Diet-induced and mono-genetic obesity alter volatile organic compound signature in mice

Martin Kistler¹,²,⁵, Andreea Muntean³, Wilfried Szymczak³, Nadine Rink⁴, Helmut Fuchs¹,², Valerie Gailus-Durner¹,², Wolfgang Wurst⁶,⁷, Christoph Hoeschen³, Martin Klingenspor⁴, Martin Hrabě de Angelis¹,²,⁵,⁸, Jan Rozman¹,²,⁵,*

¹Institute of Experimental Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstrasse 1, 85764 Neuherberg, Munich, Germany
²German Mouse Clinic, Institute of Experimental Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstrasse 1, 85764 Neuherberg, Munich, Germany
³Research Unit Medical Radiation Physics and Diagnostics, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstrasse 1, 85764 Neuherberg, Munich, Germany
⁴ZIEL Department of Molecular Nutritional Medicine, Else Kröner-Fresenius Center, Technische Universität München, 85350 Freising, Germany
⁵German Center for Diabetes Research (DZD)
⁶Institute of Developmental Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstrasse 1, 85764 Neuherberg, Munich, Germany
⁷Developmental Genetics, Centre of Life and Food Sciences Weihenstephan, Technische Universität München, Freising, Germany
⁸Chair for Experimental Genetics, Life and Food Science Center Weihenstephan, Technische Universität Munich, D-85354 Freising-Weihenstephan, Germany

*Corresponding author. phone: +49 89 3187 3807; E-mail: jan.rozman@helmholtz-muenchen.de (J. Rozman), Postal: Institute of Experimental Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstrasse 1, 85764 Neuherberg, Munich, Germany

Abbreviations: VOCs, volatile organic compounds; HFD, high fat diet; LFD low fat diet; AUC, area under the curve; RF, random forest; ROC, receiver operating characteristic; FDR, false discovery rate; PTR-MS, proton transfer reaction mass spectrometry; TOF, time of flight; MC4R, melanocortin 4 receptor; MC4R-ki, MC4R W16X knock in; qNMR, quantitative nuclear magnetic resonance pk, peak; gt, genotype; adlib, ad libitum; MTMT, (methylthio)methanethiol; DMS, dimethyl sulfide; OGTT oral glucose tolerance test.
Abstract

Objective: The prevalence of obesity is still rising in many countries of the world and the multitude of consequential metabolic dysregulations such as type II diabetes mellitus that occur in patients exacerbate the impact on quality of life and public health systems. The analysis of volatile organic compounds (VOC) in breath that originate from multiple metabolic pathways provides an extraordinary potential to identify or monitor obese patients with increased risk profile for associated later diseases. In this study we aimed to describe the VOC patterns symptomatic for both general and model-specific obesity by analyzing samples of exhaled breath in diet-induced obese and mono-genetic obese mice.

Methods: We induced obesity by feeding a high fat diet to male C57BL/6J mice for 12 weeks (HFD). In addition, we analyzed male C57BL/6J mice carrying a global knock-in mutation in melanocortin-4 receptor (W16X, MC4R-ki). In both experimental groups, the source strengths of volatile organic compounds were analyzed ad libitum fed and after overnight food restriction. Volatiles altered in obese mice were selected using the AUC-RF algorithm and tested using false discovery rate-controlled mixed effects model. A Gaussian graphical model was employed to identify chemical and metabolic links among the selected volatiles.

Results: In both models for obesity, volatiles relevant for the separation of obese and lean mice were detected (26 in MC4R-ki, 22 in HFD mice). Eight volatiles were found to be important in both obesity models. Interestingly, by creating a partial correlation network of the volatile metabolites, the chemical and metabolic origins of several volatiles were identified. HFD-induced obese mice showed an elevation in the ketone body acetone and acrolein, a marker of lipid peroxidation, and several unidentified volatiles. In MC4R-ki mice, several yet-unidentified VOCs were found to be altered. Remarkably, the pheromone (methylthio)methanethiol was found to be reduced, linking metabolic dysfunction and reproduction.

Conclusions: The signature of volatile metabolites can be instrumental to identify and monitor metabolic disease states, as shown in this screening of two obese mouse models. Our findings show the potential of breath gas analysis to non-invasively assess metabolic alterations for personalized diagnosis. Furthermore, breath gas analysis could aid in the stratification of patients with heterogeneous metabolic phenotypes and risk profiles.
Graphical abstract

Keywords
High-fat diet; melanocortin 4 receptor; volatile organic compound; Gaussian graphical model; mouse pheromone; non-invasive metabolic phenotyping;

Highlights
- Both common and specific VOCs patterns emitted by HFD and MC4R mice
- HFD: Altered oxidative stress-associated and ketone body volatiles
- MC4R: MTMT reduction and a cluster altered likely consisting of (mono-)terpenes
- Gaussian graphical model is a valuable tool to detect VOC identity and pathways
1. **Introduction**

Obesity has progressed to a world-wide epidemic linked to a number of co-morbidities such as diabetes, cardiovascular disease, dyslipidemia and certain types of cancers (Guh et al, 2009). Easily accessible biomarkers are central for the assessment of individual risks of patients to suffer from such pathologies and to develop personalized medicine approaches for prevention and treatment. The dysregulation of metabolic pathways and the associated changes in body fluid metabolite concentrations are increasingly studied and also used for risk prediction (Mahendran et al, 2013; Elliott et al, 2015; Wahl et al, 2015). A variety of normal and disease-associated metabolic reactions produce small volatile organic compounds (VOCs) which can be detected in body fluids but also non-invasively in exhaled breath. Over the past decade, advances in the methodology made it possible to determine VOCs online in a concentration range of ppm to ppt and led to studies linking VOC signatures to various pathologies (Boots et al, 2012). Regarding diseases associated with energy metabolism, several human studies were conducted trying to monitor glucose levels (Lee et al, 2009; Minh et al, 2011), identify gestational, type1, or type 2 diabetes (Halbritter et al, 2012; Novak et al, 2007; Greiter et al, 2010), and characterize non-alcoholic fatty liver disease and liver cirrhosis (Morisco et al, 2013; Alkhouri et al, 2014). Prerequisites for broader clinical application are (I) clear identification of the molecules that are exhaled as VOCs and (II) a better understanding of the considerable inter- and intra-individual variation in VOCs found in breath even in healthy humans (Phillips et al, 1999; Basanta et al, 2012; Martinez-Lozano Sinues et al, 2014). Environmental “wash-in”, the diet and associated microbial changes as well as circadian rhythm might increase this variation. Animal models, and especially rodent models, for human diseases are a tremendously valuable tool to deepen the understanding of molecular mechanisms and decipher the various sources of volatiles in a controlled environment (Rosenthal & Brown, 2007). In this study, we were particularly interested in differences in VOC signatures of normal weight mice and mice with manifested obesity induced either by feeding a high fat diet (HFD) or induced by targeted loss of function mutation of the melanocortin 4 receptor (Bolze et al, 2011). VOC signatures could reflect the degree of obesity but could also be due to changes in diet composition which is a confounding interaction that cannot be avoided in diet induced obesity studies (Baranska et al, 2013; Kistler et al, 2014). In contrast to diet-induced modifications of VOC signatures, to our knowledge no investigation on the effects of genetically-induced obesity on the volatilome, defined as the total amount of all VOCs emitted, was conducted so far. Therefore, the aim of this study was to characterize alterations in exhaled volatile organic compounds both in a diet-induced and a mono-genetic obese mouse model and to evaluate whether a symptomatic pattern of VOCs related to
obesity can be determined. In addition, individual changes in the volatilome of two specific obese models with distinct metabolic deregulations are of interest. We employed statistical analysis methods that identified typical correlations between VOC emission rates, in the following called source strengths that could be used to unravel the biochemical origin of the respective molecules.

2. Material and Methods

2.1. Mice, animal housing and challenge experiments

Mice were housed in in type II L polycarbonate cages in individually ventilated cages (Tecniplast, Italy). A 12:12h light/dark cycle at a temperature of 24 ± 1 degree Celsius and air humidity of 50 – 60% were maintained. Animals were housed in groups of 2 to five animals per cage in specific pathogen-free conditions in the German Mouse Clinic (GMC) (Fuchs et al, 2009). Wood shavings were used for bedding (Altromin GmbH, Germany). For the generation of the diet-induced obesity model, 20 male C57BL/6J mice from in-house breeding were fed a pelleted laboratory chow from weaning on with ad libitum access to food and drinking water (no. 1314, Altromin, Lage, Germany).

From the age of twelve weeks until the start of the VOC measurement (24± 2 weeks), the diet was changed to pelleted purified low fat and high fat diets (low fat: E 15000-04; high fat: E 15741-34; both: Ssniff, Soest, Germany). Assignment to diet groups was performed randomly using existing cage stocking to avoid single housing while ensuring balanced group numbers. A mono-genetic hyperphagic obesity model having a melanocortin-4-receptor nonsense allele W16X was used (Mc4r-ki mouse, as previously published (Bolze et al, 2011)). 15 homozygous MC4R-ki BL6/J mice as well as 15 controls were transferred to the GMC from the provider’s lab at the age of 5 weeks and analyzed at the age of 24± 2 weeks. Mice had ad libitum access to drinking water and a pelleted laboratory chow from weaning on (<5 weeks: “RM-Z autoklavierbar”, ssniff; >5 weeks: no. 1314, Altromin, Lage, Germany). All experiments were performed following animal welfare regulations with permission from the district government of Upper Bavaria (Regierung von Oberbayern).

For the analysis of VOCs from ad libitum fed mice, gas measurements took place between 1 pm and 6 pm. During this time, food consumption is low compared to nighttime. Therefore, this period was chosen to reduce contribution of food-derived volatiles to measured VOC patterns. Mice were measured in random order and alternating between control and obese mice to remove potential systemic bias. For the fasted VOC measurements, mice were food deprived overnight beginning
around 5-6 pm and were measured in the same order as in ad libitum state between 8 to 12 am the following day. Mice were weighed before every VOC measurement to the nearest 0.1 g and body composition was monitored by non-invasive qNMR scans in ad libitum fed state (Bruker Minispec LF50 body composition analyser, Ettlingen, Germany). The comparisons of body, lean and fat mass between groups were performed using a linear regression model.

2.2. Proton-transfer reaction time-of-flight mass spectrometry and protocol for real-time breath gas analysis in unrestrained mice

A high-sensitivity Proton Transfer Reaction Mass Spectrometer (PTR-MS, e.g. benzene 100 cps/ppbV; PTR-MS, Ionicon Analytic GmbH, Innsbruck, Austria) with a resolution of $\Delta m/m \leq 2000$ was used. The principle of PTR mass spectrometry using $\text{H}_3\text{O}^+$ ions to softly ionize and detect VOCs was developed in the late 1990s (Lindinger et al, 1998; Petersson et al, 2009). A drift tube temperature of 80 degrees Celsius, a drift tube voltage of 600 V and a drift pressure 2.3 mbar were applied. A mass range from $m/z$ 0 to 349.5 was recorded (repetition rate of 77 kHz); the sum spectra with integration time of 3 s were stored (TOF-DAQ, Tofwerk AG, Switzerland). For the integration of peaks from the TOF-spectra the software PTR-MS Viewer was used (Version 3.2.6, Ionicon analytic GmbH, Innsbruck, Austria). An internal calibration with the known peaks $\text{H}_3^{18}\text{O}^+$ ($m/z$ 21.0221), $\text{NO}^+$ ($m/z$ 29.9971) and protonated acetone ($m/z$ 59.0491, $\text{C}_3\text{H}_6\text{O.H}^+$) was performed. 306 peaks were selected manually from the spectra. The deconvolution of overlapping peaks was performed fitting a gaussian distribution to the peaks in the PTR-MS Viewer. VOC concentrations were calculated using a constant $k$-rate of $2 \times 10^{-9} \text{[cm}^3 \text{ s}^{-1}]$ in the semi-quantitative estimation formula (Lindinger et al, 1998). The system sensitivity was controlled using a gas calibration unit (GCU, Ionicon Analytic GmbH, Innsbruck, Austria) with a mixture of substances (VOC gas standard, Ionicon Analytic GmbH, Innsbruck, Austria) regularly. From a set of compounds, a linear calibration curve obtained from multiple concentrations was used to calculate the individual transmission factors.

A setup and protocol for real-time measurement of breath gas analysis in unrestrained mice was used as described previously (Szymczak et al, 2014; Kistler et al, 2014). In brief, mice were acclimatized to a training respiratory chamber for 7 minutes. A measurement chamber connected to the mass spectrometer is flushed 2 minutes with a flow rate 3 l/min to dilute enclosed laboratory room air. After flushing, the VOCs from the empty respiratory chamber are measured as a blank (5 min, flow 60 mL min-1) to the detect system leakage and background VOCs. Signals monitored for leakage from laboratory air are acetone ($m/z = 59.05$) and propanol ($m/z = 41.06$) concentration.
Following this blank measurement, the system is switched to flushing state and the mouse is placed into the chamber. After the 2 min flushing, measurement phases alternate with flushing of the chamber, allowing the volatiles to accumulate in the gaseous phase.

During accumulation of VOCs in the headspace of unrestrained and non-anaesthetized mice, contaminations arise e.g. from urination and defecation. The measurement chamber was monitored for signs of contamination repeatedly during measurements. In addition, several volatiles were used as marker substances for urine or feces. Urine was detected online due to sudden changes in humidity (determined as water-cluster \((H_2^{18}O)_2H^+, m/z = 39.05\)), concentration of trimethylamine \((m/z = 60.07)\) and pk127B (tentatively dimethyl trisulfide, \(m/z = 127.02\)). Concentration of methanethiol \((m/z = 49.02)\) indicated presence of feces. In case of feces contamination, feces were removed and accumulation phase was skipped. If urine was present, the respirometry chamber was replaced, the mouse was gently cleaned using soft tissue paper (Kimtech Science, Kimberly-Clark) and the measurement was restarted.

### 2.3. Data analysis and statistics

#### 2.3.1. Calculation of source strength and data pre-processing

Measurement start and end were defined manually using an in-house web-application based on R and shiny package (R Core Team (2014); Chang et al, 2015). A compartment model was used to describe the emission of a certain peak from recorded saturation curves (non-linear regression, described in (Szymczak et al, 2014)). This model resulted in a source strength [ppb*ml/min], which we used for further analyses. The information of group membership was not recorded in raw data files but added later on to ensure fully blinded analysis of saturation curves. As further contamination control step, data was filtered for high concentrations or sudden increases of urinary and feces markers (pk127B, pk60 and pk49 > 1 ppb). The individual source strengths (1-5 per mouse and feeding state) were differentially corrected against the respective blank source strength of an empty box to account for possible micro-leakage or background system emission. For every peak, outliers (defined as greater 5 standard deviations from mean) were removed. Peak data and single measurement data with more than 10% missing values were excluded. Peaks with source strengths not different or lower compared to corresponding blank source strengths were excluded as well (linear regression modelling with \(p < 0.1\) to ensure enclosure of low signal candidate VOCs). An
exception was made for known oxygen isotopes, as the negative source strength (=consumption) is expected. This filter steps resulted in a final group sizes of 15 for fasted MC4R-wt mice, 14 for fasted MC4R-ki mice, 11 for ad libitum fed MC4R-wt mice and 9 for every other experimental group. As a complete data-matrix is required to calculate random forest and gaussian graphical models, missing data was imputed using chained equations (mice R package (van Buuren & Groothuis-Oudshoorn, 2011)), which accounted for 0.29% of data.

2.3.2. **Feature selection and statistical testing of individual VOCs**

Using a time-of-flight mass spectrometric detection of volatiles, the dataset consisted of a large number of peaks relative to animal numbers. We applied the AUC-RF (Area under the curve – random forest) algorithm as recently published (Urrea & Calle, 2012) to find a reduced set of candidate volatiles. In this algorithm, an initial random forest is computed to obtain a ranking of predictors and an area-under-the receiver operating characteristic (ROC) curve. During the elimination process, less important variables are removed and AUCs of the resulting RFs are computed; an optimal set of predictors based on the AUC is finally reported. We used this algorithm for both HFD-fed and MC4R-ki datasets independently; setting strata to allow only one measurement per mouse and fasting status in every decision tree. A five-fold cross validation was applied 20 times to avoid over-fitting of each of the resulting RF model (using again a modified version of the algorithm allowing for stratification). Peaks with a selection probability higher than 70% in the cross-validation AUC-RFs were used for further analysis.

For the analysis of genotype-induced and diet-induced effects on the VOC source strengths, two-sided mixed effects models were applied (Pinheiro et al, 2015). Both diet and genotype subsets of data were log-transformed to approximate a normal distribution (tested visually by qq-plotting). The variance between groups was controlled using both boxplots of source strength as well as residuals and residual versus fitted data plots. For every peak of both subsets of data, effects of the corresponding intervention variable (diet respectively genotype), the fasting status as well as the interaction of both were tested using a mixed effects model accounting for repeated measures. If a significant interaction could be detected, individual group comparisons were performed using multcomp r package (Hothorn et al, 2014). As a larger number of tests leads to summation of Type I Error, control of false discovery rate after Benjamini and Hochberg (Benjamini & Hochberg, 1995) was applied and all p-values were adjusted according to a 10% FDR.
2.3.3. **Data visualization and VOC identification using Gaussian graphical modelling**

2.3.3.1. Heatmaps and boxplots

Both sets of data were visualized in a clustered heatmap using the Heatplus (Ploner, 2014) package from Bioconductor (Gentleman *et al.*, 2004). Mean ad libitum fed as well as mean fasted source strength data per mouse was used and shown individually. Boxplots were created using the R package ggplot2 using all repeatedly measured source strength data (Wickham, 2009).

2.3.3.2. Gaussian graphical model

We applied a gaussian graphical model to log-transformed source strengths of breath volatiles to visualize information about fragmentation, isotopic, water cluster and/or metabolic correlations. The complete dataset features more variables than number of mice, therefore we used a shrinkage approach to estimate a partial correlation matrix (Schäfer & Strimmer, 2005). As the data is of longitudinal structure, we created a network accounting for that using dynamic (partial) correlation (Opgen-Rhein & Strimmer, 2006). A network was extracted from the estimated partial correlation matrix using a local false discovery rate of 3% (GeneNet R package (Schaefer *et al.*, 2015). A “dummy” variable to correct for inter-experimental differences between HFD fed and MC4R-ki mice was included in the network but not plotted. For every peak within a selected subset with significant fasting state, genotype or diet effect, the percentaged coefficients from mixed effects model are shown in the nodes as a pie-chart. Top 20% of connections are shown with bold lines; minor 20% with grey lines, negative partial correlations with dotted lines. Direct positive connections of significant nodes were highlighted and combined for overlapping subnetworks containing multiple significant nodes. Peaks included in AUC-RF model data-subsets but without significant connections were included in the graphical model for illustration purposes.

2.3.4. **Data availability**

Data is accessible as a supplementary document.
### Results

#### 3.1. Obesity state of high-fat diet fed and mono-genetic mice

**Figure 1:** Body mass, lean mass and fat mass. Body mass (A), lean mass (B) and fat mass (C) for High Fat Diet fed (HFD, dark red) and melanocortin-4-receptor W16X knock-in (MC4R-ki, dark blue) mice as well as corresponding controls (Low-fat diet LFD, red; melanocortin-4-receptor wild type, MC4R-wt, blue) is shown in boxplots. Significant differences between controls and respective obesity mouse models are shown as black lines over individual boxes (linear regression model, p<0.05). Group sizes: MC4R-wt ad lib (n=11), other groups (n=9).

Both HFD and the MC4R-ki mutation resulted in clear states of obesity as was evident from increased body mass as well as lean and fat mass. HFD fed mice were heavier compared to littermate controls (44.21 ± 4.17 g vs 27.94 ± 1.47 g, p=4.01*10^{-8}, Fig. 1A). This gain in mass was partly due to an increase in lean mass (24.64 ± 1.7 g vs 18.47 ± 1.08 g, p=2.56*10^{-7}, Fig. 1 B) as well as an increase in fat mass (16.83 ± 2.88 g vs 5.69 ± 0.65 g, p=2.16*10^{-8}, Fig. 1 C). For MC4R-ki mice, also a considerable difference in body mass compared to littermate controls could be detected (50.99 ± 3.19 g vs 28.01 ± 1.28 g, p=2.00*10^{-14}, Fig. 1 A). This difference was in part attributed to lean mass (26.28 ± 1.4 g vs 17.8 ± 0.64 g, p=5.67*10^{-13}, Fig. 1 B), but largely due to elevated fat mass (21.3 ± 2.09 g vs 6.2 ± 0.83 g, p=1.77*10^{-14}, Fig. 1 C). Overall, the impact of the MC4R-ki on body mass and body composition was more pronounced compared to the HFD model.

#### 3.2. Selection of VOCs relevant for classification
Figure 2: Variable importance for peaks selected for alterations in obesity models. Variable importance recursive feature selection using AUC-OOB of random forest models in high fat diet fed mice (HFD, A) and melanocortin-4-receptor W16X knock-in (MC4R-ki, B). Color gradients indicate selection probability after 20 iterations of a five-fold cross validation procedure. Peaks with more than 70% selection probability are shown here and were selected for further analysis. Overlapping of selected peaks is shown as Venn diagram (C).

For both obesity models, a feature selection was performed to classify obesity biomarkers by optimization of the ROC area under the curve in a series of random forest models. A cutoff of at least 70% selection probability was used to select 22 candidate peaks with the highest variable importance in HFD mice (Fig. 2A). For MC4R-ki mice, 26 candidate peaks fulfilled the cutoff criterion (Fig. 2B). Interestingly, within these peaks with the highest classification importance an overlap of 8
peaks could be detected between the two mouse models (Fig. 2C). The eight peaks present in both groups were pk33B (methanol), pk50 (unassigned), pk61 (acetic acid), pk62 (MTMT), pk63 (CO2, DMS), pk65B (CO2, DMS isotopes), pk81B (unassigned) and pk117B (unassigned).

3.3. Visualization of selected source strength data

**Figure 3: Heatmaps of selected VOCs.** Heatmap of selected peaks in HFD fed (A) and MC4R-ki mice (B) are shown with hierarchical clustering of individual mice (mean data, rows, sub-clusters colored) and VOC peaks (columns, labels according to nominal mass). Data is scaled and centered. Color-coding legend shown on the left. Classification of individual mice is
annotated on the right (A: diet = LFD or HFD; B: gt = MC4R-wt or MC4R-ki; both feed = ad libitum fed or fasted, body mass [g], subcluster-membership colored). Group sizes: MC4R-ki fasted (n=15), MC4R-wt fasted (n=14), MC4R-wt ad lib (n=11), other groups (n=9).

Heatmaps consisting of RF-selected peaks for both models were created to get further insight into data structure by using unsupervised hierarchical clustering (Fig. 3). HFD fed mice clustered in the top half of the heatmap with a remote subgroup within the dark blue sub cluster (Fig. 3A). Interestingly, despite the selection for obesity relevant peaks, a clustering according to fasting status was observed (fasted within light red, light green and dark blue, predominantly). Contrary to the findings in HFD fed mice, the feeding status seemed to be the dominant clustering principle with MC4R wild type and knock-in mice showing fasted mice in the light green, dark blue and light blue sub clusters (Fig. 3B). Notably, fasted MC4R-ki mice clustered together mostly in the dark blue subcluster, whereas ad libitum fed knock-in mice showed a weaker clustering in the “warm colored” sub clusters.

3.4. Effects on VOC signature in diet-induced and monogenetic obesity
Figure 4: VOC source strengths affected by high fat diet or fasting. Source strengths for nominal mass-labelled peaks 18 (A), 57B (B), 65B (C), 63 (D), 75 (E), 62 (F), 50 (G), 249B (H), 253A (I), 117B (J), 57A (K), 34A (L), 81B (M), 74B (N), 55B (O), 36B (P), 43B (Q), 61 (R), 43A (S), 33B (T), 59 (U) and 64B (V) are shown as boxplots (ordered after selection probability in cross-validated AUC-RF algorithm). Box fill corresponds to diet (red: low fat diet; dark red: high fat diet). Box border corresponds to fasting state (black: ad libitum fed; grey: fasted). Significant main effects in mixed effects model are shown as dotted.
In HFD mice several VOC source strengths were affected as shown by linear mixed effects modelling for diet and fasting effects (Fig. 4, Supplementary table 1). A significant increase in source strength could be found in eleven peaks. Those peaks are 65B (unassigned, Fig. 4 C), 63 (CO₂*H₂O/ DMS, Fig. 4 D), 75 (methyl acetate, Fig. 4 E), 253A (unassigned, Fig. 4 I), 117B (unassigned, Fig. 4 J), 57A (unassigned, Fig. 4 K), 81B (unassigned, Fig. 4 M), 74B (unassigned, Fig. 4 N), 55B (H₃O⁺.H₂O/ C₄H₆.H⁺, Fig. 4 O), 43A (C₂H₂O.H⁺, Fig. 4 S), 59 (acetone, Fig. 4 U) and 64B (¹³CO₂.H₃O⁺/¹³CCH₆S.H⁺, Fig. 4 V). Source strength in peak 34A (¹⁶O¹⁸O, Fig. 4 L) was decreased in diet-induced obese mice.

In the HFD model, several volatiles were affected by the fasting status of the mouse. Fasting induced higher emitted source strength in the seven peaks: 50 (unassigned, Fig. 4 G), 55B (H₃O⁺.H₂O/ C₄H₆.H⁺, Fig. 4 O), 43B (C₃H₆.H⁺, Fig. 4 Q), 43A (C₂H₂O.H⁺, Fig. 4 S), 33B (methanol, Fig. 4 T), 59 (acetone, Fig. 4 U) and 64B (¹³CO₂.H₃O⁺/¹³CCH₆S.H⁺, fig. 4 V). Three volatiles were reduced after overnight food restriction: 249B (unassigned, Fig. 4 H), 34A (¹⁷O₂, Fig. 4 L) and 81B (unassigned, Fig. 4 M).

In four peaks, an interaction of HFD feeding and food restriction was present. Ammonia (Fig. 4A, pk18) was decreased in HFD mice (with a larger decrease in source strengths in the ad libitum fed state) and upon overnight fasting. Acrolein (Fig 4B, pk57B (2-propenal, C₃H₄O.H⁺)) was elevated in obese mice in both states. Upon fasting source strength was increased in wild type mice but decreased in HFD fed animals. Peak 62 was assigned to a thiol-loss fragment of (Methylthio)methanethiol (MTMT, CH₃SCH₂.H⁺) as described in (Da Yu Lin et al, 2005) (Fig. 4F). MTMT was increased upon fasting in both groups. Furthermore, ad libitum fed HFD mice showed higher source strength of MTMT compared to LFD fed mice, an effect which was no longer present when fasted mice were measured in the morning. pk36B was decreased in HFD ad libitum fed mice and fasted LFD fed mice (Fig 4P, unassigned).
Figure 5: VOC source strengths affected by genotype or fasting. Source strengths for nominal mass-labelled peaks 63 (A), 62 (B), 81B (C), 65B (D), 123B (E), 151 (F), 153 (G), 50 (H), 135B (I), 61 (J), 36A (K), 122B (L), 117B (M), 137B (N), 44A (O), 113 (P), 94B (Q), 35B (R), 80B (S), 126A (T), 87B (U), 33B (V), 109A (W), 95B (X), 211 (Y) and 48B (Z) shown as boxplots.
In MC4R-ki mice several VOC source strengths differed between genotypes as shown by linear mixed effects modelling (Fig. 5, detailed model results in Supplementary table 1). A significant increase was found in eight volatiles: peak 63 (CO2*H2O/ DMS, Fig. 5 A), 81B (unassigned, Fig. 5 C), 151 (unassigned, Fig. 5 F), 135B (unassigned, Fig. 5 I), 122B (unassigned, Fig. 5 L), 137B (unassigned, Fig. 5 N), 44A (C,H,O / 13CCH2O.H+, Fig. 5 O) and 95B (unassigned, Fig. 5 X). In five volatiles, a decrease in source strength was observed in MC4R-ki mice: peak 62 (MTMT, Fig. 5 B), 50 (unassigned, Fig. 5 H), 36A (unassigned, Fig. 5 K), 80B (unassigned, Fig. 5 S) and 109A (unassigned, Fig. 5 W).

In addition to genotype effects, overnight food restriction affected four peaks positively and eleven peaks negatively. An increase in fasted state was observed in peaks 62 (MTMT, Fig. 5 B), 44A (C,H,O / 13CCH2O.H+, Fig. 5 O), 87B (unassigned, Fig. 5 U) and 95B (unassigned, Fig. 5 X). Decreased fasting source strengths were found for peaks 63 (CO2*H2O/ DMS, Fig. 5 A), 81B (unassigned, Fig. 5 C), 123B (unassigned, Fig. 5 E), 61 (acetic acid, Fig. 5 J), 36A (unassigned, Fig. 5 K), 117B (unassigned, Fig. 5 M), 137B (unassigned, Fig. 5 N), 113 (unassigned, Fig. 5 P), 94B (unassigned, Fig. 5 Q), 35B (CH16OH.H+, Fig. 5 R) and 211 (unassigned, Fig. 5 Y).

An interaction of MC4R-ki genotype and food restriction was present in peak 65B (Fig. 5D, unassigned). Here, source strength was increased in fasted but not in ad libitum fed MC4R-ki mice and reduced in wild type mice in response to fasting.

3.5. Gaussian graphical modelling as a tool to identify VOCs
Gaussian graphical models were proposed recently to identify metabolites and model metabolic pathways from metabolomics data (Krumsiek et al, 2012, 2011). As in a large p, smaller n dataset a
full partial correlation matrix cannot be directly applied, we estimated a partial correlation network using a shrinkage approach for longitudinal data (Schäfer & Strimmer, 2005; Opgen-Rhein & Strimmer, 2006). In this graphical model, several obesity relevant subnetworks can be detected with additional information on VOC identity (Fig. 6). In addition to genotype and diet effects, fasting effects are visualized in grey (increase) and dark grey (decrease) pie slices. If both cohorts showed fasting coefficients, we calculated mean coefficients. Fasting coefficients from MC4R cohort contribute to peaks 33B, 35B, 36A, 44A, 61, 62, 63, 65B, 80B, 81B, 87B, 94B, 95B, 109A, 113, 117B, 123B, 137B, 153 and 211 while fasting coefficients from HFD models are contributing to 18, 33B, 34A, 36B, 43B, 50, 55B, 57B, 59, 62, 64B, 81B and 249B. Interpretations of peak-peak-connections are given in table 1.

Table 1: Tentative assignment of peaks with significant diet, genotype or fasting effects and interpretation of partial correlations to other peaks as depicted in the gaussian graphical model (Fig. 6).

<table>
<thead>
<tr>
<th>Subnet</th>
<th>peak 1</th>
<th>formula / assignment(s) (mass)</th>
<th>prot. mass</th>
<th>peak 2</th>
<th>formula / assignment(s) (mass)</th>
<th>prot. mass</th>
<th>partial cor.</th>
<th>Interpretation/pathway</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 / red</td>
<td>18</td>
<td>NH₃.H⁺ Ammonia (18.03)</td>
<td>18.03</td>
<td>46C</td>
<td>C₅H₇N structure: CH₃CH₇(NH₃.H⁺)/ Ethyamine (46.07)</td>
<td>46.06</td>
<td>0.086</td>
<td>(bio)chemical reaction</td>
</tr>
<tr>
<td></td>
<td>63</td>
<td>CO₂.H₂O⁺ / Carbon dioxide water cluster (63.00)</td>
<td>63.01</td>
<td>64B</td>
<td>¹³CO₂.H₂O⁺ / Carbon dioxide water cluster (64.00)</td>
<td>64.01</td>
<td>0.081</td>
<td>C-13 Carbon isotope</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CH₅SH₂.H⁺ / Dimethyl sulfide (63.02)</td>
<td>62.02</td>
<td>44B</td>
<td>CH₅¹³CH₂.H⁺ / propanol fragment (44.07)</td>
<td>44.07</td>
<td>0.077</td>
<td>Both increase in fasting/morning</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CH₅¹³CH₂.H⁺ / MTMT fragment (64.02)</td>
<td>62.02</td>
<td>44B</td>
<td>CH₅¹³CH₂.H⁺ / propanol fragment (44.07)</td>
<td>44.07</td>
<td>0.074</td>
<td>S-34 Sulfur isotope</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CH₃O.H⁺ / Acetaldehyde (45.03)</td>
<td>45.05</td>
<td>0.078</td>
<td>Unknown/ (Bio) chemical reaction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CH₃H₂.H⁺ / Propane (45.07)</td>
<td>45.05</td>
<td>0.078</td>
<td>Unknown/ (Bio) chemical reaction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>¹²CC₅H₄.H⁺ / propanol fragment (42.05)</td>
<td>42.05</td>
<td>0.081</td>
<td>Carbon isotope of other propanol fragment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>¹²CC₅H₄.H⁺ / propanol fragment (42.05)</td>
<td>42.05</td>
<td>0.081</td>
<td>Carbon isotope of other propanol fragment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>¹²CC₅H₄.H⁺ / propanol fragment (42.05)</td>
<td>42.05</td>
<td>0.081</td>
<td>Carbon isotope of other propanol fragment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>¹²CC₅H₄.H⁺ / propanol fragment (42.05)</td>
<td>42.05</td>
<td>0.081</td>
<td>Carbon isotope of other propanol fragment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>¹²CC₅H₄.H⁺ / propanol fragment (42.05)</td>
<td>42.05</td>
<td>0.081</td>
<td>Carbon isotope of other propanol fragment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>¹²CC₅H₄.H⁺ / propanol fragment (42.05)</td>
<td>42.05</td>
<td>0.081</td>
<td>Carbon isotope of other propanol fragment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>¹²CC₅H₄.H⁺ / propanol fragment (42.05)</td>
<td>42.05</td>
<td>0.081</td>
<td>Carbon isotope of other propanol fragment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>¹²CC₅H₄.H⁺ / propanol fragment (42.05)</td>
<td>42.05</td>
<td>0.081</td>
<td>Carbon isotope of other propanol fragment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>¹²CC₅H₄.H⁺ / propanol fragment (42.05)</td>
<td>42.05</td>
<td>0.081</td>
<td>Carbon isotope of other propanol fragment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 / orange</td>
<td>43B</td>
<td>C₅H₆.O.H⁺(43.06) / propanol fragment (after –OH loss)</td>
<td>43.07</td>
<td>60B</td>
<td>¹²CC₅H₆.O.H⁺ / acetone (60.05)</td>
<td>60.05</td>
<td>0.077</td>
<td>C-13 Carbon isotope</td>
</tr>
<tr>
<td></td>
<td></td>
<td>¹²CC₅H₆.O.H⁺ / acetone (60.05)</td>
<td>60.05</td>
<td>0.077</td>
<td>C-13 Carbon isotope</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CH₃(SCH₂)H⁺ / (Methylthio) methanethiol fragment (62.02)</td>
<td>62.02</td>
<td>0.083</td>
<td>Both increase in fasting/morning</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CH₃(SCH₂)H⁺ / (Methylthio) methanethiol fragment (62.02)</td>
<td>62.02</td>
<td>0.083</td>
<td>Both increase in fasting/morning</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>⁴¹⁴CH₅H₂.H⁺ / propadiene, propanol fragment (41.04)</td>
<td>41.05</td>
<td>0.204</td>
<td>Alternative propanol fragment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>⁴¹⁴CH₅H₂.H⁺ / propadiene, propanol fragment (41.04)</td>
<td>41.05</td>
<td>0.204</td>
<td>Alternative propanol fragment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>⁴¹⁴CH₅H₂.H⁺ / propadiene, propanol fragment (41.04)</td>
<td>41.05</td>
<td>0.204</td>
<td>Alternative propanol fragment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>⁴¹⁴CH₅H₂.H⁺ / propadiene, propanol fragment (41.04)</td>
<td>41.05</td>
<td>0.204</td>
<td>Alternative propanol fragment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>⁴¹⁴CH₅H₂.H⁺ / propadiene, propanol fragment (41.04)</td>
<td>41.05</td>
<td>0.204</td>
<td>Alternative propanol fragment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44B</td>
<td>( ^{13}\text{C}\text{C}_2\text{H}_4\text{H}^- ) / propanol fragment (44.07)</td>
<td>44.07</td>
<td>0.094</td>
<td>(Bio) chemical reaction catalyzed by alcohol dehydrogenase (adh1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45B</td>
<td>( \text{C}_3\text{H}_7\text{O}_2\text{H}^- ) / acetaldehyde (45.03)</td>
<td>45.05</td>
<td>0.096</td>
<td>(Bio) chemical reaction/unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45B</td>
<td>( \text{C}_3\text{H}_7\text{H}^- ) / propane (45.07)</td>
<td>45.05</td>
<td>0.096</td>
<td>(Bio) chemical reaction/unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43A</td>
<td>( \text{C}_2\text{H}_3\text{O}^+ ) / acylium ion (43.02)</td>
<td>43.03</td>
<td>0.087</td>
<td>Acetate fragment ((-\text{H}_2\text{O})) loss</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>( \text{C}_3\text{H}_6\text{O}_2\text{H}^- ) / acetic acid (61.03)</td>
<td>61.03</td>
<td>0.090</td>
<td>Both high in chow ad libitum / methyl group fragmentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>( \text{C}_2\text{H}_4\text{O}^+ ) / aldehyde (75.03)</td>
<td>75.03</td>
<td>0.090</td>
<td>Acetate fragment ((-\text{H}_2\text{O})) loss</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>124A</td>
<td>Not assigned</td>
<td>123.84</td>
<td>0.074</td>
<td>Relation unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>126A</td>
<td>Not assigned</td>
<td>126.05</td>
<td>0.078</td>
<td>Relation unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>138B</td>
<td>Not assigned</td>
<td>138.01</td>
<td>0.080</td>
<td>Relation unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>142B</td>
<td>Not assigned</td>
<td>142.05</td>
<td>0.076</td>
<td>Relation unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>146A</td>
<td>Not assigned</td>
<td>146.03</td>
<td>0.076</td>
<td>Relation unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>151</td>
<td>Not assigned</td>
<td>150.99</td>
<td>0.099</td>
<td>Relation unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**
- **44B** propanol fragment
- **45B** acetaldehyde
- **43A** acylium ion
- **61** acetic acid
- **75** acetaldehyde
- **124A** \( \text{C}_10\text{H}_16\text{H}^+ \) / Monoterpenes
- **138B** \( \text{C}_10\text{H}_18\text{H}^+ \) / Monoterpenes
- **151** \( \text{C}_10\text{H}_19\text{H}^+ \) / Monoterpenes

---

## Table: Color-Related Assignments

<table>
<thead>
<tr>
<th>Color</th>
<th>Assignment</th>
<th>Isotope</th>
<th>Value</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 / yellow</td>
<td>( \text{C}_3\text{H}_6\text{O}_2\text{H}^- ) / acetic acid (61.03)</td>
<td>51B</td>
<td>51.05</td>
<td>Water cluster formation</td>
</tr>
<tr>
<td>4 / olive green</td>
<td>( \text{C}_3\text{H}_6\text{O}_2\text{H}^- ) / acetic acid (61.03)</td>
<td>50</td>
<td>50.01</td>
<td>Relation unknown</td>
</tr>
<tr>
<td>5 / green</td>
<td>( \text{C}_3\text{H}_6\text{O}_2\text{H}^- ) / acetic acid (61.03)</td>
<td>57A</td>
<td>56.94</td>
<td>Relation unknown</td>
</tr>
<tr>
<td>6 / aquamarine</td>
<td>( ^{17}\text{O}_2 ) / oxygen isotope (34.00)</td>
<td>34A</td>
<td>34.00</td>
<td>Relation unknown</td>
</tr>
<tr>
<td>7 / cyan</td>
<td>Not assigned</td>
<td>81B</td>
<td>81.05</td>
<td>Relation unknown</td>
</tr>
<tr>
<td>151</td>
<td>Not assigned (pot. ( \text{C}_3\text{H}_7\text{N}_2\text{H}^- ) / Pyrazine (81.05), ( \text{C}_3\text{H}_6\text{H}^- ) / hexenal fragment / monoterpene fragment (81.07))</td>
<td>151</td>
<td>151.00</td>
<td>Relation unknown</td>
</tr>
<tr>
<td>Mass</td>
<td>Assignment</td>
<td>Formula</td>
<td>Comment</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>------------</td>
<td>---------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>81B</td>
<td>Not assigned</td>
<td>C4H4N2.H+</td>
<td>Pyrazine (81.05), C6H8.H+ / hexenal fragment (81.07)</td>
<td></td>
</tr>
<tr>
<td>81.04</td>
<td>Not assigned</td>
<td>C4H4N2.H+</td>
<td>Pyrazine (81.05), C6H8.H+ / hexenal fragment (81.07)</td>
<td></td>
</tr>
<tr>
<td>137B</td>
<td>Not assigned</td>
<td>C10H16.H+</td>
<td>Monoterpenes (137.14)</td>
<td></td>
</tr>
<tr>
<td>137.00</td>
<td>Not assigned</td>
<td>C10H16.H+</td>
<td>Monoterpenes (137.14)</td>
<td></td>
</tr>
<tr>
<td>95.97</td>
<td>Not assigned</td>
<td>C10H16O</td>
<td>(153.12)</td>
<td></td>
</tr>
<tr>
<td>153.00</td>
<td>Not assigned</td>
<td>C10H16O</td>
<td>(153.12)</td>
<td></td>
</tr>
<tr>
<td>151.00</td>
<td>Not assigned</td>
<td>C10H14O</td>
<td>(151.10)</td>
<td></td>
</tr>
<tr>
<td>152.99</td>
<td>Not assigned</td>
<td>C10H16.H+</td>
<td>Monoterpenes (137.14)</td>
<td></td>
</tr>
<tr>
<td>137.00</td>
<td>Not assigned</td>
<td>C10H16.H+</td>
<td>Monoterpenes (137.14)</td>
<td></td>
</tr>
<tr>
<td>134.98</td>
<td>Not assigned</td>
<td>C10H16.H+</td>
<td>Monoterpenes (137.14)</td>
<td></td>
</tr>
<tr>
<td>137.00</td>
<td>Not assigned</td>
<td>C10H16.H+</td>
<td>Monoterpenes (137.14)</td>
<td></td>
</tr>
<tr>
<td>121.98</td>
<td>Not assigned</td>
<td>C10H16.H+</td>
<td>Monoterpenes (137.14)</td>
<td></td>
</tr>
<tr>
<td>123.00</td>
<td>Not assigned</td>
<td>C10H16.O</td>
<td>(123.04)</td>
<td></td>
</tr>
<tr>
<td>128.01</td>
<td>Not assigned</td>
<td>C10H16.O</td>
<td>(123.04)</td>
<td></td>
</tr>
<tr>
<td>128.01</td>
<td>Not assigned</td>
<td>C10H16.O</td>
<td>(123.04)</td>
<td></td>
</tr>
<tr>
<td>134.98</td>
<td>Not assigned</td>
<td>C10H16.O</td>
<td>(123.04)</td>
<td></td>
</tr>
<tr>
<td>134.98</td>
<td>Not assigned</td>
<td>C10H16.O</td>
<td>(123.04)</td>
<td></td>
</tr>
<tr>
<td>134.98</td>
<td>Not assigned</td>
<td>C10H16.O</td>
<td>(123.04)</td>
<td></td>
</tr>
</tbody>
</table>

**No significant positive connections to other VOCs:**

- CH3OH.H+ (35.04)
- C3H3O (44.03) / HCHO.H+ (44.03)
- H2O.H+ / Water cluster (55.04)
- C3H6.H+ / butadiene as fragment of aldehydes (55.05)
- acrolein (2-propenal, C3H5O.H) (57.03)
- Unassigned (pot. C3H5S.H+ / DMS sulfur isotope (65.03) C3SO.H+ oxygen isotope (65.01))
The success of personalized medicine approaches for metabolic diseases depends on inexpensive and minimally invasive but also sensitive and specific diagnostic tools. The analysis of volatile organic compounds in human breath has the potential to provide such an “easy-access” view on a broad range of metabolic pathways. However, the origin and the link to physiological functions of many volatiles are still unknown thus hindering the implementation of a breath gas screen in clinical settings. Here in this study, we screened breath of obese mice for disease related alterations in VOC patterns under controlled and standardized circumstances. We found that a variety of volatiles were affected in obese mice or varied depending on the feeding status of the animals. For both visualizing and identifying altered patterns of volatile organic compounds, we estimated a gaussian graphical model as a data-driven approach novel in PTR-MS based breath gas analysis. Known from metabolomics studies (Krumsiek et al., 2012, 2011), graphical modeling here helped to identify contributions to single peaks; for example, isotopes of MTMT (pk62) and CO$_2$.H$_2$O'/DMS (pk63) to peak 64B. In addition, visualizing fragmentation patterns like propanol fragments at peaks 41 and 43B is possible (Schwarz et al., 2009). In addition to chemical properties, biochemical pathway information could be included, as e.g. the conversion of acetone to 2-propanol by ADH1 (Lewis et al., 1984). The combination of this information can shed light on previously unknown peaks and the corresponding volatiles. In addition, the graphical model resembles the hierarchical clustering shown in the two heatmaps in several edges but emphasizes numerous further connections as it can be built on both complete data sets with correction for intra-experimental effects. In the following, the tentatively identified volatiles altered in obese mice are discussed individually with a special emphasis on the available obesity and metabolic disease relevant literature.

4.1. A VOC signature altered in the volatilome both obesity models

In the present study we found that in both diet-induced and genetically induced obesity source strengths of several emitted VOCs were altered. Interestingly, a set of eight peaks was changed in both obesity models, of which four were tentatively identified and four remain unknown.

4.1.1. Acetic acid (pk61, pk43A)

A volatile identified in both groups by the feature selection algorithm as potentially relevant was acetic acid (pk61 and fragment pk43A in HFD mice). In this untargeted screening approach mixed effects modeling did not find significantly changed source strengths in acetate but in the fragment
pk43A in HFD mice (pk61 \( p_{\text{MC4R}} = 0.27, p_{\text{HFD}} = 0.12; \) pk43A \( p_{\text{HFD}} = 0.015 \)). Interestingly, despite the non-significant increase in source strength the increased variance and outlying data in both obesity models might be used for stratification of obesity related pathologies like disturbances in glucose homeostasis. The enzymes acetyl-CoA synthetase and acetyl-CoA hydrolase regulate free acetic acid levels, in addition exogenous sources like gut fiber fermentation contribute to serum levels predominantly after food intake (Wolever et al., 1997). Serum acetic acid levels were reported to be inversely correlated to insulin levels in mice and humans (Layden et al., 2012; Sakakibara et al., 2009). Acetate reduces glucose-induced insulin secretion via pancreatic FFAR2 and FFAR3. This is likely mediated by pancreas secreted acetate produced from glucose as a negative feedback as well as from overall systemic acetic acid levels (Tang et al., 2015). It is therefore coherent that acetic acid in breath could be used to model glucose levels during an OGTT and to detect individuals with gestational diabetes (Halbritter et al., 2012). In addition, short chain fatty acids like acetate have been recognized to induce a PPARγ-dependent switch from lipid synthesis to lipid utilization in white adipose tissue and liver (Besten et al., 2015). Thus, if the variance in acetic acid could be attributed to associated (patho)physiological states in obesity, it might be a relevant non-invasive marker.

4.1.2. Methanol (pk33B)

Methanol was selected in both models (peaks 33B and MC4R-ki mice also 18O-isotope 35B). A massive increase of methanol was found in the MC4R experiment during the fed state. These mice were fed a so-called chow diet comparably high in pectin/ fiber content thus affecting VOC signatures mediated by altered microbial digestion as previously described (Kistler et al., 2014). We could not detect a general effect of obesity status on methanol. Thus, the effect of diet on methanol breath levels seems to exceed the endogenous variation. This is in accordance to other studies, identifying methanol to originate mainly from microbial digestion of consumed pectins with a smaller fraction from other dietary and endogenous sources as aspartame or S-Adenosylmethionine (Axelrod & Daly, 1965; Siragusa et al., 1988; Lindinger et al., 1997; Dorokhov et al., 2012). Notably, methanol was used together with a set of VOCs to model blood glucose in type 1 diabetics (Minh et al., 2011) and was found to be reduced in HFD-diet fed rats (Aprea et al., 2012), inversely correlated to BMI in humans (Turner et al., 2006; Halbritter et al., 2012) but increased in liver cirrhosis (Morisco et al., 2013). Those findings may be related to differences in life style or eating habits (e.g. reduced fruit (pectin) consumption) affecting gut microbiota in obese patients. In addition, a reduction of methanol detoxification capacity could be present in those states. However, to verify this in a mouse model, it has to considered that detoxification of methanol in humans is primarily adh1 driven, while
in rodents, peroxidative activity of catalase is relevant for degradation (Dorokhov et al., 2015; Karinje & Ogata, 1990).

4.1.3. Carbon dioxide \( ^4\text{H}_2\text{O} / \text{Dimethyl sulfide} \) (pk63, pk64B, pk65)

Another peak elevated in both obese models is peaks 63 as well as isotopes at 64B (in HFD mice) and probably 65, which are probably a mixed signal from carbon dioxide - water cluster and dimethyl sulfide. Carbon dioxide, as a terminal mitochondrial oxidation product of most energy-containing molecules, is directly related to the amount of energy used in the organism. Indeed the utilized obese models do have an increased overall amount of metabolic active tissue (Figure 1). This increase consists not only of fat mass, which is considered to have a lower but not negligible metabolic activity per gram (Kaiyala et al., 2010), but also of highly active lean mass which is elevated. Hence a higher emission of carbon dioxide in heavier mice is not surprising (Butler & Kozak, 2010; Tschöp et al., 2012).

DMS, the second candidate, was found to be increased in obese rats with steatohepatitis, obese children, liver cirrhotic patients and is a known constituent of the fetor hepaticus (Aprea et al., 2012; Alkhouri et al., 2015; Morisco et al., 2013; Van den Velde et al., 2008). DMS can be found in breath after methionine ingestion and is altered in hepatitis and cirrhosis patients showing an increased half-life (Kaji et al., 1979). Furthermore, rat skeletal muscle cells were observed to be releasing DMS, possibly produced by the transamination pathway out of methionine and cysteine (Mochalski et al., 2014). Thus, after further insight in its metabolism, DMS could be used as a non-invasive biomarker of altered systemic or hepatic metabolism of sulfur containing amino acids.

4.1.4. \((\text{Methylthio})\text{methanethiol} \) (MTMT, pk62)

The source strength of a fragment of \((\text{methylthio})\text{methanethiol} \) (MTMT, pk62 and part of pk64B signal in HFD fed mice) is ~100 times higher in male mice compared to females (as shown for MC4R-ki animals in Appendix 1). This long-distance pheromone was initially described to be detected in the main olfactory bulb, being involved in the attractiveness of male urine to female mice and can be reduced in urine by castrating male mice (Da Yu Lin et al., 2005). Unexpectedly, we found the source strength of MTMT significantly increased in ad libitum fed diet-induced obese mice but reduced in MC4R-ki mice. The tissue and mechanism of endogenous MTMT synthesis are currently unknown. However, a link between the melanocortin 4 receptor and sexual reproduction has been shown (Van
der Ploeg et al., 2002) and reduced mating success is observed in models with reduction in melanocortin production (Faulkner et al., 2015). In contrast to the MC4R-ki mice, upon HFD feeding an elevation in MTMT in ad libitum state but not in the fasted state is observed. Upon HFD feeding, an acute compensatory activation of MC4R signaling is known (Butler et al., 2001) and could also modulate MTMT levels via a MC4R-dependent mechanism.

4.2. Significant HFD specific peaks

4.2.1. Ammonia (pk18)

Ammonia is elevated in liver pathologies (Adeva et al., 2012) and was found increased in obese children (Alkhouri et al., 2015). In addition, in a rat study featuring diet-induced obesity measured with similar instrumentation, breath ammonium was increased in purified HFD versus low fat standard diet fed rats (Aprea et al., 2012). Unexpectedly, we found reduced breath ammonia source strength of in HFD-fed mice in comparison to control mice. Generally, values from HFD-fed mice seem to be lower than both groups in the MC4R experiment as well. In the field of breath research, the reproducibility of breath ammonia measurements is in discussion (Blanco Vela & Bosques Padilla, 2011), as aside from changed blood ammonia concentrations, breath ammonia altered by physical activity level (Solga et al., 2014), mode of breathing, airway or mouth pH (Solga et al., 2013) and mouth bacteria expressing urease (Chen et al., 2014). As mice show a strong preference for nasal breathing, some of the above mentioned effects should not be present here. A higher dietary protein load in this particular HFD (24.1 % in HFD versus 20.8 % in LFD) can contribute to a metabolic acidosis. This in combination with the ketoacidosis in ad libitum HFD and fasting could lead to an increased urinary ammonia excretion to compensate acidosis and therefore reduced breath ammonia.

4.2.2. Acrolein (pk57B)

Interestingly, another VOC increased in HFD fed mice both in ad libitum as well as in fasted state but not selected as relevant in MC4R-ki mice is likely acrolein. In humans, acrolein exposure from exogenous sources as diet as well as inhalation of polluted air and smoking are known to be relevant. In addition, endogenous production from lipid peroxidation in oxidative stress, degradation of methionine/threonine and spermine/spermidine can contribute to the observed concentrations (Stevens & Maier, 2008). Notably, in HFD mice a slight reduction in fasted state is observed, probably
indicating that both directly diet-derived and endogenous produced acrolein contribute to the elevation compared to LFD fed littermates. Acrolein is contributing to metabolic pathologies via a wide range of mechanisms and target tissues, including protein adduction, induction of oxidative stress, mitochondrial dysfunction, inflammation and immune alterations, ER stress, structural and membrane effects and deregulated signal transduction as reviewed by Moghe et al. (Moghe et al., 2015). Hence it can be an interesting breath resource for monitoring carbonyl stress and redox state.

4.2.3. Methyl acetate (pk75)

An increase in pk75 in HFD mice was observed, which is possibly predominantly methyl acetate as suggested by the gaussian graphical model. In addition to the individually discussed literature on acetate and methanol, methyl acetate in breath is only described to be increased acutely after exercise (King et al., 2010). In obese patients with non-alcoholic fatty liver disease, an increase in various fecal volatile esters including methyl acetate could be observed and associated to a gut microbial shift (Raman et al., 2013). Emitted methyl acetate therefore indicates such a shift, or alternatively can be created from acetate and methanol within the mouse metabolism.

4.2.4. $^{18}$O$^{16}$O oxygen (pk34A)

An increased consumption of oxygen isotope $^{18}$O-$^{16}$O is observed in HFD fed mice. This can likely be explained by a higher amount of metabolic active tissue and therefore higher absolute oxygen demand in the heavier HFD mice (Tschöp et al., 2012).

4.2.5. H$_3$O$^+$.H$_2$O$_2$ water cluster and fragments of aldehydes (pk55B)

Pk55B is increased in HFD mice and is likely to consist of both H$_3$O$^+$.H$_2$O$_2$ water cluster and fragments of aldehydes e.g. butanal, hexanal, octanal or nonanal (Buhr et al., 2002). Although it is unclear why humidity and water clustering should be increased in obese mice, aldehydes in breath (and breath condensate) can be increased in oxidative stress pathologies with associated lipid peroxidation (Amann et al., 2014), which can be elevated in obese state.

4.2.6. Acetone and propanol (pk59, pk43B)
In HFD induced obesity, energy demands are to an extended portion satisfied by lipid oxidation and hepatic ketogenesis. One of the ketone bodies is acetone, which is thought to be produced by spontaneous decarboxylation of acetoacetate. Therefore, in both states of increased fatty acid oxidation, namely HFD and food restriction, acetone source strengths are elevated. In humans, fasting breath acetone levels were shown to be highly correlated to β-hydroxybutyrate and acetoacetate blood concentrations (Qiao et al., 2014; Musa-Veloso et al., 2006). Both were associated to increased fasting and 2h plasma glucose levels and acetoacetate could be used to predict both an increased GTT AUC and 5-year diabetes incidence (Mahendran et al., 2013). Notably, in a subnetwork of fasting responsive volatiles (Figure 6), acetone showed a significant partial correlation to both propanol fragments at nominal masses 41 and 43B (which showed a significant fasting but no diet effect (p=0.118)). The conversion from acetone to iso-propanol is known and can be enhanced in a ketogenic setting (Lewis et al., 1984; Petersen et al., 2012). As conversion from propanol to acetone is also possible and breath propanol is highly correlated to environmental concentrations in a clinical setting (Ghimenti et al., 2013), this can be one reason why in human breath analysis high variance in acetone levels is observed.

4.3. MC4R-ki specific peaks

4.3.1. Cluster of unknowns 151 (F), 153 (G), 137B (N) and 81B (C)

In the gaussian graphical model, a subnetwork affected in MC4R-ki mice was observed, featuring peaks 151 (F), 153 (G), 137B (N) and 81B (C). A literature search on similar PTR-MS fragmentation patterns revealed that monoterpenes like α- and β-pinene, 3-carene, limonene and camphor produced fragment ions of masses 67, 81 and 95 as well as a protonated molecular ion of mass 137 or 153 (Tani et al., 2003). Notably, in a human lipid infusion study to predict plasma TG and FFA levels from breath volatiles, β-Limonene and β-pinene were relevant for the models (Minh et al., 2012). Also monoterpenes (137.137) and terpene-related peak (135.119) have been found as breath markers for liver cirrhosis in a human study (Morisco et al., 2013). An altered diet composition as one explanation the authors named can be excluded here. So either increased food consumption in MC4R-ki mice or the suggested alteration hepatic terpene metabolism in this study can explain the elevated levels in obese mice. However, it has to be noted that the observed peaks do not match exactly the theoretical masses. Possibly, the fact that those peaks are far from the internal calibration masses typically applied in PTR-MS measurements using H$_3$O$^+$ (21.02), NO$^+$ (30.00) and protonated acetone (59.05) should be causing this mass shift. Especially in the used PTR-TOF-2000
instrument with a resolution of $\leq 2000 \text{ m/}\Delta \text{m}$, an added high molecular internal calibration gas can be useful to be included in future studies.

### 4.4. Unassigned volatiles

In addition, three further volatiles - namely peaks 50, 81B and 117B - were identified to be altered in both obese mouse models. Unidentified peaks were also found in HFD fed mice, including peaks 249B, 253A, 57A, 36B and 74B. Even more volatiles are considered unknown in MC4R-ki mice, including peaks 44A, 123B, 135B, 36A, 122B, 137B, 113, 94B, 80B, 87B, 211 and 95B. Those candidates are worth further exploration using complementary methods like classical pre-concentration combined with GC-MS-MS or a novel combination of a fast-GC device to the PTR-MS for additional chemical information (Romano et al., 2014). However, altered experimental settings leading to increased VOC concentration can be necessary due to the lower sensitivity of those methods. Alternatively, a nose mask sampling could be applied despite its obvious need for extensive acclimation of rodents to avoid stress induced effects on the measured volatiles (Aprea et al., 2012). Interestingly, data-driven models like the applied gaussian graphical model can at least in part contribute to the identification of VOCs. In addition to the volatiles showing effects in mixed effects models, peaks 126A and 48B did not show effects and are considered false positives in the selection process.

### 5. Conclusion

In this study we characterized alterations in exhaled volatile organic compounds in both diet-induced and mono-genetic obese mouse models and to evaluate whether a common pattern of VOCs altered in obesity can be determined. Alterations in the volatilome could be detected with a common obesity VOC signature. Notably, different adiposity models do create distinct shifts in the volatilome as well, thus showing the potential of VOC analysis to monitor and distinguish different obesogenic mechanisms. Identified VOCs originate from various metabolic pathways and biological processes including ketone body metabolism, lipid peroxidation and pheromones allowing a broad overview over metabolic state in a fast a non-invasive way. In addition, we suggest gaussian graphical models as a helpful tool in understanding and characterizing the volatilome. Thus, the
analysis of the volatile metabolome has the potential to contribute to a personalized medicine by aiding in the stratification of patients with heterogeneous metabolic phenotypes and risk profiles.

6. Conflicts of Interest

The authors declare that they have no conflict of interest.

7. Author contributions

M. Ki. conceived and designed the experiments, researched data, reviewed and analyzed the data and wrote the manuscript. N.R. and A.M. took care of animal management, researched data (A.M.), reviewed and edited the manuscript. W.S. and J.R. conceived and designed the experiments, reviewed the data, wrote (JR), reviewed and edited the manuscript. W.W. contributed to mouse line generation. C.H., M. Kl., H.F., V.GD., W.W. and M.H.A. contributed to discussion, reviewed and edited the manuscript.

8. Acknowledgments

We thank Brigitte Herrmann, Ann Elisabeth Schwarz and all animal caretakers in the GMC for their technical contribution to mouse care and phenotyping. Furthermore, our thanks go to Florian Bolze and Ralf Kühn, who contributed to generation and initial characterization of the MC4R-ki mouse line. This work was partly funded by the FP7 Marie Curie Initial Training Network PIMMS, Grant Agreement No. 287382, by German Federal Ministry of Education and Research (Infrafrontier grant 01KX1012) and by the German Center for Diabetes Research (DZD).
9. References


Ploner A (2014) Heatplus: Heatmaps with row and/or column covariates and colored clusters


Supplementary figure 1: Source strengths of pk62 in males and females. Source strengths of pk62 (MTMT) are shown as boxplots. Box fill corresponds to genotype and sex (blue: male melanocortin-4-receptor wild type; dark blue: male melanocortin-4-receptor W16X knock-in, yellow: female melanocortin-4-receptor wild type; dark yellow: female melanocortin-4-receptor W16X knock-in). Box border corresponds to fasting state (black: ad libitum fed; grey: fasted).

Supplementary table 1: Linear mixed effects model testing showing genotype, diet and fasting effects on VOC source strengths.
<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>sd</th>
<th>mean</th>
<th>sd</th>
<th>mean</th>
<th>sd</th>
<th>mean</th>
<th>sd</th>
<th>mean</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>fasted</td>
<td></td>
<td>ad lib</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>haversian</td>
<td>1246.68</td>
<td>267.05</td>
<td>2166.76</td>
<td>4020.37</td>
<td>639.59</td>
<td>346.57</td>
<td>119.32</td>
<td>576.09</td>
<td>488.37</td>
<td>40.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Osteoblast</td>
<td>1246.68</td>
<td>267.05</td>
<td>2166.76</td>
<td>4020.37</td>
<td>639.59</td>
<td>346.57</td>
<td>119.32</td>
<td>576.09</td>
<td>488.37</td>
<td>40.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Osteoclast</td>
<td>1246.68</td>
<td>267.05</td>
<td>2166.76</td>
<td>4020.37</td>
<td>639.59</td>
<td>346.57</td>
<td>119.32</td>
<td>576.09</td>
<td>488.37</td>
<td>40.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Osteocyte</td>
<td>1246.68</td>
<td>267.05</td>
<td>2166.76</td>
<td>4020.37</td>
<td>639.59</td>
<td>346.57</td>
<td>119.32</td>
<td>576.09</td>
<td>488.37</td>
<td>40.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Osteoclast</td>
<td>1246.68</td>
<td>267.05</td>
<td>2166.76</td>
<td>4020.37</td>
<td>639.59</td>
<td>346.57</td>
<td>119.32</td>
<td>576.09</td>
<td>488.37</td>
<td>40.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Osteocyte</td>
<td>1246.68</td>
<td>267.05</td>
<td>2166.76</td>
<td>4020.37</td>
<td>639.59</td>
<td>346.57</td>
<td>119.32</td>
<td>576.09</td>
<td>488.37</td>
<td>40.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Osteoclast</td>
<td>1246.68</td>
<td>267.05</td>
<td>2166.76</td>
<td>4020.37</td>
<td>639.59</td>
<td>346.57</td>
<td>119.32</td>
<td>576.09</td>
<td>488.37</td>
<td>40.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Osteocyte</td>
<td>1246.68</td>
<td>267.05</td>
<td>2166.76</td>
<td>4020.37</td>
<td>639.59</td>
<td>346.57</td>
<td>119.32</td>
<td>576.09</td>
<td>488.37</td>
<td>40.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Osteoclast</td>
<td>1246.68</td>
<td>267.05</td>
<td>2166.76</td>
<td>4020.37</td>
<td>639.59</td>
<td>346.57</td>
<td>119.32</td>
<td>576.09</td>
<td>488.37</td>
<td>40.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>--------</td>
<td>-------</td>
<td>---</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>3.11</td>
<td>5.05</td>
<td>1.21</td>
<td>0.42</td>
<td>0.45</td>
<td>0.33</td>
<td>1.09</td>
<td>B</td>
<td>4.05</td>
<td>1.83</td>
<td>0.16</td>
</tr>
<tr>
<td>7.12</td>
<td>4.17</td>
<td>0.73</td>
<td>0.23</td>
<td>0.01</td>
<td>0.09</td>
<td>0.17</td>
<td>1.0</td>
<td>0.13</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>1.23</td>
<td>4.99</td>
<td>0.91</td>
<td>0.08</td>
<td>0.08</td>
<td>0.03</td>
<td>0.07</td>
<td>0.0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>0.97</td>
<td>2.87</td>
<td>0.45</td>
<td>0.21</td>
<td>0.71</td>
<td>0.03</td>
<td>0.03</td>
<td>0.0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>0.56</td>
<td>3.96</td>
<td>0.91</td>
<td>0.08</td>
<td>0.08</td>
<td>0.03</td>
<td>0.07</td>
<td>0.0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>0.76</td>
<td>3.96</td>
<td>0.91</td>
<td>0.08</td>
<td>0.08</td>
<td>0.03</td>
<td>0.07</td>
<td>0.0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>0.22</td>
<td>0.85</td>
<td>0.33</td>
<td>0.21</td>
<td>0.71</td>
<td>0.03</td>
<td>0.03</td>
<td>0.0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

**Note:** The table contains numerical data, likely related to scientific or technical measurements.