



Machine Learning Reveals Time-Varying Microbial Predictors with Complex Effects on Glucose Regulation

Oliver Aasmets,^{a,b} Kreete Lüll,^{a,b} Jennifer M. Lang,^c Calvin Pan,^c Johanna Kuusisto,^d Krista Fischer,^e Markku Laakso,^d Aldons J. Lusis,^{c,f,g}  Elin Org^a

^aInstitute of Genomics, Estonian Genome Centre, University of Tartu, Tartu, Estonia

^bDepartment of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia

^cDepartment of Medicine, University of California, Los Angeles, California, USA

^dInstitute of Clinical Medicine, Internal Medicine, University of Eastern Finland, Kuopio, and Kuopio University Hospital, Kuopio, Finland

^eInstitute of Mathematics and Statistics, University of Tartu, Tartu, Estonia

^fDepartment of Human Genetics, University of California, Los Angeles, California, USA

^gDepartment of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, California, USA

ABSTRACT The incidence of type 2 diabetes (T2D) has been increasing globally, and a growing body of evidence links type 2 diabetes with altered microbiota composition. Type 2 diabetes is preceded by a long prediabetic state characterized by changes in various metabolic parameters. We tested whether the gut microbiome could have predictive potential for T2D development during the healthy and prediabetic disease stages. We used prospective data of 608 well-phenotyped Finnish men collected from the population-based Metabolic Syndrome in Men (METSIM) study to build machine learning models for predicting continuous glucose and insulin measures in a shorter (1.5 year) and longer (4 year) period. Our results show that the inclusion of the gut microbiome improves prediction accuracy for modeling T2D-associated parameters such as glycosylated hemoglobin and insulin measures. We identified novel microbial biomarkers and described their effects on the predictions using interpretable machine learning techniques, which revealed complex linear and nonlinear associations. Additionally, the modeling strategy carried out allowed us to compare the stability of model performance and biomarker selection, also revealing differences in short-term and long-term predictions. The identified microbiome biomarkers provide a predictive measure for various metabolic traits related to T2D, thus providing an additional parameter for personal risk assessment. Our work also highlights the need for robust modeling strategies and the value of interpretable machine learning.

IMPORTANCE Recent studies have shown a clear link between gut microbiota and type 2 diabetes. However, current results are based on cross-sectional studies that aim to determine the microbial dysbiosis when the disease is already prevalent. In order to consider the microbiome as a factor in disease risk assessment, prospective studies are needed. Our study is the first study that assesses the gut microbiome as a predictive measure for several type 2 diabetes-associated parameters in a longitudinal study setting. Our results revealed a number of novel microbial biomarkers that can improve the prediction accuracy for continuous insulin measures and glycosylated hemoglobin levels. These results make the prospect of using the microbiome in personalized medicine promising.

KEYWORDS T2D, gut microbiome, machine learning, prediction analysis, gut microbiome, type 2 diabetes

The prevalence of type 2 diabetes (T2D) has more than doubled since 1980, resulting in a huge burden on the health care system worldwide (1). In order to fight the

Citation Aasmets O, Lüll K, Lang JM, Pan C, Kuusisto J, Fischer K, Laakso M, Lusis AJ, Org E. 2021. Machine learning reveals time-varying microbial predictors with complex effects on glucose regulation. *mSystems* 6:e01191-20. <https://doi.org/10.1128/mSystems.01191-20>.

Editor Mani Arumugam, University of Copenhagen

Copyright © 2021 Aasmets et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Elin Org, elin.org@ut.ee.

Received 13 November 2020

Accepted 22 January 2021

Published 16 February 2021

epidemic of T2D and improve public health, an understanding of the first stages of this disease is necessary for preventive actions. Recently, the bacterial communities residing in our intestines have become a topic of interest as a potential way to prevent the development of glucose dysregulation. The microbiome has been shown to modulate a variety of physiological functions, such as gut permeability, inflammation, glucose metabolism, and fatty acid oxidation, supporting an important role of the microbiome in the pathophysiology of T2D (2).

Numerous studies have already reported changes in the gut microbiome in subjects with T2D or prediabetes compared to healthy individuals (3–5). Although there is information that the abundance of bacteria such as *Roseburia* and *Bifidobacteria* is altered in subjects with T2D (2), compelling evidence that supports the use of gut microbiome as a predictive tool for T2D is lacking, as a majority of the findings are based on cross-sectional studies. However, in order to assess the microbiome as a prognostic tool for T2D, prospective studies are needed.

T2D is a heterogeneous disease with multiple pathophysiological pathways involved (6). Thus, in order to fully understand the role of the microbiome in the risk of T2D, a case-control design might not be sufficient. As the progression of the disease is a continuous process, detailed data about metabolic outcomes such as continuous glucose and insulin measurements could help to unravel the disease mechanisms involving the microbiome.

Together with heterogeneity in the first stages of T2D, the gut microbiome itself is known to be highly personalized (7, 8). Variability in continuous metabolic outcomes and gut microbiome lead to difficulties in reproducing the results obtained and raises the need for robust modeling strategies. Machine learning methods have been shown to capture various complex association patterns from different data types. Although machine learning has become popular in microbiome studies as well, the ability of the algorithms to provide robust results remains unclear (9, 10).

We now report the application of a random forest algorithm on microbiome data to predict multiple continuous metabolic outcomes that influence the development of T2D in a longitudinal study setting. We identify microbial biomarkers for the metabolic outcomes and describe their effects on the predictions using interpretable machine learning techniques. In addition, we show that there are significant differences in the identified biomarkers between long and short follow-up periods. We also show how the modeling procedure significantly influences the results.

RESULTS

Study design. We used prospective data of well-phenotyped Finnish men collected from a population-based Metabolic Syndrome in Men (METSIM) study. A comprehensive machine learning strategy was implemented to identify microbial biomarkers and their effect on numerous metabolic traits. A graphical overview of the study design and modeling procedure is shown in Fig. 1. Random forest models were trained to predict the metabolic outcomes of interest in the follow-up using the baseline microbiome (MB), metabolic outcomes (MO), and additional covariates (CoV) such as body mass index and age as predictors. To evaluate the effect of the microbiome, models including microbial predictors were compared to models excluding microbial predictors. In order to assess the temporal changes in biomarker selection and predictive performance, independent prospective models were trained for the 18-month and 48-month follow-up period. To evaluate the model generalizability and stability, model training was repeated 200 times with a different train-test split made each run. Permutation feature importance metrics were used to identify microbial biomarkers. Finally, accumulated local effects methodology was used to plot the effect of the microbial biomarkers for predicting the corresponding metabolic trait.

Model stability and generalizability. In the first step, we tested whether we could improve the prediction of metabolic outcomes using microbiome data as an additional predictor. The human gut microbiome is known to be highly variable and personalized (7, 8). Thus, estimating the robustness of the predictive models is essential. The

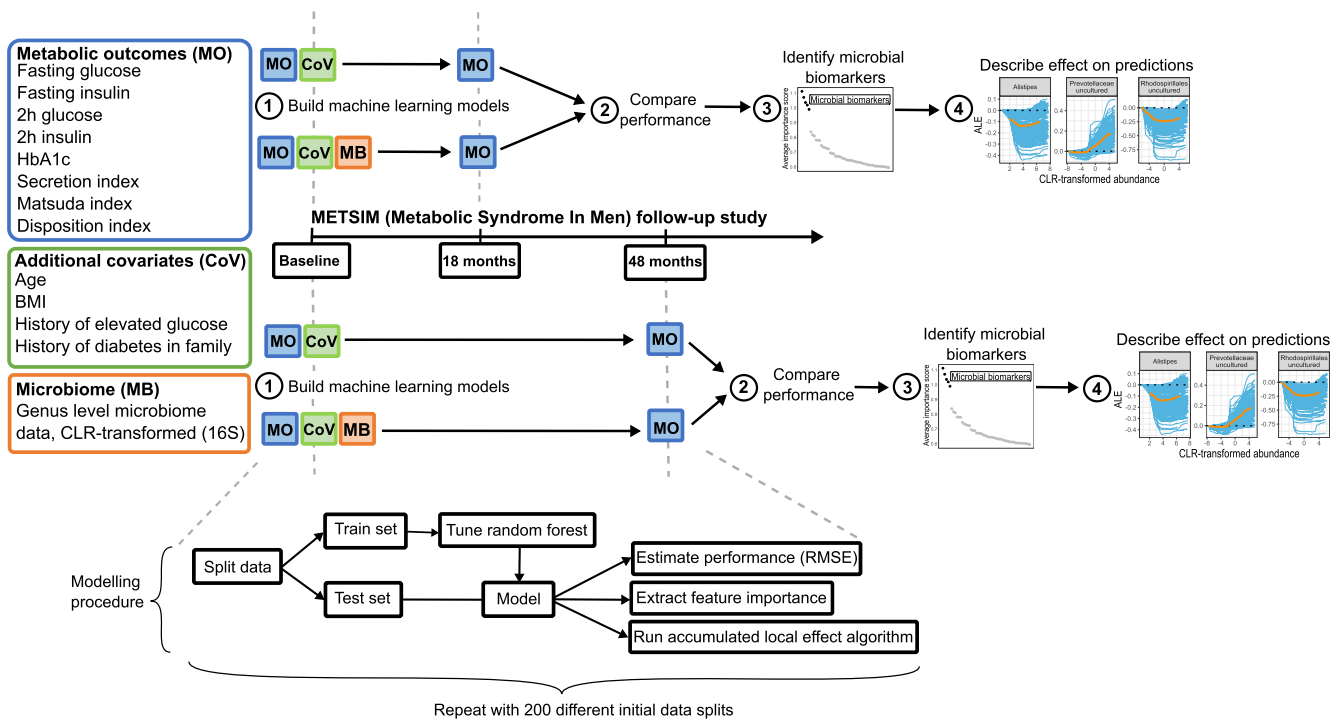


FIG 1 Study design and modeling procedure.

problem with microbiome data based on our experience is that the performance of the model might be highly dependent on the initial data split for training and test sets. The models were run 200 times with different initial splits to assess the impact of the data split. Table 1 summarizes the obtained results.

These results highlight the variability in performance estimates occurring due to the data split. Out of 200 data splits, the number of models that took advantage of using microbial predictors varies around 100, which implies that the data split plays an important role in the outcome. Our results suggest that for the 18-month time frame, microbiome as a predictor can improve the prediction accuracy for secretion index, glycosylated hemoglobin (HbA1c), and 2-h insulin levels. For secretion index, models including microbial predictors outperformed simpler models in 61% of the cases, for 2-h insulin in 70.5% of the cases, and for HbA1c in 64.5% of the cases. The

TABLE 1 Model stability and generalizability

Trait	18-mo time frame			48-mo time frame		
	Mean (SD) difference in RMSE ^a	No. of models including microbiome performing better (%)	P (model including microbial predictors performs better) Bonferroni adjusted	Mean (SD) difference in RMSE ^a	No. of models including microbiome performing better (%)	P (model including microbial predictors performs better) BONFERRONI adjusted
Fasting glucose	0.001 (0.0594)	99 (49.5)	1.0000	-0.006 (0.0641)	112 (56)	0.8291
2-h glucose	-0.02 (0.217)	118 (59)	0.1050	0.07 (0.332)	73 (36.5)	1.0000
Fasting insulin	0.20 (1.04)	73 (36.5)	1.0000	-0.29 (1.080)	137 (68.5)	1.432 · 10 ⁻⁶⁶
2-h insulin	-3.23 (10.840)	141 (70.5)	5.046 · 10 ^{-8b}	-1.42 (12.304)	122 (61)	0.0182 ^b
HbA1c	-0.005 (0.0305)	129 (64.5)	0.0004 ^b	-0.002 (0.0360)	111 (55.5)	1.000
Secretion index	-0.36 (4.949)	122 (61)	0.0182 ^b	-0.77 (3.254)	138 (69)	6.422 · 10 ^{-7b}
Matsuda index	0.07 (0.573)	90 (45)	1.0000	-0.01 (0.569)	103 (51.5)	1.0000
Disposition index	4.42 (26.590)	77 (38.5)	1.0000	2.01 (16.251)	86 (43)	1.0000

^aMean differences in root-mean-square error (RMSE) between models including microbial predictors and models excluding microbial predictors. Negative value indicates a model including microbial predictors outperforming the model excluding microbial predictors.

^bStatistically significant results according to the binomial test after Bonferroni correction.

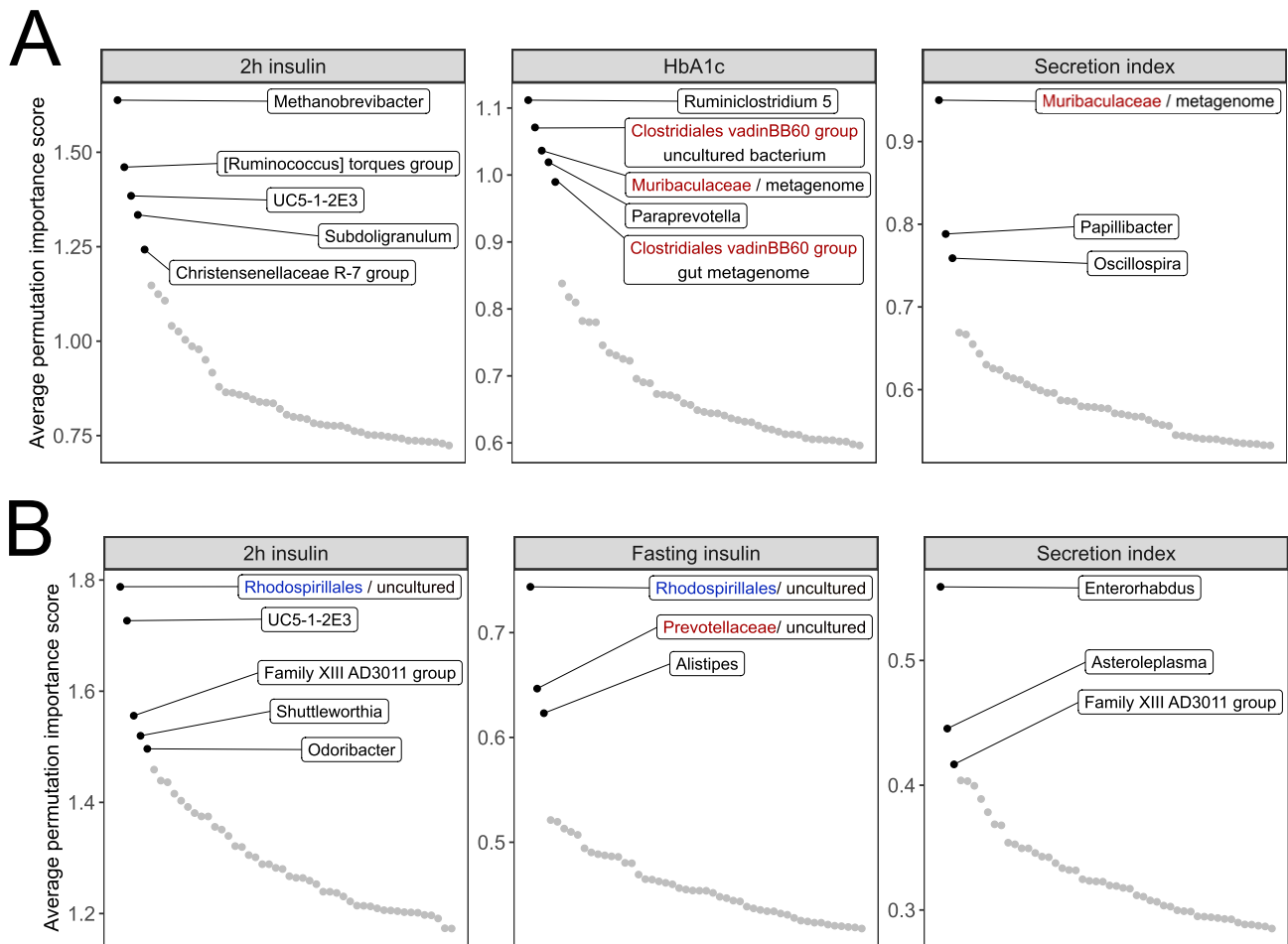


FIG 2 Average feature importance scores for the top 50 microbial markers. The highlighted taxa are considered the most significant biomarkers. Blue represents order-level taxonomy, and red represents family-level taxonomy. (A) Predictors for 18-month follow-up. (B) Predictors for 48-month follow-up.

Bonferroni-corrected P values for testing whether the model including microbial predictors outperforms the model excluding microbial predictors based on a random data split are 0.0182 for secretion index, $5.046 \cdot 10^{-8}$ for 2-h insulin, and 0.0004 for HbA1c. Average improvements in root-mean-square error (RMSE) were -0.36 for secretion index, -3.23 mU/liter for 2-h insulin, and -0.005% for HbA1c. For a 48-month time frame, the microbiome improves the prediction model for the secretion index, fasting insulin, and 2-h insulin. For secretion index, models including microbial predictors outperformed simpler models in 69% of the cases, for 2-h insulin in 61% of the cases, and for fasting insulin in 68.5% of the cases. The adjusted P values for testing whether the model including microbial predictors outperforms the model excluding microbial predictors based on a random data split are $6.422 \cdot 10^{-7}$ for secretion index, 0.0182 for 2-h insulin, and $1.432 \cdot 10^{-6}$ for fasting insulin. Average improvements in RMSE were -0.77 for secretion index, -1.42 mU/liter for 2-h insulin, and -0.29 mU/liter for fasting insulin.

Remarkably, the variation in differences in RMSE between the model including microbial predictors and the model excluding microbial predictors over the 200 runs is large. Due to the high variability, the potential of improvement in prediction accuracy when microbiome data are used remains largely unclear.

Novel predictive microbial biomarkers for metabolic outcomes. In order to find microbial markers that are predictive for the metabolic outcomes, average feature importance scores over 200 runs were compared. Figure 2 shows the average importance

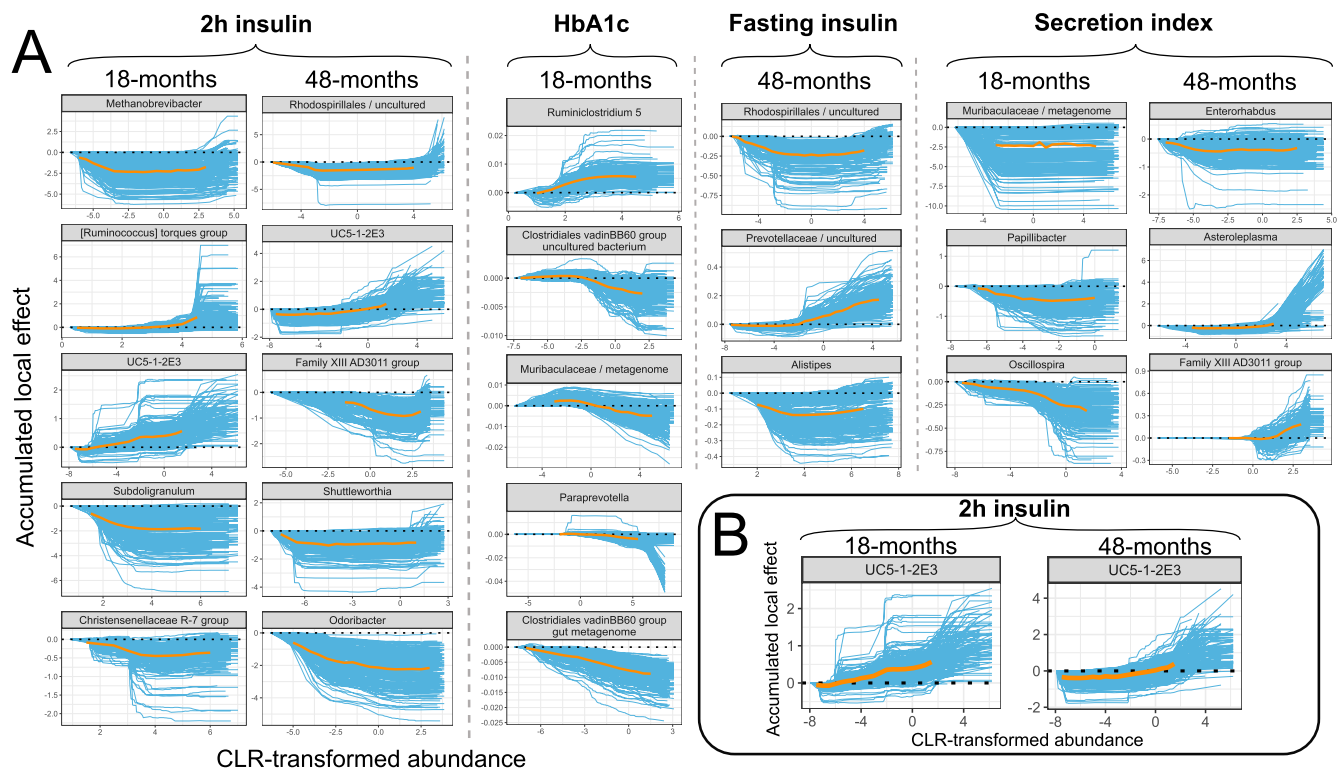


FIG 3 Accumulated local effect (ALE) plots. (A) ALE plots for the found microbial biomarkers. (B) ALE plots for genus *UC5-1-2E3* found to predict 2-h insulin in 18-month and 48-month follow-ups. Blue lines represent effects for each run out of 200; orange lines represent aggregated effects. Aggregated effect is displayed between the 2.5% and 97.5% quantiles of CLR-transformed abundance for the corresponding microbial marker.

score of the top 50 microbial predictors for metabolic outcomes that took advantage of using microbial predictors. It can be seen that certain microbial predictors significantly stand out for each metabolic outcome and time frame combination.

For an 18-month time frame (Fig. 2A, Table S2), the most important microbial predictors for 2-h insulin include genus *Methanobrevibacter* and numerous genera from the phylum *Firmicutes*, such as *[Ruminococcus] torques group*, *UC5-1-2E3*, *Subdoligranulum*, and *Christensenellaceae R-7 group*. Predictors for HbA1c are the genus *Ruminiclostridium 5*, the genus *Paraprevotella*, an unclassified member of the family *Muribaculaceae*, and members of *Clostridiales vadinBB60 group*. An unclassified member of the family *Muribaculaceae* together with *Papillibacter* and *Oscillospira* are significant predictors for secretion index.

For the 48-month time frame (Fig. 2B, Table S3), top predictors for 2-h insulin include uncultured *Rhodospirillales* and *UC5-1-2E3*. Distinguishable genera according to the average importance score are also *Family XIII AD3011 group*, *Shuttleworthia* and *Odoribacter*. Significant predictors for fasting insulin are uncultured *Rhodospirillales*, uncultured *Prevotellaceae*, and the genus *Alistipes*. For secretion index, the genus *Enterorhabdus* together with *Asteroleplasma* prove to be the most important predictors, with *Family XIII AD3011 group* slightly standing out.

There is overlap in the most important microbial markers found for predicting different metabolic outcomes. In the 18-month follow-up period, unclassified *Muribaculaceae* is a significant predictor for secretion index and HbA1c. For the 48-month follow-up period, *Family XIII AD3011 group* is a predictor for secretion index and 2-h insulin, and uncultured *Rhodospirillales* is an important predictor for fasting insulin and 2-h insulin. Additional overlap can be seen among the top 10 microbial predictors according to the average permutation importance score (Tables S2 and S3).

Interpreting the effect of microbial biomarkers on the predictions. Together with finding the relevant biomarkers, understanding how they influence the

predictions is necessary. This task is complicated for most of the machine learning algorithms, which is why they are considered “gray-box” or “black-box” methods. Recently, much attention has been put into explaining the predictions of such models. Here, we implemented accumulated local effect (ALE) plots that aim to describe the effect of a certain predictor on the metabolic outcome independently of the remaining predictors (11). Accumulated local effect plots for the previously highlighted most significant microbial biomarkers are shown in Fig. 3. Accumulated local effect plots for the top 10 microbial predictors are shown in Fig. S1 and S2. In most cases, ALE plots show nonlinear associations between a microbial predictor and metabolic outcome of interest. Although large variability in the effect estimates between the different data splits can be seen, the shape of the effect stays relatively stable for all microbial predictors.

Considering the 18-month time frame (Fig. 3A, Fig. S1), higher centered log-ratio (CLR)-transformed abundances of genera from the *Lachnospiraceae* family—[*Ruminococcus*] *torques* group and *UC5-1-2E3*—lead to higher predictions for 2-h insulin. High CLR-transformed abundances of the genera *Subdoligranulum*, *Methanobrevibacter*, and *Christensenellaceae R-7* group lower the predictions for 2-h insulin. For HbA1c, higher CLR-transformed abundance of *Ruminiclostridium 5* leads to higher predictions. In contrast, high CLR-transformed abundances of bacteria from the family *Muribaculaceae*, members of *Clostridiales vadinBB60* group, and *Paraprevotella* reduce the levels of HbA1c. For secretion index, the prediction might depend on the presence-absence of the unclassified genus from the family *Muribaculaceae*, because the ALE plot stays relatively stable after an initial decrease from the minimum values of CLR-transformed abundances. High CLR-transformed abundances of *Oscillospira* and *Papillibacter* decrease the predictions for secretion index.

Considering the 48-month follow-up period (Fig. 3A, Fig. S2), high CLR-transformed abundances of the genera *Firmicutes Family XIII AD3011* group, *Odoribacter*, and unclassified *Rhodospirillales* lead to lower predictions for 2-h insulin. In contrast, extremely high CLR-transformed values of the genus *UC5-1-2E3* lead to higher predictions. *Shuttleworthia* seems to show a presence-absence effect, as the drop from the lowest CLR-transformed values lowers the predictions for 2-h insulin. For fasting insulin, higher CLR-transformed abundances of unclassified *Rhodospirillales* and *Alistipes* lower the predictions. In contrast, high CLR-transformed abundances of an unclassified genus from the *Prevotellaceae* family leads to higher predictions for fasting insulin. Interestingly, extremely low values of *Alistipes* lead to higher predictions for fasting insulin than when *Alistipes* levels are within 2.5% and 97.5% quantiles. A similar effect for the genus *Asteroleplasma* on secretion index can be seen as extremely high CLR-transformed abundance of *Asteroleplasma* leads to drastically higher predictions. The genus *Enterorhabdus* might show presence-absence effects, with the presence of *Enterorhabdus* leading to decreased predictions. Lastly, high CLR-transformed abundance of the genus *Family XIII AD3011* group leads to higher predictions.

Comparison of microbial predictors in different time points. Independently modeling the two scenarios with various follow-up times allowed us to compare the most relevant predictors to see if the effect and choice of microbial biomarkers remains the same. Considering metabolic outcomes that the microbiome data helped to predict, only one microbial predictor for the same metabolic outcome was shared (Fig. 3B). The genus *UC5-1-2E3* from the *Lachnospiraceae* family was found to be among the top predictors for 2-h insulin in the 18-month and 48-month time frame. Among the top 10 predictors for each target variable, *Escherichia-Shigella* was also shared for 2-h insulin (Fig. S1 and S2).

The shape of the effect for *UC5-1-2E3* stays relatively stable, with extreme values for the genus showing higher predictions for both follow-up periods. This suggests that the genus *UC5-1-2E3* could be considered a robust biomarker for predicting 2-h insulin. Nevertheless, all other genera from the top microbial predictors were specific for a certain time frame.

DISCUSSION

We used machine learning to predict multiple metabolic outcomes (continuous glucose and insulin measures, HbA1c) over time periods of various lengths using the gut microbiome as a predictive measure. Furthermore, the modeling strategy carried out allowed us to understand the variability in performance estimates and biomarker selection. We described how high variability and personalization of the human gut microbiome leads to large variations in the performance estimates. We showed that microbial predictors can improve the prediction accuracy for continuous insulin measures and glycosylated hemoglobin in addition to conventional risk factors, additionally highlighting differences in short- and long-term cases. Finally, we identified microbial biomarkers that contribute to the improved performance and described their effect on the outcome.

Most of the current studies describing the role of bacteria in diabetes have been case-control studies, with diabetes being a binary trait defined by setting a cutoff to some continuous glucose measure (3, 4, 12). Type 2 diabetes, however, is a disease preceded by a long-lasting prediabetic state, and the development of the disease is a continuous process (13). Detailed phenotyping is definitely a strength of this study, as it allows us to study the first stages of disease progression. Our results suggest that bacteria provide a means for predicting changes in insulin secretion and insulin response to glucose intake. A causal effect of microbiome-produced short-chain fatty acids (SCFA) has been confirmed with respect to various insulin measures, primarily insulin secretion (14). Figure S3 shows that 2-h insulin levels first increase in subjects with prediabetes, defined by the WHO classification, as a compensatory mechanism to keep glucose levels in the normal range. Thus, 2-h insulin values are among the first indicators for the development of diabetes. Therefore, our results provide valuable insight into the potential application of the microbiome as a predictive measure for T2D and highlight the need for detailed phenotyping in order to fully understand the role of the microbiome in this disease.

Recently, Gou et al. (12) used a similar interpretable machine learning strategy and found bacteria that effectively differentiated type 2 diabetes cases from healthy controls in the Chinese population. Additionally, they built a microbiome risk score (MRS) and showed the causal role of identified bacteria on diabetes development after fecal microbiota transplantation to mice. The microbial predictors found do not show significant overlap with our findings. Only *Alphaproteobacteria* found by Gou et al. can be considered overlapping. We found one taxon from the class *Alphaproteobacteria*—an uncultured genus from the order *Rhodospirillales*—to predict fasting insulin and 2-h insulin in a 48-month time frame. We found a higher CLR-transformed abundance of an unclassified *Rhodospirillales* genus decreasing type 2 diabetes risk, which is consistent with the findings of Gou et al. Multiple reasons might explain the observed inconsistencies. Importantly, our study was specifically designed to find prospective predictors for continuous measures. Another possible difference is the cohort structure. Our study included men exclusively, compared to 33.1% in Gou et al. The effect of sex on the gut microbiome is not clear but cannot be ruled out (15, 16). Also, the metagenomic analyses of European women and Chinese subjects have shown differences, which is why geographic differences in microbiome are also a possibility (3, 4).

Rhodospirillales, one of the strongest predictors in the current study, was found to be predictive for fasting and 2-h insulin in a 48-month follow-up. The order *Rhodospirillales* consists of bacteria that are known to produce acetic acid (17), which has been shown to improve insulin sensitivity (18, 19). Several other detected microbial predictors have been previously described elsewhere as being associated with T2D or glucose regulation. Zhou et al. (20) showed that the genus *Odoribacter* was negatively associated with steady-state plasma glucose, which is consistent with our results for predicting 2-h insulin. Krych et al. carried out a study on mice and identified *Muribaculaceae* (previously classified as *S24-7*) to be

protective against T2D (21), which corresponds to the protective effect for HbA1c seen in our study.

Previously inconsistent associations have also been reported. We found a higher CLR-transformed abundance of *Alistipes* to predict lower values for fasting insulin, which is not consistent with the results obtained by Wu et al. (22), who showed positive associations with type 2 diabetes. *Subdoligranulum* has been found to be enriched in type 2 diabetes cases (23), which is inconsistent with our results, as higher CLR-transformed abundance predicts lower values for 2-h insulin. Similar to the work by Gou et al. (12), the main reasons behind these inconsistencies are likely study design and population structure. We note that microbiome composition has been shown to be associated with numerous factors not considered confounders in our study due to data availability (24, 25). Thus, inconsistencies with previously reported results and added predictive value for metabolic traits could be explained by uncontrolled covariates.

We are not aware of any population with a similar follow-up period and where microbiome data are available and an oral glucose tolerance test has been carried out at the baseline and at the follow-up. Therefore, we could not replicate our findings in other populations using similar study design.

How machine learning techniques can best utilize microbiome data is still an open question (26). Therefore, the true potential of the gut microbiome for predicting T2D remains unknown. Additionally, taking the compositional nature of microbiome data into account is crucial for all types of analysis and machine learning applications (26). Previous studies have shown the advantage of using log-ratio transformations for overcoming the limitations of working with compositional data. For example, Quinn and Erb (27) and Tolosana-Delgado et al. (28) showed how centered log-ratio (CLR)-transformed data can outperform raw proportions. Moreover, Tolosana-Delgado et al. (28) showed how pairwise log-ratio transformation can greatly outperform CLR transformation when a random forest algorithm is used. Thus, novel methods and strategies for handling compositionality might substantially improve the prediction accuracy for continuous metabolic outcomes. Also, shotgun metagenomics can provide more accurate taxonomic resolution than 16S sequencing (29).

It needs to be highlighted that even in the best scenario, when predicting 2-h insulin values in the 18-month scenario, incorporating microbial predictors did not improve prediction accuracy in 29.5% of the data splits. The high variability in performance estimates shows the necessity for robust modeling strategies to achieve reliable and generalizable performance. Our data clearly indicate that conventionally used 10-fold cross-validation might not be sufficient to obtain generalizable models when sample sizes stay relatively small compared to the number of microbial features. Taken together, the variability of model performance estimates in microbiome studies can be large and needs to be given attention in order to gain a proper understanding of the predictive ability of the microbiome. To reduce the variability in performance estimates and increase prediction accuracy, future studies can take advantage of combining shotgun sequencing with the best-performing modeling strategies, including additional microbiome-associated covariates, and integrating clinical and microbiome data from multiple time points as predictors.

Conclusions. In summary, our findings provide a clear indication that the microbiome, together with conventional risk factors, can be used to predict multiple metabolic outcomes. The detailed clinical characterization and longitudinal study design of the METSIM cohort make it particularly useful for understanding host-microbiome relationships. We have identified a number of novel microbial biomarkers which could predict metabolic traits associated with the prediabetic state. Our data provide a significant resource for further studies to determine the causal relationship of the identified biomarkers to the progression of T2D. Therefore, the prospect of using the microbiome in personalized medicine is promising.

MATERIALS AND METHODS

Study population and characterization. METSIM (Metabolic Syndrome in Men) is a randomly selected cohort of men from eastern Finland, aged 45 to 73 years, who have been carefully phenotyped

for different metabolic traits such as T2D, hypertension, and obesity. We investigated a subset of the METSIM cohort that took part in the METSIM follow-up study and from whom stool samples were collected ($n = 608$). The data resource consists of samples taken from three time points—at baseline (baseline of METSIM 5-year follow-up study), at 18-month follow-up, and at 48-month follow-up. At each time point, the subjects went through a 1-day outpatient visit, during which they provided blood samples after an overnight fast, and various parameters such as height, weight, and blood pressure were measured and an oral glucose tolerance test (OGTT) was performed. Additionally, at the baseline visit, the subjects were interviewed about their history of diseases and drug usage. The full study protocol and data resources are described in Laakso et al. (30). All subjects have given written informed consent, and the study was approved by the Ethics Committee of the University of Kuopio and was in accordance with the Helsinki Declaration.

In contrast to case-control studies, continuous “metabolic outcomes” (MO) were used as target variables in the modeling framework. The advantage of using continuous metabolic outcomes is that the phenotype is more distinct and there are no borderline cases with similar abilities of handling glucose as there likely are in the case-control setting (6). In total, two glucose measures, two insulin measures, glycosylated hemoglobin, and three calculated glucose regulation indices were considered (Fig. 1). Glycosylated hemoglobin (HbA1c), fasting insulin, 2-h insulin, fasting glucose, and 2-h glucose were measured according to the study protocol (30). The Matsuda insulin sensitivity index was calculated according to reference 31. The insulin secretion index was calculated as $\text{Secretion index} = \text{AUC}_{\text{Insulin}(0-30\text{min})} / \text{AUC}_{\text{Glucose}(0-30\text{min})}$, where the area under curve (AUC) was calculated using the trapezoidal formula. The disposition index was calculated as $\text{Disposition index} = \text{Secretion index} \cdot \text{Matsuda}$. The Matsuda insulin sensitivity index and the insulin secretion index have been previously shown to be best estimates for insulin sensitivity and insulin secretion in the METSIM cohort (32). Summary statistics for metabolic outcomes and additional covariates considered predictors in the machine learning models are shown in Table S1.

Microbiome data collection, sequencing, and data processing. Stool samples were collected at the baseline visit during the evaluation at the University of Kuopio Hospital and immediately stored at -80°C . Microbial DNA was extracted using the PowerSoil DNA isolation kit (MoBio Laboratories, Carlsbad, CA, USA) following the manufacturer’s instructions. The fecal microbiota composition was profiled by amplifying the V4 region of the 16S rRNA gene with 515F and 806R primers as previously described (33). PCR products were quantified with a Quant-iT™ PicoGreen double-stranded DNA (dsDNA) assay kit (Thermo Fisher). Samples were combined in equal amounts (~ 250 ng per sample) into pools and purified with the UltraClean PCR clean-up kit (MoBio). Sequencing was performed on an Illumina HiSeq 3000 instrument.

Raw demultiplexed data were imported into the open-source software QIIME2 version 2019.7 using the `q2-tools-import` script with the `CasavaOneEightSingleLanePerSampleDirFmt` input format (34). DADA2 software was used for denoising (35). DADA2 uses a quality-aware model of Illumina amplicon errors to attain an abundance distribution of sequence variance, which has a difference of a single nucleotide. The `q2-dada2-denoise-single` script was used to truncate the reads at position 123; trimming was not applied. Chimera removal was done with the “consensus” filter, in which chimeras are detected in each sample individually, and sequences established as chimeric in a certain fraction of samples are removed. After the denoising step, amplicon sequence variants (ASVs), equivalent to operational taxonomic units (OTUs), were aligned using MAFFT (36), and the phylogeny was constructed with FastTree (37). Taxonomy assignment was done using the `q2-feature-classifier` with the pretrained naive Bayes classifier based on reference reads from the SILVA 16S V3-V4 v132_99 database with a similarity threshold of 99%. Seven samples did not pass quality control during the sequencing process and were removed from further analysis.

The average number of reads per sample was 1,351,289, and samples with less than 100,000 reads were excluded from further analysis. The rest of the samples were aggregated to the genus level, which resulted in 553 genera. An additional filtering procedure was carried out to include only the most common genera for the prediction task. Genera that appeared in at least 50% of the samples were included in the final modeling task, 172 in total.

Due to the nature of sequencing, read counts are uninformative and must be considered relative to the total sum of reads for a given sample (38). In order to compensate for the compositional nature of the data, centered log-ratio (CLR) transformation was used as first proposed by Aitchison (39):

$$\text{CLR}(\vec{x}) = \ln \left[\frac{x_1}{g(\vec{x})}, \frac{x_2}{g(\vec{x})}, \dots, \frac{x_D}{g(\vec{x})} \right]$$

where $g(\vec{x}) = \sqrt[D]{x_1 * x_2 * \dots * x_D}$.

Zero replacement was carried out using the R package `zCompositions` (40).

Random forest implementation and statistical analysis. For modeling, we used samples with microbiome data available at the study baseline that did not include missing values on any of the metabolic parameters considered. In addition, subjects who had reimbursement for drug treatment of diabetes were excluded. This resulted in 529 participants for the 18-month follow-up visit and 482 participants for the 48-month follow-up visit.

All random forest models were implemented in R using the `caret` package and fast implementation of the random forest algorithm named `ranger` (41). Data sets were repeatedly split in a 75/25 ratio to training/test data sets, respectively, using a different seed each time. Models were tuned on training data using 10-fold cross-validation and random hyperparameter search with 100 hyperparameter combinations.

Performance of the models was evaluated on the test data set using root-mean-square error (RMSE). In the case of random forest models, out-of-bag (OOB) error is also widely used to evaluate model performance. Although using out-of-bag error for evaluation can increase the sample size for model training, it has been shown that in some cases the OOB error is largely overestimated and unreliable (42). Thus, for robust estimates, test data were used for evaluation. Permutation feature importance was used for selecting the microbial biomarkers. For explaining the obtained random forest models, accumulated local effects (ALE) plots were implemented using the R package DALEX (43). ALE plots aim to describe the effect of a certain predictor on the metabolic outcome independently of the remaining predictors (11).

A one-tailed binomial test was carried out to test whether the probability of the model including microbial predictors outperforming the model excluding microbial predictors is greater than 0.5. The Bonferroni correction was applied to assess significance (eight metabolic outcomes and two time points; $P < 0.05/16$).

Data availability. Individual-level 16S RNA sequencing data are available in the Sequence Read Archive (SRA) under BioProject number [PRJNA644655](https://doi.org/10.5281/zenodo.4422486). All remaining phenotype data in this study are available upon request through application to the METSIM data access committee. R codes used for the analysis are available at <https://doi.org/10.5281/zenodo.4422486>.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 0.2 MB.

FIG S2, PDF file, 0.2 MB.

FIG S3, PDF file, 0.4 MB.

TABLE S1, DOCX file, 0.02 MB.

TABLE S2, DOCX file, 0.03 MB.

TABLE S3, DOCX file, 0.03 MB.

ACKNOWLEDGMENTS

We thank all METSIM study participants.

All subjects have given written informed consent, and the study was approved by the Ethics Committee of the University of Kuopio and was in accordance with the Helsinki Declaration.

We declare that we have no competing interests.

This work was funded by Estonian Research Council grants PUT 1371 (to E.O.) and PUT1665 (to K.F.), EMBO Installation grant 3573 (to E.O.), NIH grants HL28481 (to A.J.L. and M.L.), HL144651 (to A.J.L.), and DK117850 (to A.J.L.). E.O. was supported by European Regional Development Fund project number 15-0012 GENTRANSMED and Estonian Center of Genomics/Roadmap II project number 16-0125. The METSIM study was supported by grants from the Academy of Finland (321428), the Sigrid Juselius Foundation, the Finnish Foundation for Cardiovascular Research, Kuopio University Hospital, and the Centre of Excellence of Cardiovascular and Metabolic Diseases, supported by the Academy of Finland (to M.L.). O.A. was supported by the European Regional Development Fund (smart specialization Ph.D. scholarship).

O.A. designed the study, performed the data analyses, and wrote the manuscript. J.K. and M.L. designed the METSIM study and oversaw collection of the METSIM samples. J.L. prepared samples and performed 16S microbiome sequencing. K.L. and C.P. carried out the bioinformatic analyses from raw microbiome data. E.O. wrote and reviewed the manuscript and supervised the data analysis. A.J.L., K.F., and M.L. reviewed the manuscript.

REFERENCES

- World Health Organization. 2016. Global report on diabetes. WHO, Geneva, Switzerland.
- Gurung M, Li Z, You H, Rodrigues R, Jump DB, Morgun A, Shulzhenko N. 2020. Role of gut microbiota in type 2 diabetes pathophysiology. *EBioMedicine* 51:102590. <https://doi.org/10.1016/j.ebiom.2019.11.051>.
- Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D, Peng Y, Zhang D, Jie Z, Wu W, Qin Y, Xue W, Li J, Han L, Lu D, Wu P, Dai Y, Sun X, Li Z, Tang A, Zhong S, Li X, Chen W, Xu R, Wang M, Feng Q, Gong M, Yu J, Zhang Y, Zhang M, Hansen T, Sanchez G, Raes J, Falony G, Okuda S, Almeida M, Lechatelier E, Renault P, Pons N, Batto JM, Zhang Z, Chen H, Yang R, Zheng W, Li S, et al. 2012. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* 490:55–60. <https://doi.org/10.1038/nature11450>.
- Karlsson FH, Tremaroli V, Nookaew I, Bergström G, Behre CJ, Fagerberg B, Nielsen J, Bäckhed F. 2013. Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature* 498:99–103. <https://doi.org/10.1038/nature12198>.
- Allin KH, Tremaroli V, Caesar R, Jensen BAH, Damgaard MTF, Bahl MI, Licht TR, Hansen TH, Nielsen T, Dantoft TM, Linneberg A, Jørgensen T, Vestergaard H, Kristiansen K, Franks PW, Hansen T, Bäckhed F, Pedersen O, the IMI-DIRECT Consortium. 2018. Aberrant intestinal microbiota in individuals with prediabetes. *Diabetologia* 61:810–820. <https://doi.org/10.1007/s00125-018-4550-1>.

6. Gale EAM. 2013. Is type 2 diabetes a category error? *Lancet* 381:1956–1957. [https://doi.org/10.1016/S0140-6736\(12\)62207-7](https://doi.org/10.1016/S0140-6736(12)62207-7).
7. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. 2009. Bacterial community variation in human body habitats across space and time. *Science* 326:1694–1697. <https://doi.org/10.1126/science.1177486>.
8. Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, Creasy HH, Earl AM, Fitzgerald MG, Fulton RS, Giglio MG, Hallsworth-Pepin K, Lobos EA, Madupu R, Magrini V, Martin JC, Mitreva M, Muzny DM, Sodergren EJ, Versalovic J, Wollam AM, Worley KC, Wortman JR, Young SK, Zeng Q, Aagaard KM, Abolude OO, Allen-Vercoe E, Alm EJ, Alvarado L, Andersen GL, Anderson S, Appelbaum E, Arachchi HM, Armitage G, Arze CA, Ayvaz T, Baker CC, Begg L, Belachew T, Bhonagiri V, Bihan M, Blaser MJ, Bloom T, Bonazzi V, Paul Brooks J, Buck GA, Buhay CJ, Busam DA, Campbell JL, Canon SR, Cantarel BL, Chain PSG, Chen IMA, et al. 2012. Structure, function and diversity of the healthy human microbiome. *Nature* 486:207–214. <https://doi.org/10.1038/nature11234>.
9. Pasolli E, Truong DT, Malik F, Waldron L, Segata N. 2016. Machine learning meta-analysis of large metagenomic datasets: tools and biological insights. *PLoS Comput Biol* 12:e1004977. <https://doi.org/10.1371/journal.pcbi.1004977>.
10. Topçuoğlu BD, Lesniak NA, Ruffin M, Wiens J, Schloss PD. 2019. Effective application of machine learning to microbiome-based classification problems. *bioRxiv* <https://doi.org/10.1128/mBio.00434-20>.
11. Apley DW, Zhu J. 2016. Visualizing the effects of predictor variables in black box supervised learning models. *arXiv* 1612.08468. <https://arxiv.org/abs/1612.08468>.
12. Gou W, Ling C, Jiang Z, He Y, Fu Y, Xu F, Miao Z, Sun T, Lin J, Zhu H, Zhou H, Chen Y, Zheng J-S. 2020. Interpretable machine learning framework reveals novel gut microbiome features in predicting type 2 diabetes. *bioRxiv* <https://doi.org/10.1101/2020.04.05.024984>.
13. Tabák AG, Herder C, Rathmann W, Brunner EJ, Kivimäki M. 2012. Prediabetes: a high-risk state for diabetes development. *Lancet* 379:2279–2290. [https://doi.org/10.1016/S0140-6736\(12\)60283-9](https://doi.org/10.1016/S0140-6736(12)60283-9).
14. Sanna S, van Zuydam NR, Mahajan A, Kurilshikov A, Vich Vila A, Vösa U, Mujagic Z, Masclee AAM, Jonkers DMAE, Oosting M, Joosten LAB, Netea MG, Franke L, Zhernakova A, Fu J, Wijmenga C, McCarthy MI. 2019. Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. *Nat Genet* 51:600–605. <https://doi.org/10.1038/s41588-019-0350-x>.
15. Org E, Mehrabian M, Parks BW, Shipkova P, Liu X, Drake TA, Lusi AJ. 2016. Sex differences and hormonal effects on gut microbiota composition in mice. *Gut Microbes* 7:313–322. <https://doi.org/10.1080/19490976.2016.1203502>.
16. Kim YS, Unno T, Kim BY, Park MS. 2020. Sex differences in gut microbiota. *World J Men's Health* 38:48–60. <https://doi.org/10.5534/wjmh.190009>.
17. Mamlouk D, Gullo M. 2013. Acetic acid bacteria: physiology and carbon sources oxidation. *Indian J Microbiol* 53:377–384. <https://doi.org/10.1007/s12088-013-0414-z>.
18. Mitrou P, Petsiou E, Papakonstantinou E, Maratou E, Lambadiari V, Dimitriadis P, Spanoudi F, Raptis SA, Dimitriadis G. 2015. The role of acetic acid on glucose uptake and blood flow rates in the skeletal muscle in humans with impaired glucose tolerance. *Eur J Clin Nutr* 69:734–739. <https://doi.org/10.1038/ejcn.2014.289>.
19. Johnston CS, Kim CM, Buller AJ. 2004. Vinegar improves insulin sensitivity to a high-carbohydrate meal in subjects with insulin resistance or type 2 diabetes [10]. *Diabetes Care* 27:281–282. <https://doi.org/10.2337/diacare.27.1.281>.
20. Zhou W, Sailani MR, Contrepois K, Zhou Y, Ahadi S, Leopold SR, Zhang MJ, Rao V, Avina M, Mishra T, Johnson J, Lee-McMullen B, Chen S, Metwally AA, Tran TDB, Nguyen H, Zhou X, Albright B, Hong BY, Petersen L, Bautista E, Hanson B, Chen L, Spakowicz D, Bahmani A, Salins D, Leopold B, Ashland M, Dagan-Rosenfeld O, Rego S, Limcaoco P, Colbert E, Allister C, Perelman D, Craig C, Wei E, Chaib H, Hornburg D, Dunn J, Liang L, Rose SMSF, Kukurba K, Piening B, Rost H, Tse D, McLaughlin T, Sodergren E, Weinstock GM, Snyder M. 2019. Longitudinal multi-omics of host–microbe dynamics in prediabetes. *Nature* 569:663–671. <https://doi.org/10.1038/s41586-019-1236-x>.
21. Krych L, Nielsen DS, Hansen AK, Hansen CHF. 2015. Gut microbial markers are associated with diabetes onset, regulatory imbalance, and IFN- γ level in NOD mice. *Gut Microbes* 6:101–109. <https://doi.org/10.1080/19490976.2015.1011876>.
22. Wu X, Ma C, Han L, Nawaz M, Gao F, Zhang X, Yu P, Zhao C, Li L, Zhou A, Wang J, Moore JE, Cherie Millar B, Xu J. 2010. Molecular characterisation of the faecal microbiota in patients with type II diabetes. *Curr Microbiol* 61:69–78. <https://doi.org/10.1007/s00284-010-9582-9>.
23. Zhang X, Shen D, Fang Z, Jie Z, Qiu X, Zhang C, Chen Y, Ji L. 2013. Human gut microbiota changes reveal the progression of glucose intolerance. *PLoS One* 8:e71108. <https://doi.org/10.1371/journal.pone.0071108>.
24. Zhernakova A, Kurilshikov A, Bonder MJ, Tigchelaar EF, Schirmer M, Vatanan T, Mujagic Z, Vila AV, Falony G, Vieira-Silva S, Wang J, Imhann F, Brandsma E, Jankipersadsing SA, Joossens M, Cenit MC, Deelen P, Swertz MA, Weersma RK, Feskens EJM, Netea MG, Gevers D, Jonkers D, Franke L, Aulchenko YS, Huttenhower C, Raes J, Hofker MH, Xavier RJ, Wijmenga C, Fu J, LifeLines Cohort Study. 2016. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science* 352:565–569. <https://doi.org/10.1126/science.aad3369>.
25. Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, Kurilshikov A, Bonder MJ, Valles-Colomer M, Vandeputte D, Tito RY, Chaffron S, Rymanans L, Verspecht C, De Sutter L, Lima-Mendez G, D'hoë K, Jonckheere K, Homola D, Garcia R, Tigchelaar EF, Eeckhaut L, Fu J, Henckaerts L, Zhernakova A, Wijmenga C, Raes J. 2016. Population-level analysis of gut microbiome variation. *Science* 352:560–564. <https://doi.org/10.1126/science.aad3503>.
26. Quinn TP, Erb I, Gloor G, Notredame C, Richardson MF, Crowley TM. 2019. A field guide for the compositional analysis of any-omics data. *GigaScience* 8:giz107. <https://doi.org/10.1093/gigascience/giz107>.
27. Quinn T, Erb I. 2019. Using balances to engineer features for the classification of health biomarkers: a new approach to balance selection. *bioRxiv* <https://doi.org/10.1101/600122>.
28. Tolosana-Delgado R, Talebi H, Khodadadzadeh M, Boogaart K. 2019. On machine learning algorithms and compositional data, p 172–175. *Proc 8th Int Work Compos Data Anal*.
29. Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, Gonzalez A, Kosciolek T, McCall LI, McDonald D, Melnik AV, Morton JT, Navas J, Quinn RA, Sanders JG, Swafford AD, Thompson LR, Tripathi A, Xu ZZ, Zaneveld JR, Zhu Q, Caporaso JG, Dorrestein PC. 2018. Best practices for analysing microbiomes. *Nat Rev Microbiol* 16:410–422. <https://doi.org/10.1038/s41579-018-0029-9>.
30. Laakso M, Kuusisto J, Stančáková A, Kuulasmaa T, Pajukanta P, Lusi AJ, Collins FS, Mohlke KL, Boehnke M. 2017. The metabolic syndrome in men study: a resource for studies of metabolic and cardiovascular diseases. *J Lipid Res* 58:481–493. <https://doi.org/10.1194/jlr.0072629>.
31. Matsuda M, DeFronzo RA. 1999. Insulin sensitivity indices obtained from oral glucose tolerance testing: comparison with the euglycemic insulin clamp. *Diabetes Care* 22:1462–1470. <https://doi.org/10.2337/diacare.22.9.1462>.
32. Stančáková A, Javorsky M, Kuulasmaa T, Haffner SM, Kuusisto J, Laakso M. 2009. Changes in insulin sensitivity and insulin release in relation to glycaemia and glucose tolerance in 6,414 Finnish men. *Diabetes* 58:1212–1221. <https://doi.org/10.2337/db08-1607>.
33. Org E, Blum Y, Kasela S, Mehrabian M, Kuusisto J, Kangas AJ, Soininen P, Wang Z, Ala-Korpela M, Hazen SL, Laakso M, Lusi AJ. 2017. Relationships between gut microbiota, plasma metabolites, and metabolic syndrome traits in the METSIM cohort. *Genome Biol* 18:70. <https://doi.org/10.1186/s13059-017-1194-2>.
34. Bolyen E, Rideout JR, Dillon M, Bokulich N, Abnet C, Al-Ghalith G, Alexander H, Alm E, Arumugam M, Asnicar F, Bai Y, Bisanz J, Bittinger K, Brejnrod A, Brislawn C, Brown T, Callahan B, Caraballo-Rodríguez AM, Chase J, Cope E, Da Silva R, Dorrestein P, Douglas G, Durall D, Duvallet C, Edwardson C, Ernst M, Estaki M, Fouquier J, Gauglitz J, Gibson D, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley G, Janssen S, Jarmusch A, Jiang L, Kaehler B, Bin Kang K, Keefe C, Keim P, Kelley S, Knights D, Koester I, Kosciolek T, Kreps J, et al. 2018. QIIME 2: reproducible, interactive, scalable, and extensible microbiome data science. *PeerJ Preprints* 6:e27295v2. <https://peerj.com/preprints/27295/>.
35. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods* 13:581–583. <https://doi.org/10.1038/nmeth.3869>.
36. Katoh K. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30:3059–3066. <https://doi.org/10.1093/nar/gkf436>.
37. Price MN, Dehal PS, Arkin AP. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* 26:1641–1650. <https://doi.org/10.1093/molbev/msp077>.

38. Gloor GB, Reid G. 2016. Compositional analysis: a valid approach to analyze microbiome high-throughput sequencing data. *Can J Microbiol* 62:692–702. <https://doi.org/10.1139/cjm-2015-0821>.
39. Aitchison J. 1982. The statistical analysis of compositional data. *J R Stat Soc Ser B Stat Methodol* 44:139–177.
40. Palarea-Albaladejo J, Martín-Fernández JA. 2015. ZCompositions: R package for multivariate imputation of left-censored data under a compositional approach. *Chemom Intell Lab Syst* 143:85–96. <https://doi.org/10.1016/j.chemolab.2015.02.019>.
41. Wright MN, Ziegler A. 2017. ranger: a fast implementation of random forests for high dimensional data in C++ and R. *J Stat Softw* 77. <https://doi.org/10.18637/jss.v077.i01>.
42. Janitza S, Hornung R. 2018. On the overestimation of random forest's out-of-bag error. *PLoS One* 13:e0201904. <https://doi.org/10.1371/journal.pone.0201904>.
43. Biecek P. 2018. DALEX: explainers for complex predictive models in R. *J Mach Learn Res* 19:1–5.