HUMAN URINARY METABOLIC SIGNATURES BY PARALLEL DUAL SECONDARY COLUMN-DUAL DETECTION TWO-DIMENSIONAL COMPREHENSIVE GAS CHROMATOGRAPHY: RELIABLE TARGETED AND UNTARGETED COMPARATIVE ANALYSIS

Davide Bressanello[1], Erica Liberto[1], D. W. Rempe[2], Q. Tao[3], Stephen E. Reichenbach[2], Stefano Balducci[4], Elisa Benetti[1], Carlo Bicchi[1], Chiara E. Cordero[1]

[1]*Dipartimento di Scienza e Tecnologia del Farmaco, Università degli Studi di Torino, Via Pietro Giuria 9, I-10125 Torino, Italy*
[2]*Computer Science and Engineering Department, University of Nebraska* 1400 R Street, Lincoln, NE 68588-0115, *United States*
[3]*GC Image LCC, PO Box, 57403 Lincoln, United States*
[4] *La Sapienza University, Piazzale Aldo Moro 5, 00185 Roma, Italy*

SAPIENZA UNIVERSITÀ DI ROMA

Aim and Scope

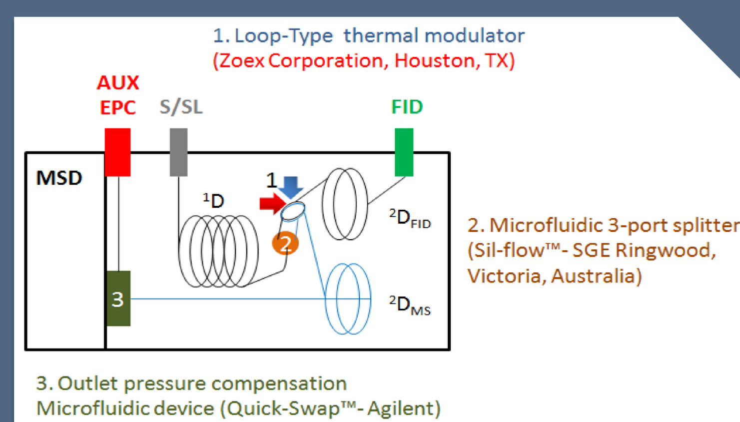
This study investigates the potential of a parallel dual secondary column-dual detection two dimensional comprehensive GC platform (GC×2GC–MS/FID) in delineating informative metabolic signatures in human urine samples. The platform was firstly tested on a dietary manipulation study on mice and demonstrated to be effective thanks to the synergic combination of full scan mass spectrometry for analytes identification with flame ionization detection for (semi)quantitative pattern comparison [1]. The present study is indeed focused on human urine samples collected from adults (males and females) with type-2 diabetes [2]. Samples submitted to a standard derivatization protocol [3] were analyzed by GC×2GC-MS/FID for profiling and fingerprinting of low-molecular weight metabolites (acids, aminoacids, monosaccharides ect..). Thanks to an effective template matching algorithm that adopts a newly implemented global mapping function [1], the parallel separation patterns from the two detectors (FID and MS) were cross-aligned and untargeted region-features adopted for comparative analysis. Contemporarily a targeted analysis was run by focusing on known informative metabolites aimed at corroborate and validate the untargeted approach.
By comparing targeted profiling with untargeted fingerprinting results it is possible to select relevant (peak-region) features showing statistically consistent variations between subjects belonging to different Gaussian quartiles.

[1] D.Bressanello, E.Liberto, M.Collino, S.E. Reichenbach, E.Benetti, F. Chiazza, C.Bicchi, C. .Urinary metabolic fingerprinting of mice with diet-induced metabolic derangements by parallel dual secondary column-dual detection two-dimensional comprehensive gas chromatography, Journal of Chromatography A 1361 (2014) 265
[2] S. E. Reichenbach ; D.W. Rempe ; Qingping Tao ; D.Bressanello ; E.Liberto ; C. Bicchi ; S. Balducci ; C. Cordero. Alignment for Comprehensive Two-Dimensional Gas Chromatography with Dual Secondary Columns and Detectors, Analytical Chemistry 87 (2015) 10056
[3] Q. Zhang, G. Wang, Y. Du, L. Zhu, A. Jiye, GC/MS analysis of the rat urine for metabonomic research, Journal of Chromatography B 854 (2007) 20–25.
[4] Medina et al; metabolomics markers in acute and endurance/resistan phisica activity effect: of the diet, Cap 14,Foodomics: Advanced Mass Spectrometry in Modern Food Science and Nutrition,April 2013;

Experimental

GC×2GC-MS/FID platform

Column set consisted of **primary column of 30 m × 0.25 mm d_c × 0.25 µm d_f SE52 (95% polydimethylsiloxane, 5% phenyl)** connected to **two secondary columns of equivalent length of 1.4 m × 0.1 mm d_c × 0.10 µm d_f OV1701 (86% polydimethylsiloxane, 7% phenyl, 7% cyanopropyl)**. The secondary column toward the MS detector was connected to a **Quick Swap unit (G3185, Agilent, Little Falls, DE, USA)** and to an auxiliary electronic pressure controller (EPC) consisted of a one channel Pneumatics Control Module (G2317A, Agilent, Little Falls, DE, USA).
The **restrictor capillary was of 0.17 m x 0.1 mm d_c**.



Columns and capillaries were from Mega (Legnano, Milan, Italy). **Carrier gas**: helium delivered at constant flow; **initial head pressure p_i 296.0 KPa**; **auxiliary gas for MS outlet pressure correction (He) was delivered at 39.9 KPa** (relative). Split ratio (MS/FID) 50:50.
Connections between the primary and the two secondary columns were by a SilFlow™ GC 3 Port Splitter (SGE Ringwood, Victoria, Australia).
The system was equipped with a **two-stage KT 2004 loop thermal modulator** (Zoex Corporation, Houston, TX) cooled with liquid nitrogen controlled by Optimode™ V.2 (SRA Instruments, Cernusco sul Naviglio, MI, Italy). **Hot jet pulse time was set at 350 ms, modulation time was 5 s** and cold-jet total flow progressively reduced with a linear function from 30% of Mass Flow Controller (MFC) at initial conditions to 5% at the end of the run. Loop dimensions 0.6 m of the ²Ds were wrapped in the metal slit of the loop-type modulator.

GC×GC analyses were run with a system configured as follows: a HT280T multipurpose sampler (HTA, Brescia, Italy) was integrated with an **Agilent 6890 GC unit coupled to an Agilent 5975C MS detector** (Agilent, Little Falls, DE, USA) operating in EI mode at 70 eV. The GC transfer line was set at 300°C. An *Auto Tune* option was used and the **scan range was set to m/z 50-350** with a **scanning rate of 12,500 amu/s** to obtain a spectra generation frequency of 25 Hz. The Flame Ionization Detector (FID) was operated as follows: base temperature 300°C, H_2 flow 40 mL/min; air flow 240 mL/min; make-up (N_2) 450 mL/min; sampling frequency 150 Hz. Oven temperature programme was as follows: 50°C (1 min) to 300°C (10 min) at 4.0°C/min.

Samples and Data Treatment

[Urine Samples]

The urine samples were gathered for the Italian Diabetes Exercise Study 2 (IDES_2), which is assessing the effect of a behavioral intervention strategy on the promotion and maintenance of physical activity in adults with type 2 diabetes. IDES_2 is a randomized clinical trial that monitors objective measurable changes in sedentary time and physical activity over a 3-year period after behavioral intervention as compared with usual care. The study also monitors physical fitness, modifiable cardiovascular risk factors (HbA1c, lipids, blood pressure, C-reactive protein), and health related quality of life. The samples analyzed by GCx2GC are for the first and fourth quartile of physical activity objectively measured at baseline.

[Derivatization protocol]

Urine samples from 78 individuals were submitted to a standard derivatization protocol oximation/silylation [3] consisting of the following steps: 200 µL of urine and a suitable volume of ISTD (gallic acid solution at 10 g/L) were diluted with methanol up to 1000 µL and carefully mixed (Whirlimixer vortex, Fisher Scientific, Loughborough, Leicestershire, UK) thus 30 µL of MOX were added to 200 µL of the resulting solution and incubated for 2 hours at 60°C. Next, 30 µL of MSTFA were added and the mixture incubated at 100 °C for 60 minutes. The resulting sample solution diluted in n-hexane was immediately analyzed in duplicate and/or stored at -80°C until analysis.
Derivatization reagents: O-methylhydroxylamine hydrochloride (MOX) and N-methyl-N-(trimethylsilyl)trifluoroacetamide (MSTFA), and HPLC-grade solvents: methanol, pyridine, n-hexane, and dichloromethane, were supplied by Sigma-Aldrich (Milan, Italy).
Calibration standards of pyruvic acid, lactic acid, malonic acid, succinic acid, malic acid, 2-ketoglutaric acid, hippuric acid, L-alanine, L-valine, glycine, L-threonine, L-tyrosine, creatinine, phenylalanine, xylitol, ribitol, glycerol, fructose, galactose, glucose, mannitol, and myo-inositol; and the internal standard (ISTD) 4-fluorophenylalanine were from Sigma-Aldrich (Milan, Italy). Calibration solutions for quantitative determination of relevant analytes were prepared as in a previous protocol [1] at 2 mg/L, 10 mg/L, 50 mg/L, and 100 mg/L. The ISTD for data normalization and quality control, 4-Fluorophenylalanine, was at 10 mg/L.

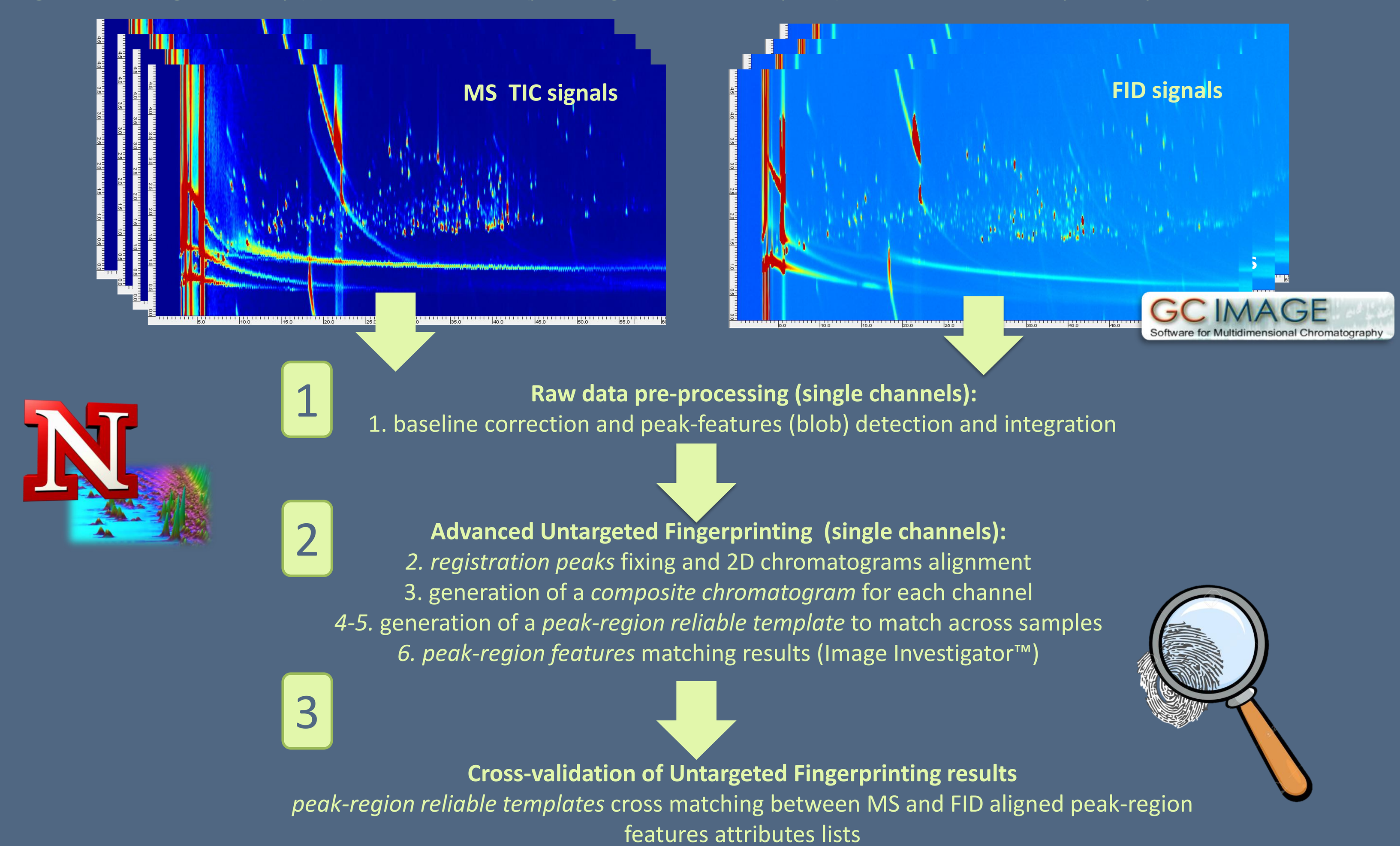
[Statistical Elaboration]

Multivariate analysis PCA (Principal Component Analysis) and OPLS-DA (Ortogonal Partial Least Square-Discriminant Analysis) was carried out by Pirouette® version 4.0 rev. 2 (Bothell, WA 98011, USA). Kruskal-Wallis test was performed with XLSTAT Version 2013.5.05 Addinsoft (PARIS, France).

Results and Discussion

Peak-region feature fingerprinting work-flow

This tool attempts to define one region (i.e., a small 2D retention-times window) per peak over the chromatographic plane to achieve the one-feature-to-one-analyte selectivity of peak features methods but with the implicit matching of region features.
Briefly this approach: (*1*) detects and records the peak patterns in individual chromatograms, (*2*) fixes a few peaks (named *registration peaks*) that can be reliably matched across samples, (*3*) aligns and combines the sample chromatograms to create a *composite chromatogram,* (*4*) defines a pattern of *region features* from the peaks detected in the cumulative chromatogram.
Then, when a target chromatogram is analyzed, (*5*) the *registration peaks* are matched to target chromatogram pattern, the feature regions are aligned relatively to those peaks, and the characteristics of those features are computed to create a feature vector for the target chromatogram finally (*6*) the feature vector (*peak-region reliable template*) is used for cross-sample analysis.

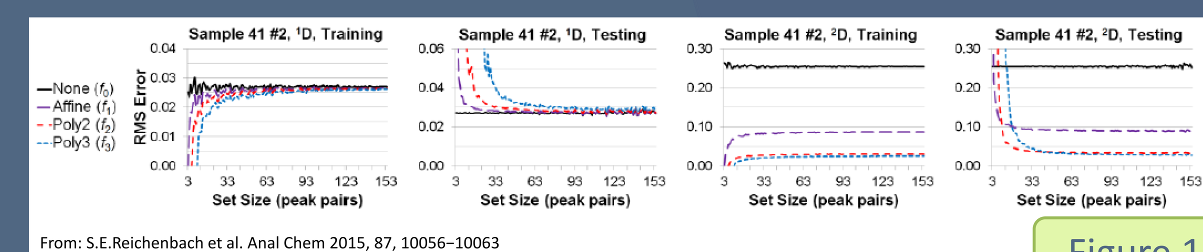


Alignment for GC×2GC-MS/FID

In order to cross-validate untargeted MS and FID fingerprinting and/or chromatographic features and data alignment is required because of the differences between the two dimensional (2D) retention-times pattern of the two detectors in particular due to system without auxiliary flow/pressure control at the outlet of the secondary column for the MS. The coherence of 2D patterns produced by FID/MS parallel detection can be managed by **Global** or **Local alignment methods**, i.e., whether the geometric differences between chromatograms are characterized by a single function for the entire chromatogram or by a combination of many functions for different regions of the chromatogram [2].

Global functions may be able to capture systemic properties and structure that underlie retention-time differences and when modeled with many parameters have unlimited representational power and are computationally simple. On the other hand, **Local functions** may be able to capture retention-time variations that are not related to systemic properties and structure. Typically, Local functions offer greater representational power than simple global functions, which allows them to capture small-scale variations, but also are more susceptible to overfitting of confounding input differences (e.g., compositional differences, artifacts, and noise) and so may be less robust than global functions. Performance of the models are quantified by the root-mean square error (RMSE) for retention times of matched peaks in paired FID/MS chromatograms from a single GCx2GC run.

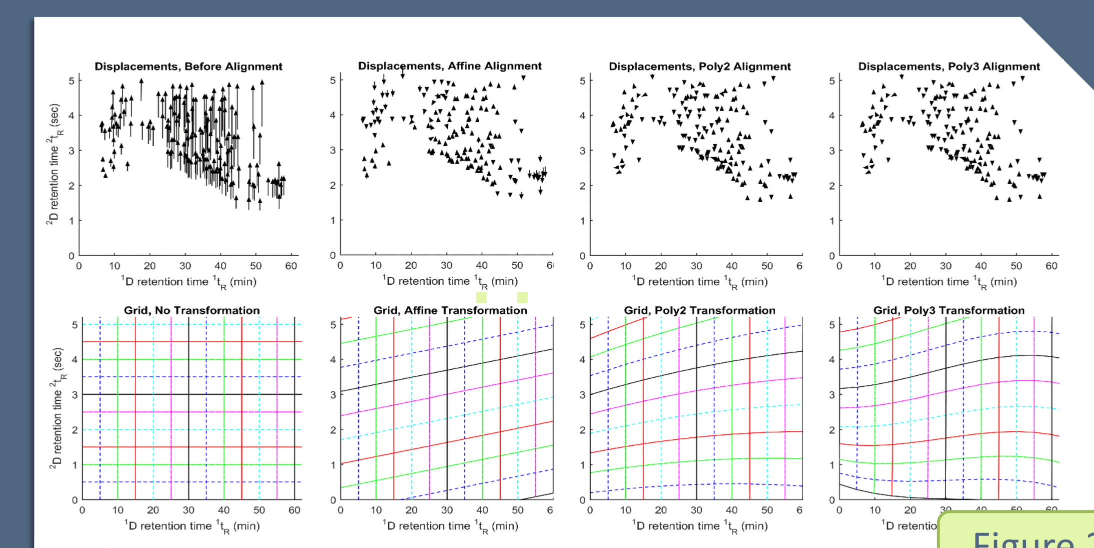
The performance of the transformation models for this task is compared to a benchmark computed as the RMSE between matched peaks for the same detector (i.e., FID with FID and MS with MS) in consecutive replicate runs.



From: S.E.Reichenbach et al. Anal Chem 2015, 87, 10056–10063

Figure 1

With the expectation that consecutive replicate runs exhibit only negligible sample and chromatographic differences (which are regarded as noise), any transformation model for them should provide only negligible reductions in RMSE (**Figure1**).



From: S.E.Reichenbach et al. Anal Chem 2015, 87, 10056–10063

Figure 2

If the performance of a transformation model to align FID and MS chromatogram pairs can approach the benchmark of consecutive replicate runs, then the residual differences can be attributed to noise rather n chromatographic misalignment.
Misalignment vectors (from FID to MS) for 156 peak pairs in chromatograms from GCx2GC analysis of a urine sample. Transformation of grid lines (FID to MS mapping for GCx2GC analysis of a urine sample). Columns from left to right are for the four alignment transformations: none (f_0), affine (f_1), second-degree polynomial (f_2), and third-degree polynomial (f_3) **Figure 2**.

Untargeted fingerprinting results

The **accuracy** of fingerprinting based on data from the two detection channels was confirmed by comparing FDRs results **Figures 3A-B**, bubble dimensions corresponds to FDR. Although these detectors show different analyte response factors and are characterized by different dynamic ranges, results were coherent as the ranking of most informative features. Cross-matching between detectors covers up to 70% of features.

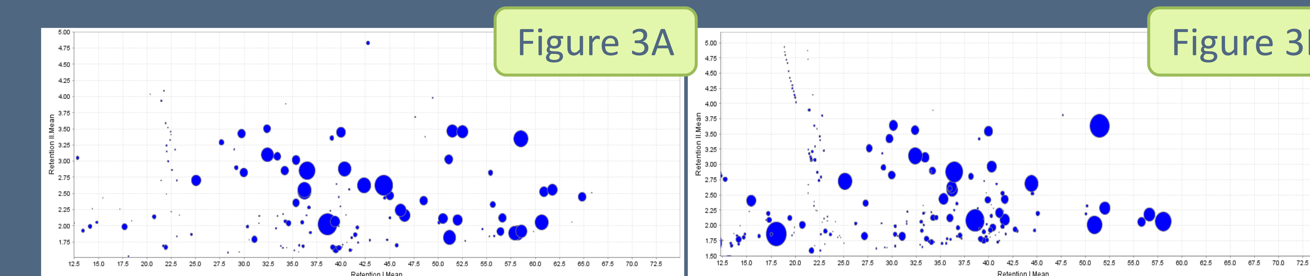


Figure 3A

Figure 3B

Untargeted fingerprinting cross-validation

Multivariate data analysis showed that the effect of the treatment (modest physical activity *vs* sedentary) is gender dependent.

Two-by-two **OPLS-DA** on **MS features** through **VIP** (Variable Influence on Projection) allowed to highlight the most relevant discriminant features of the treatment within the gender **Figures 4A-B.**

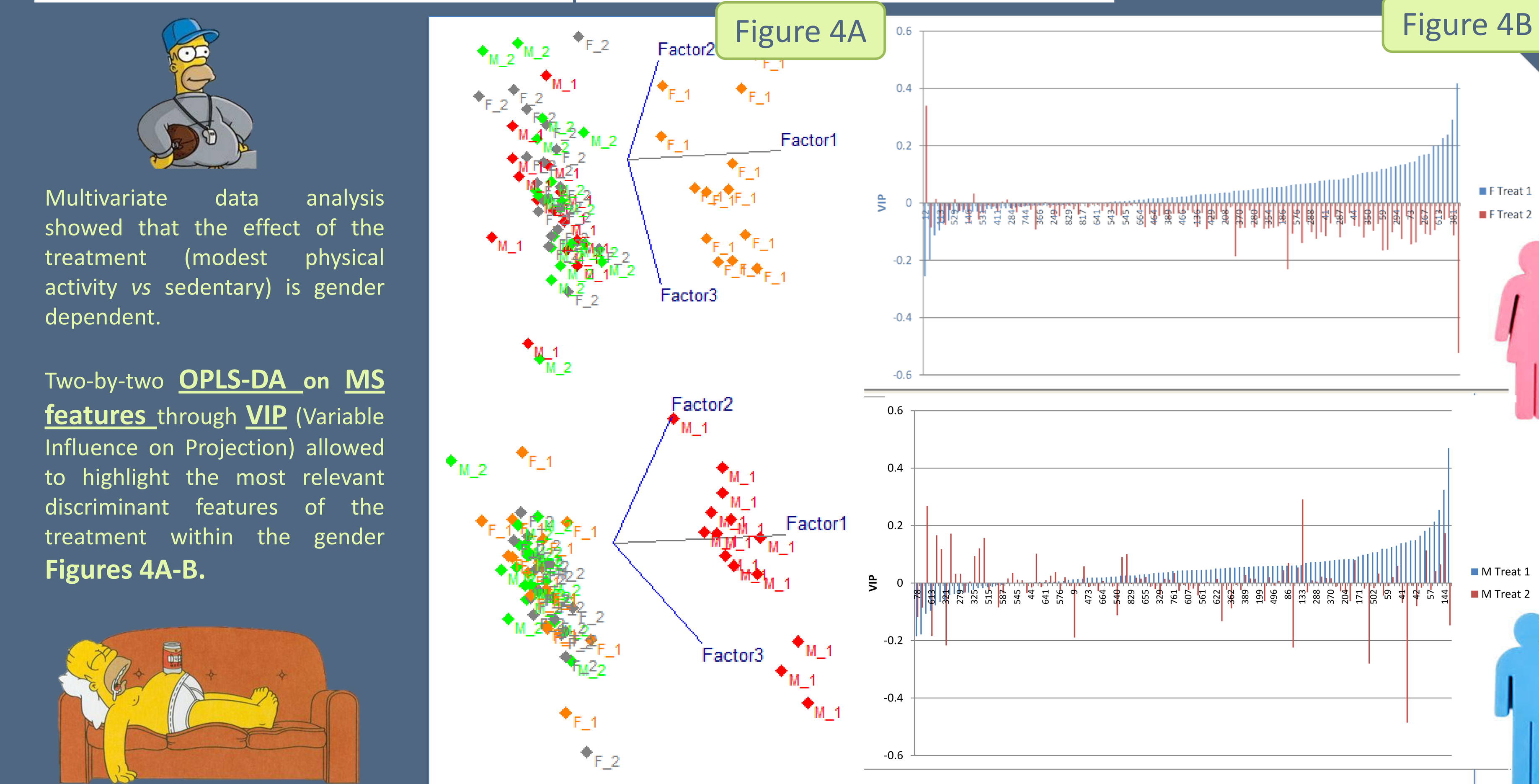


Figure 4A

Figure 4B

Targeted Quantitative fingerprinting

	Discriminant Urine metabolites	Trend	% of variation	P-value*
Females	Valine	↑↑	79	0.010
	Leucine	↑	58	0.030
	Glycine	↑↑	360	0.005
	Succinic acid	↑	68	0.039
	2-Ketoglutaric acid	↑↑	109	0.008
	L-Threonine	↑↑	83	0.018
	Xylitol	↑	41	0.023
	Rybitol	↑	37	0.008
	L-Alanine	↑↑	135	0.033
	Glycerol	↑	28	0.876
	L-phenylalanine	↑	61	0.014
Males	L-phenylalanine	↓	15	0.551
	Xylitol	↓	36	0.408
	Mannitol	↑	17	0.947
	L-Alanine	↓	29	0.085
	Glycerol	↓	43	0.354
	L-Threonine	↓	26	0.337
	Creatinine	↑	19	0.921
	Malonic acid	↑	-	0.040

*α=0.05

Analytes submitted to quantitative metabolomics were selected from untargeted fingerprinting results and from a careful survey of the scientific literature within those metabolic markers with relevant information potential.
PLS-DA carried out on FID quantitative data confirms untargeted results on the **gender's influence** on the evaluation of the treatment. At the same time PLS-DA emphasizes those **marker metabolites correlated to the treatment 1 (modest physical activities)** within the gender **Table 1**. The % of variation refers to the median analyte's relative increment/decrement of the treatment 1 *versus* treatment 2.
Females appear more responder to the modest physical activity than males both in terms of qualitative and quantitative variations. Some discriminant analytes may indeed related to an increasing of the catabolic metabolism activated by the modest physical activity (e.g Alanine to Glucose metabolism; a-ketogulatic acid to lipidic metabolism etc…)[4]. A non parametric Kruskall-Wallis test and the associate p-Values highlight the significance of these results. Some common most discriminant analytes like Phenylalanine, Threonine, Xylitol, Glycerol change with treatment 1 but they behave in opposite way between genders.

Conclusions

Parallel dual secondary column-dual detection GC × GC system exploits greater information capacity and the complementarity of two detectors and offers a powerful tool to delineate the human urine metabolic signatures. The consistencies of the joint information offered by the two chromatograms requires simple and robust data fusion provided by a global low-order polynomial transformation. Untargeted fingerprinting results shows several discriminant features related to the physical activity, mostly influenced by the gender, and some of them are also in agreement with the targeted results. In addition, even if the treatments under study was very fine it support the evidence that other small biomarker can be related to the metabolic derangement in diabetes type 2 individuals and to the positive effect of the physical activities. Finally Targeted quantitative fingerprinting underline females as more responder to the exercise.