



Quantifying the Impact of Human Leukocyte Antigen on the Human Gut Microbiota

Stijn P. Andeweg,^a Can Keşmir,^a  Bas E. Dutilh^a

^aTheoretical Biology and Bioinformatics, Science for Life, Utrecht University, Utrecht, the Netherlands

ABSTRACT The composition of the gut microbiota is affected by a number of factors, including the innate and adaptive immune system. The major histocompatibility complex (MHC), or the human leukocyte antigen (HLA) in humans, performs an essential role in vertebrate immunity and is very polymorphic in different populations. HLA determines the specificity of T lymphocyte and natural killer (NK) cell responses, including those against the commensal bacteria present in the human gut. Thus, it is likely that our HLA molecules, and thereby the adaptive immune response, can shape the composition of our microbiota. Here, we investigated the effect of HLA haplotype on the microbiota composition. We performed HLA typing and microbiota composition analyses on 3,002 public human gut microbiome data sets. We found that individuals with functionally similar HLA molecules are also similar in their microbiota composition. Our results show a statistical association between host HLA haplotype and gut microbiota composition. Because the HLA haplotype is a readily measurable parameter of the human immune system, these results open the door to incorporating the genetics of the immune system into predictive microbiome models.

IMPORTANCE The microorganisms that live in the digestive tracts of humans, known as the gut microbiota, are essential for hosts' survival, as they support crucial functions. For example, they support the host in facilitating the uptake of nutrients and give colonization resistance against pathogens. The composition of the gut microbiota varies among humans. Studies have proposed multiple factors driving the observed variation, including diet, lifestyle, and health condition. Another major influence on the microbiota is the host's genetic background. We hypothesized the immune system to be one of the most important genetic factors driving the differences observed between gut microbiotas. Therefore, we searched for a link between the polymorphic molecules that shape human immune responses and the composition of the microbiota. HLA molecules are the most polymorphic molecules in our genome and therefore makes an excellent candidate to test such an association. To our knowledge for the first time, our results indicate a significant impact of the HLA on the human gut microbiota.

KEYWORDS gut microbiome, human leukocyte antigen, immune-microbiome interactions, quantitative biology


The gut microbiota is essential for the existence of the mammalian host and performs crucial functions, such as nutrition and colonization resistance against pathogens (1, 2). In addition, commensal bacteria shape host immunity (3, 4): commensal bacteria influence the development and homeostasis of the immune system, and in the absence of a gut microbiota the innate and adaptive immune system is impaired (5). These functions are fulfilled by the 10^{13} to 10^{14} bacteria making up the human gut microbiota of a typical adult (6). The composition of the gut microbiota varies substantially over the human population (7, 8), and numerous exogenous and intrinsic factors

Citation Andeweg SP, Keşmir C, Dutilh BE. 2021. Quantifying the impact of human leukocyte antigen on the human gut microbiota. *mSphere* 6:e00476-21. <https://doi.org/10.1128/mSphere.00476-21>.

Editor June Round, University of Utah

Copyright © 2021 Andeweg et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Bas E. Dutilh, bedutilh@gmail.com.

 It has remained challenging to measure an effect of HLA on gut microbiome in humans. In this article, Stijn Andeweg @CKesmir and @BEDutilh develop a new functional measure and show that people with a similar HLA profile have similar microbiota composition.

Received 1 June 2021

Accepted 25 July 2021

Published 11 August 2021

are proposed to influence the microbial community composition, leading to observed microbiome diversity (9, 10). These factors include dietary components, mode of delivery during birth, breastfeeding, disease history, and host genetics (10–15). However, the greater part of population diversity is still unexplained. To obtain a better understanding of the microbiota composition and function, a comprehensive list of associated factors is required.

Studies with monozygotic twins have highlighted the importance of host genetics in shaping the microbiota composition (16, 17), where the immune system is thought to be one of the most important contributing factors (18). The immune system is a well-known but not fully explored mechanism influencing the microbiome (18–21). Conversely, the microbiota also imprints its signature on the host immune system, e.g., by influencing the abundances of mucosa-associated invariant T cells by exposure to microbial antigens (22). Immune response related factors that generate variation between microbiotas are likely to originate from functional differences among polymorphic immune molecules.

Major histocompatibility complex (MHC) genes, known as the human leukocyte antigen (HLA) genes in humans, are one of the most polymorphic genes found in vertebrates (23, 24). They encode cell surface glycoproteins that present antigens to T cells (25–27) and form the first essential step in the generation of an adaptive immune response. Influence of the MHC haplotype on the microbiome has been observed in mice (18, 28–30) and in stickleback (31). In humans, specific polymorphisms in the HLA region have been associated with several common infections (32), and in a genome-wide association study of over 1,500 individuals, two variants located in the HLA region showed suggestive association with microbiome composition or function (12). Moreover, an association with microbiota composition has been observed for the HLA-DQ gene in relation to celiac disease (33, 34). However, other genome-wide association studies did not find associations between HLA and the microbiome among many associated genes (35, 36).

Here, we analyzed 3,002 human gut microbiome data sets and found that HLA haplotype similarity correlates with microbiota similarity. Importantly, functionally stratifying individuals on the basis of HLA-presented microbial antigens improved the association with microbiota similarity. Overall, these results provide the first statistical support for the functional association between HLA and the gut microbiota in humans.

RESULTS

Mining gut microbiome data sets. To link the HLA haplotype to microbiota composition of individuals, we simultaneously determined the presence of human HLA genes and bacterial small-subunit (SSU) rRNA genes in human intestinal samples. We composed a combined reference database consisting of HLA gene sequences and SSU rRNA sequences and used this to map 3,002 gut microbiome data sets extracted from the Sequence Read Archive (SRA). As expected, we observed a large variability in the number of HLA reads recovered from the different data sets (Table 1; also, see Fig. S1A in the supplemental material). This variability could be due to differences in sequencing techniques between studies or due to differences between individuals. We designed a decision tree (see Materials and Methods) to predict the HLA haplotype for the classical HLA class I (HLA-A, HLA-B, and HLA-C) and HLA class II (HLA-DP, HLA-DQ and HLA-DR) genes. The decision tree predicted alleles from replicate samples of the same individual with a mean overall consistency of 75%, and zygosity was predicted with a mean consistency of 80% (Table 1). This performance is particularly notable for the specific alleles, as they were predicted based on a reference database consisting of between 5 (HLA-DQ) and 35 (HLA-B) two-digit alleles depending on the HLA gene, while zygosity was only a binary call (heterozygote/homozygote, i.e., a 50% guessing rate). Together, these results provide a rough reliability estimate for our HLA typing pipeline.

Application of this scheme to all 3,002 data sets resulted in prediction of the complete HLA class I and II profiles for 127 and 309 individuals, respectively (Fig. S1B). HLA

TABLE 1 Consistency of HLA prediction over the available data sets from the same individual^a

Individual	Gene	Alleles predicted	Zygosity	Dataset 1 allele (reads)	Dataset 2 allele (reads)	Dataset 3 allele (reads)	Dataset 4 allele (reads)	Dataset 5 allele (reads)
1	HLA-A	A*02: 5/5 A*68: 5/5 Consistency: 100%	Heterozygote: 5/5 Homozygote: 0/5 Consistency: 100%	A*02 (17.0)	HLA-A*02 (8.0)	HLA-A*02 (8.0)	HLA-A*02 (20.0)	HLA-A*02 (16.0)
				A*68 (3.0)	HLA-A*68 (7.0)	HLA-A*68 (4.0)	HLA-A*68 (4.0)	HLA-A*68 (6.0)
1	HLA-B	B*07: 5/5 B*07: 3/5 B*51: 2/5 Consistency: 80%	Heterozygote: 2/5 Homozygote: 3/5 Consistency: 60%	B*07 (49.0)	HLA-B*51 (17.0)	HLA-B*07 (24.0)	HLA-B*07 (75.0)	HLA-B*07 (62.0)
				B*07 (49.0)	HLA-B*07 (9.0)	HLA-B*51 (20.0)	HLA-B*07 (75.0)	HLA-B*07 (62.0)
1	HLA-C	C*05: 5/5 C*05: 2/5 C*04: 2/5 No allele: 1/5 Consistency: 70%	Heterozygote: 2/5 Homozygote: 2/5 No zygosity: 1/5 Consistency: 50%	C*05 (7.0)	HLA-C*04 (22.0)	HLA-C*04 (22.0)	HLA-C*05 (10.0)	HLA-C*05 (16.0)
				2 nd allele not called	HLA-C*05 (9.0)	HLA-C*05 (9.0)	HLA-C*05 (10.0)	HLA-C*05 (16.0)
2	HLA-A	A*24: 4/4 A*29: 2/4 A*02: 1/4 A*03: 1/4 Consistency: 75%	Heterozygote: 4/4 Homozygote: 0/4 Consistency: 100%	A*24 (7.0)	HLA-A*24 (10.0)	HLA-A*24 (6.0)	HLA-A*24 (4.0)	
				A*29 (3.0)	HLA-A*29 (2.0)	HLA-A*02 (3.0)	HLA-A*03 (1.0)	
2	HLA-B	B*07: 4/4 B*51: 3/4 No allele: 1/4 Consistency: 88%	Heterozygote: 3/4 Homozygote: 0/4 No zygosity: 1/4 Consistency: 75%	HLA-B*07 (7.0)	HLA-B*51 (3.0)	HLA-B*07 (5.0)	HLA-B*51 (3.0)	
				2 nd allele not called	HLA-B*07 (3.0)	HLA-B*51 (4.0)	HLA-B*07 (1.0)	
2	HLA-C	C*08: 3/4 C*05: 2/4 C*16: 2/4 C*07: 1/4 Consistency: 63%	Heterozygote: 4/4 Homozygote: 0/4 Consistency: 100%	HLA-C*08 (6.0)	HLA-C*05 (4.0)	HLA-C*08 (6.0)	HLA-C*08 (4.0)	
				HLA-C*05 (1.0)	HLA-C*16 (2.0)	HLA-C*07 (1.0)	HLA-C*16 (2.0)	
3	HLA-A	A*03: 4/4 A*29: 2/4 No allele: 2/4 Consistency: 75%	Heterozygote: 2/4 Homozygote: 0/4 No zygosity: 2/4 Consistency: 50%	HLA-A*03 (4.0)	HLA-A*03 (8.0)	HLA-A*03 (8.0)	HLA-A*03 (8.0)	
				HLA-A*29 (3.0)	2 nd allele not called	HLA-A*29 (2.0)	2 nd allele not called	
3	HLA-B	B*07: 4/4 B*40: 3/4 No allele: 1/4 Consistency: 88%	Heterozygote: 3/4 Homozygote: 0/4 No zygosity: 1/4 Consistency: 75%	HLA-B*07 (9.0)	HLA-B*07 (14.0)	HLA-B*07 (13.0)	HLA-B*07 (2.0)	
				2 nd allele not called	HLA-B*40 (6.0)	HLA-B*40 (5.0)	HLA-B*40 (2.0)	
3	HLA-C	C*04: 4/4 C*07: 3/4 No allele: 1/4 Consistency: 88%	Heterozygote: 3/4 Homozygote: 0/4 No zygosity: 1/4 Consistency: 75%	HLA-C*04 (2.0)	HLA-C*04 (3.0)	HLA-C*07 (3.0)	HLA-C*04 (4.0)	
				2 nd allele not called	HLA-C*07 (3.0)	HLA-C*04 (2.0)	HLA-C*07 (2.0)	
4	DPB1	DPB1*03: 2/2 DPB1*04: 1/2 DPB1*15: 1/2 Consistency: 50%	Heterozygote: 2/2 Homozygote: 0/2 Consistency: 100%	DPB1*03 (1.0)	DPB1*03 (1.0)			
				DPB1*04 (1.0)	DPB1*15 (1.0)			
4	DQB1	DQB1*02: 2/2 DQB1*06: 1/2 DQB1*03: 1/2 Consistency: 50%	Heterozygote: 2/2 Homozygote: 0/2 Consistency: 100%	DQB1*02 (5.0)	DQB1*03 (2.0)			
				DQB1*06 (2.0)	DQB1*02 (1.0)			
4	DRB1	DRB1*12: 2/2 DRB1*12: 1/2 DRB1*04: 1/2 Consistency: 75%	Heterozygote: 1/2 Homozygote: 1/2 No zygosity called	DRB1*12 (23.0)	DRB1*12 (21.0)			

^aHLA predictions from different data sets compared to the most likely HLA alleles prediction of the individual. Each two-digit allele was given a unique color, showing that most alleles were consistently predicted for different individuals. Frequently occurring alleles (e.g., B*07) in this selection of predictions were also observed more frequently in the general population (Fig. S2). In the set of samples from the same individual, no class II predictions were made because of a lack of class II reads found in this data set.

class I is expressed in all nucleated cells, while HLA class II expression is restricted to antigen-presenting cells (APCs) (25). Therefore, we expected a higher number of reads mapping to HLA class I in human gut data sets, but the difference between HLA class I and II reads is not significant ($P = 0.17$, two-sample Kolmogorov-Smirnov [KS] test). The observation of similar read counts between HLA class I and II may be explained by a

large number of lymphocytes (especially B cells, which have high MHC class II expression) being present in the gut as a part of mucosal immunity (37). Moreover, intestinal epithelial cells also express MHC class II (38), and HLA class I and class II expression levels vary across different cell types (39, 40). Further differences in the gene expression pattern of gut-associated cell types, as well as stochastic differences in, e.g., sampling and analysis protocols, might contribute to the observed variation in HLA mapped reads between the SRA data sets.

Inferred HLA prevalence reflects the Caucasian population profile. As an additional quality assessment of our HLA typing pipeline, the predicted population allele frequencies of the HLA class I and class II groups were compared to the HLA prevalence in the National Marrow Donor Program (NMDP) database, a database containing high-resolution HLA allele frequencies (41). The frequency distribution of HLA class I in our data set is similar to that in the NMDP database of Caucasian ethnicity (Pearson's $r=0.806$, 0.726 , and 0.812 for HLA-A, HLA-B, and HLA-C, respectively), but the frequency distribution of HLA class II is less so (Pearson's $r=0.537$ and 0.127 for HLA-DQB1 and HLA-DRB1, respectively) (Fig. S2A to E). Finally, we observed a high frequency of HLA heterozygosity and low frequency of homozygosity for any of the HLA genes (Fig. S2F), consistent with the high polymorphism of HLA genes in the population (41). All genes except HLA-C had zygosity frequencies that were consistent with the NMDP data (Fig. S2F).

Individuals with similar HLA genes have similar microbiota. Next, we compared microbiota similarity between the individuals in search for a link to similarity in HLA alleles. To do this, we determined the similarity in microbiota composition for pairs of data sets that were stratified by the number of shared HLA class I alleles (Fig. 1). As expected, microbiota similarity was generally higher between individuals who shared more HLA alleles, although the effect size was modest, possibly as a result of the low number of sample pairs with highly similar HLA profiles.

We expect that two individuals with the same HLA molecules may have similar immune control of their microbiota. However, the opposite is not true, i.e., individuals with different sets of HLA molecules can nevertheless generate similar immune responses, because different HLA molecules may have similar binding motifs. To address this issue of functional similarity, HLA molecules have been grouped into supertypes, consisting of HLA molecules that present similar peptides on the cell surface (42). In a first attempt to represent the functional HLA similarity between individuals and increase statistical power at this highly polymorphic gene, we performed association analyses at the level of HLA supertypes. While this increased the number of sample pairs with similar HLA profiles, the relation between HLA similarity and microbiota beta diversity was lost (Fig. S3). This may be because the HLA supertypes are a very generalized representation of HLA molecules, while the functional effects of the HLA molecules on the microbiota composition are likely very specific (43).

In a second attempt to quantify the functional HLA similarity between individuals, we devised a novel approach specific to human gut microbiota by focusing on the overlap in peptides presented by the different HLA molecules. To do this, we used NetMHCpan4.0, a state-of-the-art peptide-MHC binding affinity prediction tool (44), to predict the binding affinity of over two million gut bacterial oligopeptides to the HLA class I molecules. The peptides were extracted from the human gut catalog (45) in order to best represent the HLA functionality in the context of their potential response to the human gut microbiota. This approach allowed us to calculate a presented peptide similarity score (PPSS) between the HLA haplotypes of each pair of individuals (see Materials and Methods). Individuals with a high PPSS present very similar peptides, while individuals with a low PPSS present different peptides. When we stratified pairs of individuals by PPSS, we observed a significantly higher microbiota similarity for pairs with a high PPSS (Fig. 2).

Microbiota alpha diversity is higher in individuals with homozygous HLA alleles. Although we found that our zygosity predictions were less reliable than our haplotype predictions (Table 1), we still analyzed the effect of zygosity on the microbiota diversity. To this end, we used a subset of data sets of which we could determine the full

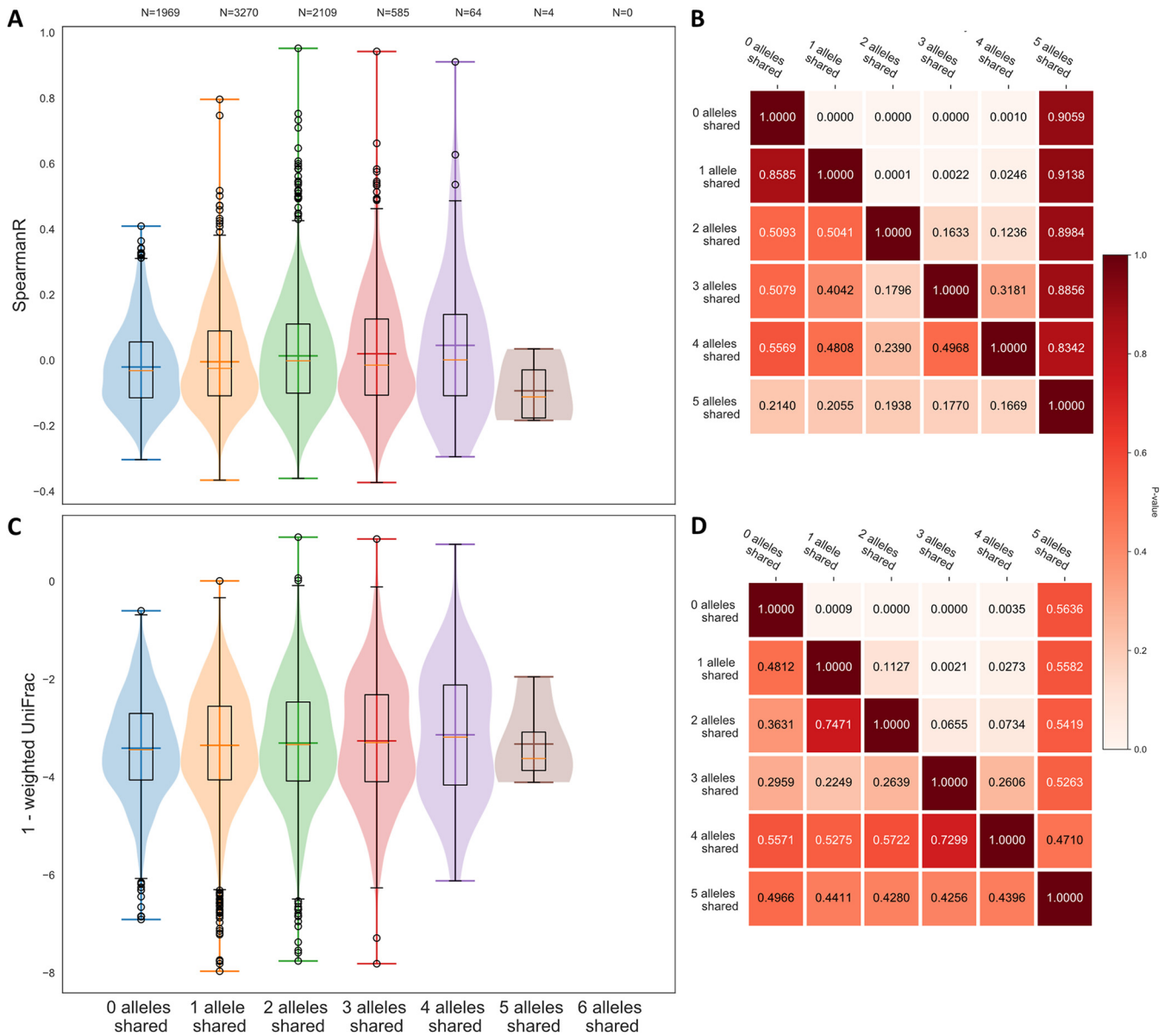


FIG 1 Microbiota beta diversity for sample pairs with an increasing number of shared HLA class I alleles. (A) Spearman rank correlation and (C) weighted UniFrac of the microbiota composition for sample pairs with 0 (blue), 1 (orange), 2 (green), 3 (red), 4 (purple), or 5 (brown) alleles shared. (B and D) Heat maps displaying the *P* values for the two-sample Kolmogorov-Smirnov test under the null hypothesis that the item in the row is drawn from an equal or smaller distribution compared to the item in the column.

zygosity of all HLA class I or II genes ($n = 60$ and $n = 205$, respectively). Although we found only very few individuals who were fully homozygous for all HLA genes, we observed a significantly higher microbiota richness in homozygous individuals than in individuals with one or more heterozygous HLA class I genes (Fig. S4A and B). Moreover, we observed a trend of a decrease of the richness as the homozygosity decreased (Fig. S4A and B). Similar results were obtained for HLA class II (Fig. S4E and F). We also investigated the correlation between HLA zygosity and microbiota evenness, but this yielded inconsistent results (Fig. S4C, D, G, and H).

Next, we analyzed whether the zygosity effect on alpha diversity was equally distributed over the HLA class I genes. For HLA class I genes, individuals homozygous for HLA-B genes had a higher richness than HLA-A and HLA-C genes (Fig. S5A). We also found a significant difference for the HLA-DQB1 and HLA-DPB1 class II genes (Fig. S5C). Association between the gut microbiota and HLA-B was observed previously (12, 29)

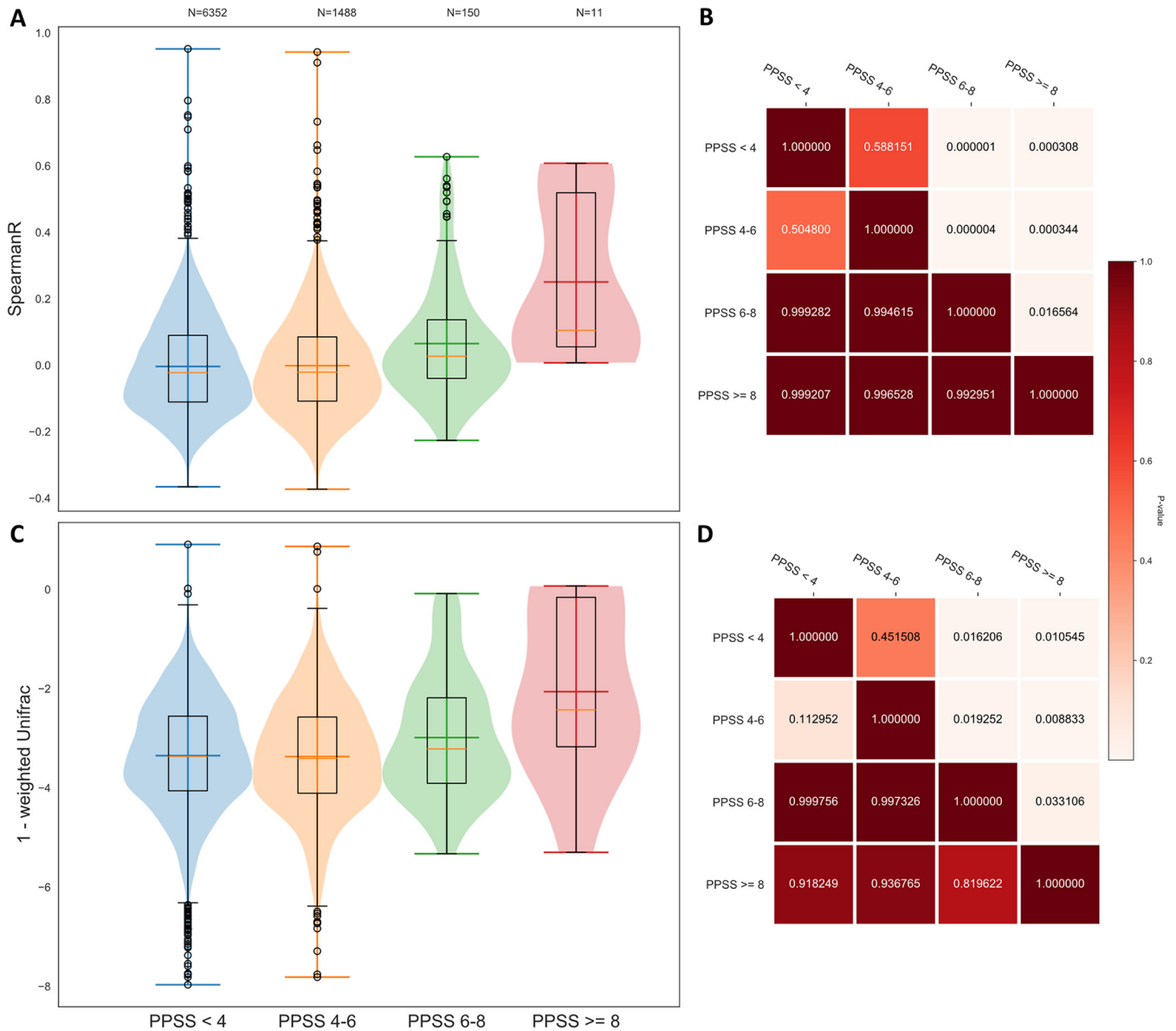


FIG 2 Microbiota beta diversity for sample pairs with increasing PPSS. Spearman rank correlation (A) and weighted UniFrac distance (C) of the microbiota composition for sample pairs with PPSS of <4 (blue), 4 to 6 (orange), 6 to 8 (green), and ≥ 8 (red). (B and D) Heat map displaying the *P* values for the two-sample Kolmogorov-Smirnov test under the null hypothesis that the item in the row is drawn from an equal or smaller distribution compared to the item in the column.

and is consistent with HLA-B often generating immunodominant T cell responses (46–49) and the most polymorphic one (24). We also observed a higher microbiota evenness associated with HLA-DQB1 and -DPB1 homozygosity (Fig. S5D). Finally, we reemphasize that our zygosity predictions might suffer from reduced consistency, and therefore, the above results on the effect of HLA zygosity on human gut microbiota should be interpreted carefully.

DISCUSSION

We present the first statistical evidence in humans that HLA significantly affects the composition of the gut microbiota, in line with previous results in other animals (18, 20, 28). To do this, we developed the PPSS, a score that uses cutting-edge machine learning tools (44), to quantify the functional similarity between HLA haplotypes of different individuals. We observed the highest correlation between the microbiota and

HLA class I profiles using the PPSS, suggesting that this score outperforms simpler measures of HLA type similarity, such as the number of shared alleles between individuals. Thus, if the association holds up in large, well-controlled cohorts, we expect that the PPSS will be useful for comparing HLA/MHC profiles in relation to commensal and pathogenic microorganisms in humans.

HLA class I molecules are involved in intracellular antigen presentation, leading to a cytotoxic response by CD8⁺ T cells. Cross-presentation of extracellular antigens on MHC class I by dendritic cells is common (26, 27, 50, 51), but it remains unknown how antigens presented on HLA class I shape mucosal immunity. In addition, the influence of HLA class I on the microbiota could be related to the MHC class I chain-related antigen A (MIC-A) and antigen B (MIC-B) genes. These polymorphic genes are constitutively expressed in the gut epithelium (52). The recognition of MIC-A/B by its ligand NKG2D, an activating receptor on NK cells, NKT cells, and T cells, leads to cytotoxic response and cytokine production (53). MIC-A and -B are genomically localized next to HLA-B on chromosome 6 and therefore in strong linkage disequilibrium with HLA-B (54, 55). Because we found a microbiota association on HLA class I and the strongest zygosity effect on the gut microbiota for HLA-B (Fig. S5), it will be of interest to dissect the association between HLA-B or MIC-A/B and the microbiota in large cohorts.

MHC class II molecules are suggested to play a role via antibody-mediated response (18, 28, 31). One possibility is that a less diverse HLA repertoire in homozygous individuals may reduce the number and diversity of immune responses to the microbiota by T cells and natural killer (NK) cells. As a result, homozygous individuals would recognize and potentially remove fewer bacterial species from their intestinal microbiota than heterozygous individuals. This could allow a higher number of different bacterial species to colonize the gut than the individuals with a more diverse set of HLA molecules. A similar effect has been observed in the stickleback, where fish with more diverse MHC class II molecules had a less diverse gut microbiota (31). Interestingly, this raises a potential conflict between the fitness benefits of a diverse immune system and of a diverse microbiome. Diverse immune responses are considered beneficial, especially in the protection against rapidly evolving pathogens (56), while a diverse microbiome is considered healthy, since it provides ecological protection against pathogen invasion and avoids dysbiosis (57). Therefore, it remains an open question what the potential effect is of zygosity on the fitness of the host. We focused our analysis on HLA class I molecules because the imputed HLA class II haplotypes did not match the expected population profile and consistency (Table 1 and Fig. S2D and E). Thus, we consider the class II typing less reliable than the class I typing. Moreover, the prediction of presented peptides is not as reliable for class II as the prediction for class I, in part because the binding is more promiscuous for class II than for class I. The seven class II supertypes show a >30% overlap in presented peptides between supertypes, which is similar to the overlap between alleles within a class II supertype (58). We expect that measurements with increased accuracy will reveal similar results with HLA class II haplotypes.

Conclusion. An important standing conundrum in microbiome research is the vast interindividual differences in gut microbiome composition between people. While many factors are known that contribute to shaping the microbiome, the relative importance of these factors remains unclear and will be the subject of study for years to come. We expect that innovative data analysis methods, designed to directly probe the functionality of factors such as the human immune system (12, 14), will allow these factors to be elucidated and their contributions mapped. As HLA is a readily measurable parameter of the immune system, the exploration of HLA and HLA-driven peptide presentation opens the door for incorporating these factors into future predictive models of the gut microbiome.

MATERIALS AND METHODS

Human gut microbiome data sets. We retrieved 3,002 human gut data sets from 37 different bio-projects from the public Sequence Read Archive (SRA) database (59). Data sets were collected using “metatranscriptome” search terms (Table 2), but due to inconsistencies in SRA metadata annotation, our

TABLE 2 Search terms for the SRA query^a

Search term	Filter			No. of data sets
	Source	Organism	Selection	
Human bowel	Metatranscriptomic		Random	841
Human gut	Metatranscriptomic		Random	925
	Transcriptomic	Human gut metagenome orgn txid:408170	Random	130
	Metatranscriptomic	Human gut metagenome orgn txid:408170	Random	897
Human stool	Metatranscriptomic		Random	778
RNA sequencing		Human gut metagenome orgn txid:408170	Random	2,879
Metatranscriptomic		Human gut metagenome orgn txid:408170		2,179

^aOverlapping data sets from the different search terms were filtered out.

selection included 750 metagenomes as well. Microbiota and HLA haplotype profiling are performed on both data types. Repeating the analyses with the subset of 2,250 metatranscriptomes led to a steep decrease in the number of paired individuals with highly similar HLA profiles and loss of signal, and therefore, we decided to continue with the original data set. Metadata fields are listed in Table S2. We are grateful to the associated studies for making these data available for reuse.

All sequence runs were preprocessed as follows. Data sets were downloaded using prefetch (version 2.9.2) utility, and SRA data were converted to fastq files using fastq-dump (version 2.9.2), both from the SRA toolbox. The quality of the reads was assessed using FastQC (version 0.11.8) (60), and reads were trimmed using AdapterRemoval (version 2.1.7) (61). For studies with multiple data sets from the same individual, we randomly selected one data set to include in our analysis to avoid biasing the data set with intraindividual sample comparisons.

Mapping reads to a composite reference database. To assess the microbiota composition and the HLA haplotype at once, reads were mapped with BWA-MEM (version 0.7.12) (62) to a reference database consisting of bacterial small-subunit (SSU) rRNA (SILVA SSU Ref nonredundant [NR] 99, release 132) (63) and genomic HLA sequences from IMGT/HLA database (release 3.31.0, 24). 16S reads were analyzed to obtain the bacterial composition of each data set. Only bacterial operational taxonomic units (OTUs) with a relative abundance of ≥ 0.0001 were taken into account. Therefore, the data sets should contain $\geq 1,000$ 16S reads. We developed a sorting scheme for multimapping reads, which were divided across target sequences as a fraction of the number of primary mapping (unique) reads (64–66).

HLA imputation. We performed two-digit HLA typing on the classical HLA loci (HLA-A, HLA-B, HLA-C, HLA-DR, HLA-DP, and HLA-DQ). Although we expect a maximum of two alleles for each gene (25), we often found more than two HLA alleles with reads mapped. This is likely a result of the high sequence similarity of the different alleles from the same HLA gene. In order to make the most reliable HLA imputation, we defined an *R* score as the ratio of the number of reads mapped to the two most frequently observed HLA alleles over the number of reads mapped to all other alleles of this gene (Fig. 3A). We determined *R* for each of the six classical HLA loci in all 3,002 data sets. Only HLA genes with an *R* score of 2 or higher were considered for HLA imputation (Fig. 3B). An individual was considered homozygotic at a given locus if at least 10 reads aligned to the top HLA allele and the second mapped allele had less than 15% of the reads

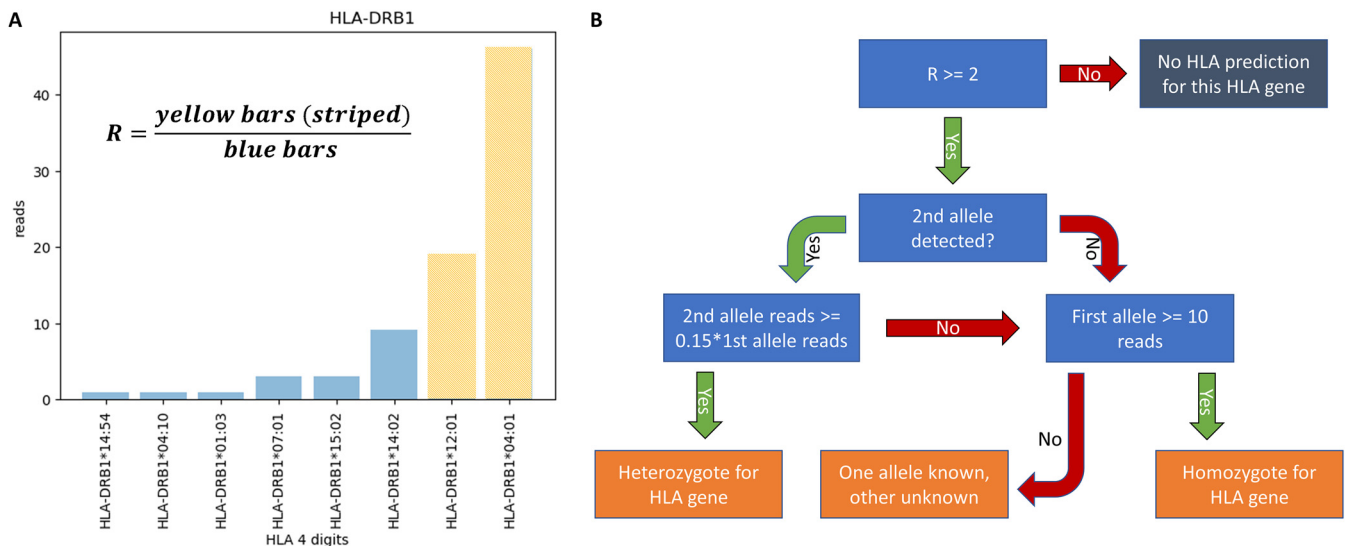


FIG 3 Decision tree genotyping of the HLA gene. (A) Visual explanation of the *R* score. (B) Decision tree for HLA typing from fecal microbiome data sets. Blue boxes show the decisions, orange boxes represent possible positive outcomes, and the gray box shows the negative outcome when no prediction is made for this gene.

of the first. A locus was considered heterozygotic if at least two alleles were found and the second mapped to at least 15% of the reads of the first. This number was optimized based on the highest number of correct predictions in data sets originating from the same individual (Table S1).

Data set selection and population-based HLA typing quality control. Individual HLA loci may be in linkage disequilibrium with other loci (67, 68). To maximize the representation of the HLA profile, we included data sets only if the complete HLA haplotype was assessed, resulting in two groups for the classical class I and II ($n = 127$ and 309 , respectively) (Table S2).

Quality control of our haplotyping approach was performed by comparing the population allele and zygosity frequencies of our data with the NMDP High-Resolution HLA Alleles and Haplotypes in the U.S. Population database (41; <https://bioinformatics.bethematchclinical.org/hla-resources/haplotype-frequencies/high-resolution-hla-alleles-and-haplotypes-in-the-us-population/>). Data were extracted in September 2019. Based on the data set localities, we expected a majority of individuals in our data sets to be of European descent (Table S2). Therefore, we compared our HLA frequencies with the cohort of European origin in NMDP database. A second quality control approach was performed using multiple data sets originating from the same individual; this information was obtained from the metadata. In the final analysis, only one data set per individual was used.

HLA supertypes and PPSS. Most of the 450 haplotyped data sets had unique HLA profiles, reflective of the high population-level diversity at the HLA loci. To increase statistical sensitivity, we performed two separate experiments on the data. First, we clustered HLA-A and HLA-B alleles into six supertypes each (42). Second, we devised a presented peptidome similarity score (PPSS) between two haplotypes to estimate the functional similarity in immune profile, based on the overlap in peptides presented by their HLAs. To do this, first we created a peptidome data set consisting of human gut bacterial peptides sampled from the 9,878,647 proteins in the human gut catalogue (45). By using a sliding window of length nine amino acids, we randomly included one in every thousand oligopeptides into the peptidome and removed duplicates, resulting in a peptidome of 2,371,920 peptides. Next, we used NetMHCpan4.0 to predict which peptides were presented by each HLA class I allele (44). We used the recommended cutoff 2.0%Rank for classifying binding peptides.

The PPSS between two individuals, i and j , is defined as the summed Jaccard indices, $J_{g,ai,aj}$ of the sets of peptides presented by all combinations of the alleles ai and aj of HLA gene g (HLA-A, HLA-B, and HLA-C). The PPSS ranges from 0 (no overlap between presented peptides) to 12 (presented peptides are identical). The Jaccard-based similarity of peptide pools presented of allele pair a,b is defined as the overlap of peptide pools ($O_{a,b}$) divided by the not-shared peptide pools (equation 1).

$$J_{ab} = \frac{O_{a,b}}{(P_a + P_b - O_{a,b})} \quad (1)$$

The J_{ab} scores between all two-digit alleles are shown in Fig. S6. PPSS_{ij}, the similarity of peptide pools of the pair of individuals i and j , is defined by the allele pair similarity for all combinations of alleles within one gene, added up for all three HLA genes (equation 2).

$$PPSS_{ij} = \sum_{g=A,B,C} \sum_{ai=1,2} \sum_{aj=1,2} J_{g,ai,aj} \quad (2)$$

The distribution of PPSS scores across all data sets is shown in Fig. S7.

Data Set S1 contains the Jaccard indices for every pair of four-digit alleles for HLA gene HLA-A, HLA-B, and HLA-C, allowing readers to readily calculate the PPSS upon HLA typing.

Alpha diversity. To assess alpha diversity, we used evenness and richness measures from the skbio.diversity.alpha package (version 0.5.5) (69). We did not rarefy the data, as this may increase type I and II errors (70). We did not find a correlation between data set size and microbiota richness ($r = 0.027$; $P = 0.16$, Pearson), and we therefore assume that data set size does not bias our results. To approximate the number of different species in a sample, we used Margalef's richness index (D_{mg}), which attenuates sampling size effects (equation 3) (71), where S is the number of different OTUs observed in the sample and N is the number of 16S reads mapped:

$$D_{mg} = \frac{(S-1)}{\ln N} \quad (3)$$

The microbiota evenness was calculated using Heip's evenness measure (E) (equation 4) (72), where H is the Shannon-Weiner entropy of the 16S reads and S is the number of OTUs in the data set:

$$E = \frac{(e^H - 1)}{(S - 1)} \quad (4)$$

For the comparison between the alpha diversity and the number of heterozygote genes, only the fully classified data sets were used (Table S2). In the analysis of the effect of homozygous HLA genes on alpha diversity, all data sets with at least one homozygous gene were used.

Beta diversity. The microbiota similarity between individuals was calculated using two beta diversity measures: weighted UniFrac distance and Spearman's rank correlation (73). We adapted the SILVA taxonomy with a branch length of 1 between taxonomic ranks and used these distances for weighted UniFrac. The Spearman rank correlation was calculated using the abundances of the OTUs observed in each microbiome.

Statistics. We used the Kolmogorov-Smirnov (KS) test for two samples from the python package `scipy.stats` (version 1.3.2) (74) for statistical analysis, with alternative hypothesis, “two-sided,” “less,” and “greater,” depending on the relevance for the comparison, as stated in the figure legends.

Code availability. Scripts for downloading the data sets, mapping the data sets to the composed reference genome, and further analysis can be accessed on GitHub (github.com/Stijn-A/HLA_microbiome).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

DATA SET S1, XLSX file, 17.8 MB.

FIG S1, TIF file, 0.5 MB.

FIG S2, TIF file, 1.1 MB.

FIG S3, TIF file, 1.1 MB.

FIG S4, TIF file, 1 MB.

FIG S5, TIF file, 1 MB.

FIG S6, TIF file, 1.5 MB.

FIG S7, TIF file, 0.8 MB.

TABLE S1, XLSX file, 0.01 MB.

TABLE S2, XLSX file, 0.04 MB.

ACKNOWLEDGMENTS

We thank R. E. Bontrop and Jelle Slager for helpful discussions.

B.E.D. was supported by the Netherlands Organization for Scientific Research (NWO) Vidi grant 864.14.004 and by the European Research Council (ERC) Consolidator grant 865694: DiversiPHI.

REFERENCES

- Kamada N, Chen GY, Inohara N, Núñez G. 2013. Control of pathogens and pathobionts by the gut microbiota. *Nat Immunol* 14:685–690. <https://doi.org/10.1038/ni.2608>.
- Lawley TD, Walker AW. 2013. Intestinal colonization resistance. *Immunology* 138:1–11. <https://doi.org/10.1111/j.1365-2567.2012.03616.x>.
- Hill DA, Artis D. 2010. Intestinal bacteria and the regulation of immune cell homeostasis. *Annu Rev Immunol* 28:623–667. <https://doi.org/10.1146/annurev-immunol-030409-101330>.
- Renz H, Brandtzaeg P, Hornef M. 2011. The impact of perinatal immune development on mucosal homeostasis and chronic inflammation. *Nat Rev Immunol* 12:9–23. <https://doi.org/10.1038/nri3112>.
- Sonnenberg GF, Artis D. 2012. Innate lymphoid cell interactions with microbiota: implications for intestinal health and disease. *Immunity* 37: 601–610. <https://doi.org/10.1016/j.immuni.2012.10.003>.
- Sender R, Fuchs S, Milo R. 2016. Revised estimates for the number of human and bacteria cells in the body. *PLoS Biol* 14:e1002533. <https://doi.org/10.1371/journal.pbio.1002533>.
- The Human Microbiome Project Consortium. 2012. Structure, function and diversity of the healthy human microbiome. *Nature* 486:207–214. <https://doi.org/10.1038/nature11234>.
- Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA. 2005. Diversity of the human intestinal microbial flora. *Science* 308:1635–1638. <https://doi.org/10.1126/science.1110591>.
- Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, Kurilshikov A, Bonder MJ, Valles-Colomer M, Van deputte D, Tito RY, Chaffron S, Rymenans L, Verspecht C, De Sutter L, Lima-Mendez G, D'hoel K, Jonckheere K, Homola D, Garcia R, Tigchelaar EF, Eeckhaudt L, Fu J, Henckaerts L, Zhernakova A, Wijmenga C, Raes J. 2016. Population-level analysis of gut microbiome variation. *Science* 352:560–564. <https://doi.org/10.1126/science.aad3503>.
- Zhernakova A, Kurilshikov A, Bonder MJ, Tigchelaar EF, Schirmer M, Vatanen T, Mujagic Z, Vila AV, Falony G, Vieira-Silva S, Wang J, Imhann F, Brandsma E, Jankipersadsing SA, Joossens M, Cenit MC, Deelen P, Swertz MA, Weersma RK, Feskens EJM, Netea MG, Gevers D, Jonkers D, Franke L, Aulchenko YS, Huttenhower C, Raes J, Hofker MH, Xavier RJ, Wijmenga C, Fu J, LifeLines cohort study. 2016. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science* 352:565–569. <https://doi.org/10.1126/science.aad3369>.
- Belkaid Y, Hand TW. 2014. Role of the microbiota in immunity and inflammation. *Cell* 157:121–141. <https://doi.org/10.1016/j.cell.2014.03.011>.
- Bonder MJ, Kurilshikov A, Tigchelaar EF, Mujagic Z, Imhann F, Vila AV, Deelen P, Vatanen T, Schirmer M, Smekens SP, Zhernakova DV, Jankipersadsing SA, Jaeger M, Oosting M, Cenit MC, Masclee AAM, Swertz MA, Li Y, Kumar V, Joosten L, Harmsen H, Weersma RK, Franke L, Hofker MH, Xavier RJ, Jonkers D, Netea MG, Wijmenga C, Fu J, Zhernakova A. 2016. The effect of host genetics on the gut microbiome. *Nat Genet* 48:1407–1412. <https://doi.org/10.1038/ng.3663>.
- Goodrich JK, Davenport ER, Beaumont M, Jackson MA, Knight R, Ober C, Spector TD, Bell JT, Clark AG, Ley RE. 2016. Genetic determinants of the gut microbiome in UK twins. *Cell Host Microbe* 19:731–743. <https://doi.org/10.1016/j.chom.2016.04.017>.
- Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, Beaumont M, Van Treuren W, Knight R, Bell JT, Spector TD, Clark AG, Ley RE. 2014. Human genetics shape the gut microbiome. *Cell* 159:789–799. <https://doi.org/10.1016/j.cell.2014.09.053>.
- Kolde R, Franzosa EA, Rahnava G, Hall AB, Vlamakis H, Stevens C, Daly MJ, Xavier RJ, Huttenhower C. 2018. Host genetic variation and its microbiome interactions within the Human Microbiome Project. *Genome Med* 10:6. <https://doi.org/10.1186/s13073-018-0515-8>.
- Zhou S, Xu R, He F, Zhou J, Wang Y, Zhou J, Wang M, Zhou W. 2016. Diversity of gut microbiota metabolic pathways in 10 pairs of Chinese infant twins. *PLoS One* 11:e0161627. <https://doi.org/10.1371/journal.pone.0161627>.
- Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, Gordon JL. 2010. Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* 466:334–338. <https://doi.org/10.1038/nature09199>.
- Khan AA, Yurkovskiy L, O'Grady K, Pickard JM, de Pooter R, Antonopoulos DA, Golovkina T, Chervonsky A. 2019. Polymorphic immune mechanisms regulate commensal repertoire. *Cell Rep* 29:541–550.E4. <https://doi.org/10.1016/j.celrep.2019.09.010>.
- van Baarlen P, Troost FJ, van Hemert S, van der Meer C, de Vos WM, de Groot PJ, Hooiveld GJEJ, Brummer R-JM, Kleerebezem M. 2009. Differential NF- κ B pathways induction by *Lactobacillus plantarum* in the duodenum of healthy humans correlating with immune tolerance. *Proc Natl Acad Sci U S A* 106:2371–2376. <https://doi.org/10.1073/pnas.0809919106>.
- Mao K, Baptista AP, Tamoutounour S, Zhuang L, Bouladoux N, Martins AJ, Huang Y, Gerner MY, Belkaid Y, Germain RN. 2018. Innate and adaptive lymphocytes sequentially shape the gut microbiota and lipid metabolism. *Nature* 554:255–259. <https://doi.org/10.1038/nature25437>.

21. Schluter J, Peled JU, Taylor BP, Markey KA, Smith M, Taur Y, Niehus R, Staffas A, Dai A, Fontana E, Amoretti LA, Wright RJ, Morjaria S, Fenelus M, Pessin MS, Chao NJ, Lew M, Bohannon L, Bush A, Sung AD, Hohl TM, Perales MA, van den Brink MRM, Xavier JB. 2020. The gut microbiota is associated with immune cell dynamics in humans. *Nature* 588:303–307. <https://doi.org/10.1038/s41586-020-2971-8>.
22. Constantinides MG, Link VM, Tamoutounour S, Wong AC, Perez-Chaparro PJ, Han S-J, Chen YE, Li K, Farhat S, Weckel A, Krishnamurthy SR, Vujkovic-Cvijin I, Linehan JL, Bouladoux N, Merrill ED, Roy S, Cua DJ, Adams EJ, Bhandoola A, Scharschmidt TC, Aubé J, Fischbach MA, Belkaid Y. 2019. MAIT cells are imprinted by the microbiota in early life and promote tissue repair. *Science* 366:eaax6624. <https://doi.org/10.1126/science.aax6624>.
23. Mungall AJ, Palmer SA, Sims SK, Edwards CA, Ashurst JL, Wilming L, Jones MC, Horton R, Hunt SE, Scott CE, Gilbert JGR, Clamp ME, Bethel G, Milne S, Ainscough R, Almeida JP, Ambrose KD, Andrews TD, Ashwell RIS, Babbage AK, Bagguley CL, Bailey J, Banerjee R, Barker DJ, Barlow KF, Bates K, Beare DM, Beasley H, Beasley O, Bird CP, Blakey S, Bray-Allen S, Brook J, Brown AJ, Brown JY, Burford DC, Burrill W, Burton J, Carder C, Carter NP, Chapman JC, Clark SY, Clark G, Clee CM, Clegg S, Copley V, Collier RE, Collins JE, Colman LK, Corby NR, Coville GJ, Culley KM, Dhami P, Davies J, Dunn M, Earthrohl ME, Ellington AE, et al. 2003. The DNA sequence and analysis of human chromosome 6. *Nature* 425:805–811. <https://doi.org/10.1038/nature02055>.
24. Robinson J, Halliwell JA, Hayhurst JD, Flicek P, Parham P, Marsh SGE. 2015. The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Res* 43:D423–D431. <https://doi.org/10.1093/nar/gku1161>.
25. Rock KL, Reits E, Neefjes J. 2016. Present yourself! By MHC class I and MHC class II molecules. *Trends Immunol* 37:724–737. <https://doi.org/10.1016/j.it.2016.08.010>.
26. Joffre OP, Segura E, Savina A, Amigorena S. 2012. Cross-presentation by dendritic cells. *Nat Rev Immunol* 12:557–569. <https://doi.org/10.1038/nri3254>.
27. Heath WR, Carbone FR. 2001. Cross-presentation in viral immunity and self-tolerance. *Nat Rev Immunol* 1:126–134. <https://doi.org/10.1038/35100512>.
28. Kubinak JL, Stephens WZ, Soto R, Petersen C, Chiaro T, Gogokhia L, Bell R, Ajami NJ, Petrosino JF, Morrison L, Potts WK, Jensen PE, O'Connell RM, Round JL. 2015. MHC variation sculpts individualized microbial communities that control susceptibility to enteric infection. *Nat Commun* 6:8642. <https://doi.org/10.1038/ncomms9642>.
29. Lin P, Bach M, Asquith M, Lee AY, Akileswaran L, Stauffer P, Davin S, Pan Y, Cambronne ED, Dorris M, Debelius JW, Lauber CL, Ackermann G, Baeza YV, Gill T, Knight R, Colbert RA, Taugro JD, Van Gelder RN, Rosenbaum JT. 2014. HLA-B*27 and human β 2-microglobulin affect the gut microbiota of transgenic rats. *PLoS One* 9:e105684. <https://doi.org/10.1371/journal.pone.0105684>.
30. Toivanen P, Vaahtovuori J, Eerola E. 2001. Influence of major histocompatibility complex on bacterial composition of fecal flora. *Infect Immun* 69:2372–2377. <https://doi.org/10.1128/IAI.69.4.2372-2377.2001>.
31. Bolnick DI, Snowberg LK, Caporaso JG, Lauber C, Knight R, Stutz WE. 2014. Major histocompatibility complex class IIb polymorphism influences gut microbiota composition and diversity. *Mol Ecol* 23:4831–4845. <https://doi.org/10.1111/mec.12846>.
32. Tian C, Hromatka BS, Kiefer AK, Eriksson N, Noble SM, Tung JY, Hinds DA. 2017. Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. *Nat Commun* 8:599. <https://doi.org/10.1038/s41467-017-00257-5>.
33. De Palma G, Capilla A, Nadal I, Nova E, Pozo T, Varea V, Polanco I, Castillejo G, Lopez A, Garrote JA, Calvo C, Garcia-Novo MD, Cilleruelo ML, Ribes-Koninckx C, Palau F, Sanz Y. 2010. Interplay between human leukocyte antigen genes and the microbial colonization process of the newborn intestine. *Curr Issues Mol Biol* 12:1–10.
34. Olivares M, Neef A, Castillejo G, Palma GD, Varea V, Capilla A, Palau F, Nova E, Marcos A, Polanco I, Ribes-Koninckx C, Ortigosa L, Izquierdo L, Sanz Y. 2015. The HLA-DQ2 genotype selects for early intestinal microbiota composition in infants at high risk of developing coeliac disease. *Gut* 64:406–417. <https://doi.org/10.1136/gutjnl-2014-306931>.
35. Kurilshikov A, Medina-Gomez C, Bacigalupe R, Radjabzadeh D, Wang J, Demirkan A, Le Roy CI, Raygoza Garay JA, Finnicum CT, Liu X, Zhernakova DV, Bonder MJ, Hansen TH, Frost F, Rühlemann MC, Turpin W, Moon JY, Kim HN, Lüll K, Barkan E, Shah SA, Fornage M, Szopinska-Tokov J, Wallen ZD, Borisevich D, Agreus L, Andreasson A, Bang C, Bedrani L, Bell JT, Bisgaard H, Boehnke M, Boomsma DI, Burk RD, Claringbould A, Croitoru K, Davies GE, van Duijn CM, Duijts L, Falony G, Fu J, van der Graaf A, Hansen T, Homuth G, Hughes DA, Ijzerman RG, Jackson MA, Jaddoe VVV, Joossens M, Jørgensen T, et al. 2021. Large-scale association analyses identify host factors influencing human gut microbiome composition. *Nat Genet* 53:156–165. <https://doi.org/10.1038/s41588-020-00763-1>.
36. Rühlemann MC, Hermes BM, Bang C, Doms S, Moitinho-Silva L, Thingholm LB, Frost F, Degenhardt F, Wittig M, Kässens J, Weiss FU, Peters A, Neuhaus K, Völker U, Völzke H, Homuth G, Weiss S, Grallert H, Laudes M, Lieb W, Haller D, Lerch MM, Baines JF, Franke A. 2021. Genome-wide association study in 8,956 German individuals identifies influence of ABO histo-blood groups on gut microbiome. *Nat Genet* 53:147–155. <https://doi.org/10.1038/s41588-020-00747-1>.
37. McDermott AJ, Huffnagle GB. 2014. The microbiome and regulation of mucosal immunity. *Immunology* 142:24–31. <https://doi.org/10.1111/imm.12231>.
38. Wosen JE, Mukhopadhyay D, Macaubas C, Mellins ED. 2018. Epithelial MHC class II expression and its role in antigen presentation in the gastrointestinal and respiratory tracts. *Front Immunol* 9:2144. <https://doi.org/10.3389/fimmu.2018.02144>.
39. Handunnetthi L, Ramagopalan SV, Ebers GC, Knight JC. 2010. Regulation of major histocompatibility complex class II gene expression, genetic variation and disease. *Genes Immun* 11:99–112. <https://doi.org/10.1038/gene.2009.83>.
40. Johnson DR. 2000. Differential expression of human major histocompatibility class I loci: HLA-A, -B, and -C. *Hum Immunol* 61:389–396. [https://doi.org/10.1016/S0198-8859\(99\)00186-X](https://doi.org/10.1016/S0198-8859(99)00186-X).
41. Maier M, Gragerl L, Klitz W. 2007. High-resolution HLA alleles and haplotypes in the United States population. *Hum Immunol* 68:779–788. <https://doi.org/10.1016/j.humimm.2007.04.005>.
42. Sidney J, Peters B, Frahm N, Brander C, Sette A. 2008. HLA class I super-types: a revised and updated classification. *BMC Immunol* 9:1. <https://doi.org/10.1186/1471-2172-9-1>.
43. van Deutekom HWM, Keşmir C. 2015. Zooming into the binding groove of HLA molecules: which positions and which substitutions change peptide binding most? *Immunogenetics* 67:425–436. <https://doi.org/10.1007/s00251-015-0849-y>.
44. Jurtz V, Paul S, Andreatta M, Marcattili P, Peters B, Nielsen M. 2017. NetMHCpan-4.0: improved peptide–MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J Immunol* 199:3360–3368. <https://doi.org/10.4049/jimmunol.1700893>.
45. Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S, Arumugam M, Kultima JR, Prifti E, Nielsen T, Juncker AS, Manichanh C, Chen B, Zhang W, Levenez F, Wang J, Xu X, Xiao L, Liang S, Zhang D, Zhang Z, Chen W, Zhao H, Al-Aama JY, Edris S, Yang H, Wang J, Hansen T, Nielsen HB, Brunak S, Kristiansen K, Guarner F, Pedersen O, Doré J, Ehrlich SD, Bork P, Wang J, Le Chatelier E, Batto J-M, Kennedy S, Haimet F, Winogradski Y, Pelletier E, LePaslier D, Artiguenave F, Bruls T, Weissenbach J, Turner K, Parkhill J, Antolin M, MetaHIT Consortium, et al. 2014. An integrated catalog of reference genes in the human gut microbiome. *Nat Biotechnol* 32:834–841. <https://doi.org/10.1038/nbt.2942>.
46. Bihl F, Frahm N, Di Giammarino L, Sidney J, John M, Yusim K, Woodberry T, Sango K, Hewitt HS, Henry L, Linde CH, Chisholm JV, Zaman TM, Pae E, Mallal S, Walker BD, Sette A, Korber BT, Heckerman D, Brander C. 2006. Impact of HLA-B alleles, epitope binding affinity, functional avidity, and viral coinfection on the immunodominance of virus-specific CTL responses. *J Immunol* 176:4094–4101. <https://doi.org/10.4049/jimmunol.176.7.4094>.
47. Lewinsohn DA, Winata E, Swarbrick GM, Tanner KE, Cook MS, Null MD, Cansler ME, Sette A, Sidney J, Lewinsohn DM. 2007. Immunodominant tuberculosis CD8 antigens preferentially restricted by HLA-B. *PLoS Pathog* 3:e127. <https://doi.org/10.1371/journal.ppat.0030127>.
48. Weiskopf D, Yauch LE, Angelo MA, John DV, Greenbaum JA, Sidney J, Kolla RV, De Silva AD, de Silva AM, Grey H, Peters B, Shrestha S, Sette A. 2011. Insights into HLA-restricted T cell responses in a novel mouse model of dengue virus infection point toward new implications for vaccine design. *J Immunol* 187:4268–4279. <https://doi.org/10.4049/jimmunol.1101970>.
49. Lacey SF, Villacres MC, La Rosa C, Wang Z, Longmate J, Martinez J, Brewer JC, Mekhoubad S, Maas R, Leedom JM, Forman SJ, Zaia JA, Diamond DJ. 2003. Relative dominance of HLA-B*07 restricted CD8+ T-lymphocyte immune responses to human cytomegalovirus pp65 in persons sharing HLA-A*02 and HLA-B*07 alleles. *Hum Immunol* 64:440–452. [https://doi.org/10.1016/S0198-8859\(03\)00028-4](https://doi.org/10.1016/S0198-8859(03)00028-4).
50. Houde M, Bertholet S, Gagnon E, Brunet S, Goyette G, Laplante A, Princiotta MF, Thibault P, Sacks D, Desjardins M. 2003. Phagosomes are competent organelles for antigen cross-presentation. *Nature* 425:402–406. <https://doi.org/10.1038/nature01912>.
51. Kamphorst AO, Guerinonprez P, Dudziak D, Nussenzweig MC. 2010. Route of antigen uptake differentially impacts presentation by dendritic cells

- and activated monocytes. *J Immunol* 185:3426–3435. <https://doi.org/10.4049/jimmunol.1001205>.
52. Groh V, Bahram S, Bauer S, Herman A, Beauchamp M, Spies T. 1996. Cell stress-regulated human major histocompatibility complex class I gene expressed in gastrointestinal epithelium. *Proc Natl Acad Sci U S A* 93:12445–12450. <https://doi.org/10.1073/pnas.93.22.12445>.
 53. Bauer S, Groh V, Wu J, Steinle A, Phillips JH, Lanier LL, Spies T. 1999. Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MIC6. *Science* 285:727–729. <https://doi.org/10.1126/science.285.5428.727>.
 54. Ando H, Mizuki N, Ota M, Yamazaki M, Ohno S, Goto K, Miyata Y, Wakisaka K, Bahram S, Inoko H. 1997. Allelic variants of the human MHC class I chain-related B gene (MICB). *Immunogenetics* 46:499–508. <https://doi.org/10.1007/s002510050311>.
 55. Petersdorf EW, Shuler KB, Longton GM, Spies T, Hansen JA. 1999. Population study of allelic diversity in the human MHC class I-related MIC-A gene. *Immunogenetics* 49:605–612. <https://doi.org/10.1007/s002510050655>.
 56. Panagioti E, Redeker A, van Duikeren S, Franken KLMC, Drijfhout JW, van der Burg SH, Arens R. 2016. The breadth of synthetic long peptide vaccine-induced CD8+ T cell responses determines the efficacy against mouse cytomegalovirus infection. *PLoS Pathog* 12:e1005895. <https://doi.org/10.1371/journal.ppat.1005895>.
 57. Costello EK, Stagaman K, Dethlefsen L, Bohannan BJM, Relman DA. 2012. The application of ecological theory toward an understanding of the human microbiome. *Science* 336:1255–1262. <https://doi.org/10.1126/science.1224203>.
 58. Greenbaum J, Sidney J, Chung J, Brander C, Peters B, Sette A. 2011. Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes. *Immunogenetics* 63:325–335. <https://doi.org/10.1007/s00251-011-0513-0>.
 59. Leinonen R, Sugawara H, Shumway M, International Nucleotide Sequence Database Collaboration. 2011. The sequence read archive. *Nucleic Acids Res* 39:D19–D21. <https://doi.org/10.1093/nar/gkq1019>.
 60. Andrews S. 1973. FastQC: a quality control tool for high throughput sequence data. *Soil* 5:47–81.
 61. Schubert M, Lindgreen S, Orlando L. 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes* 9:88. <https://doi.org/10.1186/s13104-016-1900-2>.
 62. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
 63. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41:D590–D596. <https://doi.org/10.1093/nar/gks1219>.
 64. Clausen PTL, Aarestrup FM, Lund O. 2018. Rapid and precise alignment of raw reads against redundant databases with KMA. *BMC Bioinformatics* 19:307. <https://doi.org/10.1186/s12859-018-2336-6>.
 65. Coutinho FH, Silveira CB, Gregoracci GB, Thompson CC, Edwards RA, Brussaard CPD, Dutilh BE, Thompson FL. 2017. Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. *Nat Commun* 8:15955. <https://doi.org/10.1038/ncomms15955>.
 66. Iverson V, Morris RM, Frazar CD, Berthiaume CT, Morales RL, Armbrust EV. 2012. Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. *Science* 335:587–590. <https://doi.org/10.1126/science.1212665>.
 67. Hiller C, Bischoff M, Schmidt A, Bender K. 1978. Analysis of the HLA-ABC linkage disequilibrium: decreasing strength of gametic association with increasing map distance. *Hum Genet* 41:301–312. <https://doi.org/10.1007/BF00284764>.
 68. Jeffreys AJ, Kauppi L, Neumann R. 2001. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat Genet* 29:217–222. <https://doi.org/10.1038/ng1001-217>.
 69. scikit-bio Development Team. 2014. Scikit-bio diversity.
 70. McMurdie PJ, Holmes S. 2014. Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput Biol* 10:e1003531. <https://doi.org/10.1371/journal.pcbi.1003531>.
 71. Magurran AE. 2004. Measuring biological diversity, p 76–77. Blackwell Publishing, Hoboken, NJ.
 72. Heip C. 1974. A new index measuring evenness. *J Mar Biol Assoc* 54:555–557. <https://doi.org/10.1017/S0025315400022736>.
 73. Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. 2011. UniFrac: an effective distance metric for microbial community comparison. *ISME J* 5:169–172. <https://doi.org/10.1038/ismej.2010.133>.
 74. scipy.stats. 2019. scipy.stats.ks_2samp—SciPy v1.3.2 reference guide.