The data was preprocessed as follows: (1) Metabolites with more than 50% missing values were filtered out. (2) Metabolite abundances were normalized using the probabilistic quotient approach1, using only variables with less than 20% missing values to construct the reference sample. (3) Normalized values were log-scaled to improve normality, as metabolite abundances are typically log-normally distributed. (4) The remaining missing values were imputed using a k-Nearest-Neighbors-based algorithm (knn with variable preselection and k=10)2.

1. Dieterle, F., Ross, A., Schlotterbeck, G. & Senn, H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in1H NMR metabonomics. *Anal. Chem.* **78**, 4281–4290 (2006).

2. Do, K. T. *et al.* Characterization of missing values in untargeted MS-based metabolomics data and evaluation of missing data handling strategies. *Metabolomics* **14**, 128 (2018).