more cognitive assessments administered by the research team. Some researchers record health information related to participants’ conduct during research that is not relevant to study outcomes, such as when a researcher conducting an interview study notes in the research record that an interview has been rescheduled or canceled due to a participant’s illness or accident. In all of these examples, the recorded information presumably falls within the DRS because it was used to make decisions about individuals, but inconsistent institutional policies and practices suggest that there is confusion about these determinations. Without guidance from HHS, there is the potential that institutions will be sanctioned for unintentional violations of HIPAA and that research participants will be denied access to all of the information to which they are entitled.

We therefore propose that HHS clarify the kinds of research data that are eligible for access on request. The guidance should describe use cases involving research data that are part of the DRS, research data that are not part of the DRS (and the justification for their exclusion), and what changes to those use cases would compel different conclusions.

Finally, we propose that the guidance be developed by a task force comprising regulators, researchers, and members of the public. In the past, HHS issued FAQs detailing covered entities’ obligations with respect to the HIPAA access right without robust expert or public input. We believe that such input would improve the value of the guidance to regulated entities and also that public participation in the process is warranted given that the access right was adopted for the public’s benefit. After publishing this guidance, the task force should regularly review and update it to account for emerging practices and technical capabilities that are certain to further expand the kinds of data that are generated and used in research. The public-minded spirit of the HIPAA access right seems to compel no less than a long-term commitment to resolving new questions about access as those data evolve.

Conclusions
Because of a clear trend toward more engagement and transparency with research participants, we should expect more research participants to exercise their HIPAA access right in coming years. The committee’s recommendations ensure that researchers are able to return information meeting high quality standards without fear of violating CLIA. Our additional recommendations help researchers satisfy their HIPAA obligations more efficiently and consistently.

Christi J. Guerrini, Jeffrey R. Botkin and Amy L. McGuire

1 Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, TX, USA.
2 The University of Utah School of Medicine, Salt Lake City, UT, USA.
*e-mail: amcguire@bcm.edu

Published online: 23 July 2019
https://doi.org/10.1038/s41587-019-0190-3

References
1. 45 C.F.R. § 164.524.
2. National Academies of Sciences, Engineering, and Medicine. Returning Individual Research Results to Participants: Guidance for a New Research Paradigm (National Academies Press, 2018).
3. 45 C.F.R. § 160.103.

4. 45 C.F.R. § 164.501.
5. Federal Register 65, 82461–82829 (2000).
6. Department of Health and Human Services. Individuals’ right under HIPAA to access their health information. 45 C.F.R § 164.524. https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/access/index.html#newlyreleasedfaqs (2016).
7. 45 C.F.R. §§ 160.312, 160.402, 160.404.
8. Federal Register 79, 7289–7316 (2014).
9. Meyers, P. CLIA and research results. HHS.gov http://wayback.archive.org/2015/05/082419111343/http://www.hhs.gov/ohrp/sachoap/meyers/clia0111r1r6-ly_g_meyers.pdf (2011).
10. Centers for Medicare and Medicaid Services. Research testing and Clinical Laboratory Improvement Amendments of 1988 (CLIA) regulations. https://www.cms.gov/Regulations-and-Guidance/Legislation/CLIA/Downloads/Research-Testing-and-CLIA.pdf (2014).
11. Wolf, S. M. & Evans, B. J. Return of results and data to research participants. Science 362, 159–160 (2018).
12. Barrie, M. et al. The CLIA-HIPAA conundrum of returning test results to research participants. Medical Research Law & Policy Report https://www.ropesgray.com/~/media/Files/articles/2015/July/2015-07-15-Bloomberg-BNA.ashx (2015).
13. Evans, B. J., Dorschner, M. O., Burke, W. R & Jarvik, G. P. Regulatory changes raise troubling questions for genomic testing. Genet. Med. 16, 799–803 (2014).
14. Secretary’s Advisory Committee on Human Research Protections. Attachment C: return of individual results and special considerations of issues arising from amendments of HIPAA and CLIA. https://www.hhs.gov/ohrp/sachoap-committee/recommendations/2015-september-28-attachment-c/index.html (2015).
15. Evans, B. J. & Wolf, S. M. A Faustian bargain that undermines research participants’ privacy rights and return of results. Florida Law Rev. https://ssrn.com/abstract=1368555 (2019).

Acknowledgements
The views expressed in this article are those of the authors and, except where noted, do not represent the official position of the US National Academies of Sciences, Engineering, and Medicine or of the committee that produced the report discussed in this article.

Author contributions
C.J.G., J.R.B., and A.L.M. contributed to the conception and drafting of the manuscript. C.J.G. and A.L.M. conducted legal analysis.

Competing interests
J.R.B. chaired, and A.L.M. was a member of, the committee that produced the report discussed in this article. C.J.G. was a consultant to the committee.

Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2

To the Editor — Rapid advances in DNA-sequencing and bioinformatics technologies in the past two decades have substantially improved understanding of the microbial world. This growing understanding relates to the vast diversity of microorganisms; how microbiota and microorganisms affect disease; and medical treatment; how microorganisms affect the health of the planet; and; the nascent exploration of the medical, forensic, environmental and agricultural applications of microbiome biotechnology. Much of this work has been driven by marker-gene surveys (for example, bacterial/archaeal 16S RNA genes, fungal internal-transcribed-spacer regions and eukaryotic 18S RNA genes), which profile microbiota with varying degrees of taxonomic specificity and phylogenetic information. The field is now transitioning to integrate other data types, such as metabolite, metaproteome or metatranscriptome profiles.

The QIIME 1 microbiome bioinformatics platform has supported many microbiome studies and gained a broad user and developer community. Interactions with QIIME 1 users in our online support forum, our workshops and direct collaborations have shown the platform’s potential to serve an increasingly diverse array of microbiome researchers in academia, government and industry. Here, we present QIIME 2, a completely reengineered and rewritten system that is expected to...
**Fig. 1 | QIIME 2 provides many interactive visualization tools.** The products of four worked examples are presented here, and interactive versions of these screen captures are available in Supplementary File 1 and at https://github.com/qiime2/paper1. Detailed descriptions and methods, including the commands used to generate each of these visualizations, are provided in Supplementary Methods. 

**a.** Unweighted UniFrac principal coordinate analysis plot containing 37,680 samples, illustrating the scalability of QIIME 2. Colors indicate sample type, as described by the Earth Microbiome Project ontology (EMPO). 

**b.** Interactive taxonomic composition bar plot illustrating the phylum-level composition of microbial-mat samples collected along a temperature gradient in Yellowstone National Park Hot Spring outflow channels (Steep Cone Geyser). The many interactive controls available in this plot vastly decrease the burden of exploratory analysis over QIIME 1. 

**c.** Feature volatility plot (https://msystems.asm.org/content/3/6/e00219-18) illustrating the change in *Bifidobacterium* abundance over time in breast-fed and formula-fed infants. Temporally interesting features can be interactively discovered with this visualization. Bar charts rank the importance (predictive power for time point) and mean abundance of all microbial features. These bar charts provide an interface for visualizing volatility plots (line plots) of individual features in the context of their importance and abundance; clicking on a bar will display the volatility plot of that feature and highlight in blue that feature’s importance and abundance in the bar charts below. 

**d.** Molecular cartography of the human skin surface. Colored spots represent the abundance of the small-molecule cosmetic ingredient sodium laureth sulfate on the human skin. Sample data can be interactively visualized in three-dimensional models, thus supporting the discovery of spatial patterns.
facilitate reproducible and modular analysis of microbiome data to enable the next generation of microbiome science.

QIIME 2 was developed on the basis of a plugin architecture (Supplementary Fig. 1) that allows third parties to contribute functionality (https://library.qiime2.org). QIIME 2 plugins exist for latest-generation tools for sequence quality control from different sequencing platforms (DADA2 (ref. 11) and Deblur12), taxonomy assignment13 and phylogenetic insertion14, which quantitatively improve the results over QIIME 1 and other tools (as detailed in the corresponding tool-specific publications). The plugins also support qualitatively new functionality, including microbiome paired-sample and time-series analysis16 (which are critical for studying the effects of treatments on the microbiome), and machine learning17. Trained machine learning models can be saved for application to new data and interrogated to identify important microbiome features. Several recently released plugins, including q2-cscs15, q2-metabolomics18, q2-shogun19, q2-metaphlan2 (ref. 20) and q2-picrust2 (ref. 21), provide initial support for analysis of metabolomics and shotgun metagenomics data. We are currently working with teams developing bioinformatics tools for metatranscriptomics and metaproteomics, and we expect to add new plugins supporting these data types to the ecosystem shortly. Additionally, many of the existing ‘downstream’ analysis tools, such as q2-sample-classifier16, can already work with these data types individually or in combination if they are provided in a feature table. Thus, QIIME 2 has the potential to serve not only as a marker-gen analysis tool but also as a multidimensional and powerful data science platform that can be rapidly adapted to analyze diverse microbiome features.

QIIME 2 provides many new interactive visualization tools facilitating exploratory analyses and result reporting. Static versions of interactive visualizations resulting from four worked examples are provided in Fig. 1. QIIME 2 View (https://view.qiime2.org) is a unique new service (Supplementary Methods) that allows users to securely share analyses and result reporting. Static versions of the visualizations presented in Fig. 1 are provided in Supplementary File 1 to allow readers to interact with QIIME 2 View. Corresponding worked QIIME 2 example code is provided in the Supplementary Methods.

Reproducibility, transparency and clarity of microbiome data science are guiding principles in QIIME 2 design. To this end, QIIME 2 includes a decentralized data-provenance tracking system: details of all analysis steps with references to intermediate data are automatically stored in the Data-provenance Tracking system (Supplementary File 1). The resulting visualization (Fig. 1c) has the complete data provenance and correctly identifies shared processing of inputs. This simplified representation was created manually from the complete provenance graph for the purpose of illustration. An interactive and complete version of this provenance graph (as well as those for other Fig. 1 panels) can be accessed through Supplementary File 1. The provenance is no longer reliable and the results no longer contain information enabling reproducibility. The provenance of the visualizations presented in Fig. 1 can be interactively reviewed by loading the contents of Supplementary File 1 with QIIME 2 View, providing far more detailed information than can typically be provided in Methods text. QIIME 2 results are also semantically typed (Fig. 2), and actions indicate acceptable input types, clarifying the data that actions should be applied to and making complex workflows less error prone. Complex workflows can be created and shared by using Jupyter Notebooks22 or Common Workflow Language (CWL)23, and support for other workflow engines is currently in development.

Finally, QIIME 2 provides a software-development kit (https://dev.qiime2.org) that can be used to integrate it as a component of other systems (such as QiTria24 or Illumina BaseSpace) and to develop interfaces targeted toward users with different levels of computational sophistication (Supplementary Fig. 2). QIIME 2 provides the QIIME 2 Studio graphical user interface and QIIME 2 View, interfaces designed for end-user biologists, clinicians and policy-makers; the QIIME 2 application programming interface, designed for data scientists who want to automate workflows or work interactively in Jupyter Notebooks25; and q2cli and q2cwl wrappers for QIIME 2, designed for experts in high-performance computing. At present, computationally expensive steps support parallel computing at the individual-action level (for example, many actions including de-noising and taxonomy assignment support multiple threads).

We are currently developing deeper integration with parallelism strategies available in third-party workflow engines, and workflow-level parallelism is currently possible through CWL.

There are many other powerful open-source software tools for microbiome data science, including mothur25, phyloseq26 and related tools available through Bioconductor27, and the biobakery suite28,29,30. The microbiome bioinformatics platform mothur is often compared to QIIME 1 and QIIME 2. A major difference between mothur and QIIME lies in the interactive visualizations: QIIME 2 provides many interactive visualization tools (several examples are provided in Fig. 1), whereas mothur focuses on generating data that can be easily loaded and visualized with other tools. The phyloseq tool focuses on microbiome statistical analysis and generating publication-ready visualizations.
but, unlike QIIME 2, begins with a feature or operational-taxonomic-unit table, leaving ‘upstream’ processing steps, such as sequence demultiplexing and quality control, to other processing pipelines, many of which (like phyloseq) are available through Bioconductor. The biobakery suite provides analytic functionality that complements that of QIIME 2, and we are actively working with biobakery developers to support interoperability by making their tools accessible as QIIME 2 plugins (for example, the q2-metaphlan2 plugin allows users to run MetaPhlAn2 through QIIME 2). QIIME 2 provides the only Python-based microbiome data-science platform that supports retrospective data-provenance tracking to ensure reproducibility, multi-omics analysis support, interfaces geared toward different user types to enhance usability and an extensibility-focused design through the plugin architecture and software-development kit. We share feedback from users of QIIME 2 on these and other features in Supplementary Methods.

The tools described in the preceding paragraph are all interoperable through plugins, exchange of files in standard formats or using multi-language environments, such as Jupyter Notebooks. For example, the BIOM format is supported by all of them. A diverse ecosystem of interoperable software is beneficial for the field, because it allows both experienced users to obtain multiple perspectives on their data and novice bioinformaticians to work in the programming environments that they are most comfortable with (for example, phyloseq allows users to work in R, whereas QIIME 2 allows users to work in Python). We plan to continue working with the developers of these tools, and with organizations such as the Genomics Standards Consortium, on plugins and standards to ensure interoperability, as well as developing tools to automatically import data from microbiome data-sharing platforms such as Qita, the European Bioinformatics Institute (EBI) European Read Archive and the National Center for Biotechnology Information (NCBI) Sequence Read Archive.

Advances in microbiome research promise to improve many aspects of health and the world, and QIIME 2 will help drive those advances by enabling accessible, community-driven microbiome data science.

Data availability
Data for the analyses presented in Fig. 1 are available as follows: Earth Microbiome Project data in Fig. 1a were obtained from ftp://ftp.microbio.me/emp/release1, and the American Gut Project (AGP) data were obtained from Qita (http://qiita.microbio.me) study ID 10317. Sequence data in Fig. 1b are available in Qita under study ID 10249 and the EBI under accession number ERP016173. Sequence data in Fig. 1c are available in Qita under study ID 925 and the EBI accession number ERP022167. Data in Fig. 1d are available in the q2-ili GitHub repository (https://github.com/biocore/q2-ili). Interactive versions of the Fig. 1 visualizations can be accessed at https://github.com/qiime2/paper1.

Code availability
QIIME 2 is open source and free for all use, including commercial. It is licensed under a BSD three-clause license. Source code is available at https://github.com/qiime2. Help for QIIME 2 is provided at https://forum.qiime2.org.

Editor’s Note: This paper has been peer-reviewed.
documentation, educational materials and/or user/developer support content. E.B., J.R.R., M.R.D., N.A.B., R.K. and J.G.C. wrote the manuscript; all authors assisted with revision of the manuscript. E.B., J.R.R., M.R.D., N.A.B. and J.G.C. designed and developed the QIIME 2 framework. D.M.D., A.G., R.L., E.L., S.C.M., R.S., J.R.S., W.W., C.H.D.W. and R.K. contributed data used in the manuscript and/or testing of QIIME 2. C.C.A., C.T.B., E.K.C., P.C.D., S.H., P.K., E.L., T.P., R.S., E.V., Y.W. and R.K. contributed to the design of analytical methods. E.B., J.R.R., M.R.D., N.A.B., G.A.A.-G., H.A., E.J.A., M.A., F.A., K.B., A.B., B.J.C., J.C., G.M.D., G. Duvallet, M. Ernst, J.E., A.G., K.G., J.G., S.M.G., B.H., H.H., C.H., G.H., S.J., L.J., B.D.K., C.R.K., D.K., I.K., M.G.I.L., C.L., M.M., C.M., B.D.M., D.M., L.J.M., J.T.M., A.T.N., J.A.N.-M., S.L.P., M.L.P., E.P., L.B.R., A.R., M.S.R., P.R., N.S., M.S., P.T., A.T., J.J.v.d.H., Y.V.-B., M.V., M.W., K.C.W., A.D.W., Z.Z.X., J.R.Z., Y.Z., Q.Z. and J.G.C. contributed software to QIIME 2 plugins, interfaces, framework and/or build and test systems.

Additional information

**Supplementary information** is available for this paper at https://doi.org/10.1038/s41587-019-0209-9.