

# Deriving three one dimensional NMR spectra from a single experiment through machine learning

Alessia Vignoli<sup>1,2</sup>, Stefano Cacciatore<sup>3</sup>, Leonardo Tenori<sup>1,2</sup>

## Affiliations:

<sup>1</sup> Department of Chemistry “Ugo Schiff”, University of Florence, Via della Lastruccia 3-13, 50019 Sesto Fiorentino, Italy

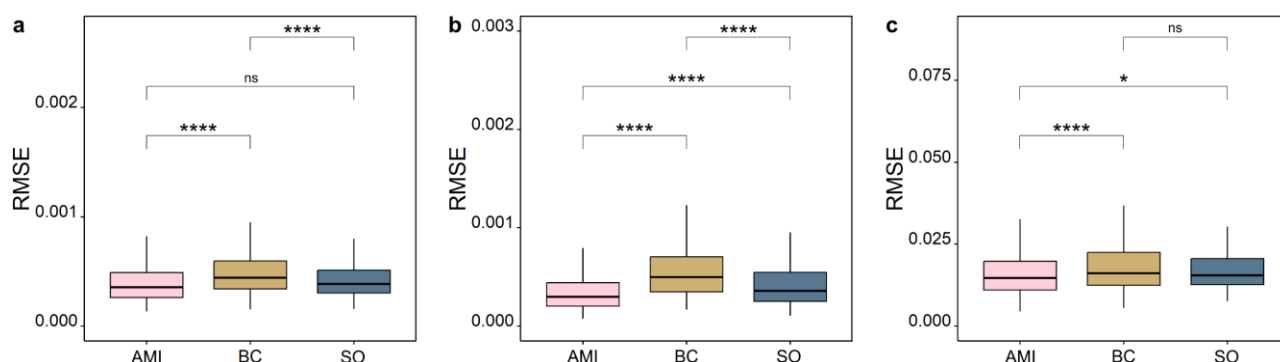
<sup>2</sup> Magnetic Resonance Center (CERM), University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Italy

<sup>3</sup> Bioinformatics Unit, International Centre for Genetic Engineering and Biotechnology, Anzio Road, Cape Town, 7925, South Africa

*Materials & Correspondence:* Prof. Leonardo Tenori, leonardo.tenori@unifi.it

Dr. Stefano Cacciatore, Stefano.Cacciatore@icgeb.org

## Supplementary Information



**Figure S1:** The plots show the root mean square error (RMSE) distribution across the three datasets used (10-fold cross-validated results, n=1753), MTBLS395 (Acute Myocardial Infarction, pink, n=940), MTBLS424 (Breast Cancer, green, n=636), and MTBLS242 (Severe Obesity, blue, n=177), for the prediction of a) CPMG (Carr-Purcell-Meiboom-Gill), b) diffusion-edited, and c) pJRES spectra. Box plot center line: median value; box bounds: interquartile range (IQR); upper whisker: 3rd quartile + 1.5 IQR; lower whisker: 1st quartile – 1.5 IQR). *P*-value symbols indicating statistical significance: \*\*\*\*  $\leq 0.0001$ , \*\*\*  $\leq 0.001$ , \*\*  $\leq 0.01$ , \*  $\leq 0.05$ , ns  $> 0.05$ .

**Table S1:**  $^1\text{H}$  chemical shifts of assigned and quantified metabolites in all 1D  $^1\text{H}$  CPMG (Carr-Purcell-Meiboom-Gill) spectra (original and predicted). Abbreviations: d: doublet; m: multiplet; q: quadruplet; s: singlet.

| Metabolite        | $^1\text{H}$ Chemical Shift (ppm) | Multiplicity     |
|-------------------|-----------------------------------|------------------|
| Leucine           | 0.96                              | m                |
| Isoleucine        | 1.02                              | d                |
| Valine            | 1.04                              | d                |
| 3-hydroxybutyrate | 1.21                              | d                |
| Alanine           | 1.49                              | d                |
| Acetate           | 1.92                              | s                |
| GlycA             | 2.04                              | multiple signals |
| GlycB             | 2.08                              | multiple signals |
| Glutamate         | 2.35                              | m                |
| Glutamine         | 2.48                              | m                |
| Citrate           | 2.55                              | d                |
| Choline           | 3.2                               | s                |
| Creatine          | 3.93                              | s                |
| Creatinine        | 4.06                              | s                |
| Lactate           | 4.11                              | q                |
| Glucose           | 5.24                              | d                |
| Tyrosine          | 6.91                              | m                |
| Histidine         | 7.06                              | d                |
| Phenylalanine     | 7.43                              | m                |
| Formate           | 8.46                              | s                |