Environmental Pollution

Blood, urine and semen VOC pattern analysis for assessing health environmental impact in highly polluted areas in Italy

--Manuscript Draft--

Manuscript Number:	
Article Type:	Research Paper
Keywords:	VOCs; GC-MS; highly polluted areas; blood; urine; human semen
Corresponding Author:	Valentina Longo, Ph.D. CNR: Consiglio Nazionale delle Ricerche Lecce, Lecce ITALY
First Author:	Valentina Longo, Ph.D.
Order of Authors:	Valentina Longo, Ph.D.
	Angiola Forleo
	Alessandra Ferramosca
	Tiziana Notari
	Sebastiana Pappalardo
	Pietro Siciliano
	Simonetta Capone
	Luigi Montano
Abstract:	Volatile Organic Compound (VOC) analysis is usually applied in pollution assessment by checking for toxic or harmful volatile compounds volatiles in air, water and soil samples. In this study, exogenous VOCs and their derivatives, metabolized by cells, were valued into specific body fluids. In particular, the VOC profiles of blood, urine and human semen samples collected from young men living in two high pollution areas in Italy, i.e. Land of Fires and Valley of Sacco River, were fingerprinted. The analysis is based on Headspace Solid Phase Micro-extraction (HS-SPME) followed by Gas Chromatography-Mass Spectrometric detection (GC-MS). The volatile composition of the three body fluids showed that some VOCs are in common between blood, urine and human semen samples, whereas others are present only in a body fluid. Some compounds, as well as also some chemical classes show a higher affinity for a specific body fluid. Statistical analysis allowed to discriminate the two contaminated areas and identify those compounds which significantly contribute to the two areas classification. Some of these compounds are toxic and found prevalently in Valley of Sacco River samples, correspondingly to sperm analysis results for young men living in this zona worse than those living in Land of Fires.
Suggested Reviewers:	Rosaria Cozzolino rcozzolino@isa.cnr.it Expert of urinary VOC analysis
	Kamila Schmidt k.schmidt@edu.salford.ac.uk Expert in VOC analysis
	Aleksandra Fucic afucic@imi.hr Expert in genotoxicology and biomonitoring
	Francesco Forastiere f.forastiere@deplazio.it Epidemiology of Regional Health Authority
	Maura Lodovici maura.lodovici@unifi.it

	Expert in air pollution and oxidative stress

Valentina Longo
Institute for microelectronics and microsystems of
National Research Council (IMM-CNR)
Via per Monteroni, Ecotekne campus, Lecce 73100
Email: valentinalongo@le.imm.cnr.it
Tel: +39 0832/422547

Editor-in-Chief
Environmental Pollution-Elsevier

Re:Submission of a new manuscript as Full Research Paper
22th of February, 2021

Dear Editor,
we are submitting to your attention the manuscript entitled:

Blood, urine and semen VOC pattern analysis for assessing health environmental impact in highly polluted areas in Italy

authored by:

Valentina Longo, Angiola Forleo, Alessandra Ferramosca, Tiziana Notari, Sebastiana Pappalardo, Pietro Siciliano, Simonetta Capone, Luigi Montano

which we aim at publishing in as an Environmental Pollution as a "Full Research Paper".

In the research here described, we evaluated the Volatile Organic Compounds (VOCs) in the headspace of blood, urine and human semen samples collected from young men living in two high pollution areas in Italy (Land of Fires and Valley of Sacco River). The aim was to value different body fluid VOC fingerprinting and, after that, to identify potential volatile compounds that can discriminate the two polluted Italian areas.

Each author has contributed significantly to the submitted work.

On behalf of the Authors, we state that
1. the paper is not under consideration elsewhere;
2. none of the paper's contents have been previously published
3. all authors have read and approved the manuscript.

We suggest as scientific reviewer:

- Rosaria Cozzolino -Institute of Food Science, CNR, Avellino, Italy- rcozzolino@isa.cnr.it
- Kamila Schmidt - Biomedical Science Research Centre, School of Environment and Life Sciences, University of Salford, Manchester M5 4WT, UK - k.schmidt@edu.salford.ac.uk
- Aleksandra Fucic- Institute for Medical Research and Occupational Health, Zagreb, Ksaverska- afucic@imi.hr
- Francesco Forastiere – Regional Health Authority, Rome, Italy, f.forastiere@deplazio.it
- Maura Lodovici- Department of Pharmacology and Toxicology, University of Florence, maura.lodovici@unifi.it

We look forward to knowing about the suitability of this manuscript for publication in Environmental Pollution.

Sincerely,

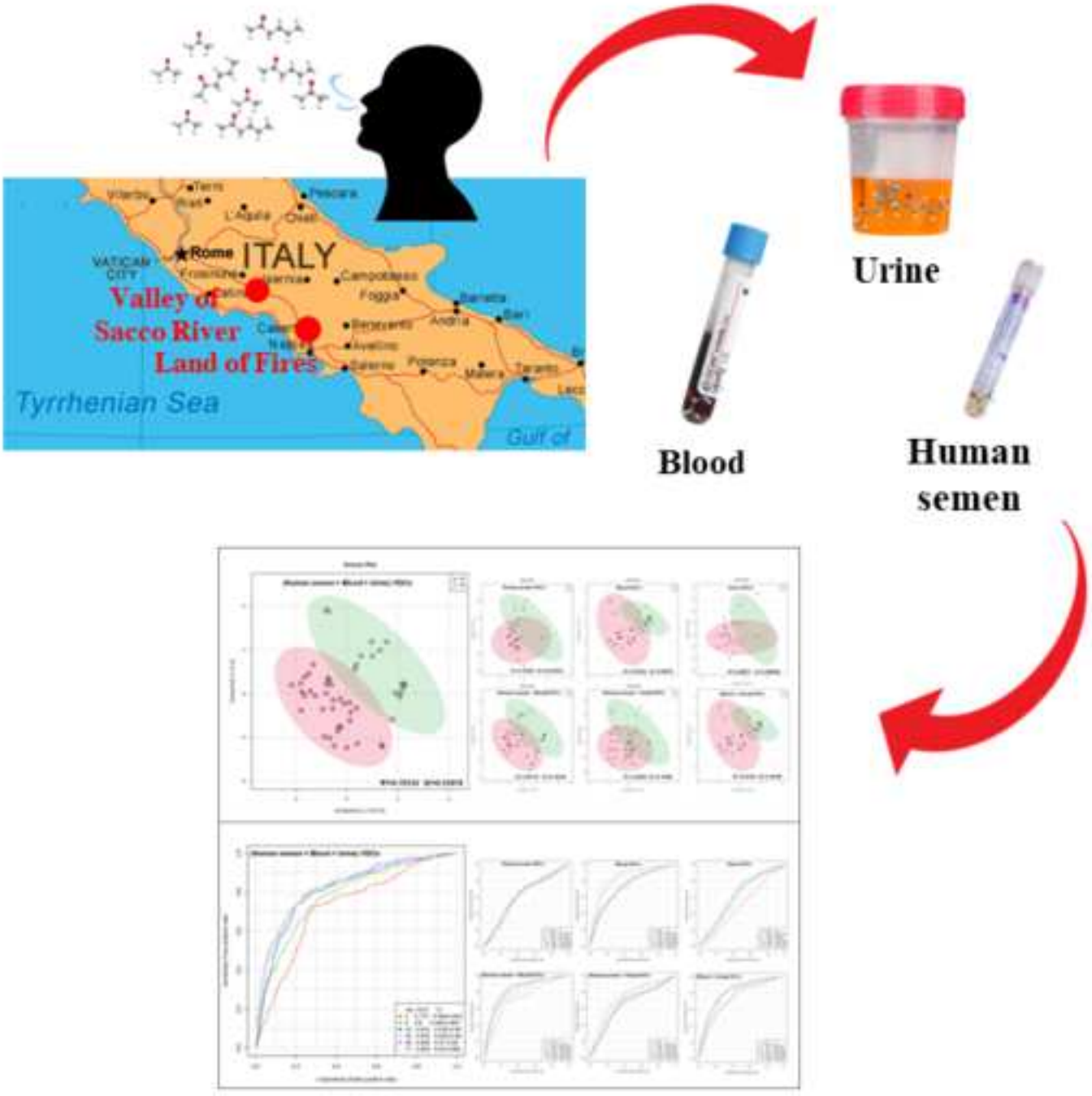
Valentina Longo, D. Biol. PhD

on behalf of the Authors

- Land of Fires and Valley of Sacco River are two highly polluted areas in Italy
- VOC pattern analysis in body fluids is a powerful approach to highlight the exposure within the body of any compounds harmful to health
- Volatile fingerprinting of blood, urine and human semen samples provides a lot of information about men exposition
- HS-SPME-GC/MS is the gold standard method for volatile organic compound analysis

Graphical Abstract

Click here to access/download;Graphical Abstract;Graphical abstract.png



Blood, urine and semen VOC pattern analysis for assessing health environmental impact in highly polluted areas in Italy

Valentina Longo[1]*, Angiola Forleo[1], Alessandra Ferramosca[2], Tiziana Notari[3], Sebastiana Pappalardo[4], Pietro Siciliano[1], Simonetta Capone[1], Luigi Montano[5,6]

[1] National Research Council of Italy, Institute for Microelectronics and Microsystems (CNR-IMM), Lecce, Italy.
[2] Department of Environmental and Biological Sciences and Technologies, University of Salento, Lecce, Italy.
[3] Reproductive Medicine Unit of Check Up Polydiagnostic Center, Via A. De Luca 5, Salerno, Italy.
[4] Reproduction and Fertility Center – Rome, Italy.
[5] Andrology Unit and Service of Lifestyle Medicine in UroAndrology, Local Health Authority (ASL) Salerno, Coordination Unit of the network for Environmental and Reproductive Health (EcoFoodFertility Project), Italy "Oliveto Citra Hospital", Salerno, Italy;
[6]PhD Program in Evolutionary Biology and Ecology,University of Rome Tor Vergata, Rome, Italy.

*corresponding authors
valentina.longo@le.imm.cnr.it

Keywords: VOCs, GC-MS, highly polluted areas, blood, urine, human semen

Abbreviations			
VOCs	Volatile Organic Compounds	CAR/PDMS	Carboxen®/Polydimethylsiloxane
GC-MS	Gas Chromatography-Mass Spectrometric	HMDB	Human Metabolomic Database
HS-SPME	Headspace Solid Phase Micro-extraction	PCA	Principal Component Analysis
LF	Land of Fires	PLSDA	Partial Least Squares Discriminant Analysis
VSR	Valley of Sacco River	ROC	Receiver Operating Characteristic
BMI	Body Mass Index	AUC	Area Under the Curve

Abstract

Volatile Organic Compound (VOC) analysis is usually applied in pollution assessment by checking for toxic or harmful volatile compounds volatiles in air, water and soil samples. In this study, exogenous VOCs and their derivatives, metabolized by cells, were valued into specific body fluids. In particular, the VOC profiles of blood, urine and human semen samples collected from young men living in two high pollution areas in Italy, i.e. *Land of Fires* and *Valley of Sacco River*, were fingerprinted.

The analysis is based on Headspace Solid Phase Micro-extraction (HS-SPME) followed by Gas Chromatography-Mass Spectrometric detection (GC-MS).

The volatile composition of the three body fluids showed that some VOCs are in common between blood, urine and human semen samples, whereas others are present only in a body fluid. Some compounds, as well as also some chemical classes show a higher affinity for a specific body fluid.

Statistical analysis allowed to discriminate the two contaminated areas and identify those compounds which significantly contribute to the two areas classification. Some of these compounds are toxic and found prevalently in Valley of Sacco River samples, correspondingly to sperm analysis results for young men living in this zona worse than those living in Land of Fires.

"CAPSULE": Highest concentration of the detected toxic compounds in human semen of Valley of Sacco River men correspond to worse semen quality.

Introduction

The use of smell in diagnostic dates back to a few centuries before Christ when Hippocrates used the smell associated with faeces for tuberculosis diagnosis. Later, it was understood that the smell emanating from body fluids contains many volatile compounds that can be very different from each other imprinting specific odorous characteristics to body secretions (Sankarganesh et al., 2019).

Over time, more and more substances belonging to the class of VOCs, which stands for Volatile Organic Compounds, have been identified. In scientific research, this heterogeneous class is used principally for two purposes: the first is the study of volatile metabolome in body fluids of sick individuals to develop 'potential biomarkers' for early diagnosis and disease prognosis as in pathological states the composition and concentration of specific VOC can be considerably altered (Di Lena et al., 2016); the second is to evaluate the volatile part of external exposome, that is the totality of volatile substances due to factors of external environment to which people are exposed (air pollution, chemical exposures, workplaces environment, diet, tobacco, drug, etc.) and detect how and to what extent these substance enter into the human body and become an important

component of human volatilome (Adgate et al., 2004). The first study concerns endogenous VOCs that derived by cell metabolism, whereas the second one regards exogenous VOCs. Discrimination between endo- or exogenous origin of VOCs is not easy. In fact, some VOCs can be both endogenous and exogenous, as well as some exogenous volatiles can be metabolized into body, mixing the two different compartments of volatilome.

The most used biological matrices for VOC analysis are blood, urine, breath and faeces, followed by saliva, skin emanations and breast milk (Amann et al., 2014; de Lacy Costello et al., 2014). Principal target is to find a right compromise between the abundance of extracted volatiles and the non-invasiveness of samples collection. Blood is the most informative of the physiological state of body, since it is in continuous contact with the whole organism and it is in equilibrium with organs and tissues, but its withdrawal is invasive and trained staff and special materials are required, so that alternatives to the use of blood are searched. Urine is a readily available biological matrix, whose non-invasive collection can be done directly by the donors with no volume limitation.

For the first time, our research group detailed VOC composition in human semen (Longo et al., 2019a), evaluating the different frequency of occurrence for specific compounds in normozoospermic, asthenozoospermic and oligozoospermic men. But to date, seminal volatiles are not included in any database or compendium and this aspect makes more difficult and slow this kind of analysis.

Both arising from the outside and cellular metabolism, VOCs are conveyed in blood. In fact, the fingerprint of volatile composition of each biospecimens derived by balance with blood: exchange blood-breath through alveolar-capillary microenvironment, blood-urine through renal glomerulus, blood-human semen through blood-testis barrier and so on. These exchanges are not only simple transitions, but it is well known that compounds can be metabolized from epithelial cells of the tissue that they cross (de Lacy Costello et al., 2014). For this reason, some volatiles can be present in other fluids but not in blood.

In some case, the presence in biological fluids of endogenous VOCs can be preoccupant. In fact, long-term exposure to toxic VOCs may increase the risk for certain types of cancers and birth defects (Uddin et al., 2014).

Contaminant compounds can be released into environment by natural and anthropic sources. The most common sources of VOC exposure include tobacco smoke, petroleum products, chlorinated water, and synthetic products such as paints, lubricants, insecticides and pesticides (Lin et al., 2008). All these factors contribute to the exacerbation of air, soil and water pollution.

In Italy, there are some areas which represent real outbreaks from the point of view of pollution. One of the most sadly known is an area widespread on a huge territorial extension between the

provinces of Naples and Caserta (Campania region) with 2,5 million inhabitants, called "Land of Fires" (LF). Here, sources of pollution are multiple: illegal disposal of urban, toxic and industrial wastes, dumping practices, traffic, intensive agriculture (Bosco et al., 2018). This area is officially recognized as a high environmental impact area on the basis of the Campania Region Environmental Protection Agency report (ARPAC, 2008).

Here, the alarming rate of environmental pollution has prompted the birth of EcoFoodFertility project (http://www.ecofoodfertility.org/), to which this work is part. EcoFoodFertility is a biomonitoring multicentre and multidisciplinary research connecting human lifestyle and dietary habits to the environmental consequences of exposure to toxicants in several environmentally-challenged areas of Italy. The principal aims of this complex project is to better understand the environmental impact of toxicants on healthy humans and implement the use human semen as an early and sensitive biomarker of environmental exposures to pollutants as well as of the quality of living environment (Bergamo et al., 2016; Montano et al., 2014; Montano et al., 2018).

But unfortunately, Land of Fire is not the only high pollution area in Italy. Another district, a little further north, with high environmental impact is the Valley of Sacco River (VSR) (Lazio Region). Over the years, it has been heavily polluted principally by industrial wastes deriving from the chemical industrial plants in Colleferro. The area of Colleferro has been polluted by multiple sources and the population has been exposed to industrial chemicals, toxic substances in the workplace, and to the cumulative accumulation of organic pesticides especially through water and food (Fantini et al., 2012).

Aim of this work is to evaluate the different fingerprint of VOCs in various body fluids coming from young health men living in these two high environmental areas: Land of Fire and Valley of Sacco River. Briefly, the study was structured in two steps: (i) untargeted analysis of blood, urine and human semen VOCs of the recruited subjects, with identification and quantification of each compound in the specific biospecimen; (ii) use of most significant compounds to discriminate the geographical origin of the samples to better define which VOCs characterize Land of Fire or Valley of Sacco River men.

Compound characterization was carried out by Headspace Solid-Phase MicroExtraction followed by Gas Chromatography-Mass Spectrometry (HS-SPME/GC-MS), an elective and advantageous analytical method for studying volatile composition in clinical and environmental analysis because it requires simple and fast handling of the sample (Longo et al., 2019b).

1. Methods

2.1.Subjects

For this study, 50 healthy young male volunteers in fertile age from Land of Fires (36) and Valley of Sacco River (14) were enrolled. The study was carried out in accordance with the Code of Ethics of the World Medical Association (World Medical Association, 2013) upon approval of the Ethical Committee of the Local Health Authority Campania Sud-Salerno (protocol number: 43_r.p.s.o. june 30, 2015). All volunteers were above 18 and gave written informed consent. The recruitment, which was occurred within the EcoFoodFertility project, was conducted for six months in 2018 (May 2018-October 2018) at the San Francesco d'Assisi Hospital (Oliveto Citra, Salerno, Italy) and at the Italian Association of Blood Volunteers (AVIS, Frosinone office, Frosinone, Italy). To avoid confounding factors, young men with habit of smoking or alcohol drinking, chronic diseases (diabetes or other systemic diseases), reproductive system malfunctions (varicocele, prostatitis and so on) and high body mass index (BMI >33 Kg/m), were excluded.

A total of 102 samples (34 from human sperm, 44 from whole blood and 24 from urine) were collected from as many subjects as possible.

2.2.Sample collection

Human semen samples were collected upon morning masturbation after 3–4 days of sexual abstinence in sterile containers. Human ejaculate has undergone routine tests to evaluate sperm quality; parameters of semen quality (semen volume, sperm concentration, motility and morphology) were evaluated according to the World Health Organization (WHO) guidelines (World Health Organization, 2010).

Blood samples were collected via venipuncture into sodium citrate tubes and gently mixed.

Each individual provided a sample of morning urine (after overnight fasting) in a 50 ml sterile PVC container.

Next, aliquots of the different sampled biofluids were taken for VOC analysis; in particular, 250 µL of semen ejaculated, 1 ml of whole blood and urine were transferred into 5 mL headspace vials (Shimadzu™, cod. 27319-U Sigma-Aldrich) capped with an assembled screw cap with hole with PTFE/silicone septum and immediately frozen and stored at -80°C.

2.3.HS-SPME/GC-MS measurements

The experimental protocol for the treatment of semen, blood and urine samples, which was followed, standardized the VOC extraction methods from all three biological fluids: the frozen samples were thawed at room temperature and subsequently the vials were immersed in water bath onto a magnetic stirrer hotplate at 60 °C overnight, stirring gently through a home-made system which guarantees the same agitation to all samples.

After this incubation, a Carboxen®/Polydimethylsiloxane (CAR/PDMS) fiber (57318, Supelco) was exposed to each sample headspace for 15 min. Different extraction times have been tested and 15 minutes was found the shortest time at which a maximum extraction occurs.

GC-MS analysis of extracted volatiles was performed using GC (6890N series Agilent Technologies) coupled to MS (5973 series Agilent Technologies) equipped with ZB-624 capillary column (Phenomenex); the injector temperature was set at 250 °C to allow thermally desorption of VOCs. The carrier gas was high purity helium with a flow rate of 1ml min^{-1}.

After several technical tests to set the GC temperature program, the following temperature method was chosen: initial 34 °C held for 2 min; then ramped at 3 °C min^{-1} to 110 °C, after that 5° C min^{-1} to 220°C held for 2 min. Total run time was 51.33 min.

The MS analyses were carried out in full-scan mode with a scan range 30–500 amu at 3.2 scans/s.

2.4.VOC Analysis

Chromatograms were analysed by Enhanced Data Analysis software and the identification of the volatile compounds was achieved by comparing mass spectra with those of the data system library (NIST14, P>60%) (Cozzolino et al., 2017). The identification was confirmed with the injection of external standards corresponding to most recurrent compounds using the same fibre, column and temperature programme. Analysis of empty vial (blank sample) allowed to exclude any contamination by ambient area during sample manipulation.

To quantify the identified VOCs, a semiquantitative method based on the internal standard (I.S.) 1,4-Dichlorobenzene-D4 (EPA-8260C) was followed. Calibration solutions were prepared in methanol and diluted in human semen, blood and urine samples to final I.S. concentrations of: 2, 10, 20, 50, 100, 150 and 200 ppb. Measurements were performed using the same protocol and method applied to all samples. Standard curves were finally constructed on the basis of the 1,4-Dichlorobenzene-D4 peak areas, obtained from Enhanced Data Analysis. For each biological fluid, a specific standard curve was built. This strategy was chosen because the considered biofluids have different density and therefore different partition coefficients. Data were analysed in triplicate.

The Human Metabolomic Database (HMDB version 4.0; (Wishart et al., 2018)) was utilized to identify compounds that were endogenous to the human body, while for VOCs that were not found on HMDB, PubChem Compound (http://www.ncbi.nlm.nih.gov/pccompound) website was used.

2.5.Univariate and multivariate statistical analysis

Descriptive statistics was applied to present the data in a more meaningful way. Univariate methods were used to screen for VOCs that might be useful for multivariate analysis. First, in order to

characterize the VOC profiles, the statistical differences between the considered biological matrices (urine, blood and semen) were explored in terms of the quantified VOC patterns by non-parametric Mann–Whitney U test. Compounds with p-value < 0.05 and VOCs unique of a specific biofluid are used for multivariate analysis which was conducted using MetaboAnalyst 4.0 (Chong et al., 2018).

Next, a total VOC data matrix resulting from the fusion of the matrices from the three biofluids, containing the most significant and recurrent VOCs, was considered to compare the two geographical areas of Land of Fires (LF) and Valley of Sacco River (VSR). As the VOC data matrix was not primarily under a normal distribution, data normalization was carried out using Metaboanalyst 4.0. The data were logarithmic transformed and range scaled to align it a normal Gaussian distribution (Taunk et al.,).

PCA was used for visualization of patterns and outliers (no samples removed as outliers).

Total VOC data matrix was subjected to partial least squares discriminant analysis (PLS-DA) to check for the clustering pattern between LS and VSR groups. R^2 and Q^2 values from the PLS-DA model were employed to evaluate the quality and reliability of mathematical model generated, in which the R^2 value indicates the goodness of fit and the Q^2 value represents the predictability of the model

The total VOC data matrix was also used for ROC analysis.

The area under curve (AUC) - receiver operating characteristic (ROC) curve, also written as AUROC curve, based on the selected VOCs, was used as binary classifier, to evaluate VOCs' ability to distinguishing between the two groups of subjects living in Land of Fires (LF) and Valley of Sacco River (VSR). ROC analysis and t-test were applied to investigate the properties of single variable. We considered VOCs with t-test, p-value less than 0.05 and AUC ROC greater than 0.50.

3. Results

All samples were subjected to routine clinical tests. Blood and urine tests revealed normal parameters in all subjects (data not shown). The results of the semen analysis (seminograms) are reported in Table 1, together to anthropometric data of sample population. Anthropometric data show a very homogeneous class of subjects for age, height, weight and body mass index (BMI). The semen analysis revealed a statistically significant reduction in the progressive motility of spermatozoa and a higher percentage of spermatozoa neck and tail defects in the young men of the Valley of Sacco River (VSR).

3.1.VOC characterization

Analysed human semen, blood and urine samples revealed a high variability in the number of identified VOCs per subject (Fig.1, panel A). Human semen samples vary from a minimum of 2 to a maximum of 29 compounds with a median of 9 (25% percentile: 6.75; 75% percentile: 13.25); blood samples from 3 to 23, with median of 11 (25% percentile: 10.00; 75% percentile: 15.00); urine samples from 3 to 29 VOCs, with a median of 13 (25% percentile: 9.00; 75% percentile: 17.00).

In total, i.e. in all the samples of the considered body fluids, 695 different VOCs have been identified; in particular 241 different VOCs were found in human semen samples, 324 in whole blood and 164 in urine. Some of the identified VOCs are present only in one body fluid, others are common to two or three body fluids. Many of these VOCs were found only in one subject and so they were discarded. In this study, we considered all the compounds that have been found at least in two project participants within the same biological fluid, excluding the VOCs that occur only in one subject. We included in data analysis 60 VOCs of 241 (24.89%) for human semen, 80 VOCs of 324 (24.69%) for blood ones and 42 compounds of 164 (25.60%) for urine ones (Table 1S). In total, we have analysed 135 VOCs.

3.2.Chemical classification

The identified metabolites belong to a variety of chemical classes: aldehydes, ketones, nitrogen compounds, terpenes, acids, alcohols, alkanes and alkenes, amines, amides, immines, aminoacid derivatives, benzene derivatives, heterocyclic aromatic organic compounds, furan and sulphur- and chloro-containing compounds, peroxides, epoxides, ethers and esters. Figure 1- panel B describes the total classes of compounds and the relative numbers of compounds within a class found in each bodily fluid.

In addition to the extreme chemical variety of the detected volatiles (see all body fluids column in fig.1, panel B), there are some groups that have a different expression in a specific body fluid. Alkanes and cycloalkanes are more present in blood than in semen and urine samples (12.50% in blood vs 3.33% in semen vs 2.38% in urine), as well as chloro-containing compounds (8.75% in blood vs 3.33% in semen vs 7.14% in urine), whereas furans are exclusive of blood samples (2.5% in blood). Ketones are more present in urine samples (21.43% in urine vs about 3% in blood and human semen), as well as sulphur-containing compounds (7.14% in urine vs only 1% in semen). Instead, the chemical classes more abundant in human semen are benzene derivatives (25% in semen vs about 21% in the other fluids), aldehydes (11.67% in semen vs 8.75% in blood and 9.52%

in urine), ethers (5% in semen vs 2.5% in blood and urine) and terpenes which are present in 10% of semen samples vs 6.25% in blood whereas they are assent in urines.

3.3.VOC distribution in human semen, blood and urine samples

After chemical classification, we have evaluated the membership of the total 135 VOCs to the sets of the three biological fluids: human semen (HS), whole blood (B) and urine (U). The results are graphed in Venn diagram represented in Figure 2.

12 compounds are common to all the three biofluids (HS ∩ B ∩ U), 11 to human semen and blood (HS ∩ B), 4 to human semen and urine (HS ∩ U) and 8 to urine and blood (B ∩ U). Instead, some volatiles are exclusive of human semen (33) or whole blood (49) or urine (18). In total, we have analysed 135 VOCs. For these compounds we carried out the quantitative analysis.

The 12 compounds present in all biofluids are: 2-Anthracenamine, 2-Ethyl-oxetane, 3-Aminopyrrolidine, 4-(4-Chlorophenyl)-2,6-diphenylpyridine, 6-Chloro-2,3-dimethyl-4-phenylquinoline, Acetic acid, sodium salt, Acetone, 3-Methylbutanal, Hexanal, Hexane, Oxime-, methoxy-phenyl-, 2-(Ethenyloxy)-11H-propane.

The 11 VOCs in common between human semen and blood are: (3-Methoxy-phenyl)-(6-methyl-4-phenyl-quinazolin-2-yl)-amine, 1-(6-Methyl-benzothiazol-2-yl)-3-(4-methyl-benzoyl)-thiourea, 3-(3-Carboxy-4-hydroxyphenyl)-D-alanine, 3-Carene, Auramine, 2-Methylbutane, Corydine, D-Limonene, Heptanal, 1-Methyl-4-[4,5-dihydroxyphenyl]- hexahydropyridine, Pentanal.

The set of VOCs shared between whole blood and urine (8) is represented by 5,10-Dihydro-5-[3-(methylamino)propyl]-11H-dibenzo[b,e][1,4]diazepin-11-one, 2-Ethylacridine, 2-Heptanone, 1,5-Dihydro-1-(4-methoxyphenyl)-5,5-diphenyl-2H-pyrrol-2-one, 2-Pentanone, 9-Methyl-acridine, 2-Nitro-diaminomethylidenhydrazone-benzaldehyde, N-acetyl-3,4,5-trimethoxy-phenylpropylamine.

At last, the smallest subset (4 VOCs) composed by volatiles common at human semen and urine includes Butanal, 3,6-Dimethoxy-9-(2-phenylethynyl)-fluoren-9-ol, 1,3-Dimethyl-8-[2-nitrophenethenyl]-purin-2,6-dione, Pyrrole.

Figure 3 reports the frequency of specific VOCs identified in the three biological fluids from the subjects participating in the study. In human semen and blood, the most present compound is acetone, followed by 3-methylbutanal and hexane in human semen and by hexanal and hexane in blood, whereas in urine the first abundant VOC is 4-heptanone, followed by acetone and 2-pentanone.

3.4.Quantitative analysis

To reduce the amount of data and try to give a statistical significance from biological point of view, we focused on VOCs which have been detected in more than 10% of subjects within one of the three biofluids. Thus, inclusion criteria for the analysis were the following: presence in greater than or equal to 4 of 34 human semen samples or to 5 of 44 blood samples or 3 of 24 urine samples.

By such a way from the total 135 VOCs we filtered 42 VOCs; fig. 4A shows the membership of the 42 VOCs to the three set of biofluids. Each of them can be common to all three fluids (11 compounds) (subset A: HS $\cap$ B $\cap$ U), human semen and blood (6 compounds) (subset B: HS $\cap$ B), human semen and urine (3 compounds) (subset C: HS $\cap$ U), blood and urine (4 compounds) (subset D: B $\cap$ U) or exclusive of human semen (2 compounds) (subset E: HS – B - U)), blood (8 compounds) (subset F: B – HS - U) and urine samples (8 compounds) (subset G: U - HS - B).

Due to different distribution index, multiple standard curves were built using the of 1,4-Dichlorobenzene-D4 peak areas (I.S.) in the three different biological matrices for VOC quantification. The related results indicated good linearity in the range 2–200 ppb (R_{Blood} = 0.99844; R_{Semen}=0.99648; R_{Urine}=0.99620). Under our experimental conditions the limit of detection (LOD) for our experimental system was shown to be equal to 2 ppb.

Quantitative analysis of selected VOCs was carried out. Only most statistically significant compounds are reported in panel B of Figure 4.

Human semen is characterized by an elevate concentration of 3-Methylbutanal compared to blood and urine, whereas acetone and hexanal are in higher concentration in blood than in urine and human semen, as well as hexane is more abundant human semen and blood compared to urine samples; 2-pentanone, which is present only in blood and urine, has higher concentrations in urine samples.

3.5. Univariate and multivariate statistical analysis

The concentration of the 42 selected volatiles were used to perform multivariate statistical analysis.

PLS-DA and Receiver operating characteristic (ROC) multivariate analysis are conducted both on the submatrices of VOCs from single biological fluid (VOC variables: 22 for HS; 29 for B; 26 for U) and on the union of these (VOC variables: 44 for HS $\cup$ B; 38 HS $\cup$ U; 46 for B $\cup$ U and 77 for HS $\cup$ B $\cup$ U).

Among the different PLS-DA models, the model that gave the best results for discriminating the two groups of subjects living in Land of Fires and Valley of Sacco River were obtained by considering the concentration of selected VOCs of human semen, blood and urine VOCs together, i.e. 77 VOCs variables for HS $\cup$ B $\cup$ U, with values of goodness of fit R^2 and predictability Q^2 of

0.76 and 0.53 respectively. In Fig. 5A we report the score scatter plots of the different PLS-DA discriminant models with the related R^2 and Q^2 values.

Also for ROC analysis the best results were obtained with the matrix containing all 77 VOC variables resulting from the combination of human semen, blood and urine submatrices (Fig. 5B), followed by the combination blood and human semen VOCs.

Using ROC univariate analysis, we obtained a set of 19 putative markers, which seem to be more significant in the distinction of two lands of high environmental impact (Table 2). 7 of these VOCs derived by blood, 8 by sperm samples and only 4 by urine.

Seven compounds have a concentration higher in subject living in Land of Fires (in blood samples: Pentane, Octane, 5,10-Dihydro-5-3-methylaminopropyl-,11H-dibenzob,e1,4diazepin-11-one, 2-Nitro-diaminomethylidenhydrazone-benzaldehyde, Acetic acid; in urine samples: Acetic acid and 2-ethenyloxy-propane) and twelve in samples derived from Valley of Sacco River (in blood samples: Cyclohexane, Methylcyclopentane; in human semen samples: 1-6-Methyl-benzothiazol-2-yl-3-4-methyl-benzoyl-thiourea, 2-Methylbutane, Auramine, 3,6-Dimethoxy-9-2-phenylethynyl-fluoren-9-ol, Pyrrole, Acetic acid, D-Limonene and 3-Aminopyrrolidine; in urine samples: 3-Aminopyrrolidine, 2-Nitro-diaminomethylidenhydrazone-benzaldehyde).

4. DISCUSSION

VOC analysis of 102 human biosamples in total, including blood, human semen and urine samples derived from 50 young men living in two Italian areas at high environmental risk (Land of Fires and Valley of Sacco River) brought out both a large inter-biological fluid and inter-subject variability. The number of detected compounds change from a minimum of 2 VOCs per samples to a maximum of 29 volatiles per samples.

By comparing transversely different samples derived from the same subject, it is evident a difference in number and kind of compounds from a biological fluid to another. This aspect confirms the idea that the distribution of VOCs strongly depends from the biological matrix.

Since in this paper we explored the relationship between statistical datasets from sample classes and not the specific fingerprint of the single subject, VOCs which are present in only a subject were excluded from the analysis with a consequent drastic decrease of compounds until about 25% of the total volatiles.

Heterogeneity of detected volatiles translates in different chemical class pattern in the three biological fluids. The detection of a specific VOC in a fluid and its absence in another is not casual; at the base of this event there are biochemical reasons. For example, if a compound is detected in blood but not in urine, it is plausible that it has been chemically transformed by kidneys or bladder

(de Lacy Costello et al., 2014). Several data of our study confirm the results reported by de Lacy Costello et al. in ref. (de Lacy Costello et al., 2014) which represents the first compendium of all VOCs emanating from the human body in health conditions. In this paper, the authors analyse different biological biosamples: faeces, urine, breath, skin, milk, blood and saliva.

In the present study, for the first time, we analysed human semen, as well as blood and urine. In addition to evaluating the change of VOCs in the passage through the kidneys and bladder, we studied the VOC composition of human semen, which is influenced by seminal vesicles and prostate. It is well-known that several metabolites and drug can pass through blood–testis barrier and transported into semen, altering semen quality within the same subject over time (Sikka and Ayaz, 2018).

Some differences are very evident as the predominance of ketones in urine samples over blood and semen samples. This probably at least partially reflects the bacterial action in the gut, maybe by decarboxylation from the corresponding oxo-acids (Mills et al., 2001).

Sulphur-containing compounds seem to pass all in urine unlike volatiles containing chloro, that are in balance between urine and blood. Urine is the major excretory pathway for sulphur (Magee et al., 2004) that derived principally from protein intake (Whiting and Draper, 1980; 1981), but also from sulphiting agents (E220–228) used as food additives with antimicrobial and antioxidants action (Wedzicha et al., 1984; Saltmarsh, 2000).

The VOCs which mainly cross the blood-testis barrier or those derived from biotransformation in this anatomical compartment (more studies are needed to understand this aspect) are aldehydes, ethers, benzene derivatives and terpenes. The compound level in the three fluids are really quite similar, except for terpenes which seem to be balanced between blood and human semen, excluding urine.

Terpenes are naturally occurring hydrocarbons produced by a wide variety of plants and animals (Brahmkshatriya and Brahmkshatriya, 2013) and have antimicrobial, antifungal, antiviral, anti-inflammatory, antioxidants, antiparasitic actions.

The very dangerous benzene derivatives which probably arise from air pollution (de Lacy Costello et al., 2014) are well distributed in three biological fluids but are more presence, in terms of variety, in semen. This could be a reproductive problem since sperm cells (and sperm DNA) are in close contact with these compounds, some of which also have mutagenic activity (Jung et al., 1981).

Looking at the most detected compounds (more than 10% of samples), it can be observed that there are only 11 compounds ubiquitous to all biological fluids. Nevertheless, some of these are distributed differently in three fluids (Fig.4B).

Acetone is produced from fatty acid breakdown through action of acetoacetate-decarboxylase in the final step of the ketone-body pathway and it represents a secondary source of energy. Despite its ubiquitous presence, acetone is present at higher concentration in blood samples. This is due to the fact that acetone is produced in the mitochondrial matrix of hepatocytes and subsequently it's introduced into the bloodstream. Since it cannot be converted back to acetyl-CoA, it is excreted in the urine or exhaled (Garibaldi et al., 2010). Therefore, all the acetone that is entered in blood, could be divided in urine, breath (Capone et al., 2018) and, also, in human semen (Longo et al., 2019a).

The origins of hexanal may be both dietary (i.e. in carrots) both metabolic as major breakdown product of linoleic acid oxidation. Furthermore, elevated levels of aldehydes are considered the biomarker for enhanced oxidative stress, so that hexanal has been proposed as a measure to diagnose cancer status (Li et al., 2005). Despite the lower levels of this aldehyde in the sperm compared to the blood, these data are worrying because it's known that the hexanal, even if it doesn't make the spermatozoa completely immotile, decreases their motility considerably (Chow et al., 1980).

The very high levels in human semen of a branched-aldehyde, 3-methylbutanal, accompanied with the exclusive presence of 2-methylbutanal in this fluid, arouse attention. In fact, it's very peculiar their attitude to prefer sperm. The possible reason could reside in the metabolic nature of the two aldehydes. In fact, both aldehydes are used as flavouring agent, but can be produced also by metabolism of isoleucine and leucine. The branched-chain aminoacids aminotransfase (BCAT), which is responsible for the conversion of these aminoacids was identified in Sertoli cells, suggesting that these cells are able to metabolize aminoacids (Kaiser et al., 2005). In testes as in ovaries and brain, but not in liver, BCAT is present as cytosolic isoenzyme, while a mitochondrial isoform of BCAT is present in sperm cells (Montamat et al., 1978).

The abundance of hexane in blood compared to the quantity in urine is now fully known. In fact, hexane may penetrate the body by inhalation or absorption through the skin and it is distributed throughout the body and metabolized with the end production of 2,5 hexanedione which is the main metabolite of hexane and is often detected in urine (Mayan et al., 2002).

2-Pentanone is not detected in human semen, but only in blood and at highest levels in urine, such as the other ketones (de Lacy Costello et al., 2014).

Once defined the predominant VOCs and how they are distributed in body compartments, we evaluated the capacity of VOC patterns to discriminate the subject living in the two highly polluted areas, Land of Fires and Valle del Sacco.

Multivariate analysis using partial least-square discriminant analysis (PLS-DA) and ROC curve show that the best classification and prediction are obtained using together blood, semen and urine compounds, followed by combination of blood and human semen VOCs. Using one by one the volatiles of a single fluid, the best results are in order those obtained for blood, human semen and, at last, urine compounds. Despite the goodness of availability of blood samples, it must be underlined that the blood withdrawal is an invasive procedure operator-dependent, in comparison with the collection of human semen obtained by masturbation and urine by urination.

The identification of crucial compounds for the distinction of the two areas highlighted 8 compounds from human semen, 7 from blood and 4 from urine samples.

An important evidence regards the distribution of these selected volatiles. In fact, all the semen VOCs identified as possible biomarkers are more expressed in subjects living in Valley of Sacco River whereas blood markers are more concentrated principally in Land of Fires young men (six of eight).

Auramine is a very dangerous substance used for dyeing of leather, jute, tanned cotton, and paints, and as dye components in inking ribbons, ballpoint pastes, oils and waxes, and carbon paper. The most important areas of application are in dyeing paper and in flexographic printing (Who Health Organization international agency for research on cancer, 2010). It has also been used as antiseptic and fungicide (Pohanish, 2014). Several toxic manifestations by auramine are identified, such as mutation in *Salmonella typhimurium*, as well as generated DNA strand breaks in primary cultures of rat hepatocytes and human cell line HuF22. It also induced DNA fragmentation in the liver, in the kidney and urinary bladder of rats (Kovacic and Somanathan, 2014).

The main source is the exposure by dermal route, but its presence in biological fluids of young men is the reflection of very polluted areas. Because of its gravity, this data requires further studies.

The other identified biomarker is 2-methylbutane, also known as isopentane. It is an important commercial chemical: in fact, it is used as a solvent and to make other chemicals and polystyrene (Pohanish, 2014), it is an ingredient in many household products such as car care, home maintenance, and shaving creams. It is present also in ink, toner and colorant products.

Exposure to 2-methylbutane can occur by inhalation, ingestion, and eye or skin contact in the general population, as well as in workers with specific occupations (Galvin and Marashi, 1999). No toxic effects were reported from brief exposures to low-to-moderate air levels of 2-methylbutane and adverse reproductive effects have not been observed in experimental animals (Yu et al., 2011).

D-Limonene is a monocyclic monoterpene which is a major constituent of citrus oils such as those found in several fruits, including lemon of which it has the odour and to which it owes the name. It has been widely used as a flavor/fragrance additive in perfumes, soaps, pharmaceuticals, and foods

(Sun, 2007). Limonene can also be used as an active or inert ingredient in pesticides, solvents, degreasers, and cleaning agents (U.S. Environmental Protection Agency, 2004).

This compound is considered to be a natural substance possessing low toxicity and, in the past, it was proposed like was proposed as anticancer drug thanks to its chemopreventive and chemotherapeutic activity against many rodent solid tumour types (Crowell and Gould, 1994).

As reported in our previous paper (Longo et al., 2019a), pyrrole is one of the most concentrated VOC in human semen. In this study, it is more present in subjects living in Valley of Sacco River compared to those living in Land of Fires. Pyrrole is a flavouring ingredient with potential anti-inflammatory and anti-microbial proprieties (Raimondi et al., 2006). Furthermore, it is the precursor of a large class of compounds called pyrrole derivatives. Also 3-aminopyrrolidine is the basis for a large set of substances (3-aminopyrrolidine derivatives) at high pharmaceutical interest.

Acetic acid results to be a biomarker in the discrimination of the two Italian areas both in blood and in human semen and in urine samples, but while in blood and urine is more abundant in Land of Fires subjects and human semen is higher concentrated in Valley of Sacco River ones. The sources of this metabolite are really numerous (production of plastic, photographic film, wood glue, synthetic fibres and fabrics, cleaning agent, acidity regulator in food industry). Moreover it can be produced also by several bacteria of urinary tract (*E.Coli, P.Aeruginosa, K.Pneumonia, Enterobacter, Acinetobacter, P.Mirabilis, C.Frundii, E.Faecalis, Streptococcus group B, S.Saprophyticus*) (Gupta et al., 2012). For this reason, it is not possible to identify the reason of the higher amount of acetic acid in human semen of Valley of Sacco River men.

It is important to note that these compounds show statistically differences in human semen but not in blood and urine (Table 2S).

Focusing on Land of Fire samples among the most expressed VOCs there are two short/medium-chain alkanes: pentane and octane. These volatiles derived from peroxidation of unsaturated fatty acid. In particular, pentane is the endproduct of the reaction of ROS with various biological molecules as lipid, DNA and protein and so it is an indicator of damage to these molecules. Historically pentane was used as a marker of lipid peroxidation based on the assumption that it is produced but not metabolized. The compound can be detected in blood and exhaled breath from patients with different pathologies: asthma, COPD, acute myocardial infarction, Crohn's disease and ulcerative colitis and several kinds of cancer (Calenic et al., 2015).

For all these reasons, pentane is considered a biomarker VOC of oxidative stress.

It is well known that the mechanism of air pollution-induced health effects involves an inflammation-related cascade and oxidation stress in several body tissue. Inflammation is initially a protective mechanism which removes the injurious stimuli and produces reactive oxygen species

(ROS) but a subsequent unbalance between ROS formation and individual antioxidant activity caused oxidation stress and damage of lipids, proteins, and macromolecules such as DNA and RNA (Lodovici and Bigagli, 2011).

Acetic acid which is more concentrated in Valley of Sacco River human semen samples, in this area seem to concentrate more in blood and urine.

It's crucial to observe that higher concentration of the detected toxic compounds in human semen of Valley of Sacco River men compared to Land of Fires correspond to worse semen quality as results from seminograms data of both groups. In fact, both the reduction of progressive motility (even below the threshold to be defined asthenozoospermic) and the increased percentage of abnormalities borne by the neck and tail of Valley of Sacco River semen samples are worrying signs that suggest a decisive influence of exposure to highly contaminated living environments. This is particularly impressive considering that the boys of the sample population in Valley of Sacco River are very young boys in the prime of their fertile life. However, more studies are necessaries to investigate possible correlations between VOC patterns in semen and semen quality parameters in relation with environmental factors. Moreover, future studies should evaluate *in-vitro* the impact of these volatile taken individually on sperm parameters to understand if their alteration is the result of the single compound or a cumulative effect.

In general the results from all the considered body fluids allow to conclude that the VOC pattern analysis in body fluids is a powerful approach to highlight the exposure within the body of any compounds harmful to health. The studies on sample population living in highly contaminated areas are the preferential studies to explore the hidden, partial but very relevant, impact of ambient pollution on health; the untargeted VOC analysis can be used as a first step for a next targeted analysis if specific VOCs with particular relevance emerge. The volatilomic investigation of body fluids is hence undoubtedly a branch of the omic sciences that will increasingly affirm its fundamental role both in environmental exposure studies and in the medical diagnosis of pathologies at an early stage through the detection of volatile metabolites.

Author Contributions

LV, FA, CS and ML designed out the experiments; LV, FA performed gas chromatographic measurements; ML, NT and PS organized the collection of the samples in the two areas and carried out the physiological analysis of blood, urine and semen samples; LV performed statistical analysis; LV took the lead in writing the manuscript, supervised by CS; LV prepared all figures and tables. ML conceived the study and design of the "EcoFoodFertility" research project. All authors

discussed the results and commented on the manuscript. All authors provided critical feedback and helped shape the research, the analysis and the manuscript.

Competing interests

The author(s) declare no competing interests.

References

Adgate, J.L., Eberly, L.E., Stroebel, C., Pellizzari, E.D., Sexton, K., 2004. Personal, indoor, and outdoor VOC exposures in a probability sample of children. J Expo Anal Environ Epidemiol. 14, S4-S13. doi:10.1038/sj.jea.7500353.

Amann, A., de Lacy Costello, B., Miekisch, W., Schubert, J., Buszewski, B., Pleil, J., Ratcliffe, N., Risby, T., 2014. The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva. J Breath Res. 8(3),034001. doi:10.1088/1752-7155/8/3/034001.

ARPAC, Agenzia Regionale Protezione Ambientale Campania, 2008. Annuario dati ambientali in Campania 2006. ARPAC, Naples.
http://www.arpacampania.it/documents/30626/51722/Siti+Contaminati.pdf

Bergamo, P., Volpe, M.G., Lorenzetti, S., Mantovani, A., Notari, T., Cocca, E., Cerullo, S., Di Stasio, M., Cerino, P., Montano, L., 2016. Human semen as an early, sensitive biomarker of highly polluted living environment in healthy men: A pilot biomonitoring study on trace elements in blood and semen and their relationship with sperm quality and RedOx status. Reprod Toxicol. 66,1-9. doi: 10.1016/j.reprotox.2016.07.018.

Bosco, L., Notari, T., Ruvolo, G., Roccheri, M.C., Martino, C., Chiappetta, R., Carone, D., Lo Bosco, G., Carrillo, L., Raimondo, S., Guglielmino, A., Montano, L., 2018. Sperm DNA fragmentation: An early and reliable marker of air pollution. Environ Toxicol Pharmacol. 58, 243-249. doi: 10.1016/j.etap.2018.02.001.

Brahmkshatriya, P.P., Brahmkshatriya, P.S., 2013. Terpenes: Chemistry, Biological Role, and Therapeutic Applications. (ed. Ramawat, K., Mérillon, J.M.) Natural Products. Springer, Berlin, Heidelberg). doi:10.1007/978-3-642-22144-6_120.

Calenic, B., Miricescu, D., Greabu, M., Kuznetsov, A.V., Troppmair, J., Ruzsanyi, V., Amann, A., 2015. Oxidative stress and volatile organic compounds: interplay in pulmonary, cardio-vascular, digestive tract systems and cancer. Open Chem. 13, 1020–1030. doi:10.1515/chem-2015-0105.

Capone, S., Tufariello, M., Forleo, A., Longo, V., Giampetruzzi, L., Radogna, A.V., Casino, F., Siciliano, P., 2018. Chromatographic analysis of VOC patterns in exhaled breath fromsmokers and nonsmokers. Biomed Chromatogr. 32. doi:10.1002/bmc.4132.

570 Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G., Wishart, D.S., Xia, J., 2018.
571 MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. Nucleic Acids
572 Res. 46, 486-494. doi:10.1093/nar/gky310.

573 Chow, P.Y., Holland, M.K., Suter, D.A., White, I.G., 1980. Evaluation of ten potential organic
574 spermicides. Int J Fertil. 25, 281-286.

575 Cozzolino, R., De Giulio, B., Marena, P., Martignetti, A., Lauria, F., Russo, P., Siani, A., 2017.
576 Urinary volatile organic compounds in overweight compared to normal-weight children: results
577 from the Italian I.Family cohort. Sci Rep.7, 15636. doi:10.1038/s41598-017-15957-7.

578 Crowell, P.L., Gould, M.N., 1994. Chemoprevention and therapy of cancer by d-limonene. Crit Rev
579 Oncog.; 5, 1-22. doi:10.1615/critrevoncog.v5.i1.10.

580 de Lacy Costello, B., Amann, A., Al-Kateb, H., Flynn, C., Filipiak, W., Khalid, T., Osborne, D.,
581 Ratcliffe, N.M., 2014. A review of the volatiles from the healthy human body. J Breath Res. 8(1),
582 014001. doi: 10.1088/1752-7155/8/1/014001.

583 Di Lena, M., Porcelli, F., Altomare, D.F., 2016. Volatile organic compounds as new biomarkers for
584 colorectal cancer: a review. Colorectal Dis. 18, 654–63. doi: 10.1111/codi.13271.

585 Fantini, F., Porta, D., Fano, V., De Felip, E., Senofonte, O., Abballe, A., D'Ilio, S., Ingelido, A.M.,
586 Mataloni, F., Narduzzi, S., Blasetti, F., Forastiere, F., 2012. Epidemiologic studies on the health
587 status of the population living in the Sacco River Valley. Epidemiol Prev. 36, 44-52.

588 Galvin, J.B., Marashi, F., 1999. 2-Methylbutane (isopentane). J. Toxicol. Environ. Health. 58, 23–
589 31.

590 Garibaldi, B.T., Diallo, A.O., Bell, S.P., Files, D.C., Troy, P.J., 2010. Patient Encounters: The
591 Internal Medicine Work-Up. Philadelphia, Pa, USA: Lippincott Williams & Wilkins.

592 Gupta, A., Dwivedi, M., Mahdi, A.A., Khetrapal, C.L., Bhandari, M., 2012. Broad identification of
593 bacterial type in urinary tract infection using (1)h NMR spectroscopy. J Proteome Res. 11, 1844-
594 1854. doi:10.1021/pr2010692.

595 Jung, R., Beermann, D., Glatt, H.R., Oesch, F., 1981. Mutagenicity of structurally related oxiranes:
596 derivatives of benzene and its hydrogenated congeners. Mutat Res. 81, 11-19. doi: 10.1016/0027-
597 5107(81)90082-8.

598 Kaiser, G.R., Monteiro, S.C., Gelain, D.P., Souza, L.F., Perry, M.L., Bernard, E.A. Metabolism of
599 amino acids by cultured rat Sertoli cells. Metabolism. 54, 515-521 (2005).
600 doi:10.1016/j.metabol.2004.11.005.

601 Kovacic, P., Somanathan, R., 2014. Toxicity of imine-iminium dyes and pigments: electrontransfer,
602 radicals, oxidative stress and other physiological effects. J Appl Toxicol. 34, 825-834.
603 doi:10.1002/jat.3005.
604 Li, N., Deng, C., Yin, X., Yao, N., Shen, X., Zhang, X., 2005. Gas chromatography-mass
605 spectrometric analysis of hexanal and heptanal in human blood by headspace single-drop
606 microextraction with droplet derivatization. Anal Biochem. 342, 318-26.
607 doi:10.1016/j.ab.2005.04.024.

608 Lin, Y.S., Egeghy, P.P., Rappaport, S.M., 2008. Relationships between levels of volatile organic
609 compounds in air and blood from the general population. J Exp Sci Environ Epidemiol. 18, 421–
610 429. doi:10.1038/sj.jes.7500635.

611 Lodovici, M., Bigagli, E., 2011. Oxidative stress and air pollution exposure. J Toxicol. 2011,
612 487074. doi:10.1155/2011/487074.

613 Longo, V., Forleo, A., Pinto Provenzano, S., Coppola, L., Zara, V., Ferramosca, A., Siciliano, P.,
614 Capone, S., 2019a. HS-SPME-GC-MS metabolomics approach for sperm quality evaluation by
615 semen volatile organic compounds (VOCs) analysis. Biomed. Phys. Eng. Express. 5, 015006.
616 doi:10.1088/2057-1976/aaeb07

617 Longo, V., Forleo, A., Capone, S., Scoditti, E., Carluccio, M.A., Siciliano, P., Massaro, M.. 2019b.
618 In vitro profiling of endothelial volatile organic compounds under resting and pro-inflammatory
619 conditions. Metabolomics. 15,132. doi:10.1007/s11306-019-1602-6.

620 Magee, E.A., Curno, R., Edmond, L.M., Cummings, J.H., 2004. Contribution of dietary protein and
621 inorganic sulfur to urinary sulfate: toward a biomarker of inorganic sulfur intake. Am J Clin Nutr
622 80, 137–142. doi:10.1093/ajcn/80.1.137.

623 Mayan, O., Teixeira, J.P., Alves, S., Azevedo, C., 2002. Urinary 2,5 hexanedione as a biomarker of
624 n-hexane exposure. Biomarkers. 7, 299-305. doi:10.1080/13547500210136796.

625 Mills, G.A., Walker, V., 2001. Headspace solid-phase microextraction profiling of volatile
626 compounds in urine: application to metabolic investigations. J. Chromatogr. B Biomed. Sci. Appl.
627 753, 259–768. doi:10.1016/s0378-4347(00)00554-5.

628 Montamat, E.E., Moreno, J., Blanco, A., 1978. Branched-chain amino acid aminotransferase in
629 mouse testicular tissue. J Reprod Fertil. 53, 117-123. doi:10.1530/jrf.0.0530117

630 Montano, L., Bergamo, P., Andreassi, M.G., Lorenzetti, S., 2014. EcoFoodFertility –
631 Environmental and food impact assessment on male reproductive function. Andrology. 2, 69.
632 doi:10.5772/intechopen.73231.

633 Montano, L., Bergamo, P., Andreassi, M.G., Lorenzetti, S., 2018. The role of human semen as an
634 early and reliable tool of environmental impact assessment on human health. Full Chapter in Final
635 Book Title & ISBN: Spermatozoa - Facts and Perspectives, "978-1-78923-171-7. InTechOpen.
636 doi:10.5772/intechopen.73231.

637 Pohanish, R., 2014. Sittig's Handbook of Pesticides and Agricultural Chemicals (ed. Elsevier).

638 Raimondi, M.V., Cascioferro, S., Schillaci, D., Petruso, S., 2006. Synthesis and antimicrobial
639 activity of new bromine-rich pyrrole derivatives related to monodeoxypyoluteorin. Eur J Med
640 Chem. 41, 1439-1445. doi:10.1016/j.ejmech.2006.07.009.

641 Saltmarsh, M., 2000. Essential Guide to Food Additives, Leatherhead Publishing: Leatherhead, UK.
642 87–302. Doi:10.1046/j.1467-3010.2000.0088a.x

643 Sankarganesh, D., Suriyakalaa, U., Ramachandran, R., Achiraman, S., Arunachalam, S.,
644 Angayarkanni, J., 2019. Urinary volatile metabolomics as a viable alternative diagnostic tool for
645 polycystic ovary syndrome: An exploratory hypothesis. Med Hypotheses. 124,121-124.
646 doi:10.1016/j.mehy.2019.02.009.

647 Sikka, S.C., Ayaz, A., 2018. Standardized Semen Analysis and Quality Control Management for
648 Multicenter Male Reproductive Toxicology Clinical Trials, in Bioenvironmental Issues Affecting
649 Men's Reproductive and Sexual Health, 372-386.

650 Sun, J., 2007. d-Limonene: Safety and clinical applications. Altern. Med. Rev. 12, 259–264.

651 Taunk, K., Taware, R., More, T.H., Porto-Figueira, P., Pereira, J.A.M., Mohapatra, R., Soneji, D.,
652 Câmara, J.S., Nagarajaram, H.A., Rapole, S., 2018. A non-invasive approach to explore the
653 discriminatory potential of the urinary volatilome of invasive ductal carcinoma of the breast. RSC
654 Advances. 44. doi:10.1039/C8RA02083C

655 U.S. Environmental Protection Agency, 2004. Exposure and risk assessment on lower risk pesticide
656 chemicals; dLimonene. Arlington, VA: Office of Pesticide Programs.

Uddin, M.S., Blount, B.C., Lewin, M.D., Potula, V., Ragin, A.D., Dearwent, S.M. Comparison of blood volatile organic compound levels in residents of Calcasieu and Lafayette Parishes, LA, with US reference ranges. J Expo Sci Environ Epidemiol. 24(6), 602-607 (2014). doi:10.1038/jes.2013.94.

Wedzicha, B.L., 1984. Chemistry of Sulphur Dioxide in Foods Ist edition, London, UK: Elsevier Applied Science Publishers Ltd.

Whiting, S.J., Draper, H.H., 1981. Effect of chronic high protein feeding on bone composition in the adult rat. J Nutr 111, 178–183. doi:10.1093/jn/111.1.178.

Whiting, S.J., Draper, H.H., 1980. The role of sulfate in the calciuria of high protein diets in adult rats. J Nutr 110, 212–222. doi:10.1093/jn/110.2.212.

Wishart, D.S., Feunang, Y.D., Marcu, A., Guo, A.C., Liang, K., Vázquez-Fresno, R., Sajed, T., Johnson, D., Li, C., Karu, N., Sayeeda, Z., Lo, E., Assempour, N., Berjanskii, M., Singhal, S., Arndt, D., Liang, Y., Badran, H., Grant, J., Serra-Cayuela, A., Liu, Y., Mandal, R., Neveu, V., Pon, A., Knox, C., Wilson, M., Manach, C., Scalbert, A., 2018. HMDB 4.0: the human metabolome database for 2018. Nucleic Acids Res. 46, D608–17. doi:10.1093/nar/gkx1089.

World Health Organization international agency for research on cancer. IARC, 2010. Monographs on the Evaluation of Carcinogenic Risks to Humans. 99.

World Health Organization, 2010. WHO aboratory manual for the examination and processing of human semen 5th ed. (Cambridge: (Cambridge: Cambridge University Press).

World Medical Association, 2013. World medical association declaration of Helsinki: ethical principles for medical research involving human subject. 310, 2191–2194. doi:10.1001/jama.2013.281053

Yu,W.J., Chung, M.K., Chung, Y.H., Kim, H.C., Kim, S.H., Lee, I.C., Kim, J.C., 2011. One-generation reproductive toxicity study of 2-methylbutane in Sprague-Dawley rats. Regul Toxicol Pharmacol. 60, 136-143. doi:10.1016/j.yrtph.2011.03.003.

682

683

684 ***Table 1.*** *The participants' anthropometric and seminal data. Results are presented as range and means ± standard*
685 *deviations. *BMI=body mass index*

	Variable	Total subjects (n=50)		*Land of Fires* subjects (n=36)		*Valley of Sacco River* subjects (n=14)		
		Range (Min-Max)	**Mean ± SD**	**Range (Min-Max)**	**Mean ± SD**	**Range (Min-Max)**	**Mean ± SD**	**t-test (p-value)**
Anthopometric	Age	18-21	18.70±1.01	18-21	18.89±1.13	18-20	18.30±0.72	0.0766
	Body weight (kg)	51-100	73.70±11.00	51-100	74.30±12.00	60-84	72.20±7.79	0.5480
	Height (cm)	158-195	175±7	158-195	175±7	167-187	175±5	1
	BMI (kg/m^2)	17.87-32.84	23.95±3.18	17.87-32.84	24.15±3.35	19-28	23.4±2.55	0.4539
Standard semen variables	Volume (ml)	0.5-5	2.67±1.06	0.5-5	2.73±0.97	0.5-5	2.52±1.25	0.5297
	Sperm concentration (10^6/mL)	0-120	50.24±30.80	0-120	51.80±33.91	7-80	46.21±20.22	0.5673
	Total sperm number (10^6/ejaculate)	0-600	134.95±112.41	0-600	141.21±118.83	21-350	118.85±71	0.5141
	Total motility (%)	5-95	58.33±22.20	5-95	60.88±23.15	5-80	52.14±18.29	0.2121
	Progressive motility (%)	0-65	35.62±17.75	0-65	39.55±18.04	0-60	26.07±12.70	0.0139
	Immotility (%)	5-95	31.56±23.07	5-95	39.11±23.15	5-30	13.21±6.97	0.0002
	Morphological abnormalities (%)	87-97	92.14±2.79	87-97	92.35±2.76	87-96	91.64±2.79	0.4195
	Round cell concentration (10^6/mL)	1-13	4.29±2.87	1-13	4.73±2.91	1-10	3.21±2.45	0.0904
Sperm morphological defects	Head defects (%)	30-70	50.52±6.11	40-70	51.02±4.63	30-60	49.28±8.58	0.3589
	Neck defects (%)	14-50	24.66±5.70	14-30	22.82±2.35	20-50	29.14±8.35	0.0001
	Tail defects (%)	5-25	17.16±3.98	10-23	18.5±2.42	5-25	13.92±5.03	0.0001

686

687

Table 2. *Selected VOCs identified in SPME GC-MS analysis with power ROC >0.5 and p-value t-test >0.05.*
**LF > VSR indicates metabolites with concentration higher in Land of Fires than in Valley of Sacco River samples; VSR > LF indicates metabolites with concentration higher in Valley of Sacco River than in Land of Fires samples. AUC ROC (95% CI) = confidence interval of the Area Under the ROC curve at the level of 95%. power ROC = power for identifying the observed AUC given a level of significance α = 0.05. t-test = power for identifying the observed difference in the means given a level of significance α = 0.05. See table 2S for ROC curve results of all 77 variable.*

VOC name	Biological fluid	Trend*	AUC ROC (95% CI)	Power ROC	t-test
Pentane	Blood	LF>VSR	0.645-0.880	0.76429	1.4692E-4
Cyclohexane	Blood	VSR>LF	0.659-0.911	0.79464	2.0498E-4
1-(6-Methyl-benzothiazol-2-yl)-3-(4-methyl-benzoyl)-thiourea	Human Semen	VSR>LF	0.635-0.888	0.76071	4.2427E-4
Cyclopentane, methyl-	Blood	VSR>LF	0.478-0.758	0.61964	0.0036686
Acetic acid	Urine	LF>VSR	0.624-0.861	0.74643	0.004424
Octane	Blood	LF>VSR	0.214-0.407	0.69107	0.0071912
3-Aminopyrrolidine	Urine	VSR>LF	0.569-0.822	0.70000	0.0072461
Butane, 2-methyl-	Human Semen	VSR>LF	0.524-0.817	0.67500	0.016431
Propane, 2-ethenyloxy-	Urine	LF>VSR	0.541-0.793	0.65357	0.020706
Auramine	Human Semen	VSR>LF	0.546-0.833	0.69018	0.020854
Fluoren-9-ol, 3,6-dimethoxy-9-2-phenylethynyl-	Human Semen	VSR>LF	0.508-0.824	0.66071	0.028333
11H-Dibenzob,e1,4diazepin-11-one, 5,10-dihydro-5-3-methylaminopropyl-	Blood	LF>VSR	0.257-0.454	0.64554	0.02885
Pyrrole	Human Semen	VSR>LF	0.520-0.801	0.66071	0.029007
Benzaldehyde, 2-nitro-, diaminomethylidenhydrazone	Urine	VSR>LF	0.507-0.776	0.65714	0.03044
Acetic acid, sodium salt	Human Semen	VSR>LF	0.509-0.786	0.65893	0.033935
D-Limonene	Human Semen	VSR>LF	0.524-0.814	0.65536	0.035789
3-Aminopyrrolidine	Human Semen	VSR>LF	0.522-0.807	0.65714	0.041369
Benzaldehyde, 2-nitro-, diaminomethylidenhydrazone	Blood	LF>VSR	0.296-0.474	0.63036	0.043908
Acetic acid, sodium salt	Blood	LF>VSR	0.286-0.474	0.62857	0.048779

Figure 1. Number of VOCs per subject in the 3 different kinds of biological fluids (A) (Minimum, maximum, median, 25% and 75% percentile). Chemical classification of identified VOCs in total biofluids and singularly (B).

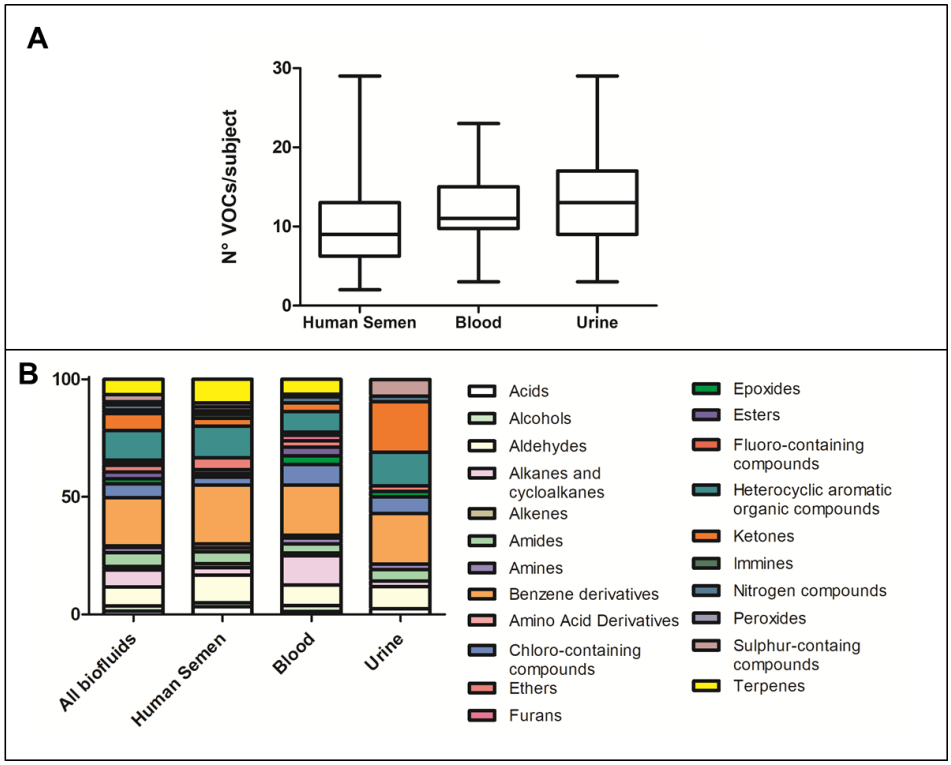


Figure 2. VOC repartition in human semen, blood and urine samples.

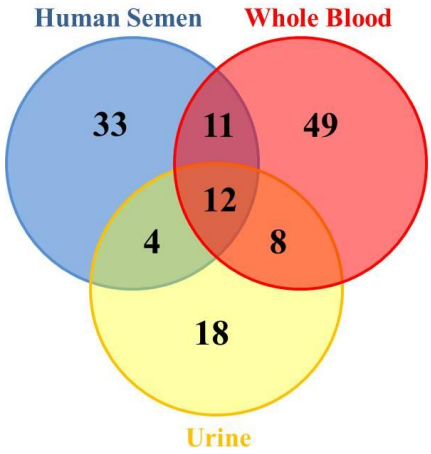


Figure 3. Percentages of VOCs found in human semen (A), blood(B) and urine (C) samples. Compounds which are detected in different biological fluid are indicated whit the same number. List of VOC name: 1:Acetone; 2:3-Methylbutanal; 3:Hexane; 4:3,6-Dimethoxy-9-(2-phenylethynyl)-fluoren-9-ol; 5:2-Methylbutanal; 6:Hexanal; 7:Pentanal; 8:2-Anthracenamine; 9:Auramine; 10:Oxime-,methoxy-phenyl-; 11:2-Ethyloxetane; 12:2-Methylbutane; 13:1-(6-Methyl-benzothiazol-2-yl)-3-(4-methyl-benzoyl)-thiourea; 14:1-Anthracenamine; 15:3-Aminopyrrolidine; 16:D-Limonene; 17:1-Benzenesulfonyl-1H-pyrrole; 18:1-Pentene; 19:3-(6-Methyl-3-pyridyl)-1,5-di(p-tolyl)-2-pyrazoline; 20:4-(4-Chlorophenyl)-2,6-diphenylpyridine; 21: 2-Cyanoacetamide; 22:Acetic acid, sodium salt; 23:Butanal; 24:Indolizine; 25:naphthalen-1-yl(1-pentyl-1H-indol-3-yl)methanone; 26:o-Cymene; 27: 2-(Ethenyloxy)-propane; 28:Pyrrole; 29:TATP; 30:(3-Methoxy-phenyl)-(6-methyl-4-phenyl-quinazolin-2-yl)-amine; 31:beta.-Myrcene; 32:beta.-Pinene; 33:gamma.-Terpinene; 34: 1,3-dimethyl-5,6-dimethoxy-2-(3,5-dimethoxyphenyl)-1H-Indole,; 35: 3-Propoxy-1-propene; 36:2-Amino-6-methylbenzoic acid; 37: 3-(Dimethylamino)-3-[(1-methylethyl)amino]-2-propenal; 38:3-(3-Carboxy-4-hydroxyphenyl)-D-alanine; 39:3-Carene; 40:4-Piperidinecarboxamide; 41: 6-methyl-5-Hepten-2-one; 42:6-Chloro-2,3-dimethyl-4-phenylquinoline; 43: 7-Methyl 7H-dibenzo[b,g]carbazole; 44: 1,3-Bis(1,1-dimethylethyl)-benzene; 45:Corydine; 46: 3,3-Dimethyldiaziridine; 47:Egtazic acid; 48:Ethanol; 49: Ethoxyethene; 50:Formic acid, ethenyl ester; 51: Methylguanidine; 52:Heptanal; 53: 1-Methyl-4-[4,5-dihydroxyphenyl]-hexahydropyridine,; 54:m-Aminophenylacetylene; 55:MDMA methylene homolog; 56:N-Isopropyl-3-phenylpropanamide; 57: 1,3-Dimethyl-8-[2-nitrophenethenyl]-purin-2,6-dione; 58: 3',4'-Dihydro-3-hydroxy-7',8'-dimethoxy-spiro[2H-indene-2,1'(2'H)-isoquinolin]-1(3H)-one; 59:Thebacon; 60:trans-4-Dimethylamino-4'-methoxychalcone; 61:4-Carene; 62:alpha.-Phellandrene; 63:alpha.-Pinene; 64:1,3-Butanediamine; 65:1,4-Dimethylazulene; 66: 5,10-dihydro-5-[3-(methylamino)propyl]-11H-Dibenzo[b,e][1,4]diazepin-11-one; 67: 2-Methyl-1-butanol,; 68: 4-[2-(2-chloro-4-nitrophenyl)diazenyl]-2-methyl-1-naphthalenol; 69:1-Nitro-9,10-dioxo-9,10-dihydro-anthracene-2-carboxylic acid diethylamide; 70:1-Sec-butyldiaziridine; 71:2-Amino-oxazole; 72:2-Butenediamide; 73:2-Chloro-4-(4-methoxyphenyl)-6-(4-nitrophenyl)pyrimidine; 74:2-Ethylacridine; 75:2-Heptanone; 76: 5,5-Dimethyl-2-hexene; 77: 1,5-Dihydro-1-(4-methoxyphenyl)-5,5-diphenyl-2H-pyrrol-2-one; 78:2-Pentanone; 79: 3-Chloro-N-ethyL-2-quinoxalinamine; 80:3-Buten-1-ol; 81:4,8-Dichloro-5-trifluoromethylquinoline; 82: 6,7-Dimethoxy-4H-3,1-benzoxazine; 83:5-Amino-2-phenyl-3,3,4(2H)-furantricarbonitrile; 84:6-Methylphenanthridine; 85: N-(1-Methylpropyl)-acetamide,; 86: 9-Methylacridine,; 87: 2-Nitro-, diaminomethylidenhydrazone-benzaldehyde; 88: 4-[2-[4-(2-Benzoxazolyl)phenyl]ethenyl]-benzoic acid; 89: 2-(2-Hydroxy-5-nitrobenzylidenamino)-benzonitrile; 90: 5-Chloro-3-methyl-2-(2-phenyl-

4-thiazolyl)-benzothiophene; 91:Cyclohexane; 92: 3-Hydroxycyclohexanone; 93:Methylcyclopentane; 94:Propylcyclopropane; 95:Dihydro-O,N-dimethyldehydrococcinine methine; 96: 2-Pentylfuran,; 97:Heptane; 98: 5-Methylhexanal; 99:Hydroperoxide, hexyl; 100:Isopropoxycarbamic acid, ethyl ester; 101:Lysergamide; 102:Methyl glyoxal; 103: 6-Chloro-1-nitronaphthalene; 104:N-Benzyl-N-ethyl-p-isopropylbenzamide; 105:Octanal; 106:Octane; 107: 2,2-Dimethyloxetane; 108: (1-Methylbutyl)-oxirane; 109:Pentane; 110: 3-Methylpentane; 111:Perazine; 112: N-Acetyl-3,4,5-trimethoxy-phenylpropylamine; 113: 4-Phenyl-pyrido[2,3-d]pyrimidine; 114: 4-(4-Chlorophenoxy)-8-fluoro-2-trifluoromethyl-quinoline; 115: Cyclohexylmethyl dodec-9-yn-1-yl ester, succinic acid; 116: N,N-Dimethyl, S-1,3-diphenyl-2-butenyl ester, thiocarbamic acid,; 117: 1,7,7-Trimethyl-tricyclo[2.2.1.0(2,6)]heptane; 118: 10-[(4-Methylphenyl)sulfanyl]-14-azatetracyclo[7.6.1.0{2,7}.0{13,16}]hexadeca-1(15),2(7),3,5,9(16),10,12-heptaen-8-one; 119: 5-Methyl-2-phenyl-1H-indole-; 120:2,4,5-Trioxoimidazolidine; 121: N-Acetyl-2-adamantylamine; 122:2-Butanone; 123:3-Hexanone; 124: 2-Methylhexanone; 125:4-Heptanone; 126:4-Phenyl-3,4-dihydroisoquinoline; 127: 6,10-Dimethyl-, 5,9-dodecadien-2-one; 128: N-Acetyl-N-(1-methylpropyl)-acetamide; 129: 5-Hydroxy-4-dimethylaminomethyl-2-phenyl-, ethyl ester, benzofurane-3-carboxylic acid; 130: 2-Allylamin cyclohexane-1,3-dione; 131:Dimethyl trisulfide; 132: Dimethyl, disulfide; 133:N-Methyl-2-(4-chlorophenyl)eth-2-en-2-yl-1H-benzimidazole; 134:Piperazine; 135: 2-Methylpropanal.

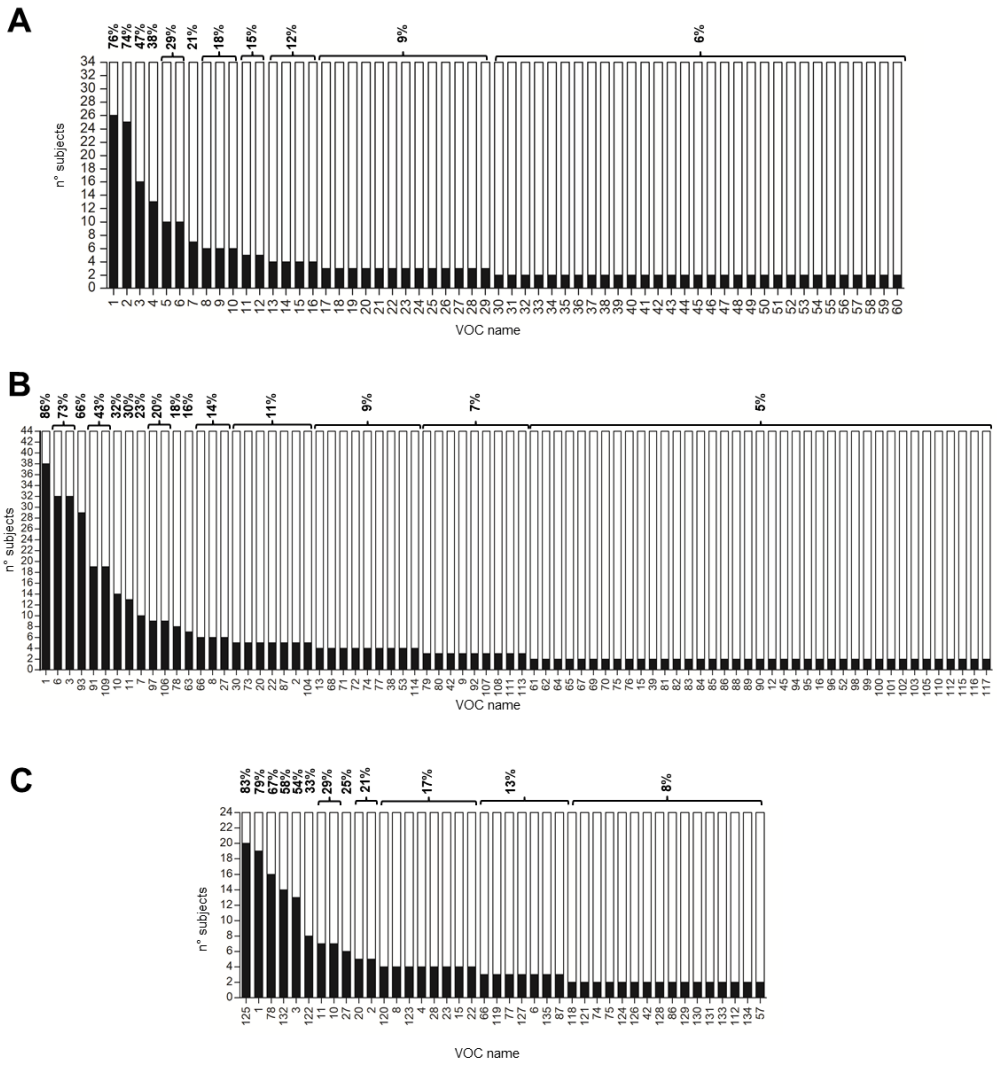

A
76%
74%
47%
38%
29%
21%
18%
15%
12%
9%
6%
n° subjects
VOC name
B
86%
73%
66%
43%
32%
30%
23%
20%
18%
16%
14%
11%
9%
7%
5%
n° subjects
VOC name
C
83%
79%
67%
58%
54%
33%
29%
25%
21%
17%
13%
8%
n° subjects
VOC name

Figure 4. VOCs detected in at least 10% of subjects within the same biofluids split into 7 subsets (A): VOCs common to A) all three body fluids, B) human semen and blood, C) human semen and urine, D) blood and urine or compounds exclusive of E) human semen, F) blood and G) urine samples. (B) Quantitative analysis of common VOCs in human semen, blood and urine samples. ** p-value< 0.001; ***p-value<0.0001. (Only statistically valid compounds are represented).

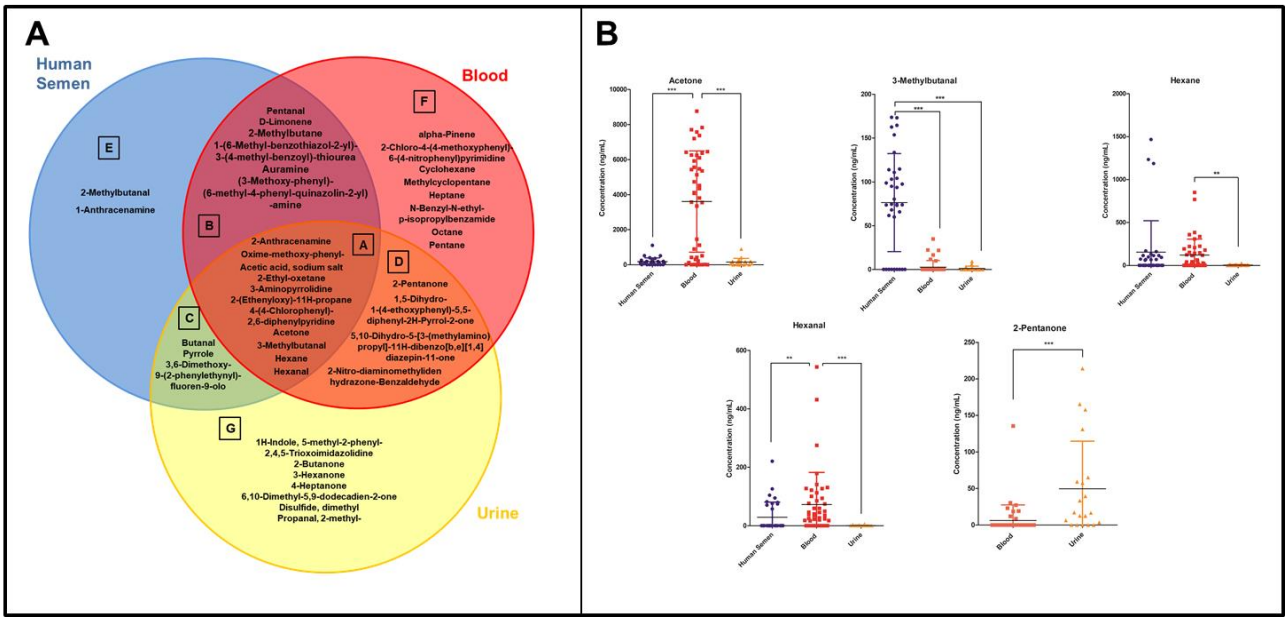
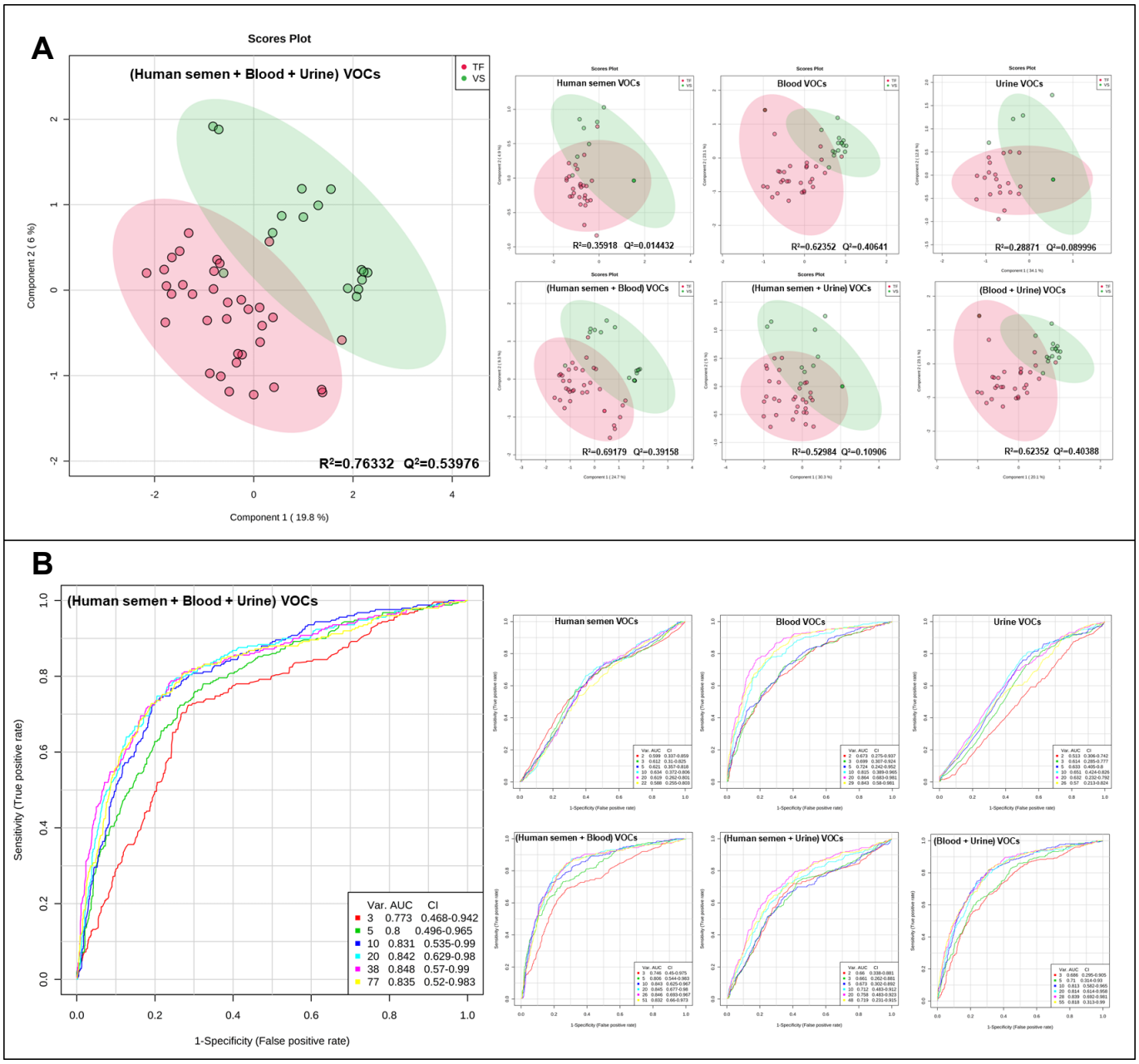


Figure 5. Score scatter plot of PLS-DA models (A) and ROC curves (B) built using concentration of 42 selected VOCs to discriminate subject living in Land of Fires and Valley of Sacco River.



Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: