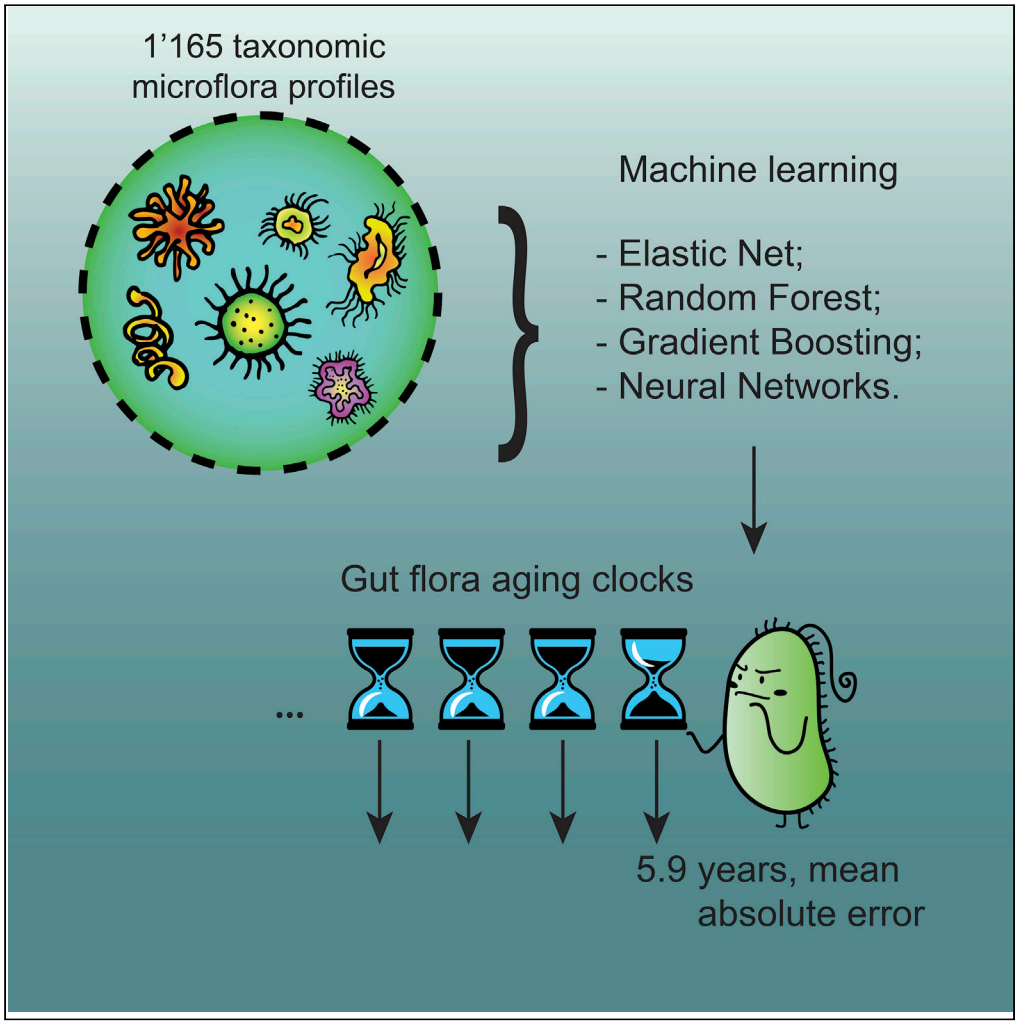
iScience

CellPress
OPEN ACCESS

Article

Human Gut Microbiome Aging Clock Based on Taxonomic Profiling and Deep Learning



Fedor Galkin, Polina Mamoshina, Alex Aliper, Evgeny Putin, Vladimir Moskalev, Vadim N. Gladyshev, Alex Zhavoronkov

alex@insilico.com

HIGHLIGHTS

DNNs are the most appropriate model to predict host age from gut microflora profiles

Our DNN models reach MAE of 5.9 years in independent verification

Feature importance analysis gives a starting point for anti-aging intervention design

Galkin et al., iScience *23*, 101199
June 26, 2020
https://doi.org/10.1016/j.isci.2020.101199

iScience

CellPress
OPEN ACCESS

Article

Human Gut Microbiome Aging Clock Based on Taxonomic Profiling and Deep Learning

Fedor Galkin,[1,2] Polina Mamoshina,[1,3] Alex Aliper,[3] Evgeny Putin,[3] Vladimir Moskalev,[3] Vadim N. Gladyshev,[4] and Alex Zhavoronkov[1,3,5,6,7,*]

SUMMARY

The human gut microbiome is a complex ecosystem that both affects and is affected by its host status. Previous metagenomic analyses of gut microflora revealed associations between specific microbes and host age. Nonetheless there was no reliable way to tell a host's age based on the gut community composition. Here we developed a method of predicting hosts' age based on microflora taxonomic profiles using a cross-study dataset and deep learning. Our best model has an architecture of a deep neural network that achieves the mean absolute error of 5.91 years when tested on external data. We further advance a procedure for inferring the role of particular microbes during human aging and defining them as potential aging biomarkers. The described intestinal clock represents a unique quantitative model of gut microflora aging and provides a starting point for building host aging and gut community succession into a single narrative.

INTRODUCTION

The human gut is colonized by a dense microbial community, calculated to consist of 10^{14} cells, which is an order of magnitude higher than the number of cells in the host (Suau et al., 1999). This gut microbiota is a complex ecosystem that carries multiple important functions in the organism such as digestion, immunity, and vitamin and other important metabolite production and even affects higher neural functions (Adak and Khan, 2019). The interaction, however, is not one-sided: microbiota is not simply determining certain host characteristics, as it also responds to signals from the host via multiple feedback loops (Lozupone et al., 2012).

The totality of processes gut microflora participates in cannot be dismissed as irrelevant. Detecting and interpreting the host-microbe interactions is essential to gain more control over host health status. Although host-microbe dialogue can be relatively easily detected with modern biological technology, its interpretation poses a much more challenging task. High-throughput methods allow to observe changes in microflora composition between host cohorts, but it is not always evident which of them are pathological and which represent transition between alternative healthy states. In this article we aspire to provide a tool that would let researchers account for normal changes associated with healthy aging.

Just as its host, human microflora matures and ages with years. This starts in infancy with gut colonization. Factors such as mode of delivery and diet (formula or breast milk) greatly affect the track along which pioneer microbiota progresses (Zhuang et al., 2019). But their effects are well documented and there are multiple studies that show consistent succession in the infant gut community (Koenig et al., 2011).

Microflora evolution does not stop in childhood but extends into adulthood and old age. However, adult hosts exhibit significantly more variable phenotypes and conditions that drive community dynamics. This variability obscures the view of microflora succession and leads to confusion and contradictory results—stark contrast with the case of infant microflora. For example, some studies report a general decrease in biodiversity as an indicator of old age, whereas other studies show that the healthy elderly may have microflora as diverse as that of the younger population (Bian et al., 2017; Nagpal et al., 2018).

Regardless of these complications, establishing regularities of intestinal succession in adults is of utmost importance, since deciphering them means grasping control over certain aspects of organismal aging.

[1]Deep Longevity Inc, Hong Kong Science and Technology Park, Hong Kong

[2]Integrative Genomics of Ageing Group, Institute of Ageing and Chronic Disease, University of Liverpool, Liverpool, UK

[3]Insilico Medicine Ltd, Hong Kong Science and Technology Park, Hong Kong

[4]Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, MA, USA

[5]Buck Institute for Research on Aging, Novato, CA, USA

[6]Biogerontology Research Foundation, London, UK

[7]Lead Contact

*Correspondence: alex@insilico.com

https://doi.org/10.1016/j.isci.2020.101199

This potential application of microflora studies has long entertained the minds of biogerontologists, but only recently the technology has allowed to see its outlines. Historically, the first theory of intestinal aging was developed by a Nobel Prize-winning Russian scientist Ilya Metchnikoff in the beginning of the 20th century (Bested et al., 2013). He proposed that malicious microbes processing undigested food, especially peptolytic bacteria, e.g., *Escherichia* and *Clostridium*, promote autointoxication. Treating autointoxication with pro- and pre-biotics, such as *Lactobacillus* preparations, was suggested as a way of alleviating the age-associated decline in organismal function. Recent studies have demonstrated promising results in line with this century-old hypothesis (Kaur et al., 2017).

Yet such findings are disorganized and need to be unified into a theory of gut community dynamics. One of the stepping stones to this theory would be a reliable way to measure the passage of time in gut community and the ability to tell two temporally different states apart—the very foundation of any dynamics theory. In this context, DNAm aging clocks provide an image of the solution. In essence, DNAm clocks illustrate that machine learning techniques can be used to define a new time dimension, which unlike chronological time, very conveniently can flow in both directions. This substitute time can be manipulated and used to tell apart biological systems that have been unequally affected by aging, even if they have the same chronological age. The success of this concept has inspired many other research groups to seek and interpret molecular footprints of aging in molecular-level features.

However, this quest has produced only limited results in the gut metagenomics field, despite there being a mass of reports on specific microbes' involvement with aging. The complexity of microbiome and its susceptibility to multiple variables apart from age complicate the basic task of aggregating the available information into an intestinal aging clock. Some researchers avoid cross-study designs, which on the one hand remove the batch effect problem, but on the other greatly reduce the power of resulting models. For example, a support vector machine model trained on human metagenomic data of 52 samples to classify samples as either young or old was shown to be only 10%–15% more accurate than the random assignment, as indicated by the area under the curve score (Lan et al., 2013). Another study, which attempted to use a co-abundance clustering approach, demonstrated general trendlines of microbiota composition for 367 individuals aged 0–100 years (Odamaki et al., 2016). According to the study, specific clades of the gut community differ significantly in abundance between different age groups. But despite the greater sample size the authors still render it impossible to put together a quantitative theory of intestinal aging, at least without controlling other important variables, such as diet.

In this study, we aim to deliver an accurate aging clock based on gut metagenomics. For this, we aggregated more than 4,000 metagenomic profiles from people aged 18–90 years. Moreover, we have used Deep Neural Network (DNN) as our model of choice. Its flexibility and ability to solve non-linear cases have made DNNs extremely useful in image, text, and voice recognition as well as in a number of biomedical applications (Aliper et al., 2016). Most recently, DNNs have permeated biogerontology, allowing the construction of aging clocks of superior quality using various data types: standard clinical blood tests (Mamoshina et al., 2018a), facial images (Bobrov et al., 2018), physical activity (Pyrkov et al., 2018), and transcriptomic data (Mamoshina et al., 2018b). As such, DNNs emerge as an optimal method to decipher the links between aging and microflora composition.

Combining an extensive cross-study data approach with DNN training, we managed to construct a clock that shows 5.91 years mean absolute error (MAE) when tested on an independent set of ~400 taxonomic profiles from healthy individuals. This accuracy is comparable with the existing DNAm solutions, which typically reach MAE <5 years. We further describe how the most abundant microbes (in terms of relative abundance) affect age prediction, which has let us identify understudied, yet important, species as potential biomarkers of intestinal aging.

RESULTS

Age Prediction Using Machine Learning

To examine the relationship between human gut taxonomic profiles and chronological age, we prepared a collection of full metagenome sequences for 1,165 healthy individuals (3,663 samples total) from 10 publicly available datasets. All individuals in our dataset were between 20 and 90 years old, with median age of 46 years (Figure S1). We applied four machine learning (ML) algorithms: Elastic Net (EN), Random Forest (RF),

		CV Host	Independent HC	Independent T1D
Neural networks	R^2	0.21 ± 0.13	0.29 ± 0.09	−0.98 ± 0.17
	MAE, years	10.60±1.12	5.91±0.40	18.02±0.82
	r	0.52 ± 0.11	0.53 ± 0.05	0.14 ± 0.08
	Baseline MAE, years	13.03	9.27	13.65
	Median age, years	50	49	33.5
	N, profiles	1,165	402	34
Machine learning	Elastic Net MAE, years	13.75±2.92	9.625±0.34	16.95±0.63
	Random Forest MAE, years	11.32 ± 4.18	7.06 ± 0.49	18.51±1.04
	XGBoost MAE, years	11.06 ± 3.91	7.11 ± 0.54	18.40±1.47

Table 1. Performance of the DNN Model on Independent Testing Sets with Healthy Individuals (HC) or Subjects with Type 1 Diabetes (T1D) and on Cross-Validation (CV) Test

Green highlights the smallest MAE in each column, and red, MAEs that are greater than baseline median age assignment (MAE).

MAE, mean absolute error; N, number. Standard deviations for the sample-based clocks are calculated over 5-fold, standard deviations for the host-based clocks are calculated over 10-fold. See Tables S1 and S2 for further details on predictions and Table S7 for metrics from more models.

Gradient Boosting (XGB), and DNN, to this dataset. All samples obtained from the same donor were collapsed into one averaged profile (1,165 profiles from 3,663 samples).

Both in cross-validation (CV) and independent verification DNNs produced the most accurate models (Table 1): 10.60 years MAE with baseline being 13.75 years (Tables S1). EN failed to predict better than median age assignment in both sample-based and host-based settings. Interestingly, DNNs also produce the most consistent models among folds, as indicated by the least standard deviations in MAE.

We then divided the samples into three age groups (20–39, 40–59, and 60–90 years) and found that the predicted age distribution generated by our DNN was significantly skewed toward the dataset median age (50 years) (Figure 1).

Independent Dataset Validation

To validate our aging models we have prepared a set of 436 metagenomic runs (18–70 years old) from three additional independent studies (PRJEB2054, PRJNA375935, PRJNA289586), one of which had 34 runs obtained from donors with type 1 diabetes (T1D) (PRJNA289586). Only the best models from each ML algorithm, as established in CV, were validated in this setting. As in CV, EN failed to produce better than baseline predictions. All other ML techniques produced results with MAE in the 7- to 8-year range with the exception of DNN, which predicted healthy donors with 5.91 years of MAE, compared with 9.27 years of MAE for median age assignment (Figure 2). Although most models outperformed baselines in healthy donors, they failed to achieve that with samples from diabetic people. Since some profiles in the verification set belong to the same donor (435 profiles for 271 donor), we also report that using MAE for averaged per donor predictions is 6.85 years for healthy people and 18.03 for diabetic people.

Unlike in CV, predictions obtained from individual folds are much more consistent for all models, which can indicate reduced data heterogeneity due to using only three studies instead of ten.

Interpreting DNN Clock Features

We applied the accumulated local effect (ALE) method (see Methods, Figure S2) to the validation set to compare how the DNN model treats its 100 most prevalent features (Figures 3 and S2). The overall

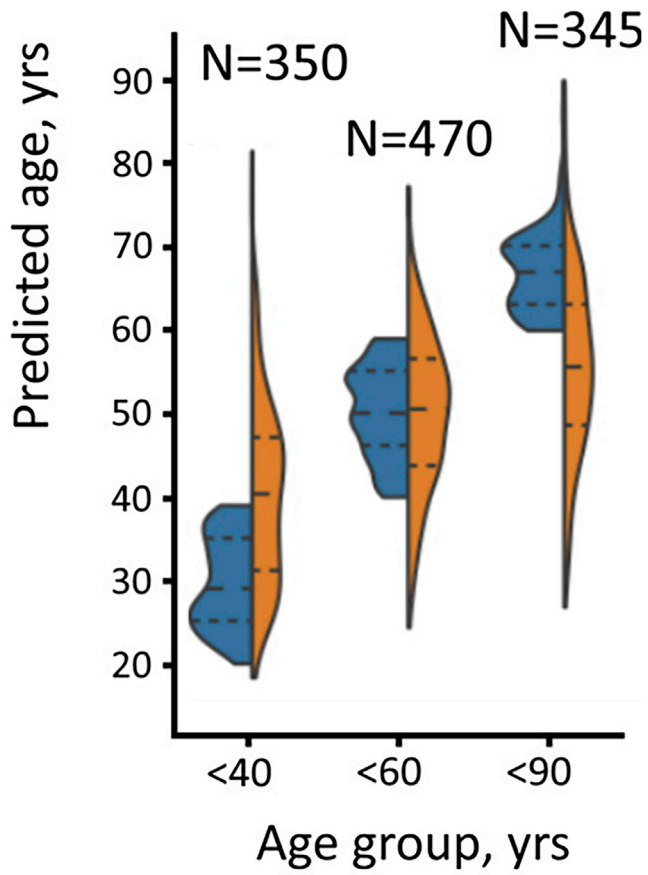


Figure 1. Predicted and Actual Age Density Distributions for Profiles Used during CV

Despite predictions being skewed toward median age (50 years), the predictor performs better than median age assignment (MAE = 13.75 years). "N" stands for the total number of samples per class. Dashed lines within violins stand for quartile borders. See Figure S1, S8 and Table S6 for original data structure.

contribution of any single microbe is less than 2 years in both models with most microbes being able to shift the prediction age by less than 0.5 years (Table S5).

Closer inspection shows that the microbes with greater influence (Figure 4) largely consist of well-studied participants of the gut community that may carry both beneficial (e.g., *Bifidobacterium spp.*, *Akkermansia muciniphila*, *Bacteroides spp.*) and negative (e.g., *Escherichia coli*, *Campylobacter jejuni*) traits. Some bacteria in the list, however, are rarely described in the context of human gut community (e.g., *Streptococcus equinus*, *Chryseobacterium gallinarum*, *Ornithobacterium rhinotracheale*).

Interestingly, within the tested dataset most microbes weakly affect the predictor until their relative abundance reaches a certain population-specific threshold. Exploring ALEs in a similar fashion for the training set of the host-based clock, we find that the steepest shifts in prediction also occur at the right end of microbe's abundance distribution (Tables S4 and S5).

DISCUSSION

This study describes a proof of concept aging clock based on microbiomic data. Among other ML algorithms, such as EN, RF, and XGB, DNN was shown to produce the best performing age predictors. A DNN trained on WGS microbiome data achieved an MAE of 10.60 years in CV and 5.91 during independent verification, which is significantly better than the baseline (median age assignment) (Table 1). Meanwhile, EN was shown to be unfit for the task of intestinal age prediction, since this ML method failed to produce better than baseline models in any setting we tested.

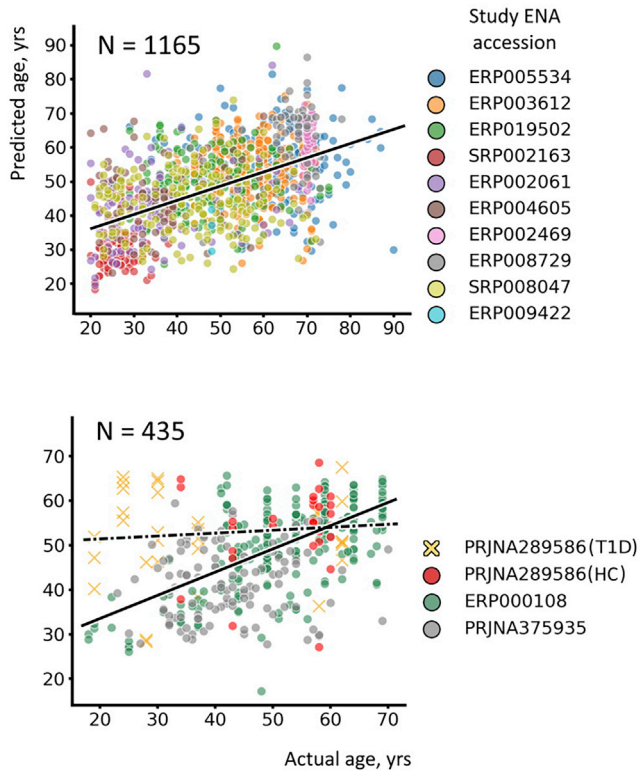


Figure 2. Microbiome Clock Predictions in CV (Top) and Independent Validation (Bottom)

Solid lines indicate linear fits for predictions derived from healthy individuals' (HC) samples: R2CV, Host = 0.64. Dash-dotted lines indicate linear fits for predictions derived from samples of people with type 1 diabetes (T1D). More metrics are available in Table 1. "N" stands for the number of metagenomic profiles used in either CV or verification stages. Note: independent verification panels have one outlier sample omitted (SRR6474267, actual age: 37 years, age predicted by the microbiome aging clock: 93 years). See Table S1 for individual predictions and Figure S9 for prediction plots of an alternative model.

Interestingly, all models assigned a higher predicted age to younger donors with T1D, which may indicate microbiome perturbations associated with this disease. Although these perturbations may be attributed to a specific diet and medications, we hypothesize that diabetes might also accelerate the aging process via a gut microbiome-associated mechanism.

Although the aging clocks' primary purpose is to measure the passage of time in biological systems, they can also be used to broaden our knowledge of the fundamental aging process. Using ALE, which explores the model's responses to minute changes in feature values, we identified gut prevalent microbes that are associated with aging. It should be noted that, although ALE is a method of feature importance analysis that can be applied to all ML algorithms, its results are not universal and depend both on the dataset structure and the specific model implementation.

Although keystone members of the gut community such as *Bacteroides*, *Eubacterium*, *Bifidobacterium* are shown to have the greatest effect on age prediction, some unexpected opportunistic and environmental bacteria also have an effect. For example, pathogenic *Campylobacter jejuni* abundance is associated with increased predicted age in the DNN model.

We also managed to show that many bacteria display a regular, monotonic effect on age prediction (Figure S4). More specifically, we show that most microbes either increase (seno-positive) or decrease (seno-negative) the estimated age. Meanwhile, microbes that are not consistently associated with changes in the predicted age occur only occasionally (Figure 4). However, a microbe's seno-status does not correspond to its beneficial or harmful function in the gut. Moreover, dataset structure can alter both the

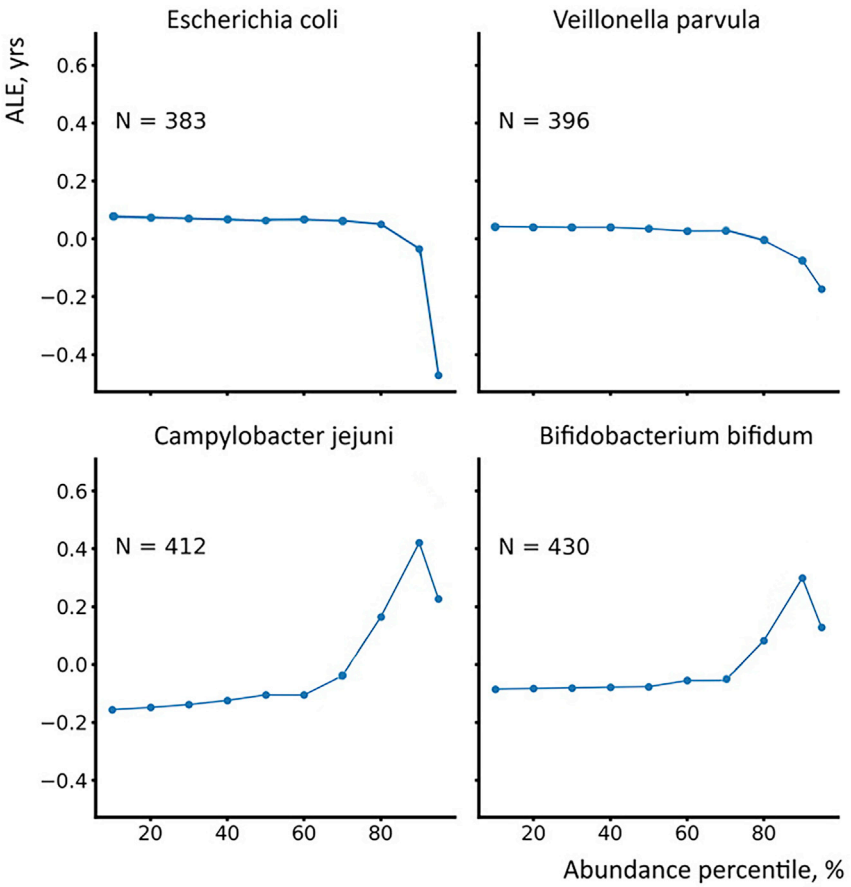


Figure 3. Accumulated Local Effect (ALE) Plots for Four Selected Microbial Taxa Derived from Perturbing Taxonomic Profiles in CV Dataset

ALE plots display how changes in a specific microbe's relative abundance (over a span of population-specific values) shift the model predictions toward acceleration or regression of age (in years); e.g., people with the abundance of *Veillonella parvula* higher than 95% of the population are predicted to be 0.25–0.3 years younger than those who have very little of this bacterium. See Figure S2 for ALE explanation, Figure S3 for an aggregated plot of more species, and Figure S10 for comparison with an alternative DNN model. Specific ALE values can be found in Tables S4 and S5.

amplitude and the direction of the effect, which means that ALEs should not be viewed as a tool for precise quantification. Nonetheless, ALEs can be used to provide a first glimpse into the prediction process of a model and derive hypotheses regarding of the nature underlying the observed regularities.

Among the 41 most important features for the DNN model (Figure 4) there were nine species belonging to the *Bacteroides/Parabacteroides* group. This reflects the high importance of the group for the gut community: their vast and frequently unique polysaccharide degrading abilities provide other members of the biota with otherwise unattainable sources of energy. However, only four of these species are seno-negative, which illustrates *Bacteroides* ambiguity regarding host-microbe interactions. For example, it contains highly virulent *B. fragilis*, which has the capacity to produce enterotoxins and cause life-threatening bacterial infections (Wexler, 2007). *B. thetaiotaomicron* and *B. ovatus* have also been associated with inflammatory gut conditions, whereas *B. dorei* might be related to occurrence of T1D in children (Davis-Richardson et al., 2014; Saitoh et al., 2002; Sitkin and Pokrotnieks, 2018). Meanwhile, various *Bacteroides* species contribute to the development of intestinal immune system, suppress pathogen growth, and form supporting trophic relations with beneficial bacteria (Sitkin and Pokrotnieks, 2018; Wexler, 2007). These circumstances make interpreting *Bacteroides* involvement in the aging process practically impossible. However, the over-representation of *Bacteroides* among the most important features indicates the need to more closely study these microbes within the aging context.

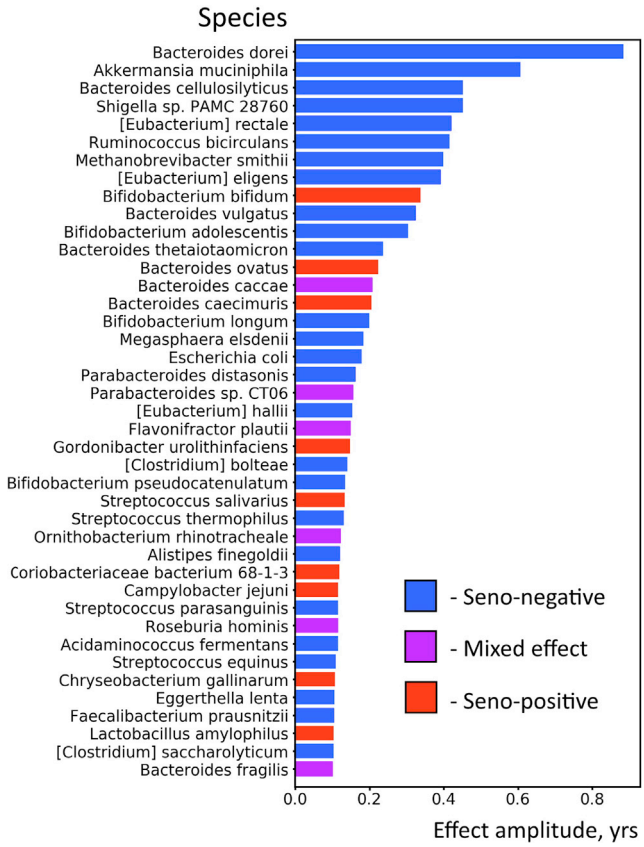


Figure 4. Microbial Species Ordered by the Shift in the Predicted Age They Can Cause in the DNN Model

Microbes whose increased relative abundance causes either a generally positive or negative shift are called seno-positive or seno-negative, respectively. Microbes that do not have a monotone effect on age prediction are called mixed. Only microbes whose effect amplitude is > 0.1 years are displayed (see Tables S4 and S5 and Figure S2).

Unlike the ambiguous *Bacteroides*, *Akkermansia muciniphila*—with the second largest effect amplitude—is commonly considered an indicator of a healthy gut community that is involved in regulating obesity, fat, and glucose metabolism, as well as regulating inflammation and mucin production (Naito et al., 2018). Some propose *A. muciniphila* as a new type of probiotic (Zhang et al., 2019). Interestingly, centenarian gut communities have a significantly higher abundance of *A. muciniphila* than other age groups, which makes it a likely candidate for a healthy aging biomarker (Biagi et al., 2016). Its high effect amplitude and seno-negative status, revealed by this study, indicate that it has an important role in intestinal aging. However, our collection of taxonomic profiles is scarce in old individuals and contains only one sample with the maximum age of 90 years. Further investigation into the role of *A. muciniphila* in aging should include a larger proportion of elderly people and centenarians.

Bifidobacterium genus is represented within the top features by three seno-negative species and one seno-positive. Generally, the seno-negative status of this genus is in line with previous findings, according to which *Bifidobacterium* tend to decrease with aging. Interestingly, the only seno-positive *Bifidobacterium bifidum* is mostly abundant in the elderly and centenarians. Thus, our findings further support that *Bifidobacterium* is directly involved into the aging process. *Bifidobacterium* probiotics convey multiple benefits to the host such as the ability to reduce inflammation in irritable bowel syndrome and prevent anxiety via the gut-brain axis (Arboleya et al., 2016). We hypothesize that *Bifidobacterium* might also possess certain geroprotective properties and should be studied in the context of biogerontology.

Another notable cluster in the ALEs top features are six species with documented butyrate-producing abilities (*Eubacterium rectale*, *E. eligens*, *E. hallii*, *Clostridium bolteae*, *Roseburia hominis*, *Faecalibacterium prausnitzii*) (Dehoux et al., 2016). Butyrate is an essential compound that regulates the gut immune system

and colon cancer apoptosis, as well as provides nutrition to colonocytes. It also affects gut-brain axis and possibly autism development. Meanwhile, a decrease in butyrate producers is characteristic of various dysbiotic states (Pequegnat et al., 2013; Rivière et al., 2016). Our results indicate that butyrate producers, akin to *A. muciniphila* or *Bifidobacterium*, can be an aging-related group and affect, or be affected by, the aging processes. Existing data show that butyrate producers do not consistently decrease in abundance with age, as we would expect from a collection of seno-negative species (Biagi et al., 2010). Although this may be attributed to sample differences, we would like to emphasize that a bacterium's seno-status should always be discussed while keeping in mind that it is determined by the performance of a DNN model that may depend on nonlinear patterns.

The presence of various bacteria that are poorly described within the human gut community context, namely, *Shigella* sp. PAMC 28760, *Gordonibacter urolithinfaciens*, *Streptococcus salivarius*, *Ornithobacterium rhinotracheale*, *Coriobacteriaceae bacterium* 68-1-3, *Streptococcus parasanguinis*, *Acidaminococcus fermentans*, *Streptococcus equinus*, *Chryseobacterium gallinarum*, *Lactobacillus amylophilus*, *Clostridium saccharolyticum*, among the above described features could be attributed to batch effects. However, there is a chance that these understudied taxa can still convey useful information regarding intestinal aging. To decide between these alternatives, the ALEs should be computed on an independent dataset of magnitude comparable with its training set while accounting for various meta variables.

The present finding shows that microbiome profiles can be used to predict the age of healthy individuals with relatively good accuracy. Moreover, patients with T1D exhibit age acceleration according to the microbiome clock. Further validation of this aging biomarker can lead to a valuable addition to the already existing aging clocks, e.g., epigenetic, blood biochemistry, and cell count clocks.

Hypothetical applications of microbiomic clocks may include "intestinal rejuvenation" with dietary interventions. This would require the ability to model microflora composition changes in response to such interventions. Previous studies have shown that microbiome can be manipulated by diets to a certain degree (Cotillard et al., 2013). But provided that microbiome itself is a major factor defining the outcome of dietary interventions, creating such a model demands a deeper understanding of gut community dynamics and metabolism before intestinal rejuvenation becomes a possibility (Zeevi et al., 2015).

Limitations of the Study

This study describes an early attempt to build an aging clock using microbiomic data from adult donors obtained elsewhere. Child microbiota profiles were not used during training, since child microbiota is expected to have significantly different dynamics and balancing the age distribution using child microbiota studies is particularly difficult. We tried predicting child intestinal age using the reported models, but they failed to perform better than baseline in this setting.

It should be noted that no normalization techniques were employed after obtaining the taxonomic profiles. Future work should be directed toward profile normalization to make the performance of such models more consistent across various datasets. However, considering how diverse human gut community is, the development of such normalization techniques poses a serious challenge in itself.

Further development of this work may include constructing an aging clock using genes or functional-level features. However, this direction requires experimentation with feature reduction methods, since the amount of genetic information contained within gut communities is overwhelming for the available sample sizes. The detected number of microbial species in this project already was too big for efficient training, and filtering out the minor species was used to bring down the number of dimensions. Such methods used in the genetic setting may cause multiple artifacts, and more sophisticated approaches are suggested. One of the promising approaches relies on Co-Abundance Group (CAG) calculation, as described in Minot and Willis (2019).

Another valuable extension to this project would be observing numerous pathologies through the lens of microbiome aging. Multiple biological processes are considered major drivers of aging, but gut microbiome's involvement in this process is still vaguely understood. Approaches such as upgrading the DNN architecture to have two output nodes (separate for age and disease) or predicting the health status using the latent representation obtained from an aging clock DNN may prove to be useful for this task.

Resource Availability

Lead Contact

Further information and requests should be directed to and will be fulfilled by the Lead Contact Alex Zhavoronkov (alex@insilico.com).

Data and Code Availability

All data used in this article come from publicly available sources. Metagenomic studies' identifiers in ENA: ERP002061, ERP008729, SRP002163, ERP004605, ERP003612, ERP002469, ERP019502, SRP008047, ERP009422, ERP005534, PRJEB2054, PRJNA375935, PRJNA289586.Only healthy individuals with age metadata available have been included in this study. These studies contain healthy individuals from Austria, China, Denmark, France, Germany, Kazakhstan, Spain, Sweden, and USA aged 20–90 years. For more details on data selection and preparation see Supplemental Information. ENA identifiers of the runs used are specified in Tables S1, S2, S3, S4, and S5.

The models are available at aging.ai in the Floro'clock section.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101199.

ACKNOWLEDGMENTS

Graphical abstract for this publication contains free vector art created by Alice Noir from the Noun Project and by eucalyp from Flaticon. The bacterium pondering the passage of time is adapted from an illustration by John Chang from American Society for Microbiology.

AUTHOR CONTRIBUTIONS

Conceptualization, A.Z. and A.A.; Methodology, F.G., E.P., and V.M.; Validation, F.G.; Formal Analysis, F.G., E.P., and V.M.; Resources, A.Z.; Data Curation, F.G.; Writing – Original Draft, F.G. and P.M.; Writing – Review and Editing, F.G., P.M., and V.N.G.; Visualization, F.G. and P.M.; Supervision, A.A., A.Z., and V.N.G.; Project Administration, P.M.

DECLARATION OF INTERESTS

F.G., P.M., and A.Z. are affiliated with Deep Longevity, Inc, a for-profit company developing biomarkers of aging and longevity commonly referred to as the "aging clocks. A.A., E.P., V.M., and A.Z. work for Insilico Medicine, a for-profit longevity biotechnology company developing the end-to-end target identification and drug discovery pipeline for a broad spectrum of age-related diseases. The microbiome aging model described in this article is a patent pending technology (application number 20200075127). The microbiomic aging clock is integrated in the Aging.AI system operated by the company. The company may have commercial interests in this publication.

Received: September 5, 2019
Revised: April 14, 2020
Accepted: May 21, 2020
Published: June 26, 2020

REFERENCES

Adak, A., and Khan, M.R. (2019). An insight into gut microbiota and its functionalities. Cell. Mol. Life Sci. *76*, 473–493.

Aliper, A., Plis, S., Artemov, A., Ulloa, A., Mamoshina, P., and Zhavoronkov, A. (2016). Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. Mol. Pharm. *13*, 2524–2530.

Arboleya, S., Watkins, C., Stanton, C., and Ross, R.P. (2016). Gut bifidobacteria populations in human health and aging. Front. Microbiol. *7*, 1204.

Bested, A.C., Logan, A.C., and Selhub, E.M. (2013). Intestinal microbiota, probiotics and mental health: from Metchnikoff to modern advances: Part I - autointoxication revisited. Gut Pathog. *5*, 5.

Biagi, E., Franceschi, C., Rampelli, S., Severgnini, M., Ostan, R., Turroni, S., Consolandi, C., Quercia,

S., Scurti, M., Monti, D., et al. (2016). Gut microbiota and extreme longevity. Curr. Biol. *26*, 1480–1485.

Biagi, E., Nylund, L., Candela, M., Ostan, R., Bucci, L., Pini, E., Nikkïla, J., Monti, D., Satokari, R., Franceschi, C., et al. (2010). Through ageing, and beyond: gut microbiota and inflammatory status in seniors and centenarians. PLoS One *5*, e10667.

Bian, G., Gloor, G.B., Gong, A., Jia, C., Zhang, W., Hu, J., Zhang, H., Zhang, Y., Zhou, Z., Zhang, J., et al. (2017). The gut microbiota of healthy aged Chinese is similar to that of the healthy young. mSphere *2*, e00327–17.

Bobrov, E., Georgievskaya, A., Kiselev, K., Sevastopolsky, A., Zhavoronkov, A., Gurov, S., Rudakov, K., Del Pilar Bonilla Tobar, M., Jaspers, S., and Clemann, S. (2018). PhotoAgeClock: deep learning algorithms for development of non-invasive visual biomarkers of aging. Aging (Albany. NY) *10*, 3249–3259.

Cotillard, A., Kennedy, S.P., Kong, L.C., Prifti, E., Pons, N., Le Chatelier, E., Almeida, M., Quinquis, B., Levenez, F., Galleron, N., et al. (2013). Dietary intervention impact on gut microbial gene richness. Nature *500*, 585–588.

Davis-Richardson, A.G., Ardissone, A.N., Dias, R., Simell, V., Leonard, M.T., Kemppainen, K.M., Drew, J.C., Schatz, D., Atkinson, M.A., Kolaczkowski, B., et al. (2014). Bacteroides dorei dominates gut microbiome prior to autoimmunity in Finnish children at high risk for type 1 diabetes. Front. Microbiol. *5*, 678.

Dehoux, P., Marvaud, J.C., Abouelleil, A., Earl, A.M., Lambert, T., and Dauga, C. (2016). Comparative genomics of *Clostridium bolteae* and *Clostridium clostridioforme* reveals species-specific genomic properties and numerous putative antibiotic resistance determinants. BMC Genomics *17*.

Kaur, H., Das, C., and Mande, S.S. (2017). In silico analysis of putrefaction pathways in bacteria and its implication in colorectal cancer. Front. Microbiol. *8*, 2166.

Koenig, J.E., Spor, A., Scalfone, N., Fricker, A.D., Stombaugh, J., Knight, R., Angenent, L.T., and Ley, R.E. (2011). Succession of microbial consortia in the developing infant gut microbiome. Proc. Natl. Acad. Sci. U S A *108*, 4578–4585.

Lan, Y., Kriete, A., and Rosen, G.L. (2013). Selecting age-related functional characteristics in the human gut microbiome. Microbiome *1*, 2.

Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K., and Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. Nature *489*, 220–230.

Mamoshina, P., Kochetov, K., Putin, E., Cortese, F., Aliper, A., Lee, W.-S., Ahn, S.-M., Uhn, L., Skjodt, N., Kovalchuk, O., et al. (2018a). Population specific biomarkers of human aging: a big data study using south Korean, Canadian, and eastern European patient populations. J. Gerontol. Ser. A *73*, 1482–1490.

Mamoshina, P., Volosnikova, M., Ozerov, I.V., Putin, E., Skibina, E., Cortese, F., and Zhavoronkov, A. (2018b). Machine learning on human muscle transcriptomic data for biomarker discovery and tissue-specific drug target identification. Front. Genet. *9*, https://doi.org/10.3389/fgene.2018.00242.

Minot, S.S., and Willis, A.D. (2019). Clustering co-abundant genes identifies components of the gut microbiome that are reproducibly associated with colorectal cancer and inflammatory bowel disease. Microbiome. https://doi.org/10.1186/s40168-019-0722-6.

Nagpal, R., Mainali, R., Ahmadi, S., Wang, S., Singh, R., Kavanagh, K., Kitzman, D.W., Kushugulova, A., Marotta, F., and Yadav, H. (2018). Gut microbiome and aging: physiological and mechanistic insights. Nutr. Heal. Aging *4*, 267–285.

Naito, Y., Uchiyama, K., and Takagi, T. (2018). A next-generation beneficial microbe: *Akkermansia muciniphila*. J. Clin. Biochem. Nutr. *63*, 33–35.

Odamaki, T., Kato, K., Sugahara, H., Hashikura, N., Takahashi, S., Xiao, J.-Z., Abe, F., and Osawa, R. (2016). Age-related changes in gut microbiota composition from newborn to centenarian: a cross-sectional study. BMC Microbiol. *16*, 90.

Pequegnat, B., Sagermann, M., Valliani, M., Toh, M., Chow, H., Allen-Vercoe, E., and Monteiro, M.A. (2013). A vaccine and diagnostic target for *Clostridium bolteae*, an autism-associated bacterium. Vaccine *31*, 2787–2790.

Pyrkov, T.V., Slipensky, K., Barg, M., Kondrashin, A., Zhurov, B., Zenin, A., Pyatnitskiy, M., Menshikov, L., Markov, S., and Fedichev, P.O. (2018). Extracting biological age from biomedical data via deep learning: too much of a good thing? Sci. Rep. *8*, 5210.

Rivière, A., Selak, M., Lantin, D., Leroy, F., and De Vuyst, L. (2016). Bifidobacteria and butyrate-producing colon bacteria: importance and strategies for their stimulation in the human gut. Front. Microbiol. *7*, 979.

Saitoh, S., Noda, S., Aiba, Y., Takagi, A., Sakamoto, M., Benno, Y., and Koga, Y. (2002). Bacteroides ovatus as the predominant commensal intestinal microbe causing a systemic antibody response in inflammatory bowel disease. Clin. Diagn. Lab. Immunol. *9*, 54.

Sitkin, S., and Pokrotnieks, J. (2018). Gut microbiota as a host defender and a foe: the 2 faces of commensal *Bacteroides thetaiotaomicron* in inflammatory bowel disease. Inflamm. Bowel Dis. https://doi.org/10.1093/ibd/izy377.

Suau, A., Bonnet, R., Sutren, M., Godon, J.J., Gibson, G.R., Collins, M.D., and Doré, J. (1999). Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. Appl. Environ. Microbiol. *65*, 4799–4807.

Wexler, H.M. (2007). Bacteroides: the good, the bad, and the nitty-gritty. Clin. Microbiol. Rev. *20*, 593–621.

Zeevi, D., Korem, T., Zmora, N., Israeli, D., Rothschild, D., Weinberger, A., Ben-Yacov, O., Lador, D., Avnit-Sagi, T., Lotan-Pompan, M., et al. (2015). Personalized nutrition by prediction of glycemic responses. Cell *163*, 1079–1094.

Zhang, T., Li, Q., Cheng, L., Buch, H., and Zhang, F. (2019). *Akkermansia muciniphila* is a promising probiotic. Microb. Biotechnol. *1751-7915*, 13410.

Zhuang, L., Chen, H., Zhang, S., Zhuang, J., Li, Q., and Feng, Z. (2019). Intestinal microbiota in early life and its implications on childhood health. Genomics. Proteomics Bioinformatics *17*, 13–25.

iScience, Volume *23*

Supplemental Information

Human Gut Microbiome Aging Clock

Based on Taxonomic Profiling and Deep Learning

Fedor Galkin, Polina Mamoshina, Alex Aliper, Evgeny Putin, Vladimir Moskalev, Vadim N. Gladyshev, and Alex Zhavoronkov

Supplementary Materials

Tables S1 to S5 are in a separate MS Excel Files

Transparent Methods

Data description

To construct the reported model, we used profiles obtained from 1165 donors and 13 studies (**Fig. S1**). All these studies are deposited to ENA and contain Illumina-generated WGS fresh-frozen stool runs (**TableS6**). Only healthy (or control groups in case-control settings) were used for training, and in case multiple profiles were available for the same donor, they were collapsed into one average profile.

The region of origin for 63% of donors is Europe, for the other 28% — Asia, and for the rest 8% — US. 53% of donors are female, 47% are male. BMI is known for 1135 donors, of them 561 are overweight (BMI>25 kg/m^2), 102 are underweight (BMI<20 kg/m^2), 271 are obese (BMI>30 kg/m^2) the rest (421) are normalweight (BMI ∈20-25 kg/m^2).

To define the CV cohorts we limited our selection of studies to only those performed with Illumina platforms with the exclusion of Illumina MiSeq. Then we performed tSNE analysis to locate the clouds of studies we could group together (learning rate = 200, perplexity = 30, random seed = 0, profiles processed with sklearn StandardScaler prior to tSNE, tSNE performed with sklearn's package manifold) (**Fig. S6**). After tSNE we excluded Illumina HiSeq 2500 platform, which unfortunately included ERP015317 (TwinsUK) study with 3288 runs, since all its samples formed a detached cluster on the tSNE plot (**Fig. S7, right**).

Files with <100’000 reads after QC were discarded, and ultimately we used 3663 files for 1165 donors, where the median file contained 4.6mln reads (**Fig. S8**).

Quality control details

k-mer entropy filtering was carried out with k-mer length of 4 nt as described in (Plaza Onate et al., 2015). To calculate normalized Shannon entropy “entropy” function from “scipy.stats” package was used. H=98% was used as the cutoof value. DSK was used to count k-mers (Rizk et al., 2013).

Quality trimming and filtering were carried out with relaxed settings to increase the amount of information available to the classifier and ultimately — DNN. BBDuk quality trimming was carried out on both ends to eliminate all nucleotides with Phred score < 4 (60% correct detection), subsequently all trimmed reads with average Phred score <8 (84% correct detection) were removed. Additionally, adapters were checked and removed using BBMerge method based on pair overlap detection. Host filtering was based on Bowtie2 default aligning with hg19 and only reads with both ends unmapped were used (Bushnell, 2016; Langmead and Salzberg, 2012).

Only runs with >1e+5 reads were used in subsequent work. The overall number of discarded reads typically did not exceed 10%.

Relative abundance calculation

All acquired sequencing files have been quality trimmed and quality filtered with BBTools (Bushnell, 2016). Human sequences have been detected using hg19 genome index. Additionally, specimen dilution test has been carried out as specified in (Plaza Onate et al., 2015). Resulting

reads have been analyzed with Centrifuge and mapped against the collection of bacterial and archaeal genomes (Kim et al., 2016).

In certain cases, abundance tables have been modified to exclude unreliably detected microbes (relative abundance < 1e-5) and minor microbial species (<1.3e-3 prevalence). No sample has lost more than 5% of its abundance. After all the modifications, individual taxonomic profiles have been renormalized to add up to one.

DNN training

All deep neural networks (DNNs) were implemented using the Python 3.6 Keras library with Tensorflow backend. DNNs were trained using a full list of species-level features, which includes 1,673 microbial taxa. Optimal architecture was established with grid search.

Two regressors were built: one using taxonomic profiles derived from individual samples (sample-based model) and a second one using taxonomic profiles averaged among all the samples belonging to the same host (host-based model). The former one was trained with five-fold CV, while the latter one — with ten-fold CV. Folds were stratified to maintain the proportion of studies observed in the complete data set. After completing grid search for various model configurations, the best performing model was selected based on MAE.

It contains three hidden layers with 512 nodes in each, with PReLU activation function, Adam optimizer, dropout fraction 0.5 at each layer, and 0.001 learning rate. The host-based model has similar settings except for having 1024 nodes in each hidden layer (**Fig. S5**).

Machine Learning Models

All other machine learning models were implemented using the following Python 3.6 libraries: sklearn.linear_model.ElasticNet, sklearn.ensemble. RandomForestRegressor, xgboost.

All models were trained in a five-fold CV setting with folds stratified to keep the proportion of studies similar to that in the whole data set.

The best performing models in CV have the following parameters:

- EN sample-based: {‘alpha’ : 0.1, ‘l1_ratio’ : 1.0};
- EN host-based: {‘alpha’ : 0.1, ‘l1_ratio’ : 1.0};
- RF sample-based: {‘max_features’ : ‘auto’, ‘min_samples_leaf’ : 1, ‘n_estimators’ : 700};
- RF host-based: {‘max_features’ : ‘auto’, ‘min_samples_leaf’ : 1, ‘n_estimators’ : 100};
- XGB sample-based: { ‘alpha’ : 0.5, ‘eta’ : 0.1. ‘eval_metric’ : ‘mae’, ‘gamma’ : 1.0, ‘lambda’ : 0.5, ‘max_delta_step’ : 0, ‘max_depth’ : 6};
- XGB host-based: { ‘alpha’ : 1.0, ‘eta’ : 0.1. ‘eval_metric’ : ‘mae’, ‘gamma’ : 1.0, ‘lambda’ : 1.0, ‘max_delta_step’ : 0, ‘max_depth’ : 6};

Accumulated local effects

The features deemed most important have been further assessed with the Accumulated Local Effects (ALE) method to determine the change in age prediction upon minor changes in a microbial species abundance. ALE has been implemented following the algorithm described below. For each of the selected species, a quantile value table (with 5% steps) has been composed. Local Effects (LE) for each quantile bin have been calculated by measuring the average change in prediction upon substituting observed abundance of a feature, with right and left bin border values. ALEs for each quantile are calculated by adding up all the previous LEs

and centering the result to make the average effect of each taxon zero (**Fig. S2**). All illustrations contain ALEs only within 5-95% range of a feature to remove outliers.

Sample-based model

We explored whether a model trained on the full set of 3663 samples, without paying attention to their donor of origin. This way the training set was significantly expanded, although sample information was leaking between the training and testing sets of each fold. Our hypothesis was that oversampling in this fashion may prove beneficial during independent verification.

As expected, MAE for this sample-based model was lower than for the host-based, which is reported in the main text (**Table S7**). Importantly, superior results obtained for the sample-based model (MAE = 3.94 years in CV) did not hold in independent verification, where the best performing sample-based DNN showed MAE similar to other ML methods — 7.28 years (**Fig. S9**).

Predicted intestinal age for sample-based DNN positively correlated (r = 0.23) with BMI, which is in line with existing data on connections between BMI and biological age (Yoo et al., 2017). However, this correlation was lower than that between donor BMI and observed age: r = 0.3.

We also carried out ALE-analysis for the sample-based DNN (**Fig. S3, S10**), and found that ALE vectors were in many cases significantly correlated across both models (r= 0.49), as well as their total effect amplitudes (r = 0.73). These results from independent verification indicate that the method of oversampling by using multiple

taxonomic profiles from the same donor as independent may produce biologically relevant models despite possible overfitting issues.

Linear Mixed Effects

We used R's *lme4* package to build a mixed-effects model for predicted age both in CV (Age ~ Age_DNN + Sex + 1|Study) and in verification (Age ~ Age_DNN + Sex + Diagnosis + 1|Study). In both cases the slope coefficient is positive within three standard deviations: 0.10±0.03 and 0.58±0.05, respectively. Study random effect std is 13.48 and 3.28, respectively.

16S model

We tried to use 16S data as our starting point, since they are more lightweight and are easier to analyze than WGS-data. We used a QIIME2 based pipeline comprising DADA2 and RDP-classifier to obtain genus-level features (Bolyen et al., 2019; Callahan et al., 2016; Wang et al., 2007). We used data from the American Gut Project (AGP): a total of 300 most prevalent geni in 8522 samples. On average only 0.46% of the community did not belong to any defined genus and were grouped into a separate feature. AGP data was additionally bloom-filtered, as suggested by its creators (Amir et al., 2017). Sex and BMI were added as predictive features as well.

The best performing model obtained with deep learning had MAE=10.8 years in verification with baseline being 12.45 years for the verification set. However, we have been unable to reproduce this in any other 16S dataset, probably due to the specifics of AGP and the suggested deblooming preprocessing.

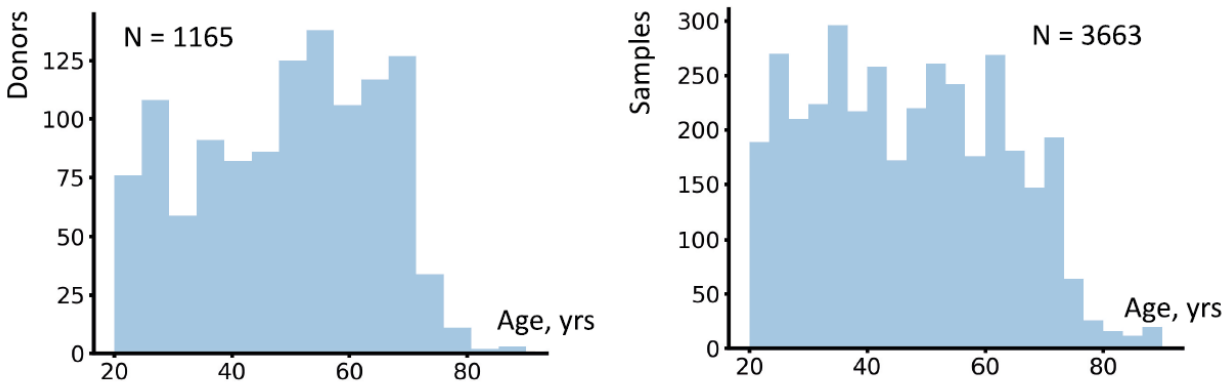


Figure S1. Related to Figure1. Age distribution in training data for sample-based (A) and host-based (B) aging clocks.

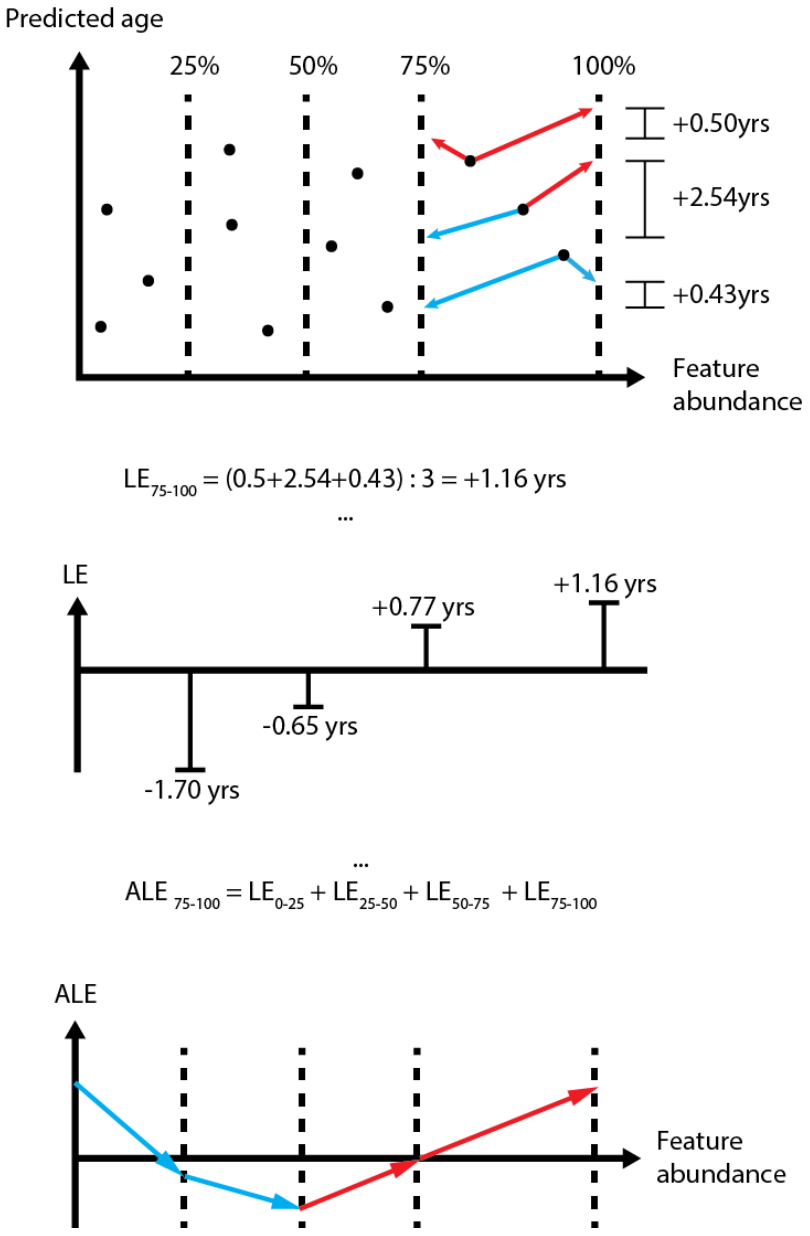
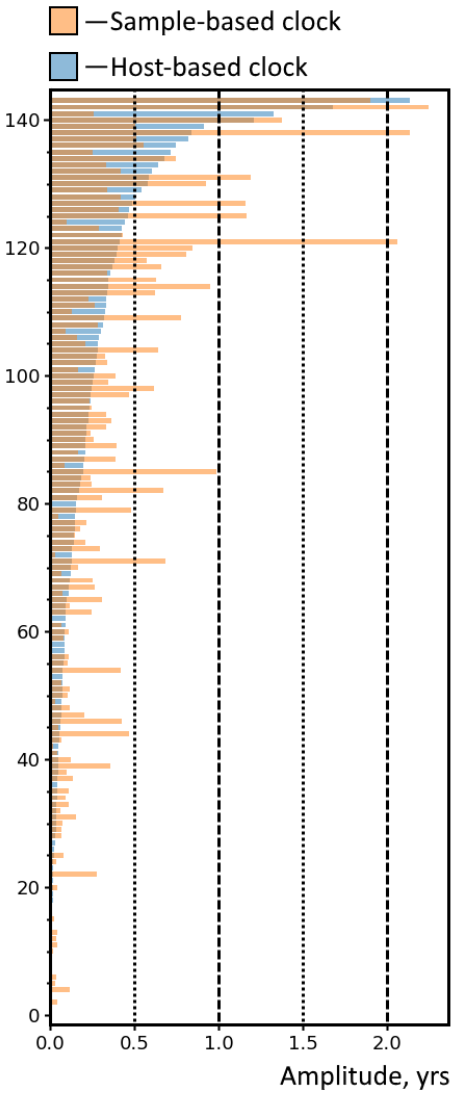


Figure S2. **Related to Figures 3,4** Accumulated Local Effects (ALE) method used in this study to assess specific taxa influence on age prediction. Changes in predicted age upon substituting observed taxon abundance with quantile values are averaged and recorded for every quantile bin. Then, they are summed to produce ALEs, which are additionally centered for convenience.



1. Streptococcus lutetiensis, 2. Treponema succinifaciens, 3. Streptococcus infantarius,
4. Erysipelotrichaceae bacterium I46, 5. Brachyspira pilosicoli, 6. Streptococcus sp. I-G2,
7. Mogibacterium diversum, 8. Enterococcus faecium, 9. Actinomyces pacaensis,
10. Streptococcus sanguinis, 11. Shigella flexneri, 12. Phoenicibacter massiliensis,
13. Lachnoclostridium sp. YL32, 14. Streptococcus oralis, 15. Burkholderiales bacterium YL45,
16. Lachnospiraceae bacterium oral taxon 500, 17. Streptococcus pasteurianus,
18. [Eubacterium] sulci, 19. Bifidobacterium breve, 20. Enterococcus faecalis,
21. Streptococcus pneumoniae, 22. Bifidobacterium catenulatum, 23. Mageeibacillus indolicus,
24. Streptococcus mitis, 25. Lawsonia intracellularis, 26. Streptococcus sp. oral taxon 431,
27. Bacteroides zoogleoformans, 28. Streptococcus sp. A12, 29. Prevotella dentalis,
30. Rothia mucilaginosa, 31. Clostridium kluyveri, 32. Leuconostoc mesenteroides,
33. Dialister pneumosintes, 34. Peptoniphilus sp. ING2-D1G, 35. Lactobacillus amylovorus,
36. Lactobacillus agilis, 37. Streptococcus agalactiae, 38. Streptococcus gallolyticus,
39. Lactococcus lactis, 40. Parvimonas micra, 41. Eubacterium limosum,
42. Bifidobacterium angulatum, 43. Haemophilus pittmaniae, 44. Bacteroides salanitronis,
45. Blautia hansenii, 46. Eggerthella sp. YY7918, 47. Campylobacter coli,
48. Prevotella jejuni, 49. Ruminiclostridium sp. KB18, 50. Victivallales bacterium CCUG 44730,
51. Klebsiella pneumoniae, 52. Megasphaera elsdenii, 53. Clostridioides difficile,
54. Bifidobacterium thermophilum, 55. Lachnoclostridium phocaeense,
56. Ruminococcaceae bacterium CPB6, 57. Mobiluncus curtisii, 58. Oscillibacter valericigenes,
59. Cloacibacillus porcorum, 60. Prevotella melaninogenica, 61. Streptococcus sp. I-P16,
62. Blautia sp. YL58, 63. Shigella sp. PAMC 28760, 64. Streptococcus acidominimus,
65. Clostridium sp. SY8519, 66. Streptococcus suis, 67. Coriobacteriaceae bacterium 68-1-3,
68. Murdochiella vaginalis, 69. Adlercreutzia equolifaciens, 70. Streptococcus mutans,
71. Lactobacillus gasseri, 72. Bifidobacterium kashiwanohense, 73. Streptococcus gordonii,
74. Shigella dysenteriae, 75. Parabacteroides sp. YL27, 76. Campylobacter lanienae,
77. Riemerella anatipestifer, 78. Streptococcus anginosus, 79. Clostridium cochlearium,
80. Myroides odoratimimus, 81. Prevotella enoeca, 82. Lactobacillus ruminis,
83. Mordavella sp. Marseille-P3756, 84. Turicibacter sp. H121, 85. Bifidobacterium longum,
86. Bacteroides heparinolyticus, 87. Bacteroides ovatus, 88. Bifidobacterium adolescentis,
89. Bacteroides helcogenes, 90. Parabacteroides distasonis, 91. Chryseobacterium taklimakanense,
92. Veillonella parvula, 93. Prevotella intermedia, 94. Salmonella enterica,
95. Streptococcus constellatus, 96. Anaerostipes hadrus, 97. Porphyromonas gingivalis,
98. Odoribacter splanchnicus, 99. Faecalibacterium prausnitzii, 100. [Clostridium] bolteae,
101. Clostridium perfringens, 102. Bifidobacterium animalis, 103. Intestinimonas butyriciproducens,
104. Christensenella massiliensis, 105. Bifidobacterium pseudocatenulatum,
106. Collinsella aerofaciens, 107. Streptococcus sp. FDAARGOS_192, 108. Bacteroides fragilis,
109. Haemophilus parainfluenzae, 110. Streptococcus thermophilus, 111. Roseburia hominis,
112. Klebsiella sp. 2N3, 113. Ornithobacterium rhinotracheale, 114. [Clostridium] saccharolyticum,
115. Bacteroides caecimuris, 116. Parabacteroides sp. CT06, 117. Streptococcus parasanguinis,
118. Bifidobacterium bifidum, 119. Eggerthella lenta, 120. Bifidobacterium dentium,
121. Chryseobacterium gallinarum, 122. Bacteroides thetaiotaomicron, 123. Ruminococcus bicirculans,
124. Desulfovibrio piger, 125. Alistipes finegoldii, 126. Barnesiella viscericola,
127. Lactobacillus amylophilus, 128. Flavonifractor plautii, 129. Escherichia coli,
130. Campylobacter jejuni, 131. Streptococcus salivarius, 132. Gordonibacter urolithinfaciens,
133. Streptococcus equinus, 134. Bacteroides caccae, 135. [Eubacterium] rectale,
136. Methanobrevibacter smithii, 137. Bacteroides cellulosilyticus, 138. [Eubacterium] eligens,
139. [Eubacterium] hallii, 140. Acidaminococcus intestini, 141. Akkermansia muciniphila,
142. Bacteroides vulgatus, 143. Bacteroides dorei,

Figure S3. Related to Figures 3,4. Amplitude of effects of 143 most prevalent bacteria (prevalence > 25%) on aging models. In the host-based clock changes in abundance of only 3 bacteria can shift the predictions by >1 year on average, while in the sample-based model there are 8 such species. Related to Figure 4

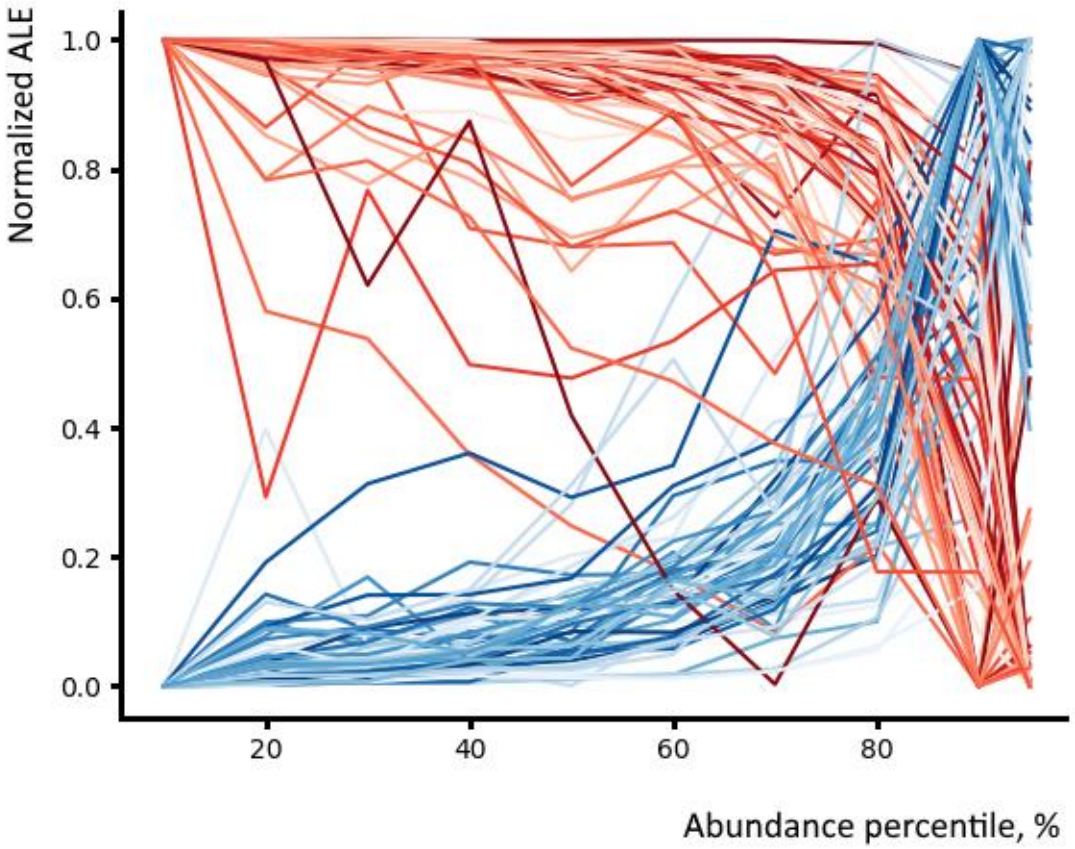


Figure S4. Related to Figure 3. Overlayed ALE plots of seno-positive (blue hues) and seno-negative (red hues) microbes. In most people, minor perturbations in a specific taxon's abundance only weakly affect the predicted age, as indicated by most ALE plots steeply increasing or decreasing once a microbe's abundance reaches values higher than in a 60-80% of the population.

Normalized ALE [$Y = ALE(X) - ALE_{min} / ALE_{max}\text{-}ALE_{min}$] was obtained for 108 highly prevalent microbes in 1,165 hosts.

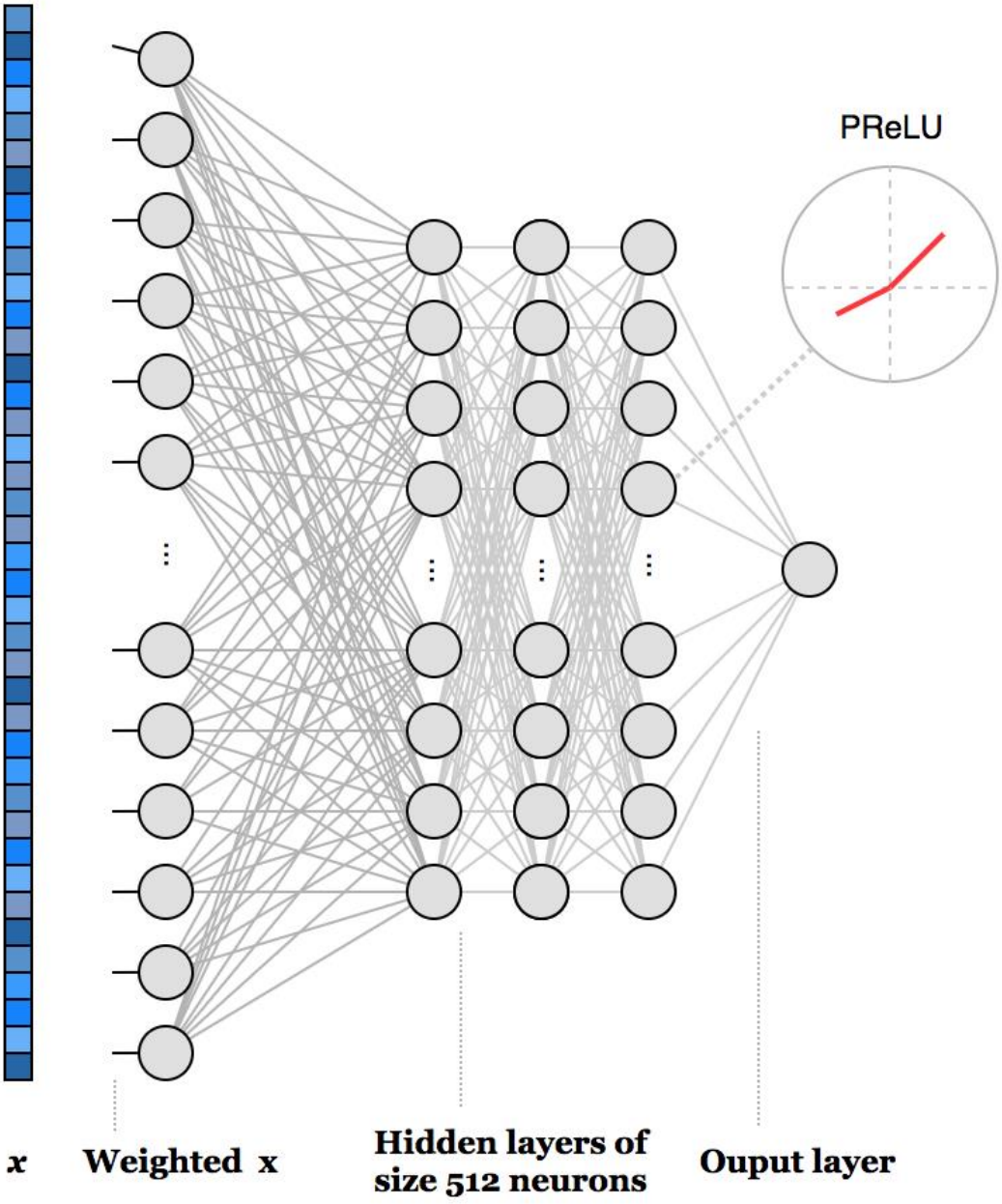


Figure S5. Related to "Transparent Methods: DNN training", Figures 1, 2. Neural network configuration for the best performing DNN sample-based regressor. The regressor takes in a full species level taxonomic profile and estimates the donor's exact chronological age. Host-based DNN regressor has a similar architecture with 1024 nodes in each layer.

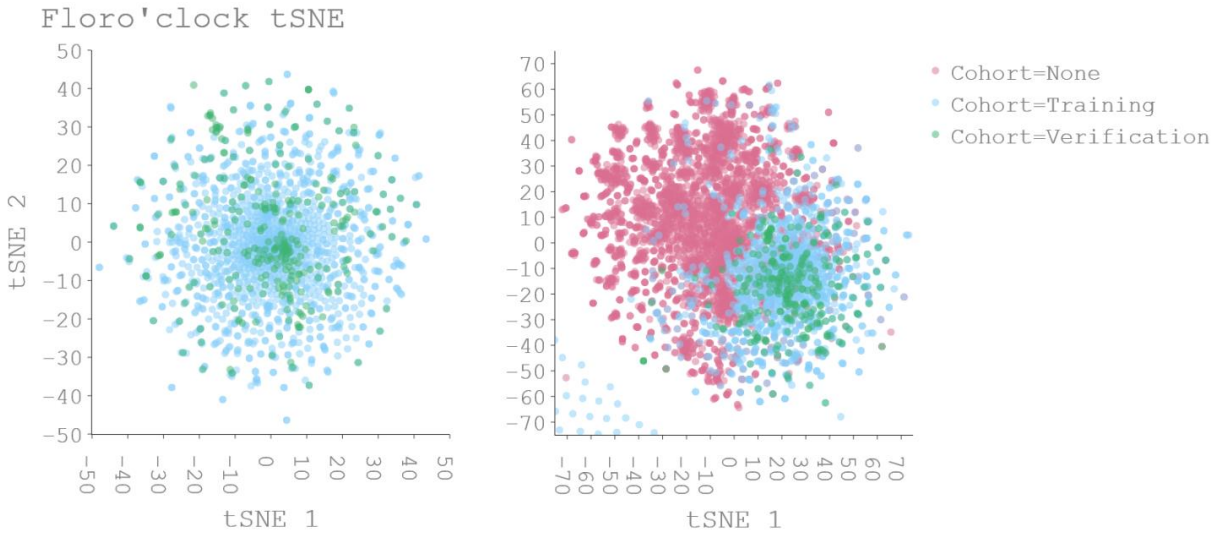


Figure S6. Related to "Transparent Methods: Data Description", Figures 1, 2. tSNE scatter plot results of samples from the CV cohort (left) and all samples considered for training (Right) colored by inclusion in this study. Red samples were removed from the study.

Figure S7. **Related to "Transparent Methods: Data Description", Figures 1, 2.** tSNE scatter plot results of samples considered for this study colored by the study of origin. ERP015317 was removed from this work, despite its great size and favorable age distribution.

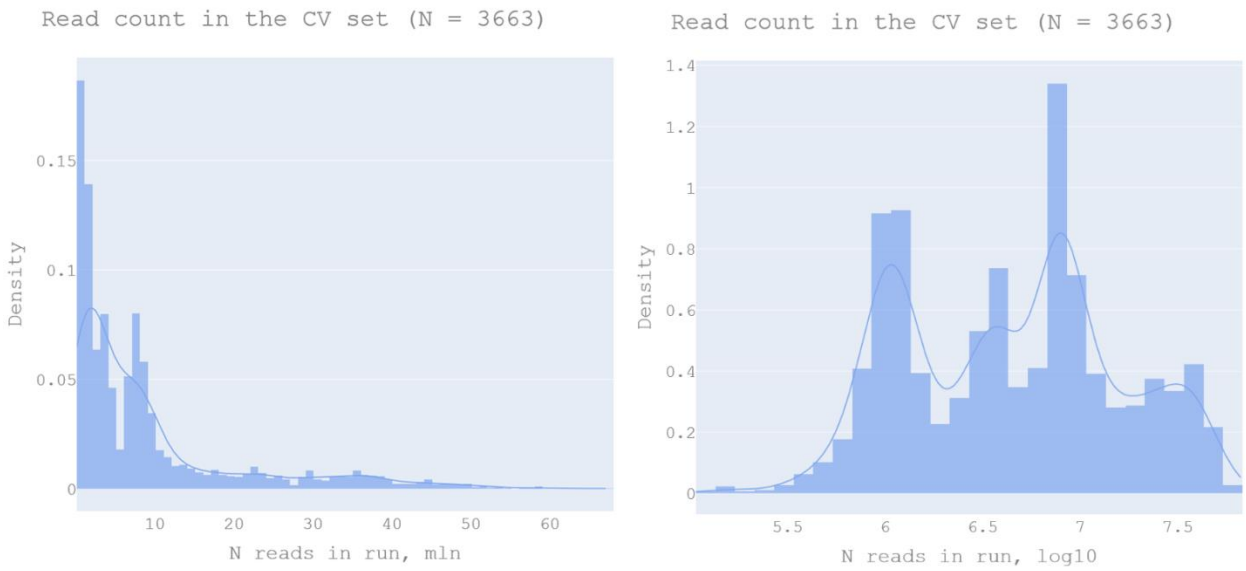


Figure S8. **Related to "Transparent Methods: Quality Control Details", Figures1, 2.**

Read number distribution in the 3,663 files used in this work in linear scale (left, 1 bin = 1), and log scale (right, 1 bin = 0.1).

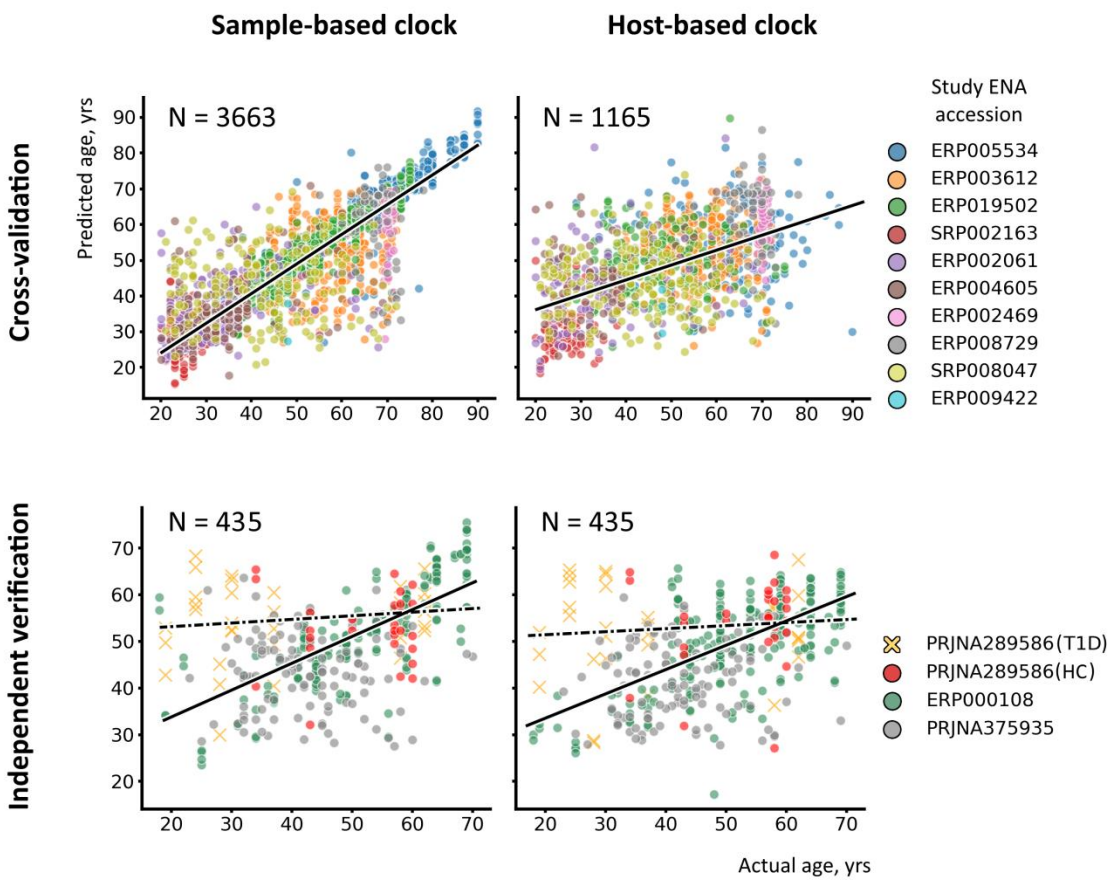


Figure S9. **Related to Figure 2**. Sample-based (left column) and host-based (right column) clock predictions in CV (top row) and independent validation (bottom row). Solid lines indicate linear fits for predictions derived from healthy individual (HC) samples: $R^2_{CV,Sample}$ = 0.97, $R^2_{CV,Host}$ = 0.64. Dash-dotted lines indicate linear fits for predictions derived from samples of people with type 1 diabetes (T1D). "N" stands for the number of metagenomic profiles used in either CV or verification stages.

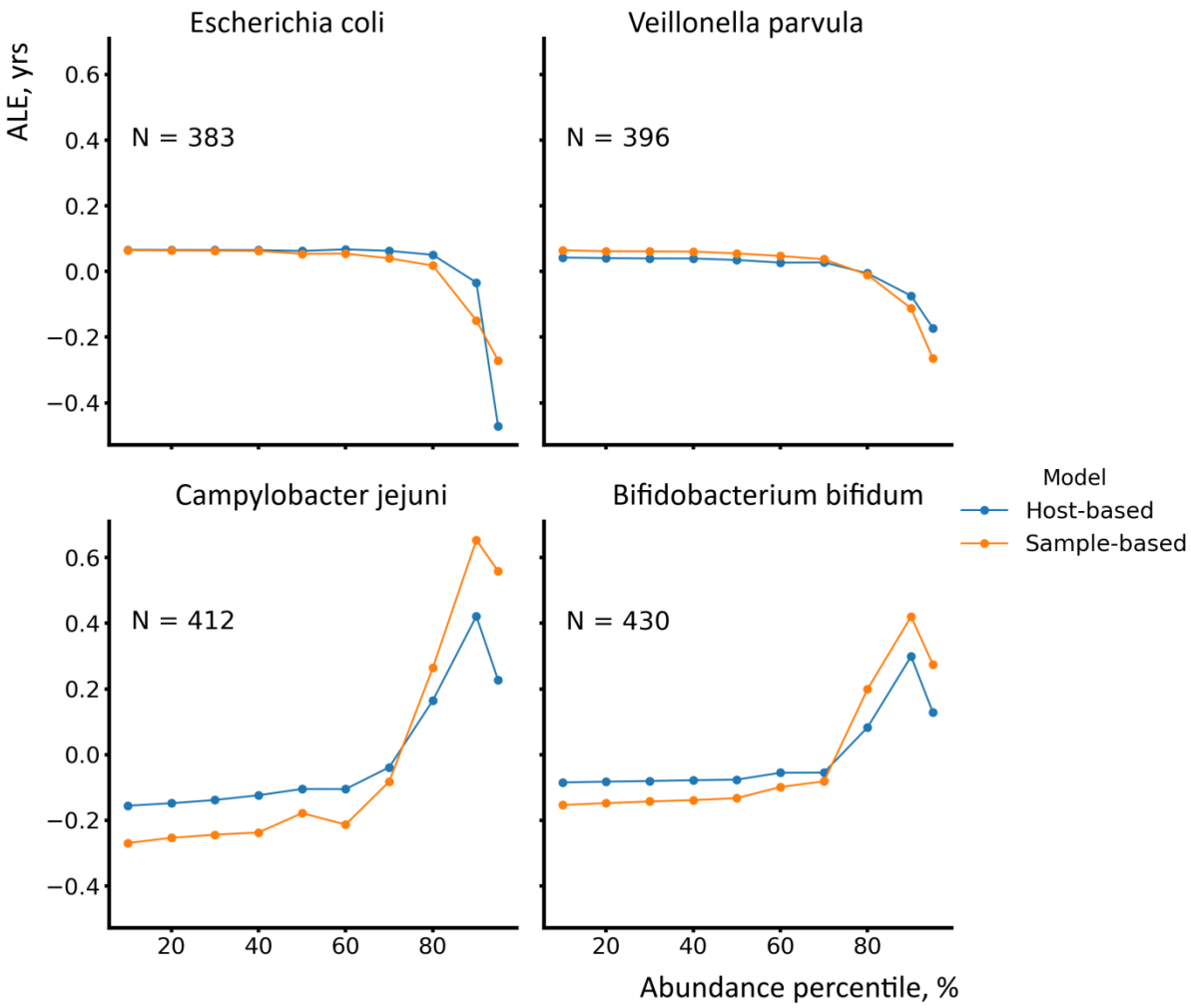


Figure S10. **Related to Figure 3.** Accumulated local effect (ALE) plots for 4 selected microbial taxa derived from perturbing taxonomic profiles in CV data set. ALE plots display how changes in a specific microbe's abundance (over a span of population-specific values) shift the model predictions towards acceleration or regression of age (in years). E.g. people with the abundance of Veillonella parvula higher than 95% of the population are predicted to be 0.25-0.3 years younger than those who have very little of this bacterium.

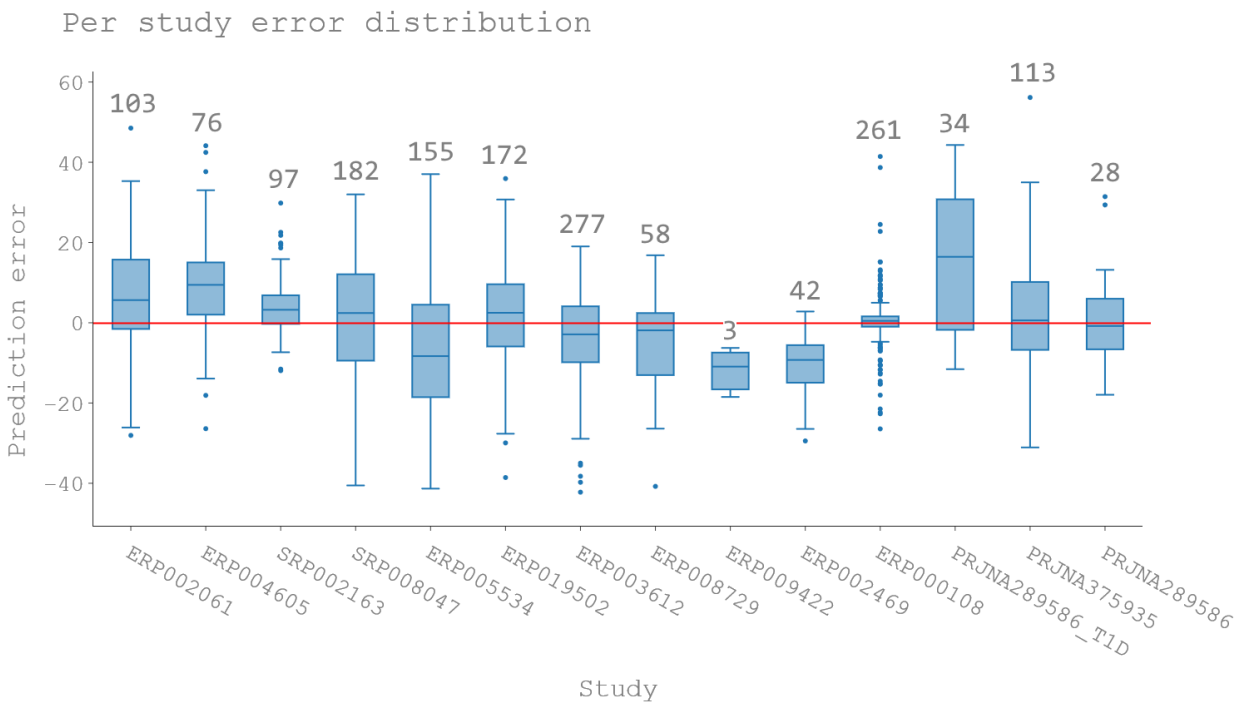


Figure S11. **Related to Figures 1, 2.** Box plot of errors produced from the reported model predictions. Last four boxes belong to the verification studies. The number above the boxes marks number of samples in the study.

Platform	N samples	N studies
Illumina HiSeq 2000	650	8
Illumina Genome Analyzer II	320	3
Illumina Genome Analyzer IIx	195	2

Table S6. Related to Figures 1, 2. — platforms used in CV cohort.

		CV Sample	CV Host	Independent Sample HC	Independent Host HC	Independent Sample T1D	Independent Host T1D
Neural Networks	R^2	0.81±0.02	0.21±0.13	0.25±0.10	0.29±0.09	-0.92±0.13	-0.98±0.17
	MAE, yrs	3.94±0.26	10.60±1.12	7.28±0.48	5.91±0.40	17.61±0.54	18.02±0.82
	R	0.90±0.01	0.52±0.11	0.49±0.04	0.53±0.05	0.08±0.06	0.14±0.08
	Baseline MAE, yrs	13.68	13.03	9.27	9.27	13.65	13.65
	Median age, yrs	46	50	49	49	33.5	33.5
	N, profiles	3663	1165	402	402	34	34
Machine Learning	Elastic Net MAE, yrs	13.94±2.65	13.75±2.92	10.25±0.56	9.625±0.34	16.12±0.60	16.95±0.63
	Random Forest MAE, yrs	11.18±4.21	11.32±4.18	7.25±0.54	7.06±0.49	18.44±1.00	18.51±1.04
	XGBoost MAE, yrs	11.03±3.95	11.06±3.91	7.78±0.50	7.11±0.54	18.11±1.37	18.40±1.47

Table S7. Related to Table 1 — Performance of host-based and sample-based models on independent testing sets for healthy individuals (HC), subjects with Type 1 Diabetes (T1D) and on cross-validation (CV) test. Green highlights the smallest MAE in each column, and red — MAEs that are greater than baseline MAE.

Supplementary references

Amir, A., McDonald, D., Navas-Molina, J.A., Debelius, J., Morton, J.T., Hyde, E., Robbins-Pianka, A., Knight, R., 2017. Correcting for Microbial Blooms in Fecal Samples during Room-Temperature Shipping. mSystems. https://doi.org/10.1128/msystems.00199-16

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J.E., Bittinger, K., Brejnrod, A., Brislawn, C.J., Brown, C.T., Callahan, B.J., Caraballo-Rodríguez, A.M., Chase, J., Cope, E.K., Da Silva, R., Diener, C., Dorrestein, P.C., Douglas, G.M., Durall, D.M., Duvallet, C., Edwardson, C.F., Ernst, M., Estaki, M., Fouquier, J., Gauglitz, J.M., Gibbons, S.M., Gibson, D.L., Gonzalez, A., Gorlick, K., Guo, J., Hillmann, B., Holmes, S., Holste, H., Huttenhower, C., Huttley, G.A., Janssen, S., Jarmusch, A.K., Jiang, L., Kaehler, B.D., Kang, K. Bin, Keefe, C.R., Keim, P., Kelley, S.T., Knights, D., Koester, I., Kosciolek, T., Kreps, J., Langille, M.G.I., Lee, J., Ley, R., Liu, Y.X., Loftfield, E., Lozupone, C., Maher, M., Marotz, C., Martin, B.D., McDonald, D., McIver, L.J., Melnik, A. V., Metcalf, J.L., Morgan, S.C., Morton, J.T., Naimey, A.T., Navas-Molina, J.A., Nothias, L.F., Orchanian, S.B., Pearson, T., Peoples, S.L., Petras, D., Preuss, M.L., Pruesse, E., Rasmussen, L.B., Rivers, A., Robeson, M.S., Rosenthal, P., Segata, N., Shaffer, M., Shiffer, A., Sinha, R., Song, S.J., Spear, J.R., Swafford, A.D., Thompson, L.R., Torres, P.J., Trinh, P., Tripathi, A., Turnbaugh, P.J., Ul-Hasan, S., van der Hooft, J.J.J., Vargas, F., Vázquez-Baeza, Y., Vogtmann, E., von Hippel, M., Walters, W., Wan, Y., Wang, M., Warren, J., Weber, K.C., Williamson, C.H.D., Willis, A.D., Xu, Z.Z., Zaneveld, J.R., Zhang, Y., Zhu, Q., Knight, R., Caporaso, J.G., 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat. Biotechnol. https://doi.org/10.1038/s41587-

019-0209-9

Bushnell, B., 2016. BBMap short-read aligner, and other bioinformatics tools. [WWW Document]. BBMap short-read aligner, other Bioinforma. tools. URL https://sourceforge.net/projects/bbmap/

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P., 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods. https://doi.org/10.1038/nmeth.3869

Kim, D., Song, L., Breitwieser, F.P., Salzberg, S.L., 2016. Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Res. 26, 1721–1729. https://doi.org/10.1101/gr.210641.116

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods. https://doi.org/10.1038/nmeth.1923

Plaza Onate, F., Batto, J.-M., Juste, C., Fadlallah, J., Fougeroux, C., Gouas, D., Pons, N., Kennedy, S., Levenez, F., Dore, J., Ehrlich, S.D., Gorochov, G., Larsen, M., 2015. Quality control of microbiota metagenomics by k-mer analysis. BMC Genomics 16, 183. https://doi.org/10.1186/s12864-015-1406-7

Rizk, G., Lavenier, D., Chikhi, R., 2013. DSK: k-mer counting with very low memory usage. Bioinformatics 29, 652–653. https://doi.org/10.1093/bioinformatics/btt020

Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. https://doi.org/10.1128/AEM.00062-07

Yoo, J., Kim, Y., Cho, E.R., Jee, S.H., 2017. Biological age as a useful index to predict

seventeen-year survival and mortality in Koreans. BMC Geriatr. 17, 7. https://doi.org/10.1186/s12877-016-0407-y