6-1-2021

Overview of the Microbiome Among Nurses study (Micro-N) as an example of prospective characterization of the microbiome within cohort studies

Christine Everett
Harvard T.H. Chan School of Public Health

Chengchen Li
Harvard T.H. Chan School of Public Health

Jeremy E. Wilkinson
Harvard T.H. Chan School of Public Health

Long H. Nguyen
Harvard T.H. Chan School of Public Health

Lauren J. McIver
Harvard T.H. Chan School of Public Health

See next page for additional authors

Overview of the Microbiome Among Nurses study (Micro-N) as an example of prospective characterization of the microbiome within cohort studies

Christine Everett,[1,2,19] Chengchen Li,[1,3,19] Jeremy E. Wilkinson,[1,3,19] Long H. Nguyen,[1,4,5] Lauren J. McIver,[1,3] Kerry Ivey,[6,7,8] Jacques Izard,[9,10] Natalia Palacios,[1,6,11] A. Heather Eliassen,[1,2,12] Walter C. Willett,[1,2,6,12] Alberto Ascherio,[1,2,6,12] Qi Sun,[1,2,6] Shelley S. Tworoger,[12,13] Andrew T. Chan,[1,2,4,5,14,15,19] Wendy S. Garrett,[1,14,15,16,17,18,19] Curtis Huttenhower,[1,3,14,15,19] Eric B. Rimm,[1,2,6,12,19] and Mingyang Song [1,4,5,6,12,19]

1 Harvard Chan Microbiome in Public Health Center, Harvard T. H. Chan School of Public Health, Boston, MA, USA.
2 Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA.
3 Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA, USA.
4 Clinical and Translational Epidemiology Unit, Mongan Institute, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA.
5 Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA.
6 Department of Nutrition, Harvard T. H. Chan School of Public Health, Boston, MA, USA.

Published in *Nature Protocols*, Vol 16 (June 2021), pp 2724–2731
doi: 10.1038/s41596-021-00519-z

Submitted 19 August 2020; accepted 16 February 2021; published 21 April 2021.

7 South Australian Health and Medical Research Institute, Infection and Immunity Theme, School of Medicine, Flinders University, Adelaide, Australia.
8 Department of Nutrition and Dietetics, College of Nursing and Health Sciences, Flinders University, Adelaide, South Australia, Australia.
9 Food Science and Technology Department, Institute of Agriculture and Natural Resources, University of Nebraska, Lincoln, NE, USA.
10 Fred and Pamela Buffett Cancer Center, University of Nebraska Medical Center, Omaha, Nebraska, USA.
11 Department of Public Health, Zuckerberg College of Health Sciences, University of Massachusetts Lowell, Lowell, MA, USA.
12 Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA, USA.
13 Department of Cancer Epidemiology, Moffitt Cancer Center, Tampa, FL, USA.
14 Broad Institute of Harvard and MIT, Cambridge, MA, USA.
15 Department of Immunology and Infectious Diseases, Harvard T. H. Chan School of Public Health, Boston, MA, USA.
16 Department of Molecular Metabolism, Harvard T. H. Chan School of Public Health, Boston, MA, USA.
17 Department of Medicine, Harvard Medical School, Boston, MA, USA.
18 Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA, USA.
19 These authors contributed equally: Christine Everett, Chengchen Li, Jeremy E. Wilkinson, Andrew T. Chan, Wendy S. Garrett, Curtis Huttenhower, Eric B. Rimm, Mingyang Song.
Correspondence — Mingyang Song, email mingyangsong@mail.harvard.edu

Abstract
A lack of prospective studies has been a major barrier for assessing the role of the microbiome in human health and disease on a population-wide scale. To address this significant knowledge gap, we have launched a large-scale collection targeting fecal and oral microbiome specimens from 20,000 women within the Nurses' Health Study II cohort (the Microbiome Among Nurses study, or Micro-N). Leveraging the rich epidemiologic data that have been repeatedly collected from this cohort since 1989; the established biorepository of archived blood, urine, buccal cell, and tumor tissue specimens; the available genetic and biomarker data; the cohort's ongoing follow-up; and the BIOM-Mass microbiome research platform, Micro-N furnishes unparalleled resources for future prospective studies to interrogate the interplay between host, environmental factors, and the microbiome in human health. These prospectively collected materials will provide much-needed evidence to infer causality in microbiome-associated outcomes, paving the way toward development of microbiota-targeted modulators, preventives, diagnostics and therapeutics. Here, we describe a generalizable, scalable and cost-effective platform used for stool and oral microbiome specimen and metadata collection in the Micro-N study as an example of how prospective studies of the microbiome may be carried out.

Over the past decade, population-scale human microbiome studies have provided tremendous evidence linking microorganisms in the gut and body-wide with various conditions from gastroenterological and periodontal illnesses to cardiovascular, neoplastic, respiratory, and neurologic disorders.[1] However, the majority of human data are limited to cross-sectional studies, making it inherently challenging to differentiate cause from effect, and are prone to reverse causation. Therefore, it is crucial to establish a biobank of samples collected prior to the onset of disease from well-characterized young and middle-aged populations with long-term longitudinal follow-up.[2] Leveraging ongoing large cohort studies provides both a relatively economical option and an ideal setting for the development of such microbiome biobanks, owing to the established infrastructure for biospecimen collection and outcome ascertainment, deep characterization of relevant risk factors, and complementary archival biospecimens that enable future linkage with the microbiome data.

Developing and deploying such platforms for microbiome epidemiology requires the establishment of reliable sample collection methods and associated metadata generation instruments.[3,4] Here, we have accomplished this in the context of the Nurses' Health Study II (NHS II), a leading epidemiologic cohort for studying risk factors and underlying mechanisms for chronic diseases among women.[5] To extend our knowledge about the role of the microbiome in health, we have recently launched a large-scale prospective collection of fecal and oral microbiome samples from 20,000 women in the NHS II, known as the Microbiome Among Nurses (Micro-N) project. In this Perspective, we provide an overview of the Micro-N project and the methods it adopts to leverage generalizable population-scale microbiome collection protocols, which have subsequently been incorporated into the Harvard Chan Microbiome in Public Health (HCMPH) Center's BIOM-Mass (Biobank for Microbiome research in Massachusetts) platform. This platform provides generalizable, scalable, and cost-effective methods for the conduct of prospective studies of the microbiome. We focus on the considerations for several key elements in microbiome specimen collection, including collection kit design, development of a comprehensive questionnaire for assessing potential major determinants of the microbiome, as well as sample shipment, handling, and storage protocols.

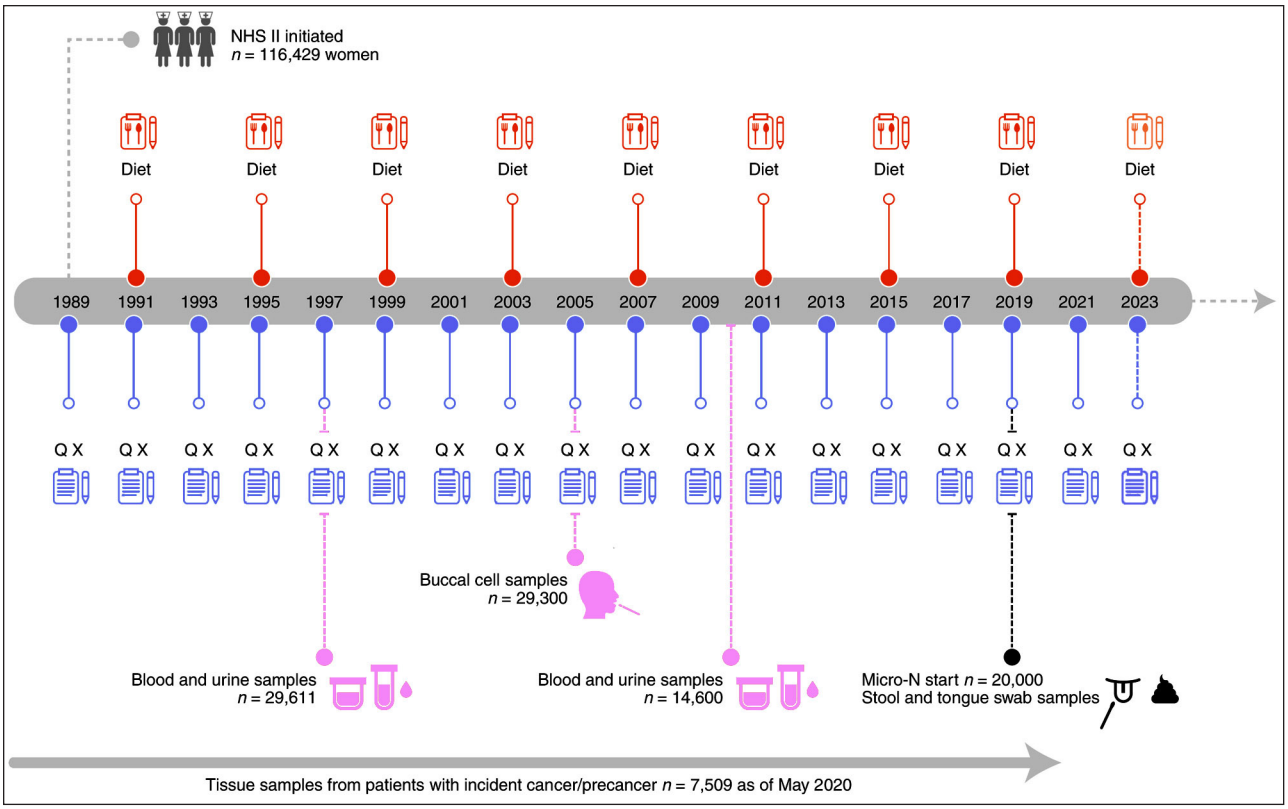

Fig. 1 Overview of the Nurses' Health Study II cohort. Qx: self-administered questionnaire covering a variety of lifestyle factors, health outcomes, and other health-related information; diet: assessed by validated food frequency questionnaire; blood and urine samples were collected from participants in 1996–1999 and again in a subset of those women in 2010–2011; buccal cell samples were collected in 2004–2006 from participants that did not provide blood samples; the collection of microbiome samples began in 2019. Tissue specimens are continuously collected from participants diagnosed with incident cancers or premalignant lesions during the ongoing follow-up.

Overview of the NHS II Cohort

The NHS II is an ongoing prospective cohort of 116,429 female registered nurses residing across the United States who were enrolled in 1989 at the age of 25–42 years (**Fig. 1**).[5] At baseline, these women completed a comprehensive questionnaire on lifestyle and medical factors and have since been followed biennially through mailed questionnaires to collect updated exposure and disease information. Diet was assessed in 1991 and updated every 4 years thereafter via a validated semiquantitative food frequency questionnaire (FFQ). A

Table 1 Outcomes that have been ascertained in the Nurses' Health Study II cohort

Group	*Health conditions*
Cardiovascular disease	Coronary heart disease, stroke, high blood pressure, cardiac arrest, congestive heart failure, arrhythmia, elevated cholesterol, peripheral artery disease, deep vein thrombosis
Cancer	Breast, colon or rectum, endometrium, ovary, melanoma, basal cell skin cancer, squamous cell skin cancer, other cancer
GI disease	Colon or rectal polyp, ulcerative colitis/Crohn's, gastric or duodenal ulcer, Barrett's esophagus, gallstones, cholecystectomy
Respiratory disease	Emphysema/chronic bronchitis, asthma
Metabolic diseases	Diabetes, obesity
Mental and neurological disorders	Multiple sclerosis, Parkinson's disease, depression
Diseases of the genitourinary system	Fibrocystic/other benign breast disease, endometriosis, kidney stones
Immune diseases	Graves' disease/hyperthyroidism, hypothyroidism, hyperparathyroidism, gout, SLE (systemic lupus), rheumatoid arthritis

These outcomes will continue to be tracked in the future, allowing fecal and oral microbiome to be associated cross-sectionally with currently prevalent cases and prospectively with incident cases.

wide array of health conditions has been assessed by the biennial questionnaires in combination with medical record review and/or supplementary questionnaires in the NHS II (**Table 1**). Deaths have been identified through the National Death Index, next-of-kin, and postal authorities. The cumulative follow-up rate of the NHS II cohort is 94%, with approximately 112,000 participants in active follow-up as of May 2020.

In addition to questionnaire data, a comprehensive biorepository has been established in the NHS II. Several types of biospecimens have been collected during follow-up, including blood and urine sample collections in 1996–1999 (n = 29,611) and again in a subset of these women from 2010 to 2011 (n = 14,600); buccal cell samples collected in 2004–2006 from participants who did not provide blood samples (n = 29,300); and tissue specimens from patients diagnosed with incident cancers or premalignant lesions (n = 7,509, as of May 2020). Over the years, these samples have been used for genotyping and assessment of numerous biomarkers through nested case–control studies, as well as deep molecular characterization of cancers.[6–8]

Overview of the Micro-N project

The Micro-N project is designed to integrate microbiome characterization into epidemiologic research within the NHS II. Specifically, it will enable us to prospectively characterize the determinants and health effects of the microbiome, as well as its interactions with environmental and genetic factors in disease development and progression. In Micro-N, stool and tongue swab samples are collected for gut and oral microbiome analysis from a targeted subset of 20,000 women under active follow-up (**Fig. 2**). To maximize the scientific yield, we have prioritized historically underrepresented women and women who have previously contributed other biospecimens (e.g., blood and urine) for genotyping and other profiling assays (e.g., metabolomics) to allow systematic interrogation of host–microbiota interactions. The study was approved by the institutional review boards of the Brigham and Women's Hospital and Harvard T. H. Chan School of Public Health.

In 2017, on the biennial questionnaire, we asked participants if they would be interested in providing a stool and oral sample if the collection was simple and relatively hygienic. Among the 83,695 participants who responded, 55,215 (66%) women answered yes, of whom 42,093 (50%) said they would definitely participate and the other 13,122 (16%) would possibly participate. These responses support the feasibility to recruit and collect microbiome samples from at least 20,000 participants. The feasibility of this large-scale collection is further supported by two previous studies, the Men's Lifestyle Validation Study (MLVS)[9,10] and the Mind Body Study (MBS).[11] The MLVS included 308 men from a parallel cohort, the Health Professionals Follow-up Study (HPFS), who provided up to four stool samples (two per week separated by 6 months) with additional data collected from two 7-day diet records, two physical activity monitors, and two FFQs. The MBS included 233 women from the NHS II who completed a detailed psychosocial assessment and were asked to self-collect stool samples from two consecutive bowel movements (1–3 days apart) at two time points 6 months apart. In the MBS, 213 (91%) women returned the first set of kits and 206 (88%) returned the second set.

Among women from the main NHS II cohort who expressed their willingness to participate, we conducted, via email, an enrollment

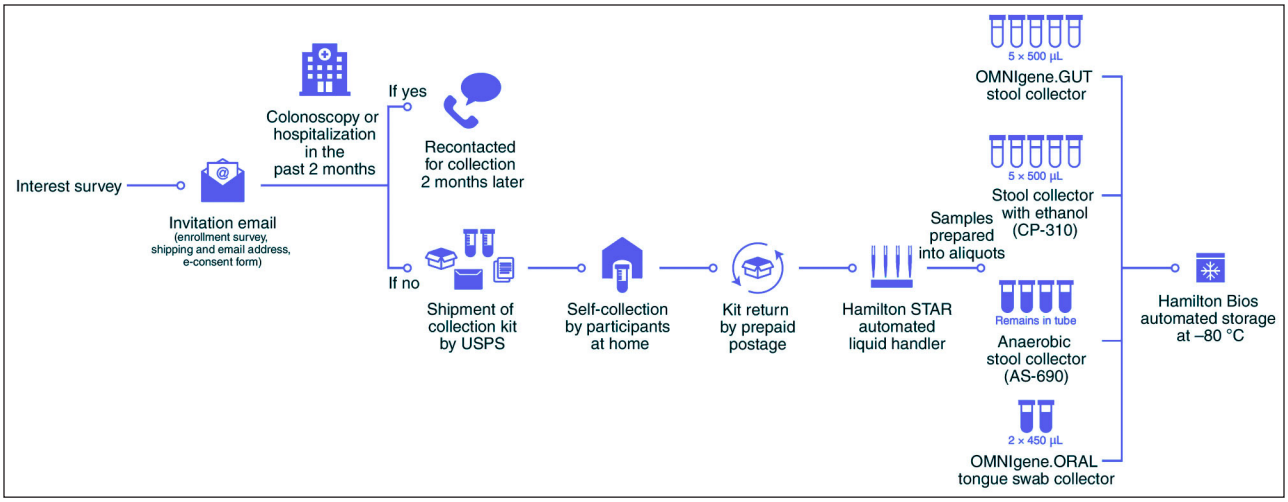

Fig. 2 Workflow of the Micro-N project generalizable by the BIOM-Mass platform. The Micro-N project aims to collect stool and oral microbiome samples from a target of 20,000 women in the NHS II, prioritizing racial minority groups and participants who have previously provided other biospecimens (e.g., blood, urine, and buccal cells). In the process, we have developed a generalizable, scalable, and cost-effective platform for microbiome specimen and metadata collection in large-scale cohorts, the BIOM-Mass platform, implemented as part of the Harvard Chan Microbiome in Public Health Center (HCMPH) and the center's Harvard Chan Microbiome Collection Core (HCMCC). For Micro-N and similar studies employing the same platform, subjects are consented on paper or electronically, after which they are provided with a sampling kit (in person or by mail) containing up to four different specimen collection modalities: stool preserved in 95% ethanol, in an OMNIgene•GUT kit, and viably in Anaerobe Systems liquid dental transport medium, and an oral sample collected with an OMNIgene•ORAL kit. Kits can be returned by participants through standard pre-paid mail, after which each non-anaerobic stool specimen is aliquoted into a target of five ~500-µl subsamples and each oral specimen into a target of two ~500-µl subsamples for molecular assays (amplicon sequencing, metagenomics, metatranscriptomics, and/or metabolomics); and each anaerobic stool sample can be used for downstream microbial isolation, culture, and/or gnotobiotics; or long-term (-80°C) storage.

survey to collect updated mailing and email addresses and to obtain electronic consent (Fig. 2; Supplementary Methods 1). We also asked participants whether they had had any overnight hospital visits or lower endoscopic evaluations within the last 2 months, because of the known influence of hospitalization[12] and bowel preparations[13,14] on the gut microbiota. Participants who reported a hospitalization or colonoscopy in the past 2 months were held in the collection queue and

approached for collection 2 months later. The collection kits with prepaid return postage were shipped by the US Postal Service (kit components are summarized in Supplementary Table 1).

We launched the collection in February 2019 with the support of an infrastructure grant from the Massachusetts Life Sciences Center (MLSC). Our collection began with a phased, 3-month ramp-up period in which we monitored both the rate of enrollment and consent as well as the rate of kit return. In this period, kits were sent to 2,321 women, and 88% returned a kit within 2 months. As of 22 January 2021, we have sent kits to 17,464 women who had consented to participate and received the kits back from 14,731 women. We currently project to complete the collection from 20,000 women by mid- August 2021. Based on the age distribution and disease rate in the NHS II, we projected the number of incident cases of selected disease outcomes in 5, 10 and 15 years after the microbiome specimen collection in Micro-N (**Table 2**). Leveraging the ongoing follow-up of the NHS II cohort, these estimations highlight the potential of Micro-N for future prospective studies elucidating the role of the microbiome in disease incidence.

Table 2 Projected number of incident cases of selected disease outcomes in 5, 10, and 15 years after the microbiome specimen collection in 20,000 NHS II women in Micro-N

Outcome	*No. of incident cases in 5 years*	*No. of incident cases in 10 years*	*No. of incident cases in 15 years*
Diabetes	510	925	1,210
Stroke	168	373	595
Myocardial infarction	149	293	418
Breast cancer	190	408	623
Lung cancer	130	274	410
Colorectal cancer	101	204	299
Endometrial cancer	81	149	197
Ovarian cancer	48	91	127
Pancreatic cancer	30	66	103
Colorectal adenoma	579	1,069	1,397
Inflammatory bowel disease	43	87	124

The age-specific incidence rates for theses outcomes and total mortality rates observed in older women in the NHS were used for the projection. Women in the Micro-N are assumed to be a random sample of women in the overall NHS II cohort.

Collection methods and kits provided by the BIOM-Mass platform as part of the HCMCC

The specimen and metadata collection protocols developed for Micro-N have been formalized under the Harvard Chan Microbiome Collection Core (HCMCC) component of BIOM-Mass, a platform that we developed, which also provides customizable fee-for-service implementations of these kits and processes for other large or small studies. As such, we sought to ensure that both fecal and oral specimen collections were reliable, flexible, and scalable, and that a minimum set of standardized participant information accompanies each collection. Although immediate freezing of stool samples at –20°C or below is often preferable for microbiome preservation, it is not feasible for such large-scale field studies, leading to the development of alternative preservation methods for self-collected samples. Prior studies from our group and others have tested several different preservation methods, including RNAlater (Thermo Fisher), 90+% ethanol, 70% ethanol, OMNIgene•GUT (DNA Genotek Inc.), Zymo DNA/RNA Shield (Zymo Research), fecal occult blood test (FOBT) cards, fecal immunochemical test (FIT) tubes, and Whatman FTA cards (GE Healthcare).[15–21] The predominant findings of these studies indicate that, for the subset of preservative methods meeting a minimum threshold for microbial fixation (i.e., the cessation of substantial metabolic activity), the remaining effects of different preservatives and temperature regimes on the microbial community profiles are small compared to interindividual variability.[15–22] Through the use of preservatives, most of the collection methods are able to prevent major compositional changes in fecal microbial community when exposed to temperature fluctuations from 4°C–40°C over as many as 8 weeks, making them appropriate for home collection and a variety of shipment conditions.[15–22]

For Micro-N, stool and oral samples are thus self-collected by participants using the provided kits. **Table 3** summarizes the applicable downstream assays and stability evidence for the collection kits used in Micro-N. For stool collection, to accommodate substantial biomass for long-term biobanking and a variety of different downstream assays, we provide three different sample tubes. One tube includes 95% ethanol, the second is the commercial OMNIgene•GUT kit, and third is a cryovial pre-filled anaerobically with liquid dental transport medium

Table 3 Applicable downstream assays and stability evidence for the collection kits used in Micro-N

Collection kit	*Applicable downstream assays*					*Stability evidence for field collections*[a]	
	16S rRNA	*Meta-genomics*	*Meta-transcriptomics*	*Meta bolomics*	*Culture/ gnotobiotic studies*	*Tested conditions*	*Ref.*
Stool collection							
95% ethanol	+	+	+	+	–	Up to 8 weeks at 4–40°C	37
OMNIgene•GUT	+	+	+	±	–	28 days at room temperature	38
Anaerobic kit	–	–	–	–	+	NA	NA
Oral sample collection							
OMNIgene•ORAL	+	+	+	–	–	At least a week at room temperature	39

a. Stability data from validation studies showing the most extreme conditions (i.e., the longest interval between sampling collection and processing, and the most extreme temperature conditions).

+, yes; –, no; ±, uncertain.

(LDTM) (Anaerobe Systems). Together, these preservative options enable amplicon, shotgun metagenomic and metatranscriptomic sequencing, stool metabolomic profiling, in addition to future culture and gnotobiotic animal model studies. Participants are asked to collect samples from the same bowel movement for all three tubes (Supplementary Methods 2). Briefly, once the toilet accessory has been affixed to the commode, participants are asked to use the included disposable spatula to transfer a small amount of stool into each collection tube up to a clearly marked, specified target level. After collection, participants are asked to shake the three tightly sealed tubes for 30 seconds to allow adequate mixture of the samples with the stabilizing liquids. Upon finishing this collection, participants complete a brief stool sample questionnaire (Supplementary Methods 3, and see next section for further details) and affix barcode labels to the questionnaire and each tube.

We have validated our ethanol-based self-collection protocol in the context of fecal metagenome and metatranscriptome profiling in a pilot study of men enrolled in the HPFS.[21] Participants self-collected stool at home using both the Human Microbiome Project 1–validated protocol (fresh frozen) and our collection kits using the US Postal Service for sample return. Consistent with 16S rRNA-based studies,[15–20]

we demonstrated that self-collected stool using the preservatives in our sample tubes (one tube with 95% molecular biology grade ethanol and one OMNIgene•GUT tube) provided statistically near-identical metagenomic data to frozen samples.[21] The second iteration of the Human Microbiome Project, the Integrative Human Microbiome Project (iHMP), and specifically the inflammatory bowel disease (IBD) cohort within iHMP, went on to use a similar protocol for self-collected stool, additionally allowing untargeted metabolomic profiling in ethanol-preserved aliquots.[23]

For the accompanying oral sample collection, we use tongue swabs from the OMNIgene•ORAL kit (DNA Genotek Inc.). The tongue microbiome can have somewhat greater microbial-to-human nucleotide ratio variability between individuals but has decreased measurement variability than other self-collectable oral locations (e.g., buccal swab, saliva).[24] Participants are asked to provide a tongue swab sample immediately upon waking on the day following the stool collection and avoid eating, drinking, smoking, using mouthwash, or brushing teeth prior to sample collection (Supplementary Methods 2). Participants use the swab to gently rub the tongue for a minimum of 30 seconds, immediately insert the swab into the bottom of the tube, and then snap the shaft off at the break point, while leaving the swab tip in the tube of liquid. To avoid contamination, participants are instructed not to touch the swab tip to any other surface.

Once the tongue swab collection is finished, participants complete an oral sample questionnaire (Supplementary Methods 3, and see next section for further details) and mail all stool and oral collection kits and questionnaires back to our laboratory using the pre-paid shipping box. Micro-N maintained a dedicated phone helpline and e-mail address for participants who had specific questions, needed to ask for replacement components, or were concerned they were not eligible for participation because of a recent hospital procedure, disease diagnosis, or change in medication use. Among ~12,000 participants who were consented and returned the collection kits by March 2020, we estimate that we received contact queries from approximately 1,200 (10%) women. While Micro-N collection kits include four types of collection tubes, along with redundant copies of consumables (toilet accessories and spatulas), the HCMCC BIOM-Mass's generalized protocol allows cost-effective, modularized configuration of these kit components to enable study customization.

Questionnaire development

To provide a minimum set of essential proximal exposure, outcome, covariate, biometric, and technical information accompanying each sample, we developed a set of standardized questionnaires (for stool and oral collections, respectively) to accompany specimen collection kits. The scannable (Scantron) questionnaires were developed through collaboration with investigators from multiple institutions across the United States that participated in a joint microbiome working group for Micro-N (see Acknowledgements). To avoid overburdening participants and facilitate implementation in similarly time-limited settings, we limited questionnaire length to two pages for stool collection and one page for oral swab collection.

We employed an iterative process for questionnaire development. First, we identified major domains for assessment through literature review of microbiome determinants, complemented by reference to existing questionnaires that were shared within the working group. Seven domains were identified for the stool questionnaire, including the timing of collection, stool consistency, bowel movement pattern, diet, major lifestyle factors, medication use and medical history. For the oral questionnaire, five domains were identified, including the timing of collection, use of the oral hygiene products, natural teeth, history of dental cleaning and surgery, and periodontal disease history. Then, for each of the domains we developed one or more potential questions that were subsequently discussed among the group to refine wording and response options. These discussions also identified the time frames most relevant to the microbiome for each question. After completing the draft questionnaires, we presented them to the working group and made further modifications based on feedback. In the end, a total of 12 overarching questions were included in each of the stool and oral questionnaires (some contained relevant subquestions, e.g., diet and medications; the questionnaires are presented in Supplementary Methods 3; the rationales and considerations for questionnaire development are summarized in the Supplementary Discussion). Of note, the questions did not include those that have already been queried in the main follow-up questionnaires in the NHS II or do not change over time. Major categories of these previously ascertained factors are early-life driving forces for the establishment

of the microbiome, including mode of delivery, breastfeeding, and household exposures (e.g., siblings and pets),[25,26] which have been included on the generalized forms of these questionnaires supported by HCMCC for other population studies.

Sample processing and storage

After return via US mail, kits are received and their components (questionnaires and specimens) tracked using the matched, affixed barcodes. This allows technical considerations such as kit incompleteness or damage to be recorded as well. For all received kits, standardized questionnaires are set aside to be coded, scanned, and programmed with existing algorithms validated in the parent NHS II study. Stool and oral specimens are loaded onto a laboratory information management system (LIMS) (LabVantage)-integrated, Hamilton Microlab STAR liquid handling robot (Fig. 2). This permits pre-separation into multiple ~500-µl aliquots per sample prior to storage, preventing the need for freeze-thaw cycles prior to assays at various times in the future. By matching specimen, questionnaire, LIMS, and aliquot cryovial barcodes, batched aliquot racks are then automatically loaded into a Hamilton BiOS robotic freezer for long-term storage at -80°C.

This storage environment permits automated retrieval of any aliquot subset based on sample type, collection date, subject characteristics, or other information recorded in the Micro-N LIMS or the parent NHS II study. Retrieval is executed and tracked automatically by the freezer, and successfully retrieved aliquots can be transferred directly for experimental work (e.g., for culturing or gnotobiotics), or shipped to molecular data generation facilities (for amplicon sequencing, metagenomics, metatranscriptomics, and/or metabolomics). Critically, this infrastructure permits long-term parent and child sample linkage with their contributing cohort participants and associated information. For example, should a population-level study result in the isolation of a microbial strain of interest, the isolate, its parent specimen, associated molecular data, and the medical history of its original donor are all available through the integrated biorepository, NHS II cohort infrastructure, and BIOM-Mass data portal.

The BIOM-Mass data portal

Finally, the BIOM-Mass Data Portal (http://portal.biom-mass.org) is an in-house data-sharing portal provided by the Harvard Chan Microbiome in Public Health Center (HCMPH) to manage and share microbiome profiles, sample, and population information from microbiome epidemiology studies carried out through the HCMPH BIOM-Mass platform, including Micro-N. It supports both open- and controlled-access dissemination of microbiome multi-omics (16S rRNA gene amplicon profiles, metagenomes, metatranscriptomes, metabolomes, etc.), raw and processed data products (sequences, taxonomic profiles, functional profiles, etc.), and sample and subject covariates (phenotypes, demographics, biometrics, technical protocols, etc.). Data can be shared publicly, controlled- access, or securely protected on a project-specific basis. Only microbial information (i.e., non-human genetic material) are shared, and sensitive covariates can be stripped, linked from an external database such as dbGaP, or secured for individual projects by Google Cloud Platform authentication. The Data Portal builds on technology from the Human Microbiome Project Data Coordinating Center (http://ihmpdcc.org)[1] and the Genomic Data Commons (https://gdc.cancer.gov)[27] and is integrated with the Terra platform (https://terra.bio) for 'omics data dissemination. The BIOM-Mass Data Portal is particularly tailored to provide raw and processed microbiome epidemiological profiles and accompanying phenotypes and covariate annotations, including for large, controlled-access projects such as Micro-N.

Summary and future prospects

The advent of inexpensive, widely available microbial community assays (particularly high-throughput, next-generation sequencing-based approaches) has underscored the extent to which prospectively banked human microbiome specimens would benefit existing long-running cohorts. With Micro-N, BIOM-Mass, the associated protocols, and the resources of the HCMPH, we aim to future-proof the NHS II and other largescale epidemiologic studies for future developments in microbiome science, while also driving near-term discoveries

in mechanism, causality, and public health. Although long-term (and especially early life) exposures are viewed as dominant driving forces for the microbiota's substantial interindividual variation,[28–30] short-term exposures also have critically important consequences (e.g., antibiotics)[31,32] and the specific effects of many such exposures on the microbiome remain to be established. Addressing these significant knowledge gaps requires prospective investigations with high-quality exposure data collected over the life course, with microbial composition and function assessed prior to disease onset.

The Micro-N project's large-scale microbiome specimen collection within the NHS II will provide unparalleled opportunities for future prospective studies to interrogate the role of the microbiome in human health and disease, as well as the interplay between environmental factors, genetics and other host characteristics, and the microbiome in disease development and progression. It has also served to establish a scalable, generalizable, and validated protocol for microbiome epidemiology, implemented by and available from the HCMCC, BIOM-Mass, and the HCMPH. Any such investigations can take advantage of the efficient sampling design of nested case–control studies, in which incident cases diagnosed after specimen collection and their matched controls are identified for microbiome assessments. These data can then be pooled to study the long-term influence on the microbiome of lifestyle, genetic and environmental exposures over the life course. While the NHS II is limited by its inclusion of female participants only, the generalizable protocol we are using allows for future pooled analysis of data from other cohorts and different study populations.[33–35] Moreover, although the current protocol covers sample collection at a single time-point, the established infrastructure in the NHS II cohort allows for future repeated collections from these participants. Finally, 27,706 children of the NHS II participants between the ages of 9 and 17 years have been enrolled and followed up since 1996 in another prospective cohort, the Growing Up Today Study (GUTS)[36] offering opportunities for transgenerational studies. Therefore, we anticipate that the Micro-N project will provide the ideal setting to elucidate how host, environment, and the microbiome interact with each other to influence health, facilitating development of microbiota-targeted preventives, diagnostics, and therapeutics.

* * * *

Acknowledgments This work was supported by the Massachusetts Life Sciences Center (MLSC), the National Institutes of Health (U01 CA176726, R24 DK110499, R00 CA215314, R35 CA253185, R01 CA202704 and R01 CA243454) and the Harvard T. H. Chan School of Public Health. J.I. is supported by Nebraska Tobacco Settlement Biomedical Research Development Funds. We thank staff at the Harvard T. H. Chan School of Public Health for their assistance (A. Spickard, M. Sinunu and S. Branstrator) and the investigators of other cohort studies for their participation of the microbiome working group and contribution to the questionnaire development. These investigators include J. Ahn (New York University), B. Blot (Vanderbilt University), R. Burk (Albert Einstein College of Medicine), M. Hullar (Fred Hutchinson Cancer Research Center), R. Kaplan (Albert Einstein College of Medicine), J. Lampe (Fred Hutchinson Cancer Research Center), L. Le Marchand (University of Hawai'i), K. Meyer (University of North Carolina at Chapel Hill), Q. Qi (Albert Einstein College of Medicine), T. Randolph (Fred Hutchinson Cancer Research Center), H. Sesso (Harvard Medical School/Brigham and Women's Hospital), M. Shrubsole (Vanderbilt University), R. Sinha (National Cancer Institute), E. Vogtmann (National Cancer Institute), L. Wilkens (University of Hawai'i) and W. Zheng (Vanderbilt University). We also thank the participants and staff of the Nurses' Health Study II for their valuable contributions—in particular, B. Hall, A. Scott, S. Al-Shanniek and E. Cornacchio for their dedication to sampling processing and handling and M. Atkinson for his database programming.

Author contributions

Study concept and design: A.H.E., W.C.W., A.T.C., W.S.G., C.H., E.B.R., M.S. Acquisition of data: C.E., C.L., J.E.W., L.H.N., L.J.M., K.I., J.I., N.P., A.H.E., W.C.W., A.A., Q.S., S.S.T., A.T.C., W.S.G., C.H., E.B.R., M.S.

Drafting of the manuscript: A.T. C., W.S.G., C.H., E.B.R., M.S.

Critical revision of the manuscript for important intellectual content: C.E., C.L., J.E.W., L.H.N., L.J.M., K.I., J.I., N.P., A.H.E., W.C.W., A.A., Q.S., S.S.T., A.T.C., W.S.G., C.H., E.B.R., M.S.

Funding acquisition: A.H.E., W. C.W., A.T.C., W.S.G., C.H., E.B.R.

Administrative, technical, or material support: A. T.C., W.S.G., C.H., E.B.R., M.S.

Study supervision: A.T.C., W.S.G., C.H., E.B.R., M.S.

Competing interests The authors declare no competing interests.

Additional information Supplementary information is attached to the online archive record for this article.

References

1. The Integrative HMP (iHMP) Research Network Consortium. The integrative human microbiome project. *Nature* 569, 641–648 (2019).
2. Sinha, R. et al. Next steps in studying the human microbiome and health in prospective studies, Bethesda, MD, May 16–17, 2017. *Microbiome* 6, 210 (2018).

3. Costea, P. I. et al. Towards standards for human fecal sample processing in metagenomic studies. *Nat. Biotechnol.* 35, 1069–1076 (2017).
4. Sinha, R. et al. Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. *Nat. Biotechnol.* 35, 1077–1086 (2017).
5. Bao, Y. et al. Origin, methods, and evolution of the three Nurses' Health Studies. *Am. J. Public Health* 106, 1573–1581 (2016).
6. Ogino, S. et al. Integrative analysis of exogenous, endogenous, tumour and immune factors for precision medicine. *Gut* 67, 1168–1180 (2018).
7. Rice, M. S. et al. Breast cancer research in the Nurses' Health Studies: exposures across the life course. *Am. J. Public Health* 106, 1592–1598 (2016).
8. Townsend, M. K., Aschard, H., De Vivo, I., Michels, K. B. & Kraft, P. Genomics, telomere length, epigenetics, and metabolomics in the Nurses' Health Studies. *Am. J. Public Health* 106, 1663–1668 (2016).
9. Abu-Ali, G. S. et al. Metatranscriptome of human faecal microbial communities in a cohort of adult men. *Nat. Microbiol* 3, 356–366 (2018).
10. Mehta, R. S. et al. Stability of the human faecal microbiome in a cohort of adult men. *Nat. Microbiol* 3, 347–355 (2018).
11. Huang, T. et al. The Mind-Body Study: study design and reproducibility and interrelationships of psychosocial factors in the Nurses' Health Study II. *Cancer Causes Control* 30, 779–790 (2019).
12. Bartosch, S., Fite, A., Macfarlane, G. T. & McMurdo, M. E. Characterization of bacterial communities in feces from healthy elderly volunteers and hospitalized elderly patients by using realtime PCR and effects of antibiotic treatment on the fecal microbiota. *Appl. Environ. Microbiol.* 70, 3575–3581 (2004).
13. Nagata, N. et al. Effects of bowel preparation on the human gut microbiome and metabolome. *Sci. Rep.* 9, 4042 (2019).
14. Jalanka, J. et al. Effects of bowel cleansing on the intestinal microbiota. *Gut* 64, 1562–1568 (2015).
15. Sinha, R. et al. Collecting fecal samples for microbiome analyses in epidemiology studies. *Cancer Epidemiol. Biomark. Prev.* 25, 407–416 (2016).
16. Carroll, I. M., Ringel-Kulka, T., Siddle, J. P., Klaenhammer, T. R. & Ringel, Y. Characterization of the fecal microbiota using high-throughput sequencing reveals a stable microbial community during storage. *PLoS ONE* 7, e46953 (2012).
17. Wu, G. D. et al. Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16S sequence tags. *BMC Microbiol.* 10, 206 (2010).
18. Lauber, C. L., Zhou, N., Gordon, J. I., Knight, R. & Fierer, N. Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated samples. *FEMS Microbiol. Lett.* 307, 80–86 (2010).
19. Dominianni, C., Wu, J., Hayes, R. B. & Ahn, J. Comparison of methods for fecal microbiome biospecimen collection. *BMC Microbiol.* 14, 103 (2014).
20. Choo, J. M., Leong, L. E. & Rogers, G. B. Sample storage conditions significantly influence faecal microbiome profiles. *Sci. Rep.* 5, 16350 (2015).

21. Franzosa, E. A. et al. Relating the metatranscriptome and metagenome of the human gut. *Proc. Natl Acad. Sci. USA* 111, E2329–E2338 (2014).
22. Drew, D. A. et al. Fecal microbiome in epidemiologic studies— letter. *Cancer Epidemiol. Biomark. Prev.* 25, 869 (2016).
23. Lloyd-Price, J. et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* 569, 655–662 (2019).
24. Human Microbiome Project, C. A framework for human microbiome research. *Nature* 486, 215–221 (2012).
25. Dong, T. S. & Gupta, A. Influence of early life, diet, and the environment on the microbiome. *Clin. Gastroenterol. Hepatol.* 17, 231–242 (2019).
26. Stewart, C. J. et al. Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* 562, 583–588 (2018).
27. Jensen, M. A., Ferretti, V., Grossman, R. L. & Staudt, L. M. The NCI Genomic Data Commons as an engine for precision medicine. *Blood* 130, 453–459 (2017).
28. Faith, J. J. et al. The long-term stability of the human gut microbiota. *Science* 341, 1237439 (2013).
29. Lloyd-Price, J. et al. Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature* 550, 61–66 (2017).
30. Wu, G. D. et al. Linking long-term dietary patterns with gut microbial enterotypes. *Science* 334, 105–108 (2011).
31. Ianiro, G., Tilg, H. & Gasbarrini, A. Antibiotics as deep modulators of gut microbiota: between good and evil. *Gut* 65, 1906–1915 (2016).
32. David, L. A. et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature* 505, 559–563 (2014).
33. Tigchelaar, E. F. et al. Cohort profile: LifeLines DEEP, a prospective, general population cohort study in the northern Netherlands: study design and baseline characteristics. *BMJ Open* 5, e006772 (2015).
34. Falony, G. et al. Population-level analysis of gut microbiome variation. *Science* 352, 560–564 (2016).
35. Fu, B. C. et al. Characterization of the gut microbiome in epidemiologic studies: the multiethnic cohort experience. *Ann. Epidemiol.* 26, 373–379 (2016).
36. Rockett, H. R., Berkey, C. S., Field, A. E. & Colditz, G. A. Cross-sectional measurement of nutrient intake among adolescents in 1996. *Prev. Med* 33, 27–37 (2001).
37. Song, S. J. et al. Preservation methods differ in fecal microbiome stability, affecting suitability for field studies. *mSystems* https://doi. org/10.1128/mSystems.00021-16 (2016).
38. Anderson, E. L. et al. A robust ambient temperature collection and stabilization strategy: Enabling worldwide functional studies of the human microbiome. *Sci. Rep.* 6, 31731 (2016).
39. Luo, T. et al. Effects of specimen collection methodologies and storage conditions on the short-term stability of oral microbiome taxonomy. *Appl. Environ. Microbiol.* 82, 5519–5529 (2016).

ORCID

Jeremy E. Wilkinson http://orcid.org/0000-0002-8024-5600
Long H. Nguyen http://orcid.org/0000-0002-5436-4219
Jacques Izard http://orcid.org/0000-0002-5904-5436
Shelley S. Tworoger http://orcid.org/0000-0002-6986-7046
Mingyang Song http://orcid.org/0000-0002-1324-0316