Million Microbiomes from Human Project (MMHP) White paper

Mission and background of the project

Since the first description of human microbiome published in 2010, the human microbiome field has moved fast from sampling hundreds of individuals to thousands. Now MMHP aims to sequence and analyze one million microbial samples in the next 3 to 5 years to draw a microbiome map of the human body and build the world's largest open access database of the human microbiome. MMHP will focus on the gastrointestinal tract and the oral cavity, but also include other body sites for which a large number of samples can be obtained.

The project was jointly initiated by the Karolinska Institute of Sweden, Shanghai National Clinical Research Center for Metabolic Diseases in China; the University of Copenhagen, Denmark; the Technical University of Denmark; MetaGenoPolis at the Institut National de Recherche pour l'Agriculture, l'alimentation et l'Environnement (INRAE), France; the Latvian Biomedical Research and Study Centre; Institute of Clinical and Preventive Medicine, University of Latvia and Shenzhen BGI Research. The project will mainly rely on MGI's DNBSEQ^{TM} microbial genome sequencing technology to generate human microbial maps of different populations and to establish an open access database at the large-scale population level. The MMHP will not only promote the field, but also lead the field further with collaborators around the world.

Project organization

An organizing committee (project steering committee) is established to decide the route of the project and set up the basic rules for project organizations and regulations. The organizing committee is composed of a representative from each country and 1-2 representative(s) from BGI. The communication will be mainly through emails and regular committee meetings. The appointed members are:

- Sweden: Prof. Lars Engstrand (Chair), Prof. Mathias Uhlén (Co-chair)
- France: Prof. Stanislav Dusko Ehrlich (Co-chair)
- Denmark: Prof. Karsten Kristiansen
- China: Prof. Guang Ning (Co-chair)
- Latvia: Prof. Jānis Kloviņš
- The Netherlands: Prof. John Penders
- BGI: Dr. Huijue Jia (Co-chair)

The life span of the organizing committee would be 2 years, from Jan. 2020 to Jan. 2022. Afterwards, the organizing committee will evolve with the establishments of scientific, executive, advisory, ethics and other necessary MMHP committees. The memberships will expand to future collaborators from the entire community.

Project implementation

Criteria to participate in the project

a. Normally contributing several thousands of samples per year. Except for special cases, the number of samples in a project should exceed 1000.
b. Provide the following metadata:
    1. Person Id (pseudonymized)
    2. Date of sampling
    3. Site of sampling (feces, oral cavity, and other body sites.)
    4. Country of residence
    5. Age (weeks for <6 months old, months for <6 years old, years thereafter)
    6. Gender (sex)
    7. BMI (if available)
    8. Healthy or Disease type (e.g. type 2 diabetes, colorectal cancer, endometriosis, periodontitis, etc.)
    9. Sample collection, processing and DNA extraction protocols. The version and reference of the protocols should be indicated.
    10. Sequencing protocol (platform, pair end, single end, read length)
    11. Protocol for reads QC
    12. A stringent alignment pipeline would be implimented to remove any human genomic sequences before all metagenomic analyses.
    13. The PI is responsible for making sure whether the informed conscent and local legislation allow analyses and reporting of proportions of human reads in each sample, or even genetic associations with the microbiome (which typically require a good coverage of the human genome and is most often sequenced using blood samples or saliva). Such analyses is not a general goal of MMHP.

Sample associated data module

Data	Type	Description
Person Id	Text	Pseudonymized text identifier that is unique ID for the sample in given database
Date of sampling	Date	The date of sample collection following ISO8601 standards (YYYYMMDD)
Site of sampling	Text	Description of the anatomical source of the sampled material (feces, oral cavity, and other body sites)
Country of residence	Predefined text	ISO 3166-1 alpha-2, two letter country code (CN, SE, FR etc.)
Age	Integer	Weeks for <6 months old, months for <6 years old, years thereafter; Unit needs to be indicated (W,M,Y)
Sex	Predefined text	Male, Female, Unknown
BMI	Text	If available, expressed in units of kg/m2 with 1 decimal number

Healthy or Disease type	Text	If available, encoded by ICD-10 (E11 - type 2 diabetes, C18 - colorectal cancer), representing diagnosis at the time or sample obtainment
Sample collection protocol	Text	The version and reference of the protocols should be indicated
Sample processing protocol	Text	The version and reference of the protocols should be indicated.
DNA extraction protocol	Text	The version and reference of the protocols should be indicated.
Sequencing protocol	Text	Including platform, pair end, single end, read length
Protocol for reads QC	Text	The protocol for elimination of human reads including filtering parameters, human genome version, human reads proportion, etc., should be indicated.

Data sharing and release

Data will be deposited in the project database as soon as possible after being generated and will be released to public annually. They will be immediately available to consortium members, notably to prepare a flagship paper accompanying public release.

a. Collaborators agree to release the data to the open-access database at the first annual release after deposition, but can embargo them until the second release, notably to prepare publication(s) stemming from their study.
b. Before data submission, participants should sign an agreement including policies on data open-access and publication with a representative of MMHP.

Note the following:
PIs are encouraged to publish separate papers on their own samples prior to a common flagship publication.

Benefits to participate in the MMHP

a. BGI Riga lab offers sequencing resources including library construction, sequencing and data analysis at an attractive discount. This would be an exclusive offer for the MMHP project.
b. Co-authorship on flagship consortium papers and ancillary consortium-related papers (depending on contribution) .
c. Invitation to annual conference.
d. Newsletter from the consortium .

Project submission

A project proposal and request for MMHP membership have to be send to the MMHP organization team (mmhp@genomics.cn). A proposal template and detailed information about project submission have been attached as Annex 1.

Sample requirement and protocol

a. DNA total Mass: ≥200ng;
b. Concentration: ≥4ng/μl; No degradation or partially degradation
c. Demonstrated Protocols:
    1. MagicPure® Stool and Soil Genomic DNA Kit is recommended for DNA extraction.
    2. SOPs of International Human Microbiome Standards (IHMS) http://www.microbiome-standards.org/#SOPS

DNBseq Sequencing QC and data storage

a. Read length: PE 150/PE100
   General data output is 40 million PE reads for stool samples. Request for higher output is accepted, notably for samples with high human DNA content (e.g. saliva).
a. Q20> 90%; Q30 >80%
b. Data storage/mirror sites: CNGB, INRA, and KI or SciLifeLab will be the preliminary data/mirror sites. More sites could be opened in the future.

Annex 1

Request for Membership in the Million Microbiomes from Human Project (MMHP)

This Request for Membership in the Million Microbiomes from Human Project (MMHP) is intended for completion and submission by PI(s) of large-scale funded microbiome research program(s), who agree to the Principles laid out in the Million Microbiome Human Project (MMHP) White paper document.

Once completed, this document should be sent by email to the MMHP organization team (mmhp@genomics.cn).

Membership to the MMHP is open at any time.

Funding Agency:

Name of funding agency:

Please provide contact information for the funding agency representative who will serve as the primary point of contact for the consortium:

Name:

Email:

Title:

Address:

Phone Number:

Scope of Supported Research:

List the Principal Investigator(s) who are currently funded to perform large-scale human microbiome studies and their respective institutions:

Provide a detailed (not to exceed one page) scientific description of the program, including the overall goal of the program and the goal(s) for individual components. Indicate if there are components of the program that do not qualify as community resource projects and

therefore will not be included in the MMHP. If a Human microbiome Program is being planned at the funding agency level, but is not yet funded, please describe the scope for this program and the timeline on which it will be funded.

Program Data Release Policy:

Describe the data release policy for both the molecular data (sequencing, genotyping, gene expression, etc.) and the clinical data, if there are any. Indicate in which public database the data will be deposited and how access to clinical data will be controlled, if known.