Spectral Relative Standard Deviation: A Practical Benchmark in Metabolomics

Helen M. Parsons[1], Drew R. Ekman[2], Timothy W. Collette[2] and Mark R. Viant[1,3]*

[1]Centre for Systems Biology, The University of Birmingham, Edgbaston, Birmingham, B15 2TT, U.K.

[2]National Exposure Research Laboratory, U.S. Environmental Protection Agency, Athens, GA 30605, U.S.A.

[3]School of Biosciences, The University of Birmingham, Edgbaston, Birmingham, B15 2TT, U.K.

Corresponding author: Mark R. Viant

Phone: +44-(0)121-414-2219

FAX: +44-(0)121-414-5925

Email: M.Viant@bham.ac.uk

Abstract

Metabolomics datasets, by definition, comprise of measurements of large numbers of metabolites. Both technical (analytical) and biological factors will induce variation within these measurements that is not consistent across all metabolites. Consequently, criteria are required to assess the reproducibility of metabolomics datasets that are derived from all the detected metabolites. Here we calculate spectrum-wide relative standard deviations (RSD; also termed coefficient of variation, CV) for ten metabolomics datasets, spanning a variety of sample types from mammals, fish, invertebrates and a cell line, and display them succinctly as boxplots. We demonstrate multiple applications of spectral RSDs for characterising technical as well as inter-individual biological variation: for optimising metabolite extractions, comparing analytical techniques, investigating matrix effects, and comparing biofluids and tissue extracts from single and multiple species for optimising experimental design. Technical variation within metabolomics datasets, recorded using one- and two-dimensional NMR and mass spectrometry, range from 1.6% to 20.6% (reported as the median spectral RSD). Inter-individual biological variation is typically larger, ranging from as low as 7.2% for tissue extracts from laboratory housed rats to 58.4% for fish plasma. In addition, for some of the datasets we confirm that the spectral RSD values are largely invariant across different spectral processing methods, such as baseline correction, normalisation and binning resolution. In conclusion, we propose spectral RSDs and their median values contained herein as practical benchmarks for metabolomics studies.

Introduction

Many factors are involved in conducting a high-quality metabolomics study, the most important of which is good experimental design and execution. A typical experiment involves collecting biological material, extracting and then measuring potentially hundreds of metabolites simultaneously by NMR spectroscopy or mass spectrometry (MS), spectral processing and multivariate statistical analysis. The principal goal is usually to discover metabolites that discriminate between two or more classes of samples. Consequently, it is essential to understand and minimise the spectrum-wide metabolic variation arising from technical sources as well as inter-individual metabolic variation within each class. Without this, the interpretation of results can be ambiguous, misleading and in the worst case, false.[1, 2]

Relative standard deviation (RSD = standard deviation / mean *100%), also termed coefficient of variation (CV), is one approach for characterising measurement variability. For example, Keun et al. reported RSDs for the measurement of citrate, taurine and hippurate in urine by NMR.[1] Clearly, if RSDs are to convey the reproducibility of whole metabolomics datasets, which by definition comprise of measurements of large numbers of metabolites, they should be calculated for all or at least the majority of detected peaks. To date there has been limited use of RSDs for capturing "spectrum-wide" variability in metabolomics studies. Previously, Ebbels et al. analysed the spectral variation both within and between rat strains in terms of the mean, standard deviation, skewness and kurtosis of each bin across an NMR spectrum.[3] RSDs of peak retention times and concentrations have been reported for the analysis of ca. 80 metabolites using HPLC with electrochemical detection.[2] RSDs of all data points, including those containing only noise, were reported graphically for the repeated NMR analysis of a blood sample.[4] Also, Want et al. reported histograms of RSD distributions of metabolite intensities for comparing the quality of extraction protocols for MS-based serum profiling.[5] A similar representation of RSDs was used to evaluate normalisation techniques for metabolomics data.[6]

Spectral RSDs have great potential to provide practical benchmarks in all fields of metabolomics for assessing and ultimately improving data quality. For example, the spectral RSD of an established protocol can provide a useful frame of reference when developing new methodologies or when applying the original protocol to different biological samples. In addition, established spectral RSDs can provide new researchers in metabolomics with "target values" for achieving high quality studies. Here we calculate spectral RSDs as a

simple measure of variance across multiple metabolomics datasets, display them succinctly in the form of boxplots, and then highlight many applications of this benchmark.† RSDs can be calculated for any spectral dataset. Here we focus on binned (or bucketed) spectra from NMR spectroscopy and mass spectrometry, which encompasses the most widely used approaches in metabolomics.

Methods

NMR spectral processing: Free induction decays from multiple datasets[7-17] were processed using standard methods, including Fourier transformation, phasing, baseline correction and calibration with either XWINNMR (Bruker), TopSpin (Bruker) or ACD/1D NMR Processor (Advanced Chemistry Development) software. 1D ^{1}H NMR spectra as well as 2D ^{1}H J-resolved spectra (JRES) and their 1D skyline projections (pJRES) were further processed using custom-written *ProMetab* software[18] in Matlab (version 7). All 1D datasets were trimmed to 0.2-10 ppm and sectioned into 1960 bins of width 0.005 ppm. 2D JRES spectra were processed in the same way along the chemical shift axis, but no binning was performed along the spin-spin coupling axis. Bins containing, for example, residual water and urea were excluded, and groups of bins containing peaks that were susceptible to shifting were compressed together. Spectra were then normalised to a total intensity of unity.

Noise estimation and removal from binned NMR spectra: For 1D data, each binned spectrum was divided into 32 sections and the standard deviation of each section calculated. The noise level for a given spectrum was estimated as 3 times the smallest standard deviation of these 32 sections.[19, 20] For 2D JRES data, a noise surface was calculated using a similar approach. For each 2D spectrum, a noise level was estimated (exactly as above) for each of the 128 spin-spin coupling increments. Next the 2D JRES spectra and corresponding noise surfaces were concatenated into 1D row vectors. Bins containing only noise were removed from the NMR spectra (which are now all represented as 1D row vectors). This was achieved by comparing the intensity of each bin to the corresponding noise level, and discarding all bins that had intensities lower than the noise threshold. Next, all spectra within one class (or dataset) were compared, and only bins containing signal in every spectrum were retained for calculation of RSDs.

† The Matlab code is freely available upon request from the corresponding author.

MS spectral processing: Fourier transform ion cyclotron resonance (FT-ICR) mass spectra of fish liver extracts were initially processed as described previously,[11] yielding a list of peaks between 70-500 Da for each sample. For each list, the total peak intensity was normalised to unity and the peaks were segmented into 430 bins of width 1 Da.

RSD calculation: RSDs were calculated for each dataset as summarised in the example in Fig. 1. Five binned NMR spectra of fish liver extracts are overlaid and plotted in Fig. 1A. Following noise removal, as described above, an RSD was calculated for every bin. Fig. 1B shows the ranked mean bin intensity (of the five spectra) plotted against the corresponding bin RSD, highlighting the consistency of the RSD values across low and high intensity peaks. The distribution of the RSDs is shown in Fig. 1C, which is re-drawn as a boxplot in Fig. 1D. The boxplot facilitates visual comparison of RSDs from multiple datasets, showing lines at the lower quartile, median and upper quartile values, whiskers to display the range of the remaining data, and outliers as individual data points. Outliers correspond to RSD values that are 1.5 times the interquartile range (or more) below the lower quartile, or 1.5 times the interquartile range (or more) above the upper quartile. Throughout the manuscript the boxplots show RSD from 0-100% to enable comparisons across all datasets, and all results are summarised in Table 1. To establish if RSD distributions differed significantly they were analysed using the Wilcoxon rank sum test for comparing two distributions and the Kruskal-Wallis test for multiple distributions (Matlab statistics toolbox).

Applications for RSDs of technical replicates

The principal use for RSDs of technical (or analytical) replicates is to evaluate spectral quality in terms of the experimental methods. Several potential uses were described above in the introduction, for example as a practical benchmark to compare the reproducibility of one method across several independent laboratories. Furthermore, this approach can provide "target values" for new practitioners in metabolomics to confirm they are achieving high data quality. Here we demonstrate four such applications using multiple technical replicates of *single* biological samples (i.e. the metabolic variability described here is purely from technical sources and not of biological origin). It is important to note at this stage that RSD values appear largely invariant to different spectral processing methods. As summarised in Table 1, we confirmed this by evaluating the effects of spectral processing software (ACD/1D NMR Processor versus Topspin), NMR bin width (0.04 versus 0.005 ppm), and

normalisation method (total spectral area of unity versus probabilistic quotient normalisation[21]), none of which altered the %RSDs by more than a few percent.

Optimising metabolite extractions: Two protocols were compared for extracting metabolites from a chub liver (*Leuciscus cephalus*), 6% perchloric acid (n=10) and 2:1 methanol:water (M:W; n=5).[7] The two distributions of RSDs derived from 1D NMR spectra are significantly different ($p<0.001$; Fig. 2), and highlight the considerably higher reproducibility of the M:W extraction (median RSD of 4.6%) versus perchloric acid (median of 20.6%). This is consistent with observations from the original study that NMR spectra of perchloric acid extracts showed considerable pH-induced peak shifting.[7]

Sampling, extraction and matrix effects: Different sample types, such as biofluids, cells or tissues, are typically collected and the metabolites extracted using protocols of differing complexities. Also, the resulting extracts will be present in different sample matrices. Together these factors will affect the reproducibility of the metabolomics data. NMR spectra of four disparate sample types were compared: urine from a dog that was collected by free-catch, divided (n=5) and prepared by simple buffer addition;[8] a flask of acute myeloid leukaemia cells (cell line K562) that was divided (n=5), centrifuged, washed and extracted using methanol:chloroform:water (M:C:W);[9] adductor muscle from a Mediterranean mussel (*Mytilus galloprovincialis*) that was rapidly dissected, frozen, homogenised from the frozen state, divided (n=6) and then extracted using M:C:W;[10] and liver from a 3-spined stickleback, (*Gasterosteus aculeatus*) that was processed in an identical manner to the mussel tissue (n=6). Fig. 2 shows the differences between the four RSDs ($p<0.001$) and confirms that the simplest collection and extraction procedure used in conjunction with the simplest aqueous matrix results in the most reproducible data (median RSD of 1.6% for urine). The rapid dissection, freeze clamping, homogenisation and extraction of more complex tissue samples increases the median RSD to 3.4% (stickleback) and 6.1% (mussel). The preparation of cell extract replicates, requiring several minutes of centrifugation and washing prior to the quenching of metabolism, yielded the least reproducible data (median RSD of 14.0%). These values serve as valuable benchmarks.

Comparison of analytical approaches: Three datasets were recorded using different NMR experiments, comprising traditional 1D ^{1}H, 1D projections of JRES spectra, and intact 2D JRES spectra. In all cases, the same five replicates of a European flounder (*Platichthys flesus*)

liver extract were analysed. The boxplots corresponding to the three sets of RSD values illustrate that 1D NMR spectroscopy yields the most reproducible data (median RSD of 3.1%), the intact 2D JRES spectra show the highest variability (19.8%), and the pJRES data is intermediate between the other two with a median RSD of 12.5% ($p<0.001$; Fig. 2). This can be explained in part by considering the number of scans used for each type of experiment: 100 averaged transients for each more quantitative 1D spectrum but only 16 for each JRES spectrum, yielding higher precision in the 1D measurement. The median RSD value for the technical variability of dab (*Limanda limanda*) liver, measured using direct infusion FT-ICR mass spectrometry, is 13.1%.[11] Since this value was not determined using the same biological samples as for the NMR measurements, a direct comparison of the MS and NMR approaches is not possible. However, 1D NMR is renowned for excellent analytical precision and therefore might be expected to yield a lower median RSD.

Applications for RSDs of inter-individual variation within one class

Variation between individuals in one class (or group) includes genuine biological inter-individual differences as well as technical variation. Quantifying this inter-individual variation can be useful for optimising experimental design, as illustrated below.

Comparison of biofluids and tissue extracts from single species: It is commonly believed (but rarely shown) that inter-individual metabolic variability is biofluid or tissue dependent. Tissue is under greater homeostatic control than biofluids and so one would predict that metabolic variability increases from tissue to plasma to urine, the latter being particularly susceptible to diet and lifestyle. Fig. 3A compares the RSDs derived from NMR datasets of three sample types, all derived from 8 marine invertebrates (red abalone, *Haliotis rufescens*).[12] The RSD distributions for foot muscle, digestive gland and haemolymph (blood) are significantly different ($p<0.001$) with haemolymph showing the greatest inter-individual variability, consistent with the prediction of homeostatic control. The RSDs of NMR spectra of rat brain and plasma, obtained from five individuals,[13] again show that the biofluid exhibits somewhat greater variability ($p<0.001$; Fig. 3B). Interestingly, these inter-individual median RSDs of 7.2% (brain) and 8.0% (plasma) are considerably smaller than for fish and marine invertebrate tissues and biofluids (that range from 16.0-58.4%), likely reflecting the conserved metabolism in laboratory-raised Sprague-Dawley rats, which were all male adults, of mass 325-375 g and were individually housed in environmentally-controlled chambers under a 12:12 hr light:dark cycle. As a further example, Fig. 3C shows the variance within

NMR spectra of testis, urine and plasma from 7 fathead minnow (*Pimephales promelas*).[17] The tissue again exhibits the lowest RSDs (median of 29.4%), while the two biofluids have similar inter-individual metabolic variability (p=0.263). It should be noted that while tissues are generally under greater homeostatic control, some exhibit considerable heterogeneity (e.g. tumour tissue) and therefore will produce larger than expected biological variation. Overall these results can help to guide experimental design, for example by identifying which sample type yields the lowest inter-individual variability.

Comparison of different species: The same sample type can also be compared across different species, which may be particularly useful for determining the feasibility of studying non-model organisms; i.e. to confirm that the inter-individual metabolic variability for that species is sufficiently low to produce interpretable data. Fig. 3D compares the RSDs from NMR spectra of rat urine (n=5) and fathead minnow urine (n=7). All samples were buffered and processed using identical procedures. However, the RSDs are quite different ($p<0.001$), potentially reflecting differing biochemistries of these organisms (which could be related to animal husbandry, described below).

Optimisation of organism "husbandry": Inter-individual variation can be used to assess protocols for the culturing or husbandry of organisms, as the condition in which animals are kept will have a large impact on their metabolic variability. This is particularly true for metabolomics studies of wildlife when organisms are housed in "foreign" laboratory environments. Inter-individual variation was compared for mussels (*M. galloprovincialis)* collected directly from the field (n=14), versus animals collected from the field and then maintained in a controlled laboratory environment for 48 hr (n=14).[10] The median RSDs derived from NMR spectra show that the laboratory introduces a small but significant increase in the metabolic variability in two tissues, the adductor muscle ($p<0.05$) and mantle ($p<0.001$) as shown in Fig. 3E. It was concluded that direct sampling from the field was preferable, as discussed in the original paper.[10]

Applications for RSDs of inter-individual variation between classes

The goal of most metabolomics experiments is to discriminate between sample classes based upon their metabolic compositions. It is therefore important and useful to confirm that technical variation is small relative to the inter-individual variation within each class, prior to interpreting the biological variation. This is clearly illustrated in the following three

examples. Furthermore, RSDs can be used to assess the differences in inter-individual variation between "control" and "exposed/stressed" classes. It is often stated by practitioners of metabolomics that the control class shows the most inter-individual variation. However, this is typically qualitative and anecdotal. For example, based upon the scatter of data points in a PCA scores plot, control female fathead minnow were determined to vary more compared to fish exposed to oestrogenic compounds.[22] This was hypothesised to arise from the control fish being in various stages of the reproductive cycle. After the exposure, however, the reproductive status of the female fish was more uniform, resulting in a metabolic response that was quite "focused". The calculation of RSDs can be used to assess and quantify such observations, as discussed in the following three examples. Of course many more biological studies would need to be characterised in this manner to assess the prevalence of the control class showing higher variability.

Effect of environment on fish liver metabolome: Flounder were collected from the Rivers Alde (n=19; clean) and Tyne (n=18; industrialised and more polluted) to determine the effects of different environments (and potentially different genetic make-ups) on the liver metabolome.[14] RSDs were derived from NMR spectra of individuals from each site as well as from technical replicates of one individual (Fig. 4A). Clearly, technical variation is considerably lower than the biological variation between fish from either site. Furthermore, the significant difference ($p<0.001$) between the inter-individual variations for the two sites suggests that the two fish populations are responding differently to their environments, with the "control" fish from the clean site showing greater metabolic variability. The metabolites responsible for this high metabolic variation in fish from the River Alde are identified in Fig. 5, which highlights the 20 NMR bins that are associated with the largest RSD values. The peaks with the greatest *relative* standard deviations are clearly distributed across the spectrum, and include both low and high intensity signals. This sort of plot could be derived for any metabolomics dataset in order to identify, for example, those metabolites that vary as a result of a particular extraction method (e.g., for technical replicates) or that vary within a particular biological class (e.g., for biological replicates).

Effect of drug treatment on cellular metabolome: K562 leukaemia cells were treated with medroxy progesterone acetate. Fig. 4B illustrates the RSDs derived from NMR spectra of both treated (n=12) and untreated (n=12) flasks of cells, as well as from five technical replicates. Again the technical replicates show the least metabolic variation. For this study,

although there is again a significant difference between the inter-individual variations ($p<0.01$), the cells treated with the drug exhibit greater metabolic variability.

Effect of disease on metabolome: NMR spectra of foot muscle were obtained from healthy abalone shellfish (n=8), animals with withering syndrome (n=5), as well as technical replicates from one individual (n=5).[12] Fig. 4C shows that the technical replicates have the smallest variation (median of 5.4%), confirming that variability arising from the methodology is minor compared to the biological differences. Furthermore, the diseased class exhibits a significantly higher inter-individual variation than the healthy controls ($p<0.01$). As withering syndrome is a chronic disease that progressively degrades the metabolic condition of the shellfish over many months, this high variation can be rationalised in terms of the animals being in various stages of the disease. Overall, considering just these few examples, it suggests that no "golden rule" exists for the control class exhibiting higher variability.

Conclusions

Following an extensive literature search we have found relatively minimal exploitation of RSD distributions for assessing the reproducibility of metabolomics datasets. Here we have calculated and reported RSDs as median values and graphically using boxplots. Ten NMR and MS datasets have been analysed in terms of technical variation or inter-individual variation, spanning a variety of sample types including fish, invertebrates, mammals and a model cell line. We have clearly demonstrated multiple applications of RSDs for assessing the quality of metabolomics datasets as well as for improving experimental design. We conclude that this benchmark could be of considerable value to both existing and new practitioners in the field, for both the interpretation of technical and biological variation, and therefore recommend that a database of RSD values be established.

Acknowledgements

HMP thanks the EPSRC and NERC for a Directed PhD studentship and MRV thanks the NERC for an Advanced Fellowship (NER/J/S/2002/00618). The authors gratefully acknowledge Dr. Dov Stekel for advice on the statistical analyses, Drs. Alessia Lodi and Stefano Tiziani for the leukaemia cell NMR spectra , Dr. Huifeng Wu, Dr. Adam Hines and Andy Southam for the fish and invertebrate spectra, Dr. David Dix for rat urine samples, and Drs. Gary Ankley, Dan Villeneuve and Dalma Martinovic for fathead minnow samples.

This paper has been reviewed in accordance with the U.S. Environmental Protection Agency's peer and administrative review policies and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

References

1. H. C. Keun, T. M. D. Ebbels, H. Antti, M. E. Bollard, O. Beckonert, G. Schlotterbeck, H. Senn, U. Niederhauser, E. Holmes, J. C. Lindon and J. K. Nicholson, *Chem. Res. Toxicol.*, 2002, **15**, 1380-1386.
2. Y. I. Shurubor, U. Paolucci, B. F. Krasnikov, W. R. Matson and B. S. Kristal, *Metabolomics*, 2005, **1**, 75-85.
3. T. M. D. Ebbels, E. Holmes, J. C. Lindon and J. K. Nicholson, *J. Pharm. Biomed. Anal.*, 2004, **36**, 823-833.
4. O. Teahan, S. Gamble, E. Holmes, J. Waxman, J. Nicholson, C. Bevan and H. Keun, *Anal. Chem*, 2006, **78**, 4307-4318.
5. E. J. Want, G. O'Maille, C. A. Smith, T. R. Brandon, W. Uritboonthai, C. Qin, S. A. Trauger and G. Siuzdak, *Anal. Chem.*, 2006, **78**, 743-752.
6. M. Sysi-Aho, M. Katajamaa, L. Yetukuri and M. Oresic, *BMC Bioinformatics*, 2007, **8**, 93.
7. C. Y. Lin, H. F. Wu, R. S. Tjeerdema and M. R. Viant, *Metabolomics*, 2007, **3**, 55-67.
8. M. R. Viant, C. Ludwig, S. Rhodes, U. L. Guenther and D. Allaway, *Metabolomics*, 2007, **3**, 453-463.
9. S. Tiziani, A. Lodi, C. Ludwig, H. M. Parsons and M. R. Viant, *Anal. Chim. Acta*, 2008, **610**, 80-88.
10. A. Hines, G. S. Oladiran, J. P. Bignell, G. D. Stentiford and M. R. Viant, *Environ. Sci. Technol.*, 2007, **41**, 3375-3381.
11. A. D. Southam, T. G. Payne, H. J. Cooper, T. N. Arvanitis and M. R. Viant, *Anal. Chem.*, 2007, **79**, 4595-4602.
12. M. R. Viant, E. S. Rosenblum and R. S. Tjeerdema, *Environ. Sci. Technol.*, 2003, **37**, 4982-4989.
13. M. R. Viant, B. G. Lyeth, M. G. Miller and R. F. Berman, *NMR in Biomedicine*, 2005, **18**, 507-516.
14. H. M. Parsons, C. Ludwig, U. L. Günther and M. R. Viant, *BMC Bioinformatics*, 2007, **8**, 234.
15. D. P. Cherney, D. R. Ekman, D. J. Dix and T. W. Collette, *Anal. Chem.*, 2007, **79**, 7324-7332.
16. D. R. Ekman, Q. Teng, K. M. Jensen, D. Martinovic, D. L. Villeneuve, G. T. Ankley and T. W. Collette, *Aquat. Toxicol.*, 2007, **85**, 104-112.

17. D. Martinović, D. R. Ekman, T. W. Collette, Q. Teng, D. L. Villeneuve, L. S. Blake, K. J. Greene, K. M. Jensen, M. D. Kahl and G. T. Ankley, in preparation.
18. M. R. Viant, *Biochem. Biophys. Res. Comm*, 2003, **310**, 943-948.
19. D. MacDougall and ACS Committee on Environmental Improvement, *Anal. Chem.*, 1980, **52**, 2242-2249.
20. S. Golotvin and A. Williams, *J. Magn. Reson.*, 2000, **146**, 122-125.
21. F. Dieterle, A. Ross, G. Schlotterbeck and H. Senn, *Anal. Chem.*, 2006, **78**, 4281-4290.
22. D. R. Ekman, T. W. Collette, Q. Teng, T. Whitehead, G. T. Ankley, K. M. Jensen, M. D. Kahl and D. L. Villeneuve, Abstracts of the Second Scientific Meeting of the Metabolomics Society, Boston, MA, USA, 2006.

Table 1. Summary of spectral RSDs for multiple metabolomics datasets.

Species	Sample	Class	Extraction method	Analytical technique	No. of 'signal bins'	Technical variation (one individual)		Inter-individual variation	
						median %RSD	%RSD range	median %RSD	%RSD range
Fish									
chub	liver	-	M:W	1D NMR	682	4.6	53		
	liver	-	perchloric acid	1D NMR	682	20.6	267		
3-spined stickleback	liver	-	M:C:W	1D NMR	914	3.4	32		
European flounder	liver	-	M:C:W	1D NMR	1166	3.1	34		
	liver	-	M:C:W	pJRES NMR	1202	12.5	58		
	liver	-	M:C:W	2D JRES NMR	1840	19.8	97		
	liver	field (Alde)	M:C:W	1D NMR	1073			30.1 [a]	183
	liver	field (Tyne)	M:C:W	1D NMR	991			24.9	198
dab	liver	-	M:C:W	FT-ICR MS	219	13.1	68		
fathead minnow	testis [b]	control	M:C:W	1D NMR	690			29.4 [c]	122
	plasma [b]	control	pH 7.4 buffer added	1D NMR	524			58.4	174
	urine [b]	control	pH 7.4 buffer added	1D NMR	273			52.9	146
Marine invertebrates									
red abalone	foot muscle	-	perchloric acid	1D NMR	989	5.4	179		
	foot muscle [d]	healthy	perchloric acid	1D NMR	964			16.0	125
	foot muscle	diseased	perchloric acid	1D NMR	691			21.5	146
	digestive gland [d]	healthy	perchloric acid	1D NMR	919			19.7	132
	haemolymph [d]	healthy	perchloric acid	1D NMR	671			25.2	92
Mediterranean mussel	adductor muscle	-	M:C:W	1D NMR	1088	6.1	56		
	adductor muscle	field	M:C:W	1D NMR	938			24.4	111
	adductor muscle	laboratory	M:C:W	1D NMR	754			26.0	138
	mantle	field	M:C:W	1D NMR	1044			23.5	90
	mantle	laboratory	M:C:W	1D NMR	854			26.9	97
Mammals									
dog	urine	-	buffer to pH 7.05±0.05	1D NMR	1177	1.6	51		
rat	brain [e]	control	perchloric acid	1D NMR	702			7.2	87
	plasma [e]	control	pH 7.4 buffer added	1D NMR	852			8.0	106
	urine [e]	control	pH 7.4 buffer added	1D NMR	981			32.2	164
human immortalised K562 cell line	cell extract	-	M:C:W	1D NMR	1054	14.0	90		
	cell extract	untreated	M:C:W	1D NMR	682			20.5	84
	cell extract	treated	M:C:W	1D NMR	682			22.0	112

[a] To confirm that spectral processing has a minimal effect on the median RSD, this dataset (originally segmented into 0.005 ppm bins and each spectrum normalised to a total area of one) was reprocessed, firstly using a bin width of 0.04 ppm (yielding a median RSD of 29.9%) then using probabilistic quotient normalisation[21] (yielding a median RSD of 27.2%).

[b] All samples obtained from the same fathead minnows.

[c] To confirm that spectral processing has a minimal effect on the median RSD, this dataset (originally apodised, Fourier transformed, phased, baseline corrected and calibrated using ACD/1D NMR Processor software) was reprocessed using Topspin software, yielding a median RSD of 26.7%.

[d] Samples obtained from the same abalone.

[e] Brain and plasma derived from one cohort of rats and urine derived from a different group of individuals.

Figure captions

Fig. 1. Calculation of spectral RSD values. (A) Five overlaid, binned 1D NMR spectra of fish liver extracts (bin width of 0.005 ppm). (B) Following removal of bins that contain only noise, RSDs are calculated for each remaining bin and then ranked according to bin signal intensity. A trendline (with running median of 10 bins) highlights that the RSDs are largely invariant to signal intensity. (C) Histogram of RSD values showing a right-skewed distribution. (D) Boxplot of RSD values which summarises succinctly the lower quartile, median and upper quartile values, whiskers to display the range of data, and outliers as individual data points.

Fig. 2. Boxplots of RSDs derived from technical replicate spectra for 10 independent datasets, as described in main text. These datasets facilitate comparisons of the reproducibilities associated with two solvent extraction methods, four types of biological sample, two analytical techniques, and three types of NMR experiment. Key: M:W = methanol:water extraction; M:C:W = methanol:chloroform:water extraction; JRES = J-resolved NMR spectroscopy; pJRES = 1D projection of 2D JRES spectrum.

Fig. 3. Boxplots of RSDs derived from several NMR datasets that facilitate comparisons of inter-individual metabolic variation across classes. (A) Three types of biological sample from abalone shellfish. (B) Tissue and biofluid samples from rat. (C) Three types of biological sample from fathead minnow. (D) Urine samples from two different species. (E) Comparison of husbandry and sampling techniques for two tissues from marine mussels. The RSD range has been clipped to show 0-100% only to facilitate comparison across all boxplots.

Fig. 4. Boxplots of RSDs derived from several NMR datasets that compare technical variation (in one sample) to inter-individual metabolic variation across classes. (A) Effect of sampling site on flounder liver extracts. (B) Effect of drug treatment on K562 leukaemia cell lines. (C) Effect of withering syndrome disease on abalone adductor muscle. In all cases, technical variability is shown to be smaller than inter-individual variability.

Fig. 5. Mean of the 19 1D NMR spectra of flounder liver extracts for the 19 fish sampled from the River Alde. Circles have been drawn around the 20 NMR bins that are associated with the largest RSD values (i.e., bins with the highest intra-class variation).

Figure 1

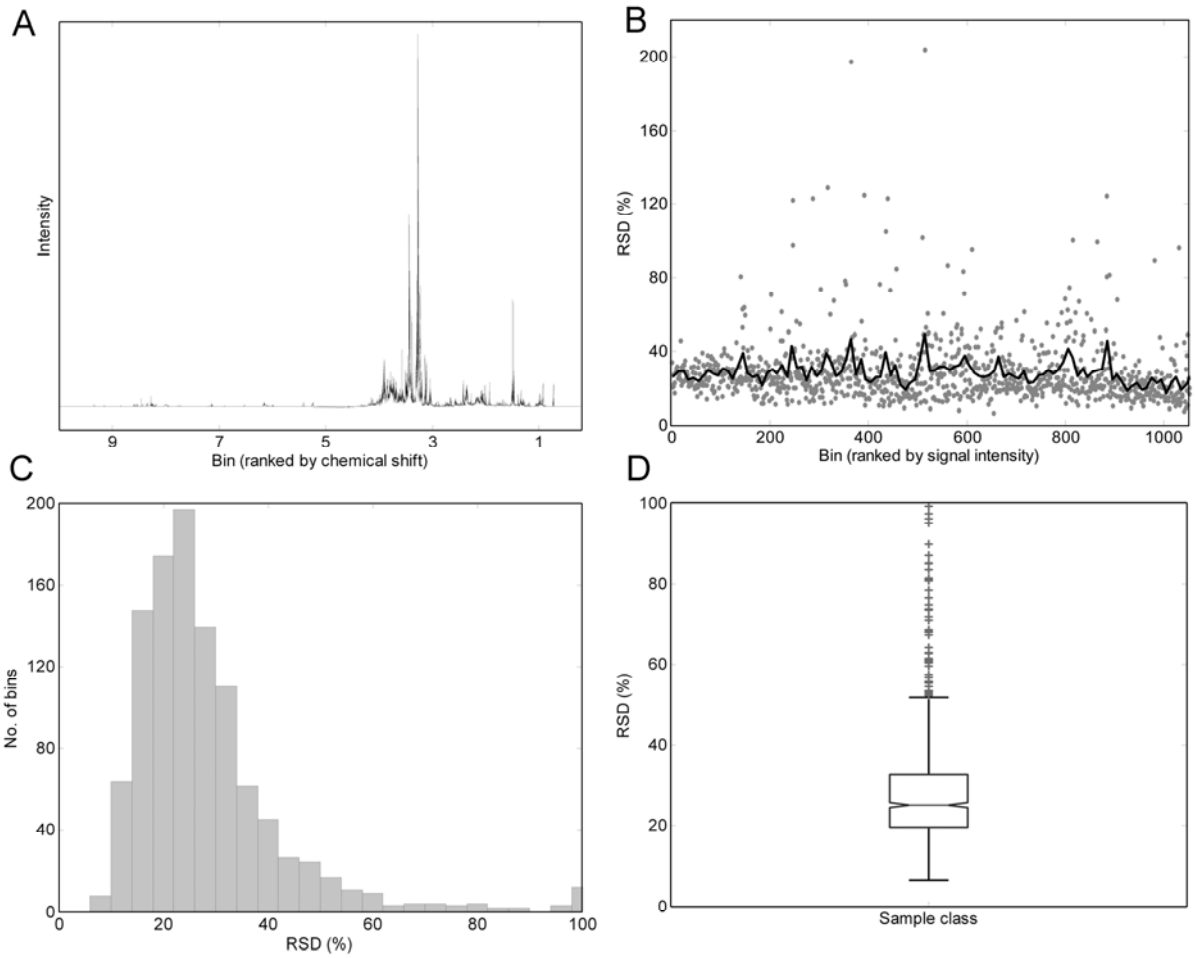
A
Intensity
9
7
5
3
1
Bin (ranked by chemical shift)
B
RSD (%)
0
40
80
120
160
200
0
200
400
600
800
1000
Bin (ranked by signal intensity)
C
No. of bins
0
40
80
120
160
200
0
20
40
60
80
100
RSD (%)
D
RSD (%)
0
20
40
60
80
100
Sample class

Figure 2

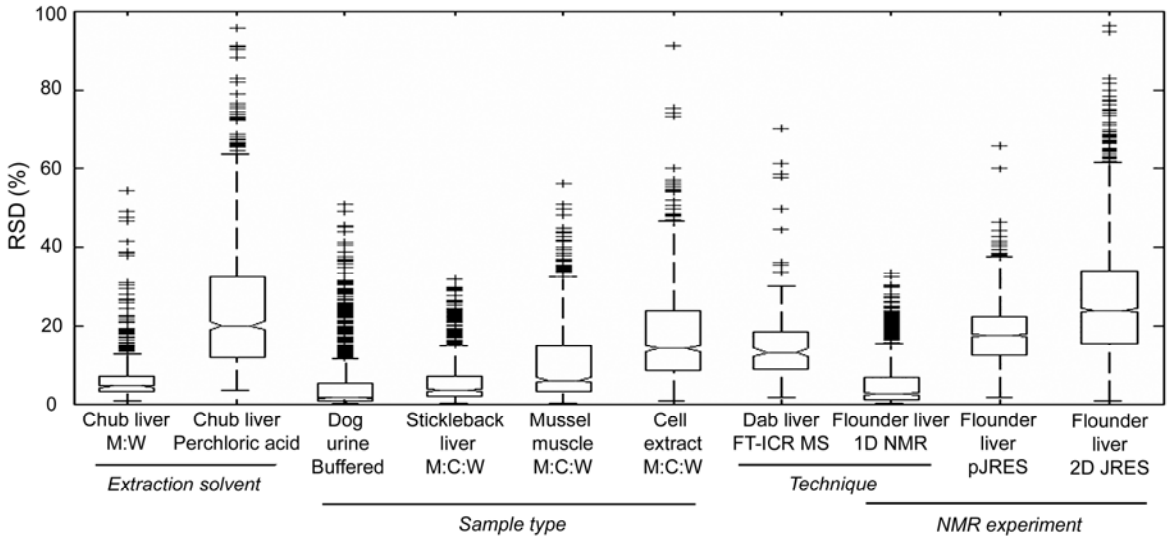
RSD (%)
100
80
60
40
20
0
Chub liver M:W
Chub liver Perchloric acid
Dog urine Buffered
Stickleback liver M:C:W
Mussel muscle M:C:W
Cell extract M:C:W
Dab liver FT-ICR MS
Flounder liver 1D NMR
Flounder liver pJRES
Flounder liver 2D JRES
Extraction solvent
Sample type
Technique
NMR experiment

Figure 3

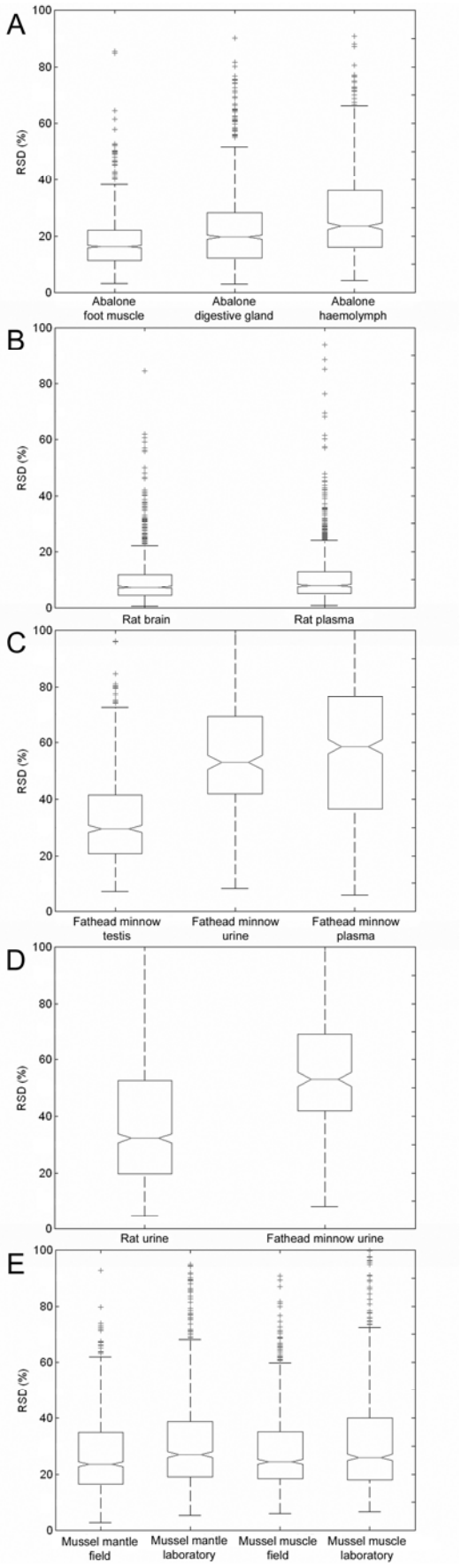
A
B
C
D
E
RSD (%)
0
20
40
60
80
100
Abalone foot muscle
Abalone digestive gland
Abalone haemolymph
Rat brain
Rat plasma
Fathead minnow testis
Fathead minnow urine
Fathead minnow plasma
Rat urine
Fathead minnow urine
Mussel mantle field
Mussel mantle laboratory
Mussel muscle field
Mussel muscle laboratory

Figure 4

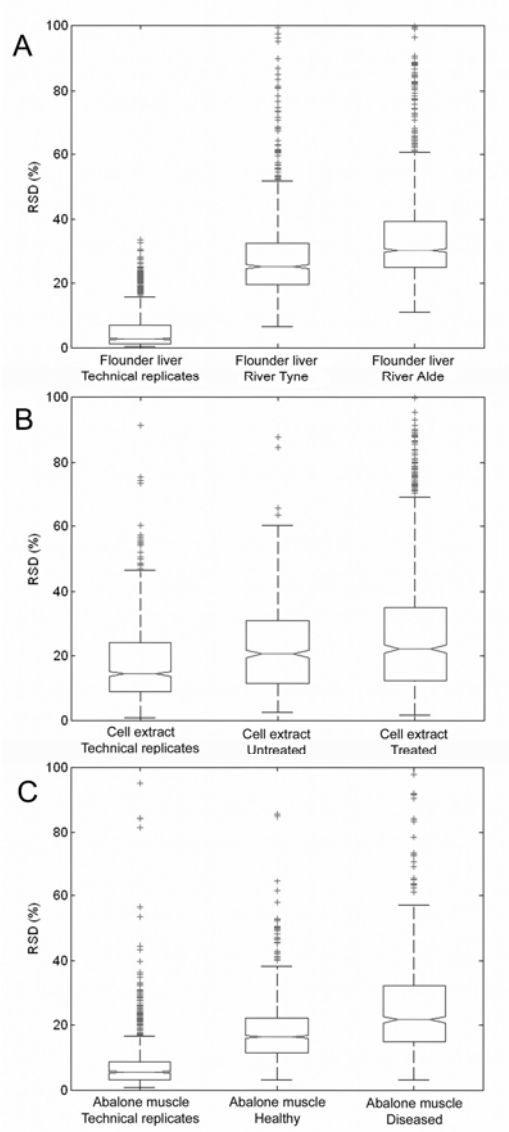
A
RSD (%)
100
80
60
40
20
0
Flounder liver
Technical replicates
Flounder liver
River Tyne
Flounder liver
River Alde
B
RSD (%)
Cell extract
Technical replicates
Cell extract
Untreated
Cell extract
Treated
C
RSD (%)
Abalone muscle
Technical replicates
Abalone muscle
Healthy
Abalone muscle
Diseased

Figure 5