Clinica Chimica Acta 557 (2024) 117895

Contents lists available at ScienceDirect

Clinica Chimica Acta

journal homepage: www.elsevier.com/locate/cca

Highly sensitive serum volatolomic biomarkers for pancreatic cancer diagnosis

María-Pilar Martínez-Moral [a,b,*,1], María Teresa Tena [b], Alfonso Martín-Carnicero [a,c], Alfredo Martínez [a]

[a] Oncology Area, Center for Biomedical Research of La Rioja (CIBIR), C/ Piqueras 98, 26006 Logroño, Spain
[b] Department of Chemistry, University of La Rioja, C/ Madre de Dios 51, 26006 Logroño, Spain
[c] Medical Oncology Department, Hospital San Pedro, C/Piqueras 98, 26006 Logroño, Spain

ARTICLE INFO

Keywords:
Volatile organic compounds
Volatolomics
SPME-GC/MS
Biomarker
Pancreas
Cancer

ABSTRACT

The discovery of new diagnostic tools for the early detection of diseases with poor prognosis such as pancreatic adenocarcinoma (PAC) is of high importance. The results from a control-case study (20 PAC patients, 19 healthy controls) for the search of new biomarkers of pancreatic cancer based in differences in the serum volatolome are presented in this work. Volatolomics were performed following a non-targeted HS-SPME-GC/MS approach, and a total of 433 volatile organic compounds (VOCs) was detected in the human serum samples. Of these, 125 VOC indexes showed a significant variation when controls and patients were compared (p-value < 0.05). Bonferroni corrected p-values < 0.05 were found for 40 features. PCA analysis showed the control-PAC discrimination capability of VOCs in serum, and PLS-DA was performed to select the best candidate biomarkers for the diagnosis of PAC. For the 40 selected VOCs, calculated areas under the curve (AUC) ranged from 0.98 to 0.85, and 11 of them were successfully validated using an independent set of samples (5 PAC patients, 5 healthy controls). Four of the proposed PAC biomarkers were identified as toluene, 2-ethyl-1-hexanol, pentylbenzene, and butoxymethylbenzene. Combinations of the identified PAC biomarkers were tested and showed AUC > 0.90, with the more promising candidate being butoxymethylbenzene (AUC $= 0.98$).

1. Introduction

Pancreatic adenocarcinoma (PAC) is one of the most lethal kinds of cancer, presenting concerning incidence and survival rate tendencies in the last years. Globocan 2020 data [1] shows PAC to be on the 12th place in the incidence rate ranking worldwide, while it presents the eighth leading mortality rate. The mortality and incidence trends of this cancer are concerning, with estimated increases on incidence (+77.7 %) and mortality (+79.9 %) from 2018 to 2040 [2,3]. Because of the pancreas location, the detection of a pancreatic tumor is not easy, and the PAC symptoms become evident only after the tumor has grown and/or spread to other organs. Those facts, and the lack of a specific diagnostic tool, make PAC to be usually diagnosed at stage III or IV, normally presenting inoperable tumors limiting their treatment and, consequently, leading to low survival rates: only 12 % of PAC patients survive after 5 years of diagnosis [4].

Biopsy is currently the only method for the unequivocal diagnosis of pancreatic cancer, supported by imaging techniques, while screening tests do not exist yet and blood tests are based in the determination of serum carcinoembryonic antigen (CEA) and cancer antigen 19–9 (CA 19–9), that are not specific [5,6].

Volatile organic compounds (VOCs) are final metabolism products, and can be detected in exhaled breath, blood, and other bodily fluids. Metabolic changes in response to external factors and intrinsic factors such as inflammation, necrosis, and diseases, including cancer, are reflected in VOCs extent; therefore, there is an increasing interest in developing diagnostic tests based in their detection [7,8]. The fluctuations of VOCs have been evidenced in individuals presenting health conditions like cystic fibrosis [9,10], cancer [11–16], diabetes [17,18], tuberculosis [19,20] , irritable bowel syndrome [21], or infectious diseases [22]. The biochemical pathways leading to the production of VOCs are not fully understood, but it is known that oxidative stress, some liver

* Corresponding author at: NYS Department of Health, Wadsworth Center, Empire State Plaza, Biggs Laboratory, Albany, NY 12201, United States.
E-mail addresses: maria-pilar.martinez-moral@health.ny.gov (M.-P. Martínez-Moral), maria-teresa.tena@unirioja.es (M.T. Tena), amcarnicero@riojasalud.es (A. Martín-Carnicero), amartinezr@riojasalud.es (A. Martínez).
[1] Present address: Wadsworth Center, New York State Department of Health, Biggs Laboratory, Corning Tower, ESP, 12201 Albany, NY, United States.

https://doi.org/10.1016/j.cca.2024.117895
Received 30 January 2024; Received in revised form 4 March 2024; Accepted 29 March 2024
Available online 30 March 2024

Table 1
Age and sex distribution of cancer patients and healthy volunteers in the study. Statistical tests employed: Chi square (*) or *t*-test (†).

Control – Case study			
	Healthy volunteers (n = 19)	Cancer patients (n = 20)	p-value
Sex (males, %)	12 (63 %)	13 (65 %)	0.77*
Age (years)	63.4 ± 7.7	65.2 ± 9.0	0.51†
Validation study			
	Healthy volunteers (n = 5)	**Cancer patients (n = 5)**	**p-value**
Sex (males, %)	2 (40 %)	2 (40 %)	1.00*
Age (years)	49.2 ± 16.9	67.0 ± 9.7	0.12†

enzymes such as alcohol dehydrogenase (ADH) or aldehyde dehydrogenase (ALDH), glycolysis, and loss of tumor suppressor genes, angiogenesis, or apoptosis, which are activated in cancer patients, alter the production of VOCs in the organism [23,24].

The study of new biomarkers for the detection of pancreatic cancer has been attempted before by the application of metabolomic and volatolomic analysis of exhaled breath and urine, showing promising results [25–31]. While blood or serum are biospecimens containing a great number of cell metabolites, studies focused in the volatolomic analysis of serum for the discovery of PAC biomarkers are lacking. Moreover, serum is collected by trained health staff in a controlled environment that avoids contaminations, and it is normally stored in biobanks, which allows the development of large retrospective studies.

The main goal of this work was the discovery of new PAC biomarkers in serum. The serum volatolome of PAC patients (n = 20) were compared with the controls (n = 19) to find those VOCs able to discriminate PAC patients from healthy population.

A sensitive PAC biomarker may be the keystone for the development of new diagnostic devices that could be used for the development of population screening strategies. The early diagnosis of pancreatic cancer would improve the treatment options and the prognosis of those patients.

2. Material and methods

2.1. Study design

Approval from the local ethic's review board (Comité Ético de Investigación Clínica de La Rioja, CEICLAR, study permit number CEICLAR P.I.260) was obtained and Informed Consent was signed by all patients and healthy donors. A total of 20 PAC patients diagnosed with unresectable locally advanced or metastatic pancreatic adenocarcinoma who had not received chemotherapy (65 % male, 65.2 ± 9.0 years of age) and a sex-age matched healthy control group (n = 19; 63 % male, 63.4 ± 7.7 years) participated in the study. Not statistical differences were found between groups in terms of sex or age. A second group of patients (n = 5, 40 % male, 67.0 ± 10.0 years) and healthy volunteers (n = 5, 40 % male, 40.2 ± 11 years) was recruited as a validation cohort (Table 1).

2.2. Sample collection

Samples for the control-case study were collected at the Hospital San Pedro facilities (Logroño, Spain) from October 2017 to November 2018. The validation set of serum samples was obtained in the same setting in 2021.

Venous blood was drawn in a tube without additives, centrifuged at 4 °C (3000*g* for 10 min) and the serum was collected and aliquoted in 1 ml portions, and stored at −80 °C for further analysis. The procedure blank was performed collecting ultrapure water in substitution of serum using the same materials, to evaluate potential contaminating volatile compounds derived from collection tubes, tubing and storage containers.

2.3. Volatolomic analysis

The HS-SPME-GC/MS analytical method for the volatolomic analysis of serum was optimized for the detection of the greatest number of signals with the best sensitivity. One of the critical steps of the SPME method development is the selection of the fiber. Different SPME fiber coatings were tested, including Carboxen/Polydimethylsiloxane (CAR/PDMS), Polydimethylsiloxane/Divinylbenzene (PDMS/DVB) and Divinylbenzene/Carboxen/Polydimethylsiloxane (DVB/CAR/PDMS). The latest, due to the mixed coating materials, was the fiber that extracted the highest number of volatile and semi-volatile compounds from serum and was selected for further analysis. The rest of analytical variables were optimized, and the final conditions are detailed below.

A Varian CP-3800 gas chromatograph coupled to a Saturn 2200 ion trap mass spectrometer equipped with a CTC CombiPal autosampler (Agilent, Madrid, Spain) was used for the non-targeted analysis of VOCs. The analysis was performed using a Zebron ZB-35 GC column (30 m, 0.25 mm ID, Phenomenex, Torrance, CA), with Helium at 1 ml min^{-1} as carrier gas.

Each serum sample was thawed at 4 °C and, after homogenisation, 250 μl were transferred to a 20 ml vial. 25 μl of surrogate standards D_{10}-ethylbenzene (d_{10}-EB) and D_4-1,4-dichlorobenzene (d_4-DCB) (Merck KGaA, Darmstadt, Germany) (1 μg/ml each) were added. The vial was closed with a magnetic screw cap for the HS-SPME analysis. The sample was incubated for 1 min at 50 °C and stirred at 250 rpm prior to extraction. Then, a 50/30 μm DVB/CAR/PDMS fiber (Supelco, Darmstadt, Germany) was used for the headspace extraction of VOCs at 50 °C for 20 min. Desorption was done at 250 °C for 30 s in the injection port. The injection was made with the closed split, that was open to 1:50 ratio at 1 min until the end of the analysis. Between analyses, the fiber was conditioned in the backout station for 15 min at 270 °C to avoid carryover effects.

The temperature program for VOCs separation started at 50 °C held for 15 min. Then, the temperature was raised 5 °C/min until 100 °C, held for 1 min at 100 °C, increased to 250 °C at 25 °C/min, and maintained for 2 min, with a total run time of 36 min. The VOCs signals were monitored in full scan mode, from 40 to 300 *m/z* and using electron ionization. Three technical replicates were processed for all samples. Analytical standards for identification tasks were provided by Merck Life Science (Madrid, Spain) and Cymit Quimica (Barcelona, Spain).

2.4. Quality assurance

An empty vial and a reagents blank were daily processed before the sample sequence started. Baseline level was checked and the labelled standards D_4-DCB and D_{10}-EB, that were added at a fixed amount (25 ng), were monitored to assure correct instrument performance.

2.5. Data collection and statistical analysis

The workflow design for data curation and statistical analysis was made according with literature [32–34]. MassHunter software (Agilent) was used for chromatogram processing including deconvolution of the signals, and exclusion criteria was set at area 150 threshold. A total of 433 components (V1-V433) were considered and corresponded with the sequential signals for the VOCs detected in the volatolome (Peak1-Peak433). For each VOC, the most abundant signal in the spectra was selected as the quantitative ion, and XIC chromatogram was manually integrated to get their area. Those components detected in the procedure blank at a similar intensity than the area found in serum were not

Table 2
Features selected after Bonferroni adjustment for multivariate analysis sorted by raw p-values. VOCs XIC normalized area average values for the control and case groups ± confidence interval at a significance level α = 0.05 (95 % CI) are shown. Fold change (FC), raw p-values and Bonferroni and BH corrected p-values are included.

VOC	Control Mean ± 95 % CI	PAC patient Mean ± 95 % CI	FC (P/C)	Raw p-value	Bonferroni corrected p-value	BH corrected p-value
V197	0.0161 ± 0.0010	0.0098 ± 0.0011	0.61	$6.39 \cdot 10^{-10}$ ✦	$2.77 \cdot 10^{-7}$	$2.77 \cdot 10^{-7}$
V328	$(4.17 \pm 0.23) \cdot 10^{-3}$	$(2.98 \pm 0.19) \cdot 10^{-3}$	0.72	$1.94 \cdot 10^{-9}$ †	$8.40 \cdot 10^{-7}$	$4.20 \cdot 10^{-7}$
V332	$(3.42 \pm 0.22) \cdot 10^{-3}$	$(2.40 \pm 0.20) \cdot 10^{-3}$	0.70	**$2.12 \cdot 10^{-7}$** †	$9.18 \cdot 10^{-5}$	$2.16 \cdot 10^{-5}$
V38	0.44 ± 0.04	0.29 ± 0.03	0.67	$2.67 \cdot 10^{-7}$ ✦	$1.16 \cdot 10^{-4}$	$2.16 \cdot 10^{-5}$
V40	94 ± 6	72 ± 4	0.77	$2.78 \cdot 10^{-7}$ ✦	$1.20 \cdot 10^{-4}$	$2.16 \cdot 10^{-5}$
V365	$(7.4 \pm 0.4) \cdot 10^{-3}$	$(5.3 \pm 0.5) \cdot 10^{-3}$	0.72	$3.92 \cdot 10^{-7}$ ✦	$1.70 \cdot 10^{-4}$	$2.16 \cdot 10^{-5}$
V39	0.74 ± 0.07	0.47 ± 0.04	0.64	$3.95 \cdot 10^{-7}$ ✦	$1.71 \cdot 10^{-4}$	$2.16 \cdot 10^{-5}$
V172	4.0 ± 0.3	2.6 ± 0.3	0.65	$4.03 \cdot 10^{-7}$ ✧	$1.75 \cdot 10^{-4}$	$2.16 \cdot 10^{-5}$
V174	0.030 ± 0.003	0.018 ± 0.003	0.58	$4.48 \cdot 10^{-7}$ ✦	$1.94 \cdot 10^{-4}$	$2.16 \cdot 10^{-5}$
V41	0.59 ± 0.06	0.37 ± 0.04	0.63	$6.59 \cdot 10^{-7}$ ✦	$2.85 \cdot 10^{-4}$	$2.85 \cdot 10^{-5}$
V390	$(2.85 \pm 0.18) \cdot 10^{-3}$	$(2.03 \pm 0.20) \cdot 10^{-3}$	0.71	$1.00 \cdot 10^{-6}$ ✧	$4.34 \cdot 10^{-4}$	$3.94 \cdot 10^{-5}$
V171	0.343 ± 0.025	0.239 ± 0.025	0.70	$1.16 \cdot 10^{-6}$ ✦	$5.02 \cdot 10^{-4}$	$4.19 \cdot 10^{-5}$
V51	$(3.3 \pm 0.3) \cdot 10^{-3}$	$(2.21 \pm 0.21) \cdot 10^{-3}$	0.68	$1.32 \cdot 10^{-6}$ ✦	$5.72 \cdot 10^{-4}$	$4.40 \cdot 10^{-5}$
V34	$(6.3 \pm 0.5) \cdot 10^{-3}$	$(4.4 \pm 0.4) \cdot 10^{-3}$	0.69	$2.02 \cdot 10^{-6}$ ✦	$8.76 \cdot 10^{-4}$	$6.25 \cdot 10^{-5}$
V338	$(4.1 \pm 0.3) \cdot 10^{-3}$	$(2.9 \pm 0.3) \cdot 10^{-3}$	0.73	$2.23 \cdot 10^{-6}$ †	$9.66 \cdot 10^{-4}$	$6.25 \cdot 10^{-5}$
V37	0.196 ± 0.015	0.139 ± 0.014	0.71	$2.31 \cdot 10^{-6}$ ✦	0.0010	$6.25 \cdot 10^{-5}$
V311	$(1.57 \pm 0.11) \cdot 10^{-3}$	$(1.14 \pm 0.16) \cdot 10^{-3}$	0.73	$2.71 \cdot 10^{-6}$ †	0.0012	$6.90 \cdot 10^{-5}$
V142	0.339 ± 0.023	0.24 ± 0.03	0.71	$3.71 \cdot 10^{-6}$ ✦	0.0016	$8.92 \cdot 10^{-5}$
V36	0.166 ± 0.015	0.108 ± 0.013	0.65	$3.95 \cdot 10^{-6}$ †	0.0017	$9.00 \cdot 10^{-5}$
V131	$(3.8 \pm 0.4) \cdot 10^{-3}$	$(2.5 \pm 0.3) \cdot 10^{-3}$	0.65	$5.82 \cdot 10^{-6}$ ✦	0.0025	$1.23 \cdot 10^{-4}$
V52	0.39 ± 0.04	0.277 ± 0.023	0.71	$5.95 \cdot 10^{-6}$ ✦	0.0026	$1.23 \cdot 10^{-4}$
V173	0.028 ± 0.003	0.018 ± 0.003	0.65	$6.48 \cdot 10^{-6}$ ✦	0.0028	$1.28 \cdot 10^{-4}$
V141	27.4 ± 1.6	20.9 ± 1.9	0.76	$6.80 \cdot 10^{-6}$ †	0.0029	$1.28 \cdot 10^{-4}$
V362	$(1.02 \pm 0.09) \cdot 10^{-3}$	$(0.71 \pm 0.07) \cdot 10^{-3}$	0.70	$7.88 \cdot 10^{-6}$ ✦	0.0034	$1.40 \cdot 10^{-4}$
V140	1.20 ± 0.07	0.93 ± 0.08	0.77	$8.11 \cdot 10^{-6}$ †	0.0035	$1.40 \cdot 10^{-4}$
V29	0.090 ± 0.006	0.065 ± 0.008	0.72	$9.89 \cdot 10^{-6}$ ✦	0.0043	$1.65 \cdot 10^{-4}$
V380	0.0282 ± 0.0015	0.0215 ± 0.0021	0.76	$1.12 \cdot 10^{-5}$ ✦	0.0049	$1.78 \cdot 10^{-4}$
V139	0.30 ± 0.03	0.21 ± 0.03	0.69	$1.15 \cdot 10^{-5}$ †	0.0050	$1.78 \cdot 10^{-4}$
V28	0.059 ± 0.003	0.044 ± 0.005	0.74	$1.24 \cdot 10^{-5}$ ✦	0.0054	$1.85 \cdot 10^{-4}$
V187	0.024 ± 0.003	0.0144 ± 0.0019	0.61	$1.36 \cdot 10^{-5}$ †	0.0059	$1.96 \cdot 10^{-4}$
V170	0.0095 ± 0.0012	0.0057 ± 0.0012	0.61	$1.89 \cdot 10^{-5}$ †	0.0082	$2.64 \cdot 10^{-4}$
V68	0.0164 ± 0.0016	0.0111 ± 0.0015	0.67	$2.38 \cdot 10^{-5}$ ✦	0.0103	$3.22 \cdot 10^{-4}$
V349	0.0047 ± 0.0005	0.0031 ± 0.0005	0.65	$2.76 \cdot 10^{-5}$ ✦	0.0120	$3.62 \cdot 10^{-4}$
V386	0.0213 ± 0.0019	0.013 ± 0.003	0.59	$3.05 \cdot 10^{-5}$ †	0.0132	$3.88 \cdot 10^{-4}$
V169	0.0133 ± 0.0017	0.0075 ± 0.0017	0.57	$3.83 \cdot 10^{-5}$ ✦	0.0166	$4.74 \cdot 10^{-4}$
V97	0.0118 ± 0.0013	0.0076 ± 0.0013	0.64	$4.15 \cdot 10^{-5}$ †	0.0180	$4.99 \cdot 10^{-4}$
V364	$(1.99 \pm 0.15) \cdot 10^{-3}$	$(1.45 \pm 0.18) \cdot 10^{-3}$	0.73	$5.96 \cdot 10^{-5}$ ✦	0.0258	$6.97 \cdot 10^{-4}$
V145	0.034 ± 0.005	0.021 ± 0.003	0.62	$8.69 \cdot 10^{-5}$ †	0.0376	$9.65 \cdot 10^{-4}$
V274	0.027 ± 0.003	0.019 ± 0.003	0.71	$8.69 \cdot 10^{-5}$ †	0.0376	$9.65 \cdot 10^{-4}$
V94	0.048 ± 0.005	0.031 ± 0.006	0.65	$1.15 \cdot 10^{-4}$ †	0.0498	0.0012

✦ *t*-test equal variances.
✧ *t*-test unequal variances.
† Mann Whitney *U* test.

considered. A peak table was constructed for the 39 samples calculating the average area of the three technical replicates for each of the 433 VOCs.

Descriptive statistics were calculated for all signals. Sample normalization was accomplished calculating the ratio of the peak area of the signal to the respective peak area of the internal standard D_4-DCB in the same chromatogram. For each feature, the control and case groups were tested for normality using a Shapiro Wilk test and homoscedasticity with an F test. Depending on normality of the signal, Mann-Whitney *U* test or two tail *t*-test were used to evaluate the significance of control-case average area differences. Data was filtered removing those signals with high variability or low differences between groups, thus after univariate statistics only those VOCs with corrected p-value < 0.05 for the inter-group mean comparison were further considered. Raw p-values were corrected for multiple comparisons by Bonferroni method

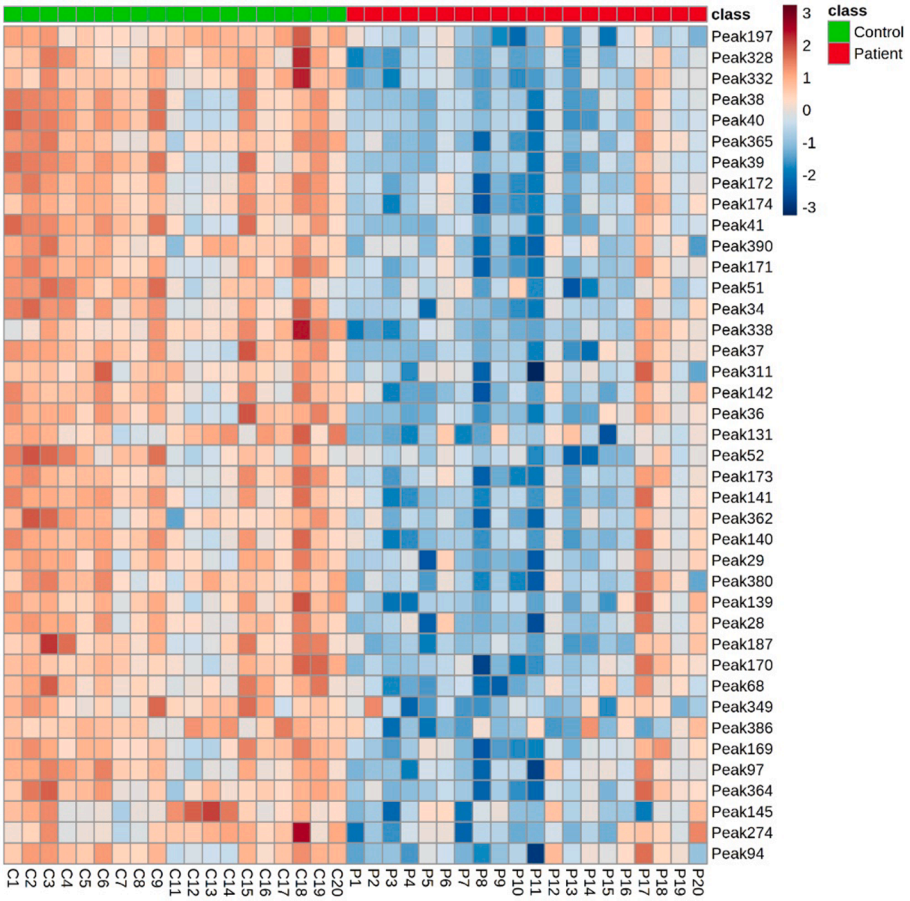

Fig. 1. Heatmap showing different patterns for the VOCs expressions on patients and healthy controls.

[33,35]. Prior to multivariate analysis, data was log transformed, mean-centered, and divided by standard deviation of each variable for scaling. Unsupervised principal component analysis (PCA) was performed to find potential outliers and to evaluate the discriminating capability of the volatolomic profile. Partial least square discriminant analysis (PLS-DA) was used to find the most significant VOCs for the discrimination of PAC patients from healthy volunteers. To evaluate the accuracy of the most suitable signals to be used as biomarkers for PAC prediction, receiver operating characteristic (ROC) analysis was performed finding the corresponding area under de curve (AUC). AUC values from 0.9 to 1.0 were considered "excellent", "good" from 0.8 to 0.9, "fair" from 0.7 to 0.8, "poor" from 0.7 to 0.6 and those values below 0.6 were considered failed [35].

Samples from the validation cohort were analysed following the proposed procedure, and their results were compared with the respective cut-off point for group assignation (control or case) and evaluation of the false positive (FP) and false negative (FN) events.

2.6. Identification

For those VOCs showing potential as PAC biomarkers, tentative identification was made using MassHunter software by mass spectra match using NIST17 library. Among NIST candidates for each VOC, for those with Rmatch above 50 %, commercially available analytical standards were purchased (Merck, Darmstadt, Germany and Cymit Quimica, Barcelona, Spain). The standards were tested using the same HS-SPME-GC/MS method; and confirmed if the unknown and standard eluted at coincident GC retention time, and with similar MS spectra match with the library record.

2.7. Software

MS Workstation (Varian) was used for data acquisition and MassHunter (Agilent) for chromatogram processing and tentative identification using NIST17 library. R software and Metaboanalyst [32,34] were used for statistics and graphs. Microsoft Excel was used as database and for graph drawing.

3. Results and discussion

3.1. Univariate statistics

The most relevant VOCs for the discrimination of pancreatic cancer from healthy individuals started with the univariately consideration of the 433 signals found in the volatolomic profile. Two normalization approaches for these signals were tested: the relative area to d_{10}-EB, and the relative area to d_4-DCB. Since d_4-DCB detection presents high inter-analysis stability, this internal standard was selected to account for HS-SPME-GC/MS method variability.

Out of the 433 signals, a p-value < 0.05 was obtained for 125 VOCs when comparing the control and case means. Within those significant features, 96 had a p-value < 0.01, and 64 a p-value < 0.001 when individual inter-group difference significance tests were applied. We found 87 % of those significant VOCs to be downregulated for the patients, whereas only 13 % were upregulated. Raw p-values were subjected to multivariate correction by Bonferroni and Benjamini&Hochberg (BH) methods. The BH method, based in false discovery rate, found adjusted p-values < 0.05 for 96 VOCs; 77 of them were below 0.01, and 40 had a p-value < 0.001. After Bonferroni correction of family-wise error, p-values were < 0.05 for 40 VOCs, from

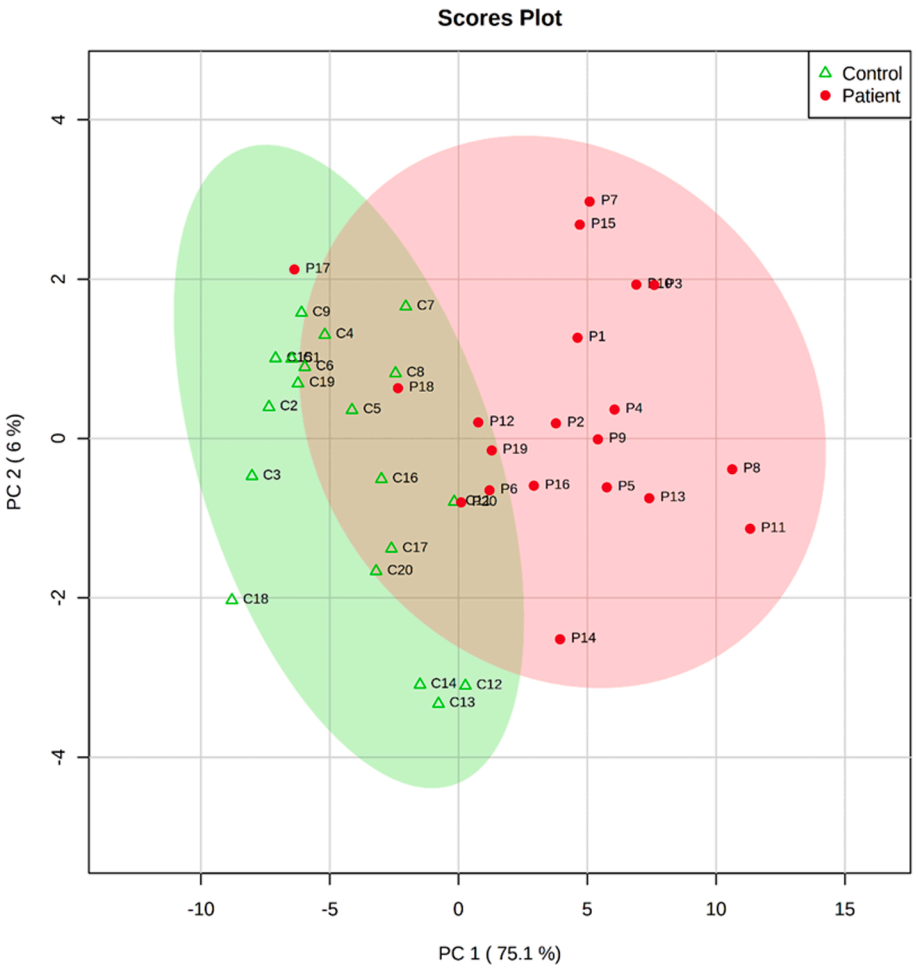

Fig. 2. Principal component analysis showing 95% CI areas for the control and patient groups. Accumulated explained variance for the two represented components is 81%.

which 31 had a p-value < 0.01, and 15 of them had a p-value below 0.001; all of them were downregulated for patients, with fold changes ranging from 0.57 to 0.77. To prevent false positive errors, the most conservative method was selected; therefore, Bonferroni adjusted p-values were used for data filtering. Those VOCs with p-value < 0.05 were further considered (Table 2), excluding from the study those signals with high intra-group variability and low inter-group change of intensity. Chromatographic parameters, precision of the method, and absence of significant signal in procedure blanks for those VOCs were checked and are reported in Table S1.

3.2. Multivariate analysis

Prior to multivariate analysis, data were normalized by the calculation of the relative response against D_4-DCB, log transformed, and scaled. Sample and feature view normalization results were visually checked (Fig. S2). In the sample view graph, certain difference between the controls and patients was visually observed (Fig. S2).

Unsupervised analysis was performed by PCA and heatmap construction to evaluate the presence of outliers among the samples. The goals of those approaches are the observation of sample patterns and the reduction of the number of variables to explain the total variability for all the samples in the study. The heatmap shows VOCs variability for the samples grouped for controls and patients (Fig. 1). Differences between the groups are appreciated, with the PAC's group presenting down-regulated VOC values. Furthermore, the signal for P17 seems to have a slightly different pattern when compared with the other patients in the heatmap.

In the PCA analysis, the first component explained 75 % of the total variance. The representation of the two most explicative components calculated by PCA with reference in the graph to the control and case groups with the 95 % CI area (Fig. 2) shows certain separation between the groups, reaching an accumulated explained variance of 81 % for the two first components, as shown in the scree plot (Fig. S3). In agreement with the conclusions from the heatmap observation, certain differentiation between controls and patients was observed, sample P17 was outside the 95 %CI for its group and, consequently, it may be considered an outlier. Since data from sample P17 is not aberrant and may be relevant to show real variability in the dataset, all further statistics and tests were performed including the integral acquired data after confirmation that similar final conclusions would be obtained if this sample was not included in the study (Fig. S4).

A supervised classification was done to get a model with a reduced number of variables that explains differences between controls and patients. Sparse PLS-DA was used to get two components built with 10 variables each, to explain intergroup variability. The separation between the control and case groups was evident, the first component explained 71 % of the total variance, and the two main components 76 % of the total variance (Fig. 3). Variables contributing to component 1 were VOCs 328, 197, 332, 39, 41, 38, 40, 365, 172 and 174, while component 2 was built with VOCs 386, 145, 169, 173, 94, 97, 171, 170, 364 and 131.

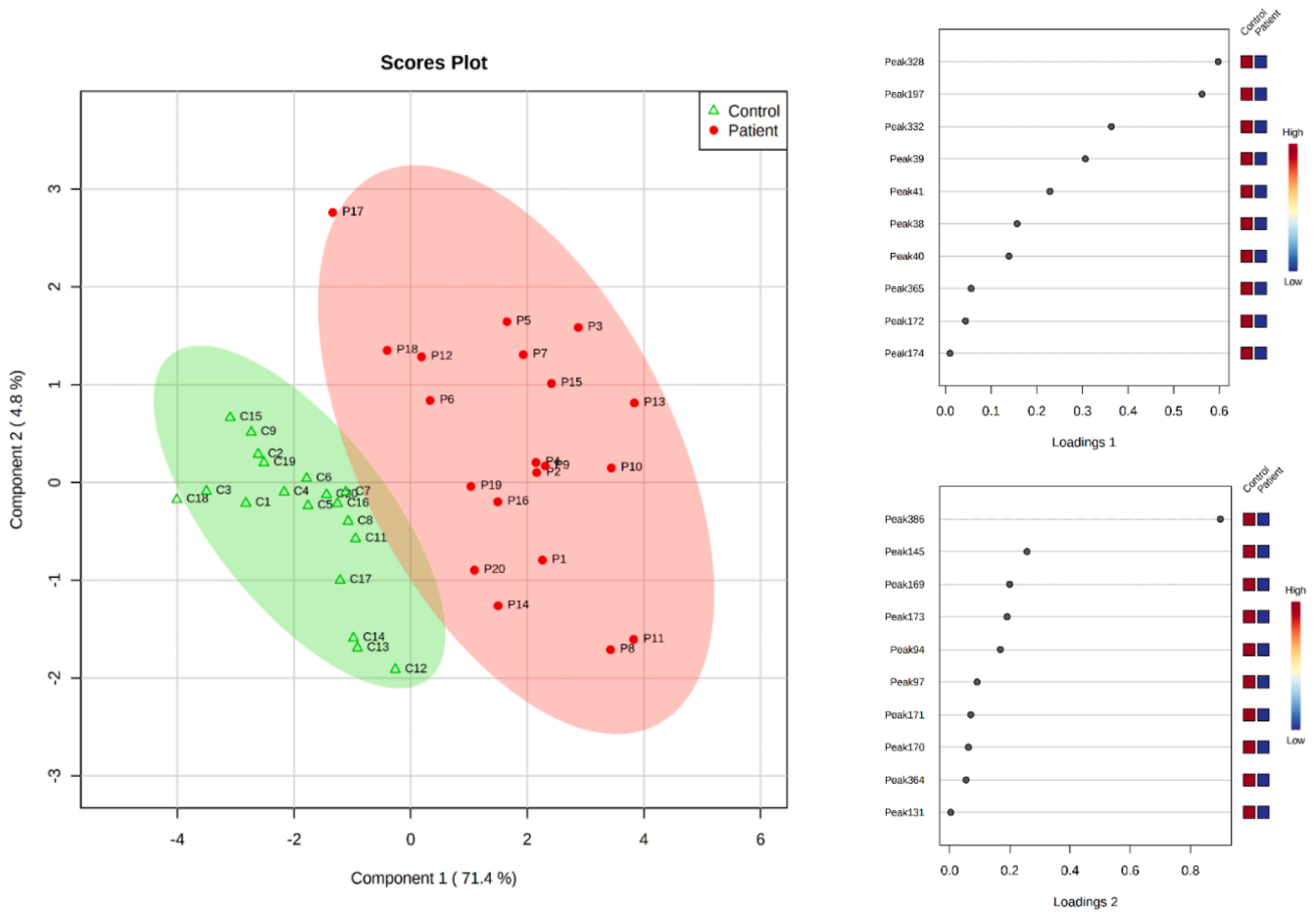

Fig. 3. Sparse PLS-DA was applied to the data to get a model with a reduced number of variables. The control-case separation is shown through the representation of the scores for the two first components, built with 10 variables each. VIP scores of the variables used for the components are shown.

3.3. Biomarker analysis

For the evaluation of their predictive capability, ROC curves were constructed for pre-selected VOCs. The AUC and the cut-off values for each signal are shown in Table 3.

Nineteen VOCs showed excellent predictive potential, with AUC above 0.90. The best results were for V197 with an AUC value of 0.98 ± 0.05, with 90 % sensitivity (true positive rate) and 95 % specificity (1-false positive rate); and for V328 with 0.98 ± 0.06, 95 % sensitivity and 95 % specificity (Fig. 4). Features V39, V332, V40, V41, V365, V172, V390, V34, V174, V38, V338, V311, and V171 also are considered excellent predictors based on their AUCs. Other VOCs showed good AUC values, above 0.85 for all cases, and may be also suitable PAC biomarker candidates.

3.4. Sex and age effects in VOC expression

It is known that the volatolome may change depending on lifestyle, diet, or disease. Our goal is to find those VOCs able to discriminate PAC disease patients from the healthy population, but since the variance of their expression may be influenced by covariates, sex and age were studied on the selected panel of VOCs as potential confusing factors. Considering all participants, differences between the male (n = 25) and female (n = 14) groups were evaluated by a *t*-test (2 tails, equal variances, $\alpha = 0.05$) for each selected VOC. The only significant difference was found for V311, with a lower signal in females ($(1.19 \pm 0.19) \cdot 10^{-3}$) than in males ($(1.44 \pm 0.14) \cdot 10^{-3}$) (p-value = 0.03).

Sex effects were also evaluated in both the control and patient groups. Interestingly, signals for the control females and males were equal, while significant differences were found for 20 VOCs in the patient's group (Table 4). The most significant differences (p-value < 0.01) were found for V34 and V68, with higher area values in males. The other 18 VOCs that showed differences between males and females for the patient group, presented p-values from 0.02 to 0.04, and were all upregulated in males. Since all diagnostic biomarkers were down-regulated for the patient group, the fact that female patients present lower signal than males ($\alpha = 0.05$) may indicate sex differences in the prediction capacity of our model. p-Values for the difference between controls and patients were calculated for the male and the female population independently; although in most cases p-values in the female population were lower (but for V34), in all cases the VOCs signal means difference between controls and patients was statistically different for both sexes (Table S5).

The relation between age and biomarker expression was studied by a Pearson correlation test. Only one feature, V349, showed a significant correlation with age (p-value = 0.018) when all participants were studied, but with a correlation factor that indicates poor negative correlation (R = −0.38). No correlations between age and VOCs intensities were observed within the control group nor for the patient group.

3.5. Tobacco and alcohol intake influence in volatolome

An additional evaluation of some lifestyle factors such as alcohol intake and smoking were done for the PAC patient group. Volatolomic profile may change depending on lifestyle, and studying if those factors may have an effect on the selected VOCs is of importance. Individuals in the smoker group (n = 10) declared to consume cigarettes daily. Results showed that smoking had a significant effect only on V274, with higher

Table 3
Results of the ROC analysis of the signals showing a significant difference between the control and case groups. AUC with 95% CI, sensitivity, specificity and cut-off point are included.

Biomarker	AUC	95 % CI	Cut-off	Sensitivity	Specificity
V197	0.984	(0.941–1.000)	0.0129	0.90	0.95
V328	0.982	(0.929–1.000)	0.00351	0.95	0.95
V39	0.953	(0.870–1.000)	0.523	0.85	0.95
V332	0.942	(0.834–1.000)	0.00278	0.90	1.00
V40	0.934	(0.826–0.989)	80.9	0.90	0.85
V41	0.934	(0.839–0.987)	0.448	0.90	0.85
V172	0.932	(0.834–0.995)	3.205	0.85	0.85
V365	0.929	(0.829–0.995)	0.00610	0.85	0.95
V390	0.929	(0.821–0.997)	0.0026	0.95	0.90
V34	0.929	(0.826–0.992)	0.00525	0.85	0.90
V174	0.921	(0.809–0.992)	0.0238	0.85	0.90
V38	0.917	(0.809–0.984)	0.348	0.90	0.85
V338	0.911	(0.800–0.992)	0.00345	0.85	0.95
V311	0.911	(0.787–0.983)	0.00131	0.85	0.90
V171	0.907	(0.797–0.987)	0.281	0.85	0.85
V173	0.905	(0.779–0.986)	0.0229	0.85	0.85
V37	0.903	(0.784–0.978)	0.172	0.90	0.80
V380	0.903	(0.766–0.989)	0.0250	0.85	0.90
V36	0.900	(0.792–0.984)	0.142	0.90	0.80
V51	0.897	(0.785–0.972)	0.0027	0.85	0.80
V141	0.897	(0.768–0.982)	23.0	0.75	0.90
V131	0.897	(0.778–0.971)	0.00305	0.80	0.80
V362	0.895	(0.754–0.987)	0.000870	0.90	0.90
V140	0.895	(0.754–0.984)	0.995	0.75	0.95
V68	0.895	(0.758–0.976)	0.0134	0.80	0.85
V142	0.892	(0.758–0.978)	0.269	0.75	0.95
V364	0.889	(0.759–0.983)	0.00165	0.80	0.90
V29	0.887	(0.757–0.971)	0.0777	0.85	0.85
V28	0.887	(0.759–0.976)	0.0515	0.85	0.85
V139	0.886	(0.748–0.982)	0.232	0.80	0.95
V187	0.884	(0.745–0.978)	0.0185	0.85	0.85
V52	0.884	(0.763–0.961)	0.322	0.90	0.75
V170	0.879	(0.755–0.971)	0.00726	0.80	0.85
V349	0.871	(0.734–0.966)	0.00359	0.75	0.90
V386	0.871	(0.738–0.968)	0.0181	0.85	0.85
V97	0.866	(0.717–0.968)	0.00904	0.80	0.80
V169	0.863	(0.710–0.963)	0.0102	0.85	0.85
V145	0.853	(0.712–0.953)	0.0249	0.75	0.85
V274	0.853	(0.716–0.963)	0.0228	0.80	0.80
V94	0.850	(0.701–0.947)	0.0430	0.90	0.75

values for smokers (0.022 ± 0.004 (n = 10)) than for non-smokers (0.016 ± 0.002 (n = 10)) (p-value = 0.02). No differences were found for the groups based on alcohol intake.

3.6. Study of markers for the prediction of survival time

The prediction of the evolution of a PAC patient would be important for the decision-making regarding treatment strategies. The complete volatolome (n = 433 signals) was included in the comparison of two groups based on the patients that survived more than one year after diagnostic or not. Out of the 20 patients that participated in the study, 19 received chemotherapy treatment. Ten patients survived more than 1 year (from 12 to 24 months), whereas the other ten patients' survival was lower than 7 months. After univariant mean comparison, 16 VOCs showed differences between the groups (α = 0.05) (Table S6). While in the control-case study we found that VOCs expression was generally downregulated for the patients, it was observed that the overall response of the significant features was downregulated for those patients that survived longer; except for V13 and V219. Only one of those signals that was related with the response to treatment, V386, was also found among the proposed biomarkers for PAC detection (AUC 0.87). Surprisingly, while the healthy controls present a higher value for V386 (0.0213 ± 0.0019) than for patients (0.013 ± 0.003), when we look within the patient group, those that have a better prognosis present lower value (0.009 ± 0.002) than those that survived less than 1 year (0.015 ± 0.004). This behaviour was also observed for other three VOCs with significant difference between controls and patients' groups: V345, V107, and V323. A different trend was found for V219, that was upregulated for controls and when the survival time was studied the longer survival corresponded with the upregulated group. Those are results from the raw p-value calculation, considering univariant comparisons; when those p-values were corrected by Bonferroni or BH methods, all tests for the comparison of groups based on survival for more than 1 year lost their significance.

The ROC analysis for those VOCs showing significant differences between the long-short survival time groups (α = 0.05), gave excellent predictive accuracy for V109 (AUC = 0.93) with 100 % sensitivity and 89 % specificity. Good accuracies were found for V345 (AUC = 0.90) and V107 (AUC = 0.87); and AUCs were < 0.8 for the other considered VOCs (Table S7). These preliminary results should be confirmed by the analysis of a greater number of samples, and ideally with a setting that minimize variability due to different treatment strategies.

3.7. Validation of the diagnostic model

The diagnostic model based in the relative intensity of the selected PAC biomarkers was validated by the analysis of an independent cohort. The validation samples (n = 10) were collected along 2021 at Hospital San Pedro (Logroño). Serum samples were collected from 5 PAC patients and 5 controls, and they were analysed following the proposed method. Using signal V197, that showed the most promising results after the modelling experiment, all samples were correctly classified. Among the other selected VOCs with biomarker potential, V362, V274, V94, V364, V145, V187, and V365 classified the 5 controls and the 5 patients correctly. For V328, that had a similar AUC with V197, out of the 10 samples one false positive was found. One false positive was also detected for V171, and one false negative for V170. Since the accuracy of the prediction was 100 % for V197, V362, V274, V94, V364, V145, V187, and V365; and 90 % for V328, V170 and V171, those eleven VOCs were considered validated biomarkers for PAC (Table 5).

The other VOCs had different results for false positive and false negative. Interestingly, for some of them (V380, V338, V172, V51, V131, V173, V68, V38, V386, V169, V332, and V97) we observed a significant difference (α = 0.05) between the signals for controls and patients (Table 5) that may indicate to be consistent regarding differences between groups, but the cut-off points were not consistent between the control-case set used to establish the model and the validation set. This may be due to a change of sensitivity of the equipment or other factors. Although good results were obtained following this strategy, ideally, those VOC levels in serum should be determined to give an accurate concentration cut-off, using a SIM acquisition method, and normalized by independent labelled ISs for each biomarker; but this would be only viable after the chemical identification of the biomarkers.

3.8. VOCs identification

The results from the control-case study are highly promising regarding the potential discovery of a biomarker capable to detect pancreatic cancer using volatolomic information. For the transference of this discovery to clinical application, for example as the basis for the design of a portable point-of-care device for the in-situ detection of PAC, the elucidation of the chemical structure of the most promising VOCs is needed. With this goal in mind, for those VOCs with AUC above 0.90 and that correctly classified the validation set, we performed NIST library search for correspondences with their MS spectra. Among the proposed compounds, we selected the highest score ranked (threshold was set at 50 %) with concordance for the proposed structure for at least 30 % of the samples (n = 10). Those analytical standards that were commercially available at reasonable conditions were purchased and an aliquot was analysed following the same HS-SPME-GC/MS method. The RT for the standard was compared with the RT found for the unknown compound in serum (Table S8). We successfully found coincidence of RT for four of

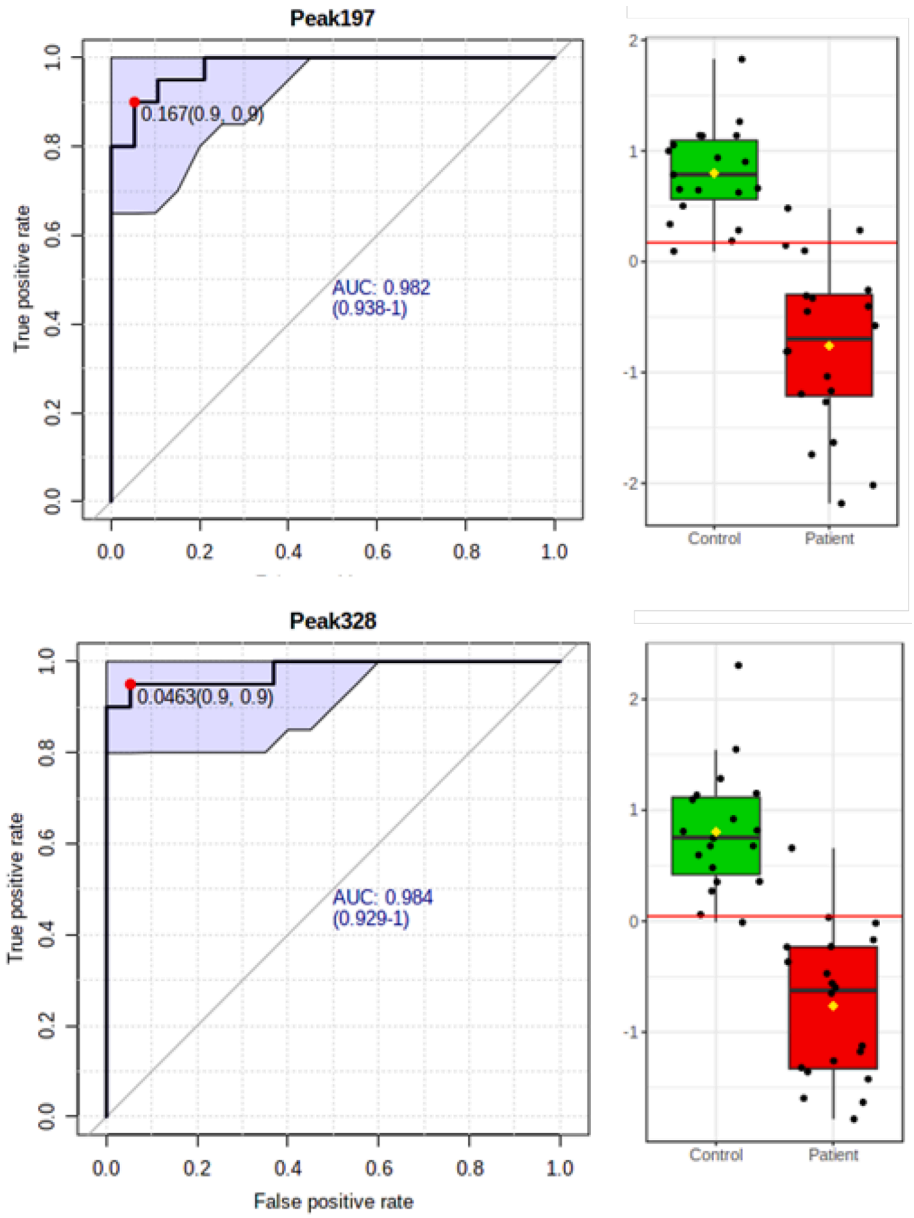

Fig. 4. ROC curves with 95% CI and boxplot for the two most significant features: V197 (AUC 0.98, 95% sensitivity and 95% specificity) and V328 (AUC 0.98, 95% sensitivity and 95% specificity). Data was log transformed and auto scaled prior to ROC analysis, raw cut-off values are shown in Table 3.

the proposed VOCs: V40 with toluene, V172 with 2-ethyl-1-hexanol, V274 with pentylbenzene, and V328 corresponded with butoxymethylbenzene, also known as benzyl butyl ether. A similar match ratio (Rmatch) after comparison with the NIST record was found for the VOC from serum and the analytical standard (Fig. 5) when the mass spectra were obtained by the proposed HS-SPME-GC/MS method. Since RT and mass spectra were matching with corresponding analytical standards, those four VOCs: toluene, butoxymethylbenzene, 2-ethyl-1-hexanol, and pentylbenzene were considered correctly identified. Among these biomarkers, pentylbenzene presented the best results in the validation study after correct classification of all the tested samples.

Toluene, pentylbenzene, and butoxymethylbenzene are benzenoids. Toluene in the body is exogenous, and it is metabolized by cytochrome P-450 (CYP) [36]. It has been detected in human blood, urine, breath, feces, and saliva [37]. Abnormal levels have been correlated with diseases [7,38], its absence has been related with *Clostridium difficile* infection [39], and other authors found positive correlation between exposure to toluene and thyroid cancer occurrence [40]. Our findings indicate that the levels of these benzenoids in pancreatic cancer patients are lower than those for healthy controls (Table 2, data for V40, V328 and V274). A possible explanation may be that, under carcinogenesis conditions, cytochrome p450 enzymes are overactivated and metabolize these benzenoids, decreasing their levels in blood [16]. But this explanation would require thorough future examination.

2-Ethyl-1-hexanol is a fatty alcohol that is metabolized in the body by ADH, and is excreted in breath, urine, sweat, feces and saliva [16,37]. Higher levels of 2-ethyl-1-hexanol in saliva were found in healthy children than in celiac children along with alterations of salivary microbiota [41]. It has been pointed out, by several authors, as a potential biomarker for lung cancer, with elevated levels in breath, urine, pleural effusions and cell lines [11,16,42]. In this work, we found significantly decreased levels of 2-ethyl-1-hexanol in pancreatic cancer serum compared with controls (Table 2, data for V172). We could propose the hypothesis that PAC patients are affected by elevated ADH activity [43], that may lead to decreased 2-ethyl-1-hexanol in blood, but this supposition would need further investigation.

Table 4
Sex and age effects in the volatolome. p-Values from the 2-tails *t*-test that were applied for all the participants and for the control and PAC groups are shown. The correlation between age and VOCs intensity was tested by Pearson correlation tests. Statistically significant p-values are indicated in red. *: $p < 0.05$; **: $p < 0.01$.

	Participants (n = 39)		Controls (n = 19)		Patients (n = 20)				
Biomarker	Male-Female, p-val	Age, R	Male-Female, p-val	Age, R	Male-Female, p-val	Age, R	Alcohol intake Yes/No, p-val	Smoke, Yes/No, p-val	Survival 1y Yes/No, p-val
V197	0.40	-0.12	0.96	0.08	0.09	-0.12	0.865	0.709	0.243
V328	0.65	-0.05	0.57	-0.05	0.51	0.11	0.959	0.122	0.828
V39	0.56	-0.14	0.82	-0.25	0.07	0.05	0.452	0.967	0.810
V332	0.49	-0.11	0.49	-0.02	0.35	-0.05	0.586	0.245	0.999
V40	0.67	-0.14	0.80	-0.25	0.17	0.05	0.531	0.992	0.862
V41	0.45	-0.17	0.99	-0.28	0.06	0.01	0.419	0.864	0.811
V172	0.23	-0.24	0.93	-0.24	0.02*	-0.24	0.797	0.430	0.932
V365	0.44	-0.18	0.35	-0.11	0.03*	-0.17	0.531	0.821	0.486
V390	0.42	-0.11	0.39	-0.06	0.04*	-0.04	0.518	0.737	0.558
V34	0.10	-0.26	0.63	-0.32	0.003**	-0.21	0.745	0.548	0.425
V174	0.17	-0.24	0.82	-0.08	0.03*	-0.30	0.756	0.207	0.984
V38	0.45	-0.17	0.95	-0.26	0.06	-0.02	0.570	0.732	0.939
V338	0.61	-0.01	0.61	0.22	0.53	0.02	0.743	0.161	0.705
V311	0.03*	-0.19	0.12	-0.21	0.02*	-0.14	0.107	0.970	0.627
V171	0.26	-0.25	0.94	-0.25	0.03*	-0.23	0.733	0.391	0.928
V173	0.16	-0.24	0.84	-0.18	0.02*	-0.24	0.688	0.326	0.880
V37	0.37	-0.11	0.70	-0.21	0.18	0.06	0.667	0.526	0.775
V380	0.49	-0.05	0.13	0.15	0.04*	-0.02	0.264	0.971	0.791
V36	0.18	-0.16	0.60	-0.20	0.03*	-0.07	0.760	0.515	0.800
V51	0.77	-0.10	0.60	-0.18	0.93	0.07	0.651	0.509	0.370
V141	0.30	-0.25	0.76	-0.22	0.11	-0.25	0.396	0.641	0.172
V131	0.72	-0.07	0.85	0.04	0.59	-0.02	0.533	0.099	0.331
V362	0.57	-0.24	0.26	-0.22	0.04*	-0.22	0.579	0.748	0.560
V140	0.31	-0.24	0.68	-0.22	0.15	-0.23	0.474	0.558	0.151
V68	0.10	-0.30	0.90	-0.18	0.008**	-0.37	0.543	0.900	0.792
V142	0.16	-0.31	0.61	-0.14	0.04*	-0.41	0.569	0.550	0.213
V364	0.37	-0.21	0.22	-0.04	0.02*	-0.25	0.482	0.516	0.609
V29	0.10	-0.19	0.71	-0.25	0.02*	-0.13	0.811	0.552	0.923
V28	0.16	-0.17	0.81	-0.25	0.04*	-0.08	0.806	0.404	0.951
V139	0.27	-0.22	0.34	-0.21	0.25	-0.19	0.659	0.223	0.091
V187	0.84	-0.10	0.69	-0.04	0.38	-0.02	0.841	0.769	0.443
V52	0.63	-0.11	0.39	-0.20	0.91	0.06	0.943	0.730	0.636
V170	0.13	-0.15	0.92	0.13	0.02*	-0.22	0.755	0.474	0.643
V349	0.08	-0.38*	0.22	-0.35	0.04*	-0.44	0.849	0.690	0.068
V386	0.90	-0.17	0.89	0.01	0.99	-0.20	0.820	0.746	0.009**
V97	0.20	-0.18	0.83	-0.18	0.03*	-0.12	0.264	0.782	0.891
V169	0.17	-0.28	0.84	-0.17	0.04*	-0.33	0.759	0.194	0.823
V145	0.75	0.04	0.77	-0.05	0.51	0.27	0.916	0.606	0.539
V274	0.70	-0.02	0.44	0.11	0.91	0.02	0.602	0.020*	0.341
V94	0.34	-0.10	0.50	-0.01	0.06	-0.06	0.103	0.724	0.786

Building a diagnostic method based in known VOCs is advantageous since their serum concentration for controls and patients may be accurately determined, improving robustness of the diagnostic method. Best individual response was found for butoxymethylbenzene, with AUC = 0.98. Toluene and 2-ethyl-1-hexanol AUC were 0.93, pentylbenzene's AUC was 0.85. With the goal of improving the diagnostic method, combinations of these normalized and scaled VOCs were considered and subjected to ROC analysis (Table S9). All tested options gave AUC > 0.90, with values ranging from 0.97 to 0.91, showing excellent predictive accuracy. While those results were slightly better than for individual toluene, 2-ethyl-1-hexanol, or pentylbenzene, none of the combinations showed better AUC than butoxymethylbenzene (0.98). The combination 2-ethyl-1-hexanol + pentylbenzene (V172 + V274) had the best results after analysis of the validation batch, correctly classifying all samples with an AUC value of 0.91 (Table S9), but that was still lower than butoxymethylbenzene alone.

4. Conclusions

The serum volatolomic profile reflects the organism lifestyle and health status, therefore its study for the discovery of disease biomarkers is meaningful. All univariant, unsupervised multivariate, and supervised multivariate analyses carried out in this control-case prospective study are congruent and drive to similar conclusions: the human serum volatolome contains information valuable for the discrimination of PAC patients from the healthy population, and consequently is a valuable tool to discover new PAC biomarkers. Among the great number of VOCs in the volatolomic profile (n = 433), those signals found at the RT and *m/z* values indicated for V197 and V328 are the most promising candidates to be biomarkers of PAC (AUC = 0.98, p-val < 0.001). Nonetheless, there are other VOCs with statistically significant results worthy to be considered: a total of 40 VOCs may be used as PAC biomarkers based on the significant differences for the control and case groups. Their ROC results give AUCs above 0.85 that means that are excellent (19 VOCs with AUC > 0.9) or good (21 VOCs with AUC 0.8–0.9) predictors of PAC. Among these biomarkers, V197, V362, V274, V94, V364, V145, V187, V365, V328, V170 and V171 have been validated using an independent set of samples. Some other VOCs are worthy to be further studied, because although the validation sample set classification was not satisfactory, the differences between the control and patient group

Table 5
Results of the validation set. False positive (FP) and false negative (FN) tests out of 5 controls and 5 PAC patients' samples. The p-value obtained from a *t*-test comparison of the samples for this set is shown. (*) Eleven VOCs in serum were validated as pancreatic cancer biomarkers. Statistically significant p-values are indicated in red.

Biomarker	FP	FN	p-value
V197*	0	0	0.0009
V362*	0	0	0.001
V274*	0	0	0.004
V94*	0	0	0.01
V364*	0	0	0.01
V145*	0	0	0.01
V187*	0	0	0.01
V365*	0	0	0.02
V328*	0	1	0.049
V170*	0	1	0.003
V171*	1	0	0.003
V380	1	1	0.04
V52	1	1	0.09
V338	2	0	0.015
V172	2	0	0.002
V51	2	0	0.03
V131	2	0	0.03
V40	2	0	0.10
V173	3	0	0.004
V68	3	0	0.03
V38	3	0	0.048
V39	3	0	0.05
V41	3	0	0.05
V140	3	0	0.06
V390	0	3	0.14
V386	0	4	0.006
V169	4	0	0.005
V332	4	0	0.006
V174	4	0	0.06
V141	4	0	0.07
V37	4	0	0.07
V142	4	0	0.08
V36	4	0	0.10
V139	4	0	0.5
V28	4	0	0.5
V29	3	1	0.4
V311	2	2	0.11
V34	1	3	0.6
V349	1	3	0.9
V97	5	0	0.003

are still present in the new set of samples so they may be good biomarkers upon establishment of accurate cut-off points. The main objective of this study was to account for the maximum number of volatolomic signals in order to increase the chances of finding meaningful information for the discrimination of PAC patients; therefore, MS full scan detection was used. However, the monitoring of those selected set of VOCs with a SIM method is recommended for further studies since it would enhance the sensitivity and precision; and, in consequence, the robustness of the analysis and would contribute to set the critical values for the diagnostic of PAC.

Some variables that may affect volatolome, i.e., sex, age, smoking, and alcohol intake, were studied. Age or alcohol intake did not have a significant effect in the selected signals. Smokers presented an enhanced signal for one of the selected VOCs: pentylbenzene (V274). Regarding sex, differences between male and female PAC patients were observed (no sex differences were observed in controls). The significance of the control-patient test was not affected by those sex differences.

For the improvement of the accuracy and robustness of the PAC diagnostic method, the identification of the biomarker candidates is crucial. We successfully identified four highly sensitive biomarkers of pancreatic cancer, namely toluene, 2-ethyl-1-hexanol, pentylbenzene, and butoxymethylbenzene. The quantification of those VOCs by SPME-GC–MS with SIM acquisition and using individual ISs may serve to establish the concentration levels in healthy volunteers and patients, and the assessment of those levels to discriminate PAC patients may serve as an initial step for designing a portable device able to detect pancreatic cancer in screening campaigns in populations with a high risk of developing PAC.

Further studies should be directed to elucidate the specificity of the VOC panel proposed for pancreatic cancer detection in an experiment where samples from patients suffering different kinds of cancer and including other pancreatic affections may be tested. Moreover, the inclusion of samples collected at different stages of the disease would show the capability to detect PAC at early stages, which would be of great interest for improving the prognosis of those patients.

CRediT authorship contribution statement

María-Pilar Martínez-Moral: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **María Teresa Tena:** Writing – review & editing, Validation, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Alfonso Martín-Carnicero:** Writing – review & editing, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Alfredo Martínez:** Writing – review & editing,

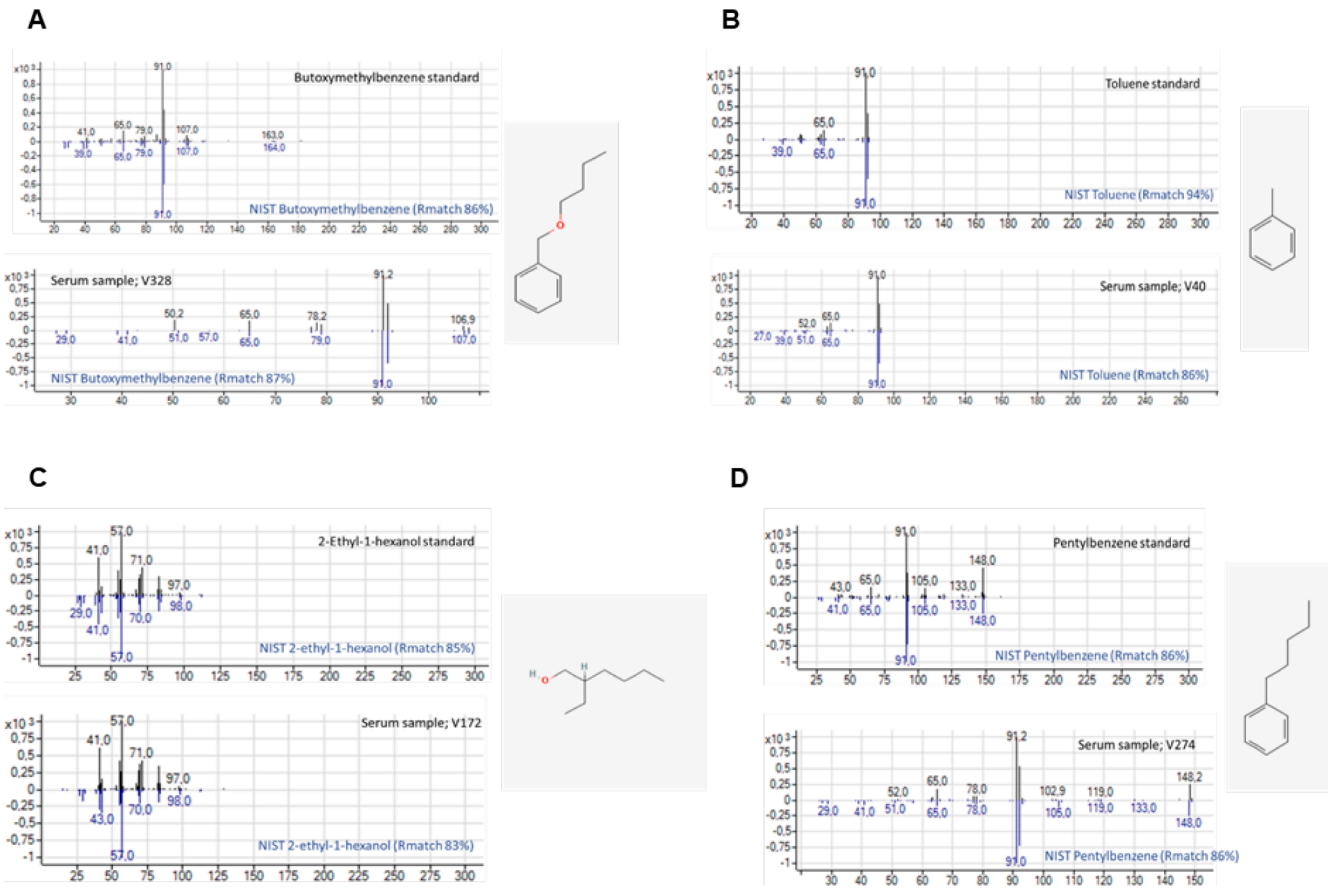

Fig. 5. Comparative of mass spectra obtained for the confirmation of the proposed structures by standards analysis with the coincident retention time VOC in a serum sample against NIST library. A: V328 vs butoxymethylbenzene; B: V40 vs toluene; C: V172 vs 2-ethyl-1-hexanol and D: V274 vs pentylbenzene.

Validation, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: All authors are inventors of a patent application (P202330772) covering the results presented in this study, including the determination of the identified biomarkers for the diagnosis of pancreatic cancer.

Data availability

Data will be made available on request.

Acknowledgements

This work has been funded by “Agencia de Desarrollo Económico de la Rioja” (ADER) through the *hequipa* project (2018-I-IDD-00059) and “Instituto de Salud Carlos III” (ISCiii) through a Sara Borrell research contract CD19/00224.

We would like to express our sincere appreciation to the serum donors that participated in this study, including pancreatic cancer patients and healthy individuals.

We recognize the valuable sample collection work done by the residents at Hospital San Pedro: Ricardo Zafra, Laura Samaniego, Leticia Thomlimson and Sara Fernandez.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cca.2024.117895.

References

[1] Globocan, https://gco.iarc.fr/ (accesed 7 November 2023).
[2] P. Rawla, T. Sunkara, V. Gaduputi, Epidemiology of pancreatic cancer: global trends, etiology and risk factors, World J. Oncol. 10 (2019) 10–27, https://doi.org/10.14740/wjon1166.
[3] A. McGuigan, P. Kelly, R.C. Turkington, C. Jones, H.G. Coleman, R.S. McCain, Pancreatic cancer: a review of clinical diagnosis, epidemiology, treatment and outcomes, World J. Gastroenterol. 24 (2018) 4846–4861, https://doi.org/10.3748/wjg.v24.i43.4846.
[4] American Cancer Society. https://www.cancer.org/ (accessed 7 November 2023).
[5] H. Xing, J. Wang, Y. Wang, M. Tong, H. Hu, C. Huang, D. Li, Diagnostic value of CA19-9 and carcinoembryonic antigen for pancreatic cancer: a meta analysis, Gastroenterol. Res. Pract. (2018) 8704751, https://doi.org/10.1155/2018/8704751.
[6] B. Zhao, B. Zhao, F. Chen, Diagnostic value of serum carbohydrate antigen 19–9 in pancreatic cancer: a systematic review and meta-analysis, Eur. J. Gastroenterol. Hepatol. 34 (2022) 891–904, https://doi.org/10.1097/meg.0000000000002415.
[7] J. Hou, X. Liu, C. Hou, D. Huo, J. Li, A PVDF-based colorimetric sensor array for noninvasive detection of multiple disease-related volatile organic compounds, Anal. Bioanal. Chem. 415 (2023) 6647–6661, https://doi.org/10.1007/s00216-023-04941-y.
[8] A. Vasilescu, B. Hrinczenko, G.M. Swain, S.F. Peteu, Exhaled breath biomarker sensing, Biosens. Bioelectron. 182 (2021) 113193, https://doi.org/10.1016/j.bios.2021.113193.
[9] R. Kramer, A. Sauer-Heilborn, T. Welte, C.A. Guzman, M.G. Höfle, W.R. Abraham, A rapid method for breath analysis in cystic fibrosis patients, Eur. J. Clin. Microbiol. Infect. Dis. 34 (2015) 745–751, https://doi.org/10.1007/s10096-014-2286-5.
[10] M. Barker, M. Hengst, J. Schmid, H.J. Buers, B. Mittermaier, D. Klemp, R. Koppmann, Volatile organic compounds in the exhaled breath of young patients

with cystic fibrosis, Eur. Respir. J. 27 (2006) 929–936, https://doi.org/10.1183/09031936.06.00085105.
[11] E. Janssens, J.P. van Meerbeeck, K. Lamote, Volatile organic compounds in human matrices as lung cancer biomarkers: a systematic review, Crit. Rev. Oncol. Hematol. 153 (2020) 103037, https://doi.org/10.1016/j.critrevonc.2020.103037.
[12] K. Schmidt, I. Podmore, Solid phase microextraction (SPME) method development in analysis of volatile organic compounds (VOCS) as potential biomarkers of cancer, J Mol Biomark Diagn. 6 (2015) 1000253, https://doi.org/10.4172/2155-9929.1000253.
[13] M. Phillips, R.N. Cataneo, B.A. Ditkoff, P. Fisher, J. Greenberg, R. Gunawardena, C. S. Kwon, F. Rahbari-Oskoui, C. Wong, Volatile markers of breast cancer in the breath, Breast J. 9 (2003) 184–191, https://doi.org/10.1046/j.1524-4741.2003.09309.x.
[14] I.J. Kwon, T.Y. Jung, Y. Son, B. Kim, S.M. Kim, J.H. Lee, Detection of volatile sulfur compounds (VSCs) in exhaled breath as a potential diagnostic method for oral squamous cell carcinoma, BMC Oral Health 22 (2022) 268, https://doi.org/10.1186/s12903-022-02301-3.
[15] X. Zou, W. Zhou, Y. Lu, C. Shen, Z. Hu, H. Wang, H. Jiang, Y. Chu, Exhaled gases online measurements for esophageal cancer patients and healthy people by proton transfer reaction mass spectrometry, J. Gastroenterol. Hepatol. 31 (2016) 1837–1843, https://doi.org/10.1111/jgh.13380.
[16] M. Hakim, Y.Y. Broza, O. Barash, N. Peled, M. Phillips, A. Amann, H. Haick, Volatile organic compounds of lung cancer and possible biochemical pathways, Chem. Rev. 112 (2012) 5949–5966, https://doi.org/10.1021/cr300174a.
[17] S. Das, S. Pal, M. Mitra, Significance of exhaled breath test in clinical diagnosis: a special focus on the detection of diabetes mellitus, J. Med. Biol. Eng. 36 (2016) 605–624, https://doi.org/10.1007/s40846-016-0164-6.
[18] F.H. Tyas, J.G. Nikita, D.K. Apriyanto, M.N.A. Mitrayana, The performance of CO_2 laser photoacoustic spectrometer in concentration acetone detection as biomarker for diabetes mellitus type 2, J. Phys. Conf. Ser., IOP Pub. 1011 (2018) 012056, https://doi.org/10.1088/1742-6596/1011/1/012056.
[19] N.M. Zetola, C. Modongo, O. Matsiri, T. Tamuhla, B. Mbongwe, K. Matlhagela, E. Sepako, A. Catini, G. Sirugo, E. Martinelli, R. Paolesse, C. Di Natale, Diagnosis of pulmonary tuberculosis and assessment of treatment response through analyses of volatile compound patterns in exhaled breath samples, J. Infect. 74 (4) (2017) 367–376, https://doi.org/10.1016/j.jinf.2016.12.006.
[20] A.M.I. Saktiawati, D.D. Putera, A. Setyawan, Y. Mahendradhata, T.S. van der Werf, Diagnosis of tuberculosis through breath test: a systematic review, EBioMedicine 46 (2019) 202–214, https://doi.org/10.1016/j.ebiom.2019.07.056.
[21] I. Ahmed, R. Greenwood, B.L. de Costello, N.M. Ratcliffe, C.S. Probert, An investigation of fecal volatile organic metabolites in irritable bowel syndrome, PLoS One 8 (2013), https://doi.org/10.1371/journal.pone.0058204.
[22] S. Sethi, R. Nanda, T. Chakraborty, Clinical application of volatile organic compound analysis for detecting infectious diseases, Clin. Microbiol. Rev. 26 (2013), https://doi.org/10.1128/CMR.00020-13.
[23] M. Hakim, Y.Y. Broza, O. Barash, N. Peled, M. Phillips, A. Amann, H. Haick, Volatile organic compounds of lung cancer and possible biochemical pathways, Chem. Rev. 112 (2012) 5949–5966, https://doi.org/10.1021/cr300174a.
[24] M. Woollam, M. Teli, P. Angarita-Rivera, S. Liu, A.P. Siegel, H. Yokota, M. Agarwal, Detection of volatile organic compounds (VOCs) in urine via gas chromatography-mass spectrometry QTOF to differentiate between localized and metastatic models of breast cancer, Sci. Rep. 9 (2019) 2526, https://doi.org/10.1038/s41598-019-38920-0.
[25] S. Nishiumi, M. Shinohara, A. Ikeda, T. Yoshie, N. Hatano, S. Kakuyama, S. Mizuno, T. Sanuki, H. Kutsumi, E. Fukusaki, T. Azuma, T. Takenawa, M. Yoshida, Serum metabolomics as a novel diagnostic approach for pancreatic cancer, Metabolomics 6 (2010) 518–528, https://doi.org/10.1007/s11306-010-0224-9.
[26] S. Urayama, W. Zou, K. Brooks, V. Tolstikov, Comprehensive mass spectrometry based metabolic profiling of blood plasma reveals potent discriminatory classifiers of pancreatic cancer, Rapid Commun. Mass Spectrom. 24 (2010) 613–620, https://doi.org/10.1002/rcm.4420.
[27] S.R. Markar, B. Brodie, S.T. Chin, A. Romano, D. Spalding, G.B. Hanna, Profile of exhaled-breath volatile organic compounds to diagnose pancreatic cancer, Br. J. Surg. 105 (2018) 1493–1500, https://doi.org/10.1002/bjs.10909.
[28] A. Princivalle, L. Monasta, G. Butturini, C. Bassi, L. Perbellini, Pancreatic ductal adenocarcinoma can be detected by analysis of volatile organic compounds (VOCs) in alveolar air, BMC Cancer 18 (2018) 529, https://doi.org/10.1186/s12885-018-4452-0.
[29] S.I. Nissinen, A. Roine, L. Hokkinen, M. Karjalainen, M. Venäläinen, H. Helminen, R. Niemi, T. Lehtimäki, T. Rantanen, N. Oksala, Detection of pancreatic cancer by urine volatile organic compound analysis, Anticancer Res 39 (2019) 73–79, https://doi.org/10.21873/anticanres.13081.
[30] R.P. Arasaradnam, A. Wicaksono, H. O'Brien, H.M. Kocher, J.A. Covington, T. Crnogorac-Jurcevic, Noninvasive diagnosis of pancreatic cancer through detection of volatile organic compounds in urine, Gastroenterology 154 (2018) 485–487, https://doi.org/10.1053/j.gastro.2017.09.054.
[31] E. Daulton, A.N. Wicaksono, A. Tiele, H.M. Kocher, S. Debernardi, T. Crnogorac-Jurcevic, J.A. Covington, Volatile organic compounds (VOCs) for the non-invasive detection of pancreatic cancer from urine, Talanta 221 (2021) 121604, https://doi.org/10.1016/j.talanta.2020.121604.
[32] Y. Chen, E.M. Li, L.Y. Xu, Guide to metabolomics analysis: a bioinformatics workflow, Metabolites 12 (2022) 357, https://doi.org/10.3390/metabo12040357.
[33] J. Antonelli, B.L. Claggett, M. Henglin, A. Kim, G. Ovsak, N. Kim, K. Deng, K. Rao, O. Tyagi, J.D. Watrous, K.A. Lagerborg, P.V. Hushcha, O.V. Demler, S. Mora, T. J. Niiranen, A.C. Pereira, M. Jain, S. Cheng, Statistical workflow for feature selection in human metabolomics data, Metabolites 9 (2019) 1–15, https://doi.org/10.3390/metabo9070143.
[34] A. Cambiaghi, M. Ferrario, M. Masseroli, Analysis of metabolomic data: tools, current strategies and future challenges for omics data integration, Brief. Bioinform. 18 (2017) 498–510, https://doi.org/10.1093/bib/bbw031.
[35] R.H. El Khouli, K.J. Macura, P.B. Barker, M.R. Habba, M.A. Jacobs, D.A. Bluemke, Relationship of temporal resolution to diagnostic performance for dynamic contrast enhanced MRI of the breast, J. Magn. Reson. Imaging 30 (2009) 999–1004, https://doi.org/10.1002/jmri.21947.
[36] M. Shou, T. Lu, K.W. Krausz, Y. Sai, T. Yang, K.R. Korzekwa, F.J. Gonzalez, H. V. Gelboin, Use of inhibitory monoclonal antibodies to assess the contribution of cytochromes P450 to human drug metabolism, Eur. J. Pharmacol. 14 (2000) 199–209, https://doi.org/10.1016/S0014-2999(00)00079-0.
[37] Human Metabolome Database (HMDB). https://hmdb.ca/ (accesed 7 November 2023).
[38] S. Dragonieri, R. Schot, B.J.A. Mertens, S. Le Cessie, S.A. Gauw, A. Spanevello, O. Resta, N.P. Willard, T.J. Vink, K.F. Rabe, E.H. Bel, P.J. Sterk, An electronic nose in the discrimination of patients with asthma and controls, J. Allergy Clin. Immunol. 120 (2007) 856–862, https://doi.org/10.1016/j.jaci.2007.05.043.
[39] C.E. Garner, S. Smith, B. Lacy Costello, P. White, R. Spencer, C.S.J. Probert, N. M. Ratcliffem, Volatile organic compounds from feces and their potential for diagnosis of gastrointestinal disease, FASEB J. 21 (2007) 1675–1688, https://doi.org/10.1096/fj.06-6927com.
[40] S. Kim, E. Park, S.H. Song, C.W. Lee, J.T. Kwon, E.Y. Park, B. Kim, Toluene concentrations in the blood and risk of thyroid cancer among residents living near national industrial complexes in South Korea: a population-based cohort study, Environ. Int. 146 (2021) 106304, https://doi.org/10.1016/j.envint.2020.106304.
[41] R. Francavilla, D. Ercolini, M. Piccolo, L. Vannini, S. Siragusa, F. De Filippis, I. De Pasquale, R. Di Cagno, M. Di Toma, G. Gozzi, D.I. Serrazanetti, M. De Angelis, M. Gobbetti, Salivary microbiota and metabolome associated with celiac disease, Appl. Environ. Microbiol. 80 (2014) 3416–3425, https://doi.org/10.1128/AEM.00362-14.
[42] S.W. Cho, H.J. Ko, T.H. Park, Identification of a lung cancer biomarker using a cancer cell line and screening of olfactory receptors for biomarker detection, biotechnol, Bioprocess Eng. 26 (2021) 55–62, https://doi.org/10.1007/s12257-020-0132-4.
[43] W. Jelski, E. Kutylowska, M. Laniewska-Dunaj, M. Szmitkowski, Alcohol dehydrogenase (ADH) and aldehyde dehydrogenase (ALDH) as candidates for tumor markers in patients with pancreatic cancer, J. Gastrointestin. Liver Dis. 20 (2011) 255–259. https://pubmed.ncbi.nlm.nih.gov/21961092/.