bial communities are strongly patterned by a continuous gradient between relative abundance of *Prevotella copri* and *Bacteroides spp* (**Fig. 1**, example 2) with little or no support for any discrete clusters (**Supplementary Fig. 1**). This is consistent with the analysis of Koren *et al.*[5] but stands in contrast with the three-enterotypes hypothesis of Arumugam *et al.*[6]. Third, we visualized the continuum of the fir-micutes–bacteroidetes gradient in gut microbiomes (**Fig. 1**, example 3). This continuum has been previously reported[4], but not for thousands of microbial species on thousands of individuals as analyzed here. Finally, we ranked all taxa–pathway pairs by magnitude of correlation in samples. The highest correlation pair shown demonstrates a strong relationship between *Prevotella copri* abundance and inosine 5 phosphate biosynthesis (**Fig. 1**, example 4), and this suggests functional differences along the gradient shown in example 2. These and other analyses (**Supplementary Figs. 1–5**) would be very large undertakings using less curated databases such as IMG/M or EBI Metagenomics; but they are straightforward, documented, and reproducible using curatedMetagenomicData.

Our large-scale, curated integration of metagenomic data is well documented and readily usable by broad scientific communities for efficient hypothesis testing and methods development. The automated pipeline allows the resource to be continually expanded by the team and community contributors (**Supplementary Methods**, package maintenance). By allowing researchers to bring their expertise to the analysis of metagenomic data without the need for extensive bioinformatic experience, curatedMetagenomicData greatly expands the accessibility of public data for study of the human microbiome.

Data availability statement. The curatedMetagenomicData data package is available online at https://waldronlab.github.io/curated-MetagenomicData/ and through the Bioconductor package installer.

A **Life Sciences Reporting Summary** is available.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper (doi:10.1038/nmeth.4468).

Edoardo Pasolli[1,8], Lucas Schiffer[2,3,8], Paolo Manghi[1,8], Audrey Renson[2,3], Valerie Obenchain[3], Duy Tin Truong[1], Francesco Beghini[1], Faizan Malik[2], Marcel Ramos[2–4], Jennifer B Dowd[2,5], Curtis Huttenhower[6,7], Martin Morgan[4], Nicola Segata[1] & Levi Waldron[2,3]

[1]Centre for Integrative Biology, University of Trento, Trento, Italy. [2]Graduate School of Public Health and Health Policy, City University of New York, New York, New York, USA. [3]Institute for Implementation Science and Population Health, City University of New York, New York, New York, USA. [4]Roswell Park Cancer Institute, University of Buffalo, Buffalo, New York, USA. [5]Department of Global Health and Social Medicine, King's College London, London, UK. [6]Biostatistics Department, Harvard School of Public Health, Boston, Massachusetts, USA. [7]The Broad Institute, Cambridge, Massachusetts, USA. [8]These authors contributed equally to this work. Correspondence should be addressed to L.W. (levi.waldron@sph.cuny.edu) or N. S. (nicola.segata@unitn.it)

1. Huber, W. *et al. Nat. Methods* **12**, 115–121 (2015).
2. Truong, D.T. *et al. Nat. Methods* **12**, 902–903 (2015).
3. Abubucker, S. *et al. PLoS Comput. Biol.* **8**, e1002358 (2012).
4. Human Microbiome Project Consortium *Nature* **486**, 207–214 (2012).
5. Koren, O. *et al. PLoS Comput. Biol.* **9**, e1002863 (2013).
6. Arumugam, M. *et al. Nature* **473**, 174–180 (2011).

The Experimental Design Assistant

To the Editor: The quality and reliability of much animal research is in question. Unreliable or low-quality research represents an unacceptable waste of animals and research resources. In the US alone, the cost of irreproducible research has been estimated at $28 billion annually, and issues with research design and reporting are estimated to account for half of that waste[1]. To address these issues, the National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs) developed the ARRIVE guidelines to improve the reporting of *in vivo* research[2,3]. We now present the Experimental Design Assistant (EDA; https://eda.nc3rs.org.uk), a freely accessible web-based tool, which was launched to help researchers improve the design, conduct, analysis and reporting of animal experiments.

The system was developed by progressive interaction between an expert group experienced in providing advice on experimental design to researchers and a software development team. It includes a computer-aided design tool through which the user develops a diagram that embodies the experimental plan. The diagram offers a new standard notation for describing experiments in which methodological details and analysis plans are explicit (**Fig. 1**). This facilitates communication between collaborators, funding bodies, ethical review committees, journal editors and peer reviewers; it also allows detailed record keeping and may serve as an *ex ante* registered protocol[4].

The structure of EDA diagrams is based on a series of relationships between the different components of the experiment. This allows the use of computer-based logical reasoning to provide feedback and advice on the experimental plans[5]. The feedback helps researchers improve their experimental design by, for instance, highlighting missing information or problems with internal consistency. It also provides assistance with identifying and characterizing the independent variables and outcome measures to be included in the analysis. Advice about common nuisance variables, which threaten the internal validity of many animal experiments, is also provided, along with practical suggestions to account for such variables in randomization and data analysis. The feedback does not restrict researchers to using a particular design type, but it promotes a better understanding of the implications of common design pitfalls so that researchers can make informed decisions. The feedback also suggests methods of statistical analysis that are appropriate for a given design, along with advice on any data requirements ('assumptions') for a given test as well as possible data transformations.

Other features of the EDA include support for randomization, blinding and power calculations, and procedures which are still underused in animal research[6]. Based on the diagram, the system generates a randomization sequence for the study, taking into
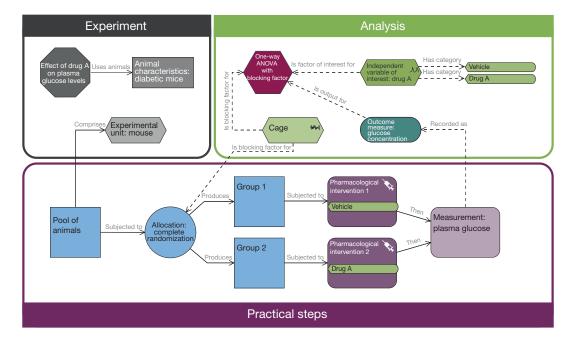
Figure 1 | Example of an EDA diagram. EDA diagram representing a two-group comparison in which each cage contains mice randomized to either of two treatments. Diagrams are composed of nodes and links to represent an entire experimental plan. The gray nodes contain high-level information about the experiment such as the null and alternative hypotheses, the effect of interest, the experimental unit and the animal characteristics. The blue and purple nodes represent the practical steps carried out in the laboratory such as the allocation to groups, the group sizes and role in the experiment, the treatments and the measurements taken. The green and red nodes represent the analysis, the outcome measures and the independent variables of interest and nuisance variables (e.g., blocking factors). For more details, see https://eda.nc3rs.org.uk.

account any blocking factors. The sequence can be sent directly to a third party nominated by the investigator, thus allowing the investigator to remain unaware of the animals' group allocation until the data have been collected and analyzed. Animal experiments are often too small to yield meaningful results[7]; the EDA's power calculation tool—along with extensive guidance on how to choose the appropriate calculation and identify the parameters required—will help researchers determine optimal sample sizes for each experiment.

In conclusion, the EDA is a new resource to help improve the quality of animal research. It can help researchers design robust and reliable experiments in two ways. First, it ensures that the experimental plans are explicit and transparent, thus allowing detailed scrutiny before and after data are collected. Second, it encourages improvements by providing researchers with critical feedback, targeted information and access to randomization and power analysis tools. We will continue to incorporate user input in planned future developments to ensure that the EDA evolves in line with the needs of the research community.

Nathalie Percie du Sert[1], Ian Bamsey[2], Simon T Bate[3], Manuel Berdoy[4], Robin A Clark[5], Innes C Cuthill[6], Derek Fry[7], Natasha A Karp[8,9], Malcolm Macleod[10], Lawrence Moon[11], S Clare Stanford[12] & Brian Lings[2]

[1]National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs), London, UK. [2]Certus Technology Associates Ltd, Exeter, UK. [3]GlaxoSmithKline Pharmaceuticals, Stevenage, UK. [4]Biomedical Services, University of Oxford, Oxford, UK. [5]Envigo, Alconbury, UK. [6]School of Biological Sciences, University of Bristol, Bristol, UK. [7]University of Manchester, Manchester, UK. [8]Wellcome Trust Sanger Institute, Hinxton, UK. [9]Quantitative Biology IMED, AstraZeneca R&D, Cambridge, UK. [10]Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK. [11]Wolfson Centre for Age Related Diseases, King's College London, London, UK. [12]Department of Neuroscience, Physiology and Pharmacology, University College London, London, UK. e-mail: nathalie.perciedusert@nc3rs.org.uk

1. Freedman, L.P., Cockburn, I.M. & Simcoe, T.S. *PLoS Biol.* **13**, e1002165 (2015).
2. Cressey, D. *Nature* http://www.nature.com/news/surge-in-support-for-animal-research-guidelines-1.19274 (2016).
3. Kilkenny, C., Browne, W.J., Cuthill, I.C., Emerson, M. & Altman, D.G. *PLoS Biol.* **8**, e1000412 (2010).
4. Nosek, B.A., Ebersole, C.R., DeHaven, A.C. & Mellor, D.T. Preprint at http://dx.doi.org/10.17605/OSF.IO/2DXU5 (2017).
5. Percie du Sert, N. *et al. PLoS Biol.* **15**, e2003779 (2017).
6. Macleod, M.R. *et al. PLoS Biol.* **13**, e1002273 (2015).
7. Button, K.S. *et al. Nat. Rev. Neurosci.* **14**, 365–376 (2013).