

Breath biomarkers of insulin resistance in pre-diabetic Hispanic adolescents with obesity

Running title: Breath biomarkers of insulin resistance

*Mohammad S. Khan,^{1,2} PhD, Suzanna Cuda^{3,4} MD, Genesis M. Karere^{1,2} PhD, Laura
Cox^{1,2} PhD, and Andrew C. Bishop^{1,2} PhD**

¹Department of Internal Medicine, Section on Molecular Medicine, Wake Forest School of
Medicine, Winston Salem, NC 27157

²Center for Precision Medicine. Wake Forest School of Medicine, Winston Salem, NC 27157

³Health and Weight Management Clinic, Children's Hospital of San Antonio, San Antonio, TX
78207

⁴Baylor College of Medicine, Houston, TX 77030

*Corresponding author

Andrew C. Bishop, Ph.D.

Department of Internal Medicine, Section on Molecular Medicine

Wake Forest School of Medicine, Winston-Salem, NC

E: abishop@wakehealth.edu

O: 336-713-7148

ABSTRACT

Background: Insulin Resistance (IR) affects a quarter of the world's adult population and is a major factor in the pathogenesis of cardio-metabolic disease. Non-invasive sampling of exhaled breath contains metabolic markers indicative of underlying systemic metabolic abnormality.

Method: In this pilot study, we implemented a non-invasive breathomics approach, combined with random forest machine learning, to investigate metabolic markers from pre-diabetic Hispanic adolescents with obesity as indicators of abnormal metabolic regulation.

Findings: Exhaled breath collection using the ReCIVA breathalyzer is feasible in an adolescent population. We have identified a signature of breath metabolites (breath-IR model) which correlates with Homeostatic Model Assessment of Insulin Resistance (HOMA-IR) ($R = 0.95, p < 0.001$). A strong correlation was also observed between the breath-IR model and the blood glycemic profile (fasting insulin $R=0.91, p < 0.001$ and fasting glucose $R=0.80, p < 0.001$). Among tentatively identified metabolites, limonene, undecane, and 2,7-dimethyl-undecane, significantly cluster individuals based on HOMA-IR ($p = 0.003, p = 0.002, \text{ and } p < 0.001$, respectively). Our breath-IR model differentiates between adolescents with and without IR with an area under the receiver operating characteristic curve of 0.87, after cross-validation.

Interpretation: Identification of a breath metabolite signature indicative of IR in prediabetic Hispanic adolescents with obesity provides evidence of the utility of exhaled breath metabolomics for assessing systemic metabolic dysregulation. A simple and non-invasive breath-based test has utility as a diagnostic tool for monitoring IR progression, potentially allowing for

earlier detection of IR and implementation of early interventions to prevent onset of type 2 diabetes mellitus.

Funding: This study was funded by The Healthy Babies Project, Texas Biomedical Research Institute, San Antonio, TX.

Keywords: insulin resistance; breath volatiles; homa-ir; metabolic biomarker

1. INTRODUCTION

An estimated 1.26 billion adults worldwide have insulin resistance (IR), a complex pathophysiological state connected to an imbalance between insulin and glucose metabolism, which is strongly associated with the development of cardio-metabolic disease¹⁻³. Although aging is considered a strong predictor of IR, a meta-analysis of 18 population-based studies from 13 countries indicated that approximately 312 million children globally have IR⁴, suggesting that other factors also contribute to IR development earlier in life. One major contributing factor is obesity and in the United States between 2015 and 2016, 18.5% of youth between the ages of 2-19 were obese^{5,6}. Severe obesity is also racially and ethnically disproportionate with two-to-four fold higher rates among African American or Hispanic populations compared to their Caucasian counterparts⁶, therefore more research is needed to investigate these populations to slow or prevent long term health consequences.

The exact pathophysiology of IR is unclear but involves multiple associations of cellular dysfunction. IR is inversely related to the insulin sensitivity in insulin-dependent cells such as skeletal muscle, adipocytes, and cardiomyocytes⁷. When insulin sensitivity is low, these cells fail to respond to insulin signaling transduction, which is required for glucose uptake⁸. As a result, IR is a common feature in multiple metabolic disorders including obesity, dyslipidemia, hypertension, atherosclerosis, nonalcoholic fatty liver disease (NAFLD), type 2 diabetes mellitus (T2DM), and some cases of type 1 diabetes mellitus (T1DM)^{1,8,9}. One potential candidate promoting IR is obesity-induced inflammation, as many cytokines and inflammatory mediators, particularly tumor necrosis factor- α (TNF- α), monocyte chemoattractant protein-1 (MCP-1), C-reactive protein (CRP), and interleukins are upregulated in individuals with IR¹⁰. Another factor that plays an important

role in IR is reactive oxygen species (ROS). Individuals with IR have reduced antioxidants and increased production of the ROS leading to increased systemic oxidative stress¹¹. ROS induced oxidative stress metabolites such as hydrogen peroxide, protein carbonyls, protein oxidative products, carbohydrate metabolites, short chain aldehydes, ketones, and hydrocarbons have been found to strongly associated with IR¹¹. Further research into these factors associated with IR development at the earlier stages or prior to the onset of clinical symptoms may uncover early biomarkers of IR. The addition of these biomarkers into the clinical setting would add to the clinical tools available for clinicians to identify at risk individuals prior to onset of disease and initiate earlier interventions.

Current clinical practice for assessing IR utilizes blood-based measures requiring fasting for 8 h and multiple visits to a medical facility for blood sampling and consultation. This is a burden for individuals who may benefit from frequently monitoring IR status, particularly young children and adolescents who rely on parents or guardians to facilitate this medical care. Individuals with IR may not have blood glucose levels that are in a range to be clinically diagnosed as T2DM but may be in a pre-diabetic stage of metabolic dysfunction. This stage is particularly important because it is a window of time when health and weight management strategies can be implemented to improve long-term cardio-metabolic health outcomes. Evidence suggests 70% of IR individuals will develop cardio-metabolic disease within 15-20 years^{12,13} without any intervention. This emphasizes the need to develop clinically relevant diagnostics tools to identify at risk individuals for intervention to slow or prevent progression to cardio-metabolic disease.

Non-invasive sampling, such as exhaled breath, provides metabolic information that may be indicative of an individual's health status. Exhaled breath metabolites have been reported in the assessment of infection^{14,15}, cognition¹⁶, metabolic disease¹⁷, and lifestyle¹⁸. These studies provide evidence that metabolic markers, which strongly correlate with metabolic dysregulation, can be identified in breath samples. Previously, we reported an exhaled breath metabolomics study investigating a young cohort of non-human primates (NHP) who developed insulin insensitivity due to developmental programming¹⁹. Following the development of IR, a volatile organic compound (VOC) breath signature was identified that discriminated between NHPs with IR compared to controls. It is suggested the altered breath signature arose from an altered cardiometabolic state potentially due to increased ROS production, beta-oxidation, inflammation, and lipid peroxidation¹⁹. Based on these findings, we hypothesized that metabolic markers contained in exhaled breath could be a potential diagnostic resource for defining metabolic status in humans and furthermore could be used to diagnose IR in pre-diabetic adolescents. Our goals were to show feasibility of adolescent breath collections using the ReCIVA breath collection device (Owlstone medical, Cambridge, UK) and to identify a small set of breath biomarkers in a pre-diabetic population that are important for the detection of IR and the risks for cardio-metabolic disease development.

2. MATERIALS AND METHODS

Ethical Approval of Study: All study procedures were approved by the Institutional Review Board for Baylor College of Medicine and Affiliated Hospitals (IRB-H-40940). Informed consent was obtained from parents or guardians and assent from participants. This work was performed in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

Study design and participants: Participants were recruited from the Pediatric Health and Weight Management Clinic at the Children's Hospital of San Antonio, TX. This cross-sectional study recruited Hispanic adolescents during their initial clinical visit. Criteria for inclusion were male and female ages ranging from 13-17 years with a BMI \geq of the 95th percentile. Exclusion criteria of the patients included a diagnosis of T2DM, asthma, or other chronic respiratory condition and any use of neurohormonal medications.

Clinical Assessments: Upon arrival to the clinic, a trained medical technologist collected anthropometric measures. Height and weight were collected using a BSM170 digital stadiometer and Scale 570 (InBody, Cerritos, CA). BMI was calculated as Kg/m^2 and z score calculated for all patients using their sex, precise age based on the date of birth and date of assessment, height, and weight. BMI percentile is expressed as the percent of the 95th percentile. Participants and parents/guardians verified fasting status prior to blood samples being collected as a standard of clinical care blood work. Blood (3 mL) was collected in red-top vacutainer tubes (BD, Franklin Lakes, NJ). Clinical blood measures included: fasting insulin (mU/L) and fasting glucose (mg/dL), Hemoglobin A1c (HbA1c), alanine aminotransferase test (ALT),

aspartate aminotransferase test (AST), fasting lipid profile, and high sensitivity CRP. The Homeostatic Model Assessment of Insulin Resistance, (HOMA-IR) of the participants was calculated based on the formula: fasting insulin (microU/L) \times fasting glucose (nmol/L)/22.5²⁰. The HOMA-IR categories are based on an adult Hispanic population based study²¹. **Table 1** reports the demographic and clinical measurement of all participants involved in this study.

Exhaled Breath Sample Collection:

On the same day of blood collection, participants provided exhaled breath samples through the breath collection protocol described below. The exhaled breath samples were collected using a commercially available breath sampler device, ReCIVA. This device safely monitors exhaled air pressure and CO₂ production allowing for consistent exhaled breath collections. The ReCIVA device was supplied medical-grade air (Airgas, Radnor, PA) at a rate of 40 L/min per manufacture recommendations to avoid artifacts and contamination from room air. Exhaled breath metabolites were concentrated onto bio-monitoring thermal desorption tubes (TDT) (CAT# C2-CAXX-5149, Markes International, Sacramento, CA) designed to capture VOCs from breath. Exhaled breath was collected with the following protocol:

- 1) Device collection parameters were set prior to starting of the collection in ReCIVA Breath Sampling Controller Software (Owlstone medical, Cambridge, UK). Total collection volume was set to 500 mL per tube at a rate of 200 mL/ min for lower and upper airways.
- 2) Upon entering the participant room, a trained research assistant opened a new one-time use breath collection mask and assembled the four TDT into the ReCIVA device. The ReCIVA mask was placed over the participant's mouth and nose and they were advised to breathe normally and familiarize themselves prior to the start of collection.
- 3) Once the participant acknowledged readiness, the breath collection was started by the

research assistant.

4) Once the software acknowledged the collection was complete, the ReCIVA mask was removed from the participant. TDTs were capped, transported to the research laboratory at the Texas Biomedical Research Institute and stored at 4 °C until analysis.

Thermal desorption and Analytical Instrumentation:

The exhaled breath VOCs collected on the TDT were processed within 1 month of collection. TDT was desorbed at 260 °C for 10 minutes to a cryotrap (general-purpose CAT#U-T11GPC-2S) using a thermal desorption unit, TD100xr (Markes Int., Sacramento, CA). VOCs concentrated on the cryotrap were rapidly heated at 40 °C/min to 280 °C and desorbed to a comprehensive two-dimensional gas chromatography tandem Time-of-Flight mass spectrometer (GC×GC-TOFMS, LECO Corp, St. Joseph MI, USA). The column set was configured as follows: ¹D Rxi- 5Sil-MS (95% polydimethylsiloxane, 5% phenyl; 30 m × 0.25 mm dc, 0.25 μm df) coupled with ²D Stabilwax column (Crossbond polyethylene glycol; 2.21 m × 0.18 mm dc, 0.18 μm df)(Restek, Bellefonte PA, USA). The inlet and ¹D column were connected by a PressFit column connector and the ¹D and ²D were connected by an MXT-union connector kits (Restek, Bellefonte PA, USA). The ²D column end is feed into the transfer line to the MS. The helium carrier gas flow was set to 2 mL/min for the entire run. The GC oven temperature program was set to ramp from 40 °C to 100 °C by 5 °C /min and then ramp by 8 °C /min to 220 °C with a final 10 minute hold for a total run time of 38 minutes.

Spectral data analysis:

Raw data were processed using Chromatof software (version 4.72, LECO Corp, St. Joseph MI,

USA). Spectra were collected over the range of m/z 40-400 at a rate of 100Hz. For peak findings, a signal-to-noise (S/N) cutoff was set to 100 and the NIST 2020 library was used for the identification of the analytes with a cut off of 700 match similarity. Analytes were aligned across all participants using Statistical-Compare tool contained within the Chromatof software. A chemical name was assigned if the analytes matched the following four criteria, 1) high mass spectral match, 2) group separation based on the structural formula 3) Retention Index match and 4) the extracted ion chromatogram (EIC) ionization patterns among all observed samples²². An analyte was not given a name if it did not match any of the four criteria. Possible contaminants were manually removed before further data analysis (**Supplementary Table S2**). Aligned data were exported for further data analysis.

Data availability statement

Raw spectral data will be made available upon request.

Data reduction and normalization

A brief summary of our data cleaning and feature reduction process is shown in **Supplementary Figure S1**. All statistical analyses were conducted in R 4.0.3 (R Core Team, Vienna, Austria)²³. Data cleaning was performed as described^{14,15}. In short, a frequency of observation (FOO) cut-offs of 80% was implemented to remove sparse features. The remaining features were normalized using Probabilistic Quotient Normalization (PQN)²⁴, \log_{10} transformed, and mean-centered. Missing values were imputed using a half-minimum approach²⁵.

Feature Selection, Machine learning and Statistical Analysis

A Random Forest (RF) regression model was implemented to select features clustering between individuals with different IR status. The important features were selected by considering all four

samples collected from each participant. Combining male (n=18) and female (n=10) samples, a total of 112 (n=28×4) samples were analyzed. The samples were then divided randomly, 3:1, for a training and validation dataset.

The RF model was built using a 10-fold cross-validation (CV) scheme in the ‘caret’ package²⁶. CV split the data into ten equal size pieces, building a model on nine of the ten pieces, and testing on the one remaining piece. The algorithm then leaves a different piece out and repeats this process for all pieces allowing for parameter tuning across the model, as well as estimating the model’s generalizability by examining accuracy statistics across the left-out pieces. Features were selected by an elbow cutoff of the RF variables important measure. The variables were given an importance measure by mean decrease accuracy method, which leaves one variable and builds a model with the rest of the variables.

Orthogonal partial least-squares discriminant analysis (OPLS-DA) was performed using the “ropls” package²⁷. Network analysis was performed in the MetScape 3²⁸ for Cytoscape²⁹ platform using the Kyoto Encyclopedia of Genes and Genomes (KEGG) ID and Human Metabolome Database (HMDB) ID. Additional statistical analysis was generated using Arsenal³³ package of R²³. For Pearson correlations, data were log₁₀-transformed and correlations were conducted in R²³ using ‘ggpubr’³⁵. Boxplots were generated using the Wilcoxon non-parametric test, performed comparing each IR group type. Figure 3 created in R²³ using ‘ggplot2’³⁷ and ‘ggpubr’³⁵. The ROC curve is generated in R²³ using the ‘MLevel’⁴⁰.

Biorender was used for generating Figure 6³⁰.

3. RESULTS

3.1 Feasibility of breath sampling from prediabetic adolescents

Twenty-eight adolescents participated in this study with a mean age of 15.46 years old (range 13-17 years old). All participants were obese or severely obese with an average BMI percent of the 95th percentile of 137.96 ± 20.15 (range 95-186), which expresses the extent of obesity (BMI \geq the 95th percentile are considered obese) among those adolescents³¹. The average exhaled breath sample collection time was 4.10 ± 0.05 minutes. At the time of collection all participants had not developed T2DM based on clinical blood measurements and clinician assessment.

3.2 Increased cardiovascular risk factors in Hispanic adolescents with obesity

As part of the routine clinical assessment, clinical blood tests were performed, and clinical characteristic are included in **Table 1**. Fasting glucose and insulin measures were used to calculate HOMA-IR, a surrogate measure of insulin sensitivity³². Based on HOMA-IR measures, three groups could be defined in our cohort, adolescents without IR (normal), adolescents with IR and adolescents with borderline IR with a statistically significant difference in IR value at < 0.001 in a Chi-square test of multiple groups. Adolescents with IR had a statistically significant higher BMI percentile ($p < 0.001$) with an average of 145.20 ± 21.69 compared with the normal adolescents 127.83 ± 16.57 . Further emphasizing a prediabetic status of the cohort, measures of cardiovascular disease (CVD) risk factors, were normal for total cholesterol (average 160.67 ± 31.84 mg/dL) and LDL-cholesterol (average 107.42 ± 25.79 mg/dL) and low for HDL-cholesterol (43.30 ± 11.48 mg/dL). However, normal fasting glucose (91.32 ± 14.07 mg/d) and fasting insulin (21.68 ± 12.54 mU/L) levels were significantly different ($p < 0.001$) in adolescents with IR. Detailed clinical measures are presented in **Table 1**.

Table 1: Demographic and clinical characteristics of the study cohort. The HOMA-IR cut-offs, are based on a study in a Hispanic population²¹. Data are presented as mean and standard deviation (SD).

	<i>Borderline (N=7[‡]) HOMA-IR 2.60–3.80</i>	<i>Insulin Resistance (N=15) HOMA-IR>3.80</i>	<i>Normal (N=6) HOMA-IR<2.60</i>	<i>Total (N=28)</i>	<i>p value[#]</i>
<i>BMI (% of the 95th Percentile)</i>					< 0.001
<i>Cholesterol</i>	131.14 (12.33)	145.20 (21.69)	127.83 (16.57)	137.96 (20.15)	0.951
<i>LDL</i>	162.00 (20.74)	160.66 (28.40)	159.16 (48.11)	160.67 (31.84)	0.137
<i>HDL</i>	114.14 (22.38)	107.333 (23.177)	99.83 (33.59)	107.42 (25.79)	< 0.001
<i>Triglycerides</i>	36.614 (5.287)	43.82 (6.05)	49.81 (20.25)	43.30 (11.48)	0.13
<i>Glucose (mg/dL)</i>	107.42 (34.26)	115.26 (79.19)	85.00 (21.10)	106.82 (62.07)	< 0.001
<i>Insulin mU/L</i>	86.28 (7.27)	96.73 (16.44)	83.66 (5.58)	91.32 (14.07)	< 0.001
<i>Glucose/Insulin ratio</i>	14.87 (1.56)	29.32 (12.72)	10.56 (1.84)	21.68 (12.54)	< 0.001
<i>[‡]HOMA-IR</i>	0.32 (0.05)	0.20 (0.05)	0.45 (0.10)	0.28 (0.12)	< 0.001
<i>[‡]FIRI</i>	3.16 (0.39)	7.40 (5.16)	2.16 (0.33)	5.22 (4.46)	< 0.001
<i>AIC</i>	0.23 (0.03)	0.14 (0.03)	0.32 (0.07)	0.20 (0.08)	0.008
<i>AST</i>	5.00 (0.26)	5.30 (0.59)	5.06 (0.17)	5.17 (0.47)	< 0.001
<i>ALT</i>	20.00 (9.09)	31.46 (17.38)	21.00 (4.71)	26.59 (14.80)	< 0.001
<i>CRP</i>	24.66 (16.27)	50.40 (32.80)	24.83 (9.94)	39.00 (28.90)	0.068
<i>Sex</i>	0.65 (0.69)	0.78 (0.94)	0.29 (0.30)	0.65 (0.80)	0.559
<i>F</i>	2 (28.6%)	6 (40.0%)	2 (33.3%)	10 (35.7%)	
<i>M</i>	5 (71.4%)	9 (60.0%)	4 (66.7%)	18 (64.3%)	

Note:

[‡]Each participant provided 4 technical replicate samples for the analysis.

[#]P-value is calculated by Chi-square test of multiple groups.

[‡] Homeostatic model assessment of insulin resistance (HOMA-IR) is calculated by the formula: fasting insulin (microU/L) x fasting glucose (nmol/L)/22.5.

[‡] Fasting insulin resistance index (FIRI) is calculated by the formula: (fasting glucose × fasting insulin)/25

3.3 Machine learning approach identifies important breathprint for IR

The RF regression model identified the most important breath features for identifying individuals with IR based on their HOMA- IR value. Ten features were selected with the highest importance as shown on the RF importance plot (**Supplementary Figure S2**). The elbow cut-off >18 were

selected based on the large drop on the important measures on the mean decrease in Gini in the RF model³⁴.

The ten important analytes include limonene, decane-2,4,6-trimethyl, undecane, undecane-2,7-dimethyl, pentylbenzene, octamethyloctane, eicosane and three unknown analytes (**Supplementary Table S1**). Identification of these analytes was confirmed by retention indices and mass spectral match score with NIST 2020 library's analytical standard (level 2 and 3 identification)²². The unknowns reported are those whose retention indices were consistent, but the mass spectra matching was inconsistent or below 700 similarity score, hence a compound name was not assigned. Chromatographic and mass spectral information on the ten compounds is provided in the **Supplementary Table S1**. An example of a peak detection and identification used in this study are shown in **Supplementary Figure S3**.

3.4 The breath-IR model correlates with the standard blood based HOMA-IR

The ten breath VOCs identified in the training dataset, were evaluated against the remaining samples in the test dataset (**Figure 1a**). Test dataset comprised of a total of 37 samples that were not used for the training dataset. The test dataset also resulted in a positive correlation with blood based HOMA-IR values with a Pearson correlation, $R=0.87$, $p < 0.001$. Combining the training and test datasets in a total dataset resulted in a higher correlation with the blood based HOMA-IR values with a Pearson correlation, $R = 0.95$, and $p < 0.001$ (**Figure 1b**).

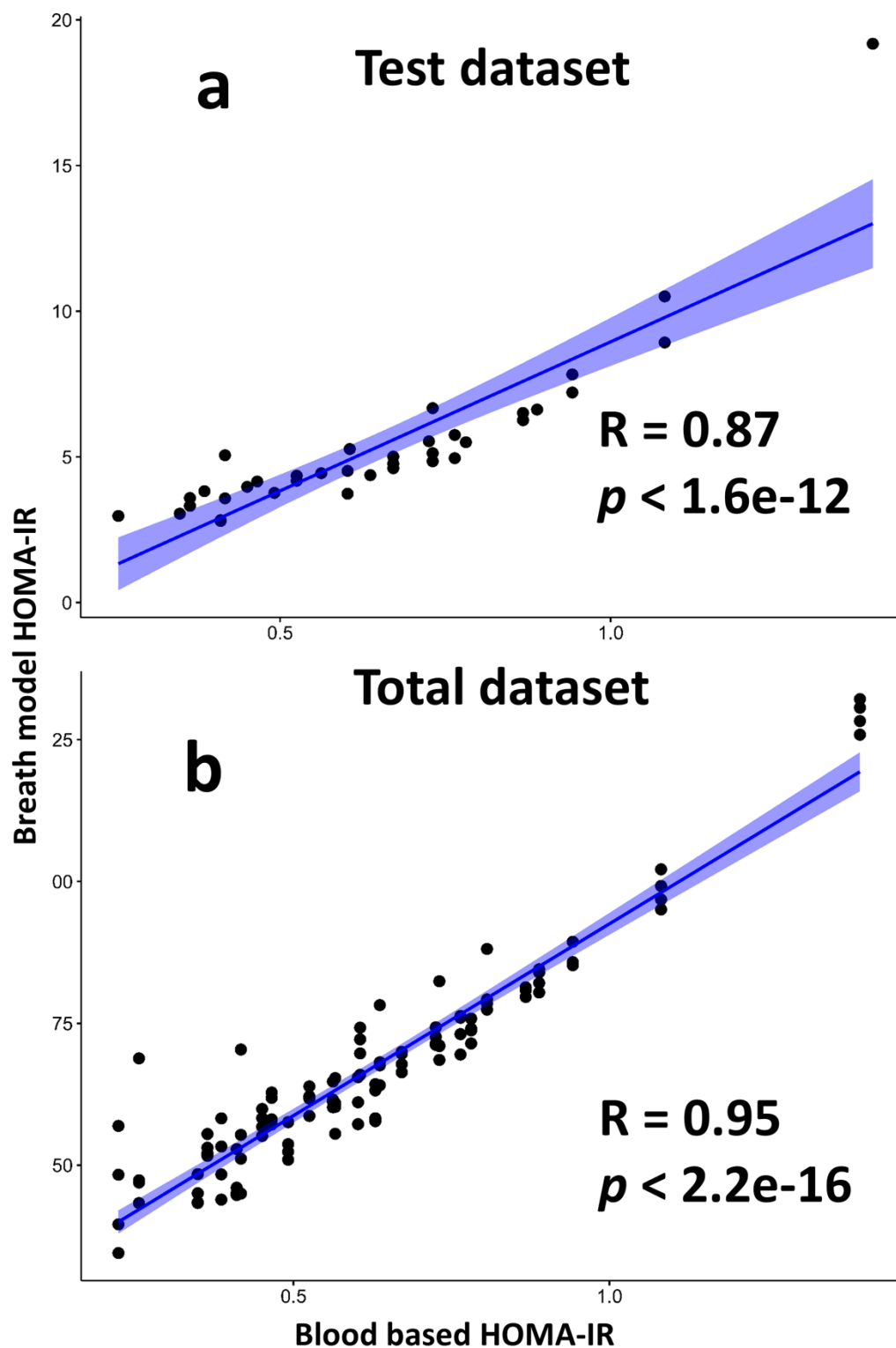


Figure 1: Pearson correlation of the breath-based IR with the blood based HOMA-IR measurement. (a) Correlation of a test dataset against the blood based HOMA-IR measures

testing the Breath IR training set on independent samples. (b) Correlation of the total dataset combining training and test datasets for the breath-IR model against blood based HOMA-IR.

3.5 Breath-IR model correlation with the glycemic profile but not with the lipid profile

The breath IR model shows a strong correlation with measures of glucose metabolism including fasting glucose (mg/dL) and fasting insulin (mU/L) with $R=0.91$, $p < 0.001$ and $R=0.80$, $p < 0.001$, respectively (**Figure 2a and b**). A weak positive correlation was observed with triglycerides (mg/dL), $R=0.35$, $p < 0.001$ and weak negative but non-significant correlation with the LDL levels at $R=-0.028$, $p < 0.77$ (**Figure 2c and d**). The breath IR model did not correlate with the total cholesterol (mg/dL) or HDL level mg/dL **Supplementary Figure S4**. The breath-IR model also correlated with additional indices of IR; fasting insulin resistance index (FIRI)³⁶ and the glucose/insulin (G/I) ratio showed in the **Supplementary Figure S4**. The HbA1c test positively correlated with the breath IR model, $R = 0.63$ and $p = < 0.001$ (**Supplementary Figure S4**).

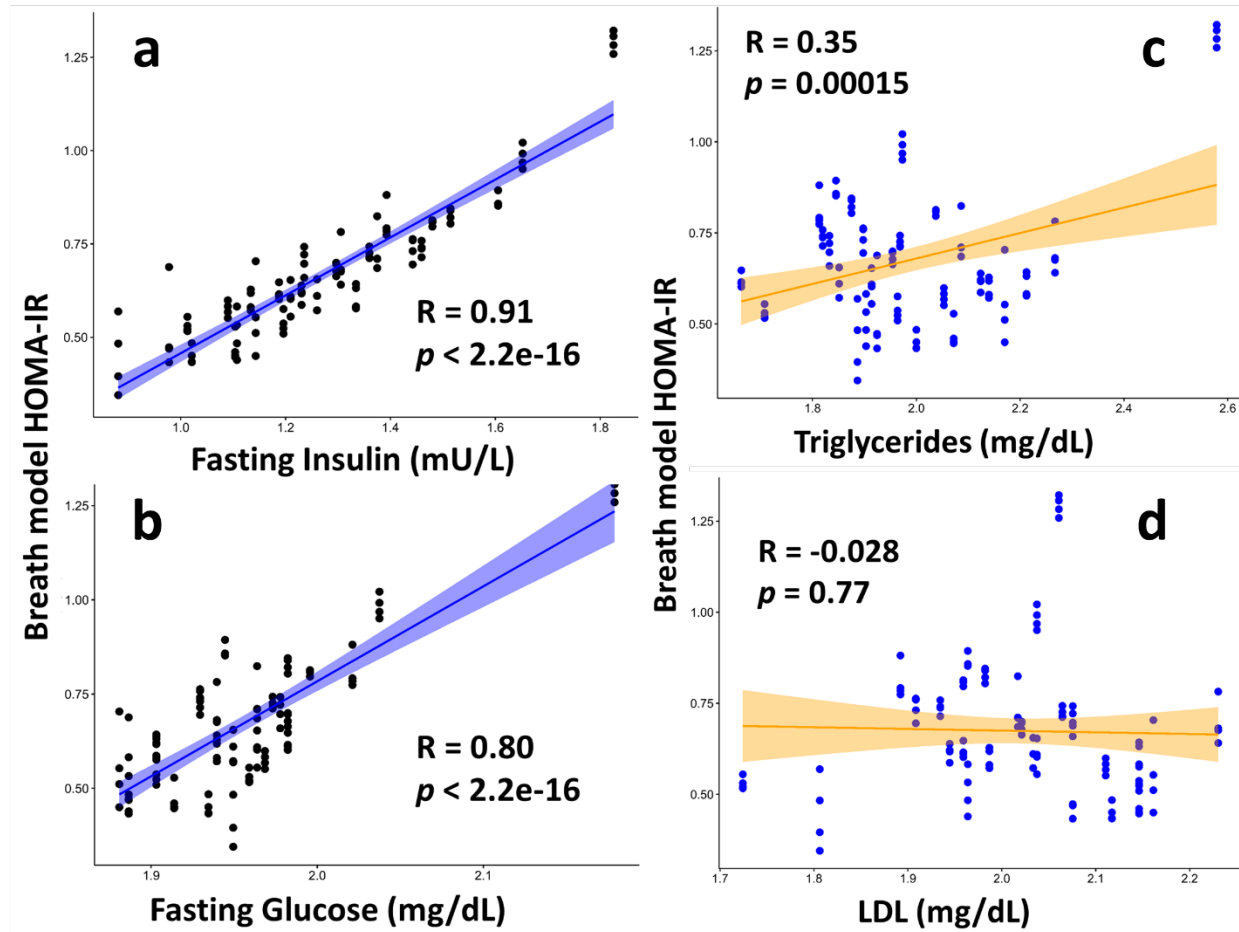


Figure 2: Pearson correlation of the breath based HOMA-IR with glycemic and lipid measures.

Pearson correlation was performed to compare the breath based HOMA-IR model against (a) fasting insulin (mU/L), (b) fasting glucose (mg/dL), (c) triglycerides (mg/dL) and (d) LDL (mg/dL).

3.4 Breath-IR model accurately identifies the IR classification

To determine the breath-IR model accuracy for identifying individuals with IR, the ten features identified by RF were further investigated. Four of the ten VOCs, limonene, undecane, undecane, 2,7-dimethyl and unknown 1 were statistically significantly different for individuals with IR ($p=0.003$, $p=0.002$, $p<0.001$ and $p=0.033$, respectively) compared to those without IR (**Figure 3**). A trend was observed for the remaining analytes but was not statistically significant **Supplementary Figure S5**. A moderate difference was observed for limonene when comparing borderline IR individuals to normal ($p=0.068$), whereas comparison of borderline individuals to IR group was not significant ($p=0.82$).

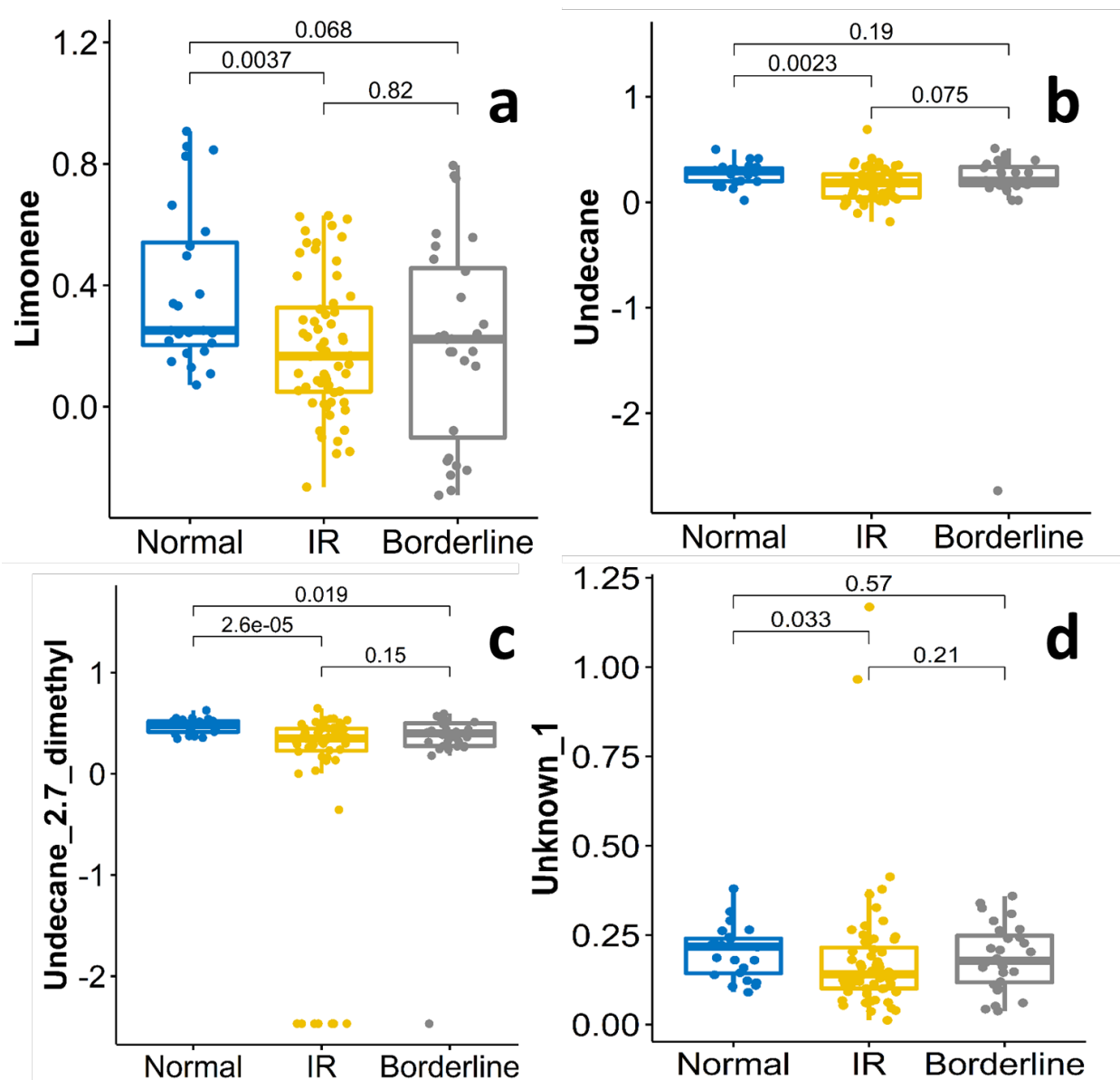


Figure 3: Boxplots of mean centered and normalized peak area for four compounds selected from the breath-IR model across normal, IR and borderline IR adolescents. Each compound's median observed peak area between the groups were different indicating a univariate difference which may be contributing the discrimination of the IR group in the combined 10 compounds model. Boxplots represent the quartiles of the data (first line is the first quartile, midline is the median, third line is the third quartile) where whiskers represent $1.5 \times$ IQR (inter-quartile range).

To evaluate the combined performance of the breath-IR model for clustering between the HOMA-IR categories, an Orthogonal Partial Least Squares Discriminatory Analysis (OPLS-DA) was performed^{27,38,39}. The OPLS-DA showed a clear clustering by the HOMA-IR score, shown in **Figure 4**. Two extreme HOMA-IR values (24.92, 12.07) were removed from the dataset to show clearer clustering, but even with those extreme values, a similar pattern of clustering of the individuals based on the breath-IR model was observed (**Supplementary Figure S6**).

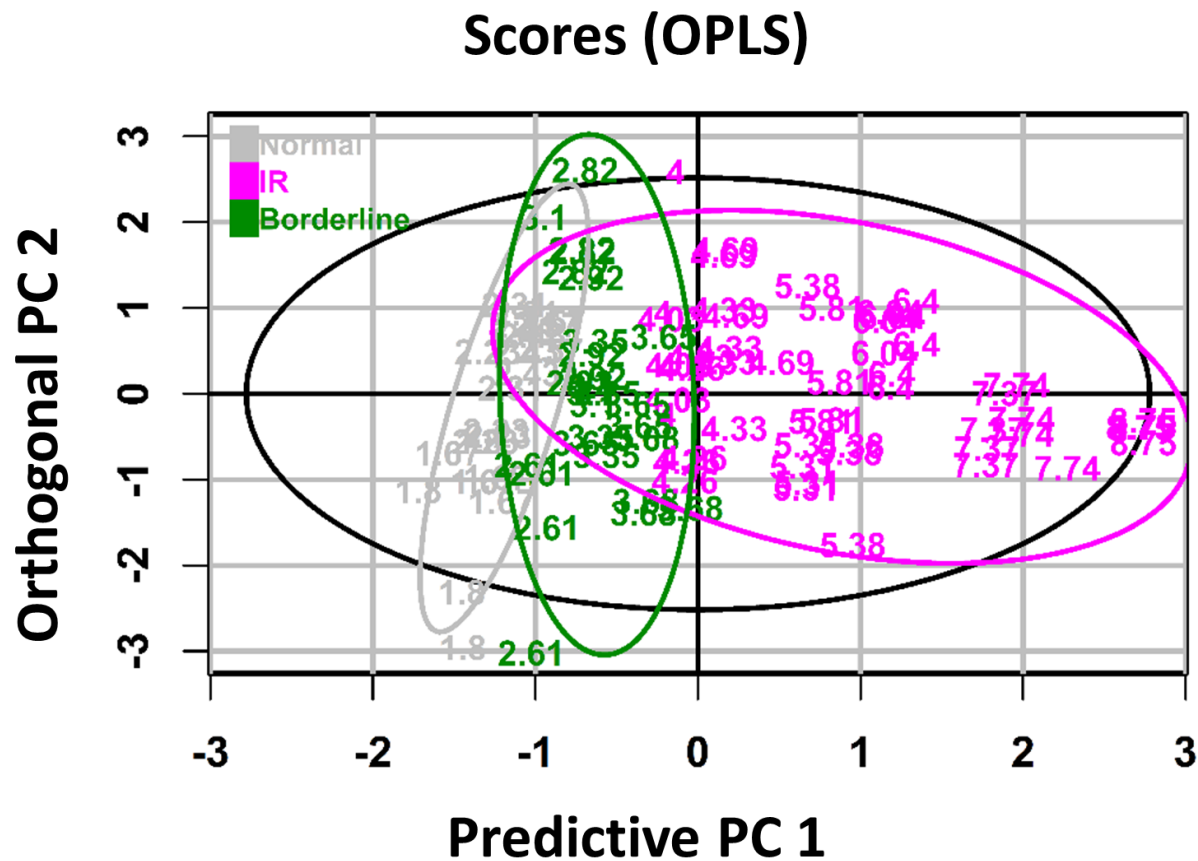


Figure 4: The supervised Orthogonal partial least-squares discriminant (OPLS) analysis of breath HOMA-IR model. The score scatter plot of OPLS model showing the clustering of samples in horizontally by predictive principal component (PC) 1 and vertically by orthogonal PC 2. The IR cut-off, is based on a study in Hispanic population²¹. The purple color indicates the IR group with HOMA-IR>3.80, the green color represents the borderline IR group with HOMA-IR 2.60–3.80

and HOMA-IR<2.60 indicate normal group which is presented as grey color. The 95% tolerance region corresponds to the ellipse that is defined by the Hotelling's T2 parameter.

To evaluate the performance of the breath-IR model for classifying individuals with IR, the area under the receiver operating characteristic curve (AUC - ROC) was plotted resulting in a breath IR model value of 0.87 (**Figure 5**). This value is generated by an RF iteration based on the selected ten breath compounds. The process was repeated for only four VOCs that were significantly different for individuals with IR, limonene, undecane, undecane, 2,7-dimethyl and unknown 1 reducing to 0.83. The RF model's importance was also assessed by randomly selecting features in the dataset resulting in a AUC of 0.57. The breath-IR model observed sensitivity and specificity as 73.1% (60.0 to 83.0% in cross-validation) and 81.7% (70.0 to 89.0% in cross-validation), respectively.

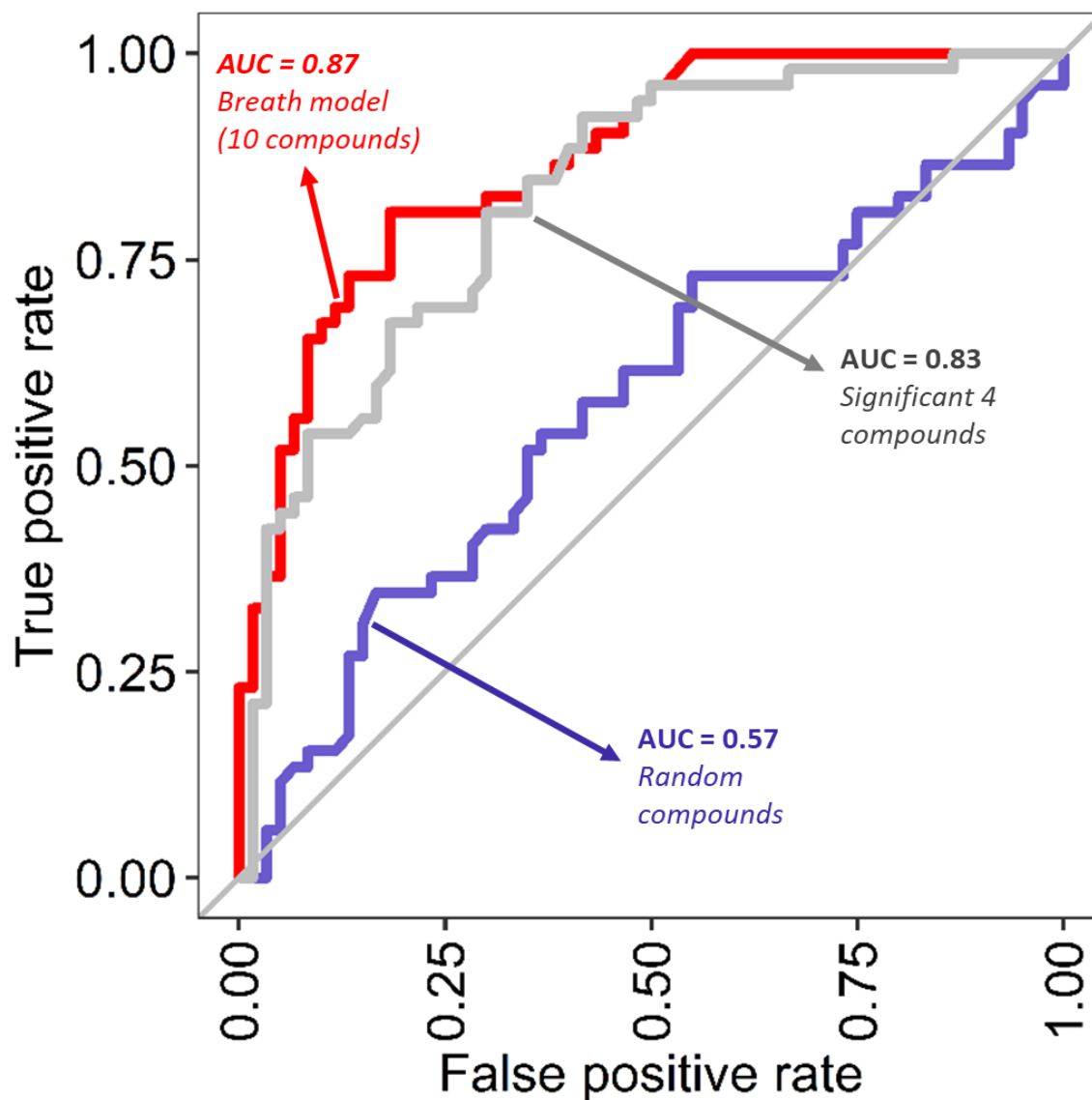


Figure 5: Diagnostics performance of the Breath-IR model. The receiver operating characters (ROC) curves using the random forest model between the IR (HOMA-IR>3.80) and normal (HOMA-IR \leq 3.80) adolescents. The ‘red line’ indicates the model’s performance based on 10 compounds in breath IR model, the ‘grey line’ indicate performance by only four compounds with individual statistically significant difference, limonene, undecane, undecane, 2,7-dimethyl and unknown 1. The ‘blue line’ indicates randomly selected 4 breath metabolites from the original

dataset. The AUC-ROC across folds after cross validation are 0.87 (0.80-0.94) for 10 features breath model, 0.83 (0.75-0.91) for significant 4 compounds, and 0.57 (0.46-0.68) for random features model.

4. DISCUSSION

4.1 Non-invasive exhaled breath as IR diagnosis

Our study shows the feasibility of collecting non-invasive exhaled breath metabolites from an adolescent population, which can then be used to assess an individual's metabolic health status. Utilizing an untargeted metabolomics approach allowed for the identification of all breath metabolites in detectable range of our mass spectrometer. By implementing a machine learning model to assess the normalized data, a predictive model based on the selected features could be built to inform diagnostic potential of exhaled breath signatures. Our previous studies in nonhuman primates identified exhaled breath signatures in animals with IR¹⁹. The adoption of the HOMA-IR categories was based on a population-based study reported from the adult Mexican Americans population²¹, as no clear definition exist for the Hispanic adolescents with IR. Our breath-IR model strongly correlates with all glycemic clinical measurements, but the model did not strongly correlate other clinical measures (i.e., circulating lipids). Lipid levels were not in a range that would have clinically categorized the participants as having T2DM, therefore the breath IR model is specific to measures impacting IR. Given the breath-IR model resulted in a strong correlation with blood based HOMA-IR measurements, information contained in exhaled breath samples has utility for a non-invasive IR diagnostic.

This pilot study is exploratory in nature and our breath metabolites distribution is based on relative abundance of each VOC. Therefore, metabolite levels are based on the normalized peak area not the actual quantity of metabolites. Four of the ten important features identified in the breath-IR model limonene, undecane, undecane, 2,7-dimethyl and unknown 1 showed a statistically significant difference between the normal and IR group. These four metabolites are the major contributors to the accuracy of the model. The OPLS-DA model filtered out noise in the dataset that is not correlated with the outcome variables⁴¹. Given two of the adolescents had an exceptionally high HOMA-IR (24.92 and 12.07), these two values put a biased weight on the classification problem. Although those values are extreme within the current cohort, two different OPLS-DA models, one without those values and one with those values, as shown in **Figure 2** and **Supplementary Figure S6**, respectively. These plots show clear clustering of individuals by their IR levels based on the breath metabolites, as IR individuals with HOMA-IR (>3.80) were clustered on the right side of the plots whereas the normal HOMA-IR individual (<2.60) on the left. The border line IR group with HOMA-IR 3.8-2.6 were also clustered in-between of the IR and normal group although there is some overlap. Although the other six breath metabolites did not reach significance between the groups individually (**Supplementary Figure S5**), their contribution to the breath-IR model is evident when assessing the diagnostic capability of the model (**Figure 5**) to identify individuals with IR. A combined signature of breath metabolites is clearly more informative than using a single metabolite which may be indicative of some normal variability in metabolites within and between individuals^{15,19,42}.

Our breath IR model shows good accuracy 77.8% with sensitivity of 73.1% (60%-83% within cross-validation) and specificity of 81% (70%-89% within cross-validation) (**Figure 5**). As a comparison, the American Diabetes Association (ADA) and International Expert Council of the

ADA define the diabetes diagnostic criteria for hemoglobin A1c (HbA1c); sensitivity, 73.3%; specificity, 88.2%^{43,44}. Although our study needs to be verified in a larger more diverse independent cohort, our results are promising for the breath-based IR diagnostic tool for monitoring IR and metabolic health, which could be developed further.

4.2 Biological insight of the breath biomarkers

Among the seven named breath biomarkers, limonene, undecane, pentylbenzene, and eicosane have been previously reported in the human metabolome according to the KEGG⁴⁵ and HMDB⁴⁶ databases. Undecane was previously detected as a biomarker of nonalcoholic fatty liver disease (NAFLD)⁴⁷, asthma⁴⁸, and gastrointestinal disease⁴⁹. Pentylbenzen and eicosane are hydrocarbon molecules and subcellular component of membrane epithelium^{50,51}. These two analytes were previously reported from human metabolomics studies but were never quantified⁴⁶. Similar hydrocarbon metabolites were reported as oxidative stress markers of disease in exhaled breath related to infectious disease¹⁵, heart disease^{52,53}, psychiatric disease⁵⁴, and cancer⁵⁵. This mechanism involves hydrocarbons being triggered or released in to the breath by ROS, a form of oxygen with a reactive electron from the by-product of oxidation metabolism in the mitochondria. These highly reactive molecules when released to the cytoplasm have the potential to damage DNA, protein, and other cellular metabolites, producing a range of small chain saturated or unsaturated hydrocarbon⁵⁶. These hydrocarbons then have the potential to enter blood circulation and be detected in the breath as non-specific oxidative markers due to normal blood gas exchange in the lungs. IR has been established as a strong contributor to oxidative stress and is known to damage the vascular lining of cells, increasing the risk of atherosclerosis⁵⁷. By identifying oxidative markers in exhaled breath from the obese adolescents that correlate with altered

metabolic status within this cohort, suggests the underlying oxidative stress may be specific to the IR development.

Several studies have reported limonene from blood, feces, urine, saliva, and breath⁵⁸⁻⁶². Limonene is a monocyclic monoterpene, with one isoprene chain. It is naturally abundant in nature and used in the flavor and fragrance of foods and drinks. In exhaled breath studies of liver cirrhosis patients elevated level of limonene were observed compare to the controls⁶³⁻⁶⁵. As an exogenous metabolite, limonene is ingested during dietary intake and then metabolized in the liver by P450 enzymes CYP2C9 and CYP2C19 and their enzyme products (S)-limonene 6-monooxygenase, (S)-limonene 7-monooxygenase and (R)-limonene 6-monooxygenase (**Figure 6a**). Limonene can then be metabolized to trans-carveol and perillyl alcohol⁶⁶. It has been noted that in patients with liver disease, particularly NAFLD at any stage, have a reduced capacity to produce both CYP2C and CYP2C19 enzymes⁶⁷. Reduced liver capacity to metabolize limonene results in abnormal levels in circulation and excretion leading to a difference on the limonene which can be identified in exhaled breath (**Figure 6b**).

ROS arises from vascular organs including heart cells⁶⁸ or kidney cells⁶⁹ leading to a higher systemic level of oxidation and peroxidation products. This may include aliphatic or aromatic hydrocarbon, straight-chain mono- or poly- unsaturated aldehyde, straight-chain -mono or -poly unsaturated carboxylic acids, ester, epoxides, MUFAs and PUFAs⁷⁰. Higher levels in the blood may cross the blood–air barrier within the gas exchanging region of the lungs and exhaled through breath. Thus a set of metabolites produced from the multiple organs as a result of cardio-metabolic disease development could potentially lead to a set of breath based cardio-metabolic biomarkers

as indicated in **Figure 6b**. Our detection of limonene as an important breath metabolite for determining IR status, may be due to early signs of NAFLD in obese adolescents' breath used in this study.

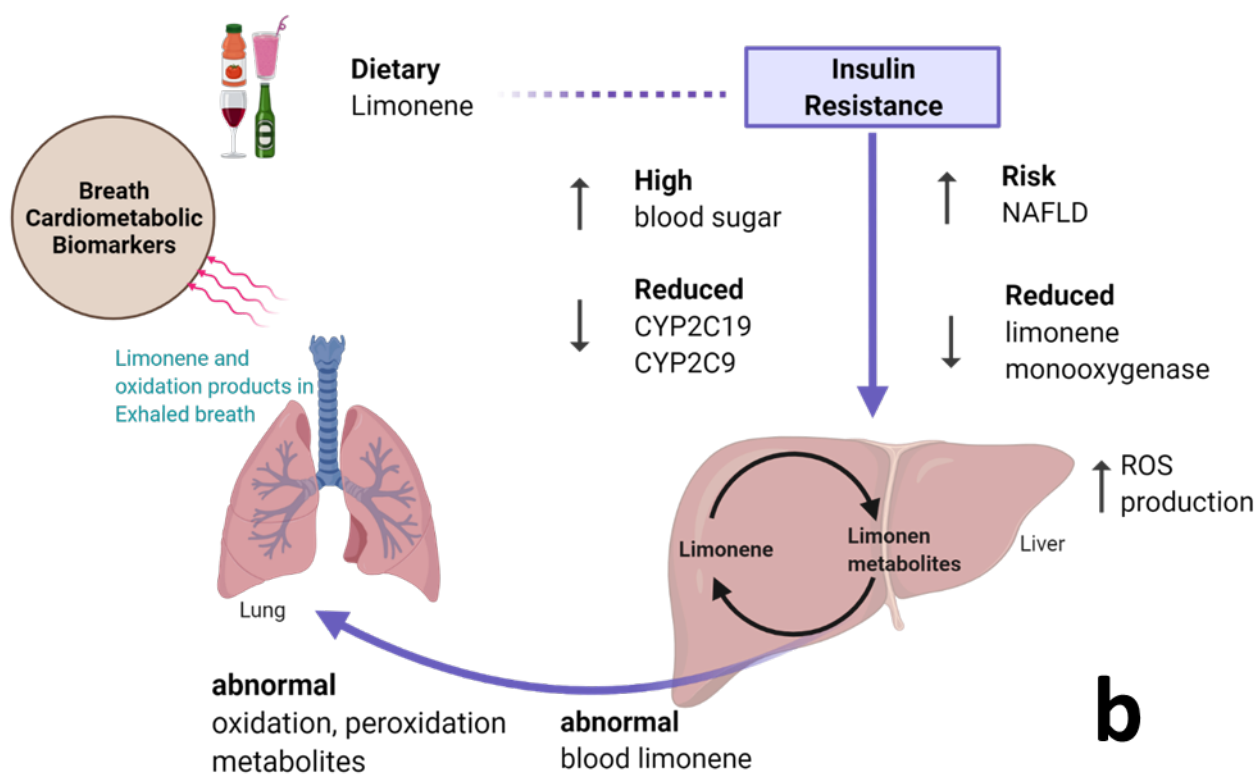
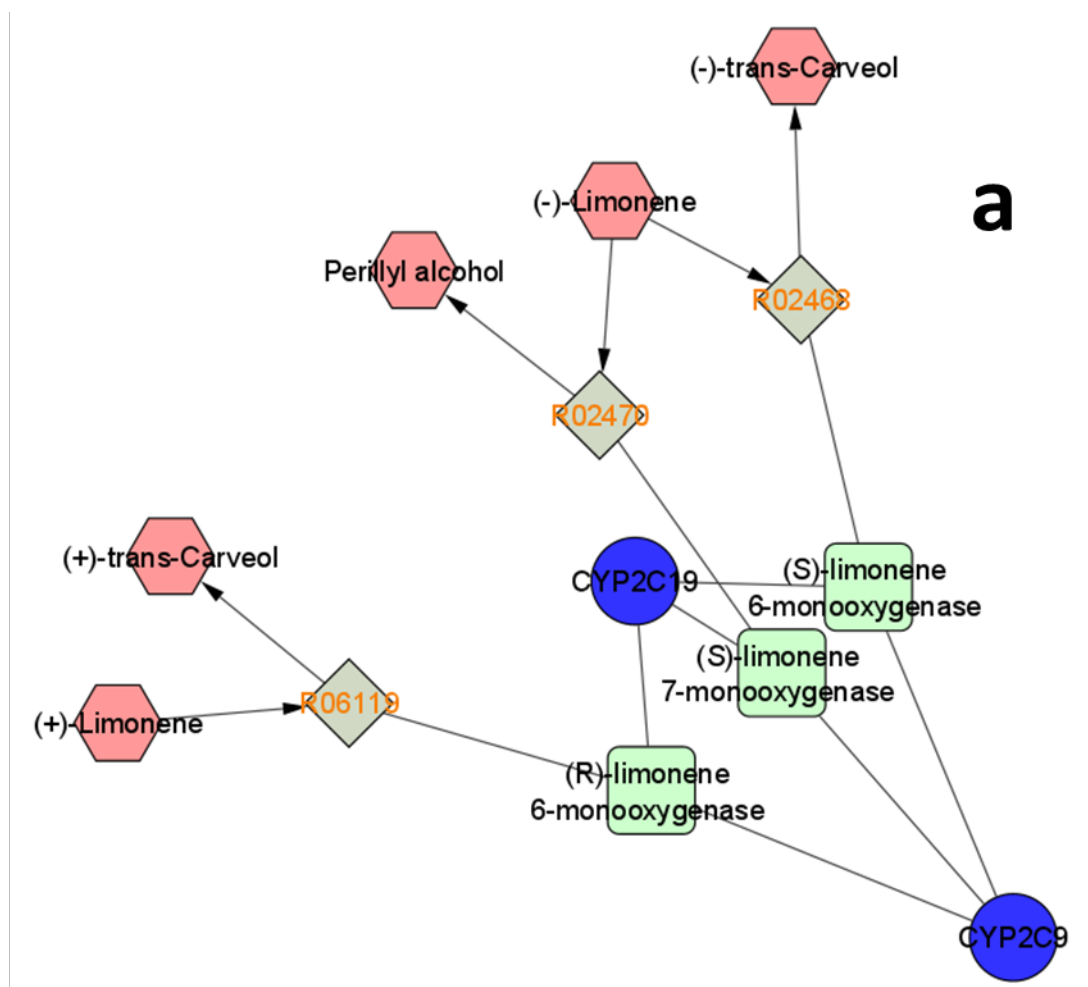


Figure 6: Molecular network analysis of limonene. The complex interaction of the gene, enzymes and the metabolites are presented by Cytoscape²⁹ platform using Metscape 2²⁸ network analysis (a). The reaction R02468, R02470 and R06119 connect the injected (+) limonene and (-) limonene to its oxidized metabolites, (+) trans-carveol, (-) trans-carveol and perillyl alcohol. The enzymes responsible for these reactions are P450 enzymes CYP2C9 and CYP2C19 and their products (S)-limonene 6-monooxygenase, (S)-limonene 7-monooxygenase and (R)-limonene 6-monooxygenase and (b) the proposed mechanism of the breath cardio-metabolic biomarkers. Increased ROS production, abnormal limonene and oxidation metabolites due to liver damage may lead to an abnormal level of breath metabolites. The figure is created using biorender³⁰ platform.

Given that the abnormal levels of limonene are present in individuals with IR suggest potential NAFLD and liver damage, we investigated clinical measures of two liver specific enzymes aspartate aminotransferase (AST) and alanine aminotransferase (ALT) and systemic inflammation marker c-reactive protein (CRP). AST and ALT are both liver specific enzymes that become elevated in blood as a result of liver damage. Our breath-IR model was slightly negative but significantly correlated with the AST/ALT ratio (**Supplementary Figure S7**). The AST/ALT ratio less than <1 is suggestive of NAFLD/NASH whereas the score >2 is suggestive of alcoholic liver disease⁷¹. A negative correlation between breath IR and AST/ALT ratio thus indicates adolescents have early NAFLD/NASH. Evaluation of CRP levels were moderately positive and statistically significant correlated with our breath-IR model (**Supplementary Figure S7**). This confirms that the presence of systemic inflammation that could negatively impact multiple organ systems throughout the body including the liver of those adolescents.

4. CONCLUSION

This pilot study demonstrates the feasibility of studying breath in adolescents and the potential of non-invasive breath metabolites as an important tool to detect IR in a unique cohort of Hispanic pre-diabetic adolescents with obesity. The breath model developed by a machine learning-based approach showed a strong correlation with the surrogate blood test for IR that is currently used in the clinical environment. The breath IR model confidently clustered individuals with and without IR, showing a promising diagnostics performance of the breath metabolites as evident by the ROC-AUC = 0.87. The detection breath metabolites as important features for IR status may result from early liver damage or due to increased cellular damage from ROS production as a result of obesity and lifestyle. Future studies will focus on validating the usefulness of breath signatures for detection and monitoring of IR in a larger population of at-risk individuals.

Acknowledgements

We would like to acknowledge the Julee Carlton and Veronica Duran from the Health and Weight Management Clinic, Children's Hospital of San Antonio, San Antonio, TX who help collect the clinical measurement and facilitate sample collection. We also acknowledge the Texas Biomedical Research Institute for providing instrumental support of this project.

Author Contribution

Conceptualization: ACB, LAC, GMK SC

Data curation: ACB, SC

Formal Analysis: MSK

Funding acquisition: LAC

Methodology: MSK, ACB, SC, LAC

Project administration: ACB

Resources: LAC, ACB
Visualization: MSK
Writing – original draft: MSK
Writing – review & editing: ACB, LAC, GMK, SC

Conflict of Interest

Authors declare no conflict of interest.

Funding

This work is funded by Healthy Babies Project, Texas Biomedical Research Institute, San Antonio, TX.

5. REFERENCES

1. Ormazabal V, Nair S, Elfeky O, Aguayo C, Salomon C, Zuñiga FA. Association between insulin resistance and the development of cardiovascular disease. *Cardiovascular Diabetology*. 2018;17(1):1-14.
2. Utzschneider KM, Kahn SE. The role of insulin resistance in nonalcoholic fatty liver disease. *The Journal of Clinical Endocrinology & Metabolism*. 2006;91(12):4753-4761.
3. Navaneethan SD, Kirwan JP, Remer EM, et al. Adiposity, Physical Function, and Their Associations With Insulin Resistance, Inflammation, and Adipokines in CKD. *American Journal of Kidney Diseases*. 2020;77(1):44-55.
4. van der Aa MP, Fazeli Farsani S, Knibbe CAJ, de Boer A, van der Vorst MMJ. Population-Based Studies on the Epidemiology of Insulin Resistance in Children. *Journal of Diabetes Research*. 2015;2015:362375.
5. CDC (Centers for Disease Control and Prevention). Prevalence of Obesity Among Adults and Youth: United States. 2017; <https://www.cdc.gov/nchs/data/databriefs/db288.pdf>.
6. Skinner AC, Ravanbakht SN, Skelton JA, Perrin EM, Armstrong SC. Prevalence of Obesity and Severe Obesity in US Children, 1999–2016. *Pediatrics*. 2018;141(3):e20173459.
7. Yaribeygi H, Farrokhi FR, Butler AE, Sahebkar A. Insulin resistance: Review of the underlying molecular mechanisms. *Journal of Cellular Physiology*. 2019;234(6):8152-8161.

8. Samuel VT, Shulman GI. The pathogenesis of insulin resistance: integrating signaling pathways and substrate flux. *The Journal of Clinical Investigation*. 2016;126(1):12-22.
9. Hirosumi J, Tuncman G, Chang L, et al. A central role for JNK in obesity and insulin resistance. *Nature*. 2002;420(6913):333-336.
10. Dandona P, Aljada A, Bandyopadhyay A. Inflammation: the link between insulin resistance, obesity and diabetes. *Trends in Immunology*. 2004;25(1):4-7.
11. Furukawa S, Fujita T, Shimabukuro M, et al. Increased oxidative stress in obesity and its impact on metabolic syndrome. *The Journal of Clinical Investigation*. 2017;114(12):1752-1761.
12. ADA (American Diabetes Association). Standards of medical care for patients with diabetes mellitus. *Diabetes Care*. 2003;26(suppl 1):s33-s50.
13. Tabák AG, Herder C, Rathmann W, Brunner EJ, Kivimäki M. Prediabetes: a high-risk state for diabetes development. *The Lancet*. 2012;379(9833):2279-2290.
14. Khan MS, Bobak C, Mellors TR, et al. The early breath biomarkers of tuberculosis using model macaque monkey. Breath Summit; 2019; Leicestershire, UK.
15. Bobak CA, Kang L, Workman L, et al. Breath can discriminate tuberculosis from other lower respiratory illness in children. *Scientific Reports*. 2020;11(1).
16. Tiele A, Wicaksono A, Daulton E, et al. Breath-based non-invasive diagnosis of Alzheimer's disease: a pilot study. *Journal of Breath Research*. 2020;14(2):026003.
17. Li W, Liu Y, Lu X, et al. A cross-sectional study of breath acetone based on diabetic metabolic disorders. *Journal of Breath Research*. 2015;9(1):016005.
18. Krilaviciute A, Leja M, Kopp-Schneider A, et al. Associations of diet and lifestyle factors with common volatile organic compounds in exhaled breath of average-risk individuals. *Journal of Breath Research*. 2019;13(2):026006.
19. Bishop AC, Libardoni M, Choudary A, et al. Nonhuman primate breath volatile organic compounds associate with developmental programming and cardio-metabolic status. *Journal of Breath Research*. 2018;12(3):036016.
20. Matthews DR, Hosker J, Rudenski A, Naylor B, Treacher D, Turner R. Homeostasis model assessment: insulin resistance and β -cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia*. 1985;28(7):412-419.
21. Qu H-Q, Li Q, Rentfro AR, Fisher-Hoch SP, McCormick JB. The definition of insulin resistance using HOMA-IR for Americans of Mexican descent using machine learning. *PloS One*. 2011;6(6):e21041.
22. Sumner LW, Amberg A, Barrett D, et al. Proposed minimum reporting standards for chemical analysis. *Metabolomics*. 2007;3(3):211-221.
23. R Core Team. R: A Language and Environment for Statistical Computing. 2019.
24. Noonan MJ, Tinnesand HV, Buesching CD. Normalizing Gas-Chromatography-Mass Spectrometry Data: Method Choice can Alter Biological Inference. *BioEssays : news and reviews in molecular, cellular and developmental biology*. 2018;40(6):e1700210.

25. Wei R, Wang J, Su M, et al. Missing Value Imputation Approach for Mass Spectrometry-based Metabolomics Data. *Scientific Reports*. 2018;8(1):663.
26. Kuhn M. Building predictive models in R using the caret package. *Journal of Statistical Software*. 2008;28(5):1-26.
27. Thévenot EA, Roux A, Xu Y, Ezan E, Junot C. Analysis of the Human Adult Urinary Metabolome Variations with Age, Body Mass Index, and Gender by Implementing a Comprehensive Workflow for Univariate and OPLS Statistical Analyses. *Journal of Proteome Research*. 2015;14(8):3322-3335.
28. Karnovsky A, Weymouth T, Hull T, et al. Metscape 2 bioinformatics tool for the analysis and visualization of metabolomics and gene expression data. *Bioinformatics*. 2012;28(3):373-380.
29. Shannon P, Markiel A, Ozier O, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*. 2003;13(11):2498-2504.
30. BioRender.com. Adapted from “Regulation of Blood Glucose”. 2021; <https://app.biorender.com/Regulation-of-blood-glucose>
31. Kuczumski RJ. CDC Growth Charts for the United States: methods and development. *Centers for Disease Control and Prevention*. 2002(246).
32. Gutch M, Kumar S, Razi SM, Gupta KK, Gupta A. Assessment of insulin sensitivity/resistance. *Indian Journal of Endocrinology and Metabolism*. 2015;19(1):160-164.
33. Heinzen E, Sinnwell J, Atkinson E, Gunderson T, Dougherty G. Arsenal: An Arsenal of ‘R’ Functions for Large-Scale Statistical Summaries. *R package version*. 2018;1(0).
34. Genuer R, Poggi J-M, Tuleau-Malot C. Variable selection using random forests. *Pattern Recognition Letters*. 2010;31(14):2225-2236.
35. Kassambara H. ggpubr: ‘ggplot2’ Based Publication Ready Plots. 2020.
36. Duncan MH, Singh BM, Wise PH, Carter G, Alaghband-Zadeh J. A simple measure of insulin resistance. *The Lancet*. 1995;346(8967):120-121.
37. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer; 2016.
38. Trygg J, Wold S. Orthogonal projections to latent structures (O-PLS). *Journal of Chemometrics*. 2002;16(3):119-128.
39. Khan MS, Rahman MM, Azmir J, et al. Multivariate analysis of PRISMA optimized TLC image for predicting antioxidant activity and identification of contributing compounds from *Pereskia bleo*. *Biomedical Chromatography*. 2015;29(12):1826-1833.
40. John C. MLeval: Machine Learning Model Evaluation. *R package version*. 2020;3.
41. Wold S, Antti H, Lindgren F, Öhman J. Orthogonal signal correction of near-infrared spectra. *Chemometrics Intelligent laboratory systems*. 1998;44(1-2):175-185.
42. Beccaria M, Mellors TR, Petion JS, et al. Preliminary investigation of human exhaled breath for tuberculosis diagnosis by multidimensional gas chromatography–Time of

- flight mass spectrometry and machine learning. *Journal of Chromatography B*. 2018;1074:46-50.
43. Cox ME, Edelman D. Tests for screening and diagnosis of type 2 diabetes. *Clinical Diabetes*. 2009;27(4):132-138.
 44. International Expert Committee. International Expert Committee report on the role of the A1C assay in the diagnosis of diabetes. *Diabetes Care*. 2009;32(7):1327-1334.
 45. Kanehisa M, Araki M, Goto S, et al. KEGG for linking genomes to life and the environment. *Nucleic Acids Research*. 2007;36:D480-D484.
 46. Wishart DS, Tzur D, Knox C, et al. HMDB: the Human Metabolome Database. *Nucleic Acids Research*. 2007;35:D521-526.
 47. Raman M, Ahmed I, Gillevet PM, et al. Fecal Microbiome and Volatile Organic Compound Metabolome in Obese Humans With Nonalcoholic Fatty Liver Disease. *Clinical Gastroenterology and Hepatology*. 2013;11(7):868-875.e863.
 48. Dallinga JW, Robroeks CMHHT, Van Berkel JJBN, et al. Volatile organic compounds in exhaled breath as a diagnostic tool for asthma in children. *Clinical & Experimental Allergy*. 2010;40(1):68-76.
 49. Garner CE, Smith S, de Lacy Costello B, et al. Volatile organic compounds from feces and their potential for diagnosis of gastrointestinal disease. *The FASEB Journal*. 2007;21(8):1675-1688.
 50. Wishart DS, Tzur D, Knox C, et al. Eicosane. *Human Metabolome Database (HMDB)*, 2021; <https://hmdb.ca/metabolites/HMDB0059909>.
 51. Wishart DS, Tzur D, Knox C, et al. Pentylbenzene. *Human Metabolome Database (HMDB)*, 2021; <https://hmdb.ca/metabolites/HMDB0059834#references>.
 52. Sobotka PA, Gupta DK, Lansky DM, Costanzo MR, Zarling EJ. Breath pentane is a marker of acute cardiac allograft rejection. *The Journal of Heart Lung Transplantation*. 1994;13(2):224-229.
 53. Weitz Z, Birnbaum A, Skosey J, Sobotka P, Zarling E. High breath pentane concentrations during acute myocardial infarction. *The Lancet*. 1991;337(8747):933-935.
 54. Phillips M, Sabas M, Greenberg J. Increased pentane and carbon disulfide in the breath of patients with schizophrenia. *Journal of Clinical Pathology*. 1993;46(9):861-864.
 55. Hietanen E, Bartsch H, Bereziat J, et al. Diet and oxidative stress in breast, colon and prostate cancer patients: a case-control study. *European Journal of Clinical Nutrition*. 1994;48(8):575-586.
 56. Butterfield DA, Koppal T, Howard B, et al. Structural and Functional Changes in Proteins Induced by Free Radical-mediated Oxidative Stress and Protective Action of the Antioxidants N-tert-Butyl- α -phenylnitron and Vitamin E a. *Annals of the New York Academy of Sciences*. 1998;854(1):448-462.
 57. Semenkovich CF. Insulin resistance and atherosclerosis. *The Journal of Clinical Investigation*. 2006;116(7):1813-1822.

58. de Lacy Costello B, Amann A, Al-Kateb H, et al. A review of the volatiles from the healthy human body. *Journal of Breath Research*. 2014;8(1):014001.
59. Silva CL, Passos M, Câmara JS. Solid phase microextraction, mass spectrometry and metabolomic approaches for detection of potential urinary cancer biomarkers--a powerful strategy for breast cancer diagnosis. *Talanta*. 2012;89:360-368.
60. De Preter V, Van Staeyen G, Esser D, Rutgeerts P, Verbeke K. Development of a screening method to determine the pattern of fermentation metabolites in faecal samples using on-line purge-and-trap gas chromatographic-mass spectrometric analysis. *Journal of Chromatography A*. 2009;1216(9):1476-1483.
61. Wang S, Chen Y, Gao Z, Xiong M, Zhong Z, Ye L. Gas chromatographic-mass spectrometric analysis of d-limonene in human plasma. *Journal of pharmaceutical and biomedical analysis*. 2007;44(5):1095-1099.
62. Gahleitner F, Guallar-Hoyas C, Beardsmore CS, Pandya HC, Thomas CP. Metabolomics pilot study to identify volatile organic compound markers of childhood asthma in exhaled breath. *Bioanalysis*. 2013;5(18):2239-2247.
63. Morisco F, Aprea E, Lembo V, et al. Rapid “Breath-Print” of Liver Cirrhosis by Proton Transfer Reaction Time-of-Flight Mass Spectrometry. A Pilot Study. *PLoS One*. 2013;8(4):e59658.
64. Dadamio J, Van den Velde S, Laleman W, et al. Breath biomarkers of liver cirrhosis. *Journal of Chromatography B*. 2012;905:17-22.
65. Friedman MI, Preti G, Deems RO, Friedman LS, Munoz SJ, Maddrey WC. Limonene in expired lung air of patients with liver disease. *Digestive Diseases Sciences*. 1994;39(8):1672-1676.
66. Miyazawa M, Shindo M, Shimada T. Metabolism of (+)-and (-)-limonenes to respective carveols and perillyl alcohols by CYP2C9 and CYP2C19 in human liver microsomes. *Drug Metabolism Disposition*. 2002;30(5):602-607.
67. Frye RF, Zgheib NK, Matzke GR, et al. Liver disease selectively modulates cytochrome P450-mediated metabolism. *Clinical Pharmacology Therapeutics*. 2006;80(3):235-245.
68. Moris D, Spartalis M, Tzatzaki E, et al. The role of reactive oxygen species in myocardial redox signaling and regulation. *Annals of Translational Medicine* 2017;5(16):324-324.
69. Ling XC, Kuo K-L. Oxidative stress in chronic kidney disease. *Renal Replacement Therapy*. 2018;4(1):53.
70. Ratcliffe N, Wiczorek T, Drabińska N, Gould O, Osborne A, De Lacy Costello B. A mechanistic study and review of volatile products from peroxidation of unsaturated fatty acids: an aid to understanding the origins of volatile organic compounds from the human body. *Journal of Breath Research*. 2020;14(3):034001.
71. Cleveland E, Bandy A, VanWagner LB. Diagnostic challenges of nonalcoholic fatty liver disease/nonalcoholic steatohepatitis. *Clinical Liver Disease*. 2018;11(4):98.