Effects of CPAP therapy withdrawal on exhaled breath pattern in obstructive sleep apnoea

Esther I Schwarz*, Pablo Martinez-Lozano Sinues*, Lukas Bregy, Thomas Gaisl, Diego Garcia Gomez, Martin T Gaugg, Yannick Suter, Nina Stebler, Yvonne Nussbaumer-Ochsner, Konrad E Bloch, John R Stradling, Renato Zenobi, and Malcolm Kohler

*These authors contributed equally to this work

Supplementary File

Content

E-METHODS

Eligibility criteria

Patients aged between 20 and 75 years with known moderate to severe obstructive sleep apnoea (OSA), having an oxygen desaturation index (ODI) of >20/h at the time of diagnosis, registered in the database of the Sleep Disorders Center of the University Hospital Zurich were eligible if they currently had an ODI >20/h (≥4%-dips) during an ambulatory nocturnal pulse oximetry on the forth night off CPAP, were treated with CPAP for more than 12 months, and had a minimum CPAP-compliance of ≥4h/night as well as an apnoea-hypopnoea index (AHI) <10/h with treatment (according to CPAP machine download). Patients with previous ventilatory failure (awake SpO2 <93% and PaCO2>6kPa), unstable coronary or peripheral artery disease, severe arterial hypertension or hypotension (>180/110 or <90/60mmHg), Cheyne-Stokes breathing, acute inflammatory disease, any previous sleep-related accident or use of inhaled drugs, as well as professional drivers, were excluded.

Sample size

The sample size was estimated on the assumption that a clinically relevant difference in the secondary outcome ODI – reflecting the disease of interest – between therapeutic and subtherapeutic CPAP is 15/h (SD 12).[1-3] Based on this assumption, power calculation indicated that 28 patients are required in total to not miss a clinically relevant difference in the secondary outcome with a power of 90%, and a two-tailed significance level (alpha = 0.05). Taking into account a possible drop-out rate of approximately 6%, the total number of patients who had to be included was adjusted to 30 (15 patients in each arm).

Sleep studies and CPAP devices

Home sleep studies using the ApneaLink™ Plus device (ResMed Corp, San Diego, USA) were performed the night before the baseline and the follow-up breath-analysis. OSA severity was quantified by ODI (≥4%-dips). Data on treatment adherence were downloaded from the internal memory of the CPAP device. Recurrence of OSA was defined as an ODI>15/h in the follow-up sleep study.

After randomisation, patients received a ResMed S8 AutoSet Spirit II CPAP device (ResMed Corp, San Diego, CA, USA). Subtherapeutic pressure in the sham-device was achieved by setting the CPAP machine to the lowest pressure, insertion of a flow-restrictor at the machine outlet, and of extra holes in the collar of the tube at the end of the mask as previously described and validated.[1]

Experimental set-up of mass spectrometric breath analysis

Real-time breath analysis was performed with secondary electrospray ionisation-mass spectrometry (SESI-MS).[4-12] **Figure E1** shows a schematic of the SESI set-up used in this study. SESI-MS allows for real-time breath-printing by detection of both volatile and non-volatile trace components in breath without any sample pre-treatment. Only the last few seconds (typically around 6 sec.) of each exhalation were considered for analysis, thus excluding the initial part of the exhalation, which reflects mostly the dead space in the upper respiratory tract. Participants were examined in the fasting state and were asked to abstain from smoking, chewing gum, alcohol, tobacco, or caffeine on the day of the measurements. Room temperature and lighting were set at the same level for all measurements. Breathprints were collected in real-time in triplicate or quadruple. Participants exhaled three- or four-times through a disposable mouthpiece into a heated Teflon tube (50 cm long, 3 mm inner diameter) connected to the curtain gas port of a time-of-flight mass spectrometer (TripleTOF 5600, AB Sciex, Concord, ON, Canada). The sampling tube was surrounded with a heating tape at 90°C to prevent water condensation and to minimize losses of exhaled compounds onto the walls. While performing full exhalations, the subjects kept the pressure through the sampling line at 10 mbar (i.e. ~ 1 L/min)), as monitored by a digital manometer. A lab-built nano electrospray plume was used, where compounds of exhaled breath are ionised and subsequently mass analysed. Water (0.1% formic acid) was electrosprayed at a flow rate of 100 nl/min. The mass spectrometric breath signatures were subsequently analysed.

Mass spectra preprocessing

Figure E2 shows an overview of the procedure followed to deliver a final working matrix containing intensity values for mass spectrometric features for each of the 28 subjects that completed the trial analysed twice (i.e. baseline and follow-up). The 56 raw *.wiff mass spectra

were transformed to *.mzXML format using the online tool msConvert.[13] The *.mzXML files were imported into Matlab (R2014b). Each mass spectrum was interpolated (shape-preserving piecewise cubic interpolation method) in the range m/z 41-449 using a linearly spaced vector of $7x10^5$ points. The interpolated mass spectra for the 56 files were concatenated resulting in a matrix of 700,000 x 10,465 (m/z x scans). Each of the 10,465 spectra was mass calibrated using an iterative grid search algorithm. The algorithm aligns the spectrum to a set of reference peaks by scaling and shifting the domain such that the cross-correlation between the input signal and a synthetic target signal is maximal. The synthetic target signal was built with Gaussian pulses centered at the locations specified by the vector of reference peaks. After obtaining the new m/z scale, the algorithm computes the calibrated mass spectrum by shape-preserving piecewise cubic interpolation of the shifted input signal to the original m/z vector. As reference peaks we chose acetone (m/z 59.04914) and commonly present contaminants pthalic anhydride (m/z 149.02332), dibutylphthalate (m/z 279.15909) and polysiloxane (m/z 445.1201).[14] The calibrated mass spectra were then centroided (threshold=200 counts), resulting in 4,068 m/z values. To increase the signal/noise ratio, the peak intensities around +/- 0.002 m/z were summed. The centroiding procedure reduced the matrix to 4,068 x 10,465 (m/z x scans). Each of the 4,068 time traces were smoothed using a moving average filter (span = 5). In order to identify the signals rising during the exhalation maneuvers in each individual, we subjected the time traces for each of the 56 files to hierarchical cluster analysis. A total of 3,922 different signals were found in the breath of the entire study population. Further consideration was only given to the signals present in at least 11 subjects, reducing the data matrix to 2,504 x 10,465 (m/z x scans). To account for instrumental drift (e.g. detector sensitivity) across the eight months of measurements, the signals were normalized. Thus, the total ion current (TIC) for each of the 56 samples (typically 2,504 x ~200, m/z x scans) was computed. The median of the TIC signal during the three replicate exhalations was computed and the 2,504 x ~200 (m/z x scans) matrix was divided by this value. This process was done for the 56 sample matrices. Finally, the adjusted 2,504 x 10,465 (m/z x scans) matrix was multiplied by the median of the 56 TIC normalizing values to recover intensity values of the same order as the pre-normalized matrix. **Figure E2** shows the non-normalized TIC (blue trace) for 9 subjects (separated by dashed lines) and the normalized TIC (red trace). Finally, the replicate signal intensities at the plateau of the exhalation phase for each subject was averaged resulting in a matrix of 2,504 x 56 (m/z x subjects). The number of features was further reduced to 911 by excluding some redundant signals such as isotopic peaks. This 911 x 56 matrix was

finally autoscaled (i.e. z-score). The 56 columns corresponded to 14 patients on therapeutic CPAP measured twice (baseline and follow-up) and 14 patients on subtherapeutic CPAP measured twice (baseline and follow-up).

Statistical methods

Changes in exhaled breath pattern

The primary outcome, the mass spectrometric profile of exhaled breath pattern, was analysed by standard univariate and multivariate statistical methods. Since this is a pathophysiologic metabolomic study, only those patients that completed the trial per protocol (26/30) were used for between group comparisons (primary outcome: change in exhaled breath pattern). Between groups comparisons were performed using a two-sample t-test (two-sided; 5% significance level) using 5,000 bootstrap samples. Bootstrapping (i.e. sampling with replacement) is a non-parametric technique employed to account for distortions caused by the specific sample that may not be fully representative of the population, which can be the case for limited sample size [15]. Subsequent estimate of the false discovery rate for multiple hypothesis testing was performed by computing q-values for each p-value as instructed by Storey.[16] A cutoff value of $p < 0.05$ and $q < 0.2$ resulted in 108 significant features. The q-values were in the range of 0.15 for these 108 features, suggesting strong evidence that around 80 signals were indeed significantly altered after CPAP withdrawal. A closer inspection of the 108 significant features revealed that some of these peaks appertained to the same molecule (e.g. in-source fragments). Redundant features were removed and the 62 remaining ones are listed in **Table E1**. **Figure E3** shows the data for six of these molecules. Similarly to the between-group comparison, for the within-groups comparisons (baseline vs. follow-up), each of the 911 features was subjected to a t-test with 5,000 bootstrap samples. Subsequently, q-values were computed for both groups.[16] **Table E1** lists the within-groups p- and q-values, as well as 95% confidence intervals for the 62 significant features selected in the between-group comparisons.

Correlation between ODI and breath signals

The 28 patients completing the trial and thus providing complete data from baseline to follow-up were used for correlation analysis between changes in signal intensity of exhaled breath compounds and disease severity. Pearson's linear correlation coefficient between changes in

breath signals and changes in ODI for the 28 subjects was computed for the 911 features. Thus, the baseline breath signals were subtracted to the follow-up ones for both groups; p-values for testing the hypothesis of no correlation against the alternative that there is a nonzero correlation were computed simultaneously. 131 features were found to correlate significantly ($p < 0.05$) with ODI. In addition, 95% bootstrap confidence intervals were computed using 10,000 bootstrap samples. Only the features not containing 0 in the 95% confidence interval were further considered, reducing the number of features to 96. To further retain features with high confidence level, the leverage for the dataset was computed to exclude artificially high correlations due to outliers. Thus, if any point in the data set had a leverage value greater than 0.5, the correlation was not further considered. This final step reduced the number of features to 79. Similarly to the groups comparison described above, some of these features were found to be redundant (i.e. more than one feature per molecule). As a result, 54 features were left. **Table E2** lists the molecules suggesting a correlation between change in ODI and change in signal intensity. **Figure E4** displays eight examples. For easier visualization, the 79 features with the strongest correlation were subjected to principal component analysis (PCA). **Figure E5** shows the significant ($r = 0.59$; $p < 0.001$) correlation between the first principal component score and change in ODI.

OSA prediction

OSA prediction was accomplished in a leave-one-out-cross-validation (LOOCV) by feature selection and subsequent classification using a random forest classification algorithm.[17] It is important to note that to prevent any bias, the predicted sample was out of the training set (n= 26) at all time during the feature selection process, meaning that feature selection was also subjected to cross-validation.[18] In order to avoid unbalanced groups and thus making it cost-sensitive and unstable due to the limited sample size, two balanced groups were created by selecting the most extreme cases in terms of ODI.[19] Thus samples with ODI>15/h were labeled as “OSA” (n=14) and ODI<3 ‘no OSA’ (n=13). Prediction was computed using an ensemble of 200 decision trees using the top 15 most informative features. First an ensemble with 200 bagged decision trees was created for the training set. Predictor importance was estimated by ranking the variables according to their Gini score. The top 100 selected features were further used in a subsequent ensemble. A further feature selection over the 100 pre-selected variables was done by measuring the increase in prediction error if the values of that variable were permuted across the out-of-bag observations. Predictions for the test sample (n=1) were finally computed using the top 15 most

informative features. This process was repeated for the 27 samples. A ROC curve was computed with the classification scores. We then computed its area under the curve and it 95% confidence interval using 10,000 bootstrap samples. **Table E3** lists the features selected during the LOOCV. In an attempt to visualize and cluster the samples using the most significant features identified during the classification, a hierarchical cluster analysis (average linkage and correlation distance) using the 19 top variables (i.e. selected > 5 times) was computed. The same 19 top variables were used to perform multidimensional scaling of the proximity matrix. A scatter plot of the first and second scaled coordinates (i.e. corresponding to the two largest eigenvalues) is shown in **figure 3b**.

Molecule identification by mass spectrometry

In order to carry out compound identification, the most significant m/z features found in the statistical analysis phase, were chemically analysed *a posteriori*. Fragmentation (i.e. MS/MS) experiments were performed with the same instrument (i.e. AB Sciex Tripletof5600+). Subjects were asked to breath into the instrument in the same way as during the full MS mode analysis to perform MS/MS analysis in real-time. Multiple reaction monitoring-like experiments were conducted with arbitrary collision energy of 30V +/- 15 V. These spectra were compared with those obtained from standards and from the MassBank database (massbank.jp) and with *in silico* MS/MS spectra obtained with MetFrag,[20] an open-source combinatorial fragmenter for identifying product ions from small organic compounds that heuristically analyses every possible fragment from chemical structure databases such as KEGG or PubChem. To gain confidence in the assignments, we conducted further real-time measurements in full MS mode using an LTQ Orbitrap mass spectrometer, which usually results in a mass accuracy below 2 ppm. This high mass accuracy, in addition to the analysis of the isotopic distribution, allowed for the assignment of a unique molecular formula for each peak with very high confidence. Finally, for further identification, exhaled breath condensate samples were collected using a customized device following the recommendations suggested by the ATS/ERS task force.[21] These samples were analysed by UHPLC-HRMS, a technique that allows identification of small molecules.[22] When available, retention times were compared with those obtained from standards. Combining all these measurements, a molecular identity was proposed for 22 molecules (**Table E4**). Thus, the identification of these structures was based on different parameters (numbered from 1 to 6) that

imply different levels or confidence. For example, the comparison with a standard MS/MS spectrum or whether the compound has been previously found in exhaled breath.

E-TABLES

Table E1

		Between-groups (n = 26)					Within-groups									
							CPAP-withdrawal (n=13)					Therapeutic CPAP (n=13)				
m/z	**Metabolite**	**p**	**q**	**mean difference (z-score)**	**95% CI**		**p**	**q**	**mean difference (z-score)**	**95% CI**		**p**	**q**	**mean difference (z-score)**	**95% CI**	
69.0681	Isoprene	0.020	0.16	1.19	0.15	2.23	0.00031	0.01	1.21	0.49	1.93	0.934	0.50	0.03	-0.81	0.86
71.0592		0.010	0.16	1.34	0.28	2.40	0.00041	0.01	1.24	0.51	1.98	0.804	0.50	-0.10	-0.93	0.74
80.0148		0.016	0.16	1.25	0.20	2.30	0.01156	0.05	1.10	0.13	2.08	0.623	0.50	-0.15	-0.67	0.38
80.0148		0.016	0.16	1.25	0.20	2.30	0.01156	0.05	1.10	0.13	2.08	0.623	0.50	-0.15	-0.67	0.38
81.0505		0.010	0.15	1.19	0.26	2.12	0.03413	0.10	0.53	0.03	1.03	0.166	0.48	-0.66	-1.50	0.18
91.0384		0.017	0.16	1.26	0.19	2.34	0.00105	0.02	1.18	0.29	2.06	0.817	0.50	-0.09	-0.80	0.62
92.04728		0.020	0.16	1.11	0.14	2.07	0.00025	0.01	1.22	0.56	1.89	0.773	0.50	0.12	-0.65	0.89
97.0261		0.033	0.18	1.01	0.04	1.98	0.00077	0.02	1.19	0.43	1.96	0.656	0.50	0.18	-0.49	0.86
97.0622	2-ethylfuran	0.023	0.17	1.13	0.12	2.14	0.00038	0.01	1.38	0.51	2.25	0.444	0.48	0.25	-0.36	0.87
99.0894		0.010	0.16	1.26	0.26	2.25	0.01044	0.05	1.18	0.22	2.14	0.729	0.50	-0.08	-0.51	0.35
102.0888	2-pentenal	0.002	0.15	1.55	0.55	2.55	0.00008	0.01	1.57	0.75	2.39	0.942	0.50	0.02	-0.65	0.68
104.0477		0.023	0.17	0.82	0.09	1.56	0.00293	0.02	1.08	0.46	1.70	0.524	0.48	0.26	-0.21	0.73
105.0514		0.031	0.18	1.06	0.06	2.07	0.00526	0.03	0.98	0.30	1.66	0.833	0.50	-0.08	-0.90	0.73
108.0094		0.009	0.15	1.26	0.28	2.23	0.00520	0.03	1.25	0.26	2.23	0.953	0.50	-0.01	-0.32	0.30
109.0079		0.006	0.15	1.23	0.32	2.13	0.00147	0.02	1.32	0.46	2.18	0.743	0.50	0.09	-0.32	0.51
109.023		0.014	0.16	1.13	0.20	2.06	0.00124	0.02	1.15	0.45	1.86	0.944	0.50	0.02	-0.67	0.71
109.0621		0.028	0.17	1.07	0.08	2.06	0.00612	0.04	1.08	0.15	2.01	0.954	0.50	0.01	-0.46	0.49
109.0743	Methylphenol	0.002	0.15	1.57	0.58	2.57	0.00001	0.00	1.45	0.75	2.16	0.760	0.50	-0.12	-0.89	0.65
110.0185		0.039	0.20	0.90	0.01	1.78	0.01276	0.05	0.97	0.32	1.61	0.851	0.50	0.07	-0.61	0.75
111.0782	2-propylfuran	0.003	0.15	1.43	0.47	2.40	0.00005	0.01	1.49	0.71	2.28	0.850	0.50	0.06	-0.59	0.71
111.0893		0.007	0.15	1.22	0.31	2.13	0.01368	0.05	1.09	0.28	1.90	0.644	0.50	-0.13	-0.65	0.39
117.0991		0.008	0.15	1.27	0.30	2.24	0.00545	0.03	1.12	0.39	1.86	0.681	0.50	-0.14	-0.86	0.57
118.0288	Homocysteine thiolactone	0.020	0.16	1.11	0.14	2.08	0.00104	0.02	1.03	0.35	1.71	0.852	0.50	-0.08	-0.85	0.69
120.0781		0.006	0.15	1.03	0.28	1.78	0.00020	0.01	1.31	0.77	1.84	0.493	0.48	0.28	-0.30	0.86

122.082		0.029	0.17	0.81	0.05	1.57	0.01279	0.05	0.94	0.30	1.59	0.735	0.50	0.13	-0.34	0.61
122.1018		0.014	0.16	1.20	0.21	2.19	0.00271	0.02	1.26	0.32	2.20	0.844	0.50	0.06	-0.40	0.51
123.0891	Ethylphenol	0.002	0.15	1.52	0.52	2.53	0.00183	0.02	0.97	0.29	1.65	0.211	0.48	-0.55	-1.36	0.26
124.0817		0.037	0.19	0.94	0.02	1.86	0.00200	0.02	1.00	0.38	1.63	0.879	0.50	0.06	-0.69	0.81
125.093	2-Butylfuran	0.009	0.15	1.30	0.29	2.31	0.00036	0.01	1.30	0.59	2.01	0.992	0.51	0.00	-0.79	0.79
125.1291		0.018	0.16	1.23	0.18	2.27	0.00196	0.02	1.20	0.39	2.01	0.929	0.50	-0.02	-0.78	0.73
127.0188		0.017	0.16	0.98	0.14	1.82	0.00238	0.02	1.21	0.51	1.90	0.511	0.48	0.23	-0.32	0.78
129.0885	4-hydroxy-2-heptenal	0.016	0.16	1.27	0.20	2.35	0.00332	0.02	1.22	0.29	2.14	0.847	0.50	-0.06	-0.72	0.60
131.0568		0.007	0.15	1.23	0.31	2.16	0.01144	0.05	1.08	0.19	1.97	0.619	0.50	-0.16	-0.56	0.25
134.0649		0.016	0.16	1.06	0.17	1.96	0.00927	0.04	1.02	0.49	1.55	0.893	0.50	-0.04	-0.83	0.74
135.0406		0.012	0.16	1.20	0.23	2.16	0.00171	0.02	1.23	0.43	2.04	0.906	0.50	0.04	-0.59	0.66
136.0193	Benzothiazole	0.027	0.17	1.05	0.08	2.01	0.00147	0.02	1.30	0.44	2.15	0.433	0.48	0.25	-0.30	0.80
136.0449		0.013	0.16	1.20	0.22	2.18	0.00187	0.02	1.22	0.39	2.04	0.950	0.50	0.01	-0.60	0.63
137.0544		0.008	0.15	1.15	0.27	2.03	0.00345	0.02	1.06	0.36	1.76	0.817	0.50	-0.09	-0.70	0.52
138.0144		0.005	0.15	1.26	0.35	2.17	0.00139	0.02	1.31	0.47	2.15	0.879	0.50	0.04	-0.42	0.51
138.0563		0.012	0.16	1.28	0.24	2.32	0.01539	0.05	0.92	0.08	1.76	0.358	0.48	-0.36	-1.07	0.35
139.109	2-pentylfuran	0.004	0.15	1.48	0.44	2.51	0.00007	0.01	1.45	0.66	2.25	0.939	0.50	-0.02	-0.77	0.73
143.1161		0.017	0.16	1.29	0.19	2.39	0.02435	0.08	0.89	0.07	1.71	0.272	0.48	-0.40	-1.22	0.42
147.0504		0.013	0.16	1.26	0.23	2.29	0.00326	0.02	1.15	0.24	2.05	0.743	0.50	-0.12	-0.72	0.49
149.0957		0.008	0.15	1.50	0.36	2.65	0.00543	0.03	1.12	0.23	2.01	0.250	0.48	-0.38	-1.19	0.44
150.0609		0.015	0.16	1.13	0.19	2.07	0.00367	0.03	1.08	0.40	1.76	0.880	0.50	-0.05	-0.78	0.67
150.1005		0.015	0.16	1.21	0.20	2.22	0.00931	0.04	1.03	0.15	1.92	0.624	0.50	-0.18	-0.77	0.41
152.0665		0.016	0.16	1.28	0.20	2.35	0.00975	0.04	0.68	0.07	1.28	0.213	0.48	-0.60	-1.56	0.36
154.0634		0.019	0.16	1.09	0.14	2.03	0.00298	0.02	1.11	0.34	1.89	0.931	0.50	0.03	-0.60	0.66
155.1522	2-decenal	0.017	0.16	0.97	0.15	1.79	0.00970	0.04	1.11	0.52	1.70	0.650	0.50	0.14	-0.50	0.78
158.123		0.036	0.19	1.24	0.04	2.45	0.00851	0.04	0.96	0.12	1.80	0.497	0.48	-0.28	-1.24	0.68
160.1298	4-hydroxy-2-octenal	0.019	0.16	1.27	0.17	2.38	0.01286	0.05	0.96	0.12	1.81	0.421	0.48	-0.31	-1.12	0.49
166.0965	Mevalonic acid	0.033	0.18	0.87	0.04	1.70	0.00465	0.03	0.94	0.26	1.62	0.874	0.50	0.06	-0.49	0.62
167.1031		0.037	0.19	1.25	0.03	2.47	0.00338	0.02	0.84	0.23	1.45	0.392	0.48	-0.41	-1.55	0.72
169.0865		0.027	0.17	1.21	0.10	2.32	0.00868	0.04	0.90	0.17	1.64	0.483	0.48	-0.31	-1.22	0.61

170.1106		0.003	0.15	1.54	0.51	2.57	0.00058	0.01	1.07	0.39	1.75	0.290	0.48	-0.47	-1.31	0.38
175.1086		0.033	0.18	1.11	0.05	2.16	0.00774	0.04	0.95	0.17	1.72	0.704	0.50	-0.16	-0.96	0.64
182.0068	2-(methylthio) benzothiazole	0.021	0.16	1.06	0.13	1.99	0.00565	0.03	1.15	0.46	1.83	0.782	0.50	0.09	-0.61	0.79
196.1102	Digitalose	0.008	0.15	1.33	0.32	2.34	0.00728	0.04	1.22	0.26	2.18	0.638	0.50	-0.11	-0.58	0.35
205.0704		0.013	0.16	-1.19	-2.15	-0.22	0.04613	0.11	-0.52	-1.00	-0.04	0.168	0.48	0.67	-0.23	1.57
209.1522		0.027	0.17	1.15	0.09	2.21	0.00391	0.03	0.95	0.29	1.61	0.648	0.50	-0.21	-1.11	0.70
221.1882		0.030	0.17	1.22	0.07	2.36	0.03879	0.10	0.79	-0.07	1.64	0.274	0.48	-0.43	-1.28	0.42
297.0996		0.029	0.17	1.12	0.08	2.17	0.03560	0.10	0.90	0.07	1.73	0.495	0.48	-0.22	-0.96	0.51

Table E1. Features significantly altered in response to CPAP-withdrawal.

Table E2

Correlation between changes of breath signals and changes of ODI (n = 28)					
m/z	ID	Pearson's correlation coefficient	95% CI		p
85.07397	2-pentenal	0.54	0.20	0.78	0.003
109.0743	Methylphenol	0.45	0.19	0.63	0.016
111.0782	propylfuran	0.40	0.08	0.65	0.033
118.0288	Homocysteine thiolactone	0.37	0.03	0.61	0.050
123.0891	Ethylphenol	0.49	0.20	0.68	0.008
123.1136		0.40	0.01	0.76	0.033
125.093	2-butylfuran	0.44	0.11	0.67	0.020
129.0885	4-hydroxy-2-heptenal	0.48	0.14	0.75	0.009
137.0544		0.44	0.18	0.65	0.019
139.109	2-pentylfuran	0.38	0.04	0.62	0.047
143.1045	4-hydroxy-2-octenal	0.42	0.05	0.70	0.025
152.0665		0.42	0.15	0.61	0.027
155.1522	2-decenal	0.41	0.07	0.66	0.029
160.1298	4-hydroxy-2-octenal	0.42	0.11	0.67	0.025
164.0716		0.41	0.06	0.78	0.030
166.0965	Mevalonic acid	0.40	0.10	0.62	0.034
169.1559	2-undecenal	0.38	0.04	0.70	0.046
170.1106		0.38	0.08	0.62	0.043
182.0837		0.38	0.11	0.60	0.045
195.1333	Hexyloxy-phenol	0.38	0.04	0.65	0.049
196.1102	Digitalose	0.59	0.18	0.84	0.001
199.0722		0.53	0.26	0.83	0.003
207.1338		0.38	0.12	0.57	0.049
209.1131		0.40	0.13	0.59	0.035
210.15		0.39	0.04	0.69	0.038
211.1263		0.38	0.06	0.61	0.044
213.0183		0.51	0.13	0.83	0.005
221.1492		0.38	0.13	0.57	0.044
223.128		0.44	0.13	0.72	0.018
227.124		0.43	0.14	0.65	0.024
233.149		0.41	0.12	0.60	0.030
236.0837		-0.39	-0.59	-0.07	0.038
236.9662		0.41	0.07	0.75	0.028
237.1066		0.40	0.18	0.57	0.037
243.1177		0.38	0.13	0.59	0.046
247.1633		0.46	0.14	0.68	0.014
249.0022		0.42	0.08	0.76	0.025
249.1438		0.38	0.18	0.57	0.046
249.1788		0.45	0.18	0.62	0.017
251.1605	2-ethylhexyl-4-hydroxybenzoate	0.43	0.09	0.68	0.023

256.9903	0.42	0.06	0.75	0.028
257.0439	0.45	0.05	0.79	0.017
262.982	0.39	0.01	0.76	0.040
263.1219	0.43	0.18	0.65	0.023
263.1604	0.38	0.08	0.60	0.048
272.0379	0.38	0.12	0.64	0.046
278.2139	0.43	0.13	0.64	0.024
294.0233	0.40	0.03	0.72	0.035
303.0174	0.37	0.04	0.60	0.050
307.0065	0.42	0.11	0.73	0.025
309.0074	0.39	0.10	0.70	0.041
313.1119	0.40	0.08	0.76	0.034
315.054	0.38	0.05	0.65	0.046
369.0855	0.46	0.16	0.71	0.014

Table E2. Features correlating with disease severity (oxygen desaturation index).

Table E3

Features selected during leave-one-out-cross-validation (n = 27)											
m/z	Selection frequency	m/z	Selection frequency	m/z	Selection frequency	m/z	Selection frequency	m/z	Selection frequency	m/z	Selection frequency
183.0104	18	93.05505	4	127.0223	2	105.01	1	177.0047	1	258.9935	1
213.9602	18	139.109	4	129.0121	2	108.0094	1	182.9446	1	274.1205	1
262.982	16	168.9793	4	147.0504	2	109.0079	1	184.0036	1	276.998	1
213.9689	15	182.9463	4	151.9598	2	110.0185	1	188.1075	1	284.0285	1
53.03835	14	216.8669	4	164.9442	2	110.0652	1	195.9709	1	294.8603	1
319.9873	13	321.9836	4	165.1243	2	111.0053	1	196.1667	1	304.0584	1
258.9889	12	43.05457	3	167.9902	2	125.0137	1	196.9349	1	327.0772	1
112.0084	10	44.02562	3	175.1086	2	128.0673	1	198.9341	1	329.9979	1
221.9646	10	79.03848	3	194.9363	2	131.0568	1	205.9698	1	330.9917	1
258.9906	9	87.04228	3	210.1541	2	132.057	1	208.1747	1	331.0056	1
346.0183	9	90.07256	3	237.9716	2	134.1267	1	211.9936	1	358.1237	1
125.1291	8	92.04728	3	258.9423	2	136.0449	1	213.9806	1	393.2893	1
111.0782	7	97.06219	3	298.0392	2	138.0144	1	213.9847	1		
139.0122	7	102.0888	3	371.319	2	138.9994	1	213.9957	1		
182.0068	7	112.0183	3	57.03326	1	140.1418	1	216.0643	1		
213.9649	7	123.1148	3	58.06375	1	141.0103	1	221.9809	1		
319.9856	7	151.1094	3	62.9889	1	146.0893	1	228.0654	1		
123.1136	6	198.9318	3	69.06811	1	157.0097	1	228.0671	1		
333.0025	6	213.9887	3	71.04749	1	157.0237	1	231.1714	1		
81.02835	5	344.975	3	83.08351	1	160.0592	1	232.9579	1		
166.983	5	89.06946	2	95.04725	1	164.0267	1	240.9931	1		
222.9508	5	90.0714	2	97.05462	1	166.9987	1	248.9649	1		
321.9865	5	103.0913	2	98.06646	1	168.9863	1	250.0222	1		
337.012	5	109.023	2	101.0571	1	170.9867	1	250.9664	1		
88.04596	4	122.1018	2	103.0924	1	172.1104	1	250.9822	1		

Table E3. LOOCV selected features. The top 19 (i.e. selected at least 6 times) were used to generate figures 3b and 3c.

Table E4

TripleTOF m/z	Orbitrap m/z	Formula	Acc. / ppm	MS/MS-Fragments	ID	ID based on
69.0681	69.0698	C_5H_9	1.1	C_3H_5	Isoprene	4, 6
85.0740 102.0888	85.0648 102.0913	C_5H_9O C_5H_9O +NH_3	0.1	C_3H_7O, C_2H_3O	2-pentenal	2, 5
87.0423	87.0440	$C_4H_7O_2$	0.6	C_4H_5O, C_3H_7O, C_2H_3O	4-hydroxy-2-butenal	2, 5
89.0695	89.0595	$C_4H_9O_2$	2.3	C_4H_7O, C_2H_3O	Acetoin	4, 6
97.0622	97.0647	C_6H_9O	0.9	C_4H_5O	2-ethylfuran	1, 5, 6
109.0743	109.0649	C_7H_9O	1	C_7H_7, C_6H_7O, C_6H_5	Methylphenol	1, 5, 6
111.0782	111.0804	$C_7H_{11}O$	0.4	C_5H_7O, C_4H_5O	2-propylfuran	2, 5, 6
118.0288	118.0320	C_4H_8NOS	0.9	C_3H_8NS, C_3H_6NO, C_3H_6N	Homocysteine thiolactone	1, 5, 6
123.0891	123.0803	$C_8H_{11}O$	1.1	C_8H_9, C_6H_7O, C_6H_5	Ethylphenol	1, 4
125.0930	125.0961	$C_8H_{13}O$	0.1	C_6H_9O, C_4H_5O	2-Butylfuran	2, 5, 6
129.0885	129.0908	$C_7H_{13}O_2$	1.6	$C_7H_{11}O$, C_5H_9O, C_4H_7O, C_2H_3O	4-hydroxy-2-heptenal	2, 5, 6
136.0193	136.0215	C_7H_6NS	0.3	C_6H_5S, C_6H_5, C_3HS, C_5H_5	Benzothiazole	1, 5, 6
139.1090	139.1117	$C_9H_{15}O$	0.3	$C_7H_{11}O$, C_6H_9O, C_4H_5O	2-pentylfuran	1, 3, 5, 6
143.1045 160.1298	143.1066 160.1332	$C_8H_{15}O_2$ $C_8H_{15}O_2$ +NH_3	0.4 0.1	$C_8H_{13}O$, C_3H_5O, C_2H_3O	4-hydroxy-2-octenal	2, 5, 6
151.1094	151.1117	$C_{10}H_{15}O$	0.3	$C_{10}H_{13}$, $C_8H_{11}O$, C_7H_9O, C_6H_7O, C_6H_5	4-butylphenol	1, 4
155.1522	155.1430	$C_{10}H_{19}O$	0.3	$C_8H_{15}O$, C_2H_3O	2-decenal	2, 5, 6
166.0965	166.1075	$C_6H_{16}NO_4$	0.7	$C_6H_{13}O_4$, $C_4H_7O_2$, C_4H_9O	Mevalonic acid	4
169.1559	169.1588	$C_{11}H_{21}O$	0.6	C_2H_3O	2-undecenal	2, 5, 6
182.0068	182.0095	$C_8H_8NS_2$	1.3	$C_7H_5NS_2$, $C_6H_4S_2$, C_7H_5NS, C_6H_5NS, C_6H_5S, C_6H_4S, C_5H_4S, C_6H_5N, C_6H_5, C_5H_5	2-(methylthio)benzothiazole	1, 3, 4
195.1333	195.1381	$C_{12}H_{19}O_2$	0.7	$C_7H_5O_2$, C_6H_7O, C_6H_5	4-(hexyloxy)phenol	1, 3 4
196.1102	196.1177	$C_7H_{18}NO_5$	1.3	$C_7H_{15}O_5$, $C_7H_{13}O_4$, $C_3H_5O_2$, C_3H_7O	Digitalose	4
251.1605	251.1639	$C_{15}H_{23}O_3$	1.1	$C_{13}H_{19}O_3$, $C_{10}H_9O_3$, $C_8H_5O_3$, $C_7H_5O_2$, C_6H_5O	2-ethylhexyl-4-hydroxybenzoate	1, 3, 4

Table E4. Combination of different approaches (1-6) allowed chemical identification of 22 molecules in exhaled breath of OSA patients. 16 of the 22 two identified molecules were shown to significantly change in response to CPAP withdrawal. 1=Comparison with standard MS/MS

spectra. 2=Comparison with similar standard MS/MS spectra. 3=Retention time in UHPLC (standard vs. EBC). 4=Comparison with in silico MS/MS spectra (MetFrag – KEGG, score: 1.0). 5=Comparison with in silico MS/MS spectra (MetFrag–PubChem, score: 1.0). 6=Previously found in the volatilome database.[23 24]

Table E5

		Between-groups (n = 26)					Within-groups									
							CPAP-withdrawal (n=13)					Therapeutic CPAP (n=13)				
m/z	Metabolite	p	q	mean difference (z-score)	95% CI		p	q	mean difference (z-score)	95% CI		p	q	mean difference (z-score)	95% CI	
69.0681	Isoprene	0.02	0.16	1.19	0.15	2.23	0.00031	0.01	1.21	0.49	1.93	0.934	0.5	0.03	-0.81	0.86
97.0622	2-ethylfuran	0.023	0.17	1.13	0.12	2.14	0.00038	0.01	1.38	0.51	2.25	0.444	0.48	0.25	-0.36	0.87
102.0888	2-pentenal	0.002	0.15	1.55	0.55	2.55	0.00008	0.01	1.57	0.75	2.39	0.942	0.5	0.02	-0.65	0.68
109.0743	Methylphenol	0.002	0.15	1.57	0.58	2.57	0.00001	0	1.45	0.75	2.16	0.76	0.5	-0.12	-0.89	0.65
111.0782	2-propylfuran	0.003	0.15	1.43	0.47	2.4	0.00005	0.01	1.49	0.71	2.28	0.85	0.5	0.06	-0.59	0.71
118.0288	Homocysteine thiolactone	0.02	0.16	1.11	0.14	2.08	0.00104	0.02	1.03	0.35	1.71	0.852	0.5	-0.08	-0.85	0.69
123.0891	Ethylphenol	0.002	0.15	1.52	0.52	2.53	0.00183	0.02	0.97	0.29	1.65	0.211	0.48	-0.55	-1.36	0.26
125.093	2-Butylfuran	0.009	0.15	1.3	0.29	2.31	0.00036	0.01	1.3	0.59	2.01	0.992	0.51	0	-0.79	0.79
129.0885	4-hydroxy-2-heptenal	0.016	0.16	1.27	0.2	2.35	0.00332	0.02	1.22	0.29	2.14	0.847	0.5	-0.06	-0.72	0.6
136.0193	Benzothiazole	0.027	0.17	1.05	0.08	2.01	0.00147	0.02	1.3	0.44	2.15	0.433	0.48	0.25	-0.3	0.8
139.109	2-pentylfuran	0.004	0.15	1.48	0.44	2.51	0.00007	0.01	1.45	0.66	2.25	0.939	0.5	-0.02	-0.77	0.73
155.1522	2-decenal	0.017	0.16	0.97	0.15	1.79	0.0097	0.04	1.11	0.52	1.7	0.65	0.5	0.14	-0.5	0.78
160.1298	4-hydroxy-2-octenal	0.019	0.16	1.27	0.17	2.38	0.01286	0.05	0.96	0.12	1.81	0.421	0.48	-0.31	-1.12	0.49
166.0965	Mevalonic acid	0.033	0.18	0.87	0.04	1.7	0.00465	0.03	0.94	0.26	1.62	0.874	0.5	0.06	-0.49	0.62
182.0068	2-(methylthio) benzothiazole	0.021	0.16	1.06	0.13	1.99	0.00565	0.03	1.15	0.46	1.83	0.782	0.5	0.09	-0.61	0.79
196.1102	Digitalose	0.008	0.15	1.33	0.32	2.34	0.00728	0.04	1.22	0.26	2.18	0.638	0.5	-0.11	-0.58	0.35

E-FIGURES

Figure E1

Top View

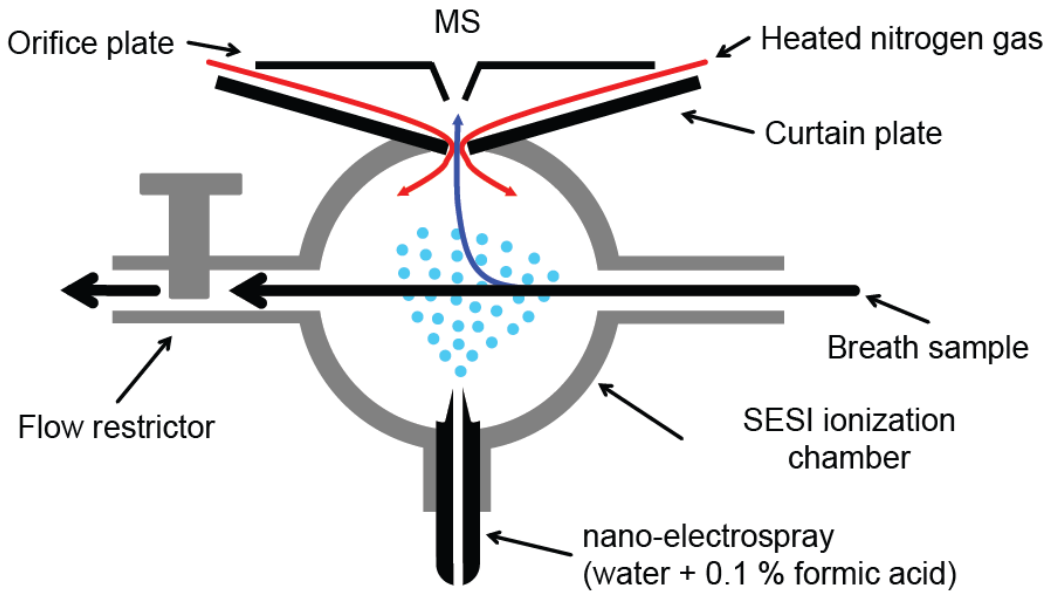

Figure E1. Scheme of the lab-built secondary electrospray ionization chamber coupled to a mass spectrometer to allow for the real-time breath analysis.

Figure E2

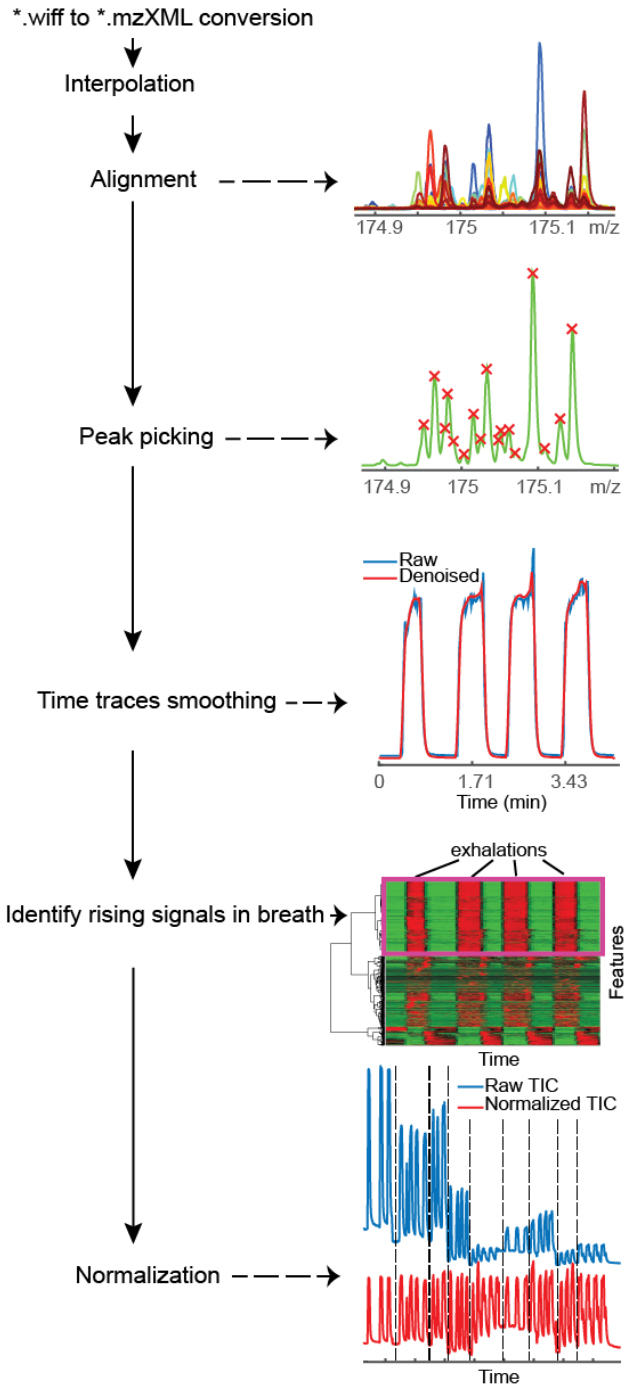

Figure E2. Workflow of mass spectra preprocessing.

Figure E3

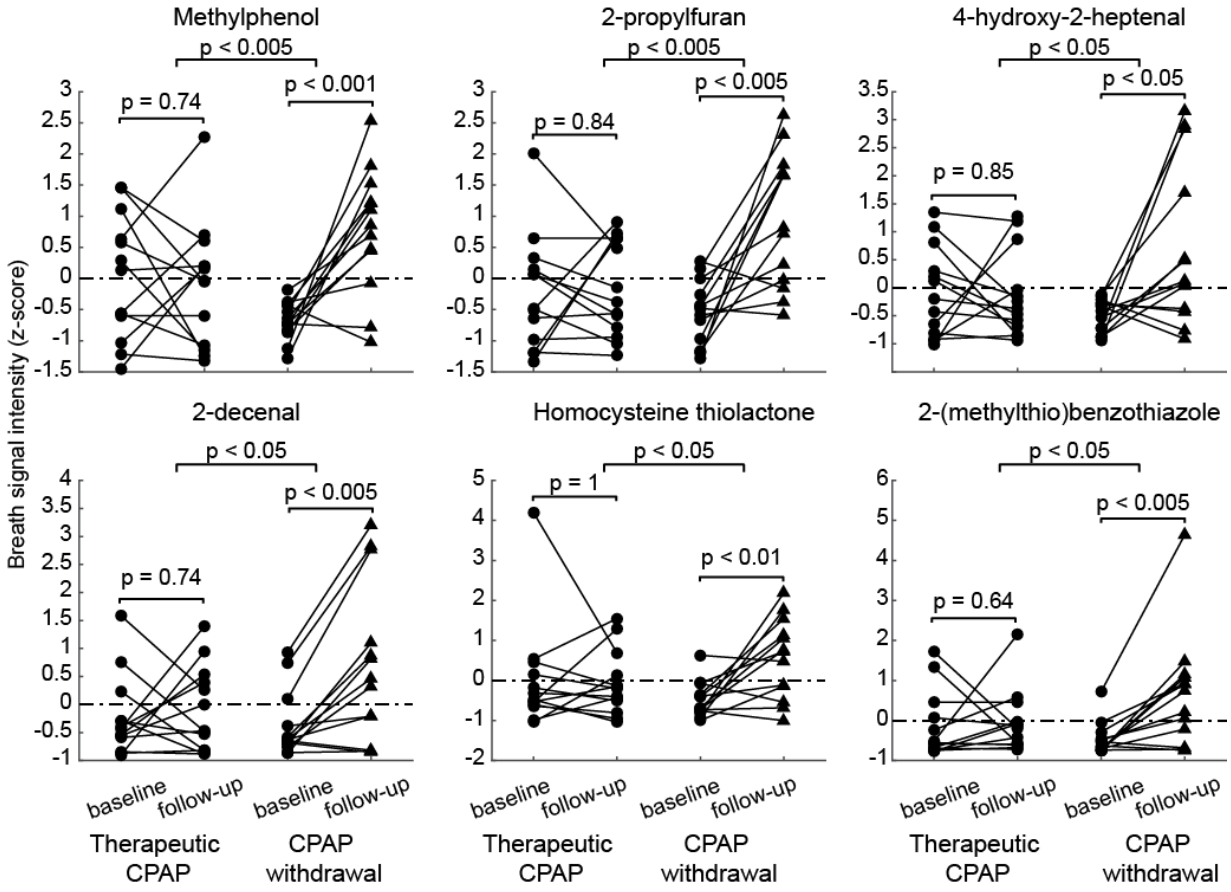

Figure E3. Changes of breath signal intensity of molecules significantly enhanced after CPAP withdrawal when compared to continuing CPAP (n=26); p-values for within and between-groups comparisons are quoted. See **Table E1** for details.

Figure E4

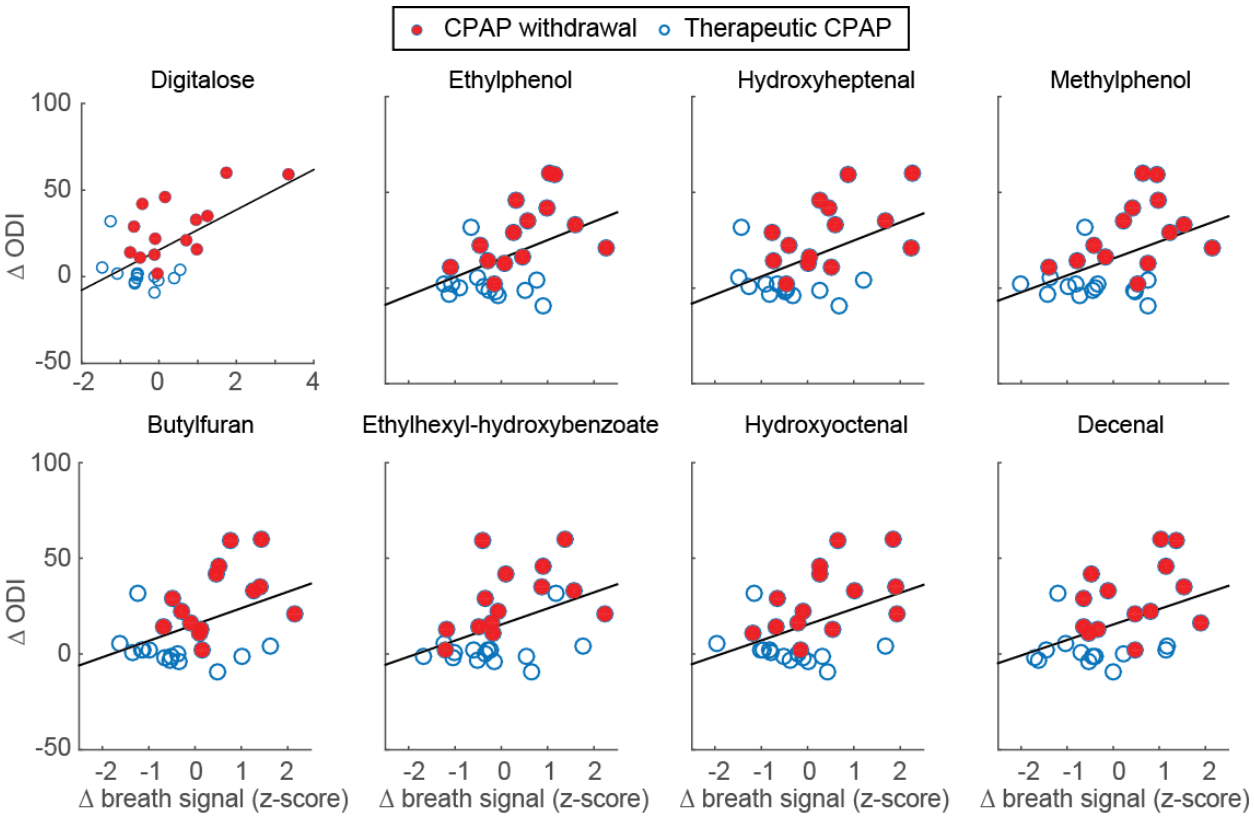

Figure E4. Selected molecules correlating with OSA severity (n=28). See **Table E2** for details.

Figure E5

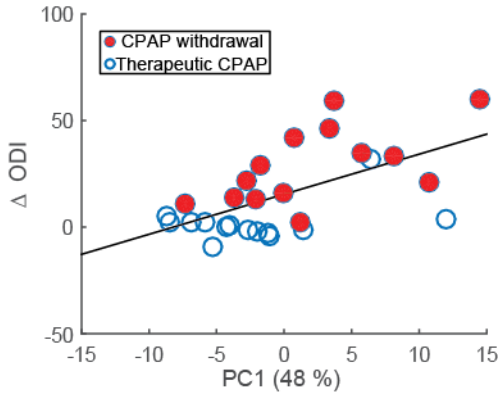

Figure E5. Change in (Δ) ODI vs. the first principal component score ($r = 0.59$; $p < 0.001$).

Figure E6

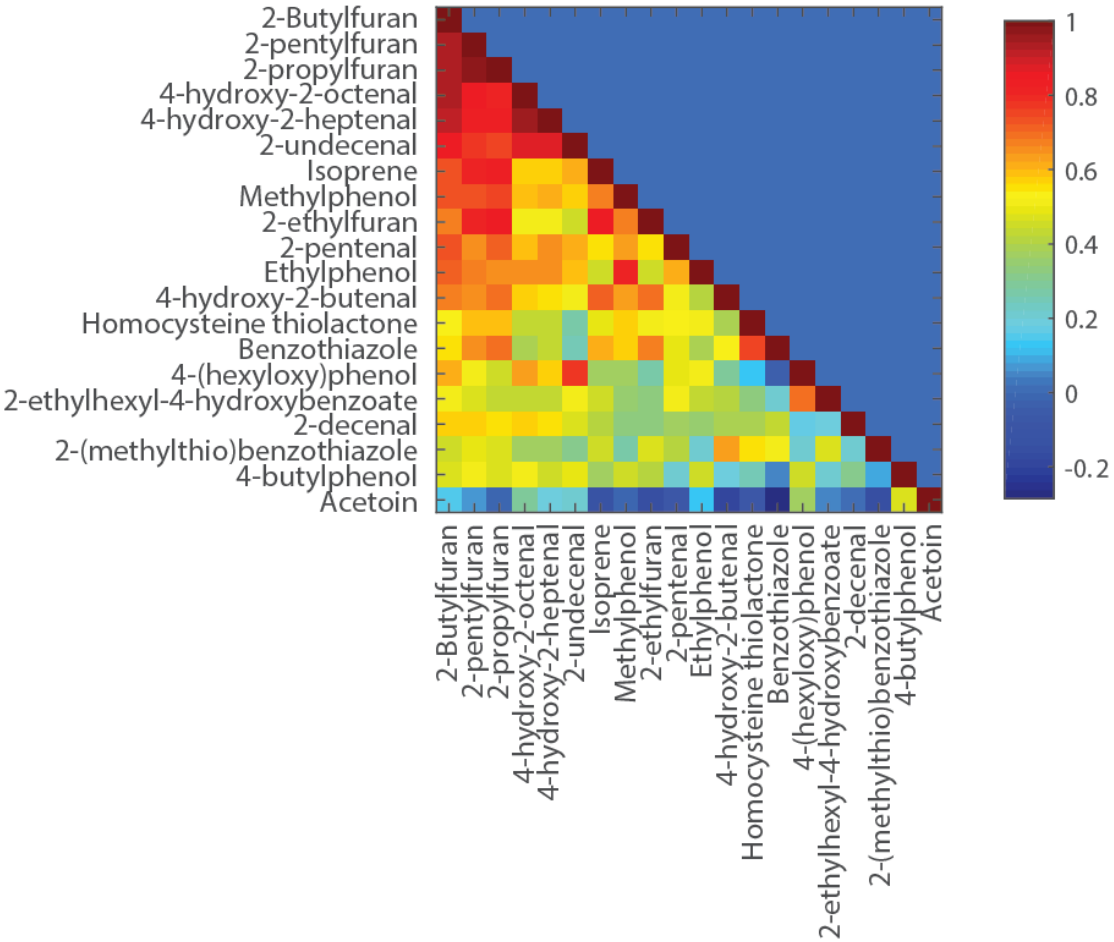

Figure E6. Heat-map of the correlation matrix for identified compounds in descending order of mean correlation from top to bottom. To reveal any latent association across the identified compounds, the correlation of each metabolite with all other metabolites was computed. A heat-map of the correlation matrix for all the identified compounds (the 20 with the highest degree of confidence) in descending order of mean correlation from top to bottom is shown. Note that the top ranked molecules are a series of furans and aldehydes. A further visualization of the connectivity network of the identified metabolites is shown in **Figure 3e**. Interconnected metabolites are based on their partial correlations (requiring $p < 0.01$).

E-REFERENCES

1. Kohler M, Stoewhas AC, Ayers L, et al. Effects of continuous positive airway pressure therapy withdrawal in patients with obstructive sleep apnea: a randomized controlled trial. American journal of respiratory and critical care medicine 2011;**184**(10):1192-9 doi: 10.1164/rccm.201106-0964OC[published Online First: Epub Date]|.
2. Rossi VA, Winter B, Rahman NM, et al. The effects of Provent on moderate to severe obstructive sleep apnoea during continuous positive airway pressure therapy withdrawal: A randomised controlled trial. Thorax 2013;**68**(9):854-59 doi: 10.1136/thoraxjnl-2013-203508[published Online First: Epub Date]|.
3. Schwarz EI, Schlatzer C, Stehli J, et al. The effects of short-term CPAP withdrawal on myocardial perfusion in OSA – A randomised placebo-controlled trial. Eur Respir J 2014 44:Suppl 58, 4664
4. Martinez-Lozano Sinues P, Meier L, Berchtold C, et al. Breath analysis in real time by mass spectrometry in chronic obstructive pulmonary disease. Respiration 2014;**87**(4):301-10 doi: 10.1159/000357785[published Online First: Epub Date]|.
5. Martinez-Lozano Sinues P, Tarokh L, Li X, et al. Circadian Variation of the Human Metabolome Captured by Real-Time Breath Analysis. PLoS ONE 2014;**9**(12):e114422 doi: 10.1371/journal.pone.0114422[published Online First: Epub Date]|.
6. Martinez-Lozano P, Fernandez de la Mora J. Direct analysis of fatty acid vapors in breath by electrospray ionization and atmospheric pressure ionization-mass spectrometry. Analytical chemistry 2008;**80**(21):8210-5 doi: 10.1021/ac801185e[published Online First: Epub Date]|.
7. Martínez-Lozano P, Fernández de la Mora J. Electrospray ionization of volatiles in breath. International Journal of Mass Spectrometry 2007;**265**(1):68-72
8. Martínez-Lozano P, Zingaro L, Finiguerra A, Cristoni S. Secondary electrospray ionization-mass spectrometry: breath study on a control group. J. Breath. Res. 2011;**5**(1):016002
9. Bean HD, Jiménez-Díaz J, Zhu J, Hill JE. Breathprints of model murine bacterial lung infections are linked with immune response. Eur. Respir. J. 2015;**45**(1):181-90 doi: 10.1183/09031936.00015814[published Online First: Epub Date]|.
10. Zhu J, Bean HD, Jiménez-Díaz J, Hill JE. Secondary electrospray ionization-mass spectrometry (SESI-MS) breathprinting of multiple bacterial lung pathogens, a mouse model study. Journal of applied physiology 2013 doi: 10.1152/japplphysiol.00099.2013[published Online First: Epub Date]|.
11. Wu C, Siems WF, Hill HH. Secondary Electrospray Ionization Ion Mobility Spectrometry/Mass Spectrometry of Illicit Drugs. Analytical chemistry 2000;**72**(2):396-403
12. Martinez-Lozano Sinues P, Zenobi R, Kohler M. Analysis of the exhalome: a diagnostic tool of the future. Chest 2013;**144**(3):746-9 doi: 10.1378/chest.13-1106[published Online First: Epub Date]|.
13. Chambers MC, Maclean B, Burke R, et al. A cross-platform toolkit for mass spectrometry and proteomics. Nat. Biotechnol. 2012;**30**(10):918-20 doi: http://www.nature.com/nbt/journal/v30/n10/abs/nbt.2377.html#supplementary-information[published Online First: Epub Date]|.
14. Keller BO, Sui J, Young AB, Whittal RM. Interferences and contaminants encountered in modern mass spectrometry. Anal. Chim. Acta 2008;**627**(1):71-81 doi: 10.1016/j.aca.2008.04.043[published Online First: Epub Date]|.

15. Riffenburgh RH. Chapter 15 - Managing Results of Analysis. In: Riffenburgh RH, ed. Statistics in Medicine (Third Edition). San Diego: Academic Press, 2012:325-43.
16. Storey JD. A direct approach to false discovery rates. J. Roy. Stat. Soc. Ser. B. (Stat. Method.) 2002;**64**(3):479-98 doi: 10.1111/1467-9868.00346[published Online First: Epub Date]|.
17. Breiman L. Random Forests. Mach. Learn. 2001;**45**(1):5-32 doi: 10.1023/A:1010933404324[published Online First: Epub Date]|.
18. Simon R, Radmacher MD, Dobbin K, McShane LM. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. Journal of the National Cancer Institute 2003;**95**(1):14-8
19. Domingos P. MetaCost: a general method for making classifiers cost-sensitive. Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining. San Diego, California, USA: ACM, 1999:155-64.
20. Wolf S, Schmidt S, Müller-Hannemann M, Neumann S. In silico fragmentation for computer assisted identification of metabolite mass spectra. BMC Bioinformatics 2010;**11** doi: 10.1186/1471-2105-11-148[published Online First: Epub Date]|.
21. Horvath I, Hunt J, Barnes PJ, et al. Exhaled breath condensate: methodological recommendations and unresolved questions. Eur. Respir. J. 2005;**26**(3):523-48 doi: 10.1183/09031936.05.00029705[published Online First: Epub Date]|.
22. García-Gómez D, Martínez-Lozano Sinues P, Barrios-Collado C, Vidal-De-Miguel G, Gaugg M, Zenobi R. Identification of 2-alkenals, 4-hydroxy-2-alkenals, and 4-hydroxy-2,6-alkadienals in exhaled breath condensate by UHPLC-HRMS and in breath by real-time HRMS. Anal. Chem. 2015;**87**(5):3087-93 doi: 10.1021/ac504796p[published Online First: Epub Date]|.
23. Costello BdL, 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
24. Amann A, Costello Bde L, Miekisch W, et al. The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva. J. Breath. Res. 2014;**8**(3):034001 doi: 10.1088/1752-7155/8/3/034001[published Online First: Epub Date]|.