

GIMICA: host genetic and immune factors shaping human microbiota

Jing Tang^{1,2,†}, Xianglu Wu^{3,†}, Minjie Mou^{1,†}, Chuan Wang², Lidan Wang², Fengcheng Li¹, Maiyuan Guo², Jiayi Yin¹, Wenqin Xie², Xiaona Wang⁴, Yingxiong Wang^{2,3}, Yubin Ding^{3,*}, Weiwei Xue^{④4,*} and Feng Zhu^{④1,*}

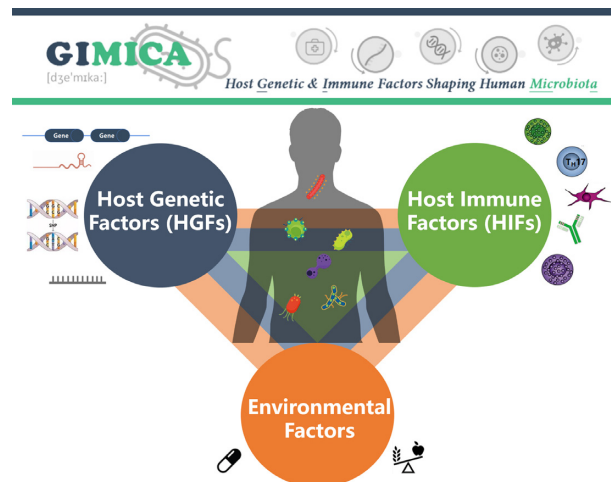
¹College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China, ²College of Basic Medicine, Chongqing Medical University, Chongqing 400016, China, ³Joint International Research Lab of Reproductive and Development, Department of Reproductive Biology, School of Public Health, Chongqing Medical University, Chongqing 400016, China and ⁴School of Pharmaceutical Sciences, Chongqing University, Chongqing 401331, China

Received August 14, 2020; Revised September 09, 2020; Editorial Decision September 18, 2020; Accepted October 08, 2020

ABSTRACT

Besides the environmental factors having tremendous impacts on the composition of microbial community, the host factors have recently gained extensive attentions on their roles in shaping human microbiota. There are two major types of host factors: host genetic factors (HGFs) and host immune factors (HIFs). These factors of each type are essential for defining the chemical and physical landscapes inhabited by microbiota, and the collective consideration of both types have great implication to serve comprehensive health management. However, no database was available to provide the comprehensive factors of both types. Herein, a database entitled ‘Host Genetic and Immune Factors Shaping Human Microbiota (GIMICA)’ was constructed. Based on the 4257 microbes confirmed to inhabit nine sites of human body, 2851 HGFs (1368 single nucleotide polymorphisms (SNPs), 186 copy number variations (CNVs), and 1297 non-coding ribonucleic acids (RNAs)) modulating the expression of 370 microbes were collected, and 549 HIFs (126 lymphocytes and phagocytes, 387 immune proteins, and 36 immune pathways) regulating the abundance of 455 microbes were also provided. All in all, GIMICA enabled the collective consideration not only between different types of host factor but also between the host and environmental ones, which is freely accessible without login requirement at: <https://idrblab.org/gimica/>

GRAPHICAL ABSTRACT



INTRODUCTION

The microbial communities (MCs) in different sites of human body are known to be instrumental in host physiology and health (1–4). The composition of these MCs has been generally considered as susceptible to a variety of environmental factors (5–7) (including: diet, lifestyle, disease, etc.), which thus mediates changes in host metabolic status (5) and regulates the development of chronic disease (6–9). Besides these environmental ones, the host factor has recently emerged as another crucial determinant that shapes human microbiota (10–12). Particularly, different from environmental ones, host factors are able to define the chemical and physical landscape (including nutrient availability and activity threshold for immune system) inhabited by host

*To whom correspondence should be addressed. Tel: +86 189 8946 6518; Fax: +86 571 8820 8444; Email: zhufeng@zju.edu.cn

Correspondence may also be addressed to Weiwei Xue. Email: xuww@cqu.edu.cn

Correspondence may also be addressed to Yubin Ding. Email: dingyb@cqmu.edu.cn

†The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

microbiota (13–18), which therefore determine the heritable nature of microbiome (12) and immune system activity (13,19).

There are two popular host factors that are frequently considered in current MC studies (19–23): (i) *host genetic factors* (HGFs, including: single nucleotide polymorphism (SNP), gene copy number variation (CNV) and non-coding ribonucleic acid (RNA) regulation) are crucial in exploring the heritability of microbiota compositions (24), revealing the mechanisms underlying the adaptation to environmental fluctuations (11), and discovering promising drug target (25) and targeted therapeutics (26,27); (ii) *host immune factors* (HIFs, including: checkpoint molecule, cytokine, antibody and lymphocyte) are critical for maintaining the symbiotic balance between host and microbiota (28), promoting the activities of immune system (13,20), and boosting the efficacy of immunotherapy (29). Since the composition of human MCs is greatly affected by their inhabited chemical/physical landscapes (13), the accumulation of human factors that define such landscapes can give significant insights into current MC studies.

So far, a variety of human MCs-related databases have been constructed and currently active, the majority of which focus on providing sequencing/metagenomic data (dbCAN-seq (30), eHOMD (31), gcMeta (32), GMrepo (33), HPMCD (34), MAHMI (35), MBGD (36), MetaRef (37), MDB (38)), while some others are specialized in containing the data of either metabolites/metabolic reactions (HMDB (39), mVOC (40), and VMH (41)) or taxonomic profiling (MMDB (42)). Only three databases are available now to describe the environmental factors affecting the compositions of MCs: Disbiome (43) linking the microbiome to disease indication, gutMDisorder (44) collecting the immense data of diet, drug and lifestyle that lead to microbial dysbiosis and aBiofilm (45) providing some environmental substances that inhibit bacteria. However, no database has yet been designed to provide the host factors shaping human microbiota.

Herein, a new database of host genetic and immune factors shaping human microbiota (GIMICA) was therefore introduced. First, a comprehensive literature review on human MC was conducted. Thousands of microbe species were discovered as inhabiting human blood, gut, lung, mucosa, nose, ocular surface, oral cavity, skin, and vagina. These microbe species were very diverse and belonged to >50 taxonomic phyla (such as: *Firmicutes*, *Actinobacteria* and *Ascomycota*). Second, based on these microbes, the HGFs and HIFs for each species were collected via literature review, and the host data of both genetic and immune factors were then provided in the GIMICA database. Third, the crosslinks of each microbe to other MCs-related knowledge bases, especially the ones describing the environmental factors affecting the composition of MCs, were systematically discovered. These crosslinks were essential because they allowed, for the first time, the collective considerations not only among different host factors but also between the host and environmental ones. All in all, the GIMICA comprehensively provided the host genetic and immune factors that shape human microbiota (as illustrated in Figure 1), which allowed the collective considerations among the different types of factor (between HGF and HIF; between

host and environmental factors). Since the composition of MCs should be considered as a complex shaped by the combination of both host and environmental factors (46–48), the GIMICA is expected to have implication for the future practice of MCs-related studies on human physiology and health (1–3,49–53).

FACTUAL CONTENT AND DATA RETRIEVAL

Collecting and confirming microbes inhabiting human

The data of microbe species inhabiting human body were first collected by literature reviews in PubMed (54) using the keyword combinations of: ‘microbial community + human’, ‘microbe + human’, ‘microbiota + human’, etc. Second, several MCs-related databases (such as: eHOMD (31), GMrepo (33), VMH (41) and gutMDisorder (44)) were reviewed, and hundreds of microbes in these databases were collected for further assessing their suitability to be included in GIMICA. In total, 6438 microbes were identified based on the researches in these steps above. Third, the names of all these identified microbes were standardized by searching NCBI Taxonomy database (55), which resulted in 5386 microbe species with well-defined Taxonomy ID. A repository that contained the diverse synonyms of each species were further constructed using NCBI Taxonomy (55) data. Fourth, for those microbes collected from MCs-related databases, literature search in PubMed (54) was performed based on the keyword combinations of ‘Microbe Name’ + human’, ‘Microbe Synonym’ + human’, etc. During this literature review and the review in the first step, the body sites in human, where each microbe species inhabited, were carefully identified and systematically recorded. Consequently, 4257 microbes were confirmed to inhabit nine sites of human body. Particularly, 281, 2436, 513, 225, 197, 164, 468, 1012 and 276 microbe species (bacteria, fungi, metamonada, amoebozoa and archaea (56–58)) were found to colonize in human blood, gut, lung, mucosa, nose, ocular surface, oral cavity, skin and vagina, respectively (59–63). These microbes are diverse and belong to 58 taxonomic phyla (such as *Firmicutes*, *Actinobacteria*, *Ascomycota*, *Euryarchaeota*, *Bacteroidete*, *Proteobacteria* and *Tenericutes*). Finally, the affiliated information for each human microbe were systematically collected using literature review, which contained taxonomic lineage (kingdom, phylum, class, order, family, genus, and species), oxygen dependency/tolerance (aerobe, anaerobe, aerotolerant and aerophilic), mechanisms of microbial metabolism (such as catalytic, fermentative, oxidative, proteolytic, respiratory and saccharolytic), sites of human body inhabited (such as gut, oral cavity and skin), Gram-staining classifications (negative, positive and variable), impact on/relation to host (probiotic, commensal, opportunistic and pathogen) and other general descriptions (disease-relevance, metabolic product, mechanism of symbiosis, the number of genes/transcripts and genome size).

Host genetic factors shaping human microbiota

The host genetic factors that shape human microbiota are characterized by a variety of regulatory types, which contained single nucleotide polymorphisms (upstream or

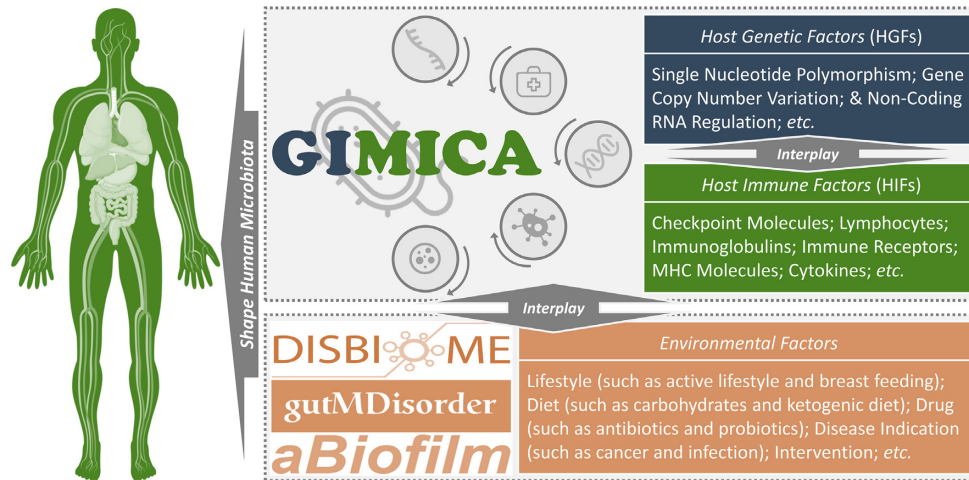


Figure 1. GIMICA comprehensively described host genetic and immune factors shaping human microbiota, which allowed the collective consideration among different types of factors (between host genetic factors and host immune factors). By providing the crosslinks to available databases (Disbiome (43), gutMDisorder (44) and aBiofilm (45)) with environmental factors information, GIMICA further enabled the collective consideration between the host and environmental factors.

downstream gene variants, untranslated region variants, intron variants, etc.), copy number variations (duplication, deletion, etc.), and non-coding RNA (lncRNA, rRNA, miRNA, miscRNA, sRNA, pseudogene, etc.). Such factors were essential in exploring the heritability of microbiota compositions (24), revealing the mechanism underlying the adaptation to environmental fluctuations (11), and discovering the drug target and targeted therapeutics (26). Because the HGFs data were largely dispersed in literatures, the PubMed (54) was systematically searched to discover the regulation of human microbiota by HGFs. Particularly, keyword combinations of ‘genetic variant + ‘Microbe Name’, ‘host genetic factor + ‘Microbe Name’, ‘genetic variation + ‘Microbe Name’, ‘gene copy number variations + ‘Microbe Name’, ‘CNV + ‘Microbe Name’, ‘nucleotide polymorphism + ‘Microbe Name’, and ‘SNP + ‘Microbe Name’ were adopted for literature reviews, and the resulting publications were manually assessed for retrieving any HGFs-related information.

As a result, the collected data included the various regulatory types of host genetic factors (single nucleotide polymorphisms, copy number variations and non-coding RNAs), the detail sub-types of each HGF type (such as upstream variant, gene duplication and lncRNA), the genetic sequence, genome location and strand type for each HGF, and the impacts of each HGF on studied microbe species (up/down-regulation of microbe abundance, promotion/inhibition of the gene expression in microbes). As a result, current version of GIMICA had 1,368 single nucleotide polymorphisms (590 intron, 74 missense, 54 synonymous variants, 43 upstream gene, 26 non coding transcript variant, 17 UTR, 7 downstream gene, etc.) regulating the relative abundances of 267 microbes that primarily inhabit host gut, mucosa, and vagina; 186 copy number variations (37 duplications and 149 deletions) modulating the expressions of 156 microbes that live mainly in human oral cavity, gut, lung, skin, blood, and mucosa; 1297 non-coding RNA regulations (514 miRNAs, 683 lncRNAs, 14 sRNAs, 9 pseudogenes, etc.) altering the gene expression of 251 mi-

crobe species that reside mostly in host lung, oral cavity, gut and skin. All these HGFs data can be assessed and retrieved using various search strategies in both the Home page and the subpage (which is entitled as ‘Host Genetic Factors’) of the GIMICA database. The example host genetic factors in GIMICA were illustrated in the Figure 2.

Host immune factors shaping human microbiota

The host immune factors that shape human microbiota contained a number of immune molecules and cells regulating microbiota compositions (such as checkpoint molecule, cytokine, complement component, eosinophils, immunoglobulin, innate lymphoid cell, macrophage, monocyte, natural killer cell, retinoic acid receptor, toll-like receptor, etc.). These factors are critical for maintaining symbiotic balance between host and microbiota (28), promoting the activities of immune system (13), and boosting the efficacy of immunotherapy (29). Similar to the collection of the HGF data, the HIF data were identified by literature search based on different keyword combinations, which included ‘host immune factors + ‘Microbe Name’, ‘immune cell + ‘Microbe Name’, ‘immune system + ‘Microbe Name’, ‘immunity + ‘Microbe Name’, ‘checkpoint molecules + ‘Microbe Name’, ‘complement component + ‘Microbe Name’, etc. The resulting publications were also manually assessed for retrieving any HIF-related information.

Consequently, the collected data included various regulatory types of host immune factors (HIF), such as checkpoint molecules, cytokines, complement component, eosinophils, immunoglobulin, etc. Moreover, other critical data were also collected and provided, which included the names of immune-related molecules/cells, the sequence, protein family, signaling pathway, and molecular function for each HIF, and the impacts of each HIF on a studied microbe (up/down-regulation of microbe abundance, promotion/inhibition of the gene expression in microbe). All in all, GIMICA contained 126 lymphocytes (including helper/cytotoxic/regulatory T-cells, plasma/regulatory

Host Genetic Factors (HGFs)				
<input type="checkbox"/> VIPR2				
HGF ID	HGF2350	HGF Info	Class	Copy Number Variation: Gene Deletion (CNV-GDe)
Description	The deletion of VIPR2 has been associated with the increased abundance of <i>Escherichia</i> .			[9]
<input type="checkbox"/> MUC2				
HGF ID	HGF2331	HGF Info	Class	Copy Number Variation: Gene Deletion (CNV-GDe)
Description	The deletion of MUC2 was significantly increased the abundance of <i>Escherichia coli</i> (p-value<0.05).			[10]
<input type="checkbox"/> C4B				
HGF ID	HGF2348	HGF Info	Class	Copy Number Variation: Gene Duplication (CNV-GDu)
Description	The low C4B-CN has been significantly associated with the decreased abundance of <i>Escherichia coli</i> (p-value<0.05).			[15]
<input type="checkbox"/> hsa-miR-1226-5p				
HGF ID	HGF2670	HGF Info	Class	Non-coding RNA: Micro (ncRNA-miRNA)
Description	The hsa-miR-1226-5p miRNA in human could increase the mRNA expression of <i>yegH</i> in the facultative anaerobic <i>Escherichia coli</i> and promote the growth of <i>Escherichia coli</i> .			[14]
<input type="checkbox"/> hsa-miR-146a-5p				
HGF ID	HGF0252	HGF Info	Class	Non-coding RNA: Micro (ncRNA-miRNA)
Description	The miR-146a downregulated the expression of the genes <i>IRAK1</i> and <i>TRAF6</i> in <i>Escherichia coli</i> .			[23]
<input type="checkbox"/> rs11145753				
HGF ID	HGF1364	HGF Info	Class	Single Nucleotide Polymorphism: Intron variant (SNP-IV)
Description	The rs11145753 SNP is significantly associated with the abundance of <i>Escherichia coli</i> (p-value=2.69227E-06).			[28]
<input type="checkbox"/> rs10870199				
HGF ID	HGF1713	HGF Info	Class	Single Nucleotide Polymorphism: Synonymous variant (SNP-SV)
Description	The rs10870199 SNP is significantly associated with the abundance of <i>Escherichia coli</i> (p-value=2.11828E-06).			[28]
<input type="checkbox"/> rs10870188				
HGF ID	HGF1714	HGF Info	Class	Single Nucleotide Polymorphism: Missense variant (SNP-MV)
Description	The rs10870188 SNP is significantly associated with the abundance of <i>Escherichia coli</i> (p-value=5.02838E-06).			[28]

Figure 2. Example host genetic factors in GIMICA. 1368 single nucleotide polymorphisms regulating the relative abundance of 267 microbes that primarily inhabit human gut, mucosa and vagina; 186 gene copy number variations modulating the expression of 156 microbes that live mainly in human oral cavity, gut, lung, skin, blood and mucosa; 1297 non-coding RNA regulations altering the expression of 251 microbes that reside mostly in host lung, oral cavity, gut and skin.

B-cells, natural killer cells, *etc.*) and phagocytes (including macrophages, neutrophils, dendritic cells, monocytes, eosinophils, mast cells, *etc.*) affecting the expression of 258 microbes that are mainly located in human blood, gut, skin, oral cavity and ocular surface; 387 immune proteins (cytokines, immune receptors, checkpoint proteins, complement components, *etc.*) altering the compositions of 235 microbes that primarily reside in human gut, oral cavity, blood, ocular surface, and vagina; 36 immune-related signaling pathways (toll-like receptor, interleukin, *etc.*) modulating the gene expressions of 65 microbe species that live in human blood, gut and skin. All these HIF data can be assessed and retrieved using various search strategies in both the Home page and the subpage entitled ‘Host Immune Factors’ of GIMICA. The example host immune factors in GIMICA were illustrated in the Figure 3. Figure 4 shows the body sites distribution of microbes inhabiting human (Figure 4A) and the statistics of host genetic factors and immune factors shaping human microbiota in GIMICA (Figure 4B).

Collective consideration among different factors

Human microbiota composition was frequently discovered to be collectively shaped by multiple host factors (46,47). For example, genetic variation in host protein NOD2 (HGF) and the immune response stimulated by the NF-kappa B pathway (HIF) could collectively increase the abundance of microbes in the *Enterobacteriaceae* family, and this microbiota dysbiosis had been discovered as a leading cause of neonatal meningitis (13). In other words, when associating the alteration of microbiome composition with a studied disease, both host genetic and immune factors should be assessed. To make this collective consideration possible, both types of host factor (HGF and HIF) were systematically reviewed and collected in GIMICA for each microbe. As a result, 36.4% of all microbes in GIMICA database were identified with both HGF and HIF reported.

Moreover, human microbiota composition was also found susceptible to both environmental and host factors (46,64). For example, a genetic polymorphism in host

Host Immune Factors (HIFs)				
C-X-C motif chemokine 14				
HIF ID	HIFM0077	HIF Info	Class	Antimicrobial peptide (AMP)
Description	The recombinant CXCL14 demonstrates antimicrobial effects against <i>Escherichia coli</i> at 10g/ml.			[32]
Human neutrophil peptide-1				
HIF ID	HIFM0305	HIF Info	Class	Antimicrobial peptide (AMP)
Description	The lethal concentrations of HNP-1 for killing <i>Escherichia coli</i> is 1.0M.			[49]
Human neutrophil peptide-4				
HIF ID	HIFM0308	HIF Info	Class	Antimicrobial peptide (AMP)
Description	Human neutrophil defensin 4 (HNP4) displays more than 100-fold higher activity against <i>Escherichia coli</i> compared with other defensins.			[50]
Cytotoxic T-lymphocyte-associated protein 4				
HIF ID	HIFM0068	HIF Info	Class	Checkpoint molecule (CM)
Description	Changes in CTLA-4 activity regulate the abundance of <i>Escherichia coli</i> .			[67]
Cathelicidin antimicrobial peptide				
HIF ID	HIFM0014	HIF Info	Class	Antimicrobial peptide (AMP)
Description	LL-37 showed the high antimicrobial activity against <i>Escherichia coli</i> , and its minimal inhibition concentration (MIC) determined for <i>Escherichia coli</i> was 1.65 μ M. The LL-37 could permeabilize cell membranes of <i>Escherichia coli</i> by increasing the beta-galactosidase activity.			[30]
Immunoglobulin alpha Fc receptor				
HIF ID	HIFM0097	HIF Info	Class	Fc Receptor (FCR)
Description	After targeting FcRI (CD89), neutrophils enhanced the effect of reducing <i>Escherichia coli</i> abundance.			[84]
Immunoglobulin G1				
HIF ID	HIFM0269	HIF Info	Class	Immunoglobulin (Ig)
Description	Immunoglobulin G1 decreases the abundance of <i>Escherichia coli</i> .			[86]
Immunoglobulin A				
HIF ID	HIFM0272	HIF Info	Class	Immunoglobulin (Ig)
Description	The monoclonal IgA antibodies W27 with potent relative binding ability against <i>Escherichia coli</i> (an optical density of about 4 at an antibody concentration of 50 μ g/ml) could inhibit the enzymatic activity of the serine hydroxymethyltransferase, which resulted in growth suppression of <i>Escherichia coli</i> . When the concentrations of IgA W27 were 23 μ g/ml, 70 μ g/ml and 210 μ g/ml, the cell growth of <i>Escherichia coli</i> was inhibited significantly in a dose-dependent manner (~108 cells, ~106 cells, ~103 cells, respectively) (p-value<0.05).			[88]
Type 3 Innate Lymphoid Cells				
HIF ID	HIFC0009	HIF Info	Class	Innate lymphoid cells (ILCs)
Description	Innate lymphoid cell type 3 derived interleukin-22 boosts the expression and production of lipocalin-2 (LCN-2) in intestinal epithelial cells (IECs) and resulting inhibiting growth of <i>Escherichia coli</i> .			[91]

Figure 3. Example host immune factors in GIMICA. 126 lymphocytes/phagocytes affecting the expressions of 258 microbes that are mainly located in human blood, gut, oral cavity and ocular surface; 387 immune proteins altering the compositions of 235 microbes that primarily reside in human gut, oral cavity, blood and vagina; 36 immune-relevant signaling pathways modulating the expressions of 65 microbes that live in human blood, gut and skin.

interleukin-5 gene (host factor) and the environmental factor (such as high-fat diets) could collectively elevate the concentration of *Prevotella melaninogenica* in vaginal microbiota, which in turn led to a chronic inflammation (65). Since the environmental factors had been long studied, the collective considerations of both host and environmental factors were of great interests and significant importance. To enable this collective consideration, each microbe in GIMICA was cross-referenced and linked to available databases (Disbiome (43), gutMDisorder (44) and aBiofilm (45)) that provided environmental factor. As a result, 88.4% of all microbes in GIMICA were identified with both host and environmental factors.

Data standardization, access and retrieval

To make the access and analysis of GIMICA data convenient for all users, the collected raw data were carefully cleaned up and then systematically standardized. These standardizations included: (i) all diseases in GIMICA were standardized by the latest version of International Classification of Diseases (ICD-11, officially released by World Health Organization (66)), which was expected to serve comprehensive health management (67); (ii) all microbe species were standardized using NCBI Taxonomy database (55) and (iii) the extend data of each microbe could be accessed using crosslinks to ICD-11 (66), UniProt (68),

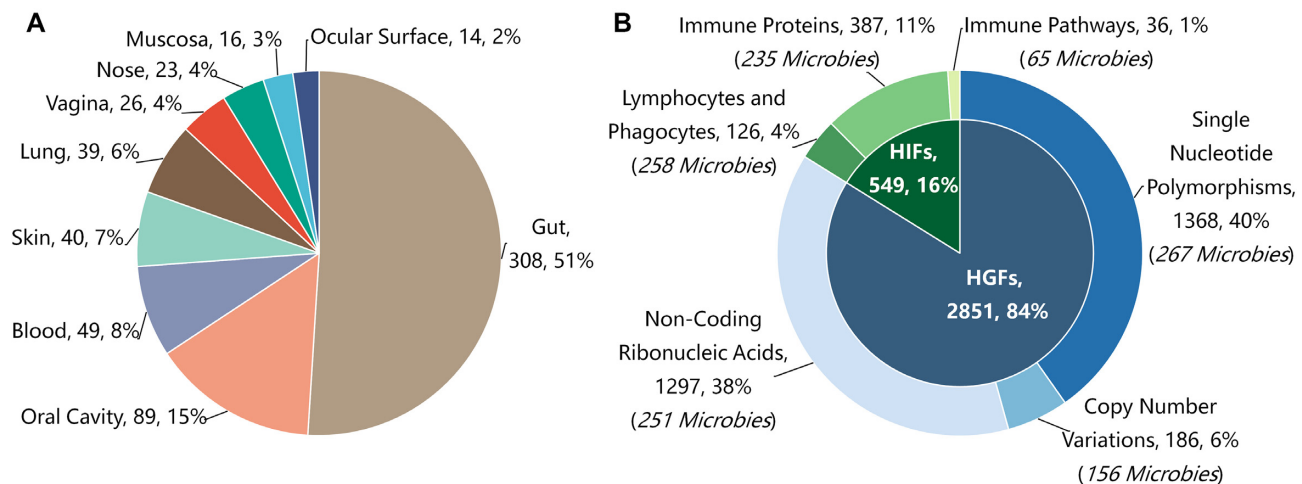


Figure 4. The statistic of microbes and host factors in GIMICA. (A) The body sites distribution of microbes inhabiting human; (B) The statistics of host genetic factors (HGFs) and host immune factors (HIFs) shaping human microbiota in GIMICA.

KEGG (69), NCBI Taxonomy (55), gutMDisorder (44), Disbiome (43), etc. GIMICA has been smoothly running for months and tested from various sites around the world. All data can be viewed, assessed and downloaded. GIMICA is freely assessable without login requirement by all users at: <https://idrblab.org/gimica/>

CONCLUSION

In sum, GIMICA is unique in: (i) providing both host genetic and immune factors shaping human microbiota and (ii) enabling collective consideration not only among various host factors but also between the host and environmental ones. With the extensive efforts made on describing the host factors that are capable of shaping human microbiota (10–12) and discussing the collective determination by both host and environmental factors (46,47,64), those immense, connected and structuralized data provided in GIMICA are expected to have implication for the future practice of MCs-related study on human physiology and health (1–3,70–73).

FUNDING

National Key R&D Program of China [2018YFC0910500]; National Natural Science Foundation of China [U1909208, 81872798]; Project supported by the China Knowledge Centre for Engineering Sciences and Technology [No. CKCEST-2019-1-12]; Leading Talent of the ‘Ten Thousand Plan’ – National High-Level Talents Special Support Plan, Key R&D Program of Zhejiang Province [2020C03010]; Technology Innovation and Application Demonstration Project of Chongqing [cstc2018jcsx-msybX0287]; Fundamental Research Funds for the Central Universities [2018QNA7023, 10611CD-JXZ238826, 2018CDQYSG0007, CDJZR14468801, 2019CDYGYB005]; Information Technology Center, Zhejiang University. Funding for open access charge: National Natural Science Foundation of China [U1909208, 81872798].

Conflict of interest statement. None declared.

REFERENCES

- Ang, Q.Y., Alexander, M., Newman, J.C., Tian, Y., Cai, J., Upadhyay, V., Turnbaugh, J.A., Verdin, E., Hall, K.D., Leibel, R.L. *et al.* (2020) Ketogenic diets alter the gut microbiome resulting in decreased intestinal Th17 cells. *Cell*, **181**, 1263–1275.
- Liu, J., Lian, Q., Chen, Y. and Qi, J. (2019) Amino acid based de bruijn graph algorithm for identifying complete coding genes from metagenomic and metatranscriptomic short reads. *Nucleic Acids Res.*, **47**, e30.
- Brown, J.M. and Hazen, S.L. (2018) Microbial modulation of cardiovascular disease. *Nat. Rev. Microbiol.*, **16**, 171–181.
- Hong, J., Luo, Y., Mou, M., Fu, J., Zhang, Y., Xue, W., Xie, T., Tao, L., Lou, Y. and Zhu, F. (2019) Convolutional neural network-based annotation of bacterial type IV secretion system effectors with enhanced accuracy and reduced false discovery. *Brief. Bioinform.*, **21**, 1825–1836.
- Sonnenburg, J.L. and Backhed, F. (2016) Diet-microbiota interactions as moderators of human metabolism. *Nature*, **535**, 56–64.
- Schloss, P.D., Iverson, K.D., Petrosino, J.F. and Schloss, S.J. (2014) The dynamics of a family’s gut microbiota reveal variations on a theme. *Microbiome*, **2**, 25.
- Spor, A., Koren, O. and Ley, R. (2011) Unravelling the effects of the environment and host genotype on the gut microbiome. *Nat. Rev. Microbiol.*, **9**, 279–290.
- Tang, J., Fu, J., Wang, Y., Li, B., Li, Y., Yang, Q., Cui, X., Hong, J., Li, X., Chen, Y. *et al.* (2020) ANPELA: analysis and performance assessment of the label-free quantification workflow for metaproteomic studies. *Brief. Bioinform.*, **21**, 621–636.
- Tang, J., Mou, M., Wang, Y., Luo, Y. and Zhu, F. (2020) MetaFS: performance assessment of biomarker discovery in metaproteomics. *Brief. Bioinform.*, doi:10.1093/bib/bbaa105.
- Deshpande, N.P., Riordan, S.M., Castano-Rodriguez, N., Wilkins, M.R. and Kaakoush, N.O. (2018) Signatures within the esophageal microbiome are associated with host genetics, age, and disease. *Microbiome*, **6**, 227.
- Poole, A.C., Goodrich, J.K., Youngblut, N.D., Luque, G.G., Rudaud, A., Sutter, J.L., Waters, J.L., Shi, Q., El-Hadidi, M., Johnson, L.M. *et al.* (2019) Human salivary amylase gene copy number impacts oral and gut microbiomes. *Cell Host Microbe*, **25**, 553–564.
- Kurilshikov, A., Wijnenga, C., Fu, J. and Zhernakova, A. (2017) Host genetics and gut microbiome: challenges and perspectives. *Trends Immunol.*, **38**, 633–647.
- Hall, A.B., Tolonen, A.C. and Xavier, R.J. (2017) Human genetic variation and the gut microbiome in disease. *Nat. Rev. Genet.*, **18**, 690–699.
- Wang, L., Ping, P., Kuang, L., Ye, S., Lqbal, F.M.B. and Pei, T. (2018) A novel approach based on bipartite network to predict human microbe-disease associations. *Curr. Bioinformatics*, **13**, 141–148.

15. Munir, A., Malik, S.I. and Malik, K.A. (2019) Proteome mining for the identification of putative drug targets for human pathogen *Clostridium tetani*. *Curr. Bioinformatics*, **14**, 532–540.
16. An, Y., Wang, Y., Zhan, J., Tang, X., Shen, K., Shen, F., Wang, C., Luan, W., Wang, X., Wang, X. *et al.* (2019) Fosfomicin protects mice from *Staphylococcus aureus* pneumonia caused by alpha-hemolysin in extracellular vesicles by inhibiting MAPK-regulated NLRP3 inflammasomes. *Front. Cell. Infect. Microbiol.*, **9**, 253.
17. Alkema, W., Boekhorst, J., Wels, M. and van Hijum, S.A. (2016) Microbial bioinformatics for food safety and production. *Brief. Bioinform.*, **17**, 283–292.
18. Tang, J., Wang, Y., Fu, J., Zhou, Y., Luo, Y., Zhang, Y., Li, B., Yang, Q., Xue, W., Lou, Y. *et al.* (2020) A critical assessment of the feature selection methods used for biomarker discovery in current metaproteomics studies. *Brief. Bioinform.*, **21**, 1378–1390.
19. Ahern, P.P. and Maloy, K.J. (2020) Understanding immune-microbiota interactions in the intestine. *Immunology*, **159**, 4–14.
20. Leshem, A., Liwinski, T. and Elinav, E. (2020) Immune-microbiota Interplay and colonization resistance in infection. *Mol. Cell*, **78**, 597–613.
21. Guzman-Castaneda, S.J., Ortega-Vega, E.L., de la Cuesta-Zuluaga, J., Velasquez-Mejia, E.P., Rojas, W., Bedoya, G. and Escobar, J.S. (2020) Gut microbiota composition explains more variance in the host cardiometabolic risk than genetic ancestry. *Gut Microbes*, **11**, 191–204.
22. Palm, N.W., de Zoete, M.R. and Flavell, R.A. (2015) Immune-microbiota interactions in health and disease. *Clin. Immunol.*, **159**, 122–127.
23. Ortega-Vega, E.L., Guzman-Castaneda, S.J., Campo, O., Velasquez-Mejia, E.P., de la Cuesta-Zuluaga, J., Bedoya, G. and Escobar, J.S. (2020) Variants in genes of innate immunity, appetite control and energy metabolism are associated with host cardiometabolic health and gut microbiota composition. *Gut Microbes*, **11**, 556–568.
24. Wen, C., Yan, W., Sun, C., Ji, C., Zhou, Q., Zhang, D., Zheng, J. and Yang, N. (2019) The gut microbiota is largely independent of host genetics in regulating fat deposition in chickens. *ISME J.*, **13**, 1422–1436.
25. Li, Y.H., Yu, C.Y., Li, X.X., Zhang, P., Tang, J., Yang, Q., Fu, T., Zhang, X., Cui, X., Tu, G. *et al.* (2018) Therapeutic target database update 2018: enriched resource for facilitating bench-to-clinic research of targeted therapeutics. *Nucleic Acids Res.*, **46**, D1121–D1127.
26. Yuan, C., Burns, M.B., Subramanian, S. and Blekhan, R. (2018) Interaction between host MicroRNAs and the gut microbiota in colorectal cancer. *mSystems*, **3**, e00205-17.
27. Wang, Y., Zhang, S., Li, F., Zhou, Y., Zhang, Y., Wang, Z., Zhang, R., Zhu, J., Ren, Y., Tan, Y. *et al.* (2020) Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics. *Nucleic Acids Res.*, **48**, D1031–D1041.
28. Zheng, D., Liwinski, T. and Elinav, E. (2020) Interaction between microbiota and immunity in health and disease. *Cell Res.*, **30**, 492–506.
29. Honda, K. and Littman, D.R. (2016) The microbiota in adaptive immune homeostasis and disease. *Nature*, **535**, 75–84.
30. Huang, L., Zhang, H., Wu, P., Entwistle, S., Li, X., Yohe, T., Yi, H., Yang, Z. and Yin, Y. (2018) dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation. *Nucleic Acids Res.*, **46**, D516–D521.
31. Escapa, I.F., Chen, T., Huang, Y., Gajare, P., Dewhirst, F.E. and Lemon, K.P. (2018) New insights into human nostril microbiome from the expanded human oral microbiome database (eHOMD): a resource for the microbiome of the human aerodigestive tract. *mSystems*, **3**, e00187-18.
32. Shi, W., Qi, H., Sun, Q., Fan, G., Liu, S., Wang, J., Zhu, B., Liu, H., Zhao, F., Wang, X. *et al.* (2019) gcMeta: a global catalogue of metagenomics platform to support the archiving, standardization and analysis of microbiome data. *Nucleic Acids Res.*, **47**, D637–D648.
33. Wu, S., Sun, C., Li, Y., Wang, T., Jia, L., Lai, S., Yang, Y., Luo, P., Dai, D., Yang, Y.Q. *et al.* (2020) GMrepo: a database of curated and consistently annotated human gut metagenomes. *Nucleic Acids Res.*, **48**, D545–D553.
34. Forster, S.C., Browne, H.P., Kumar, N., Hunt, M., Denise, H., Mitchell, A., Finn, R.D. and Lawley, T.D. (2016) HPMCD: the database of human microbial communities from metagenomic datasets and microbial reference genomes. *Nucleic Acids Res.*, **44**, D604–D609.
35. Blanco-Miguez, A., Gutierrez-Jacome, A., Fdez-Riverola, F., Lourenco, A. and Sanchez, B. (2017) MAHMI database: a comprehensive MetaHit-based resource for the study of the mechanism of action of the human microbiota. *Database*, **2017**, baw157.
36. Uchiyama, I., Mihara, M., Nishide, H., Chiba, H. and Kato, M. (2019) MBDG update 2018: microbial genome database based on hierarchical orthology relations covering closely related and distantly related comparisons. *Nucleic Acids Res.*, **47**, D382–D389.
37. Huang, K., Brady, A., Mahurkar, A., White, O., Gevers, D., Huttenhower, C. and Segata, N. (2014) MetaRef: a pan-genomic database for comparative and community microbial genomics. *Nucleic Acids Res.*, **42**, D617–D624.
38. Li, J., Jia, H., Cai, X., Zhong, H., Feng, Q., Sunagawa, S., Arumugam, M., Kultima, J.R., Pifti, E., Nielsen, T. *et al.* (2014) An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol.*, **32**, 834–841.
39. Wishart, D.S., Feunang, Y.D., Marcu, A., Guo, A.C., Liang, K., Vazquez-Fresno, R., Sajed, T., Johnson, D., Li, C., Karu, N. *et al.* (2018) HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.*, **46**, D608–D617.
40. Lemfack, M.C., Gohlke, B.O., Toguem, S.M.T., Preissner, S., Piechulla, B. and Preissner, R. (2018) mVOC 2.0: a database of microbial volatiles. *Nucleic Acids Res.*, **46**, D1261–D1265.
41. Noronha, A., Modamio, J., Jarosz, Y., Guerard, E., Sompairac, N., Preciat, G., Danielsdottir, A.D., Krecke, M., Merten, D., Haraldsdottir, H.S. *et al.* (2019) The virtual metabolic human database: integrating human and gut microbiome metabolism with nutrition and disease. *Nucleic Acids Res.*, **47**, D614–D624.
42. Yang, J., Park, J., Park, S., Baek, I. and Chun, J. (2019) Introducing murine microbiome database (MMDB): a curated database with taxonomic profiling of the healthy mouse gastrointestinal microbiome. *Microorganisms*, **7**, 480.
43. Janssens, Y., Nielandt, J., Bronselaer, A., Debonne, N., Verbeke, F., Wynneaele, E., Van Immerseel, F., Vandewynckel, Y.P., De Tre, G. and De Spiegeleer, B. (2018) Disbiome database: linking the microbiome to disease. *BMC Microbiol.*, **18**, 50.
44. Cheng, L., Qi, C., Zhuang, H., Fu, T. and Zhang, X. (2020) gutMDisorder: a comprehensive database for dysbiosis of the gut microbiota in disorders and interventions. *Nucleic Acids Res.*, **48**, D554–D560.
45. Rajput, A., Thakur, A., Sharma, S. and Kumar, M. (2018) aBiofilm: a resource of anti-biofilm agents and their potential implications in targeting antibiotic drug resistance. *Nucleic Acids Res.*, **46**, D894–D900.
46. Fan, P., Bian, B., Teng, L., Nelson, C.D., Driver, J., Elzo, M.A. and Jeang, K.C. (2020) Host genetic effects upon the early gut microbiota in a bovine model with graduated spectrum of genetic variation. *ISME J.*, **14**, 302–317.
47. Benson, A.K., Kelly, S.A., Legge, R., Ma, F., Low, S.J., Kim, J., Zhang, M., Oh, P.L., Nehrenberg, D., Hua, K. *et al.* (2010) Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *PNAS*, **107**, 18933–18938.
48. Yang, Q., Wang, Y., Zhang, Y., Li, F., Xia, W., Zhou, Y., Qiu, Y., Li, H. and Zhu, F. (2020) NOREVA: enhanced normalization and evaluation of time-course and multi-class metabolomic data. *Nucleic Acids Res.*, **48**, W436–W448.
49. Zhao, T., Hu, Y., Zang, T. and Cheng, L. (2020) MRTFB regulates the expression of NOMO1 in colon. *PNAS*, **117**, 7568–7569.
50. Wang, Y., Wu, Q., Zhang, W. and Zhang, N. (2019) Tuberculosis and HIV coinfection—the challenge in the prevention, detection and treatment of tuberculosis. *Curr. Bioinformatics*, **14**, 91–99.
51. Narang, P., Dangi, M., Sharma, D., Khichi, A. and Chhillar, A.K. (2019) An integrated chikungunya virus database to facilitate therapeutic analysis: ChkVDB. *Curr. Bioinformatics*, **14**, 323–332.
52. Mobegi, F.M., Zomer, A., de Jonge, M.I. and van Hijum, S.A. (2017) Advances and perspectives in computational prediction of microbial gene essentiality. *Brief. Funct. Genomics*, **16**, 70–79.

53. Suhr, M.J. and Hallen-Adams, H.E. (2015) The human gut microbiome: pitfalls and potentials—a mycologist’s perspective. *Mycologia*, **107**, 1057–1073.
54. Sayers, E.W., Beck, J., Brister, J.R., Bolton, E.E., Canese, K., Comeau, D.C., Funk, K., Ketter, A., Kim, S., Kimchi, A. *et al.* (2020) Database resources of the national center for biotechnology information. *Nucleic Acids Res.*, **48**, D9–D16.
55. Federhen, S. (2015) Type material in the NCBI taxonomy database. *Nucleic Acids Res.*, **43**, D1086–D1098.
56. Almeida, A., Mitchell, A.L., Boland, M., Forster, S.C., Gloor, G.B., Tarkowska, A., Lawley, T.D. and Finn, R.D. (2019) A new genomic blueprint of the human gut microbiota. *Nature*, **568**, 499–504.
57. Hallen-Adams, H.E. and Suhr, M.J. (2017) Fungi in the healthy human gastrointestinal tract. *Virulence*, **8**, 352–358.
58. Kwa, M., Plottel, C.S., Blaser, M.J. and Adams, S. (2016) The intestinal microbiome and estrogen receptor-positive female breast cancer. *J. Natl. Cancer Inst.*, **108**, djw029.
59. Nichols, R.G., Peters, J.M. and Patterson, A.D. (2019) Interplay between the host, the human microbiome, and drug metabolism. *Hum. Genomics*, **13**, 27.
60. Ozkan, J., Willcox, M., Wemheuer, B., Wilcsek, G., Coroneo, M. and Thomas, T. (2019) Biogeography of the human ocular microbiota. *Ocular Surf.*, **17**, 111–118.
61. Simon, C. (2018) Introduction: do microbes in the female reproductive function matter? *Fertil. Steril.*, **110**, 325–326.
62. Ozkan, J. and Willcox, M.D. (2019) The ocular microbiome: molecular characterisation of a unique and low microbial environment. *Curr. Eye Res.*, **44**, 685–694.
63. Ozkan, J., Willcox, M.D., de la Jara, P.L., Mandathara, P.S., Rathi, V.M., Thomas, V. and Holden, B.A. (2012) The effect of daily lens replacement during overnight wear on ocular adverse events. *Optom. Vis. Sci.*, **89**, 1674–1681.
64. Zuo, T. and Ng, S.C. (2018) The gut microbiota in the pathogenesis and therapeutics of inflammatory bowel disease. *Front. Microbiol.*, **9**, 2247.
65. Larsen, J.M. (2017) The immune response to Prevotella bacteria in chronic inflammatory disease. *Immunology*, **151**, 363–374.
66. Lancet, T. (2019) ICD-11. *Lancet*, **393**, 2275.
67. Lancet Infectious Diseases, T. (2018) ICD-11: in praise of good data. *Lancet. Infect. Dis.*, **18**, 813.
68. UniProt, C. (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.*, **47**, D506–D515.
69. Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K. and Tanabe, M. (2019) New approach for understanding genome variations in KEGG. *Nucleic Acids Res.*, **47**, D590–D595.
70. Rao, R.T. and Jayakumar, K. (2017) Genomic islands of *Mannheimia haemolytica* - in silico analysis. *Curr. Bioinformatics*, **12**, 369–377.
71. Zhuang, H., Cheng, L., Wang, Y., Zhang, Y.K., Zhao, M.F., Liang, G.D., Zhang, M.C., Li, Y.G., Zhao, J.B., Gao, Y.N. *et al.* (2019) Dysbiosis of the gut microbiome in lung cancer. *Front. Cell. Infect. Microbiol.*, **9**, 112.
72. Bayjanov, J.R., Starrenburg, M.J., van der Sijde, M.R., Siezen, R.J. and van Hijum, S.A. (2013) Genotype-phenotype matching analysis of 38 *Lactococcus lactis* strains using random forest methods. *BMC Microbiol.*, **13**, 68.
73. Banjara, N., Nickerson, K.W., Suhr, M.J. and Hallen-Adams, H.E. (2016) Killer toxin from several food-derived *Debaryomyces hansenii* strains effective against pathogenic *Candida* yeasts. *Int. J. Food Microbiol.*, **222**, 23–29.