Le French Gut – le microbiote français: a prospective French citizen science study aiming to analyze 100,000 fecal samples with associated nutritional and clinical data (Preprint)

Chloe Connan, Sébastien Fromentin, Mourad Benallaoua, Anne-Sophie Alvarez, Nicolas Pons, Benoît Quinquis, Christian Morabito, Julie-Anne Nazarre, Elise Borezée, Florence Haimet, et al.

Le French Gut – le microbiote français: a prospective French citizen science study aiming to analyze 100,000 fecal samples with associated nutritional and clinical data

Chloe Connan, Sébastien Fromentin, Mourad Benallaoua, Anne-Sophie Alvarez, Nicolas Pons, Benoît Quinquis, Christian Morabito, Julie-Anne Nazarre, Elise Borezée, Florence Haimet, S Dusko Ehrlich, Karine Valeille, Alexandre Cavezza, Hervé Blottière, Patrick Veiga, Mathieu Almeida, Joël Doré, Robert Benamouzig

Submitted to: JMIR Research Protocols
on: July 30, 2024

Table of Contents

Le French Gut – le microbiote français: a prospective French citizen science study aiming to analyze 100,000 fecal samples with associated nutritional and clinical data

Chloe Connan[1]; Sébastien Fromentin[1]; Mourad Benallaoua[2]; Anne-Sophie Alvarez[1]; Nicolas Pons[1]; Benoît Quinquis[1]; Christian Morabito[1]; Julie-Anne Nazarre[3]; Elise Borezée[1, 4]; Florence Haimet[1, 5]; S Dusko Ehrlich[1]; Karine Valeille[1]; Alexandre Cavezza[1]; Hervé Blottière[1, 6]; Patrick Veiga[1, 4]; Mathieu Almeida[1]; Joël Doré[1, 4]; Robert Benamouzig[2]

[1]Université Paris-Saclay, INRAE, MetaGenoPolis (MGP) Jouy-en-Josas FR
[2]Department of Gastroenterology, Avicenne Hospital, Assistance Publique-Hôpitaux de Paris, Université de Paris Bobigny FR
[3]Univ-Lyon, CarMeN Laboratory, Inserm, Inrae, Université Claude Bernard Lyon-1 Lyon FR
[4]Université Paris-Saclay, INRAE, AgroParisTech, Micalis Institute Jouy-en-Josas FR
[5]Mica division Jouy-en-Josas FR
[6]Nantes Université, INRAE, UMR 1280, PhAN Nantes FR

Corresponding Author:
Robert Benamouzig
Department of Gastroenterology, Avicenne Hospital, Assistance Publique-Hôpitaux de Paris, Université de Paris
Avicenne Hospital, Assistance Publique-Hôpitaux de Paris
Bobigny
FR

Abstract

Background: Over the past two decades, the gut microbiota has emerged as a key player in human health, being involved in many different clinical contexts. Yet, many aspects of the relationship with its host are poorly documented. One obstacle is the substantial variability in wet-lab procedures and data processing implemented during gut microbiota studies, which poses a challenge of comparability and potential meta-analysis.

Objective: In order to better understand the relationship between health, dietary habits, and the observed heterogeneity of gut microbiota composition in the general population, « Le French Gut – Le microbiote français » aims to collect, sequence, and analyze 100,000 fecal samples from French residents using a high-quality shotgun metagenomic pipeline, complemented with comprehensive health, lifestyle, and dietary metadata.

Methods: “Le French Gut – Le microbiote français” is a prospective, non-interventional French national study involving individuals , the creation of a biological collection (feces) and the exploitation of data from questionnaires and the National Health Data System (SNDS, Système National des Données de Santé). This national study is open to all metropolitan French adult residents, excluding those who have undergone a colectomy or digestive stoma, or who have had a colonoscopy or taken antibiotics in the last 3 months. This is a home-based trial in which volunteers complete a questionnaire with insights about their health, lifestyle, and dietary habits, and in which stool samples are self-collected.

Results: Recruitment has started in September 2022 and is still on going. As of July 2024, we enrolled 14160 participants.

Conclusions: “Le French Gut” will provide a reference database and new ecosystem tools for understanding the relationship between the gut microbiota, its host and diet. It aims to find new signatures or targets and promote the design of innovative preventive strategies, personalized nutrition, and precision medicine. Clinical Trial: ClinicalTrials.gov NCT05758961. Registered on 8 March 2023. The trial was prospectively registered.

(JMIR Preprints 30/07/2024:64894)
DOI: https://doi.org/10.2196/preprints.64894

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
✔ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

Le French Gut – le microbiote français: a prospective French citizen science study aiming to analyze 100,000 fecal samples with associated nutritional and clinical data

Chloe Connan*[1], Sebastien Fromentin*[1], Mourad Benallaoua[2], Anne-Sophie Alvarez[1], Nicolas Pons[1], Benoît Quinquis[1], Christian Morabito[1], Julie-Anne Nazarre[3], Elise Borezée[1,4], Florence Haimet[1,5], S Dusko Ehrlich[1], Karine Valeille[1], Alexandre Cavezza[1], Hervé M. Blottière[1,6], Patrick Veiga[1,4], Mathieu Almeida[1], Joël Doré[+1,4], Robert Benamouzig[+2].

* Contributed equally

+ Corresponding authors

[1] Université Paris-Saclay, INRAE, MetaGenoPolis (MGP), 78350 Jouy-en-Josas, France

[2] Department of Gastroenterology, Avicenne Hospital, Assistance Publique-Hôpitaux de Paris, Université de Paris, Bobigny, France.

[3] Univ-Lyon, CarMeN Laboratory, Inserm, Inrae, Université Claude Bernard Lyon-1, Lyon, France.

[4] Université Paris-Saclay, INRAE, AgroParisTech, Micalis Institute, 78350 Jouy-en-Josas, France

[5] Mica division, 78350 Jouy-en-Josas, France

[6]

Corresponding authors

Robert Benamouzig

Joël Doré

French Gut Consortium Members

Monique Axelos[1], Sylvie Binda[2], Pierre Cressard[3], Anne-Marie Davila-Gay[4], Christophe d'Enfert[5], Assia Dreux-Zigha[6], Erik Eckhardt[7], Etienne Formstecher[8], Maina Houssaye[9], Milan Lazarevic[10], Sophie Legrain[11], Marie-Emmanuelle Le Guern[12], Françoise Levacon[13], Pedro H. Oliveira[14], Raish Oozeer[15], Karine Roget[16]

[1]INRAE, Paris, France.

[2]Lallemand Health Solutions, Toulouse, France.

[3]Nahibu, Rennes, France.

[4] Université Paris-Saclay, AgroParisTech, INRAE, UMR PNCA, 91120, Palaiseau, France.

[5]Institut Pasteur, Direction Générale Adjointe Scientifique, F-75015 Paris, France.

[6]Greentech, Saint-Beauzire, France.

[7]Adare Biome, Houdan, France.

[8]GMT Science, Rouen, France.

[9]INSERM, Paris, France.

[10]AP-HP, Paris, France.

[11]Gnosis by Lesaffre, Marcq-en-Baroeul, France

[12]Biocodex, Compiegne, France.

[13]Biofortis, Saint-Herblain, France.

[14]Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Université Evry, Université Paris-Saclay, Evry, France.

[15]Danone Nutricia Research, Gif-sur-Yvette, France.

[16]Nextbiome Therapeutics, Clermont-Ferrand, France.

Abstract

Introduction

Over the past two decades, the gut microbiota has emerged as a key player in human health, being involved in many different clinical contexts. Yet, many aspects of the relationship with its host are poorly documented. One obstacle is the substantial variability in wet-lab procedures and data processing implemented during gut microbiota studies, which poses a challenge of comparability and potential meta-analysis. In order to better understand the relationship between health, dietary habits, and the observed heterogeneity of gut microbiota composition in the general population, « Le French Gut – Le microbiote français » aims to collect, sequence, and analyze 100,000 fecal samples from French residents using a high-quality shotgun metagenomic pipeline, complemented with comprehensive health, lifestyle, and dietary metadata.

Methods and analysis

“Le French Gut – Le microbiote français” is a prospective, non-interventional French national study involving individuals , the creation of a biological collection (feces) and the exploitation of data from questionnaires and the National Health Data System (SNDS, Système National des Données de Santé). This national study is open to all metropolitan French adult residents, excluding those who have undergone a colectomy or digestive stoma, or who have had a colonoscopy or taken antibiotics in the last 3 months. This is a home-based trial in which volunteers complete a questionnaire with insights about their health, lifestyle, and dietary habits, and in which stool samples are self-collected.

Results

Recruitment has started in September 2022 and is still on going. As of July 2024, we enrolled 14160 participants.

Discussion

“Le French Gut” will provide a reference database and new ecosystem tools for understanding the relationship between the gut microbiota, its host and diet. It aims to find new signatures or targets

and promote the design of innovative preventive strategies, personalized nutrition, and precision medicine.

Ethic and dissemination

Ethical committee approval under the number 2021-A01439-32 was obtained on October 15th, 2021. Electronic informed consent form to participate will be obtained from all study participants.

Trial registration

ClinicalTrials.gov NCT05758961. Registered on 8 March 2023. The trial was prospectively registered.

Key Words

Gut microbiome, Health, Chronic diseases, Nutrition, Symbiosis

Introduction

The gut microbiota is composed of all the micro-organisms (archaea, bacteria, microeukaryotes, viruses, etc.) that inhabit the intestine. It can be considered as an organ in its own that has co-evolved with its host in a symbiotic relationship, contributing to the physiological homeostasis known as "symbiosis”[1].

The establishment and evolution of the intestinal microbiota throughout life is multifactorial, with the environment and dietary habits playing a major role [2–5]. Advances in research demonstrate how crucial this symbiosis is for an individual's health and well-being. Conversely, an imbalance or “dysbiosis” of the gut microbiota is associated with the risk of onset of various chronic diseases, such as those observed in intestinal and liver diseases (inflammatory bowel disease, irritable bowel syndrome, colorectal cancer, non-alcoholic fatty liver disease, cirrhosis) [4,6,7], obesity [8–10], diabetes [11], autoimmune diseases[12–14] and more recently, in neurological diseases via the gut-brain axis [15–21]. Currently available microbial studies, including clinical and nutritional data in France and worldwide, are based on relatively small sample sizes, usually ranging from a few dozen to a few hundred individuals, and use different sampling and analysis methods, which makes meta-analysis complex.

To gain a better understanding of the interactions between gut microbiota, diet, and health, it is essential to significantly increase the number of samples of gut microbiota processed in a homogenous manner and covering the widest possible spectrum of phenotypes. For this reason, several large-scale initiatives have emerged, such as "The American Gut project," launched in 2012, which involves more than 10,000 participants and more recently the Million Microbiota from Humans Project (MMHP), launched in 2019, which involves several countries around the world and aims to sequence one million metagenomes. Building on this momentum, "Le French Gut" was launched in September 2022 with the aim of describing the gut microbiota and associated nutritional and clinical data of 100,000 major individuals living in France, and the ambition to accelerate

microbiota science, a source of innovation for tomorrow's medicine and nutrition.

Methods and analysis

Aim

By collecting 100,000 gut microbiota profiles combined with metadata on diet, health and lifestyle, " Le French Gut – Le microbiote français " aims at: *i*) better understanding the heterogeneity of healthy gut microbiota and its link with nutrition, lifestyle and anthropometric characteristics; *ii*) predicting changes in gut microbiota associated with diseases such as chronic diseases, neurodevelopmental disorders and neurodegenerative conditions.

Study Design

"Le French Gut – Le microbiote français" is a prospective, non-interventional national study collecting human fecal samples and connecting them with data from self-filling questionnaires and from the National Health Data System (SNDS, Système National des Données de Santé). The trial was conducted according to the recommendations of Standard Protocol Items for Clinical Trials 2013[22]. The project aims at collecting and analyzing 100,000 stool samples from adult volunteers living in France. The duration of this inclusion phase is estimated at 5 years, and the study is planned to run for 20 years, with active (additional questionnaires) and passive follow-up thanks to its link with the SNDS. Ancillary studies could be linked to this main project during this period, including extending it to French overseas departments and territories and to minors.

The study is being carried out in the general population targeting adult volunteers. Participation in the project is free of charge and takes place exclusively via the Le French Gut website (https://lefrenchgut.fr). Volunteers is informed about the study through various communication campaigns (national and local media, as well as social networks) and is invited to register on the dedicated website. Questionnaires and forms can be filled in directly online using a user-friendly web interface where all data and personal information security measures are guaranteed, in particular by

hosting data in a secure sovereign HDS (Health Data Hosting) certified cloud and GDPR (General Data Protection Regulation) compliant. After reading the information note, volunteers is asked to sign an electronic informed consent form. This procedure generates a permanent inclusion identifier.

To become a participant, volunteers must follow the participant journey (Fig. 1). After verifying eligibility for the study, the participant, is asked to complete a 53-item questionnaire about their health, lifestyle, and dietary habits. Once the questionnaire has been completed, a stool sampling kit is automatically sent to the volunteer's address. The volunteer collects the stool sample at home using the kit and then send the stool sample tube to the AP-HP (L'Assistance Publique - Hôpitaux de Paris) laboratory via regular mail using a pre-stamped envelope approved for biological material shipment. Once the integrity and quality of the sample have been checked, the samples is pseudonymized and sent to an INRAE laboratory for biobanking and metagenomic analysis, under a secured ISO 9001 and ISO 27001 certified pipeline.

It will also be possible to add to the basic protocol various customized questionnaires within ancillary protocols concerning the whole population or selected sub-samples according to a particular phenotype (age, gender, area of residence, health status, etc.).

Participants will have access to the project’s collective scientific feedback, on the website or their personal webpage, and will be kept informed of project progress via regular newsletters.

This study has been approved by the French ethics committee Paris Sud-Est IV under number 21.00225.000006 and the "Comité National Informatique et Liberté" (CNIL n°DR-2022-141).

Primary Outcome

The primary outcome is to characterize the heterogeneity and diversity of the gut microbiota of individuals living in France.

The gut microbiota of participants in the 'French Gut' project will be characterized by describing the presence and abundance in the samples of each gut microbial genes listed in our up-to-date gene

catalogue. From these data, the presence and abundance of the different taxonomic ranks will be inferred, ranging from a fine granularity at species and strain level to a more 'zoomed out' phylum level. The richness and diversity of the microbiome (genes and species) will be described. The functional modules, corresponding to the metabolic pathways encoded by the species detected in each sample, will be determined.

Secondary outcomes

The secondary outcomes are to study the variations in the gut microbiota of participant as a function of microbial profile, age, socio-demographic and anthropometric characteristics, lifestyle and eating habits, presence of known pathologies at inclusion, occurrence during follow-up of certain pathologies identified by analysis of data from health questionnaires and SNDS data.

Recruitment

The recruitment of participants is mainly based on a direct approach with "public" communication around the project. Participants learn about the project “Le French Gut” through various communication campaigns, including social media, television, radio, national and regional newspapers, magazines and events. Volunteers can find all the information they need to take part in the project on the website lefrenchgut.fr. Communication also aims to raise awareness on the intestinal microbiota, its link with health and how to care for it through diet. To this end, several communication tools and podcasts are broadcasted to the general public and widely distributed on social media.

Partnerships with existing cohorts will also enable their members to take part in this project. Specific populations, such as members of associations or networks, may be included once the scientific committee has validated the interest of these targeted recruitments. Registration procedures for individuals recruited in this way will be similar to those for the general public.

Eligibility criteria

Le French Gut is a national contribution open to anyone meeting the eligibility criteria. Eligible individuals are adult men and women living in mainland France, with an electronically signed participation agreement. Detailed inclusion and non-inclusion criteria are presented in Table 1. These eligibility criteria may be modified as the project evolves, and mays differe in ancillary projects. .

Table 1: Le French Gut eligibility criteria

Inclusion criteria	Non inclusion criteria
• Adults men and women residing in mainland France • Electronically signed consent form	• Non-adult person • Person not living in France • Persons subject to a protective measure, in particular under guardianship or curatorship or unable to express their consent • Person underwent a colectomy • Person with a digestive stoma • Person who has not signed the consent • Person who did not answer to the inclusion questionnaire • Person who did not send a compliant stool sample • Person having taken antibiotic in the 3 months prior to inclusion • Person who underwent a colonoscopy in the 3 months prior to inclusion

Sample size

Sample size was calculated taking into account the main objective, which is to specify the presence and the abundance of over 2,500 microbial species of the digestive tract and, in particular, the number of subjects required to characterize the enterotype B2[9]. The enterotype B2 is a fairly rare ecological configuration of the microbiota that is most frequently associated with low-grade inflammation, higher prevalence of cardiometabolic conditions, and predictive of non-response to a diet designed to promote weight loss. According to a power test performed on French fecal samples from the MetaCardis cohort, with a population of 100,000 individuals and an expected prevalence of enterotype B2 of 5 %, we will have the statistical power needed to characterize 74 % of the microbiome that makes up the enterotype B2 with an alpha risk of 0.05 and beta risk 0.80. 100,000 participants will provide sufficient statistical power to assess the link between the composition of the gut microbiota and diseases of interest, but also to assess composition of the intestinal microbiota of asymptomatic individuals at high risk of developing a disease. Indeed, diseases of interest include type 2 diabetes (prevalence of 5 % in 2015), ischemic heart disease (3.7 %), depression (5.6 %) but also colorectal cancer (4 % at age 70).

Participant journey

To take part, volunteers need to answer a 53-items online questionnaire about their health, lifestyle and dietary habits, and collect a compliant stool sample. The participant's journey is detailed in figure 1.

Registration and questionnaire

Registration and data collection take place directly on the project website, using an end-to-end encrypted connection, and are stored in an HDS environment (SKEZIA). After checking that the volunteer is eligible and has signed the informed consent form (ICF), the volunteer completes the online questionnaire, which consists of 53 items divided into 3 different parts. The first part of the questionnaire provides socio-demographic data (duration of residence in mainland France, mode of delivery, etc.) as well as lifestyle data (smoking habits, sedentary lifestyle, physical activity). The second part of the questionnaire provides information on anthropometric characteristics (weight, height) used to calculate the body mass index (BMI). This section also collects health data, including chronic diseases. In order to characterize the digestive health of the volunteers, the questionnaire also includes questions relating to gastrointestinal symptoms and factors influencing digestive health. This questionnaire also includes the Bristol Stool Consistency Scale. It also contains questions to define an assessment of participant's mental well-being called the "Perceived Stress Scale" 4 or PSS-4. The third part of the questionnaire is devoted to the dietary habits. Firstly, food exclusions and specific types of diet are addressed. The volunteers then indicate, for each of the food families present in the questionnaire, the frequency of their usual daily or weekly consumption. In addition, the questionnaire focuses on the specific dietary factors that have an impact on the gut microbiota. Full completion of the questionnaire is mandatory to move on to the next stage, i.e. receiving the kit at home. The participant can contact the investigation team at any time to indicate whether any of the eligibility criteria have changed (i.e., antibiotics taken between completion of the questionnaire and sample collection).

As the project progresses, additional questionnaires will be offered to participants who have agreed to be re-contacted for further contribution to scientific explorations. These questionnaires will be more detailed on health, lifestyle, and dietary habits. They will not be mandatory, but they will help us to answer specific scientific questions that emerge as the project progresses.

The validity of the data is verified by input and consistency checks. The questionnaire database is

stored in an HDS environment and the metagenomic data on the MGP's IT infrastructure, as close as possible to the sequencing instruments and the heavy bioinformatics processing.

Stool sampling

After completing the initial questionnaire, a stool sampling kit is sent to the volunteer. The stool sample is collected at home by the volunteer in a tube containing a stabilizing liquid buffer. The date and hour on which the sample was taken is handwritten by the volunteer on the pouch containing the tube. After sampling, the tube is mailed directly to the investigation team via a pre-paid envelope. The investigation team checks the conformity of the sample (e.g. presence of identification data on tube, missing sample, broken tube etc.). Samples that do not pass the initial quality control are flagged up and a new sampling is scheduled. QC-validated samples are stored at -80 °C at the MGP-SAMBO BRC (Biological Resources Centre), which is ISO9001 certified and labelled GIS-IBiSA (CRB N°242), until further processing.

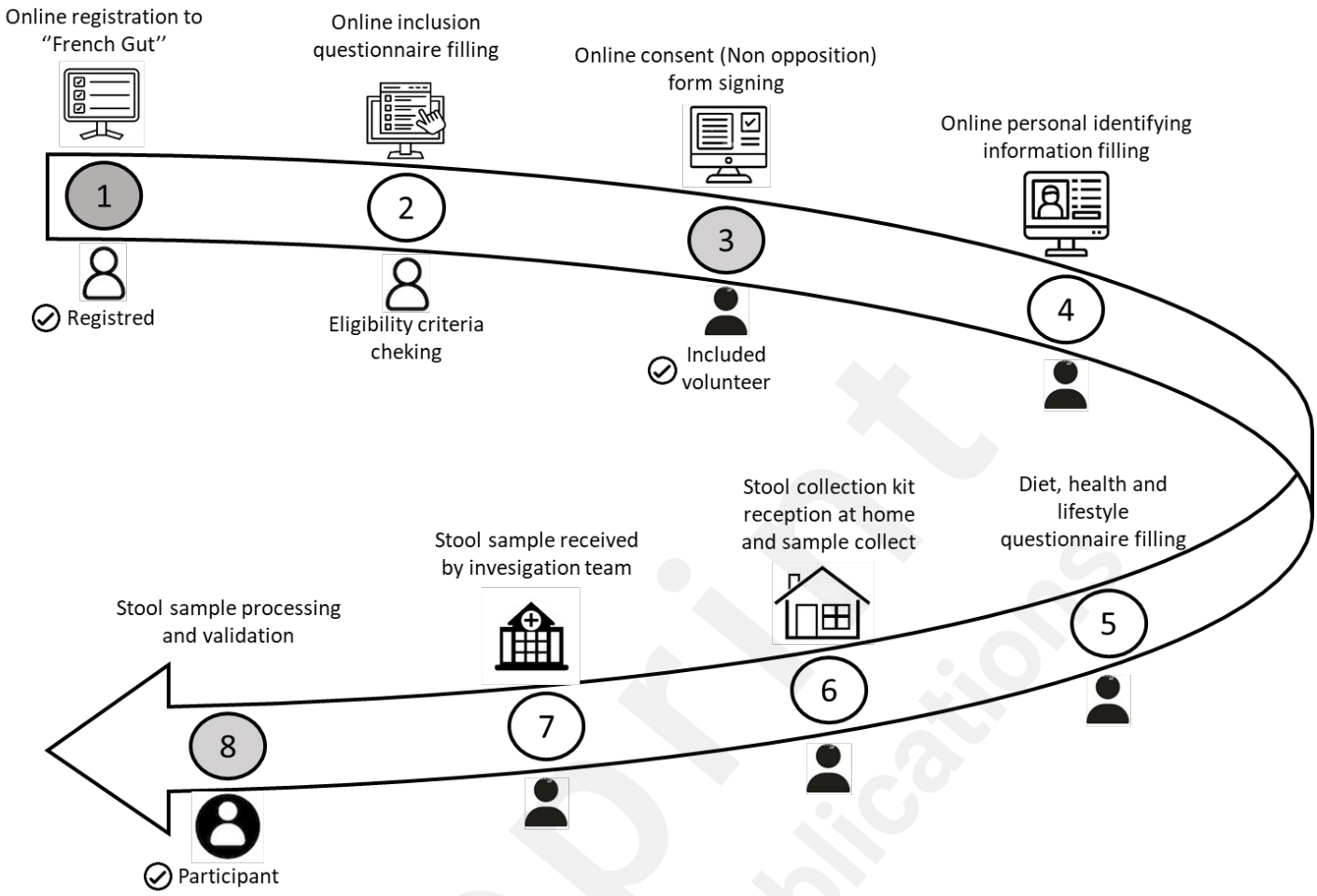

Figure 1: Le French Gut's volunteer timeline. The timeline starts with the profile registration. After filing the non-inclusion form, the non-opposition form, the personal identifying form, the registered becomes a volunteer. The volunteer becomes a participant once the questionnaire has been completed and the stool sample returned to the investigation team.

Stool sample processing

DNA extraction and shotgun sequencing

Total fecal DNA will be extracted using an in-house automated protocol that has been developed and improved on the International Human Microbiome Standards (IHMS) standard procedures (SOP 07). Extracted DNA is evaluated by fluorimetric quantification, purity assessment and fragment size analysis, then if suitable, sequenced using a high-throughput short-read DNA

sequencer. A minimum of 2 x 20 million paired-end reads of 150 nucleotides is produced for each sample.

Microbial gene abundance table and taxonomical profiling

Sequencing data are cleaned to (i) remove remaining sequencing adapters and (ii) trim low quality reads and (iii) discard reads too short (< 60 bp). Reads mapping to the human genome are removed. Microbial gene abundance and taxonomical profiling are then performed using our up-to-date gene and species[23] catalogue of the human gut microbiota.

Contribution of the French national health data system (SNDS)

The French national health data system also referred to as “Système National des Données de Santé” (SNDS) contains individual data for all reimbursements of health expenses for more than 99 % of residents of French territory, i.e. more than 65 million people. This system enables powerful analyses both in terms of the number of subjects involved and the hindsight now available (data collected since 2007). Available data are both administrative and medical. Administrative data are mainly socio-demographic data (age, sex, complementary universal health cover (named CMU-c), etc.). Medical data include, among others, the list of chronic diseases of the volunteer that are 100 % reimbursed by the French medical system or, for example, the exhaustive list of reimbursed drugs that have been prescribed over the years. Collecting the social security number during inclusion allows direct matching between the SNDS data and gut microbiota data from Le French Gut study.

This matching will allow us to gain a deep insight on the connection between the present participant’s health and its gut microbiota. It will also bring light on how the participant’s medical history is connected to its present gut microbiota profile.

Statistical methods

The first statistical step consists in characterizing the data collected on all the included volunteers.

These data will be described globally and by group using standard descriptive statistics (for example mean and standard deviation for continuous variables, contingency table for qualitative variables, etc.). Graphical visualizations coupled with univariate analyzes (mean test, analysis of variance, etc.) and multivariate analysis (Principal Component Analysis (PCA), Factorial Analysis for Mixed Data, network analysis, etc.) will be used to explore the data in order to build coherent groups of variables and identify the overall structuring of the volunteers, atypical data, and possible stratifications.

Concerning univariate analyses, conventional statistical tests (Mann-Whitney-Wilcoxon, *t*, Kruskal-Wallis, and Chi2) will be used to identify variables that differ between previously constituted groups of volunteers, be they microbial, dietary, socio-demographic, or clinical variables. The results of these tests (p-values) will be adjusted with the Benjamini-Hochberg procedure to take into account the multiplicity of tests and control the FDR (false discovery rate). Correlation analyzes will then be used to study bivariate associations between different microbial species of interest and socio-demographic, lifestyle, environmental and nutritional parameters as well as clinical features.

Multivariate analyzes will be used for both descriptive and explanatory purposes. In their descriptive part, they will be used to structure and summarize the information. PCA, factorial correspondence analysis (FCA), clustering, multidimensional positioning and network representation of bacterial species based on co-occurrences will be used to construct synthetic descriptors and to stratify individuals into homogeneous groups. One such example of complexity reduction is the concept of enterotypes, which consists in discretizing the heterogeneity of microbial profiles into a limited number of microbial states [24]. The concept of microbial guilds is another example of a synthetic descriptor, which can be used to reduce the data size. Microbial species in an ecosystem rarely live independently of each other. From their local interactions emerge guilds of species, with similar abundance profiles under varying conditions. In parallel the samples clustering will be used to create subgroups of individuals, homogeneous from a microbiota point of view, without phenotypic *a priori*. These subgroups will then be compared with each other using learning techniques (described

below) or univariate statistical tools (described above) to identify the particularities of each group. In their explanatory part, multivariate analysis methods will be used to predict one or more variables from a set of explanatory variables.

Machine learning and Artificial Intelligence methods (discriminant analysis, random forest, regression, etc.) will be trained on these data, and validated on dedicated test subsets, in order to assess the microbiota's ability to predict a phenotype of interest (presence of a pathology, nutritional profile, exposure to risk, evolution of a condition, etc.).

Integration of the "Le French Gut" project into the Million Microbiome from Human Project (MMHP)

Metagenomic data from the "Le French Gut" will be shared anonymously as part of the collaborative science project (MMHP) which involves many countries worldwide including France, Sweden, Denmark, Latvia, and China as founding members. This project aims at sequencing one million human microbial metagenomes coming from the gut, mouth, skin, and other human organs. It will make available to researchers the metagenomic data associated with the sampling date, the sampling site and protocols as well as a limited amount of metadata: age, sex, body mass index (BMI), geographical location (FR for the present study) and "healthy" or "ill" status (with ICD-10 disease code if available).

Results

Recruitment has started in September 2022 and is still on going. As of July 2024, we enrolled 14160 participants.

Discussion

Le French Gut is a national participatory science project designed to explore the heterogeneity of gut profile within the French resident's population with an unprecedented ambitious statistical power, 100,000 subjects.

One of the main advantages of this protocol is the simplicity of the participant route. Everything can be done at home. This ease of use, coupled with large public communication and awareness campaigns, is an asset towards reaching 100,000 participants.

Like the MMHP, The American Gut, The British Gut, The LifeLines Deep, The 10K, or also The Flemish gut projects, Le French Gut is part of this worldwide willingness to drastically increase the available number of gut microbiota profiles. Le French Gut will constitute an unprecedented database with 100,000 shotgun-sequenced gut microbiota profiled up to the strain levels. In addition to the microbiome data, lifestyle, medical history and an in-depth characterization of participants' dietary habits will be available in the database.

Another main strength of Le French Gut is the high quality standards used to biobank and process all microbial samples. One of the main bias when comparing metagenomics projects is the absence of sample processing information and quality measures between the projects being compared, whether in terms of sampling, storage, extraction or sequencing. Within Le French Gut project, the 100,000 microbiome profiles will be processed with protocols that will follow the International Human Microbiome Standards (IHMS)[25] or higher standards that will arise during the course of the project. Additionally, Le French Gut's flexible recruitment process offers a unique opportunity for synergistic collaborations with other cohorts aimed at deeply characterizing patients or healthy subjects. This strategic approach will facilitate the gathering of valuable data and promote cross-cohort research initiatives. In addition to the basic project, we plan to extend the recruitments to children and teenagers (Le French Gut Kids) and to overseas territories. Le French Gut will help to better understand the organization of the intestinal microbiota and the factors related to its variations.

Owing to its statistical power, Le French Gut will help validate well-known ecosystemic tools (diversity measures, enterotypes etc.) and known discoveries. It will also give the opportunity to deploy at scale novel tools such as microbial guilds, gut microbiota global and local partitioning[26] and new microbial metabolic reconstruction pathways. In conjunction with nutritional data, these ecosystemic tools offer an unparalleled opportunity to explore the connection between diet and the diverse range of ecological states of the human gut microbiota. This will establish the groundwork for precision nutrition approaches aimed at defining diets or supplements tailored to specific gut microbiota states.

Le French Gut will enable scientific community to understand the heterogeneity of microbial profiles within the healthy population, and how the gut microbiota can be a predictor of future health issues. Thanks to the SNDS, gut microbiota profiles will be integrated into each participant's overall past, present and future health landscape. The discovering of biomarkers that could lead to possible new preventive strategies through personalized nutrition, diagnostics, or therapeutics is also expected. Finally, Le French Gut is expected to provide recommendations for more systematic use of the microbiota as a tool to support clinical management.

Trial Status

Le French Gut protocol version is V4-3. The date of the first enrolment is September 15, 2022. The estimated date of primary completion is September 2027. Detailed trial registration data set is presented on Table 2.

Table 2 Le French Gut Trial details

Data Category	Information
Official Title	The French Gut : Le microbiote français
NCT Number	NCT05758961
Other Study ID Numbers	2021-A01439-32
Current Responsible Party	Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement

Current Study Sponsor	Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement
Collaborators	Assistance Publique - Hôpitaux de Paris
Investigators	Principal Investigator: Robert BENAMOUZIG, Pr AP-HP Study Chair: Joël DORE, Pr INRAE
Recruitment Status	Recruiting
Enrollment /Target Sample Size	100000
Study Start Date	2022-09-15
Primary Completion Date	2025-12-31
Study Completion Date	2042-09-15
Eligibility Criteria	Inclusion Criteria: • Men or women over the age of 18 living in France. • Consent form signed electronically. Non inclusion Criteria: • Non-adult person (declarative); • Person not living in France (declarative); • Persons subject to a protective measure, in particular under guardianship or curatorship or unable to express their consent (declarative); • Person having had a colectomy (declarative); • Person with a digestive stoma (declarative); • Person who has not signed the consent; • Person who did not answer the entry questionnaire; • Person who has not sent a compliant stool sample; • Antibiotic intake in the 3 months before inclusion (declarative); • Performing a colonoscopy in the 3 months preceding inclusion (declarative).
Location Countries	France
Study type	Observational
	Heterogeneity and diversity of the gut microbiome of 100,000 subjects residing in France

Primary outcome(s)	Variations of the gut microbiome according to age, socio-demographic and anthropometric characteristics, lifestyle and dietary habits or the presence of known diseases at inclusion
Key secondary outcomes	

Executive and steering committee

The executive committee, led by INRAE, is composed of INRAE staff and AP-HP staff, including the Coordinating Investigator. It meets at least once a month. The executive committee ensures the operational and financial management of the project under the authority of the steering committee. The steering committee is composed of one member per partner. It meets twice a year. The steering committee is the decision-making and arbitration body of the projects.

Acknowledgements

The authors thank all those who have participated in the design, elaboration and supervision of the project.

Authors' contributions

C.C contributed to the design and the regulatory aspects of the research and to the writing of the manuscript. She is involved in the supervision of the study.

S.F contributed to the design, the implementation and the promotion of the research and to the planning and the writing of the manuscript. He is involved in the supervision of the study.

MB contributed to the reviewing and correction of the manuscript. He is involved in the supervision of the study.

ASA contributed to the promotion of the research and the recruitment of volunteers by developing and implementing the communication strategy and to the writing of the manuscript.

NP contributed to the design and implementation of the research. He is involved in the supervision of the study.

BQ contributed to the design of the research.

CM contributed to the design of the research.

EB Contributed to the regulatory aspects of the research.

JAN contributed to the design of the questionnaire.

FH contributed to the design, the implementation and the regulatory aspects of the research.

DE contributed to the design of the research.

KV contributed to the design, the promotion and the funding of the research. She is involved in the supervision of the study.

AC contributed to the design and to the promotion of the research. He is involved in the supervision of the study.

HMB contributed to the design and implementation of the research. He is involved in the supervision of the study.

PV contributed to the design, the implementation and the promotion of the research and to the writing of the manuscript. He is involved in the supervision of the study.

MA contributed to the design, the implementation and the promotion of the research and to the writing of the manuscript. He is involved in the supervision of the study.

JD contributed to the design, the implementation and the promotion of the research and to the writing of the manuscript. He is involved in the supervision of the study.

RB is the principal clinical investigator of the study, contributed to the elaboration of the project and implementation of the research and to the writing of the manuscript. He is involved in the supervision of the study.

All authors have contributed to writing and reading the manuscript and approved the final version of the manuscript

Funding

Le French Gut is supported by MetaGénoPolis, an INRAE structure financially supported by national public fundings (PIA, MetaGenoPolis grant ANR-11-DPBS-0001). Others fundings come from sponsorship from private companies, members of the project consortium. J.D. acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 grant agreement ERC-2017-AdG no. 788191- Homo.symbiosus, and EU research and innovation program SC1-HCO-17-2020 grant agreement no. 964590 - IHMCSA. This work was supported by the Genoscope, the Commissariat à l'Énergie Atomique et aux Énergies Alternatives (CEA), France Génomique (ANR-10-INBS-09–08)

Availability of data and materials

The datasets analyzed during the current study and statistical code are available from the corresponding author on reasonable request, as is the whole protocol.

Ethics approval and consent to participate

Ethical committee approval under the number 2021-A01439-32 was obtained on October 15th, 2021. Electronic informed consent form to participate is obtained from all study participants. Any substantial modification made to the protocol by the promoter, or future ancillary studies must obtain prior to its implementation a favorable opinion from the Ethic committee. The information note and the new version of the informed consent form may be presented, if necessary, especially in the event of a substantial amendment to the protocol.

Consent for publication

Electronic informed consent for publication was obtained from each subject.

Competing interests

The authors declare no conflict of interest.

References

1. van de Guchte, M., Blottière, H. M. & Doré, J. Humans as holobionts: implications for prevention and therapy. *Microbiome* **6**, 81 (2018).
2. Cotillard, A. *et al.* Dietary intervention impact on gut microbial gene richness. *Nature* **500**, 585–588 (2013).
3. Meslier, V. *et al.* Mediterranean diet intervention in overweight and obese subjects lowers plasma cholesterol and causes changes in the gut microbiome and metabolome independently of energy intake. *Gut* **69**, 1258–1268 (2020).
4. Cox, S. R. *et al.* Effects of Low FODMAP Diet on Symptoms, Fecal Microbiome, and Markers of Inflammation in Patients With Quiescent Inflammatory Bowel Disease in a Randomized Trial. *Gastroenterology* **158**, 176-188.e7 (2020).
5. Partula, V. *et al.* Associations between usual diet and gut microbiota composition: results from the Milieu Intérieur cross-sectional study. *Am J Clin Nutr* **109**, 1472–1483 (2019).
6. Tap, J. *et al.* Identification of an Intestinal Microbiota Signature Associated With Severity of Irritable Bowel Syndrome. *Gastroenterology* **152**, 111-123.e8 (2017).
7. Solé, C. *et al.* Alterations in Gut Microbiome in Cirrhosis as Assessed by Quantitative Metagenomics: Relationship With Acute-on-Chronic Liver Failure and Prognosis. *Gastroenterology* **160**, 206-218.e13 (2021).
8. Le Chatelier, E. *et al.* Richness of human gut microbiome correlates with metabolic markers. *Nature* **500**, 541–546 (2013).
9. Vieira-Silva, S. *et al.* Statin therapy is associated with lower prevalence of gut microbiota

dysbiosis. *Nature* **581**, 310–315 (2020).

10. Belda, E. *et al.* Impairment of gut microbial biotin metabolism and host biotin status in severe obesity: effect of biotin and prebiotic supplementation on improved metabolism. *Gut* **71**, 2463–2480 (2022).
11. Forslund, K. *et al.* Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. *Nature* **528**, 262–266 (2015).
12. Breban, M. *et al.* Faecal microbiota study reveals specific dysbiosis in spondyloarthritis. *Ann Rheum Dis* **76**, 1614–1622 (2017).
13. Berland, M. *et al.* Both Disease Activity and HLA-B27 Status Are Associated With Gut Microbiome Dysbiosis in Spondyloarthritis Patients. *Arthritis Rheumatol* **75**, 41–52 (2023).
14. Thirion, F. *et al.* The gut microbiota in multiple sclerosis varies with disease activity. *Genome Med* **15**, 1 (2023).
15. Kelly, J. R. *et al.* Transferring the blues: Depression-associated gut microbiota induces neurobehavioural changes in the rat. *J Psychiatr Res* **82**, 109–118 (2016).
16. Xu, M., Xu, X., Li, J. & Li, F. Association Between Gut Microbiota and Autism Spectrum Disorder: A Systematic Review and Meta-Analysis. *Front Psychiatry* **10**, 473 (2019).
17. Friedland, R. P. & Chapman, M. R. The role of microbial amyloid in neurodegeneration. *PLoS Pathog* **13**, e1006654 (2017).
18. Mancuso, C. & Santangelo, R. Alzheimer's disease and gut microbiota modifications: The long way between preclinical studies and clinical evidence. *Pharmacol Res* **129**, 329–336 (2018).
19. Rosario, D. *et al.* Systematic analysis of gut microbiome reveals the role of bacterial folate and homocysteine metabolism in Parkinson's disease. *Cell Rep* **34**, 108807 (2021).
20. Fan, Y. *et al.* The gut microbiota contributes to the pathogenesis of anorexia nervosa in humans and mice. *Nat Microbiol* **8**, 787–802 (2023).
21. Thirion, F. *et al.* Alteration of Gut Microbiome in Patients With Schizophrenia Indicates

Links Between Bacterial Tyrosine Biosynthesis and Cognitive Dysfunction. *Biological psychiatry global open science* **3**, 283–291 (2023).

22. Chan, A.-W. *et al.* SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. *BMJ* **346**, e7586 (2013).

23. Plaza Oñate, F. *et al.* MSPminer: abundance-based reconstitution of microbial pan-genomes from shotgun metagenomic data. *Bioinformatics (Oxford, England)* **35**, 1544–1552 (2019).

24. Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473**, 174–180 (2011).

25. Costea, P. I. *et al.* Towards standards for human fecal sample processing in metagenomic studies. *Nature Biotechnology* **35**, 1069–1076 (2017).

26. Tap, J. *et al.* Global branches and local states of the human gut microbiome define associations with environmental and intrinsic factors. *Nat Commun* **14**, 3310 (2023).