

<https://doi.org/10.1038/s41522-025-00650-9>

Personalized prediction of glycemic responses to food in women with diet-treated gestational diabetes: the role of the gut microbiota

Check for updates

Polina V. Popova^{1,2}✉, Artem O. Isakov¹, Anastasiia N. Rusanova¹, Stanislav I. Sitkin^{3,4}, Anna D. Anopova¹, Elena A. Vasukova¹, Alexandra S. Tkachuk², Irina S. Nemikina¹, Elizaveta A. Stepanova², Angelina I. Eriskovskaya¹, Ekaterina A. Stepanova², Evgenii A. Pustozero², Maria A. Kokina¹, Elena Y. Vasilieva¹, Lyudmila B. Vasilyeva⁵, Soha Zgairi⁶, Elad Rubin⁶, Carmel Even⁶, Sondra Turjeman⁶, Tatiana M. Pervunina³, Elena N. Grineva^{1,2}, Omry Koren⁶ & Evgeny V. Shlyakhto¹

We developed a prediction model for postprandial glycemic response (PPGR) in pregnant women, including those with diet-treated gestational diabetes mellitus (GDM) and healthy women, and explored the role of gut microbiota in improving prediction accuracy. The study involved 105 pregnant women (77 with GDM, 28 healthy), who underwent continuous glucose monitoring (CGM) for 7 days, provided food diaries, and gave stool samples for microbiome analysis. Machine learning models were created using CGM data, meal content, lifestyle factors, biochemical parameters, and microbiota data (16S rRNA gene sequence analysis). Adding microbiome data increased the explained variance in peak glycemic levels (GLUmax) from 34 to 42% and in incremental area under the glycemic curve (iAUC120) from 50 to 52%. The final model showed better correlation with measured PPGRs than one based only on carbohydrate count ($r = 0.72$ vs. $r = 0.51$ for iAUC120). Although microbiome features were important, their contribution to model performance was modest.

Gestational diabetes mellitus (GDM) represents a prevalent condition, impacting a substantial portion, approximately up to 9–26%, of pregnancies¹. GDM can lead to pregnancy complications, including but not limited to increased cesarean section rate, birth trauma, nerve palsy, neonatal hypoglycemia in the short-term period², and increased risks of obesity, type 2 diabetes, and cardiovascular diseases in both mothers and their offspring throughout life³. It is crucial to uphold normal glycemic levels during pregnancy to mitigate adverse pregnancy outcomes and disrupt the cyclical transmission of predisposition to metabolic diseases across generations^{3,4}.

The most common treatment for GDM is diet and lifestyle modification, reportedly effective without adding medications for achieving glucose control in 70–85% of women⁵. However, according to real-life observations, many women with GDM do not achieve target glucose

levels, and the rate of pregnancy complications remains high in women with GDM⁶.

Nutritional studies and guidelines concerning GDM concentrate on average characteristics across populations⁵. The historically prevailing approach to predicting postprandial glucose responses (PPGRs) to food involves relying on the carbohydrate content of the meal⁵, despite evidence suggesting its inadequacy as a predictor⁷. Alternative methods include the glycemic index, which assesses the postprandial glucose response to a specific food, and the derived glycemic load⁵. Attributing a singular postprandial glucose response (PPGR) to each food implies that the response is entirely inherent to the food itself. Yet, recent studies exploring interindividual variations in PPGRs have revealed significant variability in how different individuals respond to identical foods^{7,8}. The distinctive PPGRs of an individual were shown to be influenced by their

¹World-Class Research Center for Personalized Medicine, Almazov National Medical Research Centre, Saint Petersburg, Russia. ²Institute of Endocrinology, Almazov National Medical Research Centre, Saint Petersburg, Russia. ³Institute of Perinatology and Pediatrics, Almazov National Medical Research Centre, Saint Petersburg, Russia. ⁴Department of Internal Diseases, Gastroenterology and Dietetics, North-Western State Medical University named after I.I. Mechnikov, Saint Petersburg, Russia. ⁵Institute of Molecular Biology and Genetics, Almazov National Medical Research Centre, Saint Petersburg, Russia. ⁶Azieli Faculty of Medicine, Bar-Ilan University, Ramat Gan, Israel. ✉e-mail: pvpopova5@gmail.com

biological traits (such as gut microbiome composition and genetic variation) and lifestyle factors⁷⁻⁹.

Several studies integrated gut microbial features into the models predicting PPGRs in healthy individuals^{7,8} or patients with type 1 diabetes⁹. However, only a few evaluated the impact of microbiome on the prediction of PPGRs through variance components analyses⁸ or the SHapley Additive exPlanation (SHAP) method⁹, and only one study exclusively concentrated on investigating the connections between bacterial species and the host glycemic regulation¹⁰. Furthermore, none of these studies directly compared the accuracy metrics of the models before and after the addition of microbial data into the list of the input variables. They compared basic models including carbohydrates and premeal glucose level with a full model including multiple individual parameters apart from microbial features⁷⁻⁹.

Moreover, although there is mounting evidence regarding the regulatory functions of the microbiome in normal and impaired glycemic responses among non-pregnant individuals⁷⁻¹⁰, limited knowledge exists concerning the microbiome's impact on PPGRs among pregnant women, both with and without GDM. We have previously developed PPGR prediction models based on multiple individual parameters without microbiome for pregnant women with and without GDM¹¹. The performance of our model in predicting the incremental area under the glucose curve in the 2 h after the meal (iAUC120) ($R = 0.7$) was comparable to the model created by Zeevy et al. for healthy individuals based on individual parameters and microbiome ($R = 0.7$)⁷. However, our past model was inferior to the accuracy of the microbiome-based model subsequently developed by Berry et al. for iAUC120 prediction in healthy individuals ($R = 0.77$)⁸, leaving space for the improvement of our model, potentially with the inclusion of microbiome data. Another distinguishing feature of the Berry model, in comparison to both our previous model and the Zeevi model, was the incorporation of genetic factors. This likely enhanced the predictive accuracy, as genetics constituted the second most crucial parameter group after serum glycemic markers, as indicated by the proportion of variance explained (R^2)⁸.

The aim of this study was to develop a prediction model for PPGR in pregnant women, including both those with GDM and normal glucose tolerance, and to explore the impact of microbial data on the model's performance. An accurate PPGR prediction model holds promise in optimizing personalized diet recommendations to improve glucose control and

pregnancy outcomes in women with GDM. Simultaneously, the identification of a distinct gut microbial signature affecting PPGR, a secondary aim of this project, could serve as a basis for the development of potential therapeutic interventions.

Results

In total, 152 participants were recruited for the study. After the exclusion of 3 women who did not provide CGM data, 2 women with antibiotics intake during the study period, 34 women with inaccurate food diaries, 2 women with less than 6 meals left after filtering, and 6 microbiota samples with low read count (<10,000 reads), 105 participants (77 women with GDM and 28 healthy pregnant women) were included in the final analysis (Fig. 1).

The characteristics of the participants are in Table 1. Women with GDM did not differ from the control group in terms of age and gestational age upon initiation of continuous glucose monitoring. Patients with GDM had higher body mass index (BMI) before pregnancy. As expected, healthy pregnant women had lower plasma glucose levels during OGTT and hemoglobin A1C (HbA1C) upon inclusion into the study.

Patients with GDM consumed lower amounts of carbohydrates (28.4 ± 10.9 vs 36.6 ± 10.8 g) and higher amounts of proteins (17.0 ± 5.2 vs 13.8 ± 2.9 g) per meal compared to healthy women (Table 1). Presumably due to this fact, iAUC120 and GLUmax levels did not significantly differ between the groups and even tended to be lower in women with GDM compared to their healthy counterparts who were not dieting (0.52 ± 0.29 vs 0.63 ± 0.28 and 6.2 ± 0.6 vs 6.4 ± 0.6 mmol/L, respectively) (Table 1). For comprehensive details on lifestyle assessments and baseline blood tests, please refer to Supplementary Table S3.

Microbial features in women with higher and lower PPGRs

As there was no difference in the levels of GLUmax and iAUC120 between women with and without GDM during CGM, we combined their data for selection of microbial features associated with higher and lower PPGRs. The medians for iAUC120 and GLUmax in the cohort were 0.527 and 6.254 mmol/L, respectively. Participants with median PPGR indices (iAUC120 or GLUmax, respectively) below these numbers were considered to have lower PPGRs, and those with median PPGR indices equal to or above the cohort median comprised the subgroup with higher PPGRs.

Fig. 1 | Cohort selection.

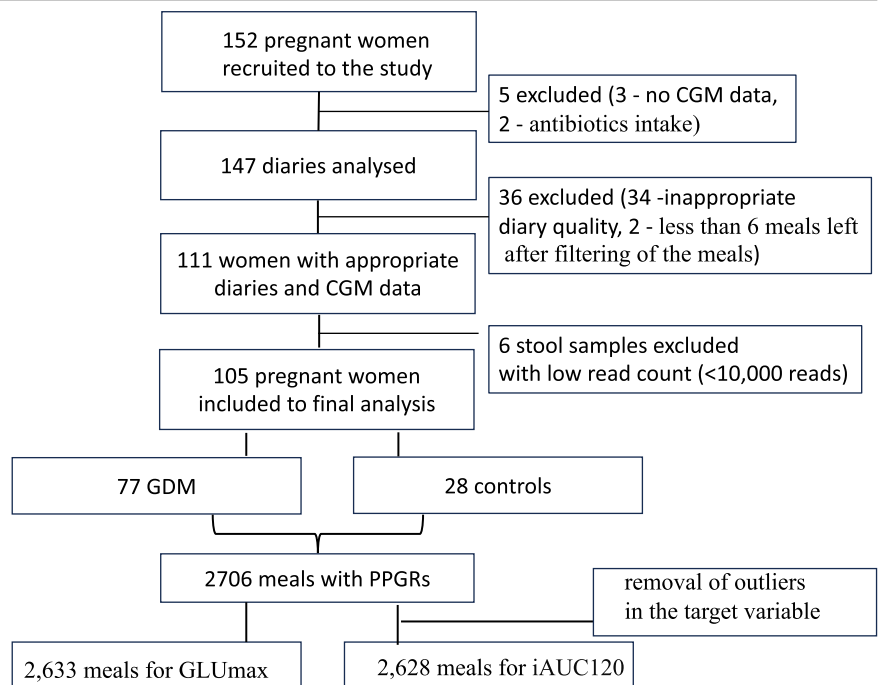


Table 1 | Characteristics of the participants

	GDM N = 77 (mean ± SD)	Healthy pregnant women N = 28 (mean ± SD)	p-value
Age, years	32.2 ± 4.3	31.4 ± 4.7	0.392
Gestational age (weeks)	30.1 ± 4.0	30.0 ± 2.0	0.947
BMI before pregnancy (kg/m ²)	24.7 ± 5.2	22.1 ± 3.6	0.017
Fasting PG (mmol/L)	5.1 ± 0.5	4.4 ± 0.4	<0.001
1-h postload glucose at OGTT (mmol/L)	9.5 ± 1.5	6.7 ± 1.6	<0.001
2-h postload glucose at OGTT (mmol/L)	8.3 ± 1.7	6.0 ± 1.2	<0.001
HbA1C, %	5.0 ± 0.4	4.8 ± 0.3	0.020
Real-time meal logging			
Days logged per participant	7.2 ± 0.6	7.0 ± 0.8	0.78
Meals logged per participant	26.5 ± 3.7	25.3 ± 4.7	0.72
Energy intake per meal (kcal)	334.8 ± 99.8	346.3 ± 95.0	0.597
Carbohydrate intake per meal (g)	28.4 ± 10.9	36.6 ± 10.8	0.001
Glycaemic load per meal (g)	14.8 ± 7.5	20.4 ± 6.8	0.001
Fat intake per meal (g)	16.4 ± 5.9	15.3 ± 5.9	0.413
Protein intake per meal (g)	17.0 ± 5.2	13.8 ± 2.9	0.003
CGM-derived indices			
Mean GLUmax, mmol/L	6.2 ± 0.6	6.4 ± 0.6	0.083
Mean iAUC120, mmol/L	0.52 ± 0.29	0.63 ± 0.28	0.092

Comparisons were performed using the bootstrap hypothesis test.
OGTT oral glucose tolerance test.

Linear discriminant analysis revealed 18 bacterial taxa exhibiting significantly higher scores in the subgroup of women with higher iAUC120 and 21 bacterial taxa with higher scores in the subgroup with lower iAUC120, $p < 0.05$ for all (Fig. 2). All bacterial taxa distinguished by LefSe were included as input variables for creation of PPGR prediction models.

When comparing women with higher and lower GLUmax, 7 taxa were enriched in the subgroup with higher GLUmax, including *Clostridia* UCG 014 and “*Lachnospiraceae*” (*Lachnospiraceae*), and 8 taxa were enriched in the subgroup with lower GLUmax, including *Methanospaera* (*Methanobacteria*), *Lachnospira eligens* (basonym: *Eubacterium eligens*) (*Lachnospiraceae*), *Butyrivibrio faecihominis* (*Oscillospiraceae*), *Intestinibacter bartlettii* (*Clostridia*, *Peptostreptococcaceae*), *Sellimonas* (*Lachnospiraceae*), *E. taylori* (*Lachnospiraceae*), *Christensenellaceae R-7 group* (*Clostridia*) (Fig. 3).

Predicting individual postprandial responses

We assessed the overall extent to which different combinations of input variables predict personal postprandial responses: iAUC120 and GLUmax. A total of 750 days of concurrent CGM usage and meal logging resulted in 3514 meals to be analyzed with their PPGRs. Meal filtering (see “Research design and methods: Meal preprocessing”) reduced the dataset to 2706 meals. After the removal of outliers in the target variable, the final dataset comprised 2633 meals with PPGRs for GLUmax prediction model and 2628 meals for iAUC120 prediction. Prediction models for both indices were developed utilizing gradient boosting algorithms, with the following combinations of input variables: (1) only carbohydrate content of the meal (carbs); (2) clinically available parameters (anthropometric, biochemical,

lifestyle questionnaire, meal content and meal context, CGM data); (3) model 2 parameters + microbial features (the full model). For the full list of features please see the Supplementary Table 1. Validation of the model was performed using a three-fold cross-validation scheme (see “Research design and methods”).

In the context of predicting GLUmax, the first model that relied solely on the amount of carbohydrates in a meal demonstrated the lowest correlation with PPGRs ($R = 0.35$) and accounted for only 5% of the variation in glycemic response (Fig. 4a). The second model based on clinically available parameters achieved a significantly higher correlation ($R = 0.62$) and explained 34% of variance (Fig. 4b). Adding microbiome features (Fig. 4c) further increased the predictive ability with an R of 0.66 and a coefficient of determination of 42%.

Likewise, in the prediction of iAUC120, a model based solely on the carbohydrate content of meals demonstrated a relatively weak correlation ($R = 0.51$) and explained only 26% of the variation in glycemic response (Fig. 5a). The addition of parameter groups, as described above, resulted in an increase in correlation between CGM-measured and predicted values ($R = 0.71$, $R^2 = 0.50$). Addition of microbial features to this model slightly increased the accuracy of prediction ($R = 0.72$, $R^2 = 0.52$) (Fig. 5b, c).

Because the performance of a model can also be affected by non-linear relationships between measured and predicted values, we also assessed MAE, MSE and RMSE for the models with higher performance (models 2–3, Table 2). As shown in Table 2, adding microbial features decreased MAE, MSE and RMSE for GLUmax prediction but did not influence these parameters characterizing prediction of iAUC120.

Exploring factors influencing the prediction of postprandial glycemic responses

Following the examination of different models predicting PPGRs, our subsequent focus was on understanding the individual factors influencing prediction accuracy, including microbial features and other parameters comprising the full model. For this purpose, we conducted feature attribution analysis employing SHAP¹².

The features that exerted the greatest influence on iAUC120 prediction, as indicated by the highest mean absolute SHAP value, encompassed the carbohydrate content of the meal, glycemic load of the meal, amount of starch in the meal, and CGM-derived parameters characterizing glucose levels preceding the meal (glucose level 10 min before meal and glucose rise from 240 min before the meal to meal start) (Fig. 6a). The most influential parameters for the prediction of GLUmax were the glucose levels at the onset of the meal (GLU0), the carbohydrate content of the meal, glycemic load of the meal, RA of *I. bartlettii*, and the amount of protein consumed up to 6 h before the meal (Fig. 6b).

Among the 20 most influential parameters for the prediction of iAUC120 or GLUmax, the algorithm selected the RA of the following bacterial taxa: *I. bartlettii*, “*L. edouardi*”, *B. faecihominis* (for iAUC120), and *I. bartlettii*, *L. eligens* (basonym: *Eubacterium eligens*), and *R. champagneensis* (for GLUmax) (Fig. 6a, b). Notably, *I. bartlettii* ranked fourth among influential parameters for the prediction of GLUmax and was selected by the algorithm among the top parameters both for iAUC120 and for GLUmax prediction.

In order to assess the cumulative influence of microbial composition and other feature groups on the model, we summed the SHAP values of associated features (Fig. 7). These examinations revealed that the meal composition had the most significant effect on prediction of iAUC120, followed by CGM-derived data, meal context, and microbial composition (Fig. 7a). On the contrary, for the prediction of GLUmax the main predictor group was the CGM-derived data, followed by meal composition, meal context, and microbial data also taking the fourth place (Fig. 7b).

Discussion

Recently, a high interpersonal difference in PPGRs was revealed, and the gut microbiota has been shown to be a factor underlying this variability. Furthermore, the gut microbiota has been used to enhance the accuracy of

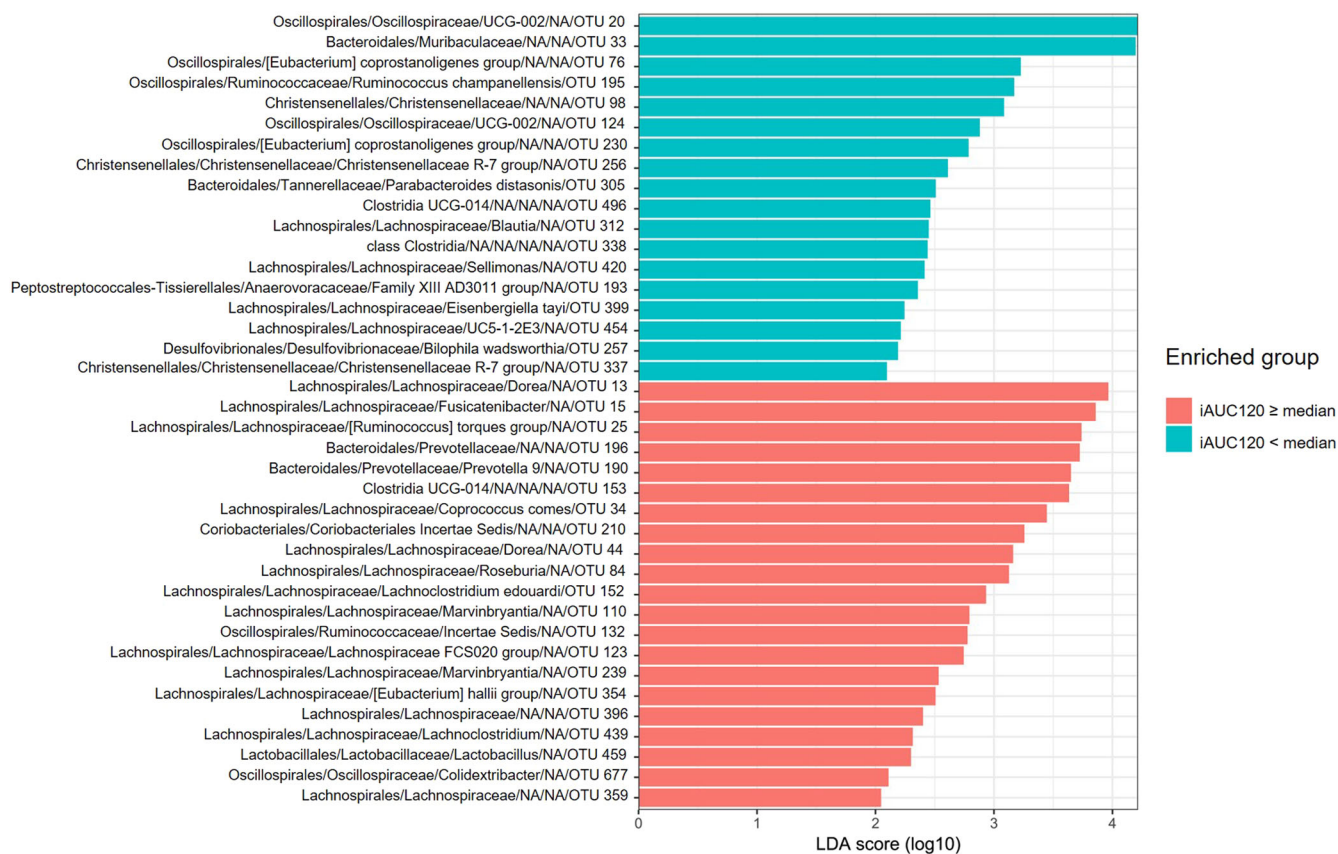


Fig. 2 | Results of LefSe analysis comparing relative abundance (RA) of microbial features of participants with PPGR (iAUC120) below and above median. Red indicates higher RA in patients with iAUC120 equal to or above median, and turquoise indicates higher RA in patients with iAUC120 below median, ranked by the effect size.

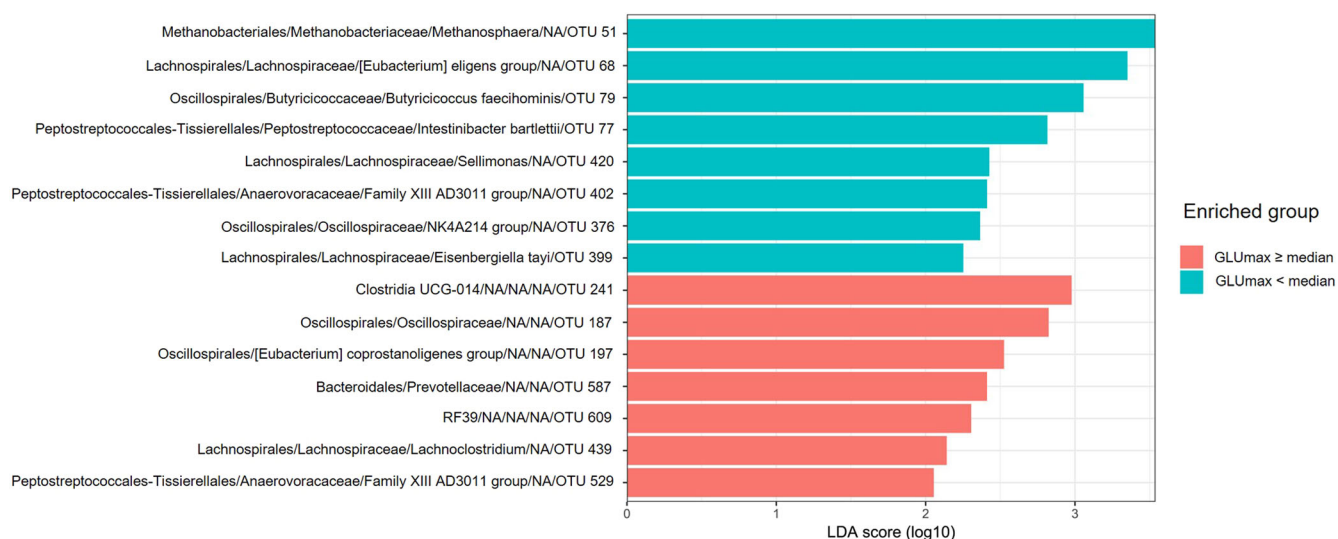


Fig. 3 | Results of LefSe analysis comparing relative abundance (RA) of microbial features of participants with GLUmax below and above median. Red indicates higher RA in patients with GLUmax equal to or above median, and turquoise indicates higher RA in patients with GLUmax below median, ranked by the effect size.

PPGR prediction in healthy volunteers and individuals with type 1 diabetes⁷⁻⁹. However, to our knowledge, no study has explored the impact of the gut microbiome on PPGRs in pregnant women with or without GDM.

Pregnancy is characterized by substantial alterations in all types of metabolism as well as by dynamic changes in gut microbial composition¹³⁻¹⁷. This fact, in line with the importance of achieving target postprandial glucose levels for improved pregnancy outcomes, underscores the importance of comprehensive evaluation of factors underlying PPGRs in pregnant

women for the construction of more accurate PPGR prediction models for personalized dietary advice in this specific population.

Our study has shown that microbiome features are among the top 5 most impactful individual parameters on the PPGR prediction in pregnant women with GDM. The cumulative influence of microbial composition was the fourth among the ten most impactful feature groups following CGM-derived data, meal composition and meal context for the prediction of GLUmax and iAUC120. Of note, adding microbiome features increased the

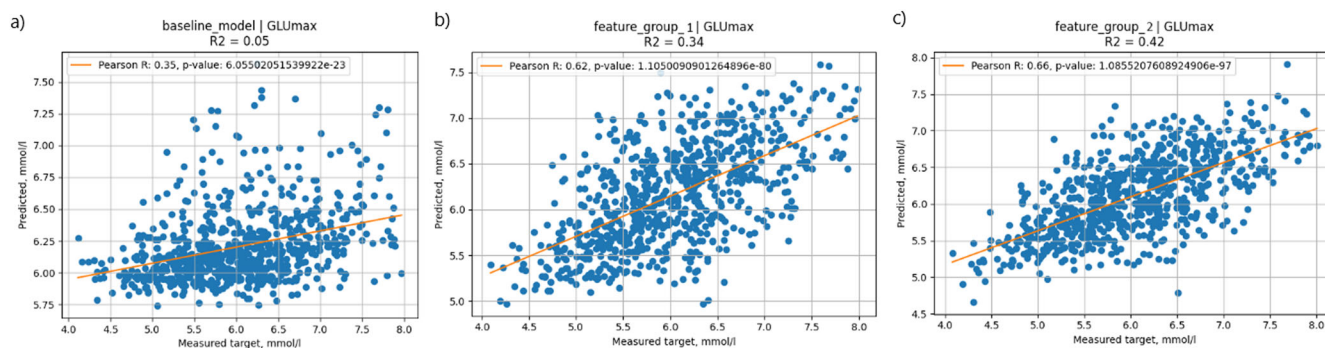


Fig. 4 | The results of peak postprandial blood glucose prediction (GLUmax) with the test set. X scale—CGM-measured values, Y scale—predicted values. **a** Baseline model—solely carbohydrate content of the meal (carbs); **b** the model based on

clinically available parameters (anthropometric, biochemical, lifestyle questionnaire, meal content, meal context, CGM data); **c** full model—clinically available parameters + microbial features.

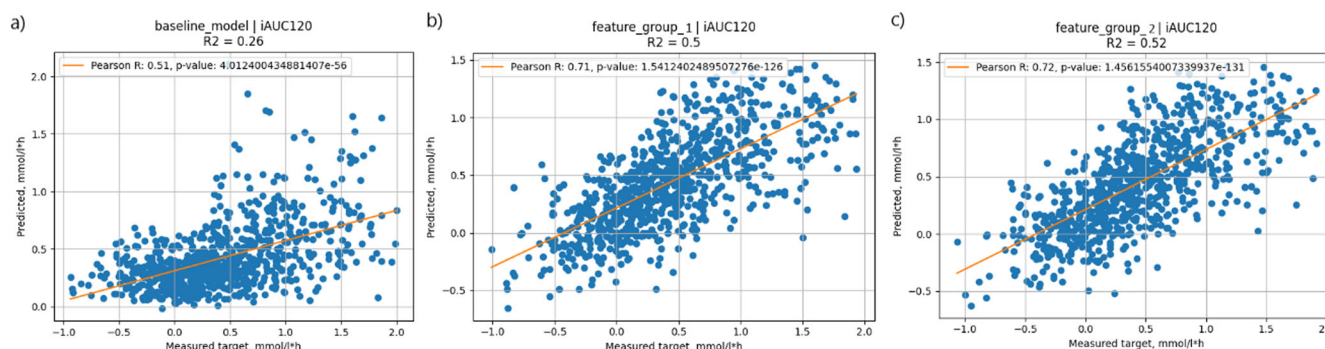


Fig. 5 | The results of iAUC120 prediction with the test set. X scale—CGM-measured values, Y scale—predicted values. **a** Baseline model—solely carbohydrate content of the meal (carbs); **b** the model based on clinically available parameters

(anthropometric, biochemical, lifestyle questionnaire, meal content, meal context, CGM data); **c** full model—clinically available parameters + microbial features.

Table 2 | Accuracy of the models for predicting PPGRs and peak postprandial glycaemic levels based on clinical data with and without the addition of bacterial features

	GLUmax		PPGR (iAUC120)	
	Without microbiome	With microbiome	Without microbiome	With microbiome
MAE	0.49	0.46	0.30	0.30
MSE	0.38	0.33	0.15	0.15
RMSE	0.62	0.57	0.39	0.39
Pearson <i>R</i>	0.62	0.66	0.71	0.72
<i>R</i> ²	0.34	0.42	0.50	0.52

CGM continuous glucose monitoring, GLUmax peak postprandial glycaemic level, PPGR postprandial glycaemic response, iAUC120 incremental area under glucose curve during 120 min after meal, MAE mean absolute error, MSE mean squared error, RMSE root mean squared error, *R* coefficient of correlation for predicted and observed values, *R*² coefficient of determination.

predictive ability of the GLUmax model with the increment of the coefficient of determination from 34 to 42%.

However, the addition of microbiome features had a lower impact on the accuracy metrics of iAUC120 prediction than on the metrics of GLUmax. This could be a consequence of relatively high accuracy of iAUC120 prediction even before microbiome addition. The addition of the microbiome had a lower impact on the models than the factors related to the meal (carbohydrates, starch, glycemic load), blood glucose levels at certain time points, and the foods eaten earlier that day (meal context). The relatively modest impact of the microbiome's addition on the performance of the

prognostic models can be attributed to the use of the 16S rRNA gene amplicon sequencing method or may reflect that the contribution of the microbiome to the regulation of the postprandial glucose levels is relatively limited.

The postprandial glycaemic predictions in our study (with *R* = 0.72 for the model predicting iAUC120) closely resembled those documented by Zeevi et al. (with an *R* value of 0.70) in healthy subjects⁷. The performance of our models was also superior to that of the models developed by Shilo et al. for patients with type 1 diabetes (*R* = 0.72 vs 0.59 for iAUC120 prediction and *R* = 0.66 vs 0.61 for GLUmax prediction)⁹. Furthermore, Shilo et al. included PPGRs from the same patient in training and validation datasets, while in our study, we separated participants between datasets so PPGRs of a participant from a training dataset could not be analyzed in the test or validation datasets. If Shilo et al. followed the same protocol, the difference in model performance might be even more pronounced. However, type 1 diabetes patients have much greater glucose variability and excursions, thus complicating the task of accurate PPGR prediction in this group of patients.

In the biggest study of PPGRs in healthy individuals, to date, Berry et al. obtained the highest accuracy of iAUC120 prediction with *R* = 0.75 in the validation cohort⁸. They likely reached the maximum accuracy which could be anticipated judging by the correlation between PPGRs to repeated standard meals (intraindividual variability) of 0.7–0.77 reported by Zeevi et al.⁷. A potential reason for the lower performance of our model is almost 10-fold smaller sample size and a lower number of included genetic variants compared to the study by Berry et al.⁸

Furthermore, taking into consideration well-known alterations in carbohydrate metabolism during pregnancy, especially in the third trimester¹³, we suppose that PPGRs may differ in a pregnant vs non-pregnant state. Compared to our previous study on PPGRs prediction

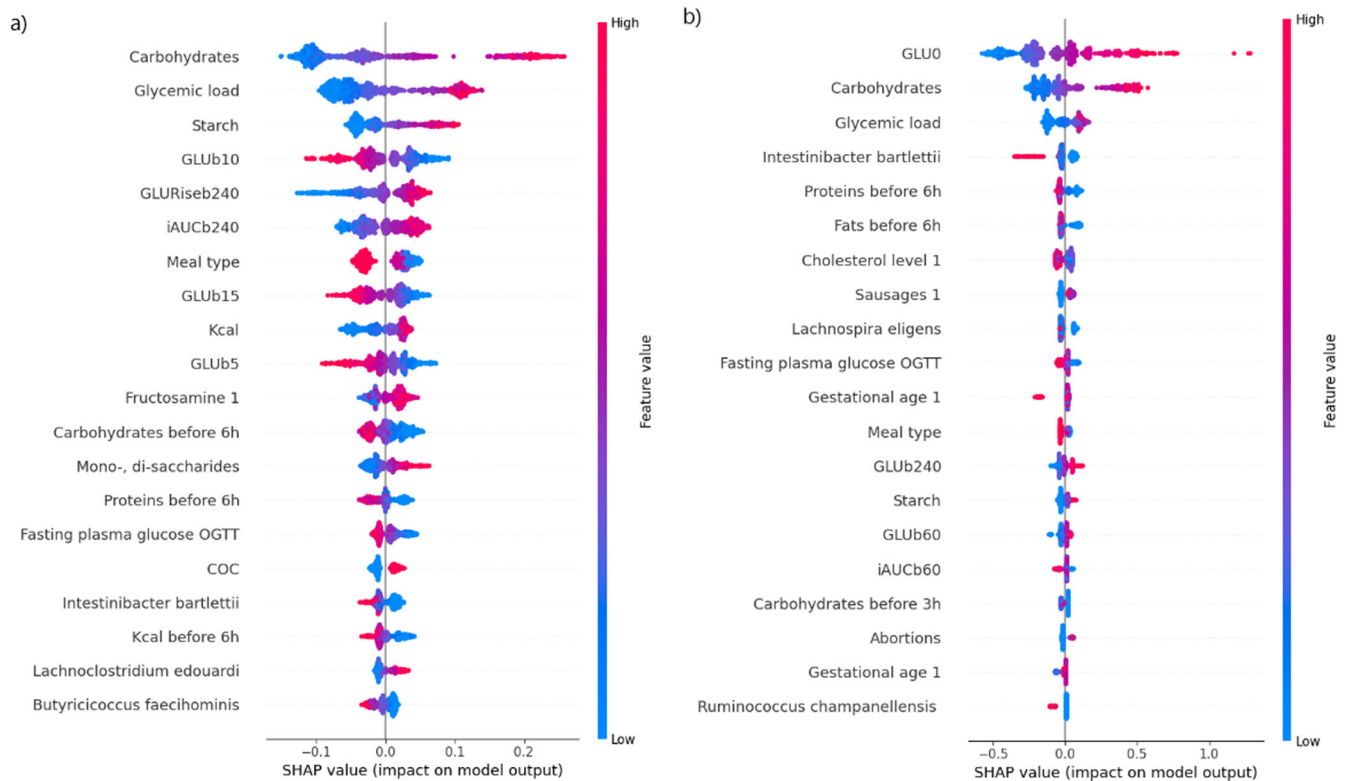


Fig. 6 | Significance level of the 20 most impactful variables of the models for predicting iAUC120 and GLUmax. iAUC120 (a) and GLUmax (b) are predicted based on full clinical data with the addition of bacterial features. Higher values of the feature are indicated by colors closer to red, lower values by colors closer to blue. If a point of a certain color is located on the left side of the central axis, the feature has a downward effect on the target variable; if the point is located on the right side, the effect will be the opposite. For example, lower values of GLU0 (the long blue tail on the left of b) correspond to lower values of the target variable (GLUmax). GLUb—

glucose level before meal initiation. Numbers near «GLUb» represent the minutes prior to meal initiation in which the measurement was obtained. For example, «GLUb10» represents the glucose level 10 min prior to the meal; Kcal—the energy value of the meal; COC—combined oral contraceptive use any time before pregnancy (1—yes, 0—no); Sausages 1—frequency of consuming sausage products before pregnancy. For a more detailed description of the input features, please refer to Supplementary Table 1.

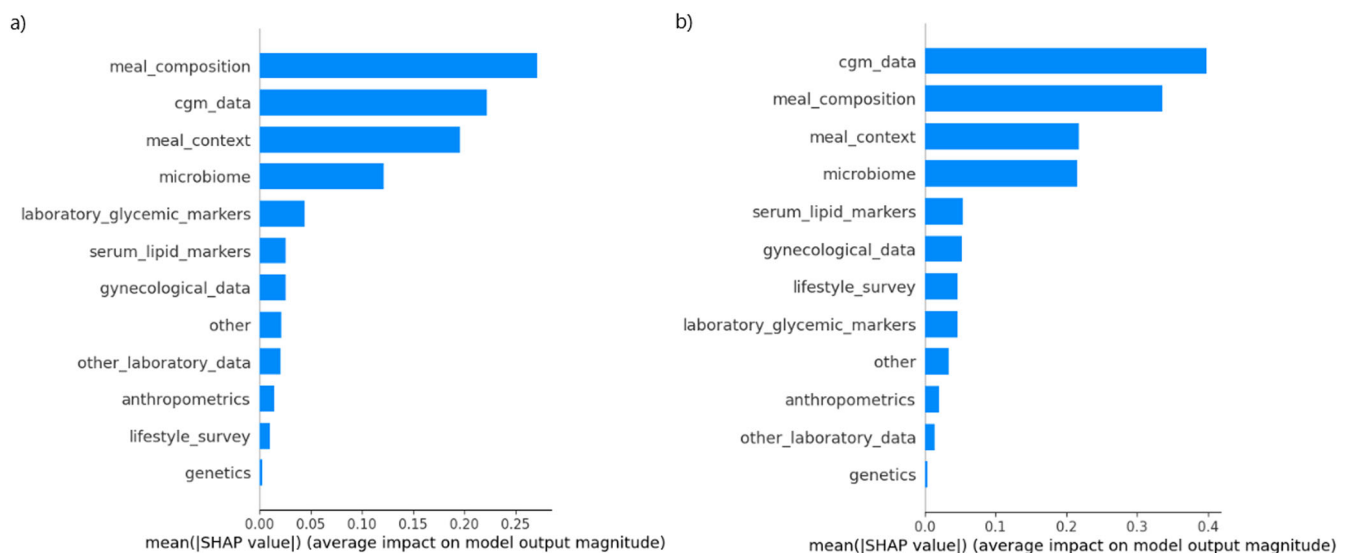


Fig. 7 | Significance level of the groups of features for the prediction. SHAP values (linear scale, absolute values) of the groups of features for the prediction of iAUC120, mmol/L·h (a) and GLUmax, mmol/L (b). The groups of features are presented as follows: «meal composition» includes the nutritional content of the meal, «cgm_data» includes glucose values obtained from CGM devices; «meal_context»

includes the nutritional content of meals consumed up to 12 h prior to the index meal; «Microbiome» includes RA of bacteria detected from stool samples; «genetics» includes rs10830963 and rs1387153 variants in MTNR1B gene. The full description of the parameters included in each feature group is listed in Supplementary Table 1.

without implementing microbiome data¹¹, our current model exhibited only a slight increase of R (0.72 vs 0.7). However, for the previous model, we had a larger dataset (3240 records of meals and corresponding PPGRs) making direct comparisons inappropriate.

Concerning the gut microbiome, our study identified bacterial taxa differentially enriched in pregnant women with higher and lower mean PPGR. According to LefSe analyses, most taxa belong to Lachnospiraceae and Oscillospiraceae, and some families of Bacteroidia. These families are represented by the most functionally active bacteria involved in dietary fiber degradation and short-chain fatty acid (SCFA) biosynthesis¹⁸. SCFAs, especially butyrate, are generally considered beneficial metabolites that reduce the risk of GDM. However, excess SCFAs can activate gluconeogenesis, leading to hyperglycemia and insulin resistance¹⁹.

Even though, in general, Lachnospiraceae and Oscillospiraceae are considered useful symbionts that interact beneficially with the host, among them, some taxa carry a “dual” function. For example, *Anaerobutyricum hallii* and some *Blautia* species, are considered pathobionts that can cause harm to the host¹⁹. Further, among the taxa with higher abundance in women with higher iAUC120 or GLUmax, several are of interest for discussion as taxa potentially contributing to GDM pathogenesis. *Prevotella 9* is now characterized as the new genus *Segatella* with the type species *Segatella copri* (basonym: *Prevotella copri*). Although *S. copri* is considered to be associated with health²⁰, a significant positive association between increases in *Prevotella 9* and higher GDM risk was identified²¹, and an increased abundance of *Prevotella* was reported in GDM patients²². *Coprococcus comes* is a butyrate producer¹⁸, usually considered beneficial. However, in the FINRISK-2002 cohort, the strongest association with higher statin-associated new-onset type 2 diabetes risk was observed for *C. comes*²³, which aligns with our results. A possible explanation for this may be the ability of *C. comes* to produce the highest butyrate levels¹⁸, which can lead to its excess. *A. hallii* (basonym: *Eubacterium hallii*) is also associated with health²⁰. However, GDM patients who failed to control glycemic levels were characterized by increased *A. hallii*²⁴, which corresponds to the results of our study.

In women with lower iAUC120 or GLUmax, some taxa also had higher relative abundance, conversely suggesting a protective effect against higher PPGR. *Oscillospiraceae* UCG-002, previously *Ruminococcaceae* UCG-002, was more abundant in the normal glucose tolerance group than in GDM. Previous research found it was reduced in early pregnancy in women with subsequent GDM and was negatively correlated with fasting blood glucose levels²⁵. *Oscillospiraceae* UCG-002 was also negatively associated with the homeostasis model assessment of insulin resistance (HOMA-IR) index and served as a marker of intestinal phytoestrogen enterolactone production²⁶.

Christensenellaceae R-7 group is a beneficial genus: elevated abundance was associated with reduced visceral adipose tissue and a healthier metabolic profile^{27,28}. *Parabacteroides distasonis* may protect against inflammation and obesity; however, increased abundance of *P. distasonis* was previously reported in GDM²². *Sellimonas* is an acetate producer, associated with a reduced type 2 diabetes risk²⁹ and has been linked to low polycystic ovary syndrome (PCOS) risk³⁰.

Eisenbergiella tayi produces butyrate, lactate, acetate, and succinate and is thought to be potentially beneficial. However, *E. tayi* was associated with the disease state²⁰. Women who developed GDM showed a significantly higher abundance of *Eisenbergiella* in early pregnancy, and *Eisenbergiella* was also positively correlated with fasting blood glucose levels²⁵, which contradicts our results. *Bilophila wadsworthia* is associated with the metabolism of fatty acid esters of hydroxy fatty acids, which improves glucose homeostasis, stimulates insulin sensitivity, and has anti-inflammatory effects³¹.

After including bacterial taxa distinguished by LefSe as input variables for the creation of PPGR prediction models, microbiome features were categorized as either advantageous or disadvantageous. As the RA of these taxa increased, the algorithm projected a decrease or increase in postprandial glucose response, respectively. Among bacterial features, the greatest contribution to iAUC120 prediction was made by the RAs of *I.*

bartlettii, *B. faecihominis*, and “*L. edouardi*”. The most impactful bacterial features for the prediction of GLUmax were *I. bartlettii*, *R. champanellensis*, and *L. eligens*. A higher abundance of these bacteria was associated with lower PPGRs.

Notably, *I. bartlettii* was selected by the algorithm both for the prediction of iAUC120 and GLUmax among the top 20 parameters. *I. bartlettii* can produce indoleacetic and phenylacetic acids, acetate, isovalerate, and isobutyrate. Due to the latter’s production, *Intestinibacter* might be beneficial to host lipid and glucose metabolism and intestinal barrier integrity, which may explain the inverse association of *Intestinibacter* with diabetes³².

L. eligens produces butyrate, acetate, and lactate, and promotes the production of the anti-inflammatory cytokine IL-10. *L. eligens* was reduced in early pregnancy in women with subsequent GDM²⁵. The abundance of *L. eligens* was significantly higher in the healthy controls than in the obese individuals³³. Additionally, *L. eligens* was positively associated with adherence to a Mediterranean diet³⁴. *B. faecihominis* is a butyrate producer and was included in the stool-derived microbial ecosystem therapeutics to combat *Clostridioides difficile* infection as a beneficial bacterium³⁵. *R. champanellensis* is a cellulose-degrading bacterium³⁶. The strongest association with lower statin-associated new-onset type 2 diabetes risk was observed for *R. champanellensis*²³. “*L. edouardi*” was associated with an increased risk of GDM²¹ and heightened type 2 diabetes risk²⁹.

The limitation of our study is a relatively small sample size. Further studies in other cohorts and populations of pregnant women are needed to confirm our findings concerning certain bacterial taxa associated with PPGRs. Another limitation of our study is the inability to confidently classify taxa at the species level which is due to sequencing only part of the 16S rRNA gene. Additionally, the participants of our study were women with GDM treated with diet only and women with normal glucose tolerance during pregnancy. Thus, our model may not be representative of pregnant women with GDM treated with insulin or other pharmacological glucose-lowering agents. Further studies are needed to explore PPGR prediction models for insulin-treated women with GDM.

There is further room for improvement, such as conducting more comprehensive assessments of contextual factors than those employed in the current study. For example, including data on physical activity preceding meals and integrating extensive ‘omics’ data could improve the predictive capacity of these algorithms.

It is essential to delve deeper into understanding the functional roles of bacterial taxa that were the most influential for PPGR prediction in our study. Additionally, circulating SCFA levels should be measured in future studies as this could have conveyed an insight into whether the SCFA actually are the important metabolites for the regulation of blood glucose levels or whether its regulation is modified by other bacterially associated metabolites. The insights gained from this data could pave the way for the future advancement of probiotic or autoprobiotic therapies aimed at enhancing glycemic regulation. Probiotics with metabolic effects that target functionally active bacteria, predominantly belonging to Clostridia and Bacteroidia, which play a key role in maintaining the balance of the intestinal microbiota, seem promising³⁷.

Our study highlights the role of the gut microbiota in the interpersonal variability of PPGRs. While previous research extensively utilized microbiota for PPGR prediction in healthy individuals and those with type 1 diabetes, our study fills an important gap by examining its impact on PPGRs in pregnant women, particularly those with GDM.

Our findings indicate that microbiome features rank among the top parameters influencing PPGR prediction in pregnant women with GDM, although the addition of the microbiome has a relatively modest impact on the performance of the prognostic models, especially for iAUC120. Specifically, certain bacterial taxa were identified as significantly associated with variations in PPGRs, highlighting the potential of microbiota-based interventions for optimizing glycemic control which should be validated in future interventional studies.

Research design and methods

Study design

We recruited pregnant women who participated in the randomized controlled trial “Genetic and Epigenetic Mechanisms of Developing Gestational Diabetes Mellitus and its Effects on the Fetus” (GEM-GDM), and consented to be connected to a continuous glucose monitoring (CGM) system (CGMS) for at least 7 days, tracked information on food consumption in a designated mobile app, and provided stool samples. The parent GEM-GDM study aimed to compare different glycemic targets for women with GDM. It was registered at ClinicalTrials.gov (Identifier: NCT03610178), and its design is described elsewhere³⁸. Briefly, gravidas with GDM were randomly assigned into two groups according to their glycemic goals: the first group had strict glycemic goals (<5.1 mmol/L for fasting blood glucose (BG) and <7.0 mmol/L for 1-h postprandial BG), and the second group had less strict glycemic goals (<5.3 mmol/L and <7.8 mmol/L, respectively). For this study, the women from both groups were combined to create the GDM group. GDM was diagnosed using a single-step 75-g OGTT according to the recommendation of the International Association of Diabetes and Pregnancy Study Groups (IADPSG)³⁹.

Apart from women with GDM, we also included healthy pregnant women with normal values of plasma glucose during oral glucose tolerance test (OGTT) (controls).

At study initiation, a physician acquired informed consent, recorded medical history and took anthropometric measurements (weight, height, waist circumference, blood pressure and heart rate). Pre gestational body mass index (BMI) was calculated by dividing self-reported pre gestational weight (in kilograms) by the square of height (in meters). Blood tests, including fasting plasma glucose, lipid profile and HbA1c, were performed in the Almazov National Medical Research Centre laboratory. Participants filled out questionnaires concerning their lifestyle before and during pregnancy and were then connected to CGMS for 7–14 days during which they tracked information on meal consumption in the proprietary mobile app DiaCompanion, as described elsewhere^{40,41}. The CGM and meal-related data were processed using a previously described algorithm⁴¹. The study was approved by the local ethics committee of the Almazov National Medical Research Centre, Russia (protocol no. 119).

Inclusion and exclusion criteria

Participation in this study was optional for participants of the GEM-GDM trial.

The GDM group included pregnant women with GDM and gestational age of ≥ 24 weeks at the start of CGM. The control group comprised pregnant women with normal glucose tolerance, confirmed by OGTT between 24 and 32 weeks of gestation.

In addition to the inclusion criteria used in the GEM-GDM study, for this particular study, consent to be connected to a CGMS for 7 days and a capability to work with a mobile phone app for the recording of dietary intake in real time were required.

Exclusion criteria included an active inflammatory or neoplastic disease, any known medical condition affecting glucose metabolism (with the exception of GDM), current use of insulin and other medications that could affect glucose metabolism, antibiotic usage 2 months prior to participation in the study, failure to provide a stool sample and submission of inaccurate food diaries through the app. Taking into account that accurate logging is crucial for analysis of PPGRs to food, a set of rules was formulated by the authors to filter negligently filled-in and misreported diaries: (1) more than 50% of the logged meals comprised of a single dish or a single dish with a single beverage; (2) the average amount of logged calories per day was less than 1000 kcal; (3) more than 50% of the logged weights of food items were rounded to the hundreds (excluding beverages); (4) the amount of logged snacks was less than 10% of all meal records¹¹. Participants with misreported diaries were excluded.

Lifestyle questionnaire

The questionnaire comprised several sections covering various aspects: frequency of consuming staple items per week (such as fruits, pastries, skimmed dairy products, legumes, meat, sausage products, dried fruits, fish, whole-grain bread, sauces, vegetables, alcohol, sweet drinks, and coffee), levels of physical activity (daily walking duration categorized as <30 min/day, 30–60 min/day, or >60 min/day; daily frequency of stair climbing categorized as <4 flights/day, 4–16 flights/day, or >16 flights/day; frequency of engaging in sports activities categorized as <2 days/week, 2–3 days/week, or >3 days/week), and smoking habits before and during pregnancy. Each section of the questionnaire was structured in a semi-quantitative manner. This questionnaire has been previously documented⁴¹. For the description of the parameters from the lifestyle questionnaire included in the final dataset, please refer to Supplementary Table S1.

Blood samples were collected by a certified nurse after 8–12 h of fasting. The blood panel included measurement of glycosylated hemoglobin (HbA1c%), plasma glucose, total and HDL cholesterol, triglycerides, insulin, and leptin, and fructosamine levels in the central lab of the Almazov National Medical Research Centre. Plasma glucose concentration was determined by the glucose oxidase method in fresh plasma samples. HbA1c was measured in fresh whole blood samples using high-performance liquid chromatography (HPLC) (D10 HbA1c). Blood for genotyping of pregnant women and serum for other biochemical analysis were stored at -80°C until the analysis. Serum fasting insulin levels were measured using the electrochemiluminescence immunoassay (Roche Diagnostics, GmbH, Germany). The homeostatic model assessment (HOMA) index was calculated using the following formula: fasting serum insulin (m IU/L) $\times$ fasting plasma glucose (mmol/L)/(22.5) as an insulin resistance indicator. Total cholesterol, HDL-C, LDL-C, VLDL-C, and triglyceride levels were measured utilizing enzymatic colorimetric methods with diagnostic reagent system designed for the Cobas Integra Autoanalyzer. Fructosamine was measured in the serum using the colorimetric nitro-blue tetrazolium assay (Roche Diagnostics Corporation, Indianapolis, IN) and the Roche cobas c 311 instrument. Serum leptin levels were measured using an enzyme-linked immunosorbent assay (ELISA) as recommended by the manufacturer (Diagnostics Biochem Canada Inc., Canada).

Continuous glucose monitoring (CGM) was conducted using the iPro2TM system from Medtronic, MN, USA. This system utilizes EnliteTM sensors placed subcutaneously to measure interstitial glucose levels. To align CGM readings with blood glucose levels, participants also utilized finger-prick measurements with the Accu Chek Performa from Roche, Germany. Participants were instructed to perform four daily blood glucose measurements. To enhance accuracy, participants were specifically asked to measure blood glucose levels before meals, following recommendations outlined in the iPro2 manual. Calibration of CGM measurements was performed using the CareLink online software from Medtronic, following the guidelines provided in the iPro2 manual.

Food diary tracking was facilitated through our proprietary mobile app DiaCompanion. Each consumed food item was recorded by selecting it from a database created by the authors on the basis of reference books of the Russian Academy of Medical Sciences and the US Department of Agriculture (USDA) Food Composition Databases (Release 28) with the expansion of additional items by certified dietitians. The distinctive feature of this food database, in addition to a wide selection of foods (more than 5500 items), is the presence of glycemic index (GI). Each food item in the database was assigned a dietary GI⁴¹. Participants were instructed to meticulously log their daily activities using this platform. They were required to document precise details, including the components and weights of each meal, sleep and wake-up times. Participants were informed of the importance of accurate logging, particularly emphasizing the correct timing of meal logging and accurate recording of food components. Research physicians conducted weekly reviews of each participant’s loggings. Any uncertainties in the logs were addressed directly with the participants.

Meal preprocessing

Before a meal and corresponding PPGR were added to the dataset for model training, the following filters were implemented (mainly to exclude recordings with incorrect timing): (1) a meal followed by a subsequent meal less than 60 min after its start; (2) a meal on the peak of a CGM-curve: an increase in glucose levels by more than 1 mmol/L during an hour preceding the index meal; (3) a meal on the falling edge of a CGM peak; and (4) a meal with inadequately low PPGR ($iAUC_{120} \leq 0.3$ mmol/L·h) to a considerable amount of carbohydrates (>40 g)¹¹. Glucose level at baseline was considered as the lowest glucose value within ± 15 min from self-reporting of the meal in the app.

DNA and genotyping of blood samples

Genomic DNA was extracted from blood samples using the FlexiGene DNA Kit from Qiagen (Hilden, Germany). Genotyping of the following variants: HKDC1 (rs10762264), MTNR1B (rs10830963 and rs1387153), GCK (rs1799884), KCNJ11 (rs5219), IGF2BP2 (rs4402960), TCF7L2 (rs7903146), CDKAL1 (rs7754840), FTO (rs9939609), and IRS1 (rs1801278), was conducted through real-time PCR utilizing custom kits from Applied Biosystems, based in the USA. The procedures recommended by the manufacturer were followed meticulously. Each primer tube contained a concentrated mixture of SNP Genotyping Assay Mix, comprising polymorphism-specific direct and reverse primers, along with two TaqMan MGB probes: one tagged with VIC dye for allele 1 identification and the other tagged with FAM dye for allele 2 identification. Following replication of 10% of the samples, the discordance rate was determined to be less than 0.1%.

Microbiome: DNA extraction

DNA was extracted from all collected samples using the PowerSoil DNA Isolation Kit (MO BIO, Carlsbad, CA, USA) according to the manufacturer's instructions and following a 2 min bead beating step (BioSpec, Bartlesville, OK, USA). Next, the variable V4 region was PCR-amplified using the 515F and 806R barcoded primers following the Earth Microbiome Project protocol⁴². Each PCR reaction contained 25 μ L with ~ 40 ng/ μ L of DNA, 2 μ L 515F (forward, 10 μ M) primer, 2 μ L 806R (reverse, 10 μ M) primer, and 25 μ L PrimeSTAR Max PCR Readymix (Takara, Mountain View, CA, USA). PCR conditions were as follows: 30 cycles of denaturation at 98 °C for 10 s, annealing at 55 °C for 5 s, and extension at 72 °C for 20 s, followed by a final elongation at 72 °C for 1 min. Amplicons were purified using AMPure magnetic beads (Beckman Coulter, Indianapolis, IN, USA) and quantified using the Picogreen dsDNA quantitation kit (ThermoFisher, Waltham, MA, USA). Equimolar amounts of DNA from individual samples were pooled and sequenced using the Illumina MiSeq platform at the Genomic Center at the Bar-Ilan University, Azrieli Faculty of Medicine. Appropriate negative and positive controls were included at all stages of analysis.

Bioinformatics and microbiome analysis

The quality of raw reads was assessed with FastQC v. 0.11.9⁴³ and MultiQC v. 1.14⁴⁴. Reads were trimmed and filtered with Trimmomatic v. 0.39⁴⁵ (SE -phred 33 HEADCROP 31 ILLUMINACLIP:2:30:10 SLIDINGWINDOW:4:15 MINLEN:150). The remaining reads were processed with the DADA2 pipeline v. 3.6.2⁴⁶, including additional trimming, denoising, and error correction. The derived sequences - amplicon sequence variants (ASVs) were clustered using MMseqs2 v. 13.45111⁴⁷ (identity 99%, coverage 80%). The resultant representative sequences were treated as operative taxonomic units (OTUs). We clustered ASVs to OTUs to reduce the number of sequencing errors inherent in ASVs and avoid false diversity. The OTUs were returned to DADA2 for taxonomy assignment with SILVA SSU database v.138.1⁴⁸. Sequences classified as eukaryotes were removed. Only samples containing more than 10,000 reads were used for downstream analysis. The bioinformatics analysis was conducted using R packages. Permutational multivariate analysis of variance (PERMANOVA) was performed with vegan v2.6.4⁴⁹. PCoA (Principal coordinates analysis) and

alpha-diversity were performed with phyloseq v1.42.0⁵⁰ and ggplot2 v. 3.3.6 (<https://github.com/tidyverse/ggplot2>). Linear discriminant analysis Effect Size (LefSe) was conducted using microbiomeMarker v1.4.0⁵¹ with default parameters.

Models for the prediction of postprandial glucose response

We used two measures of PPGR characteristics: $iAUC_{120}$ and the peak glucose level within 120 min after the meal start (GLU_{max}, mmol/L). The latter indicator was chosen because the recommended timing of blood glucose self-monitoring for pregnant women is established in the time interval when glycemic levels are highest. The peak glycemic level in pregnant women with diabetes mellitus is reached 45–75 min after a meal, which is the reason for the recommendation to measure glycemia 1 h after a meal⁵². However, the peak blood glucose (BG) level is less sensitive to inaccurate logging of meal start time than 1-hr postprandial BG.

Outliers in the target variable were removed using Tukey's Interquartile Range method⁵³ resulting in the final dataset of 2633 meals with PPGRs for GLU_{max} prediction model and 2628 meals for $iAUC_{120}$ prediction.

We used the gradient boosting algorithm LightGBM⁵⁴ to predict both indices and improved its performance with Optuna hyperparameter optimization⁵⁵.

Within Optuna, we adopted the Tree-structured Parzen Estimator (TPESampler) for sampling within the hyperparameter space, and the Asynchronous Successive Halving Algorithm (SuccessiveHalvingPruner) was implemented to eliminate underperforming trials efficiently. The optimal configuration of hyperparameters that emerged from our analysis included a "num_boost_round" of 4700, a "learning_rate" of 0.0015, "max_depth" of 11, "num_leaves" of 30, "min_sum_hessian_in_leaf" at 0.12, "bagging_fraction" of 0.55, "bagging_freq" of 10, "feature_fraction" of 0.4, "lambda_l1" at 0.006 and "lambda_l2" at 0.007. The data was divided into training and test sets with a ratio of 70:30, ensuring that records from the same patient were only included in one set to prevent data leakage and potential bias in performance metrics. To enhance the robustness of the training process, we employed 3-fold cross-validation.

The final metrics, including mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), Pearson's correlation coefficient (R), and the coefficient of determination (R^2), were calculated on the held-out test sample in the Python scikit-learn library.

Other statistical analyses. To describe the patients in our data frame, we employed the bootstrap hypothesis test⁵⁶. The main idea behind this method is to repeatedly draw random subsamples of the data with replacement in order to estimate the distribution of the test statistic and make decisions about the significance of differences. The bootstrap test does not require any assumptions about the underlying distribution of the original data, making it more robust compared to parametric tests such as the t -test.

Feature selection and evaluation of input parameters

We selected 164 features as model inputs, including features characterizing meal content, anthropometric measures, gynecological data, blood test results, CGM-derived features, lifestyle questionnaire data, and genetic and microbiome features (Supplementary Table S1).

In order to avoid model overfitting, we used several approaches to decrease the number of input variables. From the original lifestyle questionnaire characterizing the consumption of certain product groups and physical activity, described elsewhere^{11,57}, we selected the parameters with significant Spearman correlations with $iAUC_{120}$ and/or GLU_{max} (Supplementary Table S2).

Microbial features were selected based on the results of LefSe analysis. For this purpose, all participants were divided into two groups based on the average levels of PPGR indices: group 1—below the median and group 2—equal or above the median for the group. The relative abundances (RA) of

bacterial taxa differentially enriched in these groups were used as input variables.

Among genetic factors, we selected rs10830963 and rs1387153 variants in *MTNR1B* previously shown by our group to be associated with the results of OGTT in pregnant women⁵⁷.

SHAP methods were utilized for enhancing model interpretability¹². SHAP values were computed in two ways. First, calculations were made individually for each feature to denote the average alteration in the model's output when conditioning on that specific feature. Second, the additive nature of SHAP values was employed to assess the impact of various feature groups on the model.

Data availability

The datasets generated and analyzed in the current study are available in a GitHub repository <https://github.com/artemisak/MicrobesAndGlucoseAnalysis?tab=readme-ov-file>.

Code availability

The code generated in the current study is available in a GitHub repository <https://github.com/artemisak/MicrobesAndGlucoseAnalysis?tab=readme-ov-file>.

Received: 2 August 2024; Accepted: 2 January 2025;

Published online: 07 February 2025

References

- Sacks, D. A. et al. Frequency of gestational diabetes mellitus at collaborating centers based on IADPSG consensus panel-recommended criteria. *Diab. Care* **35**, 526–528 (2012).
- HAPO Study Cooperative Research Group. Hyperglycemia and adverse pregnancy outcomes. *N. Engl. J. Med.* **358**, 1991–2002 (2008).
- Hajj, N. E., Schneider, E., Lehnen, H. & Haaf, T. Epigenetics and life-long consequences of an adverse nutritional and diabetic intrauterine environment. *Reproduction* **148**, R111–R120 (2014).
- Popova, P., Castorino, K., Grineva, E. N. & Kerr, D. Gestational diabetes mellitus diagnosis and treatment goals: measurement and measures. *Minerva Endocrinol.* **41**, 421–432 (2016).
- American Diabetes Association Professional Practice Committee. 15. Management of diabetes in pregnancy: standards of care in diabetes —2024. *Diab. Care* **47**, S282–S294 (2024).
- Koning, S. H. et al. Neonatal and obstetric outcomes in diet- and insulin-treated women with gestational diabetes mellitus: a retrospective study. *BMC Endocr. Disord.* **16**, 52 (2016).
- Zeevi, D. et al. Personalized nutrition by prediction of glycemic responses. *Cell* **163**, 1079–1094 (2015).
- Berry, S. E. et al. Human postprandial responses to food and potential for precision nutrition. *Nat. Med.* **26**, 964–973 (2020).
- Shilo, S. et al. Prediction of personal glycemic responses to food for individuals with type 1 diabetes through integration of clinical and microbial data. *Diab. Care* **45**, 502–511 (2022).
- Shilo, S. et al. The gut microbiome of adults with type 1 diabetes and its association with the host glycemic control. *Diab. Care* **45**, 555–563 (2022).
- Pustozero, E. A. et al. Machine learning approach for postprandial blood glucose prediction in gestational diabetes mellitus. *IEEE Access* **8**, 219308–219321 (2020).
- Lundberg, S. M. et al. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2**, 56–67 (2020).
- Nuriel-Ohayon, M., Neuman, H. & Koren, O. Microbial changes during pregnancy, birth, and infancy. *Front. Microbiol.* **7**, 1031 (2016).
- Koren, O., Konnikova, L., Brodin, P., Mysorekar, I. U. & Collado, M. C. The maternal gut microbiome in pregnancy: implications for the developing immune system. *Nat. Rev. Gastroenterol. Hepatol.* **21**, 35–45 (2024).
- Nuriel-Ohayon, M. et al. Progesterone increases *Bifidobacterium* relative abundance during late pregnancy. *Cell Rep.* **27**, 730–736.e3 (2019).
- Koren, O. et al. Host remodeling of the gut microbiome and metabolic changes during pregnancy. *Cell* **150**, 470–480 (2012).
- Frishman, S. et al. Positive effects of diet-induced microbiome modification on GDM in mice following human faecal transfer. *Gut* **73**, e17 (2024).
- Abdugheni, R. et al. Metabolite profiling of human-originated Lachnospiraceae at the strain level. *iMeta* **1**, e58 (2022).
- Hu, R. et al. Gut microbiota and critical metabolites: potential target in preventing gestational diabetes mellitus? *Microorganisms* **11**, 1725 (2023).
- Giliberti, R., Cavaliere, S., Mauriello, I. E., Ercolini, D. & Pasolli, E. Host phenotype classification from human microbiome data is mainly driven by the presence of microbial taxa. *PLoS Comput. Biol.* **18**, e1010066 (2022).
- Wu, X. et al. Investigating causal associations among gut microbiota, gut microbiota-derived metabolites, and gestational diabetes mellitus: a bidirectional Mendelian randomization study. *Aging* **15**, 8345–8366 (2023).
- Ponzo, V. et al. Diet-gut microbiota interactions and gestational diabetes mellitus (GDM). *Nutrients* **11**, 330 (2019).
- Koponen, K. et al. Role of gut microbiota in statin-associated new-onset diabetes—a cross-sectional and prospective analysis of the FINRISK 2002 cohort. *Arterioscler. Thromb. Vasc. Biol.* **44**, 477–487 (2024).
- Ye, G. et al. The gut microbiota in women suffering from gestational diabetes mellitus with the failure of glycemic control by lifestyle modification. *J. Diab. Res.* **2019**, 6081248 (2019).
- Ma, S. et al. Alterations in gut microbiota of gestational diabetes patients during the first trimester of pregnancy. *Front. Cell Infect. Microbiol.* **10**, 58 (2020).
- Atzeni, A. et al. Taxonomic and functional fecal microbiota signatures associated with insulin resistance in non-diabetic subjects with overweight/obesity within the frame of the PREDIMED-Plus Study. *Front. Endocrinol.* **13**, 804455 (2022).
- Tavella, T. et al. Elevated gut microbiome abundance of Christensenellaceae, Porphyromonadaceae and Rikenellaceae is associated with reduced visceral adipose tissue and healthier metabolic profile in Italian elderly. *Gut Microbes* **13**, 1–19 (2021).
- Alcazar, M. et al. Gut microbiota is associated with metabolic health in children with obesity. *Clin. Nutr.* **41**, 1680–1688 (2022).
- Song, S., Zhang, Q., Zhang, L., Zhou, X. & Yu, J. A two-sample bidirectional Mendelian randomization analysis investigates associations between gut microbiota and type 2 diabetes mellitus. *Front. Endocrinol.* **15**, 1313651 (2024).
- Liang, Y. et al. Gut microbiome and reproductive endocrine diseases: a Mendelian randomization study. *Front. Endocrinol.* **14**, 1164186 (2023).
- Folz, J. et al. Human metabolome variation along the upper intestinal tract. *Nat. Metab.* **5**, 777–788 (2023).
- Luo, K. et al. Metabolic and inflammatory perturbation of diabetes associated gut dysbiosis in people living with and without HIV infection. *Genome Med.* **16**, 59 (2024).
- Hu, X. et al. Integrative metagenomic analysis reveals distinct gut microbial signatures related to obesity. *BMC Microbiol.* **24**, 119 (2024).
- Ghosh, T. S. et al. Mediterranean diet intervention alters the gut microbiome in older people reducing frailty and improving health status: the NU-AGE 1-year dietary intervention across five European countries. *Gut* **69**, 1218–1228 (2020).
- Carlucci, C. et al. Effects of defined gut microbial ecosystem components on virulence determinants of *Clostridioides difficile*. *Sci. Rep.* **9**, 885 (2019).

36. Morais, S. et al. Cryptic diversity of cellulose-degrading gut bacteria in industrialized humans. *Science* **383**, eadj9223 (2024).
37. Pokrotnieks J. & Sitkin S. He who controls Clostridia and Bacteroidia controls the gut microbiome: the concept of targeted probiotics to restore the balance of keystone taxa in irritable bowel syndrome. *Neurogastroenterol. Motil.* **36**, e14805 (2024).
38. Popova, P. et al. A randomised, controlled study of different glycaemic targets during gestational diabetes treatment: effect on the level of adipokines in cord blood and ANGPTL4 expression in human umbilical vein endothelial cells. *Int. J. Endocrinol.* **2018**, 6481658 (2018).
39. International Association of Diabetes and Pregnancy Study Groups Consensus Panel. International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. *Diab. Care* **33**, 676–682 (2010).
40. Pustozarov, E. et al. Development and evaluation of a mobile personalized blood glucose prediction system for patients with gestational diabetes mellitus. *JMIR Mhealth Uhealth* **6**, e6 (2018).
41. Pustozarov, E. et al. The role of glycemic index and glycemic load in the development of real-time postprandial glycemic response prediction models for patients with gestational diabetes. *Nutrients* **12**, 302 (2020).
42. Caporaso, J. G. et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6**, 1621–1624 (2012).
43. Andrews, S. *FastQC: A Quality Control Tool for High Throughput Sequence Data*. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (2010).
44. Ewels, P. et al. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
45. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
46. Callahan, B. J. et al. “DADA2: high-resolution sample inference from Illumina amplicon data.” *Nat. Methods* **13**, 581–583 (2016).
47. Mirdita, M., Steinegger, M., Breitwieser, F., Söding, J. & Levy Karin, E. Fast and sensitive taxonomic assignment to metagenomic contigs. *Bioinformatics* **37**, 3029–3031 (2021).
48. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2012).
49. Oksanen, F. J. et al. Vegan: community ecology package. R package version 2.4-3. <https://CRAN.R-project.org/package=vegan> (2017).
50. McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217 (2013).
51. Yang Cao et al. MicrobiomeMarker: an R/Bioconductor package for microbiome marker identification and visualization. *Bioinformatics* **38**, 4027–4029 (2022).
52. Bühling, K. J. et al. Optimal timing for postprandial glucose measurement in pregnant women with diabetes and a non-diabetic pregnant population evaluated by the Continuous Glucose Monitoring System (CGMS). *J. Perinat. Med.* **33**, 125–131 (2005).
53. Tukey, J. W. *Exploratory Data Analysis* (Addison-Wesley, 1977).
54. Ke, G. et al. LightGBM: a highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **30**, 3146–3154 (2017).
55. Akiba, T., Sano, S., Yanase, T., Ohta, T. & Koyama, M. Optuna: a next-generation hyperparameter optimization framework. In *Proc. 25th ACM SIGKDD International Conference on Knowledge Discovery & Data (KDD '19)*, 2623–2631 (2019).
56. Efron, B. Bootstrap methods: another look at the jackknife. *Ann. Stat.* **7**, 1–26 (1979).
57. Popova, P. V. et al. Effect of gene-lifestyle interaction on gestational diabetes risk. *Oncotarget* **8**, 112024–112035 (2017).

Acknowledgements

This work was financially supported by the Ministry of Science and Higher Education of the Russian Federation (Agreement No. 075-15-2022-301) and the Ministry of Innovation, Science & Technology, Israel. The funders played no role in the study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

Author contributions

P.P., E.P., E.G., O.K. and E.S. were involved in the conception, design, and conduct of the study and the analysis and interpretation of the results. A.I., A.R., and S.S. designed and conducted the analyses, interpreted the results, and wrote the manuscript. A.A., E.V., A.T., I.N., Ek.S., A.E., El.S. and T.P. provided data and interpreted the results. M.K. and E.V. performed laboratory analysis and interpreted the results. S.Z., E.R., C.E., and S.T. performed fecal specimen processing, sample sequencing and interpreted the results. L.V. performed blood specimen processing, genotyping assays and interpreted the results. P.P., A.I., A.R., and S.S. wrote the first draft of the manuscript, and all authors edited, reviewed, and approved the final version of the manuscript. P.P. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41522-025-00650-9>.

Correspondence and requests for materials should be addressed to Polina V. Popova.

Reprints and permissions information is available at <http://www.nature.com/reprints>**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025