RESEARCH ARTICLE

# An atlas of robust microbiome associations with phenotypic traits based on large-scale cohorts from two continents

**Daphna Rothschild[1,2☯], Sigal Leviatan[1,2☯], Ariel Hanemann[3☯], Yossi Cohen[3], Omer Weissbrod[4], Eran Segal[1,2]****

**1** Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel, **2** Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel, **3** DayTwo LTD, Tel Aviv, Israel, **4** Epidemiology Department, Harvard T.H. Chan School of Public Health, Boston, MA, United States of America

☯ These authors contributed equally to this work.
* eran.segal@weizmann.ac.il

OPEN ACCESS

**Citation:** Rothschild D, Leviatan S, Hanemann A, Cohen Y, Weissbrod O, Segal E (2022) An atlas of robust microbiome associations with phenotypic traits based on large-scale cohorts from two continents. PLoS ONE 17(3): e0265756. https://doi.org/10.1371/journal.pone.0265756

**Editor:** Franck Carbonero, Washington State University - Spokane, UNITED STATES

**Received:** October 17, 2021

**Accepted:** March 7, 2022

**Published:** March 24, 2022

**Data Availability Statement:** All of the source code used to generate the analyses of this paper will be made available upon publication in the following git repository: https://github.com/erans99/UniqueRelativeAbundance As commonly done in the field of human genetics, full tables of summary statistics of the correlations between bacterial species abundance and human traits will be made available. These freely downloadable summary statistics are calculated using all samples used in the study. Summary statistics are available both in supplementary (Table S12-S23) and also in the

## Abstract

Numerous human conditions are associated with the microbiome, yet studies are inconsistent as to the magnitude of the associations and the bacteria involved, likely reflecting insufficiently employed sample sizes. Here, we collected diverse phenotypes and gut microbiota from 34,057 individuals from Israel and the U.S.. Analyzing these data using a much-expanded microbial genomes set, we derive an atlas of robust and numerous unreported associations between bacteria and physiological human traits, which we show to replicate in cohorts from both continents. Using machine learning models trained on microbiome data, we show prediction accuracy of human traits across two continents. Subsampling our cohort to smaller cohort sizes yielded highly variable models and thus sensitivity to the selected cohort, underscoring the utility of large cohorts and possibly explaining the source of discrepancies across studies. Finally, many of our prediction models saturate at these numbers of individuals, suggesting that similar analyses on larger cohorts may not further improve these predictions.

## Introduction

The human gut microbiota is linked to metabolic disorders such as diabetes and obesity but these links are based on relatively small cohorts of several dozens or hundreds of individuals [1–8]. Although these studies reported many statistically significant associations, many of these effects are either moderate or do not replicate in other works [9, 10]. One such example is alpha diversity, for which there are contradicting reports regarding its association with different phenotypes. While microbiome diversity is mostly regarded as a positive indicator of health [8, 11–15], other studies found that increased diversity is associated with microbiome instability [16, 17]. Diversity was also shown to increase with age [18, 19], but this association was not conclusive in other cohorts [20]. These discrepancies call for studying these questions across larger cohorts from diverse backgrounds as was done for several studies comparing the

following repository: https://figshare.com/articles/dataset/Atlas_of_human_gut_microbiome_features_correlated_with_phenotypes/13851839 Raw metagenomic sequencing reads of samples cannot be made publicly available due to a non-negotiable clause in the DayTwo informed consent that is intended to protect the privacy of the DayTwo research participants. Bona fide academic researchers can obtain such access by emailing DayTwo (support@daytwo.com), and will be allowed to make non-commercial use of their research outcomes for replication, critique and for building further outcomes in their own research. 100 raw metagenomic samples from the Personal Nutrition Project [50] on which the code can be run to validate the pipeline and reproduce the results obtained. These samples are available under bioproject PRJEB11532, with accession numbers detailed in supplementary (Table S28). The RAs of gut microbes of these samples, as calculated by the URA method, are given in supplementary (Table S29). These 100 samples could be made publicly available as they were obtained under a different academic informed consent that allows their sharing.

**Funding:** The funders of this study are the European Research Council 786344 and the Israel Science Foundation 292/20. The rest of E.S. support in general but not directly for this project. Daytwo Ltd. provided support for this study in the form of salaries for A.H. and Y.C. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the 'author contributions' section.

**Competing interests:** The authors have read the journal's policy and the authors of this manuscript have the following competing interests: A.H. and Y. C are employees of Daytwo Ltd. E.S. is a paid consultant for Daytwo Ltd. This does not alter our adherence to PLOS ONE policies on sharing data and materials. There are no patents, products in development or marketed products associated with this research to declare.

microbiome in different populations [21–23]. Indeed, in the field of genetics, large cohorts are required since many traits are known to be polygenic and to be affected by small effects from many variants [24, 25]. Similarly, in the microbiome we expect that individual bacterial species may have a low abundance or mild associations with human phenotypes, necessitating large sample sizes. In addition, many bacterial species are present in only a relatively small fraction of the population such that the association between their abundance and traits can only be studied in large cohorts that have enough individuals that harbor them.

Apart from cohort size, there are other challenges in finding robust signals from the microbiome. One such challenge stems from the large number of genes that are shared between different bacteria through mechanisms such as horizontal gene transfer [26, 27]. Such sharing causes many short metagenomic sequencing reads to map non-uniquely to multiple bacteria, making it difficult to estimate bacterial relative abundance (RA). Several methods were devised to address this issue, e.g., by mapping to genes that appear in a single copy and are unique to a single species [28, 29]. However, these methods are not up to date with the expanded reference of bacterial species groups (SGBs) published in 2019 which added 3,796 new SGBs to the human microbiome catalog [30]. Another challenge stems from only being able to estimate microbial RAs, not absolute, which can lead to false bacteria-phenotype associations. This concern can partially be addressed with rarefaction samples and using *reference-frames [31],* with best practice being the ability to replicate results on independent datasets.

To address the above issues and with the aim of deriving robust microbiome associations, we used metagenomic sequencing to profile the gut microbiome of 34,057 individuals from both Israel and the U.S., for which we also obtained a rich set of self-reported phenotypes. We devised a novel algorithm for assessing bacterial RAs based on unique genetic elements, and applied it to the recent and much expanded SGB dataset of Pasolli et al. [30]. Using the RAs on this expanded genome set and much larger cohort, we identified numerous associations between microbiome diversity and several human traits. We were also able to develop models that predict these traits using only microbiome data with moderate accuracy, as in the case of age ($R^2 = 0.31$). Notably, these associations replicate across continents, and models derived from the Israeli cohort generalize well to the U.S. cohort, so they are not specific to a certain environment.

By subsampling our cohort to typical cohort sizes used in other studies, we show that associations and predictions derived from smaller cohorts are highly variable and thus sensitive to the selected cohort, underscoring the need for larger cohorts in the microbiome field.

## Results

### Metagenome samples for 34,057 participants from two continents

We obtained gut metagenomic profiles from 30,083 and 3,974 individuals from Israel and the U.S., respectively, who submitted their sample to a consumer microbiome company and signed an appropriate consent form (participants in this study do not necessarily represent the general population). Participants also answered questionnaires and provided self-reported phenotypic data and blood tests (**S1 Table in S2 File**). When comparing reported phenotypes, phenotypes show varying correlation with one another (**S1 Fig**), with some phenotypes being highly correlated with one another, for example HbA1C% and blood glucose levels (Spearman correlation R = 0.65).

We randomly selected 90% of the samples from the Israeli cohort (n = 27,075 samples) to be our discovery cohort on which we trained predictive models using cross-validation and set aside as independent test sets the remaining 10% of the Israeli cohort ("test1", n = 3,008 samples) and the entire U.S. cohort ("test2", n = 3,974 samples) (**Fig 1A–1E, S1 Table in S2 File**).

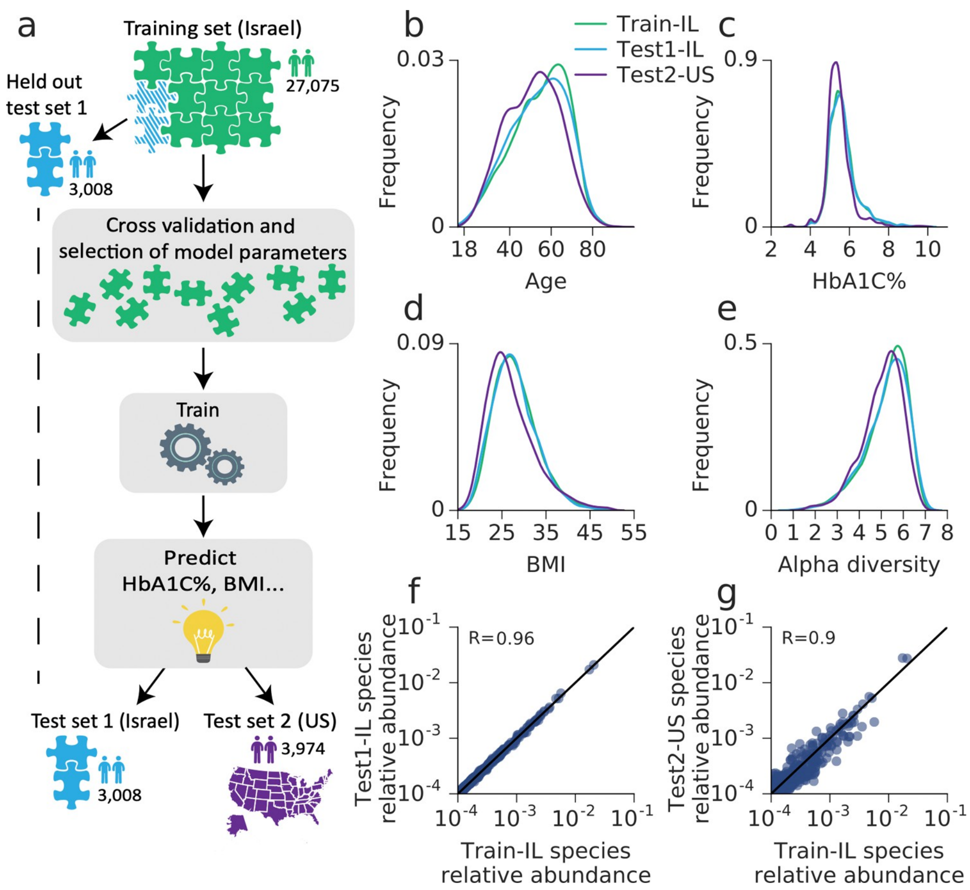

**Fig 1. Cohort description and model prediction scheme.** (a) Illustration of cohorts and machine learning process. A training set of 27,075 individuals was randomly selected out of 30,083 Israeli individuals and was used for model parameter selection using 10-fold cross validation and microbiome, age and gender features. For each phenotype the selected model was trained on the 27,075 training samples and then tested on both the held out 3,008 samples of the Israeli population and a separate U.S. test cohort of 3,974 individuals. (b) Distribution of age in the 3 cohorts, training test1 and test2. (c)-(e) Same for HbA1C%, BMI and alpha diversity. (f) A scatter plot comparing the mean log RA of each species, in the Israeli training cohort vs. the Israeli test1 cohort. R value represents Spearman correlation. (g) Same as (f) in the Israeli training cohort vs. the US test2 cohort. R value represents Spearman correlation.

https://doi.org/10.1371/journal.pone.0265756.g001

These test sets were only used once to evaluate the performance of the models developed on the discovery cohort.

To compute bacterial RA, we used a diverse set of 3,127 human microbiome species representatives from the greatly expanded species-level genome bins (SGBs) classification of Pasolli et al. [30] (**Methods**). We developed a method, Unique Relative Abundances (URA), for estimating the RA of each SGB in every sample (**Methods**), which can be applied to any set of species, and provides better predictive power than Metaphlan (**S2A, and S2B Fig**, **Methods**). Our method is based on examining only reads that map uniquely to a single SGB, since when using unique mappings, we expect uniform coverage across SGB genome bins having the same number of unique mappable regions. This property allows robust estimation of RAs, by calculating the coverage across unique-part-genomic bins for every SGB (Methods). The mean RAs of the different species are the same in the two Israeli cohorts but are somewhat different than in the US cohort (**Fig 1F, and 1G, S2 Table in S2 File**).

## Microbiome diversity increases with age and associates with metabolic parameters

We first examined the association of microbiome diversity and human phenotypes. To this end, we computed alpha diversity using the species level Shannon index and ranked individuals by deciles of alpha diversity (**Fig 2A**). When comparing the top decile and the bottom decile of alpha diversity, we found that HbA1C%, BMI, fasting glucose and fasting triglycerides are

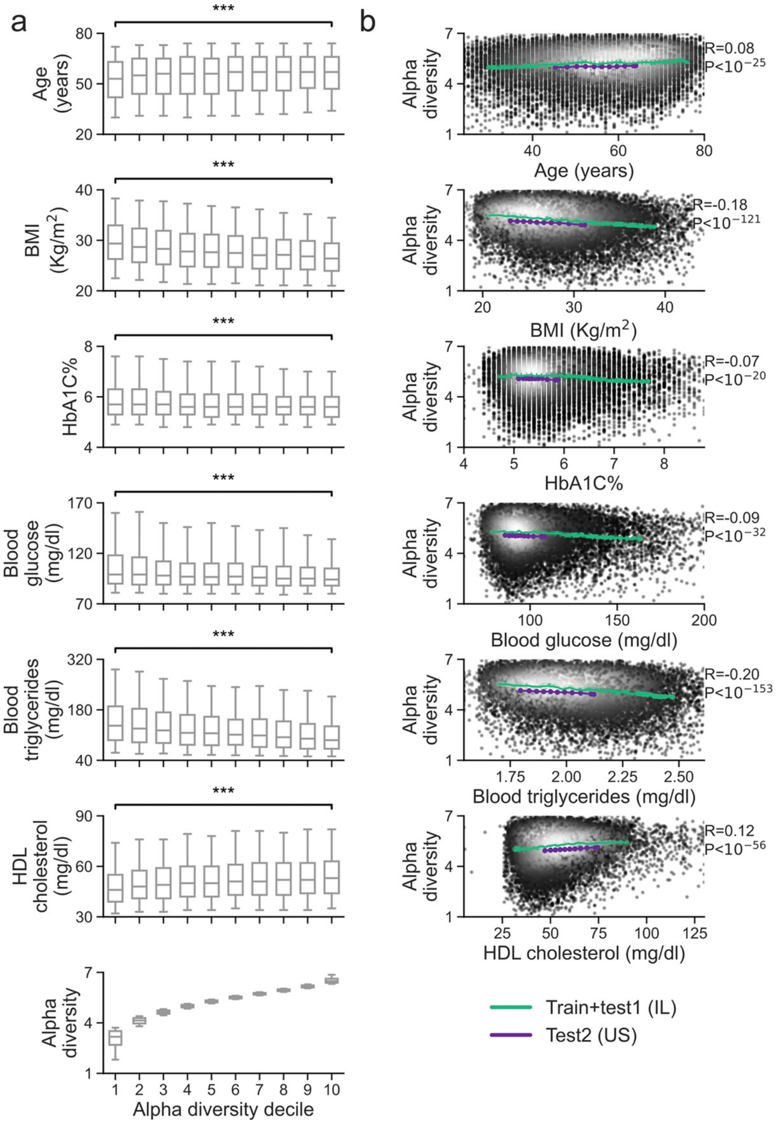

**Fig 2. Species level Shannon alpha diversity significantly associates with many phenotypes.** (a) A box-plot of the distribution of phenotype values, for each of 10 deciles of Shannon alpha diversity, on the IL (train + test1) population. Phenotype values in the first and last deciles of alpha diversity are compared using Mann-Whitney rank-sum test where $^{***}$ signifies P value$< 10^{-16}$ after FDR correction. Boxes correspond to 25–75 percentile of the distribution and whiskers bound percentiles 5–95. (b) Running average of alpha-diversity (y-axis) for the combined training and test1 cohort (green curve), and for the separate test2 cohort (purple curve), ordered by the phenotype values. For the larger Israeli cohort the average is on 1000 individuals with shift of 100 individuals; for the smaller U.S. cohort the group size and shift were chosen to obtain 10 points and the shift was 10% of the group size. The Spearman correlation and P-value shown are of the Israeli cohort, and are calculated on individual level data. Individual level data points are presented with grayscale colors. Data points with light colors have higher frequency.

https://doi.org/10.1371/journal.pone.0265756.g002

significantly higher in the bottom decile while age and HDL cholesterol are significantly lower in the bottom decile (P-value $< 10^{-16}$ after FDR correction, Mann Whitney rank-sum test), including a trend across deciles (**S3-S8 Tables in S2 File**). Similarly, examining alpha diversity as a function of these traits, we found significant correlations between alpha diversity and each of these traits (Fig 2B). Notably, these associations were consistent in both the Israeli and U.S. cohorts (Fig 2B), even though the Israeli cohort has significantly higher alpha diversity values (**Figs 1E and 2B**, mean 7.3±0.77 vs. 7.18±0.67, P-value $< 10^{-40}$, Mann Whitney rank-sum test). The higher diversity of the Israeli cohort persisted even when subsampling the Israeli cohort to match the U.S. cohort on age, gender and BMI (**Methods**, **S2 Table in S2 File**).

## Microbiome-phenotype associations are consistent across continents

We previously employed linear mixed models to estimate the fraction of phenotypic variance that can be inferred from microbiome composition, termed *microbiome-association-index* ($b^2$) *[32]*. Our previous estimates were based on a cohort of 715 individuals and therefore had wide 95% confidence intervals. We revisited these estimates for our two new and larger cohorts. We estimated explained-variance based on alpha-diversity alone (**Fig 3A, Methods**), and based on the full species RAs (**Fig 3B, Methods**). We find that alpha diversity shows significant correlations, yet overall small explained variance, to different traits. We believe that literature inconsistencies between phenotype-alpha-diversity associations are a result of lack of power detecting these associations. Notably, our new estimates agreed well with the findings in our previous study (**S9 Table in S2 File**), but the current much larger cohort of 30,083 Israeli individuals gives substantially narrower 95% confidence intervals (**S10 Table in S2 File**). We found that microbiome composition strongly associates with self-reported diabetes ($b^2$ = 53%, 58% without individuals taking metformin), age ($b^2$ = 28%), HbA1C% ($b^2$ = 16%, 14% without individuals taking metformin), fasting blood glucose ($b^2$ = 13%, 12% without individuals taking metformin), BMI ($b^2$ = 10%), fasting triglyceride ($b^2$ = 9%), HDL cholesterol ($b^2$ = 5%) and current smoking status ($b^2$ = 17%). In contrast, the blood levels of thyroid-stimulating hormone (TSH), albumin and clotting (as measured by International Normalized Ratio, INR) were not significantly associated with the microbiome in our cohort. Notably, $b^2$ estimates from our U.S. cohort of 3,974 individuals were consistent with those derived from the Israeli cohort (**Fig 3B, S10, S11 Tables in S2 File** Spearman correlation R = 0.9, P-value $< 10^{-6}$).

## Different traits are accurately predicted by microbiome composition

We next asked whether various traits can be accurately predicted based only on microbiome composition. For this we used gradient boosted decision trees (GBDT) (**Methods**), with only species RAs as input features to the model.

Our model obtained significant predictions for many traits (**Fig 4A, and 4B, S3 and S4 Figs**) such as age ($R^2$ = 0.31, for 10 fold cross validation on train IL samples), gender (AUC = 0.78), HbA1C% ($R^2$ = 0.25) and BMI ($R^2$ = 0.15). Notably, skin microbiome was recently shown to best predict age [33], and while our predictor based on gut microbiome is significantly more accurate than the reported gut microbiome predictor ($R^2$ = 0.31 versus $R^2$ = 0.17 by Huang et al. [33]) skin microbiome predictor is far more accurate ($R^2$ = 0.74 by Huang et al. [33]).

We obtained significant predictions when stratifying the analyses by gender, with the exception of height which was significantly predicted ($R^2$ = 0.13) in the entire cohort but not in the gender-separated predictions. Since metformin, the most common drug used to treat patients with type2 diabetes, is known to affect microbiome composition, we also evaluated

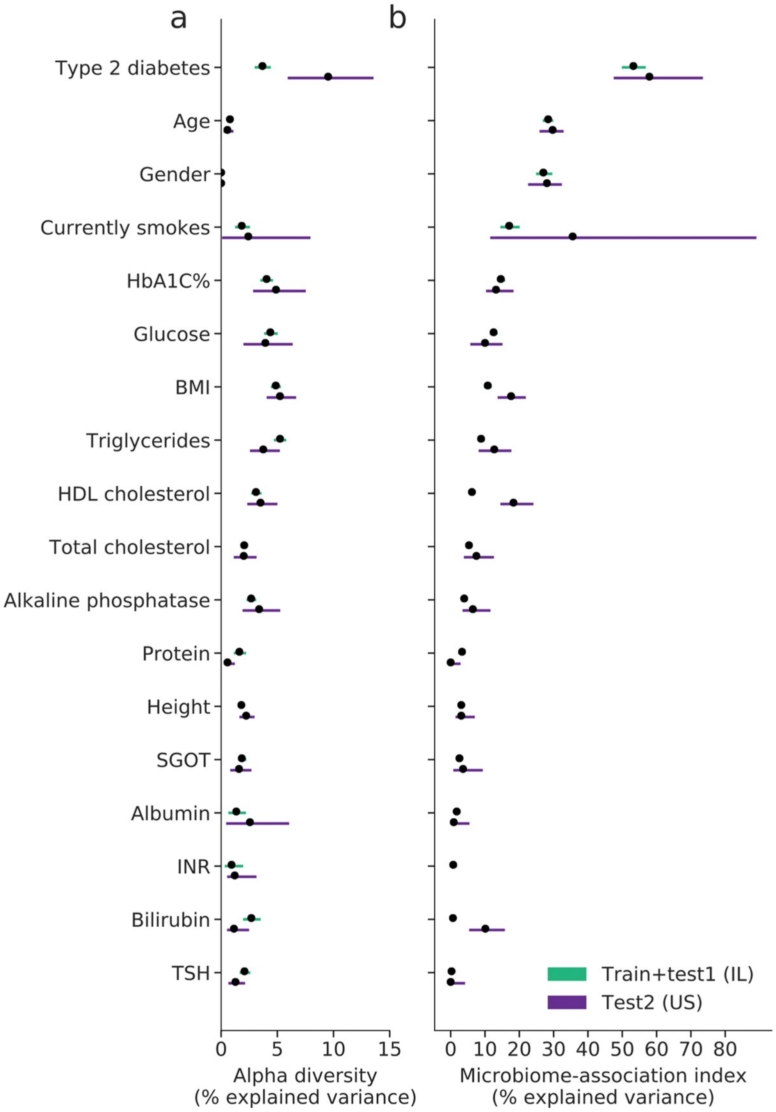

**Fig 3. Explained variance of phenotypes based on microbiome features.** (a) The proportion of variance of various phenotypes that can be explained using Shannon alpha diversity in the Israeli (green) and U.S. cohorts (purple) based on a linear model with covariates for age and gender. Also shown is the 95% confidence interval. (b) The proportion of variance of various phenotypes that can be explained using species-level RAs in the Israeli (green) and U.S. (purple) microbiome composition based on a linear mixed model estimation with covariates for age and gender (microbiome association index [32]). Also shown is the 95% confidence interval. Estimates from the larger cohort have smaller confidence intervals.

https://doi.org/10.1371/journal.pone.0265756.g003

the performance of an HbA1C% predictor only on participants who did not report taking metformin and obtained equivalent performance ($R^2 = 0.19$).

Comparing the GBDT to linear models (with Ridge regularization), GBDT slightly outperformed the linear models (overall mean $R^2$ improvement of 0.02+/-0.011, S2C and S2D **Fig**, S1 File) this may suggest that non-additive interactions between different bacteria are predictive of several traits. An additional support for the importance of non-additive interactions among bacteria in predicting traits, for both HbA1C% and BMI the $R^2$ of the GBDT predictions on held-out subjects was higher than the estimated $b^2$ for these traits (Fig 3B). As the $b^2$

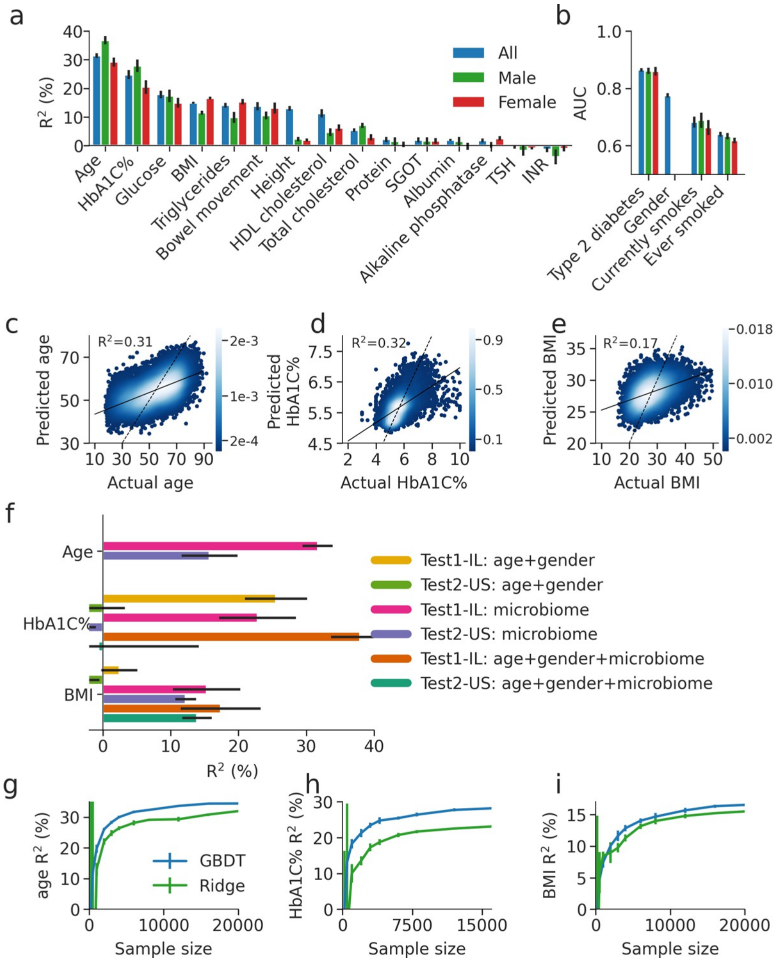

**Fig 4. Prediction of phenotypes by the microbiome.** (a) Coefficient of determination ($R^2$) of GBDT prediction of different phenotypes based only on species level gut microbiome abundance. Results are obtained in a 10-fold cross validation scheme on the training set. Predictions are shown for three models, a model using the whole cohort, and a model for each gender. (b) Same as (a), but shown is the area under the curve (AUC) for predicting binary phenotypes. (c)-(e) Scatter plot of the phenotype and 10-fold cross-validation predicted values of the phenotype, for age, HbA1C% and BMI when training on the Israeli train cohort using GBDT. $R^2$ of prediction is reported. Black line represents regression, dashed black line is x = y. (f) Coefficient of determination ($R^2$) of predictions of age, HbA1C% and BMI, for models trained with different sets of input features using GBDT, and tested on both the held-out Israel and U.S. test sets. Error bars of the test set are from bootstrapping. (g)-(i) Coefficient of determination ($R^2$) and standard deviation error bars of predictions of age, HbA1C% and BMI obtained using GBDT (purple) or Ridge regression (green) models trained on sub-samples of the cohort train IL, of different sizes, and tested of the test IL cohort. For each cohort size *k*, 10 random sub-samples of *k* individuals were obtained and the mean and standard deviation of their predictions are shown.

https://doi.org/10.1371/journal.pone.0265756.g004

estimation used linear mixed models to estimate the fraction of variation predicted by the microbiome, it does not include any non-linear interaction captured by GBDT.

We investigated if the predictive power of the microbiome is mediated through age and gender, since some of the above traits such as HbA1C% are known to increase with age [34]. We found that the microbiome composition predicted age with moderate accuracy ($R^2$ = 0.32), and age and gender alone predict HbA1C% with $R^2$ = 0.26 and BMI with $R^2$ = 0.02 (Fig 4F). Nevertheless, we found that adding microbiome to age and gender to the GBDT model significantly improved the predictions of both HbA1C% (from $R^2$ = 0.26 to 0.38, Fig 4F) and

BMI (from $R^2 = 0.02$ to 0.17, Fig 4F), demonstrating that some of the association between microbiome and these traits is not mediated through age and gender.

We also evaluated the accuracy of our above models, derived from the Israeli training set, on our two independent and held out cohorts from Israel and the U.S.. We found that microbiome base predictions for all traits are replicated in the Israeli held out cohort, and all microbiome base predictions except for that for HbA1C% are replicated in the U.S. cohort (Fig 4F), thereby validating the robustness of our models. We note that in general prediction accuracy was lower in the U.S cohort which may be explained by differences in the microbiome composition between the IL and U.S. cohorts and by lower age and HbA1C% levels in the U.S. cohort (**S1 Table in** S2 File).

Finally, to examine the importance of cohort size on prediction accuracy, we applied our above prediction pipeline to different random subsamples of our training cohort, ranging from a few hundreds of subjects to 24,000 (**Methods**). We found that prediction accuracy increases with cohort size (**Fig 4G–4I, S5 Fig**) and does not saturate even with a cohort of 1,000 individuals. For age, we observed an almost two-fold increase in the $R^2$ (from 0.18 to 0.30) when increasing the cohort from 1,000 to 12,000 individuals. For cohorts of hundreds of individuals, the standard deviation of the predictions was high, as in the case of HbA1C% for which different subsamples of 200 individuals can reach both $R^2 = 0.4$ and $R^2 = 0.0$ as likely outcomes (within 2 standard deviations). Together, these results highlight the need for obtaining large cohorts for microbiome studies, as is known to be the case in the field of human genetics.

## An atlas of bacterial species that robustly correlate with age, HbA1C% and BMI

We next sought to identify which individual bacterial species are responsible for driving the predictions of our models for age, HbA1C% and BMI, since these traits were predicted with the highest accuracy. We found many bacterial species that exhibited highly significant correlations to these traits (Fig 5A–5C, 967, 720, 1,056 bacteria out of the top 1,345 occurring bacteria had significant Spearman correlation, P-value $< 0.05$ after FDR correction for age, HbA1C% and BMI respectively, **S12–S17 Tables in** S2 File). Moreover, the Spearman correlation of the bacterial abundances with these traits was in good agreement between the Israeli and U.S. cohorts (Fig 5A–5C, Spearman-R = 0.58, 0.50, 0.74 for age, HbA1C% and BMI, P-value$<10^{-87}$). Notably, the 3 bacteria most strongly associated with BMI in both cohorts included two bacterial species from the *Eubacteriaceae* and *Clostridiaceae* family that was only recently assembled and that has no genome in public repositories (unknown SGB, Fig 5A–5C).

To further investigate these associations, prediction models were built on the subset of the most highly associated bacteria with each phenotype (Fig 5D–5F). We noticed that while the predictions of age and BMI rises significantly when the number of bacteria used grows, the prediction of HbA1C% is almost exclusively driven from the one most highly associated bacteria, E. coli, with other highly associated bacteria contributing very little new information. Interestingly it is also HbA1C% who's model does not reproduce between cohorts, even though the most significantly associated bacteria is the same for both cohorts.

Finally, we subsample the cohorts to smaller cohort sizes and observe that large cohorts are necessary in order for results to replicate **(Fig 5G–5I, Methods)**.

## Functional characterization of gut microbiome

In order to identify bacterial mechanisms that can modulate metabolic phenotypes, we looked at URA based microbiome module and pathway enrichments associated with human

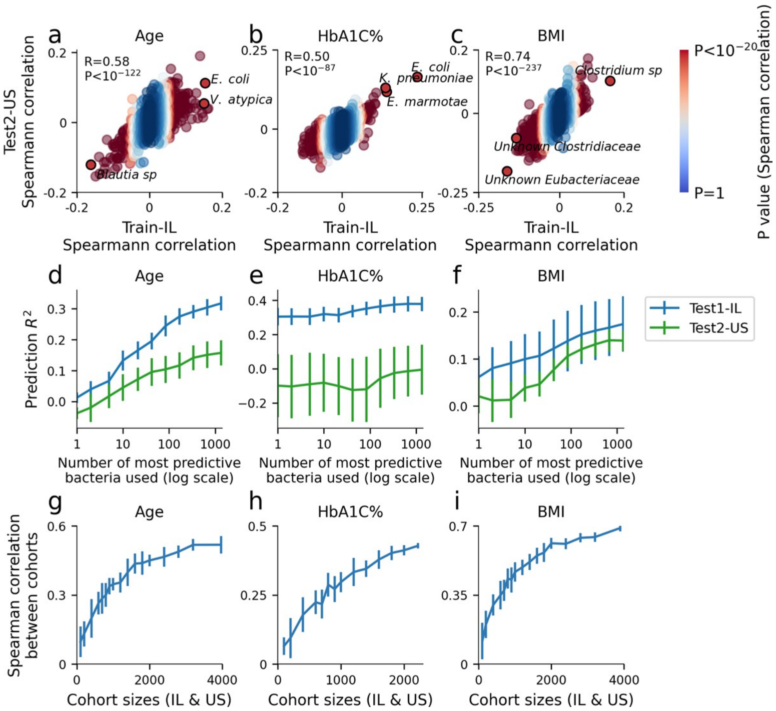

**Fig 5. Correlations of single species with age, HbA1C% and BMI.** (a) Spearman correlation of each bacterial species with age in the Israeli training cohort (x-axis, N = 27,075) and the U.S. test cohort (y-axis, N = 3,974). The correlation and P-value between the correlation coefficients of each cohort are shown. Bacteria are colored according to the P-values of the Spearman correlation coefficient in the Israeli cohort. The top three bacteria by Israeli P-values that replicate in the U.S. cohort are highlighted. (b)-(c) Same as (a) for HbA1C% and BMI. (d) $R^2$ of predictions on Test1-IL and Test2-US cohort, using a partial set of bacteria features for XGBoost. X-axis represents the number of bacteria (with the highest Spearman correlation to phenotype, on the Train-IL cohort) used to build the model. (e)-(f) Same as (d) for HbA1C% and BMI. (g) Spearman correlation between the correlation coefficients of the Israeli cohort and U.S. cohort as in (a) but for different sub-samples of cohort sizes. For each cohort size *k*, a sub-sample of *k* individuals was obtained from both the Israeli and U.S. cohorts and this procedure was repeated 10 times to obtain standard deviation error bars. (h)-(i) Same as (g) for HbA1C% and BMI.

https://doi.org/10.1371/journal.pone.0265756.g005

phenotypes (**Methods**). We found many genes whose abundance was highly correlated with human phenotypes (**Fig 6A–6C, S18–S23 Tables in S2 File**), and from them we derived several modules and pathways which were significantly associated with human phenotypes (**Fig 6D and 6E, S24, S25 Tables in S2 File**).

We found that the Lipopolysaccharide (LPS) biosynthesis pathway (ko00540) and the KDO2-lipid A biosynthesis module (M00060) were associated with adverse metabolic status, as was the citrate cycle (ko00020) and two of its sub-modules (M00009 & M00011). We also found vitamin metabolism modules, such as Menaquinone biosynthesis & Tetrahydrofolate biosynthesis (M00116 & M00126), to be associated with high BMI, high Triglycerides and low HDL (**S1 File, Fig 6D and 6E**).

Finally, as a useful resource for the community, we compiled results into an atlas of summary statistics for all bacterial species and KEGG genes to 6 top predicted phenotypes (**S12–S23 Tables in S2 File**). For each species and KEGG gene, we report bacterial associations to human phenotypes based on bacterial log RAs (**Methods**). Specifically, we report the Spearman correlation coefficient and P-value, the Pearson correlation coefficient and P-value. For bacterial species we also provide the coefficient in the linear model (trained with Ridge regularization), and bacterial feature importance in the GBDT model using the feature attribution framework of SHapley Additive exPlanations [35] (SHAP). In genetics, summary statistics of

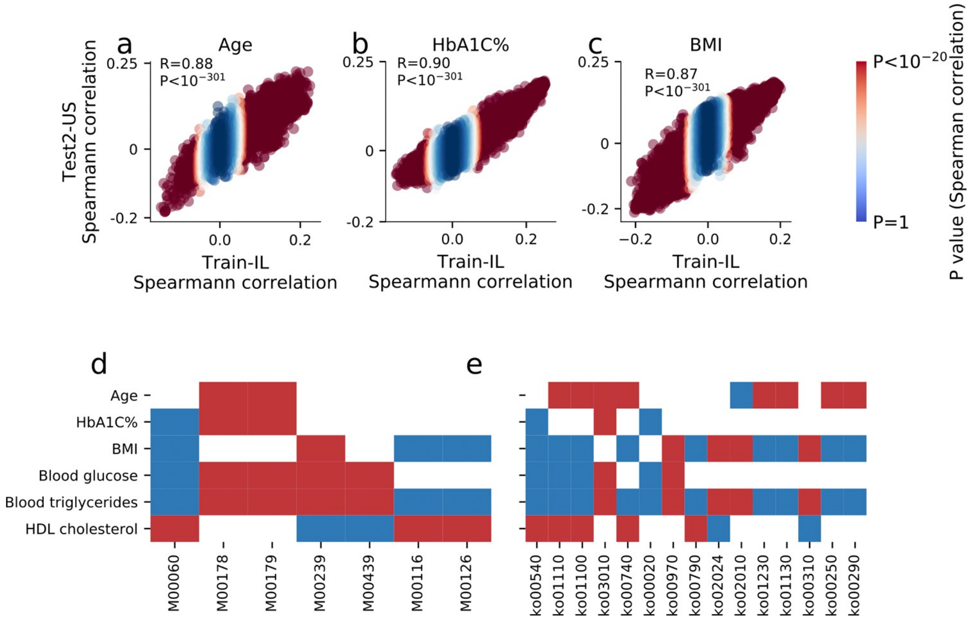

**Fig 6. Functional analysis of modules and pathways.** (a)-(c) Spearman correlation of each KEGG KO gene with age, HbA1C% and BMI in the Israeli training cohort (x-axis, N = 27,075) and the U.S. test cohort (y-axis, N = 3,974). The correlation and P-value between the correlation coefficients of each cohort are shown. Uniref are colored according to the P-values of the Spearman correlation coefficients in the Israeli cohort. (d)-(e) Heat map displaying only significant positive (blue) or negative (red) association between 6 phenotypes and top associated KEGG KO modules (d) or pathways (e). M00060 is the KDO2-lipid A biosynthesis, Raetz pathway, LpxL-LpxM type. M00116 the Menaquinone biosynthesis, chorismate → menaquinol. M00126 the Tetrahydrofolate biosynthesis, GTP → TH. Ko00540—Lipopolysaccharide biosynthesis. Ko01110—Biosynthesis of secondary metabolites. Ko01100—Metabolic pathways. Ko03010- Ribosome. Ko00740—Riboflavin metabolism. Ko00020—Citrate cycle (TCA cycle). Ko00970—Aminoacyl-tRNA biosynthesis. Ko00790—Folate biosynthesis. Ko02024—Quorum sensing. Ko02010—ABC transporters. Ko01230—Biosynthesis of amino acids. Ko00310—Lysine degradation.

https://doi.org/10.1371/journal.pone.0265756.g006

single nucleotide polymorphisms are widely used to generate polygenic risk scores which were shown to be predictive of disease [36, 37]. Similarly, researchers can now use our resource to generate microbiome-based predictions of phenotypes in their datasets by extracting our reported bacterial regression coefficients and multiplying them by the log of the RAs of the corresponding species in their dataset.

## Discussion and conclusion

In this study, we collected the largest cohort to date of metagenomic samples and phenotypic data from two continents, and analyzed it using a much-expanded set of reference microbial species. Together, this allowed us to identify highly robust associations between gut microbiome composition and phenotypes, which replicate in both cohorts. It is important to note, that the microbiome is predominantly shaped by the host environment [32] any found association between microbiome and host phenotypes cannot be assumed causal.

We compiled bacterial associations with host phenotypes into an atlas that can be used by the community to derive trait predictions on smaller datasets, akin to the use of summary statistics in the field of genetics. However, since the Israeli and U.S. populations are both westerized populations, it is yet to be discovered if bacteria that associate with human phenotypes in these populations will also apply to other populations. We show that a large fraction of the variance of several traits such as age, HbA1C% and BMI can be accurately predicted by both linear models and boosting decision trees models. For age and BMI, but not HbA1C%, the predictions replicate across continents and there is also high agreement in the set of individual bacterial species that associate with these traits in both cohorts. Prediction in the U.S. cohort was lower compared to the Israeli cohort, and while we do not have genetic and lifestyle information on these participants we hypothesize that such differences are a source of prediction accuracy differences.

When sub-sampling our large cohort into smaller sized cohorts, we found that even cohorts of 1,000 individuals have significantly lower average accuracy of associations between bacteria and phenotypes. Models derived from different sub-samples of smaller cohort sizes display high variability in the set of bacteria that associate with each trait including the direction of association, and in prediction accuracy. These results may explain the relatively low agreement that exists across studies in the set of bacteria associated with different traits and conditions, and they call for employing larger cohort sizes in microbiome studies.

Using an expanded reference set allowed us to study many bacterial species for the first time and to identify novel associations for them. Notably, even among the top associated bacteria we found unnamed bacteria that are prevalent and appear in thousands of individuals from our cohort. This also provided us a glimpse into possible functional drivers of some of these associations, supporting the hypothesis that metabolic syndrome phenotypes are associated with low-grade subclinical inflammation leading to hyperglycemia and favoring the onset of type 2 diabetes and obesity [38]. These findings emphasize the importance of expanding the reference set of the human microbiome even further, and suggests that such newly identified species may have strong associations with important host phenotypes.

Overall, by combining larger microbiome cohorts and expanded bacterial genome references we robustly characterize bacterial links to many important health parameters, serving an important first step towards unraveling the causal links and mechanisms by which bacteria affect host phenotype.

## Materials and methods

### Recruitment

All participants are paying customers of a consumer microbiome company (DayTwo), who enrolled to get personalized algorithm based dietary recommendations, from January 2017 until January 2020. Exclusion criteria includes customers using antibiotics or antifungals three months prior stool sample collection, age under eighteen, pregnancy or less than three months post-delivery, active fertility treatments or customers treated with short-acting insulin. We further excluded 2050 participants who self-reported pancreatic disease, undetermined colitis, ulcerative colitis, crohn's disease, inflammatory bowel disease, gestational diabetes or type1 diabetes. All participants provided a stool sample, and filled a medical questionnaire, before getting their dietary recommendations.

## Ethics statement

All human subjects in this study submitted their sample to a consumer microbiome company and signed an appropriate consent form. User approved terms for use of data are attached **S1 File**.

### Microbiome sample collection, processing and analysis

Participants provided a stool sample using an OMNIgene-Gut stool collection kit (DNA Genotek), and processed according to the methods described in Mendes-Soares et al. [39]: Genomic DNA was purified using PowerMag Soil DNA isolation kit (MoBio) optimized for Tecan automated platform. Illumina compatible libraries were prepared as described in [40], and sequenced on an Illumina Nextera 500 (75bp, single end), or on a NovaSeq 6000 (100bps, single end). Reads were processed with Trimmomatic [41], to remove reads containing Illumina adapters, filter low quality reads and trim low quality regions; version 0.32 (parameters used: -phred33 ILLUMINACLIP:<adapter file>:2:30:10 SLIDINGWINDOW:6:20 CROP:100 MINLEN:90 for 100bps reads, CROP:75 MINLEN:65 for 75bps reads). Reads mapping to host DNA were detected by mapping with bowtie2 [42, 43] (with default parameters and an index created from hg19) and removed from downstream analysis.

We used bowtie2 [43] to map samples from our cohort versus an index built from the set of representatives of the SGBs (demanding all mappings of length 100/75 to score -40 or above). On average 77 percent of reads (minimum of 50 percent, maximum of 86 percent) mapped to bacterial representatives. This mapping percentage is in line with the original mappability estimates of Pasolli et al. [30] for westernized gut microbiome samples on the set of representatives of the reference set.

All samples of 75bps (Nextera 500) were subsequently down-sampled to a depth of 8M mapped reads, and all samples of 100bps (NovaSeq 6000) were subsequently down-sampled to a depth of 5M mapped reads. That is, sub-sampling was performed after mapping to a bowtie2 index of the bacterial reference dataset, and only reads which mapped to one or more bacteria in the dataset were taken. For samples with fewer mapped reads, all mapped reads were taken, this accounted for about half of the 75bps samples, and less than 5% of the 100bps samples.

All atlas correlations and all predictions were performed also on the subset of samples of 75bps reads with 8M mapped reads or of 100 bps reads with 5M mapped reads. All results replicate, but with lower power, since they were performed on less samples.

## RA estimation of SGBs—Unique Relative Abundance (URA)

The bacterial reference dataset for RA estimation is based on the representative assemblies of the species-level genome bins (SGBs) and genus-level genome bins (GGBs) defined by Pasolli et al. [30]. By the process by which clusters were formed, all assemblies in each SGB are at high average nucleotide identity with one another. The representative assembly was chosen to be the best quality assembly amongst them.

Out of the 4,930 human SGBs (associated with various body sites), we chose to work with 3,127 SGBs, which were characterized by either belonging to a unique genus or with at least 5 assemblies to justify having a new SGB. We employed this restriction, since we noticed that the cutoff threshold used by Pasolli et. al. to cluster assemblies into SGBs resulted in small groups with little nucleotide difference from a large nearby SGB thus, assumed by us to be an erroneous split to a new SGB.

Abundance was calculated by counting reads that best matched to a single SGB of the set. In order to avoid sample reads which may be assigned to more than one SGB (which might mislead us to believe an SGB appears in a sample when it actually does not), we created a mapping of all 100/75-bps potential reads which are unique to a single of these representatives. We divided each representative genome assembly to non-overlapping windows such that each window includes 100 unique 100/75-bp potential reads (unique-100-bins). Since different areas of the assembly have a different proportion of uniquely mapped potential reads, these windows are not of constant length, but the number of sample reads expected to uniquely map to them is constant.

We used bowtie2 [43] to map samples from our cohort versus an index built from the set of representatives of the SGBs (demanding all mappings of length 100/75 to score -40 or above). When analyzing the mapping, we looked only at reads whose best map is unique (thus mapped to a location which is unique in the set of representatives). We count the number of reads uniquely mapped to each window of each SGB.

To assess the cover of each SGB, we first choose a window size to work with, since lower abundance species will need longer windows in order to assess coverage. Windows size is chosen as a multiple of the original windows of unique-100, so that the number of reads that map to that number of consecutive windows is about 20 reads, on average. Next, we sum the number of reads in these enlarged-windows, and test the distribution of the number of unique reads per window.

Finally, we take the dense mean of that distribution [44], in order to avoid our coverage estimation being biased by a relatively small part of the reference which is highly covered (may come about from plasmids or horizontal transfer which was not identified in the uniqueness process since it did not appear in any other representative) or lowly covered (since this is a representative of an SGB, a strain present in our sample may not include all parts of the representative). When the dense 50% of the cover distribution does not include 0 we conclude the SGB exists in the sample, and we estimate its RA. The coverage estimation for each SGB is the dense mean cover of its representative, normalized by the enlarged-window size.

The RA estimation is the coverage divided by the sum of the covers of all representatives we concluded exist in this sample.

## Step by step description of URA build and use

All the code of the algorithm, and the code for building the necessary databases, is provided in github: https://github.com/erans99/UniqueRelativeAbundance

To generate RAs on a set of samples, the pipeline performs the following steps:

## Build a URA database

1. A user selected reference species set with one representative genome per species is required. In order to get good read uniqueness, the representatives need to be far enough from each other. This means that each representative has enough unique reads (reads which do not appear in any other representative) to estimate its abundance. A Mash [45] distance of 0.05 between pairs of representatives should be enough. Adding similar genomes with Mash distance <0.05 will cause insufficient differences between genomes and thus the process cannot work for strain level abundances.
2. Build a bowtie2 index from this set of representatives.
3. Break representative genomes to reads that match the sequenced technology with window jumps of one base pair (we used two read lengths, 75 and 100 base pairs).
4. Map these reads to the index built from the set of representatives.
5. Establish the set of unique reads (those that map to a single position in the index), and create windows of 100 consecutive unique reads for each representative genome.

After the URA DB build stage, for each metagenomic sample, we run the URA method. All metagenomic samples on which URA is run are after quality control (see processing subsection in Methods) and removal of human genomic reads (see processing subsection in Methods).

## URA estimation stage

1. Map all reads to the bowtie index of representatives' genomes (see analysis subsection in Methods).

2. Take only the reads whose best map is to a unique position in a single representative, and calculate the cover of all windows created in the build process.
3. For each representative genome:
   a. Choose a number of consecutive windows (of 100 consecutive unique reads) to sum the cover over, so that most summed window will be covered by approximately 20 reads (between (2/3)*20 - (3/2)*20 = 13–30 reads). If this comes out to be 5 windows or less—the genome has abundance 0.
   b. Expected genome bin coverage distribution:
      i. Many parts have low values of 0–1 reads (parts deleted in the strain that is actually in that sample).
      ii. Some parts which are very highly covered (possibly copy number variation, or some plasmid incorrectly assigned to that genome).
      iii. If the species actually exists in the sample there will be a normal distribution of cover levels, encapsulating more than 50% of windows, around it's true cover.
      iv. If the species does not exist in the sample the dense mean window will start at 0, and will be an exponential distribution, not a normal distribution. In this case the genome has abundance 0.
   c. Estimate the true cover from the dense mean of the distribution, which is the normal distribution around the true cover. Take the mean of the normal distribution, and divide it by the number of consecutive windows summed together, to get an estimated cover of a single window (of 100 consecutive unique read windows).
4. Take the cover level of all species which are present, and normalize them to RAs (cover is proportional to abundance, and all RAs sum up to 100%).

## Predictive power of the Unique Relative Abundance (URA) method

To evaluate our URA method, we tested the quality of predictions derived from MetaPhlan [28] species RAs, vs. our URA abundances.

In order to discriminate between the predictive power derived from the estimation method vs. that coming from the expanded set of species, we created a new URA database, including only the subset of 998 SGBs that Pasolli et al. marked as known NCBI human bacterial species. This subset is not equivalent, in number (smaller) or in identity, to the set of Metaphlan species, but is merely the "known" set of SGBs.

We find that URA with this reference set achieves a higher prediction accuracy than MetaPhlan [28] for different phenotypes, and for most phenotypes a lower power than the prediction accuracy using the expanded 3,127 SGBs reference set (**S2A, and S2B Fig, S26, and S27 Tables in S2 File**). For some of the phenotypes the prediction accuracy improved dramatically, between Metaphlan prediction and URA, and for no phenotype did the accuracy decrease, for example for BMI for a Coefficient of Determination ($R^2$) of 0.12 to 0.15 (p-value of $1.4*10^{-10}$ on t-test on the 10 folds).

## Strengths and limitations of Unique Relative Abundance (URA) method

The URA method is based on taking a set of reference genomes, one representing each of the different species, and looking only at the reads that are unique to a single one of these representatives. This mapping is weighted against pre-processed uniqueness between bacterial species in the reference dataset (UREF). The expectation is that reads are randomly sampled from the gut bacterial population. In order to fulfill this assumption, the implemented method uses only single end reads.

While this process gives a good estimation of abundances, as can be seen by its high predictive power (**S1A and S1B Fig** shows comparison with predictive power of Metaphlan abundances), it does not mean that any read mapped to a species representative is necessarily a read originating from that species. Specifically, deletions, copy number variations and horizontal gene transfers may cause some of the reads originating for one bacteria in some individual's gut, to be assigned to another bacteria's UREF. Thus, the method is appropriate for species abundance estimation, but requires careful interpretation for gene level analysis.

The URA method was developed for estimating relative abundances of SGBs. Although in essence URA can be used for relative abundance estimation of any genomic content, at its current implementation URA method does not give gene abundance estimations. This is because genes shared by multiple species do not have sufficient uniqueness and for these genes the current URA implementation gives poor estimations. In order to overcome this limitation, we estimate gene relative abundances by gene presence/absence multiplied by the matching SGB relative abundance to calculate gene level relative abundance as similarly performed with PICRUSt2 [46]. This is explained in further details under "**Correlation of phenotypes with biological annotation**".

## Cohort matching

We subsampled the IL cohort to match the U.S. cohort on age, gender and BMI, using the MatchIt package from CRAN repository for r [47].

## Alpha diversity explained variance

We calculated the alpha diversity explained variance by regressing out gender and age from each phenotype, and then using ordinary least squares modeled the phenotype by alpha diversity. To get confidence intervals, we bootstrapped the data 10,000 times.

## Microbiome-association-index

We calculated $b^2$ estimates using linear mixed models as was previously described [32]. We used age and gender as fixed effects covariates, and built a microbiome genetic-relationship-matrix, using our developed SGB based RAs. The $b^2$ calculation assumes that the phenotype distributed normally, we removed sample outliers from the IL and US cohorts using the same thresholds (removing less than 5% of individuals **S30 Table in S2 File**). To account for differences between the population and study prevalence of binary traits, we applied the correction of Lee et al. [48] which has been shown to provide a lower bound on the fraction of explained variance [49]. We also provide uncorrected estimates in **S31 Table in S2 File**. Phenotype distributions of blood SGPT levels were far from normally distributed and were not estimated.

## Phenotypes prediction

We used the gradient boosting trees regressor from Xgboost [50] as the algorithm for the regression predictive model for different phenotypes. We used the gradient boosting trees classifier from Xgboost as the algorithm for the classification predictive model for phenotypes with binary values. All hyperparameters of the xgboost were fitted based only on cross validation of the train set.

The parameters of the predictors when using microbiome features were: colsample_bylevel = 0.075, max_depth = 6, learning_rate = 0.0025, n_estimators = 4000, subsample = 0.6, min_child_weight = 20. These parameters were used for regression as well as classification.

The rest of the parameters had the default values of Xgboost.

For the Ridge linear regression, we used the RidgeCV from the scikit-learn package. The parameters used for the regressor were:

alphas = [0.1,1,10,100,1000], normalize = True. The rest of the parameters were the default. The input to the Ridge linear regression was log transformed SGB abundance.

For binary phenotypes SGD classifier from the scikit-learn package was used, with default parameters (L2 normalization).

When using microbiome features for the prediction, only the top 1345 occurring SGBs were used, i.e., the SGBs that were found in at least 5% of the samples, to avoid overfitting on rare SGBs.

## Calculating prediction accuracy as a function of cohort size

For cohort size n (for n = 200, 500, 1000, 2000, 3000, 4000, 6000, 8000, 12000, 16000, 20000, 24000; for prediction of HbA1c the maximum size was 16000) we repeated the following process 10 times: we randomly selected a subset of n samples, built a predictive model for the phenotype from the subset of samples, and tested its predictions on the independent test sets (test1-IL and test2-US test-sets described in main text), and calculated the accuracy (R square) of the prediction. By repeating the procedure 10 times we received the mean and standard deviation of the prediction accuracy estimate, for each IL trained subset size.

## Predictive power of single species

For cohort size n (for n = 100, 200, 400, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2400, 2800, 3200) we repeated the following process 10 times: we randomly selected a subset of n samples from each cohort, calculated the Spearman correlations of each species with the phenotype in each cohort, and correlate the Spearman correlations of the two cohorts, over all species (not just those that pass a threshold p-value). We plot the mean and standard deviation of this correlation, as calculated over the 10 repeats.

## Correlation of phenotypes with biological annotation

Gene prediction, and annotations, of the representative genomes of the SGBs were performed in advance by Pasolli et al., including assigning Kegg Orthologies (KOs) to identified genes. Similar to PICRUSt2 [46] method, in order to calculate the pseudo-abundances of each KO in each member of the cohort we multiplied the RAs of the SGBs with the number of times each KO appeared in the representative genome.

In order to gain insights that were not captured by the single species' abundance, we looked only at KOs which appeared at least 5 times in the representatives. We calculate a pseudo-abundance as the sum of the abundances of all species it appears in. These pseudo-abundances sum the abundances of more than one species and were not derived from a single bacterial species' abundance, and thus gave new correlations to high level modules and pathways.

The vector of the gene's pseudo-abundance (one per individual in our cohort) was correlated versus different phenotypes (an atlas is provided in **S18–S23 Tables in S2 File**), and then ranked, from the most significantly negatively correlated to the most significantly positively correlated with the phenotype. This process was performed for each cohort separately, and shows significant concordance between the IL and the US cohorts (**Fig 6A–6C**).

After ranking the KOs against a phenotype, ranksum analysis was used in order to search for pathways and modules which are significantly enriched or significantly diminished (**S24, S25 Tables in S2 File**), with connection with the phenotype. We present all KOs and modules which were significantly correlated with at least 3 of the phenotypes tested (**Fig 6D and 6E, S24, S25 Tables in S2 File**).

## Supporting information

**S1 Fig. Correlation between the different phenotypes on IL cohort.** Spearman correlation of the different phenotypes, on the Train-IL cohort.
(TIFF)

**S2 Fig. Comparisons of prediction of phenotypes by the microbiome.** (a) Comparison of the predictive power of three different sets of species level abundance estimations of gut microbiome. In blue, predictions are performed using the baseline Metaphlan species abundances, in green predictions are performed on abundances calculated using the URA algorithm on the sub-set of SGBs that were known prior to the work of Pasolli et al. [30], and in red predictions are performed on URA abundances. For all three, the coefficient of determination ($R^2$) of GBDT prediction of different phenotypes are obtained in a 10-fold cross validation scheme on the training set. (b) Same as (a), but shown is the area under the curve (AUC) for predicting binary phenotypes. (c) Comparison of the predictive power of GBDT model (in blue) versus Ridge regression model (in green). Coefficient of determination ($R^2$) of prediction of different phenotypes based only on species level gut microbiome abundance. Results are obtained in a 10-fold cross validation scheme on the training set. (d) Same as (c), but shown is the area under the curve (AUC) for predicting binary phenotypes.
(TIFF)

**S3 Fig. GBDT prediction of phenotypes by the microbiome.** (a)-(n) Scatter plots of 10-fold cross-validation predicted values of quantitative phenotypes when training on the Israeli train cohort using GBDT. R2 of prediction is reported. Black line represents x = y. (o)-(r) ROC curve plots of 10-fold cross-validation predicted values of binary phenotypes when training on the Israeli train cohort using GBDT. AUC of prediction is reported. Black line represents x = y.
(TIFF)

**S4 Fig. Prediction of phenotypes by the microbiome using Ridge regression.** (a)-(n) Scatter plots of 10-fold cross-validation predicted values of quantitative phenotypes when training on the Israeli train cohort using Ridge regression. $R^2$ of prediction is reported. Black line represents x = y. (o)-(r) ROC curve plots of 10-fold cross-validation predicted values of binary phenotypes when training on the Israeli train cohort using Ridge regression. AUC of prediction is reported. Black line represents x = y.
(TIFF)

**S5 Fig. Prediction trained on sub-sampled IL cohort sizes using GBDT or Ridge tested on the US test cohort.** (a)—(b) Coefficient of determination ($R^2$) and standard deviation error bars of predictions of age (a) and BMI (b) obtained using a GBDT (purple) or Ridge regression (green) models trained on sub-samples of the cohort train IL, of different sizes, and tested of the whole test US cohort. For each cohort size *k*, 10 random sub-samples of *k* individuals were obtained and the mean and standard deviation of their predictions are shown.
(TIFF)

**S1 File.**
(DOCX)

**S2 File.**
(XLSX)

**S1 Dataset.**
(DOCX)

## Acknowledgments

We thank members of the Segal lab for useful discussions.

## Author Contributions

**Conceptualization:** Daphna Rothschild, Sigal Leviatan, Eran Segal.

**Data curation:** Daphna Rothschild, Sigal Leviatan.

**Formal analysis:** Daphna Rothschild, Sigal Leviatan, Ariel Hanemann, Yossi Cohen, Omer Weissbrod.

**Methodology:** Daphna Rothschild, Sigal Leviatan.

**Software:** Daphna Rothschild, Sigal Leviatan.

**Supervision:** Eran Segal.

**Validation:** Ariel Hanemann.

**Writing – original draft:** Daphna Rothschild, Sigal Leviatan.

**Writing – review & editing:** Daphna Rothschild, Sigal Leviatan, Eran Segal.

## References

1. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, et al. Richness of human gut microbiome correlates with metabolic markers. Nature. 2013; 500: 541–546. https://doi.org/10.1038/nature12506 PMID: 23985870

2. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, et al. A core gut microbiome in obese and lean twins. Nature. 2009; 457: 480–484. https://doi.org/10.1038/nature07540 PMID: 19043404

3. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature. 2012; 490: 55–60. https://doi.org/10.1038/nature11450 PMID: 23023125

4. Siljander H, Honkanen J, Knip M. Microbiome and type 1 diabetes. EBioMedicine. 2019; 46: 512–521. https://doi.org/10.1016/j.ebiom.2019.06.031 PMID: 31257149

5. Sanna S, van Zuydam NR, Mahajan A, Kurilshikov A, Vich Vila A, Võsa U, et al. Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. Nat Genet. 2019; 51: 600–605. https://doi.org/10.1038/s41588-019-0350-x PMID: 30778224

6. Goodrich JK, Davenport ER, Beaumont M, Jackson MA, Knight R, Ober C, et al. Genetic determinants of the gut microbiome in UK twins. Cell Host Microbe. 2016; 19: 731–743. https://doi.org/10.1016/j.chom.2016.04.017 PMID: 27173935

7. Larsen N, Vogensen FK, van den Berg FWJ, Nielsen DS, Andreasen AS, Pedersen BK, et al. Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. PLoS ONE. 2010; 5: e9085. https://doi.org/10.1371/journal.pone.0009085 PMID: 20140211

8. Zhernakova A, Kurilshikov A, Bonder MJ, Tigchelaar EF, Schirmer M, Vatanen T, et al. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science. 2016; 352: 565–569. https://doi.org/10.1126/science.aad3369 PMID: 27126040

9. Sze MA, Schloss PD. Looking for a signal in the noise: revisiting obesity and the microbiome. MBio. 2016; 7. https://doi.org/10.1128/mBio.01018-16 PMID: 27555308

**10.** Clemente JC, Ursell LK, Parfrey LW, Knight R. The impact of the gut microbiota on human health: an integrative view. Cell. 2012; 148: 1258–1270. https://doi.org/10.1016/j.cell.2012.01.035 PMID: 22424233

**11.** Scher JU, Ubeda C, Artacho A, Attur M, Isaac S, Reddy SM, et al. Decreased bacterial diversity characterizes the altered gut microbiota in patients with psoriatic arthritis, resembling dysbiosis in inflammatory bowel disease. Arthritis Rheumatol. 2015; 67: 128–139. https://doi.org/10.1002/art.38892 PMID: 25319745

**12.** Cotillard A, Kennedy SP, Kong LC, Prifti E, Pons N, Le Chatelier E, et al. Dietary intervention impact on gut microbial gene richness. Nature. 2013; 500: 585–588. https://doi.org/10.1038/nature12480 PMID: 23985875

**13.** Joossens M, Huys G, Cnockaert M, De Preter V, Verbeke K, Rutgeerts P, et al. Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. Gut. 2011; 60: 631–637. https://doi.org/10.1136/gut.2010.223263 PMID: 21209126

**14.** Tuddenham SA, Koay WLA, Zhao N, White JR, Ghanem KG, Sears CL, et al. The Impact of Human Immunodeficiency Virus Infection on Gut Microbiota α-Diversity: An Individual-level Meta-analysis. Clin Infect Dis. 2020; 70: 615–627. https://doi.org/10.1093/cid/ciz258 PMID: 30921452

**15.** Wilmanski T, Diener C, Rappaport N, Patwardhan S, Wiedrick J, Lapidus J, et al. Gut microbiome pattern reflects healthy ageing and predicts survival in humans. Nat Metab. 2021; 3: 274–286. https://doi.org/10.1038/s42255-021-00348-0 PMID: 33619379

**16.** Coyte KZ, Schluter J, Foster KR. The ecology of the microbiome: Networks, competition, and stability. Science. 2015; 350: 663–666. https://doi.org/10.1126/science.aad2602 PMID: 26542567

**17.** Srinivasan S, Hoffman NG, Morgan MT, Matsen FA, Fiedler TL, Hall RW, et al. Bacterial communities in women with bacterial vaginosis: high resolution phylogenetic analyses reveal relationships of microbiota to clinical criteria. PLoS ONE. 2012; 7: e37818. https://doi.org/10.1371/journal.pone.0037818 PMID: 22719852

**18.** Kong F, Hua Y, Zeng B, Ning R, Li Y, Zhao J. Gut microbiota signatures of longevity. Curr Biol. 2016; 26: R832–R833. https://doi.org/10.1016/j.cub.2016.08.015 PMID: 27676296

**19.** Kong F, Deng F, Li Y, Zhao J. Identification of gut microbiome signatures associated with longevity provides a promising modulation target for healthy aging. Gut Microbes. 2019; 10: 210–215. https://doi.org/10.1080/19490976.2018.1494102 PMID: 30142010

**20.** Lynch SV, Pedersen O. The human intestinal microbiome in health and disease. N Engl J Med. 2016; 375: 2369–2379. https://doi.org/10.1056/NEJMra1600266 PMID: 27974040

**21.** Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, et al. Population-level analysis of gut microbiome variation. Science. 2016; 352: 560–564. https://doi.org/10.1126/science.aad3503 PMID: 27126039

**22.** McDonald D, Hyde E, Debelius JW, Morton JT, Gonzalez A, Ackermann G, et al. American gut: an open platform for citizen science microbiome research. mSystems. 2018; 3. https://doi.org/10.1128/mSystems.00031-18 PMID: 29795809

**23.** Kurilshikov A, Medina-Gomez C, Bacigalupe R, Radjabzadeh D, Wang J, Demirkan A, et al. Genetics of human gut microbiome composition. BioRxiv. 2020. https://doi.org/10.1101/2020.06.26.173724

**24.** Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010; 42: 565–569. https://doi.org/10.1038/ng.608 PMID: 20562875

**25.** Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015; 12: e1001779. https://doi.org/10.1371/journal.pmed.1001779 PMID: 25826379

**26.** Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. Nature. 2000; 405: 299–304. https://doi.org/10.1038/35012500 PMID: 10830951

**27.** Smillie CS, Smith MB, Friedman J, Cordero OX, David LA, Alm EJ. Ecology drives a global network of gene exchange connecting the human microbiome. Nature. 2011; 480: 241–244. https://doi.org/10.1038/nature10571 PMID: 22037308

**28.** Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. Metagenomic microbial community profiling using unique clade-specific marker genes. Nat Methods. 2012; 9: 811–814. https://doi.org/10.1038/nmeth.2066 PMID: 22688413

**29.** Nayfach S, Rodriguez-Mueller B, Garud N, Pollard KS. An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. Genome Res. 2016; 26: 1612–1625. https://doi.org/10.1101/gr.201863.115 PMID: 27803195

**30.** Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, et al. Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. Cell. 2019; 176: 649–662.e20. https://doi.org/10.1016/j.cell.2019.01.001 PMID: 30661755

31. Morton JT, Marotz C, Washburne A, Silverman J, Zaramela LS, Edlund A, et al. Establishing microbial composition measurement standards with reference frames. Nat Commun. 2019; 10: 2719. https://doi.org/10.1038/s41467-019-10656-5 PMID: 31222023

32. Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D, et al. Environment dominates over host genetics in shaping human gut microbiota. Nature. 2018; 555: 210–215. https://doi.org/10.1038/nature25973 PMID: 29489753

33. Huang S, Haiminen N, Carrieri A-P, Hu R, Jiang L, Parida L, et al. Human skin, oral, and gut microbiomes predict chronological age. mSystems. 2020; 5. https://doi.org/10.1128/mSystems.00630-19 PMID: 32047061

34. Yang YC, Lu FH, Wu JS, Chang CJ. Age and sex effects on HbA1c. A study in a healthy Chinese population. Diabetes Care. 1997; 20: 988–991. https://doi.org/10.2337/diacare.20.6.988 PMID: 9167111

35. Lundberg SM, Lee S-I. A Unified Approach to Interpreting Model Predictions. 2017.

36. Yang J, Ferreira T, Morris AP, Medland SE, Genetic Investigation of ANthropometric Traits (GIANT) Consortium, DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet. 2012; 44: 369–75, S1. https://doi.org/10.1038/ng.2213 PMID: 22426310

37. Pasaniuc B, Price AL. Dissecting the genetics of complex traits using summary association statistics. Nat Rev Genet. 2017; 18: 117–127. https://doi.org/10.1038/nrg.2016.142 PMID: 27840428

38. de Luca C, Olefsky JM. Inflammation and insulin resistance. FEBS Lett. 2008; 582: 97–105. https://doi.org/10.1016/j.febslet.2007.11.057 PMID: 18053812

39. Mendes-Soares H, Raveh-Sadka T, Azulay S, Edens K, Ben-Shlomo Y, Cohen Y, et al. Assessment of a personalized approach to predicting postprandial glycemic responses to food among individuals without diabetes. JAMA Netw Open. 2019; 2: e188102. https://doi.org/10.1001/jamanetworkopen.2018.8102 PMID: 30735238

40. Suez J, Korem T, Zeevi D, Zilberman-Schapira G, Thaiss CA, Maza O, et al. Artificial sweeteners induce glucose intolerance by altering the gut microbiota. Nature. 2014; 514: 181–186. https://doi.org/10.1038/nature13793 PMID: 25231862

41. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30: 2114–2120. https://doi.org/10.1093/bioinformatics/btu170 PMID: 24695404

42. Marco-Sola S, Sammeth M, Guigó R, Ribeca P. The GEM mapper: fast, accurate and versatile alignment by filtration. Nat Methods. 2012; 9: 1185–1188. https://doi.org/10.1038/nmeth.2221 PMID: 23103880

43. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009; 10: R25. https://doi.org/10.1186/gb-2009-10-3-r25 PMID: 19261174

44. Zeevi D, Korem T, Godneva A, Bar N, Kurilshikov A, Lotan-Pompan M, et al. Structural variation in the gut microbiome associates with host health. Nature. 2019; 568: 43–48. https://doi.org/10.1038/s41586-019-1065-y PMID: 30918406

45. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol. 2016; 17: 132. https://doi.org/10.1186/s13059-016-0997-x PMID: 27323842

46. Douglas GM, Maffei VJ, Zaneveld JR, Yurgel SN, Brown JR, Taylor CM, et al. PICRUSt2 for prediction of metagenome functions. Nat Biotechnol. 2020; 38: 685–688. https://doi.org/10.1038/s41587-020-0548-6 PMID: 32483366

47. Ho DE, Imai K, King G, Stuart EA. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. J Stat Softw. 2011; 42. https://doi.org/10.18637/jss.v042.i08

48. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. Am J Hum Genet. 2011; 88: 294–305. https://doi.org/10.1016/j.ajhg.2011.02.002 PMID: 21376301

49. Golan D, Lander ES, Rosset S. Measuring missing heritability: inferring the contribution of common variants. Proc Natl Acad Sci USA. 2014; 111: E5272–81. https://doi.org/10.1073/pnas.1419064111 PMID: 25422463

50. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining—KDD '16. New York, New York, USA: ACM Press; 2016. pp. 785–794. https://doi.org/10.1145/2939672.2939785