Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 255 (2021) 119680

Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

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

A new method to predict genotoxic effects based on serum molecular profile

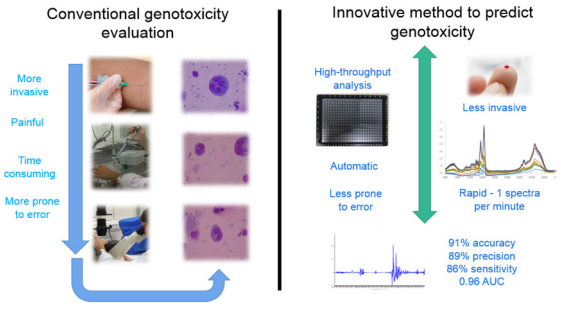
Rúben Araújo [a,*], Luís Ramalhete [a,b], Hélder Paz [a], Carina Ladeira [c,d,e,1], Cecília R.C. Calado [a,f,1]

[a] ISEL – Instituto Superior de Engenharia de Lisboa, Instituto Politécnico de Lisboa, Rua Conselheiro Emídio Navarro 1, 1959-007 Lisboa, Portugal
[b] CSTL-T – Centro de Sangue e da Transplantação de Lisboa – Instituto Português do Sangue e Transplantação, IP, Alameda das Linhas de Torres, n°117, 1769-001 Lisboa, Portugal
[c] H&TRC – Health & Technology Research Center, Escola Superior de Tecnologia da Saúde (ESTeSL), Instituto Politécnico de Lisboa, Avenida D. João II, lote 4.69.01, Parque das Nações, 1990-096 Lisboa, Portugal
[d] NOVA National School of Public Health, Public Health Research Centre, Universidade NOVA de Lisboa, Lisbon, Portugal
[e] Comprehensive Health Research Center (CHRC), Universidade NOVA de Lisboa, Portugal
[f] CIMOSM, ISEL – Centro de Investigação em Modelação e Optimização de Sistemas Multifuncionais, ISEL, Portugal

HIGHLIGHTS

- Genotoxicity prediction based on serum molecular profile.
- Genotoxicity prediction by serum FTIR spectroscopic analysis.
- Biomonitoring based on a drop of blood analysis by FTIR spectroscopy.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 11 January 2021
Received in revised form 19 February 2021
Accepted 1 March 2021
Available online 8 March 2021

Keywords:
Biomonitoring
Genotoxicity
CBMN
Serum
FTIR spectroscopy

ABSTRACT

It is critical to develop new methods to assess genotoxic effects in human biomonitoring since the conventional methods are usually laborious, time-consuming, and expensive. It is aimed to evaluate if the analysis of a drop of serum by Fourier Transform Infrared spectroscopy, allow to assess genotoxic effects in occupational exposure to cytostatic drugs in hospital professionals, as obtained by the lymphocyte cytokinesis-block micronucleus assay. It was considered peripheral blood from hospital professionals exposed to cytostatic drugs (n = 22) and from a non-exposed group (n = 36). It was observed that workers occupationally exposed presented a higher number of micronuclei ($p < 0.05$) in lymphocytes, in relation to the non-exposed group. The serum Fourier Transform Infrared spectra from exposed workers presented diverse different peaks ($p < 0.01$) in relation to the non-exposed group. The hierarchical cluster analysis of serum spectra separated serum samples of the exposed group from the non-exposed group with 61% sensitivity and 88% specificity. A support vector machine model of serum spectra enables to predict exposure with high accuracy (0.91), precision (0.89), sensitivity (0.86), F1 score (0.87) and AUC (0.96). Therefore, Fourier Transform Infrared spectroscopic analysis of a drop of serum enabled to predict in a rapid and simple mode the genotoxic effects of cytostatic drugs. The method presents therefore potential for high-dimension screening of exposure of genotoxic substances, due to its simplicity and rapid setup mode.

* Corresponding author.
E-mail address: rubenalexandredinisaraujo@gmail.com (R. Araújo).
[1] Authors contributed equally to this work.

https://doi.org/10.1016/j.saa.2021.119680
1386-1425/

1. Introduction

Cancer may be partially prevented by identifying and consequently controlling risk factors [1], as social, biological, and physical factors, that account for 39%, 18% and 31% of the causes of cancer, respectively [2]. Genotoxicity biomarkers allow the detection of early effects that result from the interaction between the individual and the environment, enabling to access health risks derived from exposure to complex mixtures as associated with occupational and environmental exposure [3–5]. The monitoring of these biomarkers is consequently used in cancer epidemiology, to identify risk factors while increasing the understanding of the complex relationship between exposure to genotoxic substances and cancer, and consequently enable better strategies for cancer prevention.

The most used matrices for human biomonitoring of exposure to genotoxic chemicals are blood and urine. These biofluids are analysed to identify the chemical substance and/or its metabolites, or its effect on biochemical targets, as proteins adducts (e.g., in albumin and haemoglobin), DNA adducts, DNA strand breaks or 8-hydroxy-2′-deoxyguanosine and chromosome aberrations, e.g., sister chromatid exchanges, micronuclei, among others [6].

One common method to analyse genotoxic effects in occupational studies, including medical based environments, chemical industry and agriculture, is based on the culture of human lymphocyte, obtained from peripheral blood, and the *in vitro* blocking of cytokinesis, and the subsequent microscope observation of nuclear abnormalities [7]. Among the wide range of cytogenetic biomarkers, the observation of micronucleus (MN) in lymphocytes after cytokinesis-blocked micronucleus (CBMN) assay provide a promising approach to assess health risk [8], being the preferred *in vitro* test in genotoxicity for human biomonitoring studies [9]. CBMN assay is a comprehensive system for measuring DNA damage, cytostasis and cytotoxicity-DNA damage events scored specifically in once-divided binucleated cells. It is a method for assessing DNA damage caused by xenobiotics, allowing detection of effects caused by clastogenic agents (that provoke chromosome breakage) and aneugenic agents (abnormal chromosome segregation associated with loss) [10–14]. Other eexamples of applications of this fundamental methodology to evaluate human health risks, supporting the establishment of a safe work environment, are reviewed e.g. in Ladeira and Smajdova [15].

Despite the high relevance of biomonitoring genotoxic exposure, as the ones based on CBMN assay, their impact in large scale epidemiological studies and consequently on public health are highly limited since, usually, the analytical techniques associated are laborious, time consuming and expensive. The present work will focus on evaluating a new method for biomonitoring, enabling to predict genotoxic exposure in a simple, economic, and high-throughput mode, but also with high sensitivity and specificity. A new biomonitoring technique presenting these characteristics, could be more easily implemented in routine analysis of populations at risk, or in large-scale monitoring studies to identify populations at risk. The method under evaluation is based on Fourier Transform InfraRed (FTIR) spectroscopy, since it is possible to conduct the technique with a simple sample processing, without expensive reagents, in a rapid mode (as a spectrum can be acquired in one minute) and in a high-throughput mode based on microplates with multi-wells [16,17]. The technique's simplicity is usually also associated to a high sensitive and specific analysis, enabling to acquire the metabolic status of the system, directly from the cells or from the cells environment, being consequently applied in a high diversity of biomedical applications, such as detection of infections [18], monitoring the impact of biocompounds consumption in plasma profile [19], identify the mechanism of cell death [20], cancer diagnosis and cancer grading [21,22], and diagnosis and prognosis of diverse other diseases and monitoring of treatments efficiency [23–25]. The technique's versatility also enables its application to a high diversity of biological samples, from tissue-based biopsies to body fluids such as blood, serum, urine, saliva and tears [24,26,27].

In the present work, it was considered serum obtained from peripheral blood of hospital workers potentially exposed to antineoplastic drugs due to their professional activity, such as pharmacists, pharmacy technicians and nurses and a control group from academia without exposure to cytostatic drugs. This sample is part of a previous research from the present group described in Ladeira et al. [28] and Ladeira et al. [29]. The hospital professionals worked in two Portuguese hospitals, with contact with working surfaces that were contaminated with the cytotoxic drugs Cyclophosphamide, 5-Fluorouracil and Paclitaxel [29]. These cytostatic drugs were evaluated as surrogate markers for surfaces contamination since they are amongst the most used in these hospitals, both in frequency and amount. In the present work, serum samples from exposed workers (n = 22) and controls (n = 36) from the previous studies [29] were analyzed by FTIR spectroscopy. Univariate and multivariate analysis of spectra were conducted to study if the new technique enabled to acquire the molecular profile of serum in comparison with the endpoints measured by the CBMN assay.

2. Materials and methods

2.1. Human volunteers

The human volunteers participating in the current study are described in the previous publications of Ladeira et al. [28,29]. Briefly, the study was conducted in accordance with standards of ethics and received necessary approvals (ACT project 036APJ/09) from the institution's administration and from each study participant with the informed consent form. From these participants, the available serum samples were analysed by FTIR spectroscopy, being included 22 serum samples from hospital workers occupationally exposed to antineoplastic drugs from two different hospitals, as nurses, pharmacists and pharmacy technicians, and 36 samples from controls, i.e. non-exposed group, with workers from an academic institution, namely teachers and office workers, chosen for having no contact with cytostatic drugs and for being statistically comparable to the exposed group, in terms of gender and age distribution.

2.2. CBMN assay

CBMN assay was conducted as described in Ladeira et al. [28]. Briefly, peripheral blood was collected by venipuncture for lymphocytes gradient isolation (Ficoll-Paque®, Sigma-Aldrich) and subsequently processed according to Fenech [30]. The cells were observed by optical microscope after air drying cells smears, in microscope slides and subsequently double-stained with May–Grünwald–Giemsa and mounted with Entellan®. Two slides were analyzed per sample, each one analyzed by an independent observer as described by Fenech et al. [31].

2.3. FTIR spectra acquisition

Serum was obtained by centrifugation of 1 mL heparinised whole blood samples at 3500 rpm for 10 min (Mikro with 1195/L rotor, Hettich) and maintained at −20 °C until further analysis. Triplicates of 25 μL of serum from each volunteer, previously diluted at 1/8 in NaCl (0.9%, w/v), were transferred to a 96-wells

Si plate and then dehydrated for about 2.5 h, in a desiccator under vacuum (Vacuubrand, ME 2, Germany). Spectral data was collected using a FTIR spectrometer (Vertex 70, Bruker, Germany) equipped with an HTS-XT (Bruker, Germany) accessory. Each spectrum represented 64 coadded scans, with a 2 cm^{-1} resolution, and was collected in transmission mode, between 400 and 4000 cm^{-1}. The first well of the 96-wells plate did not contain a sample and the corresponding spectra was acquired and used as background, according to the HTS-XT manufacturer.

2.4. Spectra pre-processing and processing

All spectra were pre-processed by atmospheric correction. Some spectra were also pre-processed by baseline correction, first derivative or second derivative using a Savitzky-Golay filter with a second order polynomial over a 15-point window. Atmospheric and baseline corrections were conducted with OPUS® software, v.6.5 (Bruker, Germany). All remaining pre-processing and processing analysis such as principal component analysis (PCA), hierarchical cluster analysis (HCA) and support vector machine (SVM) models were conducted with Orange 3 v.3.19.0 (Bioinformatics Lab, University of Ljubljana, Slovenia). HCA was based on absolute Pearson distances, complete linkage. SVM was based on a radial basis function kernel and a maximum iteration limit of 100 tests, in which the random sampling of the training size comprised of 80% of all samples, with the remainder for testing (i.e., prediction of exposed). *t*-student analysis was performed by SPSS® v.26 (IBM Corp, Armonk, NY, USA).

3. Results

3.1. General characterization of the sample and CBMN assay results

The volunteer's characteristics of exposed and non-exposed groups, including gender, age, tobacco, and alcohol consumption are shown in Table 1. It was observed that the exposed group presented lymphocyte with a significantly higher number of MN ($p < 0.05$) in relation to the non-exposed group (Table 2, Fig. 1).

3.2. Univariate data analysis of serum spectra

Fig. 2 represents the normal and derivative FTIR spectra of participants serum. It was observed a high variability of absorbances among participants, highlighting the high sensitivity of the FTIR spectroscopic technique to capture the variability of the serum biochemical composition among individuals (Fig. 2A).

To further highlight the biochemical information associated to the serum spectra, the spectra second derivative was considered (Fig. 2C-2F), as derivatives increase bands resolution, particularly relevant in biochemical complex biological samples, such as serum. From the second derivative spectra, 60 negative peaks were identified along the whole spectra. In relation to these peaks, 28 out of the 60 peaks (Table 3), were significantly ($p < 0.01$) different between the exposed and non-exposed groups, that included wavenumbers along the whole spectra, including vibrations in amide A and B f proteins (3300 and 3062 cm^{-1}), CH_3 in lipids (2960 and 2871 cm^{-1}), phospholipid esters (at 1749 cm^{-1}), amide II and I from proteins (1657, 1639, 1545 cm^{-1}), CH_3 and CH_2 vibrations present in lipids and proteins (1453, 1368 cm^{-1}), creatinine (1400 cm^{-1} [32], CO-O-C bonds in lipids (1171 cm^{-1}), carbohydrates (1053 and 1030 cm^{-1}), phosphate groups in lipids, proteins and phospholipids (1242 and 1081 cm^{-1}) and diverse peaks from the fingerprint region (936, 851, 782, 744, 723, 701, 676, 658, 548, 426 and 421 cm^{-1}) (Fig. 2). Fig. 3 points to second derivative absorbances obtained at the 3 wavenumbers that presented the lowest *p*-values between the two groups.

From the 28 wavenumbers identified as significantly different among the two groups, its standard deviation values were higher in the exposed group in relation to the non-exposed group except for 4 wavenumbers (426, 677, 713 and 782 cm^{-1}) (Table 3). Indeed, 18 out of the 28 wavenumbers of the exposed group presented 2-fold-higher standard deviations than the non-exposed group. Nine out of the 28 wavenumbers of the exposed group presented 10-fold-higher standard deviations values than the non-exposed group, 2960, 2871, 1657, 1545, 1453, 1401, 1314, 1242 and 1172 cm^{-1}, respectively. This points to a higher diversity of molecular profiles of serum, as acquired by FTIR spectroscopy, on exposed participants in relation to non-exposed participants.

3.3. Multivariate data analysis of serum spectra

To evaluate the whole spectra, diverse spectra multivariate data analysis was conducted, as principal component analysis (PCA), hierarchical cluster analysis (HCA) and support vector machines (SVM). Due to the relevance of pre-processing on spectra analysis, it was evaluated the impact of spectra baseline correction, first and second derivative on PCA (Fig. 4). PCA reduces the spectra data dimensionality, facilitating its data pattern analysis. The apparent best separation between data clusters from exposed and non-exposed participants was achieved with second derivative spectra, were the two first principal components (i.e. PC1 and PC2) represented a high data variance of the original spectra, of 88%. It was also with the second derivative as pre-processing that a lower

Table 1
Characteristics of human volunteers.

		Control	Exposed
Number of volunteers		36	22
Gender	Female	27 (75%)	19(86.4%)
	Male	9 (25%)	3 (13.6%)
Age	(mean ± SD, years)	40.6 ± 9.47	34.2 ± 8.1
	Range	20 – 61	24 – 58
Exposure time (mean and range)		Not applicable	8.8 (0.5 – 30 years)
Smoking habits – Smokers		7 (19.4%)	1 (4.5%)
Occasional drinkers		9 (25%)	4 (18.2%)

Table 2
Micronucleus (MN) in lymphocytes.

	Mean, MN lymphocytes ± SD (range)
Control (not exposed)	5.31 ± 6.42 (0–34)
Exposed	11.41 ± 11.10 (1–58)
p-value	0.03

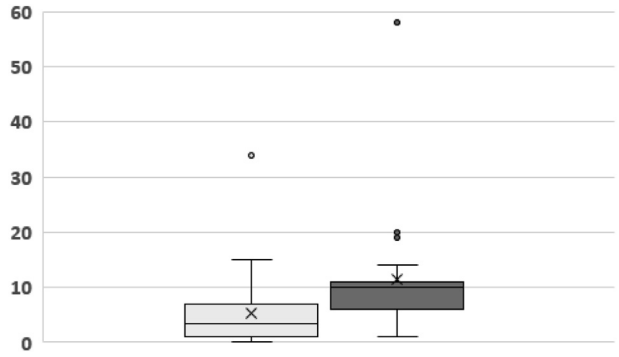


Fig. 1. Boxplots for the MN. Control group is shown in light grey, whereas the exposure group is represented in dark grey.

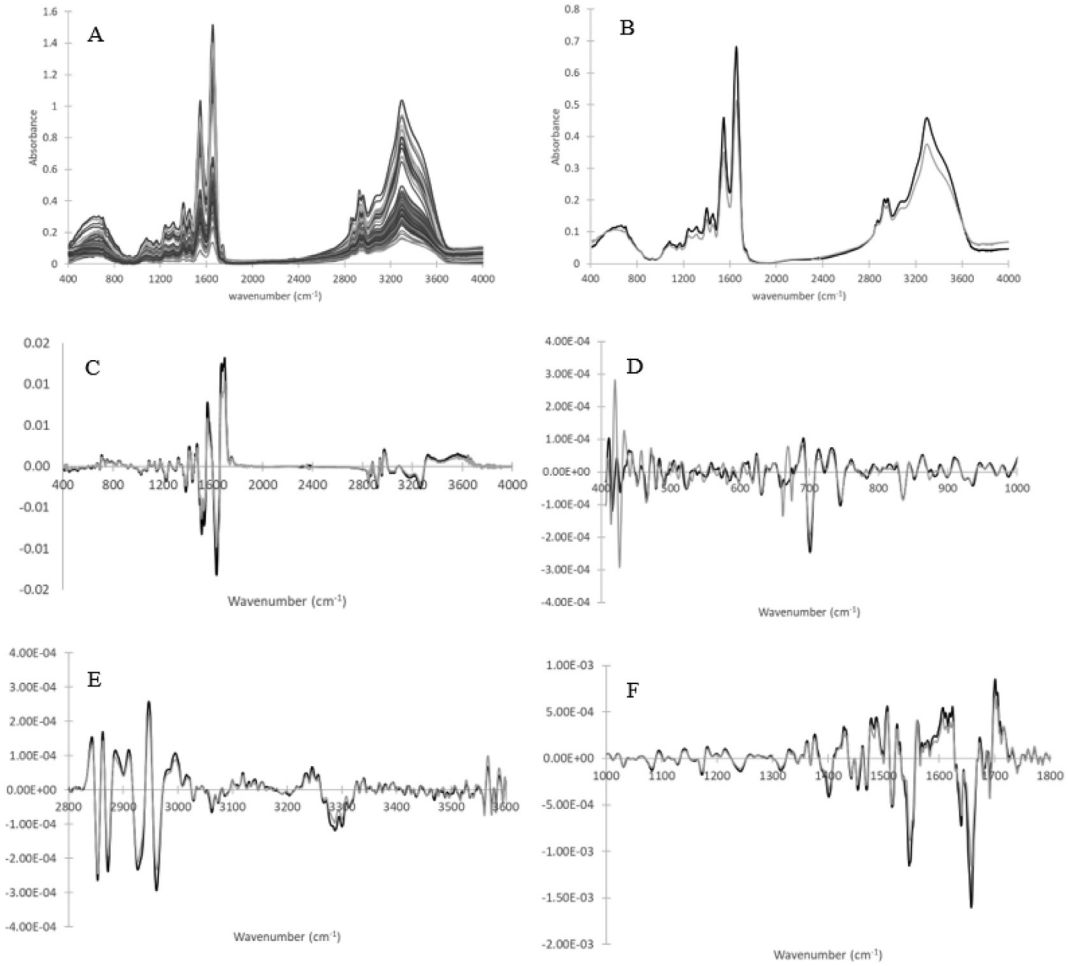


Fig. 2. Serum spectra from exposed (n = 22) and non-exposed (n = 36) individuals to cytostatic drugs after baseline correction (A). Average spectra of exposed (black line) and control (grey line) after baseline correction (B) first derivative (C) or second derivative (D to F).

Table 3
Average values and standard deviations of peaks of the serum second derivative spectra for the control and exposed group, with the corresponding *p*-value of Student's *t*-test.

Wavenumber (cm^{-1})	Control		Exposed		*p*-value (control vs exposed)
	Average	Standard deviation	Average	Standard deviation	
3299.92	−8,34E-05	1,45E-05	−1,06E-04	5,67E-05	2,09E-03
3062.70	−5,00E-05	7,77E-06	−6,08E-05	2,82E-05	3,59E-03
2960.48	−2,32E-04	3,69E-05	−2,94E-04	1,63E-04	3,70E-03
2871.76	−1,92E-04	3,05E-05	−2,38E-04	1,32E-04	6,82E-03
1740.61	−1,62E-04	5,53E-05	−1,27E-04	7,25E-05	1,20E-03
1657.68	−1,14E-03	1,66E-04	−1,61E-03	1,04E-03	6,27E-04
1639.35	−5,70E-04	1,07E-04	−7,26E-04	4,03E-04	3,12E-03
1545.81	−8,71E-04	1,18E-04	−1,16E-03	6,84E-04	1,29E-03
1500.49	−4,05E-05	8,76E-05	3,57E-05	1,23E-04	2,78E-05
1453.24	−2,79E-04	4,09E-05	−3,41E-04	1,82E-04	7,59E-03
1401.16	−3,34E-04	4,67E-05	−4,14E-04	2,33E-04	7,85E-03
1345.23	−6,05E-05	1,73E-05	−5,12E-05	2,46E-05	8,69E-03
1314.38	−1,02E-04	1,64E-05	−1,30E-04	7,73E-05	4,36E-03
1242.05	−1,10E-04	1,84E-05	−1,41E-04	8,02E-05	2,86E-03
1172.62	−1,42E-04	2,33E-05	−1,76E-04	9,82E-05	7,27E-03
1081.01	−9,63E-05	2,02E-05	−1,25E-04	7,74E-05	4,20E-03
1053.04	−8,79E-06	9,57E-06	−1,38E-05	1,33E-05	9,71E-03
936.36	−3,14E-05	1,08E-05	−4,26E-05	2,17E-05	2,01E-04
851.50	−1,43E-05	1,10E-05	−2,36E-05	2,11E-05	1,49E-03
782.07	7,60E-06	1,63E-05	−1,88E-05	1,30E-05	3,26E-23
744.46	−7,77E-05	2,23E-05	−1,03E-04	5,61E-05	8,32E-04
723.24	1,16E-05	1,83E-05	1,14E-06	1,70E-05	2,11E-04
701.06	−1,92E-04	3,17E-05	−2,47E-04	1,28E-04	1,05E-03
676.96	−1,59E-05	2,08E-05	−5,94E-06	1,98E-05	2,03E-03
658.63	−1,41E-05	1,88E-05	−4,76E-05	4,12E-05	2,50E-08
548.70	−2,96E-05	2,86E-05	−1,27E-05	4,26E-05	5,65E-03
426.23	−2,92E-04	1,18E-04	−5,40E-05	1,31E-04	1,35E-22
421.41	1,87E-04	1,23E-04	3,35E-05	1,62E-04	1,84E-09

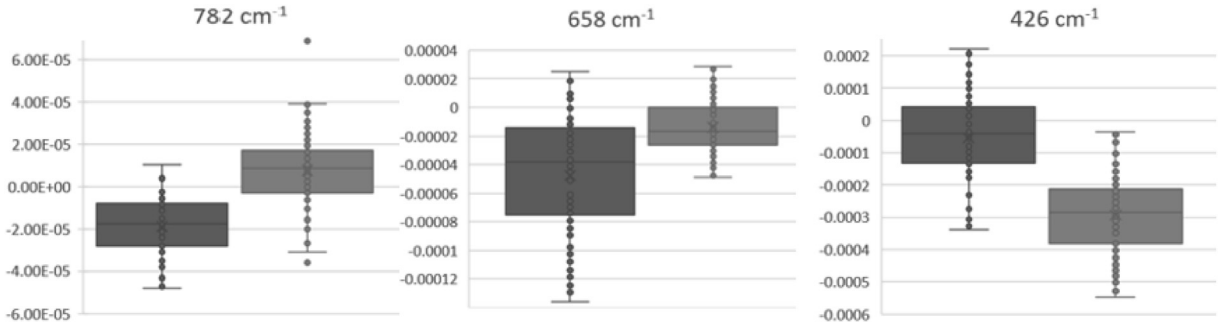


Fig. 3. Boxplots of serum second derivative absorbances at 426, 658 and 782 cm^{-1} of exposed (left bars) and non-exposed individuals (right bars).

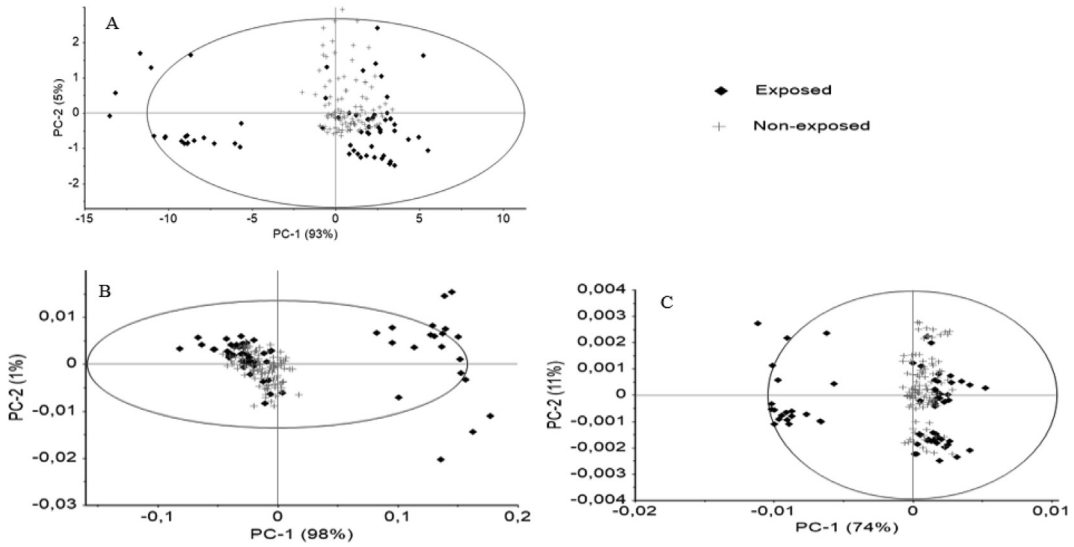


Fig. 4. PCA of serum spectra from exposed and non-exposed individuals, pre-processed with baseline correction (A), first derivative (B) and second derivative (C), and the corresponding Hotelling's ellipse at 1% significance.

number of sample outliers, that were samples outside the Hotelling's ellipse, were identified. All this highlights the relevance of second derivatives on deconvolution spectral bands, and consequently on increasing the information retrieved from the spectra as mentioned above. Consequently, the remaining analysis was based on second derivative spectra.

Fig. 5 represents the loadings of PC1 of the PCA based on second derivative spectra, pointing the spectral regions that most contributed for samples separation between exposed and non-exposed individuals, and that included regions along the whole spectra, such as between 400 and 1500 cm^{-1}, known as the fingerprint region, due to overlapped vibrations from diverse molecules (Fig. 2D); between 1500 and 1800 cm^{-1}, mostly due to double bonds as C = O, C = C, and C = N, as present in amide I (~1650 cm^{-1}) amide II (~1550 cm^{-1}) of proteins, C = O as from phospholipids esters (~1740 cm^{-1}) (Fig. 2E); and between 2000 and 3600 cm^{-1}, mostly due to vibrations between X-H (where X can be C, O, or N), as present in amide A (~3300 cm^{-1}), from proteins, CH_3 (~2955 and ~2870 cm^{-1}) and CH_2 (~2918 and ~2850 cm^{-1}) groups from lipids (Fig. 2F).

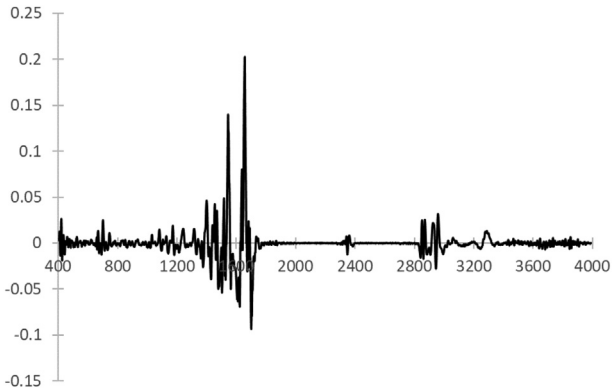


Fig. 5. PC2 loading of PCA based on second derivative spectra.

An HCA based on serum second derivative spectra was conducted to better evaluate the discrimination between samples from the exposed and non-exposed group. The HCA dendrogram represents the similarities between samples in a tree mode, where the distances relate quantitatively the samples similarities. Cluster 1 of the dendrogram of HCA based on second derivative spectra, was classified as the cluster of the exposed group, containing 60% of serum samples from exposed individuals, whereas cluster 2, was designated of non-exposed individuals since it contained 88% of samples from non-exposed individuals (Fig. 6). However, cluster 1 and 2 also presented, 40% and 12% of samples of non-exposed and exposed individuals, respectively, which resulted in a classification to detect exposed individuals with a sensitivity and specificity of 60% and 88%, respectively.

Fig. 6. HCA of serum spectra from exposed and non-exposed individuals, pre-processed with second derivative.

Supervised models based on serum second derivative spectra were also developed based on SVM to evaluate if it was possible to predict from the serum spectra if that individual was exposed to cytostatic drugs as observed by the CBMN assay. SVM uses spectra close to the class borders as support vectors to define a discriminant surface [33], presenting high advantages for high dimension input space [34,35]. A SVM model with high accuracy (0.91), precision (0.89), sensitivity (0.86) and a high weighted average of prevision and sensitivity (i.e. F1 score) (0.87) was obtained. This model also resulted in an excellent area under the curve (AUC), i.e. with a value close to 1 (0.96), which allowed to maximise true positive rate (i.e. the recall) and minimise false positive rate. In Table 4 it is presented the SVM confusion matrix.

4. Discussion

The significant higher MN ($p < 0.05$) observed by the CBMN assay in lymphocytes of the exposed group to cytostatic drugs ($n = 22$) in relation to the control (i.e. non-exposed) group ($n = 36$), are most probably due the antineoplastic drugs effect, since the exposed group was composed by professionals that handled cytostatic drugs at hospital services were 37% of the surfaces were contaminated with the 3 most used antineoplastic drugs in hospitals, i.e. Cyclophosphamide, 5-Fluorouracil and Paclitaxel as previously studied by the present research group [28,29]. All these antineoplastic drugs can interact with DNA, inducing double and single strand breaks, crosslinks, alkylation's, and DNA intercalations, which could account, at least partially, for the genotoxicity observed.

It was subsequently observed that the FTIR spectra enabled to discriminate individuals from the group presented genotoxic effects due to exposition to cytostatic drugs in relation to individuals from the control group. Indeed, the HCA based on the spectra second derivative, clustered individuals between exposed and non-exposed groups with a sensitivity and specificity of 60% and 82%, respectively. A very good SVM model, based on the second derivative spectra, also enabled to predict if a participant was exposed to cytostatic drugs with high value of accuracy (0.91) precision (0.89), sensitivity (0.86) high F1 score (0.87) and AUC (0.96).

The serum molecular profile as acquired by the spectra, also pointed that the exposition to genotoxic substances imply alterations on the serum whole molecular profile, as affected its composition on a high diversity of chemical functional groups. For example, the loading of PC2 highlighted the proteins' regions (amide I and II at ~1650 and ~1550 cm^{-1}) (Fig. 5), were 28 peaks of the second derivative spectra were significantly different between the exposed and non-exposed group, and that were associated to high diversity of molecular bonds related to different types of groups/ molecules as phosphates, proteins, lipids and polysaccharides (Table 3). These observations are according to the high impact the cytostatic drugs, as 5-FU based therapy on patients' whole blood composition, including its cells constituents and serum composition as total proteins, creatinine, triglycerides, and cholesterol [36,37].

Spectral data also indicated that the serum molecular profile of exposed individuals presented higher variability among individuals than the serum composition among non-exposed participants. This was deduced from: i) a higher data dispersion in the PCA score-plot from exposed participants in relation to non-exposed participants (Fig. 4); ii) a higher standard deviation value in 23 out of 28 wavenumbers referred above observed among exposed individuals in relation to non-exposed individuals (Table 3). This

Table 4
Confusion matrix of SVM model.

		Predicted	
		Exposed	Control
Experimental	Exposed	89.9%	8.4%
	Control	11.1%	91.6%

higher variability in serum composition in exposed individuals may result from the different exposure time, that varied from 5 months to 30 years, (Table 1) and/or due to individual different metabolic sensitivity to cytostatic drugs. For example, in relation to 5-FU, it is known that its pharmacokinetics depends on diverse factors as genetic variations of the enzyme responsible by 80% of the drugs catabolism (i.e. the dihydropyrimidine dehydrogenase), but also from patients age, serum alkaline phosphatase, among other factors [38,39].

Therefore, FTIR spectroscopy enabled to acquire the molecular profile of serum, pointing a molecular whole composition on the occupationally exposed participants substantial different from the non-exposed participants to cytostatic drugs. The technique presents the advantage of acquiring the serum molecular profile based on the analysis of a drop of serum, as acquired in an rapid (i.e. in a time scale of 1 min), and simple (with only a previous dehydration step is needed) and high-throughput (based on microplate with multi-wells) mode.

5. Conclusions

It is relevant to monitor the impact of exposure to a high diversity of genotoxic chemicals present in living and working environments. However, the routine human monitoring of these exposures, as based on CBMN assay, is impaired as the methods used are usually costly, laborious, time-consuming, and in need of human expertise, at times not found in local hospitals and even clinical laboratories. The present study considered the effect of cytostatic exposure to antineoplastic drugs on exposed hospital professionals in relation to a control group, i.e., not exposed to these cytostatic drugs, and that presented a significant different mean value of MN as analysed by the conventional CBMN assay. Univariate and multivariate data analysis of serum FTIR spectra enabled to predict if a defined individual belonged to the exposed or not exposed group, i.e., enabled to predict the individual cytostatic exposure, with in a high specificity, sensitivity, and accuracy. The serum molecular profile acquired by the spectra also pointed to a higher variability on serum composition among exposed participants in relation to non-exposed participants. It was observed that cytostatic exposure resulted in a high metabolic impact since it led to a significant alteration of a high diversity of biomolecules (as proteins, polysaccharides, and lipids) of serum, as observed by univariate analysis of specific wavenumbers and in the PCA loadings. All these information's were acquired by a rapid, simple, high-throughput and economic spectroscopic analysis, and based in a small serum sample. The major limitation of the present study is the population's low dimension (n = 22 and n = 36 for the exposed and control samples, respectively). Therefore, for a continuing work, a higher dimension population should be evaluated. Despite this limitation, the method presents, characteristics with a high potential to be applied e.g., in routine analysis of populations at risk, or in large-scale monitoring studies to identify populations at risk, and/or new genotoxic substances. It also presents high potential to evaluate the complex relationship between exposure to genotoxic substances and cancer. All of these could contribute to a decrease in exposure to genotoxic substances and consequently enable better cancer prevention strategies going forward.

Declarations

Funding

This work was supported by the *Instituto Politécnico de Lisboa* under grants IDI&CA/IPL/2017/GenTox/ESTeSL, 2018/RenalProg/ISEL and 2020/NephroMD. The present work was partially conducted in the Engineering & Health Laboratory through a collaboration protocol established between *Universidade Católica Portuguesa* and the *Instituto Politécnico de Lisboa*.

CRediT authorship contribution statement

Rúben Araújo: Conceptualization, Methodology, Data curation, Writing - original draft, Visualization, Investigation, Formal analysis, Software, Writing - review & editing. **Luís Ramalhete:** Data curation, Visualization, Software, Investigation, Funding acquisition. **Hélder Paz:** Methodology, Resources. **Carina Ladeira:** Conceptualization, Methodology, Resources, Funding acquisition. **Cecília R.C. Calado:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Supervision, Funding acquisition, Project administration.

Declaration of Competing Interest

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.

References

[1] M.C. White, D.M. Holman, J.E. Boehm, L.A. Peipins, M. Grossman, S.J. Henley, Age and cancer risk: a potentially modifiable relantionship, Am. J. Prev. Med 46 (3) (2014) S7–S15, https://doi.org/10.1016/j.amepre.2013.10.029.
[2] G.A. Colditz, E.K. Wei, Preventability of cancer: the relative contributions of biologic and social and physical environmental determinants of cancer mortality, Annu. Rev. Public Health 33 (1) (2014) 137–156, https://doi.org/10.1146/annurev-publhealth-031811-124627.
[3] J.C. Barrett, H. Vainio, D. Peakall, B.D. Goldstein, 12th meeting of the scientific group on methodologies for the safety evaluation of chemicals: susceptibility to environmental hazards, Environ. Health Perspect. 105 (suppl 4) (1997) 699–737, https://doi.org/10.1289/ehp.97105s4699.
[4] J.M. Battershill, K. Burnett, S. Bull, Factors affecting the incidence of genotoxicity biomarkers in peripheral blood lymphocytes: impact on design of biomonitoring studies, Mutagenesis 23 (6) (2008) 423–437, https://doi.org/10.1093/mutage/gen040.
[5] C. Ladeira, S. Viegas, Human Biomonitoring – An overview on biomarkers and their application in occupational and environmental health, Biomonitoring 3 (1) (2016), https://doi.org/10.1515/bimo-2016-0003.
[6] J. Angerer, U. Ewers, M. Wilhelm, Human biomonitoring: State of the art, Int. J. Hyg. Environ. Health 210 (3–4) (2007) 201–228, https://doi.org/10.1016/j.ijheh.2007.01.024.
[7] A. Nersesyan, M. Fenech, C. Bolognesi, M. Mišík, T. Setayesh, G. Wultsch, et al., Use of the lymphocyte cytokinesis-block micronucleus assay in occupational biomonitoring of genome damage caused by in vivo exposure to chemical genotoxins: Past, present and future, Mutat. Res. Rev. Mutat. Res. 770 (2016) 1–11, https://doi.org/10.1016/j.mrrev.2016.05.003.
[8] L.E. Knudsen, Å.M. Hansen, Biomarkers of intermediate endpoints in environmental and occupational health, Int. J. Hyg. Envir. Health 210 (3–4) (2007) 461–470, https://doi.org/10.1016/j.ijheh.2007.01.015.
[9] OECD, 2016. Test No. 487: In Vitro Mammalian Cell Micronucleus Test (2016), OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing. https://doi.org/10.1787/9789264264861-en
[10] M. Fenech, The advantages and disadvantages of the cytokinesis-block micronucleus method, Mutat. Res. Genet. Toxicol. Environ. Mutagen. 392 (1–2) (1997) 11–18, https://doi.org/10.1016/S0165-1218(97)00041-4.
[11] M. Fenech, The in vitro micronucleus technique, Mutat. Res. 455 (1–2) (2000) 81–95, https://doi.org/10.1016/S0027-5107(00)00065-8.
[12] M. Fenech, J.W. Crottm, Micronuclei, nucleoplasmic bridges and nuclear buds induced in folic acid deficient human lymphocytes—evidence for breakage–fusion-bridge cycles in the cytokinesis-block micronucleus assay, Mutat. Res. 504 (1–2) (2002) 131–136, https://doi.org/10.1016/S0027-5107(02)00086-6.
[13] G. Göethel, N. Brucker, M. Moro, A.F. Charão, M. Fracasso, et al., Evaluation of genotoxicity in workers exposed to benzene and atmospheric pollutants, Mutat. Res. Genet. Toxicol. Environ. Mutagen 770 (2014) 61–65, https://doi.org/10.1016/j.mrgentox.2014.05.008.
[14] R. Mateuca, N. Lombaert, P.V. Aka, I. Decordier, M. Kirsch-Volders, Chromosomal changes: induction, detection methods and applicability in human biomonitoring, Biochimie 88 (11) (2006) 1515–1531, https://doi.org/10.1016/j.biochi.2006.07.004.
[15] C. Ladeira, L. Smajdova, The use of genotoxicity biomarkers in molecular epidemiology: applications in environmental, occupational and dietary studies, AIMS Genet. 4 (3) (2017) 166–191, https://doi.org/10.3934/genet.2017.3.166.
[16] F. Rosa, K.C. Sales, B.R. Cunha, A. Couto, M.B. Lopes, C.R.C. Calado, A comprehensive high-throughput FTIR spectroscopy-based method for evaluating the transfection event: estimating the transfection efficiency and

extracting associated metabolic responses, Anal. Bioanal. 407 (26) (2015) 8097–8108, https://doi.org/10.1007/s00216-015-8983-9.
[17] S.C. Sales, F. Rosa, B.R. Cunha, P.N. Sampaio, M.B. Lopes, C.R.C. Calado, Metabolic profiling of recombinant Escherichia coli cultivations based on high-throughput FT-MIR spectroscopic analysis, Biotechnol. Prog. 33 (2) (2017) 285–298, https://doi.org/10.1002/btpr.2378.
[18] V. Marques, B. Cunha B., A. Couto, P. Sampaio, L.P. Fonseca, S. Aleixo, C. R.C. Calado, Characterization of gastric cells infection by diverse Helicobacter pylori strains through Fourier-transform infrared spectroscopy. Spectrochim. Acta A, 210, 2019, 193–202. https://doi.org/10.1016/j.saa.2018.11.001
[19] R. Araújo, L. Ramalhete, H. Da Paz, E. Ribeiro, C.R.C. Calado, A simple, label-free, and high-throughput method to evaluate the epigallocatechin-3-gallate impact in plasma molecular profile, High-Throughput 9 (2) (2020) 9, https://doi.org/10.3390/ht9020009.
[20] B.R. Cunha, L.P. Fonseca, C.R.C. Calado, Metabolic fingerprinting with fourier-transform infrared (FTIR) spectroscopy: towards a high-throughput screening assay for antibiotic discovery and mechanism-of-action elucidation, Metabolites 10 (4) (2020) 145, https://doi.org/10.3390/metabo10040145.
[21] G. Bellisola, C. Sorio, Infrared spectroscopy and microscopy in cancer research and diagnosis. American Journal of Cancer Research, 2(1) (2012) 1–21. PMCID: PMC3236568
[22] E. Gazi, M. Baker, J. Dwyer, N.P. Lockyer, P. Gardner, et al., A correlation of FTIR spectra derived from prostate cancer biopsies with Gleason grade and tumour stage, Eur. Urol. 50 (4) (2006) 750–761, https://doi.org/10.1016/j.eururo.2006.03.031.
[23] A.A. Bunaciu, S. Fleschin, V.D. Hoang, H.Y. Aboul-Enein, Vibrational spectroscopy in body fluids analysis, Crit. Rev. Anal. Chem. 47 (1) (2017) 67–75, https://doi.org/10.1080/10408347.2016.1209104.
[24] S. De Bruyne, M.M. Speeckaert, J.R. Delanghe, Applications of mid-infrared spectroscopy in the clinical laboratory setting, Crit. Rev. Clin. Lab. Sci. 55 (1) (2018) 1–20, https://doi.org/10.1080/10408363.2017.1414142.
[25] W.B. Dunn, D. Broadhurst, P. Begley, E. Zelena, F. McIntyre, et al., Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry, Nat. Protoc. 6 (7) (2011) 1060–1083, https://doi.org/10.1038/nprot.2011.335.
[26] M.J. Baker, S.R. Hussain, L. Lovergne, V. Untereiner, C. Hughes, et al., Developing and understanding biofluid vibrational spectroscopy: a critical review, Chem. Soc. Rev. 45 (7) (2016) 1803–1818, https://doi.org/10.1039/C5CS00585J.
[27] M.J. Baker, J. Trevisan, P. Bassan, R. Bhargava, H.J. Butler, et al., Using Fourier transform IR spectroscopy to analyze biological materials, Nat. Protoc. 9 (8) (2014) 1771–1791, https://doi.org/10.1038/nprot.2014.110.
[28] C. Ladeira, S. Viegas, M. Pádua, M. Gomes, E. Carolino, M.C. Gomes, M. Brito, Assessment of genotoxic effects in nurses handling cytostatic drugs, J. Toxicol. Environ. Health Part A 77 (14–16) (2014) 879–887, https://doi.org/10.1080/15287394.2014.910158.
[29] C. Ladeira, S. Viegas, M. Pádua, E. Carolino, M.C. Gomes, M. Brito, Relation between DNA damage measured by comet assay and OGG1 Ser326Cys polymorphism in antineoplastic drugs biomonitoring, AIMS Genetics 2 (3) (2015) 204–218, https://doi.org/10.3934/genet.2015.3.204.
[30] M. Fenech, Cytokinesis-block micronucleus cytome assay, Nat. Protoc. 2 (5) (2007) 1084–1104, https://doi.org/10.1038/nprot.2007.77.
[31] M. Fenech, N. Holland, W.P. Chang, E. Zeiger, S. Bonassi, The Human MicroNucleus Project—An international collaborative study on the use of the micronucleus technique for measuring DNA damage in humans, Mutat. Res. 428 (1–2) (1999) 271–283, https://doi.org/10.1016/S1383-5742(99)00053-8.
[32] D.P. Jerônimo, R.A. de Souza, F.F. da Silva, G.L. Camargo, H.L. Miranda, et al., Detection of creatine in rat muscle by FTIR spectroscopy, Ann. Biomed. Eng. 40 (9) (2012) 2069–2077, https://doi.org/10.1007/s10439-012-0549-9.
[33] C. Krafft, G. Steiner, C. Beleites, R. Salzer, Disease recognition by infrared and Raman spectroscopy, J. Biophotonics 2 (1–2) (2009) 13–28, https://doi.org/10.1002/jbio.200810024.
[34] Z.R. Yang, Biological applications of support vector machines, Brief. Bioinf. 5 (4) (2005) 328–338, https://doi.org/10.1093/bib/5.4.328.
[35] F. Deng, J. Huang, X. Yuan, C. Cheng, L. Zhang, Performance and efficiency of machine learning algorithms for analyzing rectangular biomedical data, Lab Invest (2021), https://doi.org/10.1038/s41374-020-00525-x.
[36] B.R. Ali, Physiological and biochemical changes for 5- fluorouracil drug (anti-cancer) in cancer patients, Bas. J. Vet. Res. 16 (1) (2017) 87–99, https://doi.org/10.33762/bvetr.2017.127439.
[37] B.R. Ali, Physiological and biochemical changes for 5-fluorouracil drug (anti-cancer) in cancer patients, Bas. J. Vet. Res. 16 (1) (2017) 87–99, https://doi.org/10.33762/bvetr.2017.127439.
[38] G. Bocci, C. Barbara, F. Vannozzi, A. Dipaolo, A. Melosi, et al., A pharmacokinetic-based test to prevent severe 5-fluorouracil toxicity, Clin. Pharmacol. The. 80 (4) (2006) 384–395, https://doi.org/10.1016/j.clpt.2006.06.007.
[39] G. Milano, A.L. Chamorey, Clinical pharmacokinetics of 5-fluorouracil with consideration of chronopharmacokinetics, Chronobiol. Int. 19 (1) (2002) 177–189, https://doi.org/10.1081/CBI-120002597.