J-resolved ^1H NMR 1D-projections for large-scale metabolic phenotyping studies: Application to blood plasma analysis

Andrea Rodriguez-Martinez[†], Joram M. Posma[†], Rafael Ayala[†], Nikita Harvey[†], Beatriz Jimenez[†], Ana L. Neves[†], John C. Lindon[†], Kazuhiro Sonomura^{‡§}, Taka-Aki Sato^{‡§}, Fumihiko Matsuda[§], Pierre Zalloua^{||}, Dominique Gauguier^{†θ}, Jeremy K. Nicholson^{†*}, Marc-Emmanuel Dumas^{†*}

[†] Computational and Systems Medicine, Department of Surgery and Cancer, Imperial College London, London, UK.

[‡] Life Science Research Center, Technology Research Laboratory, Shimadzu Corporation, Kyoto, Japan.

[§] Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan.

^{||} School of Medicine, Lebanese American University, Beirut, Lebanon.

^θ Cordeliers Research Centre, INSERM UMR_S 1138, Paris, France.

ABSTRACT: ^1H nuclear magnetic resonance (NMR) spectroscopy-based metabolic phenotyping is now widely used for large-scale epidemiological applications. In order to minimize signal overlap present in 1D ^1H NMR spectra, we have investigated the use of 2D J -resolved (JRES) ^1H NMR spectroscopy for large-scale phenotyping studies. In particular, we have evaluated the use of the 1D projections of the 2D JRES spectra (pJRES), which provide single peaks for each of the J -coupled multiplets, using 705 human plasma samples from the FGENTCARD cohort. Based on the assessment of several objective analytical criteria (spectral dispersion, attenuation of macromolecular signals, cross-spectral correlation with GC-MS metabolites, analytical reproducibility and biomarker discovery potential), we concluded that the pJRES approach exhibits suitable properties for implementation in large-scale molecular epidemiology workflows.

In the last two decades, metabolic phenotyping (metabotyping¹) has become a versatile approach for studying metabolism, with a broad range of applications in the fields of clinical biochemistry, toxicology and drug metabolism²⁻⁴. The main analytical techniques used for metabolic profiling are proton nuclear magnetic resonance (^1H NMR) spectroscopy and mass spectrometry (MS) coupled to liquid chromatography (LC-MS) or gas chromatography (GC-MS). NMR spectroscopy is a quantitative, non-destructive and highly reproducible technique that provides detailed information of molecular structures based on atom-centered nuclear interactions and properties^{5,6}. These qualities have made ^1H NMR spectroscopy one of the key analytical strategies for metabotyping in large-scale epidemiological studies^{2,7-9}.

The vast majority of ^1H NMR-based metabolic phenotyping studies employ one-dimensional (1D) experiments, as they allow relatively rapid spectral acquisition and therefore maximize the throughput. However, 1D NMR spectroscopy leads to complex mixture spectra, with considerable peak overlap, thus limiting the number of metabolites that can be unambiguously identified and quantified¹⁰. One possibility to reduce the burden of spectral overlap and increase metabolite specificity is to resolve the multiplet resonances into a second dimension (*i.e.* 2D-NMR). J -resolved (JRES) ^1H NMR spectroscopy¹¹ is a very useful 2D ^1H NMR method for structural assignments because it efficiently separates chemical shift (δ) in the F2 dimension from J -coupling in the F1 dimension, and data acquisition times are amenable to be comparable to those used for other 1D ^1H NMR experiments¹².

The JRES experiment is composed of an array of spin-echo pulses, where an incremented delay period is applied to define the second dimension. In a JRES spectrum, both chemical shifts and homonuclear J -couplings appear in F2 and only homonuclear couplings appear in F1 (heteronuclear coupled multiplets behave like chemical shifts). The J -couplings and chemical shifts are resolved into two orthogonal axes (F1 and F2, respectively) by tilting the spectrum through 45° and also by symmetrizing the display to choose the lowest intensity of the two data points equidistant from the F1=0 axis¹³. The resonance intensities are edited according to proton T_2 relaxation times¹⁴. The sum or skyline projection of a 2D JRES spectrum along the chemical shift axis effectively yields a virtual ^1H broadband proton-decoupled spectrum (pJRES) with interesting properties in terms of signal decomposition: i) there are fewer resonances in the pJRES spectrum than in a classical 1D spectrum for the same number of metabolites; and ii) multiplets have a smaller footprint along the chemical shift axis. Additional advantages of the JRES experiment over several other homonuclear decoupling (*i.e.* “pure shift”) methods^{15,16} are the efficient attenuation of broad resonances from macromolecules and the provision of J -coupling data, which can aid metabolite assignment^{14,17}. JRES spectroscopy is particularly suitable for metabotyping because the majority of endogenous metabolites of interest exhibit first-order spin systems, and consequently artifactual peaks associated with second-order spin systems are rare, especially at high magnetic fields¹⁴. In the last three decades, the advantages of JRES-based NMR spectroscopy have been reflected in several studies demonstrating the potential of the JRES approach for

Figure 1. Overview of the 3 typical NMR experiments for plasma analysis, using a sample from the FGENTCARD cohort. (A) Standard 1D spectrum, (B) spin-echo spectrum and (C) skyline projection of the JRES spectrum with the corresponding 2D contour plot right below. Abbreviations: L1-L6, specific groups of the lipoprotein envelope; U, unidentified signal; NAC, refers to residual composite acetyl signals from α 1-acid glycoprotein; TMAO, trimethylamine *N*-oxide; α G and β G, refer to signals from the two glucose anomers; and A, strong coupling artifacts.

metabolic fingerprinting and metabolite identification^{14,17-22}. However, the JRES experiment has limitations for metabolite quantification caused by non-standard line shapes and T_2 editing²³.

In contemporary NMR-based metabolotyping studies of plasma or serum, a short JRES experiment is routinely acquired, together with other 1D experiments, such as the standard 1D water-presaturated and spin-echo experiments^{10,12}. The classical 1D water-presaturated spectrum provides an excellent unbiased overview of all plasma molecules, but it is dominated by broad lipoprotein signals and therefore it is unrevealing with respect of most low molecular weight metabolites¹⁰. The spin-echo and JRES experiments are more informative for the study of plasma metabolites since resonances from macromolecules are filtered *via* T_2 editing. However, despite its potential, the JRES experiment tends to be used only as a complementary tool to facilitate the assignment of metabolites detected by the spin-echo experiment.

Here, we have used 705 human plasma samples from the FGENTCARD cohort²⁴ to evaluate the use of pJRES-based ^1H NMR metabolotyping for large-scale epidemiological studies, and to compare it with typical 1D ^1H NMR experiments for plasma analysis: a standard 1D experiment with water presaturation and a spin-echo experiment. Based on the assessment of several objective criteria, including: i) spectral dispersion; ii) attenuation of residual lipid resonances; iii) metabolite specificity; iv) analytical reproducibility; and v) discovery potential in a metabolome-wide association study (MWAS) setting², we show that the pJRES approach presents suitable properties for implementation in data analysis workflows for large-scale epidemiology projects.

MATERIALS AND METHODS

Patients and samples. A major goal of the FGENTCARD project was to discover biomarkers associated with coronary artery disease and metabolic risk factors, such as obesity. The

FGENTCARD cohort consists of over 7,000 patients who underwent catheterization between 2006 and 2009 at two major university hospitals in Lebanon²⁴. Plasma samples from patients who provided consent for the whole study were collected and stored at -80°C until analysis. In the current study a subset of 705 plasma samples was used. All patient medical charts were reviewed for data collection and demographics, medical history, and pharmacological interventions.

^1H NMR data acquisition. Thawed human plasma samples ($n = 705$) were centrifuged (14,000 rpm, 10 min, 4°C) and the supernatants were allocated in 96-well plates, in a randomized design. A 5 μL aliquot of each sample was pooled to prepare the quality control (QC) sample and two aliquots of the QC sample were added (positions 48 and 96) to each 96-well plate. Plates were stored at -80°C until profiling. 100 μL of sample were thoroughly mixed with 100 μL of plasma buffer (75 mM Na_2HPO_4 , pH 7.4, 4.6 mM 3-(trimethylsilyl)-[2,2,3,3- $^2\text{H}_4$]-propionate TSP, 6.2 mM sodium azide (NaN_3), as a bacteriostatic agent, and 20% (v/v) of D_2O)¹² in a different plate. 180 μL of the resulting solution were subsequently transferred into 3 mm NMR tubes using a Gilson liquid handling robot. NMR spectra were acquired on a Bruker Avance III 600 MHz spectrometer (Bruker Biospin Ltd, Germany) operating at 600 MHz and 310 K. For each sample, standard 1D, spin-echo and JRES experiments were performed as previously described¹². The standard 1D experiment was acquired using the pulse sequence: $-\text{RD}-g_{z,1}-90^\circ-t-90^\circ-\text{tm}-g_{z,2}-90^\circ-\text{ACQ}$, where RD is the relaxation delay (4 s), t is a short delay (3 μs), 90° represents the 90° radio frequency (RF) pulse, tm is the mixing time (10 ms), $g_{z,1}$ and $g_{z,2}$ are magnetic field z -gradients both applied for 1 ms, and ACQ is the data acquisition period (2.7 s). For each sample, 64 transients were collected into 96,000 data points using a spectral window of 30 ppm. Continuous wave irradiation at the water resonance frequency using 25 Hz RF pulses was applied during the RD and also during the tm . The spin-echo experiment was acquired with the pulse sequence: $-\text{RD}-90^\circ-(t-180^\circ-t)_n-\text{ACQ}$, where t is the spin-echo delay (0.3

Figure 2. Comparison of spectral dispersion of the JRES experiment and 1D experiments. Contour plot of selected regions of a 2D JRES spectrum containing metabolic signals of various multiplicities, and right above the corresponding skyline projection together with the spin-echo spectrum. Abbreviations: TMAO refers to trimethylamine-*N*-oxide.

Table 1. Signal-to-noise (SN) ratios, line width at half height (LW) and multiplet footprints (MF) of metabolic peaks with different multiplicities in the pJRES spectra and in the spin-echo spectra.

Metabolite	$\delta^1\text{H}$	pJRES		Spin-echo		
		SN	LW (Hz)	SN	LW (Hz)	MF (Hz)
Acetate	1.91 (s)	252.77 $\pm$ 62.27	3.69 $\pm$ 0.06	160.83 $\pm$ 23.27	1.24 $\pm$ 0.22	-
Valine	0.98 (d)	270.17 $\pm$ 62.03	3.51 $\pm$ 0.11	122.51 $\pm$ 13.04	1.38 $\pm$ 0.19	8.57 $\pm$ 0.23
Glucose	3.25 (dd)	753.84 $\pm$ 172.98	3.87 $\pm$ 0.11	394.72 $\pm$ 35.99	1.60 $\pm$ 0.22	19.31 $\pm$ 0.30
Lactate	4.11 (q)	197.91 $\pm$ 46.55	3.44 $\pm$ 0.10	125.126 $\pm$ 15.94	1.34 $\pm$ 0.18	21.68 $\pm$ 0.24

These parameters were calculated across the QC samples ($n = 14$) and reported as mean $\pm$ standard deviation.

ms), 90° represents the 90° RF pulse while 180° is the 180° RF pulse, and n represents the number of loops (256). For each sample, 64 transients were collected into 72,000 data points using a spectral window of 20 ppm. The sequence used for the 2D JRES experiment takes the form: -RD- 90° - t_1 - 180° - t_1 -ACQ, where the definitions are as above, plus t_1 represents the increment delay. JRES spectra were acquired using 4 scans per increment over 40 increments that were collected in 8,000 data points, using spectral windows of 16.6 ppm in F2 and 78 Hz in F1. The total data acquisition time each of the 3 NMR experiments was 8 min. The raw NMR files are accessible from MetaboLights²⁵ (study identifier: MTBLS540).

^1H NMR data preprocessing. Following spectral acquisition, the data were automatically processed using TopSpin 3.2 with Icon (Bruker Biospin Ltd, Germany). For the 1D experiments, an exponential line broadening of 0.3 Hz and a zero-filling by a factor of 2 were applied prior to Fourier transformation. For the JRES experiment, zero-filling by a factor of 2 was included in the F2 dimension and the digital resolution was increased to 256 points in F1 by zero-filling. Apodization of JRES data using a sine-bell function was applied in both dimensions prior to Fourier transformation. The JRES spectra were then titled, symmetrized and skyline- or sum-projected to obtain the pJRES spectra. Following phase and baseline corrections, all spectra (δ 0.4 – 10) were imported into MATLAB 8.3 (Mathworks, USA) using in-house routines. The spectra were calibrated to the α -glucose anomeric signal at δ 5.23²⁶ and aligned manually²⁷. Spectral regions containing signals from water and EDTA (ethylenediamine tetraacetic acid) were

removed. Metabolite assignment was performed using an in-house database^{10,28,29}, correlation-based analysis³⁰⁻³² and spiking experiments with authentic commercial standards.

Signal-to-noise ratio and line width analyses. Signal-to-noise ratios (SN) of selected ^1H NMR signals were calculated as the ratio of peak intensity at maximum height to the standard deviation of the noise (δ 9.88 – 9.94). Line widths (LW) were calculated as the peak width at half height. For the spin-echo experiment, multiplet footprints (MF) were calculated as the line width from half-height of the most up-field peak of the multiplet to the most down-field peak of the multiplet. SN, LW and MF were calculated across the QC samples ($n = 14$) and reported as mean $\pm$ standard deviation.

GC-MS data. Thawed plasma samples ($n = 36$) were centrifuged (14,000 rpm, 10 min, 4°) and 50 μL of the supernatant was added to a mix (3:125) of the internal standard solution (2-isopropylmalic acid, 0.1mg/mL) and extraction solvent (methanol:water:chloroform 2.5:1:1). The resulting solution was mixed with a shaker (1,200 rpm, 30 min, 37°) and centrifuged (14,000 rpm, 10 min, 4°), followed by lyophilization of 180 μL of the supernatant. The lyophilized sample was dissolved in 80 μL of methoxyamine solution (20 mg/mL in pyridine) and agitated (1,200 rpm, 30 min, 37°). 40 μL of *N*-methyl-*N*-trimethylsilyltrifluoroacetamide solution were added for trimethylsilyl derivatization, followed by agitation (1,200 rpm, 30 min, 37°) and centrifugation (14,000 rpm, 5 min). Finally, 1 μL derivative was injected into a GCMS-QP2010 Ultra (Shimadzu, Kyoto, Japan). A DB-5 column

Figure 3. Assessment of efficiency of attenuation of lipid signals in JRES spectroscopy through correlation analysis. STOCY plots driven by metabolic signals that overlap with lipid resonances in the 1D NMR experiments (acetone: δ 2.22 in (A-C) and mannose: δ 5.18 in (D-F)) ($n = 705$). Only NMR variables significantly correlated ($p\text{FDR} < 0.05$) with the driver signal are displayed with color matching their correlation. Abbreviations: L1-L10 indicate specific groups of mobile lipids and lipoproteins (L1: CH_3 , L2: $(\text{CH}_2)_n$, L3: $(\text{CH}_2)(\text{CH}_2)\text{CO}$, L4: CCH_2 , L5: CH_2CO , L6: $\text{C}-\text{CH}_2-\text{C}=\text{}$, L7: CH_2OCOR , L8: CH_2OCOR , L9: CHOCOR , L10: $\text{CH}=\text{CHCH}_2\text{CH}=\text{CH}$).

(Aligent Technologies, Palo Alto, CA) was used for metabolite separation. The ion source temperature was set to 200° , and the interface temperature was set to 250° . Data acquisition was carried out in full scan mode in the range of m/z 85-500. GC-MS data processing was performed with GCMSSolution 2.71 (Shimadzu, Kyoto, Japan). Identification of chromatographic peaks was performed using the NIST library or Shimadzu GC/MS database, and further confirmed with authentic commercial standards. The area of each of the confirmed metabolites ($n = 100$) was normalized to the area of the internal standard.

Statistical analyses. All statistical analyses were performed with the MWAStools R package³³. The analytical reproducibility of the NMR run was assessed *via*: i) principal component analysis (PCA) on mean-centered data; and ii) coefficient of variation (CV) (ratio of standard deviation to the mean) of the spectral variables across the QC samples⁸. Statistical total correlation spectroscopy (STOCY)³¹ was used to evaluate the confounding effect of lipid resonances on specific signals from small metabolites. Partial Spearman correlations of body mass index (BMI) and the NMR variables were performed to identify metabolites associated with obesity. The correlations were adjusted for three potential confounding risk factors: age, gender and type-2 diabetes. The p -values were corrected for multiple-testing using the Benjamini and Hochberg (BH) false discovery rate (FDR)³⁴. Significance of spectral variables was determined based on a $p\text{FDR}$ threshold of 0.05. The results of the correlations were represented in Manhattan plots showing the $-\log_{10}$ ($p\text{FDR}$) adjusted for the sign of the correlation of each spectral variable⁹. In order to avoid spurious findings, only NMR peaks with at least 3 adjacent spectral variables passing the significance threshold were retained. For metabolites with several significant NMR peaks, the most specific

peak in terms of intramolecular correlations with NMR and GC-MS data was retained and the minimum $p\text{FDR}$ value of the peak was reported. The 95 % confidence intervals (CI)

Figure 4. Evaluation of specificity of metabolic signals in pJRES spectroscopy *via* cross-correlation with GC-MS metabolites. (A) Kernel density curves showing the distribution of coefficients of Spearman correlations between 32 GC-MS metabolites and matched NMR features in pJRES and spin-echo datasets ($n = 36$). (B) Heat-map showing the individual coefficients of correlation between the 32 GC-MS metabolites and matched NMR features in the pJRES and spin-echo experiments.

Figure 5. Assessment of analytical repeatability of pJRES spectroscopy. (A-C) PCA score plots of mean-centred data with the QC samples ($n = 14$) colored in red and the experimental samples ($n = 705$) in blue. (D-F) pJRES, spin-echo, and standard 1D spectra of a QC sample colored according to the coefficient of variation of each NMR feature. NMR signals with coefficient of variation equal or above 0.30 were colored in red.

of the correlation coefficients were calculated *via* bootstrapping ($n = 1000$ iterations).

RESULTS

JRES ^1H NMR spectroscopy of human plasma. Typical examples of standard 1D, spin-echo and JRES ^1H NMR spectra are shown in Figure 1. The standard 1D spectrum is not very informative in terms of small metabolites because of the extensive overlap with broad resonances from macromolecules, such as proteins (particularly albumin) and lipids. Application of spin-echo pulse sequences in the 1D spin-echo experiment and in the JRES experiment, results in the attenuation of spectral contributions from macromolecules, which have short relaxation times. In the spin-echo spectrum metabolic signals are heavily overlapped in crowded chemical shift ranges (*e.g.* δ 3-4) and therefore cannot be easily resolved. The projection of the JRES spectrum along the chemical shift axis yields a greatly simplified virtual proton-decoupled 1D ^1H spectrum (pJRES), which is composed only of singlets. This collapse of multiplets into singlets considerably reduces peak overlap in congested spectral regions compared with the 1D-spectra (Figure 2). Importantly, the pJRES spectra preserves high intramolecular correlations regardless of peak multiplicity, even for second-order spin systems (Figure S1). Two projection modes are typically employed to obtain a pJRES spectrum: sum-projection (*i.e.* summing all intensities along F1) and skyline-projection (*i.e.* selecting the highest intensity along F1). Here we have used the skyline projection as it leads to higher SN, while conserving high intramolecular correla-

tions, analytical reproducibility and cross-spectral correlations with GC-MS metabolites (Figure S2, Table S1).

Table 1 compares the SN, LW and MF of selected metabolic peaks with different multiplicities, in the pJRES spectra and in the spin-echo spectra. These parameters were not calculated for the standard 1D spectra because the metabolic signals are confounded with the macromolecule baseline. The SN of all selected metabolites was higher in the pJRES spectra than in the spin-echo experiment. In terms of LWs, the pJRES spectra exhibited broader singlet peaks than the spin-echo spectra; however this disadvantage is compensated by the reduction in MF in the pJRES spectra. The resolution of the pJRES spectra could be further improved by increasing the data acquisition period (Figure S3). This first set of results highlights the major advantage of pJRES spectroscopy, which is the increased peak dispersion compared to the 1D NMR experiments.

Assessment of efficiency of attenuation of lipid resonances in the JRES experiment. Another major advantage of JRES spectroscopy is the nearly complete attenuation of broad resonances from macromolecules, including proteins, lipids and lipoproteins, which dominate the standard 1D spectrum (Figure 1). In the spin-echo experiment, protein ^1H resonances are mostly absent, but strong lipid resonances are still detected. In order to show how the presence of the residual lipid baseline in the spin-echo spectra may obscure specific metabolic signals, we performed STOCSY-based correlation analysis³¹ driven by two metabolic peaks that overlap with lipid resonances: acetone (δ 2.22) and mannose (δ 5.18). Figure 3 shows the STOCSY-plots corresponding to these data, which display the covariance (peak height) and the Pearson

Figure 6. Visualization of the associations between BMI and NMR metabolites in the pJRES and spin-echo datasets. (A, B) Manhattan plots showing the $-\log_{10}(\text{pFDR}) \times \text{sign}$ of correlation of each NMR feature with BMI in the pJRES dataset (A) and in the spin-echo dataset (B) ($n = 705$). Statistically significant NMR signals were colored in red if positively associated with BMI and in blue if negatively associated. Right below each Manhattan plot, the NMR spectrum of a QC sample colored based on association results is shown. (C) Bigraph showing the metabolites significantly associated with BMI detected in each NMR dataset, based on the criteria of at least 3 adjacent spectral variables with $\text{pFDR} < 0.05$.

coefficient of correlation (color) of each spectral variable with the driver signal (*i.e.* acetone or mannose). The pJRES STOCYSY plots driven by acetone (Figure 3A) and mannose (Figure 3D) signals successfully highlighted resonances from biologically related metabolites (3-hydroxybutyrate for acetone and glucose for mannose). In contrast, the spin-echo (Figure 3B, E) and standard 1D (Figure 3C, F) STOCYSY-plots driven by the same NMR signals highlighted spectral regions corresponding to lipid resonances. These results confirm that the presence of a residual lipid baseline in the spin-echo spectra confounds specific metabolic resonances, negatively impacting the reliability of these metabolic signals compared to the pJRES experiment. Altogether, we have shown that lipid attenuation in pJRES leads to cleaner correlation patterns for metabolite structural assignment purposes.

Evaluation of specificity of metabolic signals in pJRES spectroscopy through cross-correlation with GC-MS data. The improved peak dispersion and more efficient attenuation of lipid signals in the JRES experiment should be, in principle, accompanied by an increase in metabolite specificity. To confirm this, we performed correlation-based analysis of NMR (pJRES and spin-echo) data and GC-MS data (Figure 4)^{32,35}. We computed Spearman correlations between all NMR features and 100 GC-MS detected metabolites. For each GC-MS metabolite, we selected the NMR feature with the highest correlation within the range of expected chemical shifts. Based on a correlation cut-off of $r = 0.60$ in at least one of the NMR experiments, 32 GC-MS metabolites were matched with NMR

features. As expected, the paired GCMS-NMR correlations were significantly higher (Wilcoxon signed-rank p -value = 1.09×10^{-4}) in the pJRES dataset compared to the spin-echo dataset (Figure 4A), with remarkable differences for several metabolites such as 2-aminobutyrate ($r_{\text{JRES}} = 0.76$, $r_{\text{spin-echo}} = 0.50$) or 2-hydroxybutyrate ($r_{\text{JRES}} = 0.96$, $r_{\text{spin-echo}} = 0.28$) (Figure 4B). These results highlight the enhanced specificity of metabolic signals in the pJRES spectra, compared with the spin-echo experiment.

Assessment of analytical repeatability of pJRES spectroscopy. Untargeted NMR metabolotyping in epidemiological studies is based on relative comparison of spectral profiles and therefore a QC strategy is required to ensure the repeatability of analytical workflows. We assessed the analytical reproducibility of the JRES experiment, in comparison with the standard 1D and spin-echo experiments, *via* PCA and CV analysis across the QC samples ($n = 14$), prepared from a representative pool of the experimental samples and analyzed regularly throughout the run. In the PCA score plots obtained for each NMR experiment, the QC samples appeared clustered in the center of the Hotelling's T^2 ellipse, indicating good analytical stability and reproducibility (Figure 5A-C). In terms of CV, in all three experiments, the majority of NMR signals corresponding to actual metabolites exhibited CV values below 0.30 (Figure 5D-F), which is the threshold established by the FDA (Food and Drug Administration) for biomarker discovery³⁶. Overall, these results indicate that the JRES experiment has comparable reproducibility to classical 1D experiments.

Table 2. Association results of metabolites significantly associated with BMI in the pJRES and/or spin-echo data sets (n = 705).

Metabolite	$\delta^1\text{H}$	r pJRES (95% CI)	pFDR pJRES	r spin-echo (95%CI)	pFDR spin-echo
Isoleucine	0.93 (t), 0.99 (d), 1.45 (m)	0.18 (0.10, 0.25)	7.06×10^{-4}	0.17 (0.09, 0.24)	1.30×10^{-3}
Leucine	0.95 (dd), 1.71 (m), 3.73 (m)	0.15 (0.07, 0.21)	6.70×10^{-3}	0.16 (0.08, 0.23)	2.04×10^{-3}
2-Aminobutyrate	0.97 (t)	0.15 (0.08, 0.21)	6.70×10^{-3}	-	-
Valine	0.98 (d), 1.03 (d), 2.26 (m)	0.22 (0.14, 0.28)	8.86×10^{-6}	0.22 (0.15, 0.29)	2.07×10^{-5}
3-Hydroxyisobutyrate	1.06 (d)	0.17 (0.08, 0.24)	1.03×10^{-3}	0.14 (0.07, 0.22)	7.19×10^{-3}
Alanine	1.47 (d), 3.77 (q)	0.15 (0.08, 0.22)	3.65×10^{-3}	0.14 (0.06, 0.21)	9.77×10^{-3}
N-acetylated compound	2.07 (s)	0.16 (0.08, 0.22)	2.38×10^{-3}	0.12 (0.04, 0.18)	4.40×10^{-2}
Citrate	2.65 (d)	0.13 (0.07, 0.21)	1.74×10^{-2}	0.17 (0.10, 0.25)	8.41×10^{-4}
Lactate	1.31 (d), 4.11 (q)	0.17 (0.09, 0.24)	1.03×10^{-3}	0.17 (0.10, 0.24)	1.02×10^{-3}
Mannose	5.18 (d)	0.14 (0.06, 0.21)	1.38×10^{-2}	0.12 (0.04, 0.19)	4.89×10^{-2}
4-Hydroxymandelate	7.29 (d)	-0.15 (-0.21, -0.08)	4.65×10^{-3}	0.03 (-0.04, 0.10)	8.06×10^{-1}
Tyrosine	6.89(d), 7.18 (d)	0.13 (0.06, 0.20)	2.81×10^{-2}	0.16 (0.08, 0.22)	2.53×10^{-3}

For 2-aminobutyrate (δ 0.97) only the pJRES results are shown, since in the spin-echo spectra this peak was unresolved.

Application of pJRES spectroscopy in a metabolome-wide association study. Finally, we evaluated the biomarker discovery potential of the pJRES approach in a MWAS setting. For each NMR experiment, we computed partial Spearman correlations between all NMR signals and BMI, adjusted for age, gender and type II diabetes. The *p*-values were adjusted for multiple-testing using BH correction and represented in a Manhattan plot (Figure 6A-B). The results obtained with the standard 1D experiment are not shown as only lactate was statistically significant ($\text{pFDR} = 4.88 \times 10^{-2}$). Considering the pJRES and spin-echo datasets together, we identified 12 metabolites significantly associated with BMI, 7 of which were detected in both datasets (Figure 6C, Table 2). Importantly, 4 out of the 12 associations were only detected in the pJRES dataset, which might be due to the increased peak dispersion, as observed for 2-aminobutyrate (Figure 2), as well as to the more efficient removal of lipid signals, as observed for the mannose (Figure 3). On the other hand, the association between tyrosine and BMI was only found in the spin-echo dataset, which might be due to the attenuation of signals from metabolites extensively protein-bound (e.g. aromatic amino acids) in JRES spectroscopy. However, it should be noted that associations between clinical phenotypes and protein-bound plasma metabolites are difficult to interpret, as the levels of these metabolites are dependent on protein concentrations. Collectively, these results indicate that the pJRES approach can complement and augment the recovery of metabolic information of typical 1D NMR experiments for plasma analysis.

CONCLUSION

^1H NMR spectroscopy profiling of biofluids is a key component of metabolically top-down systems biology, with great potential for the discovery of novel biomarkers of human disease. One of the main challenges of ^1H NMR metabolic profiling is due to the complex nature of biological samples, which leads to highly congested spectra. To address this challenge, for the first time, we have investigated the use of pJRES spectroscopy in a large-scale population setting, in comparison with typical 1D NMR experiments for plasma analysis. We have demonstrated that the pJRES approach exhibits: i) enhanced peak dispersion; ii) more efficient attenuation of lipid resonances; iii) superior metabolite identification capacities; and iv) comparable analytical reproducibility. Given its advantageous properties, pJRES spectroscopy has the potential to

complement and augment the recovery of metabolic information of 1D NMR experiments in the context of MWAS. Altogether, we conclude that pJRES spectroscopy offers a range of sought properties for implementation in large-scale analytical workflows, such as MWAS.

ASSOCIATED CONTENT

Supporting Information

Three additional figures and one additional table are given as indicated in the text.

AUTHOR INFORMATION

Corresponding Author

* Email: m.dumas@imperial.ac.uk;
* Email: j.nicholson@imperial.ac.uk

Author Contributions

ARM, NH, BJ, ALN, KS, TS run experiments. ARM analyzed the data, with input from JMP, RA, FM, JCL, JKN, MED. PZ, DG, JKN and MED designed the study. ARM and MED wrote the paper and all authors gave approval to the final version.

Conflict of interest

KS and TAS are employees of Shimadzu, Kyoto, Japan.

ACKNOWLEDGMENT

This work was supported by PhD studentships awarded to ARM (Medical Research Council Doctoral Training Centre (MR/K501281/1), Imperial College (EP/M506345/1) and La Caixa Foundation), by grants from the European Commission (FGENTCARD, LSHG-CT-2006-037683, EURATRANS, HEALTH-F4-2010-241504) to DG, JKN and MED and by the Practical Research Project for Rare/Intractable Diseases from the Japanese Agency for Medical Research and Development (AMED) to FM. We thank the Imperial-National Institute for Health Research (NIHR) Clinical Phenome Centre, which is supported by the NIHR Imperial Biomedical Research Centre based at Imperial College Healthcare National Health Service (NHS) Trust and Imperial College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health.

REFERENCES

(1) Gavaghan McKee, C. L.; Wilson, I. D.; Nicholson, J. K. J. *Proteome Res.* 2006, 5, 378-384.

- (2) Holmes, E.; Loo, R. L.; Stamler, J.; Bictash, M.; Yap, I. K.; Chan, Q.; Ebbels, T.; De Iorio, M.; Brown, I. J.; Veselkov, K. A.; Daviglus, M. L.; Kesteloot, H.; Ueshima, H.; Zhao, L.; Nicholson, J. K.; Elliott, P. *Nature* 2008, 453, 396-400.
- (3) Nicholson, J. K.; Lindon, J. C.; Holmes, E. *Xenobiotica* 1999, 29, 1181-1189.
- (4) Nicholson, J. K.; Lindon, J. C. *Nature* 2008, 455, 1054-1056.
- (5) Beckonert, O.; Keun, H. C.; Ebbels, T. M.; Bundy, J.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. *Nat. Protoc.* 2007, 2, 2692-2703.
- (6) Lindon, J. C.; Nicholson, J. K. *Annu. Rev. Anal. Chem. (Palo Alto Calif.)* 2008, 1, 45-69.
- (7) Gavaghan, C. L.; Wilson, I. D.; Nicholson, J. K. *FEBS Lett.* 2002, 530, 191-196.
- (8) Dumas, M. E.; Maibaum, E. C.; Teague, C.; Ueshima, H.; Zhou, B.; Lindon, J. C.; Nicholson, J. K.; Stamler, J.; Elliott, P.; Chan, Q.; Holmes, E. *Anal. Chem.* 2006, 78, 2199-2208.
- (9) Elliott, P.; Posma, J. M.; Chan, Q.; Garcia-Perez, I.; Wijeyesekera, A.; Bictash, M.; Ebbels, T. M.; Ueshima, H.; Zhao, L.; van Horn, L.; Daviglus, M.; Stamler, J.; Holmes, E.; Nicholson, J. K. *Sci. Transl. Med.* 2015, 7, 285ra262.
- (10) Nicholson, J. K.; Foxall, P. J. D.; Spraul, M.; Farrant, R. D.; Lindon, J. C. *Anal. Chem.* 1995, 67, 793-811.
- (11) Aue, W. P.; Karhan, J.; Ernst, R. R. *J. Chem. Phys.* 1976, 64, 4226-4227.
- (12) Dona, A. C.; Jimenez, B.; Schafer, H.; Humpfer, E.; Spraul, M.; Lewis, M. R.; Pearce, J. T.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. *Anal. Chem.* 2014, 86, 9887-9894.
- (13) Ludwig, C.; Viant, M. R. *Phytochemical analysis : PCA* 2010, 21, 22-32.
- (14) Foxall, P. J.; Parkinson, J. A.; Sadler, I. H.; Lindon, J. C.; Nicholson, J. K. *J. Pharm. Biomed. Anal.* 1993, 11, 21-31.
- (15) Zangger, K. *Prog. Nucl. Magn. Reson. Spectrosc.* 2015, 86-87, 1-20.
- (16) Foroozandeh, M.; Adams, R. W.; Kiraly, P.; Nilsson, M.; Morris, G. A. *Chem. Commun.* 2015, 51, 15410-15413.
- (17) Viant, M. R. *Biochem. Biophys. Res. Commun.* 2003, 310, 943-948.
- (18) Fonville, J. M.; Maher, A. D.; Coen, M.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. *Anal. Chem.* 2010, 82, 1811-1821.
- (19) Viant, M. R. *Methods Mol. Biol.* 2007, 358, 229-246.
- (20) Wang, Y.; Bollard, M. E.; Keun, H.; Antti, H.; Beckonert, O.; Ebbels, T. M.; Lindon, J. C.; Holmes, E.; Tang, H.; Nicholson, J. K. *Anal. Biochem.* 2003, 323, 26-32.
- (21) Posma, J. M.; Garcia-Perez, I.; Heaton, J. C.; Burdisso, P.; Mathers, J. C.; Draper, J.; Lewis, M.; Lindon, J. C.; Frost, G.; Holmes, E.; Nicholson, J. K. *Anal. Chem.* 2017, 89, 3300-3309.
- (22) Ludwig, C.; Easton, J. M.; Lodi, A.; Tiziani, S.; Manzoor, S. E.; Southam, A. D.; Byrne, J. J.; Bishop, L. M.; He, S.; Arvanitis, T. N.; Gunther, U. L.; Viant, M. R. *Metabolomics* 2012, 8, 8-18.
- (23) Parsons, H. M.; Ludwig, C.; Viant, M. R. *Magn. Reson. Chem.* 2009, 47 Suppl 1, S86-95.
- (24) Platt, D. E.; Ghassibe-Sabbagh, M.; Youhanna, S.; Hager, J.; Cazier, J. B.; Kamatani, Y.; Salloum, A. K.; Haber, M.; Romanos, J.; Doueihy, B.; Mouzaya, F.; Kibbani, S.; Sbeite, H.; Deeb, M. E.; Chammas, E.; El Bayeh, H.; Khazen, G.; Gauguier, D.; Zal-loua, P. A.; Abchee, A. B., et al. *J. Thromb. Thrombolysis* 2015, 39, 15-22.
- (25) Haug, K.; Salek, R. M.; Conesa, P.; Hastings, J.; de Matos, P.; Rijnbeek, M.; Mahendraker, T.; Williams, M.; Neumann, S.; Rocca-Serra, P.; Maguire, E.; Gonzalez-Beltran, A.; Sansone, S. A.; Griffin, J. L.; Steinbeck, C. *Nucleic Acids Res.* 2013, 41, D781-786.
- (26) Pearce, J. T.; Athersuch, T. J.; Ebbels, T. M.; Lindon, J. C.; Nicholson, J. K.; Keun, H. C. *Anal. Chem.* 2008, 80, 7158-7162.
- (27) Veselkov, K. A.; Lindon, J. C.; Ebbels, T. M.; Crockford, D.; Volynkin, V. V.; Holmes, E.; Davies, D. B.; Nicholson, J. K. *Anal. Chem.* 2009, 81, 56-66.
- (28) Bell, J. D.; Brown, J. C.; Nicholson, J. K.; Sadler, P. J. *FEBS Lett.* 1987, 215, 311-315.
- (29) Nagana Gowda, G. A.; Gowda, Y. N.; Raftery, D. *Anal. Chem.* 2015, 87, 706-715.
- (30) Posma, J. M.; Garcia-Perez, I.; De Iorio, M.; Lindon, J. C.; Elliott, P.; Holmes, E.; Ebbels, T. M.; Nicholson, J. K. *Anal. Chem.* 2012, 84, 10694-10701.
- (31) Cloarec, O.; Dumas, M. E.; Craig, A.; Barton, R. H.; Trygg, J.; Hudson, J.; Blancher, C.; Gauguier, D.; Lindon, J. C.; Holmes, E.; Nicholson, J. *Anal. Chem.* 2005, 77, 1282-1289.
- (32) Crockford, D. J.; Holmes, E.; Lindon, J. C.; Plumb, R. S.; Zirah, S.; Bruce, S. J.; Rainville, P.; Stumpf, C. L.; Nicholson, J. K. *Anal. Chem.* 2006, 78, 363-371.
- (33) Rodriguez-Martinez, A.; Posma, J. M.; Ayala, R.; Neves, A. L.; Anwar, M.; Petretto, E.; Emanueli, C.; Nicholson, J. K.; Dumas, M. E.; Bioinformatics, 2017. doi:10.1093/bioinformatics/btx477
- (34) Benjamini, Y.; Hochberg, Y. *J Roy Stat Soc B Met* 1995, 57, 289-300.
- (35) Nicholson, J. K.; O'Flynn, M. P.; Sadler, P. J.; Macleod, A. F.; Juul, S. M.; Sonksen, P. H. *Biochem. J.* 1984, 217, 365-375.
- (36) Crews, K. *Bioanalysis* 2013, 5, 2597-2597.