Title

GutMetaNet: an integrated database for exploring horizontal gene transfer and functional redundancy in the human gut microbiome.

Journal

Nucleic Acids Research, 53(D1)

ISSN

0305-1048

Authors

Jiang, Yiqi

Wang, Yanfei

Che, Lijia

et al.

Publication Date

2025

DOI

10.1093/nar/gkae1007

Nucleic Acids Research, 2025, **53**, gkae1007
https://doi.org/10.1093/nar/gkae1007
Advance access publication date: 11 November 2024
Database issue

GutMetaNet: an integrated database for exploring horizontal gene transfer and functional redundancy in the human gut microbiome

Yiqi Jiang[1,2,†], Yanfei Wang[1,†], Lijia Che[1,2,†], Shuo Yang[1,2,†], Xianglilan Zhang[3,†], Yu Lin[3,4], Yucheng Shi[1,2], Nanhe Zou[1,2], Shuai Wang[1,2], Yuanzheng Zhang[1,2], Zicheng Zhao[5] and Shuai Cheng Li[1,2,*]

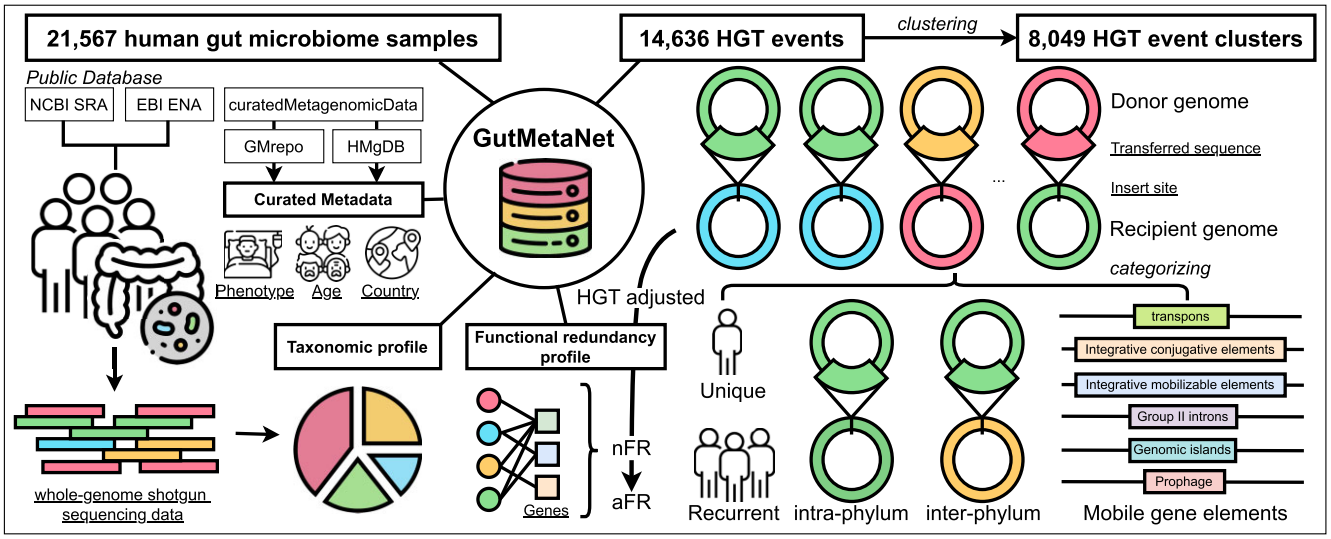
[1]City University of Hong Kong Shenzhen Research Institute, 8 Yue Xing Yi Road, Nanshan District, Shenzhen, 518057, China
[2]Department of Computer Science, City University of Hong Kong, 83 Tat Chee Ave, Kowloon Tong, Hong Kong
[3]State Key Laboratory of Pathogen and Biosafety, 20 East Street, Fengtai District, Beijing, 100071, China
[4]Beijing University of Chemical Technology, 15 Beisanhuan East Road, Chaoyang District, Beijing, 100029, China
[5]OmicLab Limited, Unit 917, 19 Science Park West Avenue, New Territories, Hong Kong

[*]To whom correspondence should be addressed. Tel: +852 34429412; Fax: +852 34420503; Email: shuaicli@cityu.edu.hk
[†]The first five authors should be regarded as Joint First Authors.

Abstract

Metagenomic studies have revealed the critical roles of complex microbial interactions, including horizontal gene transfer (HGT) and functional redundancy (FR), in shaping the gut microbiome's functional capacity and resilience. However, the lack of comprehensive data integration and systematic analysis approaches has limited the in-depth exploration of HGT and FR dynamics across large-scale gut microbiome datasets. To address this gap, we present GutMetaNet (https://gutmetanet.deepomics.org/), a first-of-its-kind database integrating extensive human gut microbiome data with comprehensive HGT and FR analyses. GutMetaNet contains 21 567 human gut metagenome samples with whole-genome shotgun sequencing data related to various health conditions. Through systematic analysis, we have characterized the taxonomic profiles and FR profiles, and identified 14 636 HGT events using a shared reference genome database across the collected samples. These HGT events have been curated into 8049 clusters, which are annotated with categorized mobile genetic elements, including transposons, prophages, integrative mobilizable elements, genomic islands, integrative conjugative elements and group II introns. Additionally, GutMetaNet incorporates automated analyses and visualizations for the HGT events and FR, serving as an efficient platform for in-depth exploration of the interactions among gut microbiome taxa and their implications for human health.

Graphical abstract



Received: August 15, 2024. **Revised:** October 9, 2024. **Editorial Decision:** October 11, 2024. **Accepted:** October 16, 2024

Introduction

The human gut microbiome has been extensively studied for its critical role in human health and disease. Recent studies have highlighted the critical importance of diverse microbiome interactions (1–4), beyond just taxonomic abundances, in understanding these complex relationships. Emerging evidence suggests that factors such as horizontal gene transfer (HGT) networks (1,5,6) and genetic content networks [reveal functional redundancy (FR)] (7) can have significant implications for an individual's overall health and resilience.

Numerous studies have highlighted the critical role of HGT in shaping microbial metabolic networks within the human gut microbiome, enabling rapid adaptation through genetic innovation (1,8). These HGTs are essential for understanding the transmission of antibiotic resistance and disease-associated genetic elements, directly impacting human health (9,10). Analyzing HGT networks reveals distinct community structures linked to various health conditions, such as early life stages and inflammatory bowel disease, offering insights into disease mechanisms (11). By capturing the dynamic nature of gene flow, HGT networks serve as crucial biomarkers for diagnosing and understanding the microbiome's role in disease progression (6). Exploring HGT networks is vital in gut metagenomic research, providing a unique perspective on the adaptive capabilities and functional potential of the human gut microbiome in disease contexts (12,13).

To facilitate large-scale analysis of HGT events in the human gut microbiome, we employed the LocalHGT method (14). Traditional HGT detection approaches, such as parametric methods and phylogenetic methods, face notable limitations when applied to metagenomic whole-genome shotgun (mWGS) data. Parametric methods (15,16) rely on uniform genomic signatures, while phylogenetic methods (17–21) need precise evolutionary trees of taxa and genes, which can be challenging to discern in mixed microbial communities present in mWGS samples (22). Moreover, all-against-all BLASTN, a commonly used phylogenetic approach, can only detect recent HGT events (23). More recently, specialized HGT detection tools have been developed for mWGS data, including MetaCHIP (24), DaisySuite (25), LEMON (26) and LocalHGT (14). These approaches aim to overcome the limitations of traditional methods by directly analyzing mWGS data and can detect both recent and ancient HGT events. However, these specialized tools still face challenges. MetaCHIP, like traditional assembly-based approaches, requires intensive computing resources for genome assembly, and it can be difficult to integrate comparable results across samples. Junction reads-based methods such as DaisySuite and LEMON, on the other hand, require significant computational resources for aligning large reference databases. In contrast, LocalHGT, also a junction reads-based approach, employs a fast, fuzzy *k*-mer matching strategy to expedite the HGT detection process and reduce resource requirements, enabling large-scale analysis. Compared to other junction reads-based methods, LocalHGT has been shown to have lower computational resource demands and faster processing times (14). Additionally, LocalHGT can identify complete HGT events, including the transferred sequence as well as the corresponding deletion and insertion sites in the donor and recipient genomes. Overall, LocalHGT provides a more efficient means of detecting both recent and ancient HGT events within large-scale mWGS datasets (Supplementary Table S1).

FR networks offer a lens to understand the stability, resilience and assembly principles of the human gut microbiome and their implications for health. Although the taxonomic composition of the gut microbiome varies greatly across individuals, its functional capacity is highly conserved, suggesting FR (27,28). Studies have demonstrated its importance in understanding microbiome dynamics. High FR in the recipient's pretreatment microbiome can pose a barrier to successful engraftment of donor microbiota during fecal transplantation, highlighting the role of FR in microbiome resilience (7). Genomic content networks (GCNs) have been used to reveal the topological features favoring FR. Specifically, the study by Tian *et al.* demonstrated that a high rate of HGT is necessary to generate a GCN with a highly nested structure and a very heterogeneous gene degree distribution, which are crucial features to maintain high FR in the human microbiome (7). This suggests the importance of combining the analysis of HGT and FR to gain a better understanding of the adaptive capabilities and functional potential of the gut microbiome.

The existing gut metagenomic databases lack specialized focus on HGT networks and FR within the human gut microbiome. HumanMetagenomeDB (29) combines metadata from multiple studies and metagenomes of body parts with collated standardized database but lacks the taxonomic abundance data. While databases such as curatedMetagenomicData (30) and GMrepo (31,32) aim to improve data accessibility, they primarily contain sample metadata and taxonomic profiles, without comprehensive information related to HGT and FR. This represents a significant gap in the current gut metagenomic data resources.

To fill this gap, we developed GutMetaNet (https://gutmetanet.deepomics.org), a comprehensive resource containing 21 567 human gut microbiome samples with WGS sequencing data, associated with various health conditions. The primary objective of GutMetaNet is to provide a comprehensive resource that enables researchers to investigate the relationships between HGT patterns and various host phenotypes, including healthy and disease states. Through systematic analysis, GutMetaNet has characterized the taxonomic profiles and FR profiles, and identified 14 636 HGT events across the dataset with LocalHGT. These HGT events have been curated into 8049 clusters and annotated with categorized mobile genetic elements (MGEs), including 1374 transposons, 336 integrative mobilizable elements, 134 genomic islands and others. GutMetaNet also incorporates automated analysis and specialized visualization capabilities to explore HGT networks and FR. One notable feature is a novel visualization approach that clearly depicts the composition of HGT events and the surrounding gene content, facilitating rapid interpretation of these key microbial interactions.

Materials and methods

Human gut mWGS sequencing data collection

We first conducted a comprehensive survey of published literature on the human gut microbiome. We focused on studies that had deposited WGS sequencing data of human gut microbiome samples in public repositories, such as the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) (33) or the European Bioinformatics Institute (EBI) European Nucleotide Archive (ENA) (34). As a result of this curation process, we collected a dataset comprising

21 567 samples from 96 projects (35–126) (Supplementary Table S2).

Metadata curation and integration

We collected metadata of the 21 567 gut microbiome samples from various sources, including NCBI BioProject and existing databases (29,30,32). To establish a consistent metadata repository, we curated and integrated the records, cross-checking and correcting any discrepancies by referring to the original literature. As part of the curation, we standardized the phenotype annotations using MeSH ID (127), converted country names to ALPHA-3 codes and categorized samples into age groups. Samples with missing metadata were marked as 'NA' for further investigation.

Raw sequencing data processing

We retrieved the raw sequencing data for the 21 567 gut microbiome samples from the NCBI SRA and the EBI ENA. For the data downloaded from NCBI SRA in SRA format, we converted them to FASTQ format using the SRA Toolkit's fasterq-dump (version 3.1.1) (128). Similarly, the data downloaded from ENA in BAM format were transformed into FASTQ using samtools (version 1.20) (129). For samples with multiple sequencing runs, we merged the reads into single paired-end FASTQ files. Finally, we performed quality control on the raw sequencing data using fastp (version 0.23.4) (130) with the default parameters.

Taxonomic profiles

We applied Kraken2 (version 2.0.7-beta) (131) to the high-quality paired-end sequencing data to assess the taxonomic profiles. We used the pre-built UHGG v2.0.2 (132) Kraken database in EBI FTP (http://ftp.ebi.ac.uk/pub/databases/metagenomics/) as the reference. The Kraken2 taxonomic assignments were further refined using Bracken (version 2.9). We then transformed the results into the MPA format, which is commonly used by the MetaPhlAn series of tools, using the KrakenTools (133). Finally, we scaled the taxonomic profiles to relative abundances using an in-house script.

FR profiles

We followed the method from the original study (7) to determine the FR profiles of the gut microbiome samples in the GutMetaNet database. We deployed an in-house script to apply the same method on the UHGG v2.0.2 reference genomes. First, we constructed a GCN based on the UHGG database. We extracted the representative genomes for each species according to the UHGG annotations and used Prokka (version 1.14.6) (134) to determine their functional genomic content in the form of KEGG Orthology (KO) identifiers (135). This allowed us to construct a GCN in the form as its definition (7), a weighted bipartite graph connecting the taxa to their associated KOs. The edge weights represent the copy number of KOs in the corresponding genomes. Next, we calculated the normalized FR (nFR) for each sample as $\mathrm{nFR}_\alpha = \mathrm{FR}_\alpha / \mathrm{TD}_\alpha$, where $\mathrm{TD}_\alpha = \sum_{i=1}^{N} \sum_{j \neq i}^{N} p_i p_j$ and $\mathrm{FR}_\alpha = \sum_{i=1}^{N} \sum_{j \neq i}^{N} (1 - d_{ij}) p_i p_j$. p represents the taxonomic profile of the sample, and d_{ij} is the functional distance between taxa i and j, calculated from the GCN using the Jaccard distance metric.

We computed HGT adjusted FR (aFR) with HGT events detected in the corresponding sample. We first constructed a sample-specific GCN. For each HGT event, the KOs contained in the transferred sequences will transfer to the recipient taxa, so we update the edge weight of those KOs and recipient taxa in GCN by adding the existing number of KOs in the transfer sequences, resulting in the sample-specific GCN′. Next, we calculate the aFR using the same formula of nFR but replacing GCN with GCN′: $\mathrm{aFR}_\alpha = \mathrm{FR}'_\alpha / \mathrm{TD}_\alpha$, where TD_α is the same, $\mathrm{FR}'_\alpha = \sum_{i=1}^{N} \sum_{j \neq i}^{N} (1 - d'_{ij}) p_i p_j$ and d'_{ij} is the functional distance calculated from the GCN′.

HGT event detection and database construction

We identified HGT events for each sample based on the LocalHGT protocol (136): 'localhgt bkp' and 'localhgt event' using the UHGG v2.0.2 representative genomes. For each HGT event, we identified three breakpoints: two representing the start and end of the donor sequence, and one representing the insertion site on the recipient genome.

To account for potential variations in the breakpoint locations, we merged breakpoints within 500-bp flanking regions into a single representative breakpoint, calculated as the average position of all related breakpoints. We then redefined the HGT events using these new representative breakpoints and merged HGT events with identical breakpoints into HGT clusters (HGTCs).

To annotate the HGT events, we first aligned the 4744 UHGG representative genomes to the ImmeDB (intestinal microbiome mobile element database) (137) and virulence factor database (VFDB) (138) using BLASTN (version 2.3.0) (139). We discarded BLAST hits with a mapping length <50% of the MGE sequence or virulence factor gene length. We then classified the matched regions in the UHGG representative genomes as partial (50–80%), substantial (80–95%) or full (>95%) based on the length of the match relative to the full sequence.

For each HGT event, we screened the donor sequence (the transferred sequence plus 5-kb flanking regions) and the recipient sequence (5 kb flanking the insertion site) to identify all overlapping MGEs in the corresponding UHGG genomes. Additionally, we annotated the genes within the transferred sequences using the GFF files provided by UHGG.

Platform development

GutMetaNet is hosted on an Ubuntu 20.04.6 LTS server, which is outfitted with 1 TB of memory and 90 TB of storage. The platform's backend functionality is supported by an in-house framework (140–142) consisting of Apache, Django, PostgreSQL, Typescript + Vue3 and Slurm. All online data visualizations are implemented with Oviz (143). We provide detailed tutorials on the platform to facilitate usage.

Results

GutMetaNet database

The GutMetaNet database hosts a comprehensive collection of 21576 whole-metagenome shotgun human gut microbiome samples, sourced from diverse public repositories (144,145) and published databases (29–31), corresponding to 96 studies (Supplementary Table S2) with 66 healthy conditions (Figure 1). GutMetaNet provides extensive metadata for the samples, including information on country, health phenotype, gender, age, age category, BMI, diet and antibiotic usage, where available. The samples originate from 46 re-

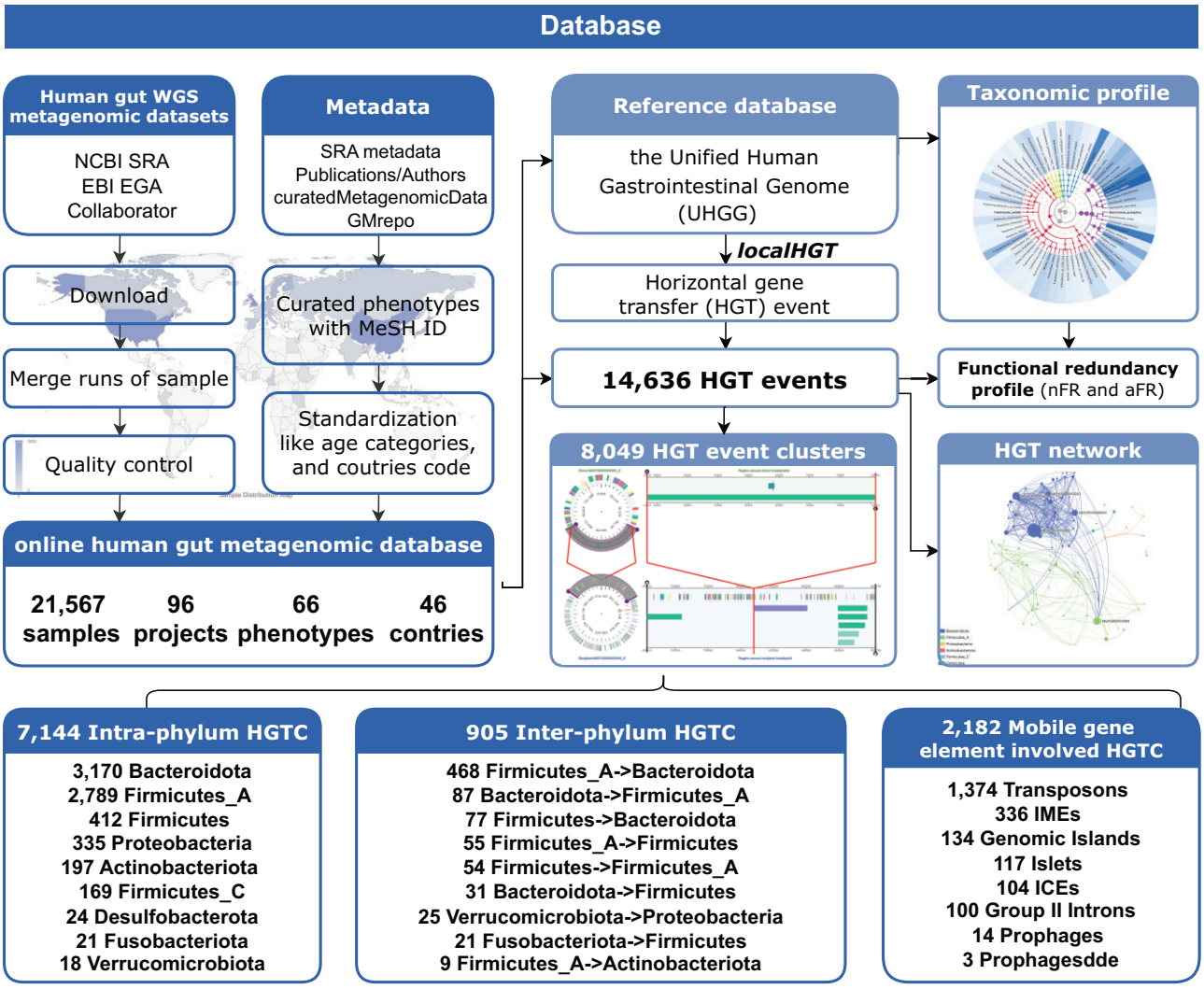


Figure 1. Overview of GutMetaNet database. The GutMetaNet database stores 21576 human microbiome samples with corresponding metadata, taxonomic profile, FR profile and 14636 identified HGT events.

gions, with the majority coming from the USA (4793 samples) and China (3800 samples) (Supplementary Figure S1). The sample distribution across age categories shows that the old-aged (45–60) category has the most samples (3703) (Supplementary Figure S2A). For health conditions, the majority of samples (14 246) are from healthy individuals, while the largest patient groups include those with irritable bowel syndrome (2090), colorectal neoplasms (759), obesity (708) and premature birth (440) (Supplementary Figure S2B and Supplementary Table S3).

GutMetaNet provides the taxonomic profiles and FR profiles for each sample. The taxonomic profiles are based on the UHGG v2.0.2 Kraken2 database (132), transformed into the MPA (MetaPhlAn) format with relative abundances at all taxonomic levels, from species to kingdom. GutMetaNet allows users to directly download the metadata, taxonomic profiles and FR profiles for the gut microbiome samples.

GutMetaNet provides the FR profiles of samples via nFR and aFR (see the 'Materials and methods' section for details). The distribution of nFR shows different patterns across different phenotypic sample groups. We examined MeSH curated phenotypes with >400 samples: health, irritable bowel syndrome, obesity, colorectal neoplasms, infant and diabetes mellitus type 2. First, there were no significant differences in nFR between the phenotypes (overall Kruskal–Wallis test *P*-value: 0.6936). Next, the healthy samples had the smallest normality *P*-value (1.62×10^{-31}) and a kurtosis >3, suggesting that FR distribution of healthy samples was the closest to a normal distribution among the phenotypes examined. The infant samples had the highest mean nFR but also the highest variance. Among the remaining phenotypes, all the disease states had lower mean nFR and lower variance compared to the health group (detailed statistics in Supplementary Table S5 and Supplementary Figure S6A).

The comparison of aFR and nFR revealed that the impact of HGT on nFR was relatively small overall, as the number of HGT events in the samples was generally low. However, there were cases where HGT had a greater impact on FR, particularly in infant samples with simpler microbial compositions, suggesting that HGT can significantly impact FR when it occurs in high-abundance species and involves long transfer sequences. Conversely, some healthy samples from the Mongolian population exhibited a decrease in aFR compared to nFR, highlighting the distinct features of their gut mi-

crobiome that warrant further investigation (detailed results in Supplementary Notes S2.1 and Supplementary Figure S6B).

GutMetaNet collected a total of 14 636 identified HGT events for the samples using the LocalHGT protocol (136). For each HGT event, GutMetaNet includes the corresponding sample, recipient genome, recipient scaffold, insertion site in the recipient, the donor genome, donor scaffold, and the start and end positions of the transferred sequence in the donor, as well as a flag indicating whether the transferred sequence is reversed or not. On average, the samples with detectable HGT events contain 4.35 HGT events. The distribution of the number of HGT events per sample is shown in Supplementary Figure S3A. The length distribution of the transfer sequences is shown in Supplementary Figure S3B, with the 1400–1600 bp range having the highest frequency.

GutMetaNet also provides the clustering of HGT events, resulting in 8049 HGTCs (see the 'Materials and methods' section for details). In addition to the basic information provided for HGT events, the HGTC metadata includes the frequency, phylum and MGE annotations. Based on the frequency, GutMetaNet categorizes the HGTC into 6261 unique clusters and 1788 recurrent clusters. Furthermore, GutMetaNet classifies the HGTC into 7144 intra-phylum and 905 inter-phylum events based on the phylum of the recipient and donor. The intra-phylum HGTCs are dominated by six phyla, with over 100 HGTC each: 3170 in Bacteroidota, 2789 in Firmicutes_A, 412 in Firmicutes, 335 in Proteobacteria, 197 in Actinobacteriota and 169 in Firmicutes_C. The inter-phylum HGTCs consist of eight phylum pairs with over 50 HGTCs each, including 468 from Bacteroidota to Firmicutes_A, 87 from Firmicutes_A to Bacteroidota, 77 from Bacteroidota to Firmicutes, 55 from Firmicutes to Firmicutes_A, 54 from Firmicutes_A to Firmicutes and 31 from Firmicutes to Bacteroidota. The top 10 most frequent donor and recipient phyla are shown in Supplementary Figure S4.

GutMetaNet has comprehensively annotated the identified HGTCs by cataloging the MGEs located in both the recipient flanking sequence (5 kb flanking the insertion site) and the donor flanking sequence (the transferred sequence and 5-kb flanking regions) (see the 'Materials and methods' section for details). In total, the 8049 HGTCs were found to involve 1374 transposons, 336 integrative mobilizable elements, 134 genomic islands, 117 islets, 104 integrative conjugative elements, 100 group II introns and 14 prophages. Notably, the MGEs involved in the donor flanking sequences showed a higher proportion of genomic islands, suggesting their prominent role in facilitating HGT events within the gut microbiome (Supplementary Figure S5 and Supplementary Table S4).

GutMetaNet also annotated the genes located within the transferred sequences and recorded the gene count for each HGTC. Six thousand three hundred forty-four HGTCs did not contain any known genes, while 6391 known genes were involved in the remaining 1705 HGTCs. The most frequently observed genes in the transfer sequences were xerC_19, xerD_2, clpP_1, noc_2 and xerC_1, present in 139, 112, 102, 102 and 100 HGTCs, respectively.

To facilitate the exploration of HGT events, GutMetaNet provides an interactive visualization tool (Supplementary Figure S8). This visualization displays the genomic context of the HGT events, including the donor and recipient scaffolds, highlighting the breakpoint pairs and the transferred sequence, as well as the flanking regions around the HGT site. Users can interactively adjust the view to examine the detailed genomic features in these flanking regions. The visualization also reconstructs the HGT haplotype, annotating genes and MGEs present within the transferred sequence. The visualization includes interactive features, such as tooltips that provide detailed information about the genomic elements. This visualization allows researchers to directly inspect the HGT events and identify potential gene fusions or rearrangements of MGEs.

Additionally, GutMetaNet provides three interactive visualizations for each sample. First, an HGT network visualization illustrates the connections between the detected HGT events and the associated microbial taxa, allowing users to explore the relationships between HGT and the taxonomic structure of the gut microbiome (Supplementary Figure S9). Second, a visualization displays the sample's FR position within the background distribution of FR values observed in healthy samples. This enables users to quickly assess the relative FR status of the samples in comparison to the healthy baseline (Supplementary Figure S10). Third, an interactive taxonomy tree visualization presents the abundance profile of the microbial community in the sample, providing context for interpreting the HGT patterns and FR status (Supplementary Figure S11).

GutMetaNet provides users with access to a comprehensive range of data related to the identified HGTCs, including the metadata, flanking sequences and transferred sequences. Additionally, GutMetaNet offers categorized data on the HGTCs, such as unique HGTCs, recurrent HGTCs, intra-phylum HGTCs per phylum, top frequency inter-phylum HGTCs, and HGTCs involved with specific MGEs. The MGE-involved HGTC data are provided in three forms: (i) MGEs involved in the donor flanking sequences; (ii) MGEs involved in the recipient flanking sequences; and (iii) MGEs involved in either the donor or the recipient. The MGE categories included in the database are genomic islands, group II introns, integrative conjugative elements, integrative mobilizable elements, islets, prophages and transposons.

Automatic analyses and visualizations

GutMetaNet supplies automatic analysis workflows for users to study their own human gut microbiome samples efficiently (Supplementary Figure S7). The HGT annotation pipeline enables the users to annotate their HGT event results with GutMetaNet HGTCs, MGE (137), VFDB (138), KEGG (135) and COG (146). Additionally, it provides functional enrichment analysis and microhomology detection analysis (136) (Supplementary Methods S1.1–S1.3). The HGT aFR calculation module enables users to upload their own taxonomic profile and detected HGT events, and then obtain the corresponding nFR and aFR metrics for their samples (Supplementary Methods S1.4). The HGT functional comparison module allows users to perform functional enrichment analysis between user-defined sample groups or conditions on the HGT-related sequences, leveraging databases such as KEGG, COG, VFDB and MGE (Supplementary Methods S1.5). The HGT–FR co-analysis pipeline provides an analysis module to investigate the correlation between the networks of HGT and FR in between user-defined sample groups, uncovering potential direct associations with HGT, FR and sample phenotypes (Supplementary Methods S1.6). This pipeline also provides interactive visualization of the average FR networks for the cohorts (Supplementary Figure S12). The abundance

biomarker pipeline enables users to upload taxonomic profile of their own cohort with phenotype, to process FR comparison, alpha diversity, beta diversity, dimension reduction, enterotype identification, differential testing and classifier detection (Supplementary Methods S1.7–S1.13). Once the submitted tasks are completed, users can easily download the resultant documents and visualizations.

Comparison to the existing databases and web servers

We compared GutMetaNet with other prominent human gut microbiome databases, including curatedMetagenomicData (30), HumanMetagenomeDB (29) and the online platform GMrepo (32). GutMetaNet is the only database that integrates both HGT event data and FR profiles for each individual sample (comparisons in Supplementary Table S6). Additionally, GutMetaNet provides automated analyses and interactive visualization tools, further enhancing its capabilities compared to the existing resources.

Discussion

The GutMetaNet database is the first of its kind to provide a comprehensive resource for the study of HGT within the human gut microbiome. Additionally, GutMetaNet integrates both HGT event data and FR profiles for individual samples, enabling researchers to investigate the intricate connections between host phenotype and gut microbial interaction.

Through the comprehensive classification of HGTCs into unique, recurrent, intra-phylum and inter-phylum categories, as well as their association with diverse MGEs, GutMetaNet provides resources for exploring the mechanisms underlying HGT in the gut microbiome. For example, previous studies have shown that intra-phylum HGT events occur much more frequently than inter-phylum transfers (136), while the mechanism difference is not clear. The identification of 7144 intra-phylum HGTCs and 905 inter-phylum HGTCs in GutMetaNet offers researchers the opportunity to delve into the distinct mechanisms driving cross-taxonomic gene exchange, in contrast to the more commonly observed intra-phylum transfers.

Additionally, GutMetaNet's provision of flanking sequences and transferred gene information further empowers investigations into the functional implications of these HGT events, such as the dissemination of antibiotic resistance or the introduction of novel metabolic capabilities within the gut ecosystem.

The integration of FR profiles in GutMetaNet empowers researchers to investigate the relationship between human gut microbiome resilience, as indicated by FR, and various disease states. By linking FR, HGT and taxonomic profiles, GutMetaNet provides a valuable platform to unravel the complex interplay between microbial genomic plasticity and the overall functional stability of the gut ecosystem, informing our understanding of microbiome responses to perturbations, such as dietary changes or disease states.

The FR distribution among samples with different phenotypes indicates that the FR distributions across the different phenotypes deviate significantly, with varying degrees of skewness and kurtosis that reflect the unique functional characteristics of the gut microbiomes in these populations. The lower mean nFR and variance observed in the disease states suggest that disease processes may contribute to a decrease in FR, potentially due to the gut microbiome being under stress and shifting toward a more uniform state (147,148). Consistent with the known characteristics of the infant gut microbiome, the infant samples had the highest mean nFR but also the highest variance. This finding is consistent with the known characteristics of the infant gut microbiome, which is less mature and has greater FR and higher variance compared to adults, due to its relatively simple composition and structure (149).

The other HGT-focused databases such as HGT-DB (150), HGTree (151) and HGTree v2.0 (152) provide collections of putative horizontally transferred genes across prokaryotic genomes. GutMetaNet offers several distinct advantages for studying HGT within the human gut microbiome. First, the data underlying GutMetaNet are derived from human gut metagenome samples, in contrast to those databases that are based on a collection of prokaryotic genomes. While the number of prokaryotic species represented in GutMetaNet may be more limited compared to the broader cross-domain scope of HGTree v2.0, the key advantage is that all the HGT events captured in GutMetaNet are specific to the human gut microbiome and the samples are associated with various host phenotypes, enabling the integration of HGT events with relevant clinical data—a capability that sets GutMetaNet apart from the more generalized HGT databases. Second, GutMetaNet leverages LocalHGT, which can reliably identify not only recent but also ancient HGT events. The phylogenetic approach used in HGTree and HGT-DB can only detect recent HGTs and is not suitable for metagenomic data as we mentioned in the 'Introduction' section. Furthermore, in addition to the transferred genes, GutMetaNet also provides detailed annotations on the corresponding donor sequences, recipient genome insertion sites and surrounding MGEs. This contextual information is crucial for elucidating the mechanisms and potential functional implications of HGT events in shaping the adaptive potential and resilience of the gut microbiome.

While the GutMetaNet database provides a comprehensive resource for studying HGT patterns in the human gut microbiome, it is important to acknowledge the limitations of the underlying methods and data. The LocalHGT approach employed to identify HGT events in GutMetaNet has several key limitations. The primary limitation of the LocalHGT approach is its inability to reliably detect HGT events between different strains of the same microbial species. This is due to the high sequence similarity between closely related strains, which makes it challenging for LocalHGT to confidently identify the distinctive genomic breakpoint signatures. Additionally, while LocalHGT can capture both recent and ancient HGT events, it does not inherently differentiate between the two, requiring additional downstream analysis to discern the evolutionary timescales of the detected HGT. These limitations of the LocalHGT method have implications for the interpretation and representation of HGT patterns within the GutMetaNet platform. Specifically, the inability to reliably detect intra-species HGT events may result in an underestimation of the extent of HGT occurring in the gut microbiome. Furthermore, the inability to directly distinguish between recent and ancient HGT events complicates the analysis of the potential impact of HGT on the functional potential of the microbial community and host phenotype.

Looking ahead, we are actively exploring strategies to further enhance the GutMetaNet HGT dataset by integrating complementary methods and resources. This includes using

additional reference genome databases, such as SGB (Strain-level Genome Bins) (153), which could provide a more comprehensive and comparable HGT detection framework across the GutMetaNet samples, potentially addressing the limitations in strain-level HGT identification. Additionally, we plan to incorporate data from existing phylogenetic HGT databases, such as HGTree v2.0 (152), which catalogs tens of thousands of horizontally transferred genes. This additional annotation could help us better differentiate recent HGT events from ancient ones within the LocalHGT results, providing deeper insights into the evolutionary dynamics of HGT in the human gut microbiome. We intend to gradually roll out these enhancements in future versions of the GutMetaNet resource, as we continue to expand and refine the database to better serve the needs of the research community.

The incorporation of automated analyses and interactive visualizations within GutMetaNet reinforces its purpose of serving the scientific community as a user-friendly and effective tool. GutMetaNet empowers researchers to conveniently explore the details of their custom gut microbiome samples, investigating the underlying HGT and FR patterns. This enables researchers to gain deeper insights and understanding of the intricate microbial interaction characteristics within the human gut microbiome.

Data availability

All the data are freely available at https://gutmetanet.deepomics.org.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

The authors thank Mr Qian Zhou for his help in data process.

Funding

Shenzhen Science and Technology Program [JCYJ20220818101201004]. Funding for open access charge: Shenzhen Science and Technology Program [JCYJ20220818101201004].

Conflict of interest statement

None declared.

References

1. Smillie,C.S., Smith,M.B., Friedman,J., Cordero,O.X., David,L.A. and Alm,E.J. (2011) Ecology drives a global network of gene exchange connecting the human microbiome. *Nature*, **480**, 241–244.
2. Venturelli,O.S., Carr,A.V., Fisher,G., Hsu,R.H., Lau,R., Bowen,B.P., Hromada,S., Northen,T. and Arkin,A.P. (2018) Deciphering microbial interactions in synthetic human gut microbiome communities. *Mol. Syst. Biol.*, **14**, e8157.
3. Berry,D. and Widder,S. (2014) Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Front. Microbiol.*, **5**, 219.
4. Faust,K., Sathirapongsasuti,J.F., Izard,J., Segata,N., Gevers,D., Raes,J. and Huttenhower,C. (2012) Microbial co-occurrence relationships in the human microbiome. *PLoS Comput. Biol.*, **8**, e1002606.
5. Pál,C., Papp,B. and Lercher,M.J. (2005) Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nat. Genet.*, **37**, 1372–1375.
6. Li,C., Chen,J. and Li,S.C. (2020) Understanding horizontal gene transfer network in human gut microbiota. *Gut Pathog.*, **12**, 33.
7. Tian,L., Wang,X.-W., Wu,A.-K., Fan,Y., Friedman,J., Dahlin,A., Waldor,M.K., Weinstock,G.M., Weiss,S.T. and Liu,Y.-Y. (2020) Deciphering functional redundancy in the human microbiome. *Nat. Commun.*, **11**, 6217.
8. Kloesges,T., Popa,O., Martin,W. and Dagan,T. (2011) Networks of gene sharing among 329 proteobacterial genomes reveal differences in lateral gene transfer frequency at different phylogenetic depths. *Mol. Biol. Evol.*, **28**, 1057–1074.
9. Forsberg,K.J., Reyes,A., Wang,B., Selleck,E.M., Sommer,M.O. and Dantas,G. (2012) The shared antibiotic resistome of soil bacteria and human pathogens. *Science*, **337**, 1107–1111.
10. Soucy,S.M., Huang,J. and Gogarten,J.P. (2015) Horizontal gene transfer: building the web of life. *Nat. Rev. Genet.*, **16**, 472–482.
11. Li,H., He,J. and Jia,W. (2016) The influence of gut microbiota on drug metabolism and toxicity. *Expert Opin. Drug Metab. Toxicol.*, **12**, 31–40.
12. Koonin,E.V., Makarova,K.S. and Aravind,L. (2001) Horizontal gene transfer in prokaryotes: quantification and classification. *Annu. Rev. Microbiol.*, **55**, 709–742.
13. Jain,R., Rivera,M.C. and Lake,J.A. (1999) Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl Acad. Sci. U.S.A.*, **96**, 3801–3806.
14. Wang,S., Jiang,Y., Che,L., Wang,R.H. and Li,S.C. (2024) Enhancing insights into diseases through horizontal gene transfer event detection from gut microbiome. *Nucleic Acids Res.*, **52**, e61.
15. Hasan,M.S., Liu,Q., Wang,H., Fazekas,J., Chen,B. and Che,D. (2012) GIST: genomic island suite of tools for predicting genomic islands in genomic sequences. *Bioinformation*, **8**, 203.
16. Langille,M.G. and Brinkman,F.S. (2009) IslandViewer: an integrated interface for computational identification and visualization of genomic islands. *Bioinformatics*, **25**, 664–665.
17. Zhu,Q., Kosoy,M. and Dittmar,K. (2014) HGTector: an automated method facilitating genome-wide discovery of putative horizontal gene transfers. *BMC Genomics*, **15**, 717.
18. Podell,S. and Gaasterland,T. (2007) DarkHorse: a method for genome-wide prediction of horizontal gene transfer. *Genome Biol.*, **8**, R16.
19. David,L.A. and Alm,E.J. (2011) Rapid evolutionary innovation during an Archaean genetic expansion. *Nature*, **469**, 93–96.
20. Bansal,M.S., Alm,E.J. and Kellis,M. (2012) Efficient algorithms for the reconciliation problem with gene duplication, horizontal transfer and loss. *Bioinformatics*, **28**, i283–i291.
21. Dessimoz,C., Margadant,D. and Gonnet,G.H. (2008) DLIGHT—lateral gene transfer detection using pairwise evolutionary distances in a statistical framework. In: Vingron,M. and Wong,L. (eds.) *Annual International Conference on Research in Computational Molecular Biology*. Springer, Berlin, Heidelberg, pp. 315–330.
22. Ravenhall,M., Škunca,N., Lassalle,F. and Dessimoz,C. (2015) Inferring horizontal gene transfer. *PLoS Comput. Biol.*, **11**, e1004095.
23. Groussin,M., Poyet,M., Sistiaga,A., Kearney,S.M., Moniz,K., Noel,M., Hooker,J., Gibbons,S.M., Segurel,L., Froment,A., *et al.* (2021) Elevated rates of horizontal gene transfer in the industrialized human microbiome. *Cell*, **184**, 2053–2067.
24. Song,W., Wemheuer,B., Zhang,S., Steensen,K. and Thomas,T. (2019) MetaCHIP: community-level horizontal gene transfer identification through the combination of best-match and phylogenetic approaches. *Microbiome*, **7**, 36.

25. Seiler,E., Trappe,K. and Renard,B.Y. (2019) Where did you come from, where did you go: refining metagenomic analysis tools for horizontal gene transfer characterisation. *PLoS Comput. Biol.*, **15**, e1007208.
26. Li,C., Jiang,Y. and Li,S. (2019) LEMON: a method to construct the local strains at horizontal gene transfer sites in gut metagenomics. *BMC Bioinformatics*, **20**, 702.
27. Moya,A. and Ferrer,M. (2016) Functional redundancy-induced stability of gut microbiota subjected to disturbance. *Trends Microbiol.*, **24**, 402–413.
28. Vieira-Silva,S., Falony,G., Darzi,Y., Lima-Mendez,G., Garcia Yunta,R., Okuda,S., Vandeputte,D., Valles-Colomer,M., Hildebrand,F., Chaffron,S., *et al.* (2016) Species–function relationships shape ecological properties of the human gut microbiome. *Nat. Microbiol.*, **1**, 16088.
29. Kasmanas,J.C., Bartholomäus,A., Corrêa,F.B., Tal,T., Jehmlich,N., Herberth,G., von Bergen,M., Stadler,P.F., de Leon Ferreira de Carvalho,A.C.P. and Nunes da Rocha,U. (2021) HumanMetagenomeDB: a public repository of curated and standardized metadata for human metagenomes. *Nucleic Acids Res.*, **49**, D743–D750.
30. Pasolli,E., Schiffer,L., Manghi,P., Renson,A., Obenchain,V., Truong,D.T., Beghini,F., Malik,F., Ramos,M., Dowd,J.B., *et al.* (2017) Accessible, curated metagenomic data through ExperimentHub. *Nat. Methods*, **14**, 1023–1024.
31. 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.
32. Dai,D., Zhu,J., Sun,C., Li,M., Liu,J., Wu,S., Ning,K., He,L.-J., Zhao,X.-M. and Chen,W.-H. (2022) GMrepo v2: a curated human gut microbiome database with special focus on disease markers and cross-dataset comparison. *Nucleic Acids Res.*, **50**, D777–D784.
33. Katz,K., Shutov,O., Lapoint,R., Kimelman,M., Brister,J.R. and O'Sullivan,C. (2022) The Sequence Read Archive: a decade more of explosive growth. *Nucleic Acids Res.*, **50**, D387–D390.
34. Leinonen,R., Akhtar,R., Birney,E., Bower,L., Cerdeno-Tárraga,A., Cheng,Y., Cleland,I., Faruque,N., Goodgame,N., Gibson,R., *et al.* (2010) The European Nucleotide Archive. *Nucleic Acids Res.*, **39**, D28–D31.
35. Qin,J., Li,R., Raes,J., Arumugam,M., Burgdorf,K.S., Manichanh,C., Nielsen,T., Pons,N., Levenez,F., Yamada,T., *et al.* (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, **464**, 59–65.
36. Rose,G., Shaw,A.G., Sim,K., Wooldridge,D.J., Li,M.-S., Gharbia,S., Misra,R. and Kroll,J.S. (2017) Antibiotic resistance potential of the healthy preterm infant gut microbiome. *PeerJ*, **5**, e2928.
37. Tanca,A., Abbondio,M., Palomba,A., Fraumene,C., Manghina,V., Cucca,F., Fiorillo,E. and Uzzau,S. (2017) Potential and active functions in the gut microbiota of a healthy human cohort. *Microbiome*, **5**, 79.
38. Gigliucci,F., von Meijenfeldt,F.B., Knijn,A., Michelacci,V., Scavia,G., Minelli,F., Dutilh,B.E., Ahmad,H.M., Raangs,G.C., Friedrich,A.W., *et al.* (2018) Metagenomic characterization of the human intestinal microbiota in fecal samples from STEC-infected patients. *Front. Cell. Infect. Microbiol.*, **8**, 25.
39. Brooks,B., Firek,B.A., Miller,C.S., Sharon,I., Thomas,B.C., Baker,R., Morowitz,M.J. and Banfield,J.F. (2014) Microbes in the neonatal intensive care unit resemble those found in the gut of premature infants. *Microbiome*, **2**, 1.
40. Gevers,D., Kugathasan,S., Denson,L.A., Vázquez-Baeza,Y., Van Treuren,W., Ren,B., Schwager,E., Knights,D., Song,S.J., Yassour,M., *et al.* (2014) The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe*, **15**, 382–392.
41. Lloyd-Price,J., Mahurkar,A., Rahnavard,G., Crabtree,J., Orvis,J., Hall,A.B., Brady,A., Creasy,H.H., McCracken,C., Giglio,M.G., *et al.* (2017) Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature*, **550**, 61–66.
42. Yachida,S., Mizutani,S., Shiroma,H., Shiba,S., Nakajima,T., Sakamoto,T., Watanabe,H., Masuda,K., Nishimoto,Y., Kubo,M., *et al.* (2019) Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nat. Med.*, **25**, 968–976.
43. Zhernakova,A., Kurilshikov,A., Bonder,M.J., Tigchelaar,E.F., Schirmer,M., Vatanen,T., Mujagic,Z., Vila,A.V., Falony,G., Vieira-Silva,S., *et al.* (2016) Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science*, **352**, 565–569.
44. Vich Vila,A., Imhann,F., Collij,V., Jankipersadsing,S.A., Gurry,T., Mujagic,Z., Kurilshikov,A., Bonder,M.J., Jiang,X., Tigchelaar,E.F., *et al.* (2018) Gut microbiota composition and functional changes in inflammatory bowel disease and irritable bowel syndrome. *Sci. Transl. Med.*, **10**, eaap8914.
45. Loman,N.J., Constantinidou,C., Christner,M., Rohde,H., Chan,J. Z.-M., Quick,J., Weir,J.C., Quince,C., Smith,G.P., Betley,J.R., *et al.* (2013) A culture-independent sequence-based metagenomics approach to the investigation of an outbreak of Shiga-toxigenic *Escherichia coli* O104:H4. *JAMA*, **309**, 1502–1510.
46. Nielsen,H.B., Almeida,M., Juncker,A.S., Rasmussen,S., Li,J., Sunagawa,S., Plichta,D.R., Gautier,L., Pedersen,A.G., Le Chatelier,E., *et al.* (2014) Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat. Biotechnol.*, **32**, 822–828.
47. Yassour,M., Lim,M.Y., Yun,H.S., Tickle,T.L., Sung,J., Song,Y.-M., Lee,K., Franzosa,E.A., Morgan,X.C., Gevers,D., *et al.* (2016) Sub-clinical detection of gut microbial biomarkers of obesity and type 2 diabetes. *Genome Med.*, **8**, 17.
48. Karlsson,F.H., Tremaroli,V., Nookaew,I., Bergström,G., Behre,C.J., Fagerberg,B., Nielsen,J. and Bäckhed,F. (2013) Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature*, **498**, 99–103.
49. Le Chatelier,E., Nielsen,T., Qin,J., Prifti,E., Hildebrand,F., Falony,G., Almeida,M., Arumugam,M., Batto,J.-M., Kennedy,S., *et al.* (2013) Richness of human gut microbiome correlates with metabolic markers. *Nature*, **500**, 541–546.
50. Li,J., Jia,H., Cai,X., Zhong,H., Feng,Q., Sunagawa,S., Arumugam,M., Kultima,J.R., Prifti,E., Nielsen,T., *et al.* (2014) An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol.*, **32**, 834–841.
51. Zeller,G., Tap,J., Voigt,A.Y., Sunagawa,S., Kultima,J.R., Costea,P.I., Amiot,A., Böhm,J., Brunetti,F., Habermann,N., *et al.* (2014) Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol. Syst. Biol.*, **10**, 766.
52. Qin,N., Yang,F., Li,A., Prifti,E., Chen,Y., Shao,L., Guo,J., Le Chatelier,E., Yao,J., Wu,L., *et al.* (2014) Alterations of the human gut microbiome in liver cirrhosis. *Nature*, **513**, 59–64.
53. Bäckhed,F., Roswall,J., Peng,Y., Feng,Q., Jia,H., Kovatcheva-Datchary,P., Li,Y., Xia,Y., Xie,H., Zhong,H., *et al.* (2015) Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host Microbe*, **17**, 690–703.
54. Bengtsson-Palme,J., Angelin,M., Huss,M., Kjellqvist,S., Kristiansson,E., Palmgren,H., Larsson,D.J. and Johansson,A. (2015) The human gut microbiome as a transporter of antibiotic resistance genes between continents. *Antimicrob. Agents Chemother.*, **59**, 6551–6560.
55. Feng,Q., Liang,S., Jia,H., Stadlmayr,A., Tang,L., Lan,Z., Zhang,D., Xia,H., Xu,X., Jie,Z., *et al.* (2015) Gut microbiome development along the colorectal adenoma–carcinoma sequence. *Nat. Commun.*, **6**, 6528.
56. Raymond,F., Déraspe,M., Boissinot,M., Bergeron,M.G. and Corbeil,J. (2016) Partial recovery of microbiomes after antibiotic treatment. *Gut Microbes*, **7**, 428–434.
57. David,L.A., Weil,A., Ryan,E.T., Calderwood,S.B., Harris,J.B., Chowdhury,F., Begum,Y., Qadri,F., LaRocque,R.C. and

Turnbaugh,P.J. (2015) Gut microbial succession follows acute secretory diarrhea in humans. *mBio*, **6**, e00381-15.
58. Xie,H., Guo,R., Zhong,H., Feng,Q., Lan,Z., Qin,B., Ward,K.J., Jackson,M.A., Xia,Y., Chen,X., *et al.* (2016) Shotgun metagenomics of 250 adult twins reveals genetic and environmental impacts on the gut microbiome. *Cell Syst.*, **3**, 572–584.
59. Yu,J., Feng,Q., Wong,S.H., Zhang,D., Liang,Q.Y., Qin,Y., Tang,L., Zhao,H., Stenvang,J., Li,Y., *et al.* (2017) Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut*, **66**, 70–78.
60. Zeevi,D., Korem,T., Zmora,N., Israeli,D., Rothschild,D., Weinberger,A., Ben-Yacov,O., Lador,D., Avnit-Sagi,T., Lotan-Pompan,M., *et al.* (2015) Personalized nutrition by prediction of glycemic responses. *Cell*, **163**, 1079–1094.
61. Vrieze,A., Van Nood,E., Holleman,F., Salojärvi,J., Kootte,R.S., Bartelsman,J.F., Dallinga-Thie,G.M., Ackermans,M.T., Serlie,M.J., Oozeer,R., *et al.* (2012) Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome. *Gastroenterology*, **143**, 913–916.
62. Vogtmann,E., Hua,X., Zeller,G., Sunagawa,S., Voigt,A.Y., Hercog,R., Goedert,J.J., Shi,J., Bork,P. and Sinha,R. (2016) Colorectal cancer and the human gut microbiome: reproducibility with whole-genome shotgun sequencing. *PLoS One*, **11**, e0155362.
63. Li,J., Zhao,F., Wang,Y., Chen,J., Tao,J., Tian,G., Wu,S., Liu,W., Cui,Q., Geng,B., *et al.* (2017) Gut microbiota dysbiosis contributes to the development of hypertension. *Microbiome*, **5**, 14.
64. Costea,P.I., Coelho,L.P., Sunagawa,S., Munch,R., Huerta-Cepas,J., Forslund,K., Hildebrand,F., Kushugulova,A., Zeller,G. and Bork,P. (2017) Subspecies in the global human gut microbiome. *Mol. Syst. Biol.*, **13**, 960.
65. Ijaz,U.Z., Quince,C., Hanske,L., Loman,N., Calus,S.T., Bertz,M., Edwards,C.A., Gaya,D.R., Hansen,R., McGrogan,P., *et al.* (2017) The distinct features of microbial 'dysbiosis' of Crohn's disease do not occur to the same extent in their unaffected, genetically-linked kindred. *PLoS One*, **12**, e0172605.
66. Jie,Z., Xia,H., Zhong,S.-L., Feng,Q., Li,S., Liang,S., Zhong,H., Liu,Z., Gao,Y., Zhao,H., *et al.* (2017) The gut microbiome in atherosclerotic cardiovascular disease. *Nat. Commun.*, **8**, 845.
67. Gopalakrishnan,V., Spencer,C.N., Nezi,L., Reuben,A., Andrews,M.C., Karpinets,T.V., Prieto,P., Vicente,D., Hoffman,K., Wei,S.C., *et al.* (2018) Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. *Science*, **359**, 97–103.
68. Lokmer,A., Cian,A., Froment,A., Gantois,N., Viscogliosi,E., Chabé,M. and Segurel,L. (2019) Use of shotgun metagenomics for the identification of protozoa in the gut microbiota of healthy individuals from worldwide populations with various industrialization levels. *PLoS One*, **14**, e0211139.
69. Wirbel,J., Pyl,P.T., Kartal,E., Zych,K., Kashani,A., Milanese,A., Fleck,J.S., Voigt,A.Y., Palleja,A., Ponnudurai,R., *et al.* (2019) Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat. Med.*, **25**, 679–689.
70. Zhu,F., Ju,Y., Wang,W., Wang,Q., Guo,R., Ma,Q., Sun,Q., Fan,Y., Xie,Y., Yang,Z., *et al.* (2020) Metagenome-wide association of gut microbiome features for schizophrenia. *Nat. Commun.*, **11**, 1612.
71. Shao,Y., Forster,S.C., Tsaliki,E., Vervier,K., Strang,A., Simpson,N., Kumar,N., Stares,M.D., Rodger,A., Brocklehurst,P., *et al.* (2019) Stunted microbiota and opportunistic pathogen colonization in caesarean-section birth. *Nature*, **574**, 117–121.
72. Keohane,D.M., Ghosh,T.S., Jeffery,I.B., Molloy,M.G., O'Toole,P.W. and Shanahan,F. (2020) Microbiome and health implications for ethnic minorities after enforced lifestyle changes. *Nat. Med.*, **26**, 1089–1095.
73. Asnicar,F., Berry,S.E., Valdes,A.M., Nguyen,L.H., Piccinno,G., Drew,D.A., Leeming,E., Gibson,R., Le Roy,C., Khatib,H.A., *et al.* (2021) Microbiome connections with host metabolism and habitual diet from 1,098 deeply phenotyped individuals. *Nat. Med.*, **27**, 321–332.
74. The Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome. *Nature*, **486**, 207–214.
75. Qin,J., Li,Y., Cai,Z., Li,S., Zhu,J., Zhang,F., Liang,S., Zhang,W., Guan,Y., Shen,D., *et al.* (2012) A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature*, **490**, 55–60.
76. Cleary,B., Brito,I.L., Huang,K., Gevers,D., Shea,T., Young,S. and Alm,E.J. (2015) Detection of low-abundance bacterial strains in metagenomic datasets by eigengenome partitioning. *Nat. Biotechnol.*, **33**, 1053–1060.
77. Obregon-Tito,A.J., Tito,R.Y., Metcalf,J., Sankaranarayanan,K., Clemente,J.C., Ursell,L.K., Zech Xu,Z., Van Treuren,W., Knight,R., Gaffney,P.M., *et al.* (2015) Subsistence strategies in traditional societies distinguish gut microbiomes. *Nat. Commun.*, **6**, 6505.
78. Kostic,A.D., Gevers,D., Siljander,H., Vatanen,T., Hyötyläinen,T., Hämäläinen,A.-M., Peet,A., Tillmann,V., Pöhö,P., Mattila,I., *et al.* (2015) The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. *Cell Host Microbe*, **17**, 260–273.
79. Rampelli,S., Schnorr,S.L., Consolandi,C., Turroni,S., Severgnini,M., Peano,C., Brigidi,P., Crittenden,A.N., Henry,A.G. and Candela,M. (2015) Metagenome sequencing of the Hadza hunter-gatherer gut microbiota. *Curr. Biol.*, **25**, 1682–1693.
80. Louis,S., Tappu,R.-M., Damms-Machado,A., Huson,D.H. and Bischoff,S.C. (2016) Characterization of the gut microbial community of obese patients following a weight-loss intervention using whole metagenome shotgun sequencing. *PLoS One*, **11**, e0149564.
81. Vincent,C., Miller,M.A., Edens,T.J., Mehrotra,S., Dewar,K. and Manges,A.R. (2016) Bloom and bust: intestinal microbiota dynamics in response to hospital exposures and *Clostridium difficile* colonization or infection. *Microbiome*, **4**, 12.
82. Sankaranarayanan,K., Ozga,A.T., Warinner,C., Tito,R.Y., Obregon-Tito,A.J., Xu,J., Gaffney,P.M., Jervis,L.L., Cox,D., Stephens,L., *et al.* (2015) Gut microbiome diversity among Cheyenne and Arapaho individuals from western Oklahoma. *Curr. Biol.*, **25**, 3161–3169.
83. Pehrsson,E.C., Tsukayama,P., Patel,S., Mejía-Bautista,M., Sosa-Soto,G., Navarrete,K.M., Calderon,M., Cabrera,L., Hoyos-Arango,W., Bertoli,M.T., *et al.* (2016) Interconnected microbiomes and resistomes in low-income human habitats. *Nature*, **533**, 212–216.
84. Schmidt,T.S., Hayward,M.R., Coelho,L.P., Li,S.S., Costea,P.I., Voigt,A.Y., Wirbel,J., Maistrenko,O.M., Alves,R.J., Bergsten,E., *et al.* (2019) Extensive transmission of microbes along the gastrointestinal tract. *eLife*, **8**, e42693.
85. Olm,M.R., Brown,C.T., Brooks,B., Firek,B., Baker,R., Burstein,D., Soenjoyo,K., Thomas,B.C., Morowitz,M. and Banfield,J.F. (2017) Identical bacterial populations colonize premature infant gut, skin, and oral microbiomes and exhibit different *in situ* growth rates. *Genome Res.*, **27**, 601–612.
86. Chu,D.M., Ma,J., Prince,A.L., Antony,K.M., Seferovic,M.D. and Aagaard,K.M. (2017) Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery. *Nat. Med.*, **23**, 314–326.
87. Liu,W., Zhang,J., Wu,C., Cai,S., Huang,W., Chen,J., Xi,X., Liang,Z., Hou,Q., Zhou,B., *et al.* (2016) Unique features of ethnic Mongolian gut microbiome revealed by metagenomic analysis. *Sci. Rep.*, **6**, 34826.
88. Asnicar,F., Manara,S., Zolfo,M., Truong,D.T., Scholz,M., Armanini,F., Ferretti,P., Gorfer,V., Pedrotti,A., Tett,A., *et al.* (2017) Studying vertical microbiome transmission from mothers

to infants by strain-level metagenomic profiling. *mSystems*, **2**, e00164-16.
89. De Filippis,F., Pasolli,E., Tett,A., Tarallo,S., Naccarati,A., De Angelis,M., Neviani,E., Cocolin,L., Gobbetti,M., Segata,N., *et al.* (2019) Distinct genetic and functional traits of human intestinal *Prevotella copri* strains are associated with different habitual diets. *Cell Host Microbe*, **25**, 444–453.
90. Vatanen,T., Kostic,A.D., d'Hennezel,E., Siljander,H., Franzosa,E.A., Yassour,M., Kolde,R., Vlamakis,H., Arthur,T.D., Hämäläinen,A.-M., *et al.* (2016) Variation in microbiome LPS immunogenicity contributes to autoimmunity in humans. *Cell*, **165**, 842–853.
91. Schirmer,M., Smeekens,S.P., Vlamakis,H., Jaeger,M., Oosting,M., Franzosa,E.A., Ter Horst,R., Jansen,T., Jacobs,L., Bonder,M.J., *et al.* (2016) Linking the human gut microbiome to inflammatory cytokine production capacity. *Cell*, **167**, 1125–1136.
92. Mehta,R.S., Abu-Ali,G.S., Drew,D.A., Lloyd-Price,J., Subramanian,A., Lochhead,P., Joshi,A.D., Ivey,K.L., Khalili,H., Brown,G.T., *et al.* (2018) Stability of the human faecal microbiome in a cohort of adult men. *Nat. Microbiol.*, **3**, 347–355.
93. Ferretti,P., Pasolli,E., Tett,A., Asnicar,F., Gorfer,V., Fedi,S., Armanini,F., Truong,D.T., Manara,S., Zolfo,M., *et al.* (2018) Mother-to-infant microbial transmission from different body sites shapes the developing infant gut microbiome. *Cell Host Microbe*, **24**, 133–145.
94. Loomba,R., Seguritan,V., Li,W., Long,T., Klitgord,N., Bhatt,A., Dulai,P.S., Caussy,C., Bettencourt,R., Highlander,S.K., *et al.* (2017) Gut microbiome-based metagenomic signature for non-invasive detection of advanced fibrosis in human nonalcoholic fatty liver disease. *Cell Metab.*, **25**, 1054–1062.
95. Wen,C., Zheng,Z., Shao,T., Liu,L., Xie,Z., Le Chatelier,E., He,Z., Zhong,W., Fan,Y., Zhang,L., *et al.* (2017) Quantitative metagenomics reveals unique gut microbiome biomarkers in ankylosing spondylitis. *Genome Biol.*, **18**, 142.
96. Kieser,S., Sarker,S.A., Berger,B., Sultana,S., Chisti,M.J., Islam,S.B., Foata,F., Porta,N., Betrisey,B., Fournier,C., *et al.* (2018) Antibiotic treatment leads to fecal *Escherichia coli* and coliphage expansion in severely malnourished diarrhea patients. *Cell. Mol. Gastroenterol. Hepatol.*, **5**, 458–460.
97. Nagy-Szakal,D., Williams,B.L., Mishra,N., Che,X., Lee,B., Bateman,L., Klimas,N.G., Komaroff,A.L., Levine,S., Montoya,J.G., *et al.* (2017) Fecal metagenomic profiles in subgroups of patients with myalgic encephalomyelitis/chronic fatigue syndrome. *Microbiome*, **5**, 44.
98. Brooks,B., Olm,M.R., Firek,B.A., Baker,R., Thomas,B.C., Morowitz,M.J. and Banfield,J.F. (2017) Strain-resolved analysis of hospital rooms and infants reveals overlap between the human and room microbiome. *Nat. Commun.*, **8**, 1814.
99. Wampach,L., Heintz-Buschart,A., Fritz,J.V., Ramiro-Garcia,J., Habier,J., Herold,M., Narayanasamy,S., Kaysen,A., Hogan,A.H., Bindl,L., *et al.* (2018) Birth mode is associated with earliest strain-conferred gut microbiome functions and immunostimulatory potential. *Nat. Commun.*, **9**, 5091.
100. Hall,A.B., Yassour,M., Sauk,J., Garner,A., Jiang,X., Arthur,T., Lagoudas,G.K., Vatanen,T., Fornelos,N., Wilson,R., *et al.* (2017) A novel *Ruminococcus gnavus* clade enriched in inflammatory bowel disease patients. *Genome Med.*, **9**, 103.
101. Hannigan,G.D., Duhaime,M.B., Ruffin,M.T. IV, Koumpouras,C.C. and Schloss,P.D. (2018) Diagnostic potential and interactive dynamics of the colorectal cancer virome. *mBio*, **9**, e02248-18.
102. Smits,S.A., Leach,J., Sonnenburg,E.D., Gonzalez,C.G., Lichtman,J.S., Reid,G., Knight,R., Manjurano,A., Changalucha,J., Elias,J.E., *et al.* (2017) Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania. *Science*, **357**, 802–806.
103. Dhakan,D., Maji,A., Sharma,A., Saxena,R., Pulikkan,J., Grace,T., Gomez,A., Scaria,J., Amato,K. and Sharma,V. (2019) The unique composition of Indian gut microbiome, gene catalogue, and associated fecal metabolome deciphered using multi-omics approaches. *Gigascience*, **8**, giz004.
104. Matson,V., Fessler,J., Bao,R., Chongsuwat,T., Zha,Y., Alegre,M.-L., Luke,J.J. and Gajewski,T.F. (2018) The commensal microbiome is associated with anti-PD-1 efficacy in metastatic melanoma patients. *Science*, **359**, 104–108.
105. Rosa,B.A., Supali,T., Gankpala,L., Djuardi,Y., Sartono,E., Zhou,Y., Fischer,K., Martin,J., Tyagi,R., Bolay,F.K., *et al.* (2018) Differential human gut microbiome assemblages during soil-transmitted helminth infections in Indonesia and Liberia. *Microbiome*, **6**, 33.
106. Ye,Z., Zhang,N., Wu,C., Zhang,X., Wang,Q., Huang,X., Du,L., Cao,Q., Tang,J., Zhou,C., *et al.* (2018) A metagenomic study of the gut microbiome in Behcet's disease. *Microbiome*, **6**, 135.
107. Thomas,A.M., Manghi,P., Asnicar,F., Pasolli,E., Armanini,F., Zolfo,M., Beghini,F., Manara,S., Karcher,N., Pozzi,C., *et al.* (2019) Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. *Nat. Med.*, **25**, 667–678.
108. Yassour,M., Jason,E., Hogstrom,L.J., Arthur,T.D., Tripathi,S., Siljander,H., Selvenius,J., Oikarinen,S., Hyöty,H., Virtanen,S.M., *et al.* (2018) Strain-level analysis of mother-to-child bacterial transmission during the first few months of life. *Cell Host Microbe*, **24**, 146–154.
109. Pasolli,E., Asnicar,F., Manara,S., Zolfo,M., Karcher,N., Armanini,F., Beghini,F., Manghi,P., Tett,A., Ghensi,P., *et al.* (2019) Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell*, **176**, 649–662.
110. Hansen,L.B., Roager,H.M., Søndertoft,N.B., Gøbel,R.J., Kristensen,M., Vallès-Colomer,M., Vieira-Silva,S., Ibrügger,S., Lind,M.V., Mærkedahl,R.B., *et al.* (2018) A low-gluten diet induces changes in the intestinal microbiome of healthy Danish adults. *Nat. Commun.*, **9**, 4630.
111. Pasolli,E., Asnicar,F., Manara,S., Zolfo,M., Karcher,N., Armanini,F., Beghini,F., Manghi,P., Tett,A., Ghensi,P., *et al.* (2019) Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell*, **176**, 649–662.
112. Tett,A., Huang,K.D., Asnicar,F., Fehlner-Peach,H., Pasolli,E., Karcher,N., Armanini,F., Manghi,P., Bonham,K., Zolfo,M., *et al.* (2019) The *Prevotella copri* complex comprises four distinct clades underrepresented in westernized populations. *Cell Host Microbe*, **26**, 666–679.
113. Gupta,A., Dhakan,D.B., Maji,A., Saxena,R., PK,V.P., Mahajan,S., Pulikkan,J., Kurian,J., Gomez,A.M., Scaria,J., *et al.* (2019) Association of *Flavonifractor plautii*, a flavonoid-degrading bacterium, with the gut microbiome of colorectal cancer patients in India. *mSystems*, **4**, e00438-19.
114. Kaur,K., Khatri,I., Akhtar,A., Subramanian,S. and Ramya,T. (2020) Metagenomics analysis reveals features unique to Indian distal gut microbiota. *PLoS One*, **15**, e0231197.
115. Rubel,M.A., Abbas,A., Taylor,L.J., Connell,A., Tanes,C., Bittinger,K., Ndze,V.N., Fonsah,J.Y., Ngwang,E., Essiane,A., *et al.* (2020) Lifestyle and the presence of helminths is associated with gut microbiome composition in Cameroonians. *Genome Biol.*, **21**, 122.
116. Yang,J., Li,D., Yang,Z., Dai,W., Feng,X., Liu,Y., Jiang,Y., Li,P., Li,Y., Tang,B., *et al.* (2020) Establishing high-accuracy biomarkers for colorectal cancer by comparing fecal microbiomes in patients with healthy families. *Gut Microbes*, **11**, 918–929.
117. Zeng,Q., Yang,Z., Wang,F., Li,D., Liu,Y., Wang,D., Zhao,X., Li,Y., Wang,Y., Feng,X., *et al.* (2021) Association between metabolic status and gut microbiome in obese populations. *Microb. Genom.*, **7**, 000639.
118. Li,D., Li,Y., Dai,W., Wang,H., Qiu,C., Feng,S., Zhou,Q., Wang,W., Feng,X., Yao,K., *et al.* (2019) Intestinal *Bacteroides* sp. imbalance associated with the occurrence of childhood undernutrition in China. *Front. Microbiol.*, **10**, 2635.

119. Liu,P., Jiang,Y., Gu,S., Xue,Y., Yang,H., Li,Y., Wang,Y., Yan,C., Jia,P., Lin,X., *et al.* (2021) Metagenome-wide association study of gut microbiome revealed potential microbial marker set for diagnosis of pediatric myasthenia gravis. *BMC Med.*, **19**, 159.
120. He,Q., Gao,Y., Jie,Z., Yu,X., Laursen,J.M., Xiao,L., Li,Y., Li,L., Zhang,F., Feng,Q., *et al.* (2017) Two distinct metacommunities characterize the gut microbiota in Crohn's disease patients. *Gigascience*, **6**, gix050.
121. Karlsson,F.H., Fåk,F., Nookaew,I., Tremaroli,V., Fagerberg,B., Petranovic,D., Bäckhed,F. and Nielsen,J. (2012) Symptomatic atherosclerosis is associated with an altered gut metagenome. *Nat. Commun.*, **3**, 1245.
122. Franzosa,E.A., Morgan,X.C., Segata,N., Waldron,L., Reyes,J., Earl,A.M., Giannoukos,G., Boylan,M.R., Ciulla,D., Gevers,D., *et al.* (2014) Relating the metatranscriptome and metagenome of the human gut. *Proc. Natl Acad. Sci. U.S.A.*, **111**, E2329–E2338.
123. Gibson,M.K., Wang,B., Ahmadi,S., Burnham,C.-A.D., Tarr,P.I., Warner,B.B. and Dantas,G. (2016) Developmental dynamics of the preterm infant gut microbiota and antibiotic resistome. *Nat. Microbiol.*, **1**, 16024.
124. Zhang,C., Yin,A., Li,H., Wang,R., Wu,G., Shen,J., Zhang,M., Wang,L., Hou,Y., Ouyang,H., *et al.* (2015) Dietary modulation of gut microbiota contributes to alleviation of both genetic and simple obesity in children. *eBioMedicine*, **2**, 968–984.
125. Schirmer,M., Franzosa,E.A., Lloyd-Price,J., McIver,L.J., Schwager,R., Poon,T.W., Ananthakrishnan,A.N., Andrews,E., Barron,G., Lake,K., *et al.* (2018) Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. *Nat. Microbiol.*, **3**, 337–346.
126. Zhou,W., Sailani,M.R., Contrepois,K., Zhou,Y., Ahadi,S., Leopold,S.R., Zhang,M.J., Rao,V., Avina,M., Mishra,T., *et al.* (2019) Longitudinal multi-omics of host–microbe dynamics in prediabetes. *Nature*, **569**, 663–671.
127. Leydesdorff,L., Comins,J.A., Sorensen,A.A., Bornmann,L. and Hellsten,I. (2016) Cited references and Medical Subject Headings (MeSH) as two different knowledge representations: clustering and mappings at the paper level. *Scientometrics*, **109**, 2077–2091.
128. International Nucleotide Sequence Database Collaboration, Leinonen,R., Sugawara,H. and Shumway,M. (2010) The Sequence Read Archive. *Nucleic Acids Res.*, **39**, D19–D21.
129. 1000 Genome Project Data Processing Subgroup, Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G. and Durbin,R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
130. Chen,S., Zhou,Y., Chen,Y. and Gu,J. (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, **34**, i884–i890.
131. Wood,D.E., Lu,J. and Langmead,B. (2019) Improved metagenomic analysis with Kraken 2. *Genome Biol.*, **20**, 1–13.
132. Almeida,A., Nayfach,S., Boland,M., Strozzi,F., Beracochea,M., Shi,Z.J., Pollard,K.S., Sakharova,E., Parks,D.H., Hugenholtz,P., *et al.* (2021) A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat. Biotechnol.*, **39**, 105–114.
133. Lu,J., Rincon,N., Wood,D.E., Breitwieser,F.P., Pockrandt,C., Langmead,B., Salzberg,S.L. and Steinegger,M. (2022) Metagenome analysis using the Kraken software suite. *Nat. Protoc.*, **17**, 2815–2839.
134. Seemann,T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, **30**, 2068–2069.
135. Kanehisa,M. and Goto,S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.*, **28**, 27–30.
136. Wang,X.-W., Sun,Z., Jia,H., Michel-Mata,S., Angulo,M.T., Dai,L., He,X., Weiss,S.T. and Liu,Y.-Y. (2024) Identifying keystone species in microbial communities using deep learning. *Nat. Ecol. Evol.*, **8**, 22–31.
137. Jiang,X., Hall,A.B., Xavier,R.J. and Alm,E.J. (2019) Comprehensive analysis of chromosomal mobile genetic elements in the gut microbiome reveals phylum-level niche-adaptive gene pools. *PLoS One*, **14**, e0223680.
138. Chen,L., Yang,J., Yu,J., Yao,Z., Sun,L., Shen,Y. and Jin,Q. (2005) VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res.*, **33**, D325–D328.
139. Boratyn,G.M., Camacho,C., Cooper,P.S., Coulouris,G., Fong,A., Ma,N., Madden,T.L., Matten,W.T., McGinnis,S.D., Merezhuk,Y., *et al.* (2013) BLAST: a more efficient report with usability improvements. *Nucleic Acids Res.*, **41**, W29–W33.
140. Jiang,Y., Wang,Y., Che,L., Zhou,Q. and Li,S.C. (2022) GutMeta: online microbiome analysis and interactive visualization with build-in curated human gut microbiome database. bioRxiv doi: https://doi.org/10.1101/2022.09.26.509484, 27 September 2022, preprint: not peer reviewed.
141. Wang,X., Chen,L., Liu,W., Zhang,Y., Liu,D., Zhou,C., Shi,S., Dong,J., Lai,Z., Zhao,B., *et al.* (2023) TIMEDB: tumor immune micro-environment cell composition database with automatic analysis and interactive visualization. *Nucleic Acids Res.*, **51**, D1417–D1424.
142. Wang,R.H., Yang,S., Liu,Z., Zhang,Y., Wang,X., Xu,Z., Wang,J. and Li,S.C. (2024) PhageScope: a well-annotated bacteriophage database with automatic analyses and visualizations. *Nucleic Acids Res.*, **52**, D756–D761.
143. Jia,W., Li,H., Li,S., Chen,L. and Li,S.C. (2020) Oviz-Bio: a web-based platform for interactive cancer genomics data visualization. *Nucleic Acids Res.*, **48**, W415–W426.
144. Schoch,C.L., Ciufo,S., Domrachev,M., Hotton,C.L., Kannan,S., Khovanskaya,R., Leipe,D., Mcveigh,R., O'Neill,K., Robbertse,B., *et al.* (2020) NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database*, **2020**, baaa062.
145. Kanz,C., Aldebert,P., Althorpe,N., Baker,W., Baldwin,A., Bates,K., Browne,P., van den Broek,A., Castro,M., Cochrane,G., *et al.* (2005) The EMBL nucleotide sequence database. *Nucleic Acids Res.*, **33**, D29–D33.
146. Galperin,M.Y., Wolf,Y.I., Makarova,K.S., Vera Alvarez,R., Landsman,D. and Koonin,E.V. (2021) COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res.*, **49**, D274–D281.
147. Zeevi,D., Korem,T., Zmora,N., Israeli,D., Rothschild,D., Weinberger,A., Ben-Yacov,O., Lador,D., Avnit-Sagi,T., Lotan-Pompan,M., *et al.* (2015) Personalized nutrition by prediction of glycemic responses. *Cell*, **163**, 1079–1094.
148. Falony,G., Joossens,M., Vieira-Silva,S., Wang,J., Darzi,Y., Faust,K., Kurilshikov,A., Bonder,M.J., Valles-Colomer,M., Vandeputte,D., *et al.* (2016) Population-level analysis of gut microbiome variation. *Science*, **352**, 560–564.
149. Yatsunenko,T., Rey,F.E., Manary,M.J., Trehan,I., Dominguez-Bello,M.G., Contreras,M., Magris,M., Hidalgo,G., Baldassano,R.N., Anokhin,A.P., *et al.* (2012) Human gut microbiome viewed across age and geography. *Nature*, **486**, 222–227.
150. Garcia-Vallve,S., Guzmán,E., Montero,M. and Romeu,A. (2003) HGT-DB: a database of putative horizontally transferred genes in prokaryotic complete genomes. *Nucleic Acids Res.*, **31**, 187–189.
151. Jeong,H., Sung,S., Kwon,T., Seo,M., Caetano-Anollés,K., Choi,S.H., Cho,S., Nasir,A. and Kim,H. (2016) HGTree: database of horizontally transferred genes determined by tree reconciliation. *Nucleic Acids Res.*, **44**, D610–D619.
152. Choi,Y., Ahn,S., Park,M., Lee,S., Cho,S. and Kim,H. (2023) HGTree v2. 0: a comprehensive database update for horizontal gene transfer (HGT) events detected by the tree-reconciliation method. *Nucleic Acids Res.*, **51**, D1010–D1018.
153. Pasolli,E., Asnicar,F., Manara,S., Zolfo,M., Karcher,N., Armanini,F., Beghini,F., Manghi,P., Tett,A., Ghensi,P., *et al.* (2019) Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell*, **176**, 649–662.

Received: August 15, 2024. **Revised:** October 9, 2024. **Editorial Decision:** October 11, 2024. **Accepted:** October 16, 2024