24 on chr4), *Clostridium_XIVa* (max LOD 3.02 on chr4, max LOD 4.39 for region 1 on chr14, max LOD 3.60 for region 2 on chr14), *Pseudoflavonifractor* (max LOD 3.13 on chr7), *Parasutterella* (max LOD 3.02 on chr14), *Ruminococcus* (max LOD 4.42 for region 1 on chr16, max LOD 3.11 for region 2 on chr16), *Roseburia* (max LOD 3.78 on chr19), and *Odoribacter* (max LOD 4.31 for region 1 on chr22, max LOD 4.26 for region 2 on chr22, max LOD 4.08 for region 3 on chr22), as well as the Shannon index of α diversity (max LOD 3.72 on chr3) (Table 2). A complete overview of the results of the genome-wide linkage analysis is provided in Supplementary Fig. 2.

SNPs located in three discovered regions (Table 3, Figure 3a-c), namely those linked to the abundance of *Barnesiella* (chr4), *Clostridium_XlVa* (chr14), and *Roseburia* (chr19), respectively, have been reported before to be associated with IBD or one of its sub-entities.[21–23] The SNPs in question are rs13126505 and rs3774937 (chr4), rs8005161 (chr14) and rs17771967 (chr19). SNPs rs13126505 and rs3774937 are intronic variants of the *BANK1* (B cell scaffold protein with ankyrin repeats 1) and *NFKB1* (nuclear factor kappa B subunit 1) genes, respectively, rs8005161 is an intronic variant of the
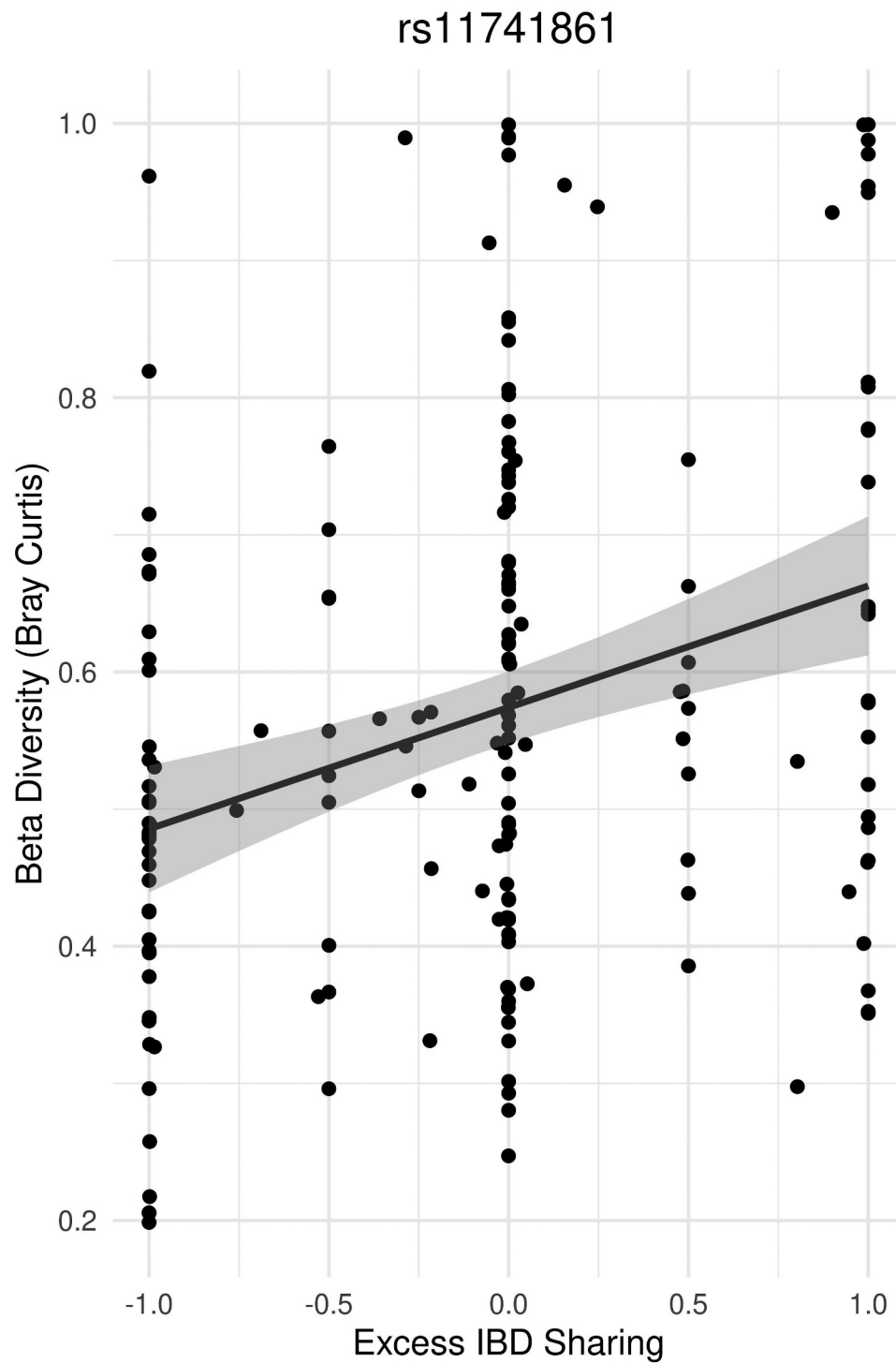
Figure 2. Statistically significant association (FDR≤0.05) between β diversity and excess identical-by-descent sharing (Δρ) for SNP rs11741861. Each dot represents a pair of blood relatives discordant for IBD.

GPR65 (G protein-coupled receptor 65) gene, and rs17771967 is an intergenic variant mapping to the region of the *RNU6-222P* (RNA, U6 small nuclear 222, pseudogene) and *KIR3DL2* (killer cell immunoglobulin like receptor, three Ig domains and long cytoplasmic tail 2) genes.

Table 2. Genetic linkage analysis of microbiome traits (i.e. genus abundance, Shannon index) in IBD families. For each significantly linked locus (LOD>3), the two-unit support interval demarcates the chromosomal region in which the surrounding LOD score differs by less than 2 units from the maximum LOD score. The number n of genes in each region was determined using biomaRt software (v.2.44.4).[20] Mb: mega base pairs; cM: centi-Morgan.

Microbiome trait	Chr.	Two-unit support interval	Physical (genetic) size of region	Genes (n)
Shannon index	3	rs838269 – rs16865790	0.61 Mb (1.47 cM)	13
Barnesiella	4	rs28532850 – rs28521501	15.17 Mb (10.34 cM)	146
Clostridium_XIVa	4	rs74648066 – rs115532888	3.51 Mb (3.86 cM)	31
Pseudoflavonifractor	7	rs73686954 – rs117808204	10.64 Mb (12.80 cM)	163
Parasutterella	14	rs10142572 – rs179667	6.11 Mb (8.26 cM)	58
Clostridium_XIVa (region 1)	14	rs79049326 – rs10138419	6.80 Mb (7.05 cM)	64
Clostridium_XIVa (region 2)	14	rs17707618 – rs1956681	12.76 Mb (13.02 cM)	226
Ruminococcus (region 1)	16	rs79541675 – rs77680845	5.21 Mb (8.01 cM)	133
Ruminococcus (region 2)	16	rs9972717 – rs10500519	11.61 Mb (16.87 cM)	194
Roseburia	19	rs117653309 – rs114419168	2.31 Mb (7.99 cM)	168
Odoribacter (region 1)	22	rs16987300 – rs111633245	0.32 Mb (0.89 cM)	38
Odoribacter (region 2)	22	rs78835794 – rs6518994	3.68 Mb (5.57 cM)	56
Odoribacter (region 3)	22	rs117342979 – rs73164039	2.02 Mb (1.71 cM)	97

Table 3. IBD-associated SNPs significantly linked to microbial abundance. Gene: nearby IBD-associated gene according to the GWAS Catalog; bp: base pairs.

SNP	Chr.	Position (bp)	Genus	Gene
rs13126505	4	102865304	*Barnesiella*	*BANK1* (intronic)
rs3774937	4	103434253	*Barnesiella*	*NFKB1* (intronic)
rs8005161	14	88472595	*Clostridium_XIVa*	*GPR65* (intronic)
rs17771967	19	55380214	*Roseburia*	Intergenic

Notably, various SNPs not genotyped in the present study, but located within the two-unit support intervals for traits *Clostridium_XIVa* (rs116135844 on chr4, rs8019270 on chr14), *Barnesiella* (rs3775467 on chr4), and *Pseudoflavonifractor* (rs10248138 on chr7), have been reported previously to be associated with *Bacteroidales, Prevotella, Weissella*, and *Lachnobacterium*, respectively.[12,24–26]

Association analysis within the linkage regions

SNP-wise association analyses covering the two-unit support intervals linked to different microbiome traits (Figure 3a-c, Supplementary Fig. 3a-g) comprised a variable number of SNPs: 1821 for the abundance of *Barnesiella* (chr4), 528 for *Clostridium_XIVa* (chr4), 1995 for *Pseudoflavonifractor* (chr7), 1221 for *Clostridium_XIVa* (chr14, interval 1), 2139 for *Clostridium_XIVa* (chr14, interval 2), 986 for *Parasutterella* (chr14), 1192 for *Ruminococcus* (chr16, interval 1), 2391 for *Ruminococcus* (chr16, interval 2), 698 for *Roseburia* (chr19), 22 for *Odoribacter* (chr22, interval 1), 763 for *Odoribacter* (chr22, interval 2), 281 for *Odoribacter* (chr22,

interval 3), and 177 for the Shannon index of α diversity (chr3). A list of all SNPs exhibiting a nominally significant trait-association is provided in Supplementary Table 5. One SNP that clearly stood out because it had a p value one order of magnitude smaller than all other SNPs in the region was rs242076 (chr22:33229830), associated with the abundance of *Odoribacter* (Supplementary Fig. 3f, SNP highlighted in red). This SNP is an intronic variant of the *SYN3* and *TIMP3* genes, according to dbSNP.[27]

Discussion

The heritability estimates from the present study were in line with the results of previous microbiome studies in twins, in other types of siblings, and among North American Hutterites, all reporting consistently lower heritability of microbial traits (at different taxa classification levels) than of other heritable human traits.[25,28–30] Our linkage study led to the identification of 12 different chromosomal regions linked to the abundance of one of seven microbial genera, namely *Barnesiella, Clostridium_XIVa, Pseudoflavonifractor, Parasutterella, Ruminococcus, Roseburia*, and *Odoribacter*, or to the Shannon index of α diversity. Subsequent association analyses within the linkage regions identified a SNP on chromosome 22 as being associated with the abundance of genus *Odoribacter*.

Based upon place of residence information, we determined that only 44 of 364 pairs of individuals relevant to our study belonged to the same household. In a linear regression analysis relating
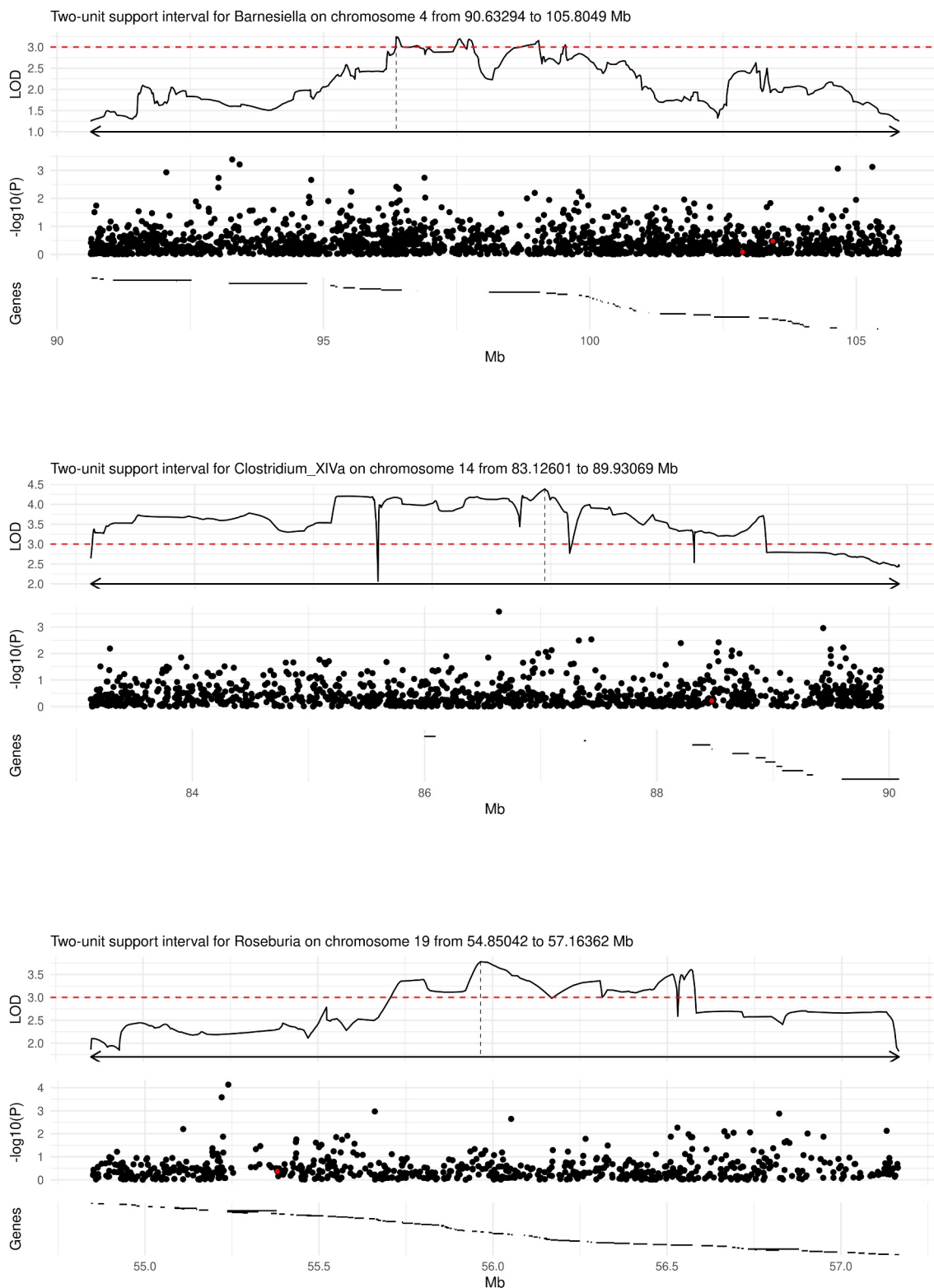
Figure 3. a-c Selected linkage regions of microbiome characteristics. Locus zoom plots span the two-unit support intervals (demarcated by horizontal arrowhead lines) of genome-wide significant linkages to the abundance of microbial genera, found with regions harboring IBD-associated SNPs (highlighted in red). Plots of the remaining linkage regions from Table 2 are provided in **Supplementary Fig. 3a-g**). Depicted are the LOD score (top panel), p value (on the log10 scale) from a SNP-wise association analysis (central panel) and the location of protein coding genes (bottom panel). The range of each gene is marked by a horizontal line; a list of gene names (from left to right) is provided in **Supplementary Table 4**. Dashed vertical lines mark the location of the max LOD in each interval.

pairwise genetic and microbiome similarity to one another, shared household was not a significant covariate. In the pedigree-based heritability analysis with MERLIN, however, only covariates at the single individual level (such as age or sex) could be accounted for, but not covariates pertaining to pairs of individuals. This means that shared familial environment may well have biased the heritability estimates obtained with MERLIN, but this bias is likely to have been small.

GWAS between microbiome and host genetics (mGWAS) are challenging due to a number of factors. First, the microbiome is a highly dynamic entity and evolves over time. Second, it is influenced by various demographic and environmental factors such as diet and lifestyle. Third, the very nature of microbiome data, such as sparseness, high dimensionality, and spurious correlation due to compositionality, renders their statistical analysis difficult in practice. Finally, owing to its mostly descriptive nature, a 16S rRNA-based inventory of taxa may not suffice to fully depict all relevant characteristics of microbiome function in a given sample. In consequence, the replicability of mGWAS has been considerably poor.[31]

This notwithstanding, mGWAS have identified several associations between host genetic variation and gut microbiome composition in the past.[12–14] Of particular interest in the context of IBD, *NOD2* gene variants have been shown to be associated with the presence or absence of certain microbial taxa.[15] Taken together with the alteration of microbiomes consistently seen in IBD patients,[4,6] this association suggests that the microbiome may function as an effector of at least some of the genetic risk of IBD. However, since such a causal link should plausibly depend upon additional disease-specific triggers as well, such as other genes and the environment, it called for studying the possible modulation, by genetic IBD risk factors, of the microbiome also in unaffected relatives of IBD patients.

Against this background, it is interesting to note that a recent study by Turpin et al.,[32] being specifically targeted at *NOD2*, demonstrated an association between variants in this gene and the fecal microbiome composition in healthy first-degree relatives of CD patients. Notably, the authors also pointed out the advantages of their study design over traditional association studies of population-representative healthy individuals. In contrast to Turpin et al.[32] however, our study (i) was not focused upon certain genetic variants, but covered all genomic regions implicated in IBD by GWAS before, and (ii) used a genetic linkage approach in IBD families to explore the possible joint role of genotype and microbiome in the etiology of the disease.

Our analysis of the relationship between intra-familial genetic similarity at IBD risk loci and microbiome dissimilarity, measured by β diversity, revealed one significant association for SNP rs11741861. IBD-discordant pairs of relatives who were genetically more similar at the respective IBD risk locus tended to be less similar in terms of their microbiome composition. This observation is consistent with similar findings in monozygotic IBD-discordant twins[16] and supports the idea that a genetic disposition to IBD may only lead to manifest disease in the presence of a susceptible microbiome. What is more, SNP rs11741861 is an intronic variant of the *IRGM* (immunity-related GTPase M) and *ZNF300* (zinc finger protein 300) genes on chromosome 5. Notably, altered levels of *IRGM* expression have been shown before to affect bacterial autophagy,[33] which in turn is linked to CD susceptibility. Moreover, IBD risk variants, including rs11741861, are known to influence the gut microbiota of healthy individuals in the form of a reduction in abundance of the butyrate producing genus *Roseburia*.[6]

Three of the 12 chromosomal regions identified as being linked to a microbiome QTL in IBD families, namely on chr4 (*Barnesiella* abundance), chr14 (*Clostridium_XlVa*), and chr19 (*Roseburia*), harbor four SNPs previously reported to be associated with either IBD or one of its sub-entities[21–23] (Table 3). Our linkage analysis was complemented by a genotype-phenotype association analysis within the two-unit support intervals, which identified SNP rs242076 (chr22 interval 2) as being associated with the abundance of genus *Odoribacter*. This SNP is an intronic variant of the *SYN3* and *TIMP3* genes. *TIMP3* has been shown to be down-regulated in CD patients[34] and a loss of *TIMP3* function was associated with gut microbiome dysbiosis, liver steatosis, and systemic inflammation.[35]

In conclusion, our study showed (i) that heritability is lower for microbiome traits than for other traits, (ii) that genetic similarity between IBD-discordant relatives is inversely correlated with microbiome similarity for at least one known IBD risk locus and (iii) that at least 12 distinct chromosomal regions in the human genome are linked to different microbiome traits in IBD families, including genera abundance and α diversity. Our study illustrates the usefulness of a family-based design to study genotype-microbiome relationships in the context of specific diseases. Future studies on IBD, in particular, should include targeted, preferably DNA sequence-based analyses of the linkage regions identified in our study to clarify further the causative role of the microbiome in disease etiology.

Methods

Cohort description

Using a multi-tiered approach (Figure 4), we analyzed genotype, phenotype, and microbiome data from the IBD-KC, a nationwide prospective study in Germany that was initiated in Kiel in 2013 and that has been ongoing since then, with regular follow-up assessments undertaken every two years.

Biomaterial (blood, stool, and hair) and questionnaire data have been collected from all study participants, currently comprising 455 IBD patients (including 11 cases of suspected IBD), all of whom were >6 years of age at the time of inclusion, and 575 first- and second-degree relatives. The IBD patients were recruited through treating physicians, clinics, study flyers or letters, and articles disseminated by patient organizations such as the Deutsche Morbus Crohn/Colitis ulcerosa Vereinigung. Each study participant received a study set containing the informed consent form, a questionnaire, and biomaterial tubes. The participants were asked to bring the biomaterial tubes to their next regular medical appointment, and their treating physicians were asked to conduct the blood sampling and to send tubes to the study laboratory at the Institute of Clinical Molecular Biology in Kiel. In addition, the IBD patients received a questionnaire to be filled out by their treating physician. Of the data collected by the IBD-KC, we used age, sex, smoking status, BMI, affection status, and household concordance as covariates in the present study.

The IBD-KC study protocol was approved by the ethics committee of the medical faculty of Kiel University (AZ 117/13). Each study participant signed an age-adapted informed consent (IC)
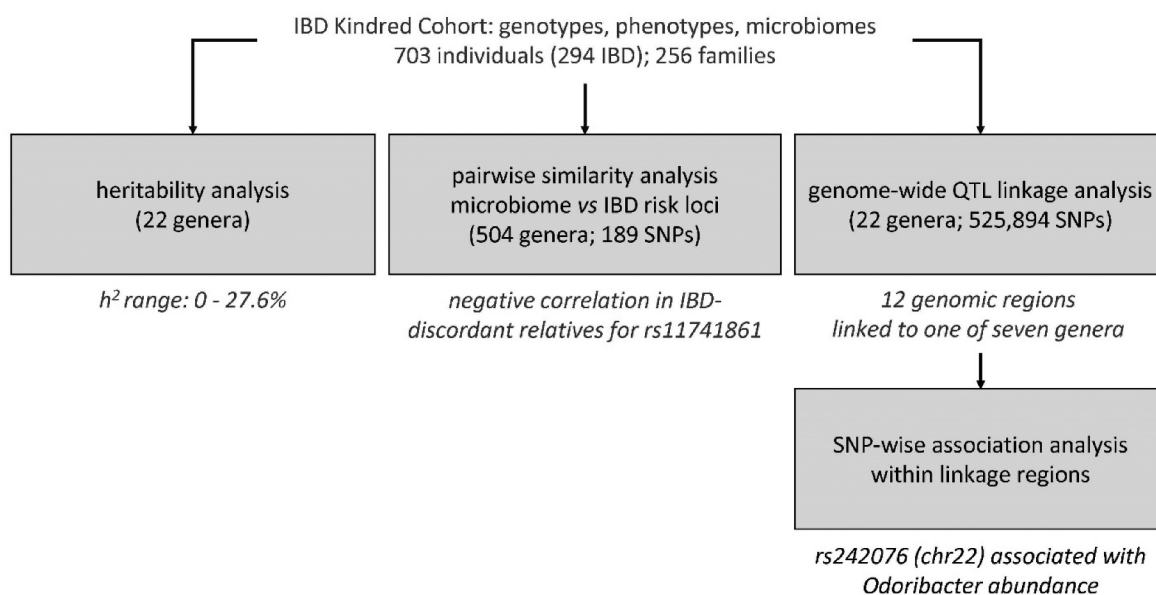

Figure 4. Multi-tiered genetic linkage and association analysis of microbiome data from the Kiel IBD Kindred Cohort (IBD-KC). The key findings of each analysis step are summarized in italic script.

form prior to their inclusion. For children <18 years of age, the IC had to be signed by the parents, and additional re-consent was sought at the child's 12th and 18th birthday.

Genotyping

All samples were genotyped with the Global Screening Array-24 Multi Disease v2.0, following the Illumina(R)Infinium HTS Assay Auto Workflow (Document #15045738v0). SNP calling was performed from idats image files using Illumina Genome Studio v2.0 with the appropriate cluster definition file (gsa-24-v2-0-A1-cluster-file.bpm). A total of 712,189 SNPs was available for analysis, including quality control (QC).

SNP alleles were converted to the plus strand of human reference genome build hg37 using strand files available at https://www.well.ox.ac.uk/~wrayner/strand/. SNPs lacking a match between the probe and reference sequence, or exhibiting more than two alleles, were removed. For SNPs mapping to the same chromosomal position, the one with the lowest number of missing genotypes in the IBD-KC data was kept for further analysis. Genetic distances were derived from the Rutgers Combined Linkage-Physical Map v3 (Kosambi, in centi-Morgan, cM)[36] and interpolated, if necessary, using a modified version of the SNPgenmap() function implemented in R package CrypticIBDcheck.[37]

QC of SNP genotypes was performed with PLINK[38] v1.9, unless stated otherwise, at both the SNP and the sample level. SNPs were excluded under one of the following conditions: call rate <0.99, minor allele frequency (MAF) equal to zero among family founders, Hardy-Weinberg equilibrium $p < 10^{-5}$ in unaffected founders, presence of alleles at odds with the EUR superpopulation of the 1000 Genomes Project, or an absolute MAF difference >0.2 to the EUR sample. Of the 686,627 SNPs for which genotype data were available, some 526,849 (76.7%) passed QC.

Samples were removed if their overall call rate was <0.97, if the proportion of heterozygous SNPs was outside mean±5SD, as estimated from all samples, or if the sex check failed.

Principal component analysis (PCA) is a standard approach in GWAS to detect outliers and to detect unknown population structure. To account for the family design of the present study, PCA was performed with PC-Air[39] from Bioconductor package GENESIS (v2.16.1).[40] Briefly, PC-Air extracts a subset of unrelated individuals and infers the axes of largest genetic variation in this subset. The genetic similarity to subset members is then used to determine the coordinates on the inferred axes for the remaining individuals. Function pedigreeMax Unrelated() of Bioconductor package GWASTools (v1.28.0)[41] was applied here to identify the required maximum set of unrelated individuals. Based upon the first two principal components derived, outlier samples were identified and removed using a bivariate generalization of boxplots, called bagplots,[42] as implemented in R package aplpack (v1.3.3), adopting a loop factor of 5. The PCA was performed twice, once using only genotypes from the IBD-KC and once after merging the data with the 1000 Genomes Project phase 1 genotype data. In total, 11 individuals were removed according to the above-mentioned criteria.

Reported family relationships were evaluated for consistency based upon the available SNP genotypes, using different strategies. First, possible relatedness between families was assessed considering the most likely level of identity-by descent (ibd) between pairs of presumably unrelated individuals from different families, estimated with PLINK from 50k SNPs with MAF >0.05 and pairwise linkage disequilibrium (LD) <0.1. Second, relationships within pedigrees were assessed for correctness with a maximized log-likelihood ratio (MLLR) test as implemented in PREST (v3_02),[43] using 5k SNPs with MAF >0.1 and pairwise LD <0.1. Third, potentially false relationships were checked further with ALTERTEST (v3_02),[43] using 2k SNPs with MAF >0.1 and pairwise LD <0.1 where the reduction in marker number became necessary due to the computational complexity and demand of the analysis. As a result, one pair of presumably unrelated individuals was removed because they could have been first cousins or at least distantly related, according to ALTERTEST. In addition, the genotypes of two samples involved in a potential sample swap were set to missing.

Mendelian inconsistencies were ascertained using level 1 and level 2 algorithms implemented in Pedcheck (v1.2).[44] Level 1 is the nuclear family algorithm that checks inconsistencies between

parents and offspring while the genotype elimination algorithm at level 2 aims to detect subtle inconsistencies by accounting for the full pedigree. Based upon level 1 results, one complete family with more than 500 Mendelian inconsistencies was removed. In addition, 955 SNPs with more than one Mendelian error were excluded. For the remaining inconsistencies, including those identified at level 2, genotypes in the respective families were set to missing as well.

16s RNA sequencing and data processing

IBD-KC participants collected fecal samples in standard stool collection tubes at home and sent them to the laboratory at room temperature within 24 hours. There, samples were stored at −80°C until further processing. DNA was extracted with the QIAamp DNA Stool Mini and QIAamp Fast DNA Stool Mini Kits (Qiagen) on a QIAcube system, adding an additional bead-beating step. Subsequent 16S rRNA gene library generation and sequencing was performed as previously described.[45] In short, the V1-V2 region of the 16S rRNA gene was sequenced on the MiSeq platform using v3 chemistry for 2x300bp paired-end reads (Illumina Inc., San Diego, CA, USA). At this point, it must be noted that the choice of 16s rRNA primers has been shown before to affect the resolution of the analysis of microbiome diversity.[46,47] In particular, primers for the V3-V4 region were found to identify more microbial taxa than V1-V2 primers in vaginal swaps,[46] and V2-V3 primers could be shown to yield the highest resolution for microbes from lower taxonomic ranks.[47] Therefore, it cannot be excluded that some bacterial taxa may have been overlooked in our study due to the lack of V3-V4 primers.

Demultiplexing after sequencing was based upon null mismatches in the barcode sequences. Data processing was performed using the DADA2 v1.10[48] workflow for big datasets (https://benjjneb. github.io/dada2/bigdata.html) resulting in abundance tables of ASVs. Briefly, all sequencing runs were handled separately (workflow adjusted for V1-V2 region can be found at https://github.com/ mruehlemann/ikmb_amplicon_processing/blob/ master/dada2_16S_workflow.R) and finally collected in a single abundance table per dataset, which underwent chimera filtering. ASVs underwent taxonomic annotation using the Bayesian classifier provided by DADA2 and the Ribosomal Database Project (RDP) version 16 release. All samples were rarefied to 10,000 counts.

Pre-processed 16s RNA sequence data were transformed into a phyloseq[49] object. The ASVs were collapsed at genus level to allow meaningful inclusion of very rare taxa. Gut microbiome abundance data are usually characterized by many low to zero counts, which results in severely skewed abundance distributions. We therefore normalized the abundance values at genus level using the log-transformed counts per million (cpm) function in R package edgeR.[50] The analysis of β diversity was carried out with non-normalized values.

Statistical analysis

Heritability analysis

Pedigree-based estimates of the heritability (h^2) of genus-specific abundance values and of overall microbiome diversity were obtained using the variance components method as implemented in MERLIN (v 1.1.2).[51] Briefly, heritability quantifies the proportion of inter-individual phenotypic variation attributable to genetic variation. For our heritability analyses, we considered the 22 'highly prominent' genera, defined by a count ≥10 in at least 50% of samples. Sex, smoking status, affection status, BMI, and age were included in this and all other statistical analyses as covariates.

Pairwise microbial and genetic similarity

The association between genetic and microbiome similarity was studied considering all blood-related pairs of eligible individuals from the IBD-KC, except parent-offspring and monozygotic twin pairs (because these types of relationship entail no variation in genetic similarity). The dissimilarity of their microbiomes was quantified by Bray-Curtis β diversity, calculated from either a) all ASVs (n = 50,167), b) all genera, or c) the 22 highly prominent genera. Pairwise genetic similarity was quantified as the local excess ibd sharing, estimated with MERLIN (v 1.1.2) for each of the 189 IBD risk loci reported by Liu et al.[21] To this end, SNPs <500kb distant from a given risk locus were used to derive the posterior probabilities p0,

p1 and p2 of the three possible ibd states at that locus, and the actual level of ibd sharing (ρ_{act}) was estimated as p1 + 2× p2. Next, the expected level of ibd sharing (ρ_{exp}) was derived for each pair of individuals from their pedigree-based familial relationship using R package *ribd*.[52] Finally, the excess ibd sharing was set equal to $\Delta\rho = \rho act - \rho exp$. Overall genetic similarity at all IBD risk loci combined was defined as the sum of the locus-specific $\Delta\rho$ values.

The relationship between microbiome and genetic similarity was analyzed by way of constructing linear models with microbiome dissimilarity as the response variable and genetic similarity as the predictor variable, treating pairs of individuals as observational units. Potential confounders and covariates with a reported effect upon the microbiome[45] were also included in the models, namely difference in age and BMI as well as concordance in terms of sex, smoking status, household, and IBD affection status. To allow for multiple testing, p values of the risk locus-wise analyses were corrected using a false discovery rate (FDR) approach at a nominal significance level of 0.05.

Genome-wide QTL linkage analysis

To identify chromosomal regions specifically linked, in IBD families, to α diversity at the genus level, quantified by either chao1 or Shannon index, or to the abundance of the 22 highly prominent genera, genome-wide multipoint QTL linkage analysis was carried out both with and without linkage disequilibrium (LD) pruning, using MERLIN-REGRESS (v 1.1.2). This software is an implementation of the regression-based method proposed by Sham et al[53] which relates the estimated local level of ibd sharing to the squared sum and the squared difference in the quantitative trait of interest. All covariates mentioned above were included in the linkage analysis as well.

Association analysis within linkage regions

Family-based association analyses between individual SNPs and microbiome traits were performed in the candidate chromosomal regions identified before through linkage analysis. The analyses were limited to the classical so-called 'n-unit support interval', defined as the chromosomal region in which the LOD score of SNPs differs by less than n units from the maximum LOD score.[54]

A LOD score >3 classically has been considered indicative of statistically significant linkage because it implies that the likelihood ratio of two loci being linked rather than unlinked is at least 1000. Yet, two randomly chosen loci in the human genome are about 50 times more probable to be unlinked so that the prior odds of linkage equal 1:50. Taken together, a LOD score >3 thus corresponds to posterior odds of linkage >20, so that p < .05.[55] The classical approach of adopting a one-unit support interval as a surrogate 95% confidence intervals for the location of a peak LOD signal is based upon the theory of maximum likelihood (ML) estimation, equating sampling variance of the ML estimate to the steepness of the likelihood curve around the ML estimate.[56] This approach has been criticized before for providing too narrow an interval at times so that we chose to adopt a two-unit support interval instead.

Statistical testing was performed with the WISARD workbench,[57] which is an implementation of the genome-wide efficient mixed-model association algorithm (GEMMA)[58] to fit a linear mixed model under the inclusion of a kinship matrix accounting for familial relationships. All covariates mentioned above were also included in the GEMMA analysis.

Acknowledgments

Disclosure statement

No potential conflict of interest was reported by the author(s).

Contributions

AS, MK; SSz and AD conceived the project; AS performed statistical analyses and wrote a first version of the manuscript; SSz performed QC of genotype data; MR and AF pre-

processed and provided microbiome data; JE and WL processed and provided genotype and demographic data; MK and AD prepared different revisions of the manuscript; all authors reviewed and approved the final version of manuscript.

ORCID

Arunabh Sharma http://orcid.org/0000-0001-7808-0942
Silke Szymczak http://orcid.org/0000-0002-8897-9035
Malte Rühlemann http://orcid.org/0000-0002-0685-0052
Sandra Freitag-Wolf http://orcid.org/0000-0002-1069-7740
Janna Enderle http://orcid.org/0000-0002-5481-5905
Stefan Schreiber http://orcid.org/0000-0003-2254-7771
Andre Franke http://orcid.org/0000-0003-1530-5811
Wolfgang Lieb http://orcid.org/0000-0003-2544-4460
Michael Krawczak http://orcid.org/0000-0003-2603-1502
Astrid Dempfle http://orcid.org/0000-0002-2618-3920

References

1. Caruso R, Lo BC, Núñez G. Host–microbiota interactions in inflammatory bowel disease. Nat Rev Immunol. 2020;20(7):411–426. doi:10.1038/s41577-019-0268-7.
2. Alatab S, Sepanlou SG, Ikuta K, Vahedi H, Bisignano C, Safiri S, Sadeghi A, Nixon MR, Abdoli A, Abolhassani H, et al. The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Gastroenterol Hepatol. 2020;5(1):17–30. doi:10.1016/S2468-1253(19)30333-4.
3. Lavoie S, Conway KL, Lassen KG, Jijon HB, Pan H, Chun E, Michaud M, Lang JK, Gallini Comeau CA, Dreyfuss JM, et al. The Crohn's disease polymorphism, ATG16L1 T300A, alters the gut microbiota and enhances the local Th1/Th17 response. eLife. 2019;8: e39982. doi:10.7554/eLife.39982.
4. Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, Andrews E, Ajami NJ, Bonham KS, Brislawn CJ, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature. 2019;569(7758):655–662. doi:10.1038/s41586-019-1237-9.
5. Ott SJ, Musfeldt M, Wenderoth DF, Hampe J, Brant O, Fölsch UR, Timmis KN, Schreiber S. Reduction in diversity of the colonic mucosa associated bacterial microflora in patients with active inflammatory bowel disease. Gut. 2004;53:685–693. doi:10.1136/gut.2003.025403.
6. Imhann F, Vila AV, Bonder MJ, Fu J, Gevers D, Visschedijk MC, Spekhorst LM, Alberts R, Franke L, van Dullemen HM, et al. The interplay of host genetics and the gut microbiota underlying the onset and clinical presentation of inflammatory bowel disease. Gut. 2018;67:108–119. doi:10.1136/gutjnl-2016-312135.
7. Liu S, Zhao W, Lan P, Mou X. The microbiome in inflammatory bowel diseases: from pathogenesis to therapy. Protein Cell [Internet]. 2020 [cited 2021 Mar 15]; Available from; doi:10.1007/s13238-020-00745-3.
8. Lee S , Sung J , Lee J , Ko G.Comparison of the gut microbiotas of healthy adult twins living in South Korea and the United Statess.Applied and Environmental Microbiology.77(20)7433–7437 doi:10.1128/AEM.05490-11
9. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, et al. Human gut microbiome viewed across age and geography. Nature. 2012;486:222–227 doi:10.1038/nature11053.
10. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, Beaumont M, Van Treuren W, Knight R, Bell JT, et al. Human genetics shape the gut microbiome. Cell. 2014;159(4):789–799. doi:10.1016/j.cell.2014.09.053.
11. Goodrich JK, Davenport ER, Clark AG, Ley RE. The relationship between the human genome and microbiome comes into view. Annu Rev Genet. 2017;51(1):413–433. doi:10.1146/annurev-genet-110711-155532.
12. Hughes DA, Bacigalupe R, Wang J, Rühlemann MC, Tito RY, Falony G, Joossens M, Vieira-Silva S, Henckaerts L, Rymenans L, et al. Genome-wide associations of human gut microbiome variation and implications for causal inference analyses. Nat Microbiol. 2020;5(9):1079–1087. doi:10.1038/s41564-020-0743-8.
13. Rühlemann MC, Hermes BM, Bang C, Doms S, Moitinho-Silva L, Thingholm LB, Frost F, Degenhardt F, Wittig M, Kässens J, et al. Genome-wide association study in 8,956 German individuals identifies influence of ABO histo-blood groups on gut microbiome. Nat Genet. 2021;53(2):147–155. doi:10.1038/s41588-020-00747-1.
14. Bonder MJ, Kurilshikov A, Tigchelaar EF, Mujagic Z, Imhann F, Vila AV, Deelen P, Vatanen T, Schirmer M, Smeekens SP, et al. The effect of host genetics on the gut microbiome. Nat Genet. 2016;48(11):1407–1412. doi:10.1038/ng.3663.
15. Kurilshikov A, Wijmenga C, Fu J, Zhernakova A. Host genetics and gut microbiome: challenges and perspectives. Trends Immunol. 2017;38(9):633–647. doi:10.1016/j.it.2017.06.003.

16. Lepage P, Häsler R, Spehlmann ME, Rehman A, Zvirbliene A, Begun A, Ott S, Kupcinskas L, Doré J, Raedler A, et al. Twin study indicates loss of interaction between microbiota and mucosa of patients with ulcerative colitis. Gastroenterology. 2011;141(1):227–236. doi:10.1053/j.gastro.2011.04.011.

17. Bailey-Wilson JE, Wilson AF. Linkage analysis in the next-generation sequencing era. Hum Hered. 2011;72:228–236. doi:10.1159/000334381.

18. Ott J, Wang J, Leal SM. Genetic linkage analysis in the age of whole-genome sequencing. Nat Rev Genet. 2015;16(5):275–284. doi:10.1038/nrg3908.

19. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, McMahon A, Morales J, Mountjoy E, Sollis E, et al. The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 2019;47(D1):D1005–12. doi:10.1093/nar/gky1120.

20. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 2009;4(8):1184–1191. doi:10.1038/nprot.2009.97.

21. Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, Ripke S, Lee JC, Jostins L, Shah T, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat Genet. 2015;47(9):979–986. doi:10.1038/ng.3359.

22. de Lange KM, Moutsianas L, Lee JC, Lamb CA, Luo Y, Kennedy NA, Jostins L, Rice DL, Gutierrez-Achury J, Ji S-G, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. Nat Genet. 2017;49(2):256–261. doi:10.1038/ng.3760.

23. Ellinghaus D, Jostins L, Spain SL, Cortes A, Bethune J, Han B, Park YR, Raychaudhuri S, Pouget JG, Hübenthal M, et al. Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. Nat Genet. 2016;48(5):510–518. doi:10.1038/ng.3528.

24. Blekhman R, Goodrich JK, Huang K, Sun Q, Bukowski R, Bell JT, Spector TD, Keinan A, Ley RE, Gevers D, et al. Host genetic variation impacts microbiome composition across human body sites. Genome Biol. 2015;16(1):191. doi:10.1186/s13059-015-0759-1.

25. Turpin W, Espin-Garcia O, Xu W, Silverberg MS, Kevans D, Smith MI, Guttman DS, Griffiths A, Panaccione R, Otley A, et al. Association of host genome with intestinal microbial composition in a large healthy cohort. Nat Genet. 2016;48(11):1413–1417. doi:10.1038/ng.3693.

26. Bubier JA, Chesler EJ, Weinstock GM. Host genetic control of gut microbiome composition. Mamm Genome. 2021;32(4):263–281. doi:10.1007/s00335-021-09884-2.

27. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001;29(1):308–311. doi:10.1093/nar/29.1.308.

28. Goodrich JK, Davenport ER, Beaumont M, Jackson MA, Knight R, Ober C, Spector TD, Bell JT, Clark AG, Ley RE. Genetic determinants of the gut microbiome in UK Twins. Cell Host Microbe. 2016;19(5):731–743. doi:10.1016/j.chom.2016.04.017.

29. Lim MY, You HJ, Yoon HS, Kwon B, Lee JY, Lee S, Song Y-M, Lee K, Sung J, Ko G. The effect of heritability and host genetics on the gut microbiota and metabolic syndrome. Gut. 2017;66(6):1031–1038. doi:10.1136/gutjnl-2015-311326.

30. Davenport ER, Cusanovich DA, Michelini K, Barreiro LB, Ober C, Gilad Y. Genome-wide association studies of the human gut microbiota. PLOS ONE. 2015;10:e0140301. doi:10.1371/journal.pone.0140301.

31. Awany D, Allali I, Dalvie S, Hemmings S, Mwaikono KS, Thomford NE, Gomez A, Mulder N, Chimusa ER. Host and microbiome genome-wide association studies: current state and challenges. Front Genet [Internet]. 2019 [cited 2021 Mar 15]; 9. Available from: https://www.frontiersin.org/articles/10.3389/fgene.2018.00637/full

32. Turpin W, Bedrani L, Espin-Garcia O, Xu W, Silverberg MS, Smith MI, Garay JAR, Lee S-H, Guttman DS, Griffiths A, et al. Associations of NOD2 polymorphisms with Erysipelotrichaceae in stool of in healthy first degree relatives of Crohn's disease subjects. BMC Med Genet. 2020;21:204. doi:10.1186/s12881-020-01115-w.

33. McCarroll SA, Huett A, Kuballa P, Chilewski SD, Landry A, Goyette P, Zody MC, Hall JL, Brant SR, Cho JH, et al. Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. Nat Genet. 2008;40:1107–1112. doi:10.1038/ng.215.

34. Monteleone I, Federici M, Sarra M, Franzè E, Casagrande V, Zorzi F, Cavalera M, Rizzo A, Lauro R, Pallone F, et al. Tissue inhibitor of metalloproteinase-3 regulates inflammation in human and mouse intestine. Gastroenterology. 2012;143:1277–1287.e4. doi:10.1053/j.gastro.2012.07.016.

35. Mavilio M, Marchetti V, Fabrizi M, Stöhr R, Marino A, Casagrande V, Fiorentino L, Cardellini M, Kappel B, Monteleone I, et al. A role for timp3 in microbiota-driven hepatic steatosis and metabolic dysfunction. Cell Rep. 2016;16:731–743. doi:10.1016/j.celrep.2016.06.027.

36. Matise TC, Chen F, Chen W, Vega FMDL, Hansen M, He C, Hyland FCL, Kennedy GC, Kong X, Murray SS, et al. A second-generation combined linkage–physical map of the human genome. Genome Res. 2007;17:1783–1786 doi:10.1101/gr.7156307.

37. Nembot-Simo A, Graham J, McNeney B. CrypticIBDcheck: an R package for checking cryptic relatedness in nominally unrelated individuals. Source Code Biol Med. 2013;8(1):5. doi:10.1186/1751-0473-8-5.

38. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience [Internet]. 2015 [cited 2021 Mar 15]; 4. 10.1186/s13742-015-0047-8.

39. Conomos MP, Miller M, Thornton T. Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. Genet Epidemiol. 2015;39(4):276–293. doi:10.1002/gepi.21896.

40. Gogarten SM, Sofer T, Chen H, Yu C, Brody JA, Thornton TA, Rice KM, Conomos MP. Genetic association testing using the GENESIS R/Bioconductor package. Bioinformatics. 2019;35:5346–5348. doi:10.1093/bioinformatics/btz567.

41. Gogarten SM, Bhangale T, Conomos MP, Laurie CA, McHugh CP, Painter I, Zheng X, Crosslin DR, Levine D, Lumley T, et al. GWASTools: an R/bioconductor package for quality control and analysis of genome-wide association studies. Bioinformatics. 2012;28 (24):3329–3331. doi:10.1093/bioinformatics/bts610.

42. Rousseeuw PJ, Ruts I, Tukey JW. The bagplot: a bivariate boxplot. Am Stat. 1999;53:382–387.

43. McPeek MS, Sun L. Statistical tests for detection of misspecified relationships by use of genome-screen data. Am J Hum Genet. 2000;66(3):1076–1094. doi:10.1086/302800.

44. O'Connell JR, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet. 1998;63(1):259–266. doi:10.1086/301904.

45. Wang J, Thingholm LB, Skiecevičienė J, Rausch P, Kummen M, Hov JR, Degenhardt F, Heinsen F-A, Rühlemann MC, Szymczak S, et al. Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. Nat Genet. 2016;48(11):1396–1406. doi:10.1038/ng.3695.

46. Graspeuntner S, Loeper N, Künzel S, Baines JF, Rupp J. Selection of validated hypervariable regions is crucial in 16S-based microbiota studies of the female genital tract. Sci Rep. 2018;8(1):9678. doi:10.1038/s41598-018-27757-8.

47. Bukin Y, Galachyants Y, Morozov IV, Bukin SV, Zakharenko AS, Zemskaya TI. The effect of 16S rRNA region choice on bacterial community metabarcoding results. Sci Data. 2019;6:190007. doi:10.1038/sdata.2019.7.

48. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13(7):581–583. doi:10.1038/nmeth.3869.

49. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE [Internet]. 2013 [cited 2021 Mar 15]; 8. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3632530/

50. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. 2012;40(10):4288–4297. doi:10.1093/nar/gks042.

51. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet. 2002;30(1):97–101. doi:10.1038/ng786.

52. Vigeland MD. Relatedness coefficients in pedigrees with inbred founders. J Math Biol. 2020;81(1):185–207. doi:10.1007/s00285-020-01505-x.

53. Sham PC, Purcell S, Cherny SS, Abecasis GR. Powerful regression-based quantitative-trait linkage analysis of general pedigrees. Am J Hum Genet. 2002;71 (2):238–253. doi:10.1086/341560.

54. Ott J. Analysis of human genetic linkage, 3rd ed. Baltimore and London: Johns Hopkins University Press; 1999.

55. Nyholt DR. All LODs are not created equal. Am J Hum Genet. 2000;67:282–288. doi:10.1086/303029.

56. Lander ES, Botstein D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics. 1989;121:185–199. doi:10.1093/genetics/121.1.185.

57. Lee S, Choi S, Qiao D, Cho M, Silverman EK, Park T, Won S. WISARD: workbench for integrated superfast association studies for related datasets. BMC Med Genomics. 2018;11(S2):39. doi:10.1186/s12920-018-0345-y.

58. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012;44:821–824. doi:10.1038/ng.2310.