



# Data-dependent normalization strategies for untargeted metabolomics—a case study

Paula Cuevas-Delgado<sup>1</sup> · Danuta Dudzik<sup>1,2</sup> · Verónica Miguel<sup>3</sup> · Santiago Lamas<sup>3</sup> · Coral Barbas<sup>1</sup>

Received: 15 December 2019 / Revised: 4 March 2020 / Accepted: 10 March 2020 / Published online: 14 April 2020  
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

## Abstract

Despite the recent advances in the standardization of untargeted metabolomics workflows, there is still a lack of attention to specific data treatment strategies that require deep knowledge of the biological problem and need to be applied after a well-thought out process to understand the effect of the practice. One of those strategies is data normalization. Data-driven assumptions are critical especially addressing unwanted variation present in the biological model as it can be the case in heterogeneous tissues, cells with different sizes or biofluids with different concentrations. Chronic kidney disease (CKD) is a widespread disorder affecting kidney structure and function. Animal models are being developed to be able to get valuable insights into the etiopathogenesis of the condition and effect of the treatments. Moreover, diagnosis and disease staging still require defining appropriate biomarkers. Untargeted metabolomics has the potential to deal with those challenges. Renal fibrosis is one of the consequences of kidney injury which greatly affects the concentration of metabolites in the same quantity of sample. To overcome this challenge, several data normalization strategies have been applied, following a multilevel normalization method with the overall aim of focussing on the relevant biological information and reducing the influence of disturbing factors. A comprehensive evaluation of the performance of the normalization strategies, both on methods assessing the intragroup variation and on the impact on differential analysis, is provided. Finally, we present evidence of the importance of biological-model-driven guided normalization methods and discuss multiple criteria that need to be taken into consideration to obtain robust and reliable data. Special concern is transmitted on the misleading conclusions that might be the consequence of inappropriate data pre-treatment solutions applied for untargeted methods.

**Keywords** Unwanted variation · Data pre-treatment · Normalization · Tissue samples · Capillary electrophoresis mass spectrometry · Biomarker discovery

Published in the topical collection featuring *Female Role Models in Analytical Chemistry*.

Paula Cuevas-Delgado and Danuta Dudzik contributed equally to this work.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00216-020-02594-9>) contains supplementary material, which is available to authorized users.

✉ Coral Barbas  
cbarbas@ceu.es

<sup>1</sup> Centro de Metabolómica y Bioanálisis (CEMBIO), Facultad de Farmacia, Universidad San Pablo-CEU, CEU Universities, Urbanización Montepríncipe, 28660 Boadilla del Monte, Madrid, Spain

<sup>2</sup> Department of Biopharmaceutics and Pharmacodynamics, Faculty of Pharmacy, Medical University of Gdańsk, al. J. Hallera 107, 80-416 Gdańsk, Poland

<sup>3</sup> Program of Physiological and Pathological Processes, Centro de Biología Molecular “Severo Ochoa” (CSIC-UAM), c. Nicolás Cabrera 1, 28049 Madrid, Spain

## Abbreviations

BCA	Bicinchoninic acid assay
BGE	Background electrolyte
CE-MS	Capillary electrophoresis–mass spectrometry
CKD	Chronic kidney disease
CTMOD	Control genetically modified with the FAO gain-of-function group
CTWT	Control wild-type group
ECM	Extracellular matrix
ESI	Electrospray ionization
FAO	Fatty acid oxidation
FC	Median fold change
HCA	Hierarchical cluster analysis
IS	Internal standard
JK	Jack-knifing uncertainty measures
OBSMOD	Obstruction genetically modified with FAO gain-of-function group
OBSWT	Obstruction wild-type group

OPLS-DA	Orthogonal partial least squares discriminant analysis
PBS	Phosphate-buffered saline
PCA	Principal component analysis
PLS-DA	Partial least squares-discriminant analysis
PQN	Probabilistic quotient normalization
QC	Quality control
Q C -	Quality control samples and support vector
SVRC	regression correction
RLA	Relative log abundance
RLE	Relative log expression
RSD	Relative standard deviation
TOF	Time of flight
TUS	Total useful signal
UUO	Unilateral ureteral obstruction
VIP	Variable importance in projection
WT	Wild type

#### Abbreviations related to the tested data matrix

All+QC	Complete data matrix, all samples from experimental groups, QC samples included
All-QC	All samples from experimental groups, QC samples excluded
QC	Matrix associated only with QC samples
2Gr	Matrix divided into two groups, (1) control group: CTWT and CTMOD; (2) obstruction group: OBSWT and OBSMOD
4Gr	Matrix divided into four groups: (1) CTWT; (2) CTMOD; (3) OBSWT; (4) OBSMOD

## Introduction

Chronic kidney disease (CKD) is estimated to be one of the major global health challenges. The worldwide burden of CKD has increased rapidly during the last few years, and its prevalence is still progressing. It is an independent risk factor for cardiovascular disease and one of the causes of premature mortality. Moreover, the majority of the CKD cases is asymptomatic and not diagnosed until late stages, with limited therapeutic options available [1, 2]. The state-of-the-art metabolomics technologies are considered as one of the most powerful analytical approaches with a big impact on the understanding of the molecular changes and pathogenesis of the disease. Untargeted metabolomics gives the ability to measure thousands of small molecules, also referred to as metabolites, that are important components of the cellular metabolism and provides a direct functional description of cellular activity and physiological status of the cell, tissue, organ, and entire organism. These molecular signatures have already made significant breakthroughs in discovering multiple biomarkers and predicting the progression of the disease. Therefore, untargeted metabolomics has been already widely applied to

study CKD on different biological samples originating both from *in vitro* and *in vivo* settings [3–6]. Although this approach has great potential and capacity to investigate and solve relevant clinical issues, it also represents certain challenges that need to be overcome, namely the considerable level of unwanted variation that enhances systematic bias and could lead to spurious correlations. In identifying sources of such variation associated with both biological and experimental factors, pre-analytical, analytical and post-analytical phases of untargeted metabolomics workflow should be considered [7, 8]. Experimental variation can appear from human error, within and between batch instrumental variation or variation due to the metabolite extraction protocols, that leads to a lack of precision which can significantly impact the quality of metabolomics data. It is only biological variation that reflects true variability among experimental cases that should be captured, and only such variation is of main interest to investigators. Therefore, when an unwanted biological variation is not recognized and removed or reduced, it could be easily confounded with the biological factor of interest. Some of the most common confounders are associated with the loss of sample integrity due to multiple freeze-thaw cycles or a variation of the sample weight, volume or the differences in the size or number of cells [8, 9]. Another source of unwanted biological variation is attributed to the complexity and heterogeneity of tissues and to the amount of extracted tissue, which finally get reflected in variations in metabolite concentration. Such a specific type of unwanted biological sample-driven variability is addressed by our research paper, and the amount of published studies that do not give any information about the normalization strategies employed is striking. It will be demonstrated that this fact can completely modify the biological meaning. Herein, we have studied mouse kidney tissue sample in an animal model of CKD, with and without a gene modification with protect role in CKD.

Data normalization aims to eliminate unwanted variations. It is considered a fundamental step in metabolomics data preprocessing. Several normalization methods have been already described, compared and discussed [8–11]. However, an adequate choice of a normalization method or strategy to remove overall undesired variations in a particular data set is still considered a challenging task. Many researchers have been focused on correcting technical variations, signal drift within an analytical batch or between batches [12, 13]. Some studies have compared the best strategies to normalize acquired data from urine analysis, a type of sample that possesses high natural variability in metabolite concentration mostly dependent on the organism's hydration and physiological status [14, 15]. However, there is still lack of deep evaluation of biological model-driven normalization strategies to minimize the level of unwanted biological variation associated with the analysis of tissue-derived samples, emphasizing that data normalization cannot be treated in an automated manner and it should be

carried out according to the biological model. In our study, we estimated that both factors together, CKD-associated tubulointerstitial fibrosis and the genetic modification introduced in the animal model, are prominent sources of nuisance variation in experimental results. Therefore, we are conscious that these variations might either bias experimental results, produce significant background noise or even mask differences arising from study questions and significantly affect research outcomes and data interpretation [16]. On these premises, we have performed data quality evaluation as an inherent part of our quality assurance protocol [7], following several normalization procedures that aim to reduce or eliminate unwanted variations in the data acquired by capillary electrophoresis mass spectrometry metabolomics analysis. We have applied a multilevel normalization method considering pre-acquisition and post-acquisition levels [14]. Finally, we have demonstrated that the data pre-treatment step has a significant impact on the reliability of reported results.

## Materials and methods

### Chemicals and reagents

All chemicals were of analytical or reagent grade and were purchased from Sigma-Aldrich (Germany). Reference mass solution was obtained from Agilent Technologies. Deionized water (Milli-Q) was used throughout the study (Millipore, Billerica, USA).

### Experimental design

All experimental procedures involving the use of animals were performed according to the Guide for the Care and Use of Laboratory Animals contained in Directive 2010/63/EU of the European Parliament [17]. Approval was granted by the local ethics review board of Centro de Biología Molecular “Severo Ochoa” (CBMSO) Madrid, Spain, after complying with the legal and ethical requirements relevant to the procedure for animal experimentation established by the Comunidad de Madrid and current Spanish legislation regarding employment, protection and care of experimental animals (RD 53/2013), taking into account the consideration of the 3 R and the Helsinki regulations in force for animal experimentation. The study was approved by the Committee and all procedures with animals and sample collection and pre-treatment were performed at CBMSO Center.

This study was performed in a conditional transgenic mouse model for kidney fatty acid oxidation (FAO) gain-of-function, 8-week-old wild-type (WT) and transgenic male mice ( $n = 8$  in each group) with C57BL6J genetic background subjected to 7-day unilateral ureteral obstruction (UUO), a model for tubulointerstitial fibrosis development [18, 19].

Control and obstructed kidneys were harvested after perfusion with phosphate-buffered saline (PBS). A quarter piece of each kidney sample (obtained after dissection in half both lengthwise and crosswise) was immediately frozen in liquid nitrogen and stored at  $-80\text{ }^{\circ}\text{C}$  until analysis. Kidney samples were classified into four experimental groups: control wild type (CTWT), obstruction wild type (OBSWT), control genetically modified (CTMOD) and obstruction genetically modified with FAO gain-of-function (OBSMOD). Figure 1 presents the general workflow of the presented study.

### Sample preparation

Tissue disruption and homogenization as well as metabolite extraction was carried out by following our previous protocol [20] with minor modifications [21]. In brief, cold methanol:water (1:1,  $v/v$ ) at a tissue weight to volume ratio of 1:10 was used for tissue homogenization (TissueLyser LT bead-mill homogenizer (QIAGEN, Hilden, Germany)). The weight range of the kidney tissue varied from 20 to 45 mg (CTWT, 25–45 mg; OBSWT, 19.5–39.4 mg; CTMOD, 24–42.6 mg; OBSMOD, 20.6–37 mg, respectively). The homogenization process was performed by using 2.8-mm (mean diameter) steel beads, vibrating at 50 Hz for 5 min, 4 repeated cycles with a 1-min break between cycles during which samples and TissueLyser adapter were cooled on ice. Subsequently, 100  $\mu\text{L}$  of the kidney homogenate was vortex-mixed with 100  $\mu\text{L}$  of 0.2 M formic acid, centrifuged (16,000 $\times g$  10 min,  $4\text{ }^{\circ}\text{C}$ ) and transferred to a Centrifree ultra-centrifugation device (Millipore Ireland Ltd., Cork, Ireland) with 30-kDa protein cutoff filter for deproteinization through centrifugation (2000 $\times g$ , 70 min,  $4\text{ }^{\circ}\text{C}$ ). The filtrate was then transferred to the chromacol vial, evaporated to dryness using a SpeedVac Concentrator (Thermo Fisher Scientific, Waltham, MA, USA) and resuspended in 50  $\mu\text{L}$  of 0.1 M formic acid containing 0.2 mM methionine sulfone (IS). Before the analysis, the samples were centrifuged at 4000 $\times g$  for 20 min at  $4\text{ }^{\circ}\text{C}$ .

Quality control samples (QC) were prepared by pooling equal volumes (10  $\mu\text{L}$ ) of each homogenized tissue following the same procedure applied for the experimental samples. Extraction blank samples (without analyte of interest) were also considered for their preparation. All experimental samples were randomized before sample preparation and their analytical run.

### Capillary electrophoresis-time of flight-mass spectrometry (CE-TOF-MS) analysis

The experiment was carried out using a 7100 capillary electrophoresis system (Agilent Technologies) coupled to a 6224 TOF Mass Spectrometer (Agilent Technologies), equipped with an electrospray ionization source (ESI). A fused-silica capillary (Agilent Technologies; total length, 96 cm; i.d.,

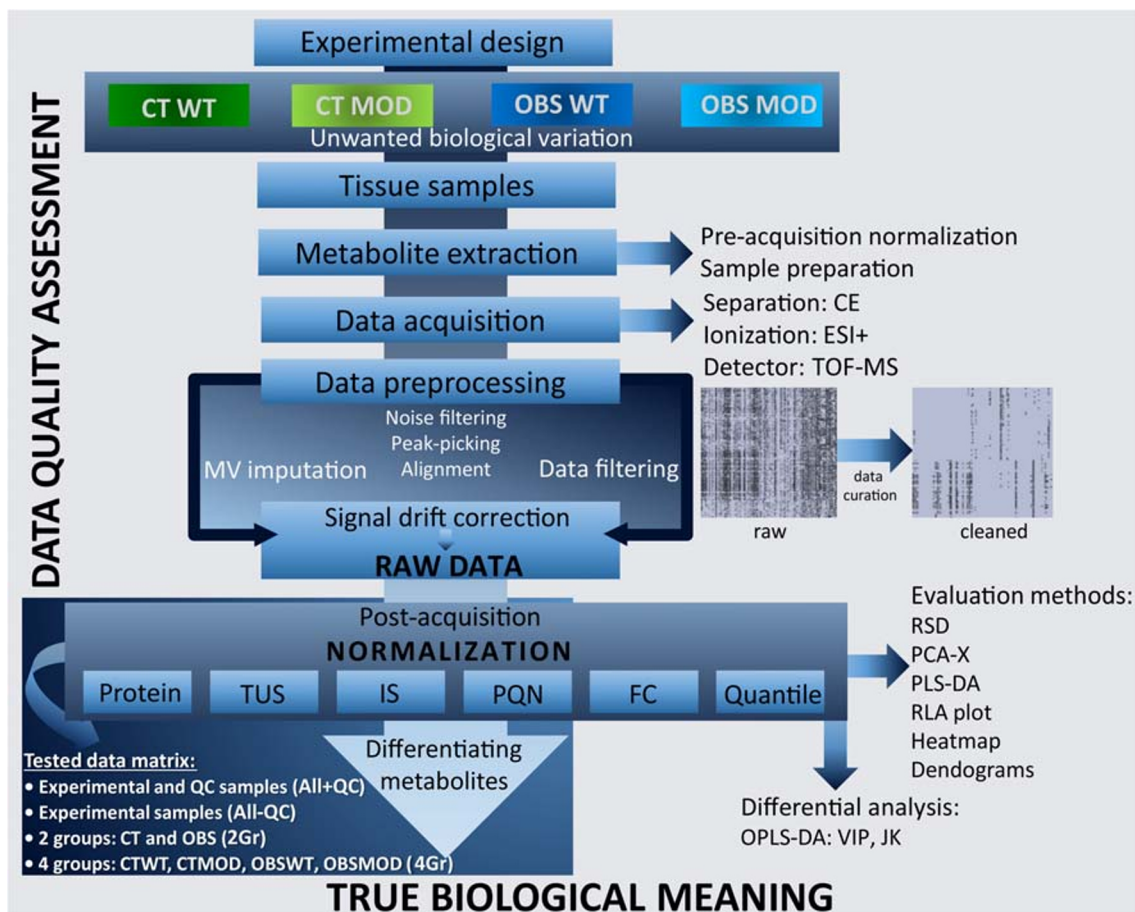


Fig. 1 Flowchart illustrating the general workflow of the study

50  $\mu\text{m}$ ) was used for metabolite separation. Before each analysis, the capillary was flushed for 5 min with background electrolyte (BGE) containing 1 M formic acid solution in 10% methanol (*v/v*). Sample injections were performed over 50 s at 50 mbar, and to improve the reproducibility of the analysis, the BGE was injected for 20 s at 100 mbar after the injection of each sample. The sheath liquid consisted of methanol/water (1/1, *v/v*), formic acid (1.0 mM) and two reference masses: purine (*m/z* 121.050873) and HP-0921 (*m/z* 922.009798). Flow rate was 0.6 mL/min and split was set to 1/100. The separation was performed at a pressure of 25 mbar and a voltage of + 30 kV, in positive ionization mode. The total time of the analytical run was 23 min. The MS was operated in positive polarity, with a full scan range from *m/z* 70 to 1000 at a rate of 1.36 scan/s. The drying gas was set to 10 L/min, nebulizer to 10 psi, voltage to 3.5 kV, fragmentor to 125 V, drying gas temperature to 200 °C and skimmer to 65 V. Data acquisition was performed with Mass Hunter Workstation Software (Ver. B.06.01, Agilent Technologies) [21].

All samples were randomized before the analytical run. Blank samples were analysed at the beginning and at the end of the worklist. To achieve fully reproducible conditions, QC samples (QCs) were used to equilibrate and condition the

system, then QCs were analysed after every 5 samples to measure the stability and performance of the analysis. Quality control and quality assurance procedures were applied according to published guidelines [7]. A representative total ion electropherogram obtained from the CE-MS analysis in positive ionization mode of a kidney-derived QC sample is presented in Electronic Supplementary Material (ESM), Fig. S1.

## Data processing

### Data pre-processing

MassHunter Qualitative analysis (Ver. B.08.00, Agilent Technologies) was used to examine the quality of acquired electropherograms. Raw data were pre-processed with Agilent MassHunter Profinder Software (Ver. B.08.00, Agilent Technologies) which performs batch molecular feature extraction on each data file that reduces data complexity by removing redundant and non-specific information and identifying important features (variables) associated with data. Related co-eluting ion signals (isotopes, common adducts, detected dimers or those with neutral loss of water) were summed and grouped into one metabolic feature. Features

were aligned (by mass and migration time) across all the samples to create an average consensus spectrum for each compound group, enabling a recursive re-extraction of the batch data files to eliminate false-positive and false-negative results. The data matrix was extensively cleaned by manual inspection of the quality of each metabolic feature, including peak area and migration time integration with Mass Hunter Profinder Software (Ver. B.08.00, Agilent Technologies). The raw data matrix consisted of 4 extraction blanks, 6 QCs and 32 experimental samples with the 334 aligned metabolic features.

### Data pre-treatment

The obtained data matrix was imported into Microsoft Excel (Microsoft Office 2016) for further calculations. Blank subtraction and removal of detected salt clusters was applied to eliminate irrelevant information, and the curated data matrix, with 199 reported features, was used for further evaluation. In the data pre-treatment workflow, metabolic features detected in less than 50% of QC samples are considered to be excluded from the data matrix. That was not the case for our data, as we did not observe any missing values in analysed QC samples. Systematic variation of the instrument's response was evaluated by plotting the sum intensity of all metabolic features over the acquisition time as well as considering the group index, as shown in ESM, Fig. S2. Unsupervised principal component analysis (PCA) was applied to detect patterns, trends and outlying observations according to Hotelling's T2 Range Plot (SIMCA-P+ 15.0, Umetrics, Umea, Sweden). A presented PCA scatter plot (ESM, Fig. S2) represents a sample that is outside of the confidence ellipse range, and indeed, it was identified as a strong outlier (T2 range value > 99%) and removed from further data treatment. Inspection of the original electropherogram additionally confirmed that the metabolome of the indicated sample was significantly different as compared to others. Quality control samples, and support vector regression with a radial basis function kernel (QC-SVRC) was applied to correct analytical signal drift, as described elsewhere [22]. The  $\varepsilon$ -insensitive loss parameter was selected for each metabolic feature as 1.5% of the median peak area value in QCs. The error penalty  $C$  was expressed as a percentile of intensity of QCs ( $C = 50$ ), the kernel parameter  $\gamma$  was in interval log-space (0, 3, 20) and  $k$ -fold cross-validation (10-fold CV) was applied. QC-SVRC function used in this work was performed using MATLAB scripts (Matlab R2015, Mathworks) kindly provided by Dr. Julia Kuligowski (Neonatal Research Unit, Health Research Institute Hospital La Fe, Valencia, Spain).

The RSD (expressed as a percentage) for each metabolic feature present in QCs was calculated, and the metabolic features with RSD in QC > 20% were considered to filter out. The RSD values in QCs for metabolic features in our data

did not exceed 20%. Missing values were imputed by  $k$ -nearest neighbours' algorithm ( $k = 3$ ).

### Metabolite identification

For tentative identification, the accurate  $m/z$  of the metabolic features were searched against online databases as Kegg, Metlin, LipidMaps and HMDB using the advanced CEU Mass Mediator tool [23]. Matched compounds were identified by using the accurate mass and isotopic distribution. An in-house-developed CE-MS standards library was used to compare relative migration time if data were available.

### Normalization strategies

This paper aims to evaluate representative post-acquisition normalization methods, with a different statistical or chemical base, namely normalization by (1) total protein content [24], (2) total useful signal [25], (3) internal standard [26], (4) probabilistic quotient normalization [27], (5) median fold change [15] and (6) quantile normalization [28], which are one of the most widely applied to mass-spectrometry-based untargeted metabolomics data.

A pre-acquisition normalization considering the weight of samples was applied to adjust individual dilution factors by which samples are diluted to a common concentration regarding mass/volume. The analytical variability was corrected using quality control samples and support vector regression (QC-SVRC) strategy for the signal drift [22]. Six different post-acquisition normalization methods specified herein were applied to CE-MS data to eliminate or reduce the remaining analytical variability and the unwanted model-driven biological variation.

- Normalization by total protein content, a method based on the measurement of the total amount of protein in each sample. The area of each peak in each sample was divided by the amount of protein measured in that sample [24, 29, 30]. The protein was measured by bicinchoninic acid assay (BCA) [31]. The general advantage of this strategy is the high linear correlation between the metabolite abundance and protein amount in the tissue, which permits comparisons between samples with different cellular masses. However, this method is not efficient when we consider highly heterogeneous samples like in our case, samples with fibrosis with a high amount of connective tissue in which proteins are not measured in the BCA assay. In such case, less cellular mass is considered; however, the weight of the samples is comparable. As a general recommendation, normalization by the total protein content is not preferred for metabolomics studies because of the relatively large errors introduced associated with poor protein recovery in the solvents used for metabolite

extraction and incomplete protein re-solubilization from the pellet [29, 32].

- Total useful signal (TUS) assumes that changes in the metabolic signals are stable across the data set which therefore forces all samples to have equal total intensity. The TUS normalization uses the total abundance of each metabolic feature that is present in all samples as the normalization factor. The abundance of each metabolic feature is divided by this factor in a given sample [11, 25]. Considering statistical assumptions of TUS normalization, large changes in the peak intensity of high-concentrated metabolites will have a substantial contribution to total peak intensity and can compromise remarkably this normalization procedure. The validity of this approach is questionable, as an increase in one metabolite concentration may not necessarily be accompanied by a decrease in another [10]. Despite that, it is a simple normalization method that has been commonly used in metabolomics studies.
- Internal standard (IS) normalization is based on a known compound added to each biological sample. The area of this compound in each sample is used to normalize the feature signals in that sample. The variation captured by IS normalization depends on its chemical properties, and such a strategy of data normalization is not able to handle unwanted biological variability, only useful to remove analytical variation [11, 16, 26].
- Probabilistic quotient normalization (PQN) method is based on the calculation of the most probable dilution factor by comparing the distribution of quotients between samples and a reference spectrum, followed by sample normalization using such dilution factor. For this strategy, it is fundamental to proper selection of the reference spectrum, and that should be dependent on the data-driven case. It is a robust method, although it presents some limitations when there are essential differences between experimental groups. The PQN method is considered as the optimal for multidimensional data sets [9, 27, 33].
- Median fold change (FC) is similar to TUS, in the assumption that measured peak intensities are directly proportional to concentrations of metabolites in solution. However, the influence of high-intensity metabolites is reduced because FC considers that the intermediate-intensity metabolites are those that will be constant across the data set. The method adjusts the median of log FC of peak intensities between samples in a set of experiments to be approximately zero. FC is more useful and practical than TUS when saturated metabolite abundances are related to the interest factor and has relaxed assumption with regard to the proportion of asymmetrical metabolite changes [9, 15, 34].
- Quantile normalization considers the metabolic feature peak intensity-dependent scaling factor and transforms

the intensity distributions of variables to be equal between all samples in a data set. Therefore, all the samples will have the same intensity values, although distributed otherwise according to the different variables. It is considered as a simple and effective method to reduce systematic variation, revealing the biological variation [27]. However, this approach can pose some problems with high-intensity values with considerable changes between samples [28, 34].

Further details and comprehensive information on the abovementioned normalization methods can be found in [9–11, 26, 27, 29, 35, 36].

In the case of median FC, PQN and quantile normalization, several additional strategies referred to the model-driven information were evaluated. This strategy assumed the differences in metabolite concentrations between specified cases and in our previous observations where data matrix composition could affect the results of normalization. Therefore, the aim was to evaluate the performance of selected normalization methods in the settings where (i) all analysed samples were considered simultaneously (All+QC); (ii) QCs samples were excluded from the data matrix (All-QC); (iii) the normalization was performed separately considering 2 groups, one with all control cases (CTWT and CTMOD) and the other with all cases with renal obstruction (OBSWT and OBSMOD); and (iv) the last strategy treated all experimental groups in a separate manner and the normalization was applied independent for each one of four experimental groups (CTWT, CTMOD, OBSWT, OBSMOD).

The same strategy was applied to select the reference spectrum in case of PQN normalization. Taking into account heterogeneity of our experimental groups, we tested different procedures that refer to the reference spectrum selection. Therefore, for the data matrix All+QC as reference, we used the distribution of metabolites across all experimental and QC samples; in the case of the data matrix All-QC, QC samples were not counted for the reference spectrum; for matrix QC, the distribution of metabolites in the QC sample was considered as the reference spectrum; in case of 2Gr, the reference spectrum was chosen separately from control cases (CTWT and CTMOD) and groups with obstruction (OBSWT and OBSMOD); in the last case 4Gr, the reference spectrum was based on the individual sample distribution in each experimental group CTWT, CTMOD, OBSWT and OBSMOD. In consequence, this work not only addresses the evaluation of the distinct normalization algorithms but also tackles the assay problems associated with the biological model.

### Methods to evaluate normalization performance

Multiple quantitative and qualitative statistical approaches were considered to choose the optimal normalization strategy,

including (i) RSD calculations; (ii) RLA plots; (iii) PCA and PLS-DA multivariate analysis; and (iv) HCA-heatmap plots. Multivariate statistical methods based on orthogonal projections to latent structures discriminant analysis (OPLS-DA) were further applied to identify metabolites with a differential presence between specified interpretations. Differential metabolites were considered significant when  $VIP \geq 1$  with jack-knifing (JK) confidence intervals at the 95% level. Multivariate analysis applied sevenfold cross-validation of the model. Relative standard deviation ( $RSD\% = \text{standard deviation}/\text{mean} \times 100\%$ ) was calculated to characterize measurement variability and was used to estimate data dispersion among different normalization methods. Such calculation of data uncertainty was proposed by Persons et al. [37] as a practical benchmark for metabolomics studies. Intragroup relative log abundance RLA plots were obtained by standardizing each metabolite by subtracting the median from each metabolite within each group. The scaled variables are illustrated as boxplots and demonstrate a median centred at zero and low variability [16].

## Results and discussion

There is a relationship between a specific biological sample and the unwanted biological variation affecting samples. Because normalization methods cannot be used with default parameters and performed automatically, it is necessary to study each biological model carefully and evaluate the suitable fit for the corresponding data pre-treatment strategy. Normalization is one of the most important procedures for handling untargeted metabolomics data and can completely transform their value and biological meaning as illustrated in Fig. 2. Data normalization can correct aspects that hinder the biological interpretation only when its application is driven by a deep knowledge and serious concern on the total unwanted variation originating from uninduced biological and analytical variation.

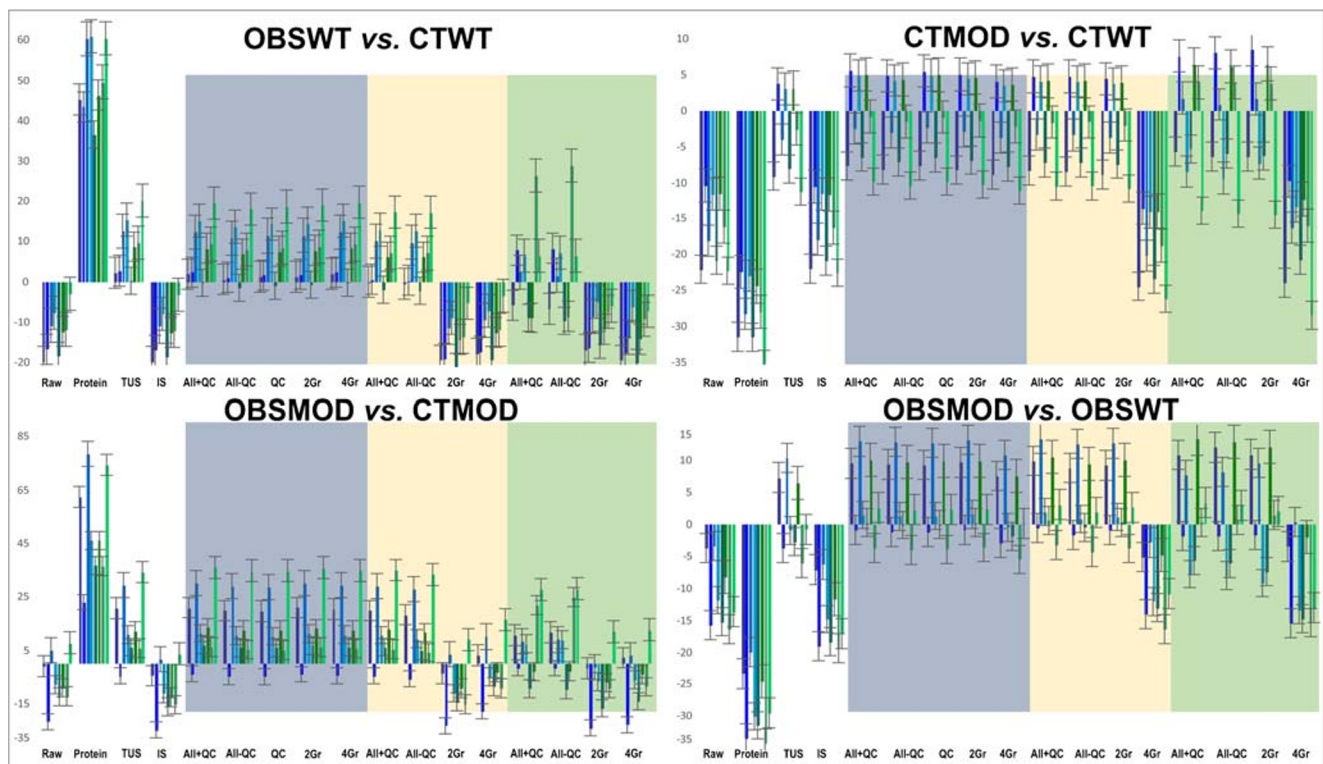
In Fig. 2, it can be observed that the magnitude of the signal for a set of selected metabolites distributed along the profile and their regulation (positive or negative) significantly vary according to different normalization strategies applied. Data normalization not only modifies the level of changes, as expected, but also, what is more important, it can also completely modify the directions from positive to negative, which has a definitive influence in the interpretation of the results. In this research paper, we have proven that in the biological model studied, data normalization by the total amount of protein, a common strategy, is not suitable. That is illustrated in Fig. 2, especially in the comparisons where obstruction-associated changes are addressed (OBSWT vs. CTWT and OBSMOD vs. CTMOD). We have observed cases where some compounds could be up- or downregulated depending on the

chosen data normalization strategy. Selected metabolites like glutamic acid, tryptophan, hypoxanthine and uracil have been described in the literature to be related to kidney disease [38]. The others like dimethylallyl pyrophosphate, L-asparagine, 2-aminoadipic acid and ValPhe were chosen randomly but covering the whole profile. In the cases of normalization based on protein content, TUS or PQN, those metabolites were increased while in raw data they were decreased. We could also see that this effect was data matrix-dependent and it did not normalize equally when all experimental samples together with or without QCs (All+QC or All-QC) were considered.

Following the idea of a sequential normalization recently reported by Gagnebin et al. [14], we adapted the proposed strategy to the specific case described in this study, with the aim of removing or reducing unwanted variability related to the analytical signal drift and the observed sample concentration fluctuations. Pre-acquisition sample normalization was performed based on the tissue weight and a specific dilution factor for each sample to level concentrations. After acquisition, raw data were evaluated according to quality assurance protocols to estimate the quality of the analytical procedure (ESM, Fig. S2). Analysis of QC samples indicated a slight level of analytical variations arising from signal drift which was then corrected by QC-SVRC algorithm, commonly encountered in untargeted metabolomics workflows [22]. Furthermore, the drift in the response was evaluated before and after applying the correction on selected metabolites (spermidine, tryptophan, guanidinoacetate and carnitine) that have been previously described to be altered due to the onset and evolution of renal fibrosis (ESM, Fig. S3) [38]. Application of that method of data normalization decreases technical variance, increases the number of metabolic features that meet the QC criteria and, what is more important, enhances overall data quality without affecting true biological variability.

## Evaluation of normalization performance

Relative standard deviation (RSD) was applied to compare the uncertainty level between different measurements for each dataset as well as for each experimental class. These results are summarized in Fig. 3. RSD is considered a good measure of the reliability of these results, where the lower the RSD value, the higher the data reliability. According to the distribution of RSD values, there is no doubt that both normalization algorithms and the data matrix composition itself affect overall data dispersion. It is especially evident in the case of data normalization by protein content and by an internal standard (methionine sulfone). When we compare calculated total average RSD values, a reduced RSD is observed after PQN, median FC and quantile normalization. Furthermore, considering the experimental sample classification, we can observe differences in inter-individual variations between the cases



**Fig. 2** Differences in changes of direction of selected metabolites ■ dimethylallyl pyrophosphate, ■ L-asparagine, ■ 2-aminoadipic acid, ■ L-glutamic acid, ■ L-tryptophan, ■ ValPhe, ■ hypoxanthine and ■

uracil, for each normalization method (lavender, beige and green panels represent the results of PQN, median FC and quantile normalization respectively) and specified interpretation

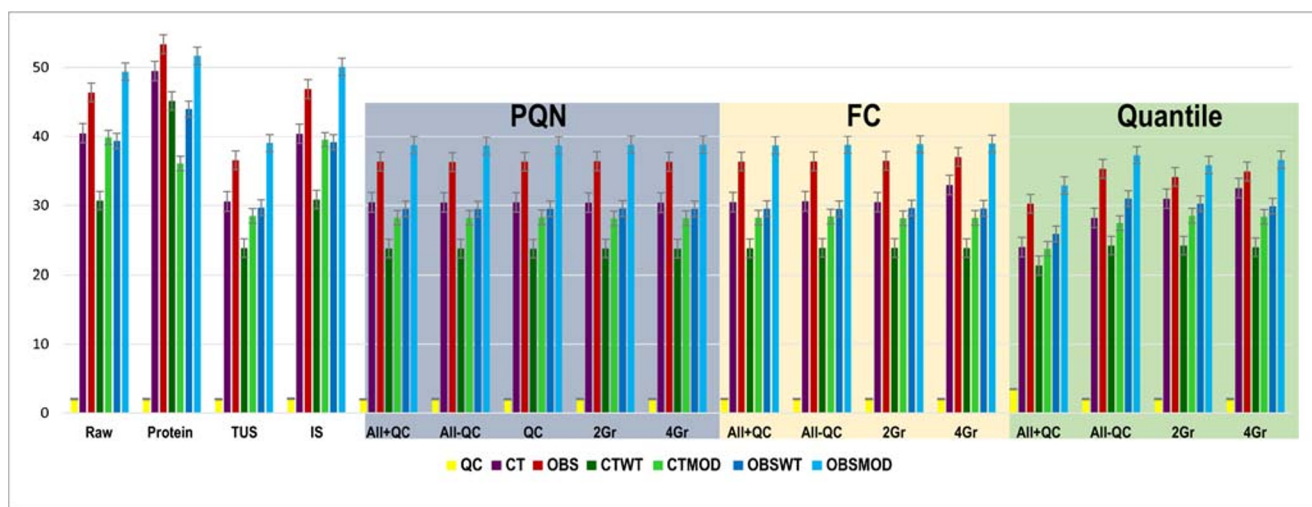
with kidney obstruction and controls, where RSD in OBS > CT. In these cases, it is remarkable that variance is even higher in CT and OBS cases with genetic modifications. The general trend referred to the observed variance in RSD according to the experimental group order: CTWT < CTMOD < OBSWT < OBSMOD could be highlighted and is illustrated in Fig. 3. That observed trend does not depend on the type of the normalization algorithm or the normalization strategy used, except when normalization by protein content is applied. The results of normalization by protein content, shown in Fig. 2 (changes in the observed regulations between cases and control) and Fig. 3 (changes in the pattern of RSD according to the group), indicate that it is necessary to take into account the biological model. It is important to highlight that considering analysis of tissue, which is subject to superior homeostatic control than other biofluids, we could expect relatively low inter-individual deviations. However, metabolic variability increases when the tissue exhibits considerable heterogeneity. Kidney fibrosis is characterized by the loss of renal cells and their replacement by excessive formation and deposition of extracellular matrix (ECM) proteins, mostly collagen, in the kidney interstitium, resulting in structural damage, loss of renal function and end stage of chronic kidney disease [39]. Thus, this tissue sample will be prone to manifest significantly higher biological variation [37]. Moreover, in the same amount of sample, cells are replaced by collagen which is

not measured in the protein assay. Therefore, when the samples from fibrotic tissue are divided by lower amounts of protein to normalize, results are biased towards higher abundances. This observation is not biologically real because the total weight of the kidney does not change. Although unwanted biological variation, especially related to the specificity of tissue-derived samples, should raise serious concerns, this issue remains underestimated and is rather poorly or not addressed at all in many untargeted metabolomic experiments.

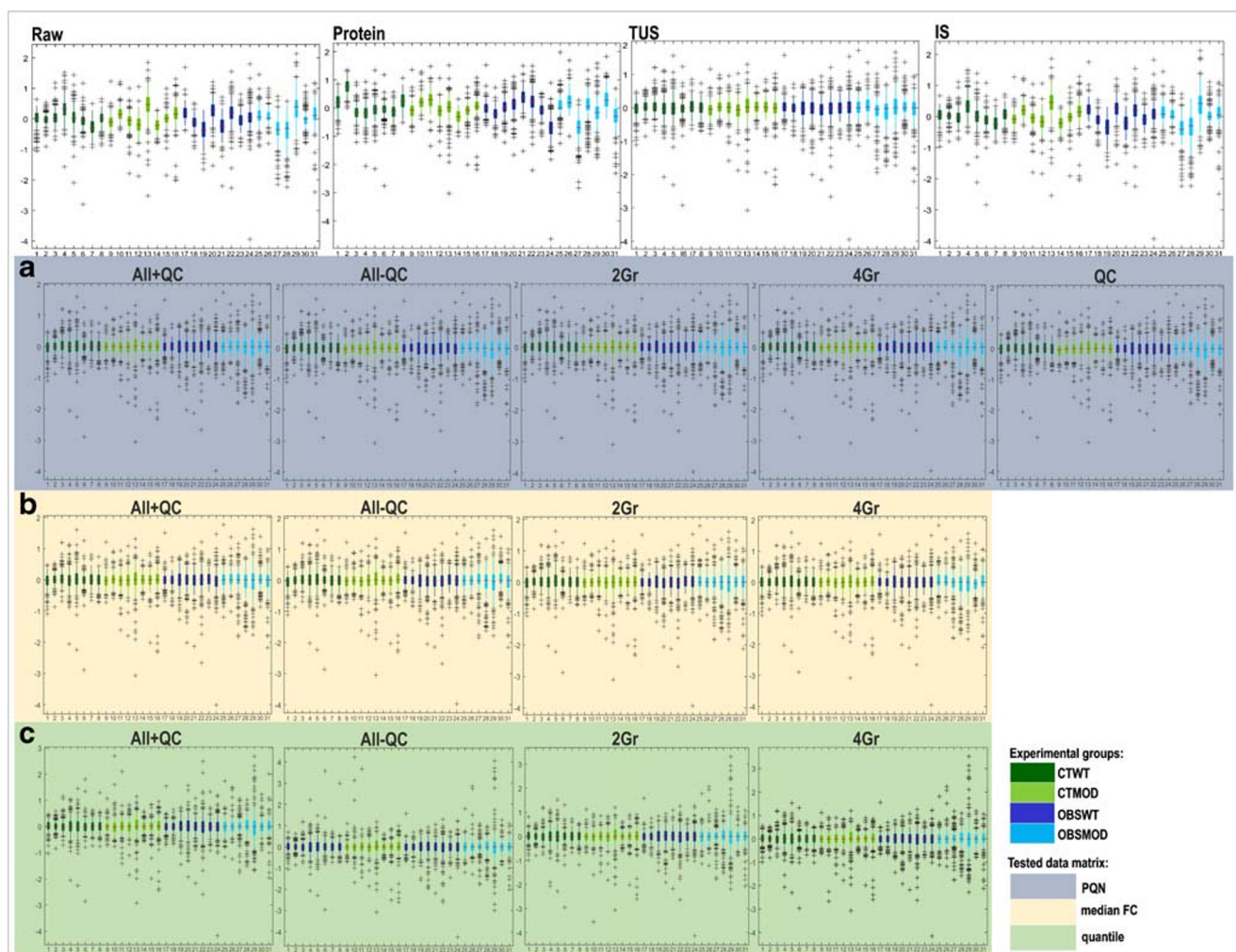
Moreover, this fact is not only related to the use of an untargeted metabolomics strategy, as it could affect in the same way to any targeted analysis where proteins are used to normalize results.

We use within-group relative log abundance (RLA) plots (Fig. 4) to reveal unwanted variations in experimental data. RLA plots are a powerful tool with a great ability to detect and visualize unwanted variation in a metabolomics data matrix. As recommended by De Livera et al., RLA plots are particularly useful to assess whether a normalization procedure has been successful [16]. Recently, Gandolfo et al. have provided a detailed examination and critical discussion about the relative log expression RLE plots, performed as RLA and their application in high-dimensional data [40]. We have observed that the RLA plot for raw data shows substantial unwanted variation between samples. This variation is still visible when data are normalized by protein content and an internal





**Fig. 3** Distribution of total average RSD values expressed in percentage (%) calculated for QC, CT, CTWT, CTMOD, OBSWT and OBSMOD data following different normalization strategies. The data indicates the impact of different normalization on the metabolite variation



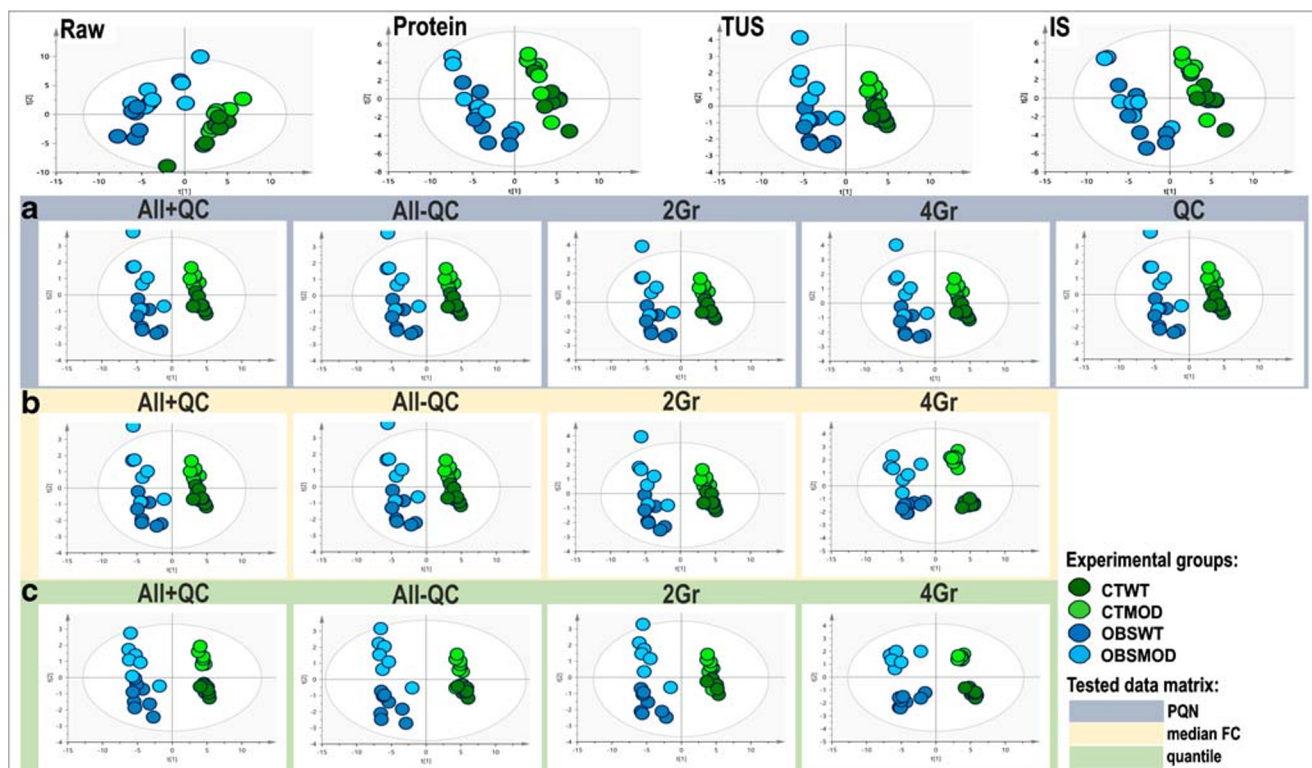
**Fig. 4** Within-group RLA plots of the raw data (Raw), data normalized by protein content (Protein), data normalized by total useful signal (TUS), data normalized according to internal standard (IS). **a** Lavender panel, **b**

beige panel and **c** green panel represent RLA plots obtained after PQN, median FC and quantile normalization, respectively, together with the evaluation of data normalization performance on the specified data matrix

standard, while it is reduced for matrix normalized by the total useful signal. Normalization with PQN, median FC and quantile successfully removed observed variations, leading to tight clustering of biological replicates. Nevertheless, Gandolfo et al. pointed out that RLA plots only give strong evidence that a normalization procedure has failed, but they are not capable of diagnosing whether the factor of interest has been retained [40]. Therefore, further careful evaluation of reported data reliability needs to be taken into consideration.

Multivariate models, principal component analysis (PCA) and partial least-squares-discriminant analysis (PLS-DA) were constructed for further evaluation of normalization strategies. Unsupervised PCA approach is commonly applied to reduce data dimensionality and to extract relevant information from a given data set. Figure S4 (see ESM) shows the resulting PCA score plots for each normalization strategy applied to this study. In all models, it can be seen that the first component of the sample scores captures the variation associated with kidney obstruction. Considering the disease severity and fibrotic remodelling of the kidney tissue, such differences are expected. Nevertheless, no clear separation attributed to gene modification was observed. In addition, an interesting performance was observed on PLS-DA-based analysis (Fig. 5). Herein, the  $X$ -matrix is connected to the  $Y$ -data matrix characterizing class membership considered following the

classification: CTWT—control wild type, CTMOD—control genetically modified, OBSWT—obstruction wild type and OBSMOD—obstruction genetically modified. This method aims to maximize the covariance between the independent variables (metabolic features) and the corresponding dependent variables (classes) by finding a linear subspace of the independent variables. The first component from all PLS-DA models indicates the discrimination between the cases with renal obstruction (OBSWT and OBSMOD) and control group (CTWT and CTMOD). Moreover, a clear tendency to separate the experimental groups due to the gene modification can be observed except in raw data, data normalized by protein content and an internal standard. For other cases, the second component was able to discriminate between control samples and those with genetic modification (CTWT vs. CTMOD). We have already commented the higher variation associated to obstruction and gene modification that hinder data interpretation. Indeed, in most cases, analysis cannot discriminate samples from the OBSWT and OBSMOD groups. Only the PLS-DA models derived from the median FC and the quantile normalization method performed for each of experimental group (CTWT, CTMOD, OBSWT, OBSMOD) separately were able to reduce within-group variation. As a consequence, we were able to capture between-group variation and moreover, what is particularly interesting, the variation



**Fig. 5** Supervised PLS-DA comparing normalization strategies for raw data (Raw), data normalized by protein content (Protein), data normalized by total useful signal (TUS) and data normalized according to internal standard (IS). (a) Lavender panel, (b) beige panel and (c) green panel

represent PLS-DA plots obtained after PQN, median FC and quantile normalization, respectively, together with the evaluation of data normalization analysis based on the specified data matrix

attributed to gene modification. Such strategy of normalization results in not only increased homogeneity of the within-group data but also a benefit on the more efficient experimental cases discrimination. Both median FC-4Gr and quantile-4Gr PLS-DA models were characterized with a significant CV-ANOVA  $p$  value ( $5.6E-10$  and  $2.74E-10$ , respectively), good model parameters for explained variation ( $R^2$  0.6 and 0.8, respectively) and variation predicted ( $Q^2$  0.5 and 0.6, respectively). It is important to highlight that the PLS-DA model based on the 4Gr quantile normalization strategy demonstrates the shortest distance between samples within the same experimental group (reduced within-group variation) and the greatest separation between samples from different groups (enhanced between-group variation). The heatmaps presented in ESM, Fig. S5, based on relative signal intensities of the metabolic features, confirm the observations derived from the interpretation of PLS-DA models. Moreover, the misclassification between control and obstruction cases could be noticed in the case of a heatmap that represents raw data, protein and internal standard normalized data.

The rapid emerging and global interest in untargeted metabolomics research caused the difficulties with the adaptation of analytical procedures and challenges for large-scale data analysis. There was a lack of standardized protocols and consensus on the general metabolomics workflow. Therefore, some issues especially related to metabolomics data processing, e.g. sample normalization, have been sometimes ignored. Normalization according to TUS was one of the most common methods applied. Although TUS is still in use, many researchers recognize that it can pose problems in case of samples with considerable differences in their concentrations. Some authors, e.g. Filzmoser & Walczak or Walach et al., focus to evaluate the performance of TUS normalization, among other methods in terms of the size effects that refer to the samples that vary in the range of concentration [41, 42]. Indeed, those studies and other authors [10, 26] are aware that TUS normalization, if not properly justified, can lead to misleading conclusions. The sample concentration as, e.g. samples derived from the tissue, can vary from one to another, and the amount of metabolites depends on several factors as sample weight, cell density or the number of cells. Therefore, the selection of a normalization method and normalization strategy should depend on the type of biological sample and should also assume the size effect. Over the past few years, we have seen a great effort to evaluate normalization protocols, especially for the analysis of urine samples. However, there is still an urgent need to evaluate normalization for highly heterogeneous samples like tissue that could have a different sample-to-sample total amount of metabolites. The assumptions of TUS normalization could be violated in a specific sample set, especially when comparing normal tissue with cancerous or tissue with fibrotic changes.

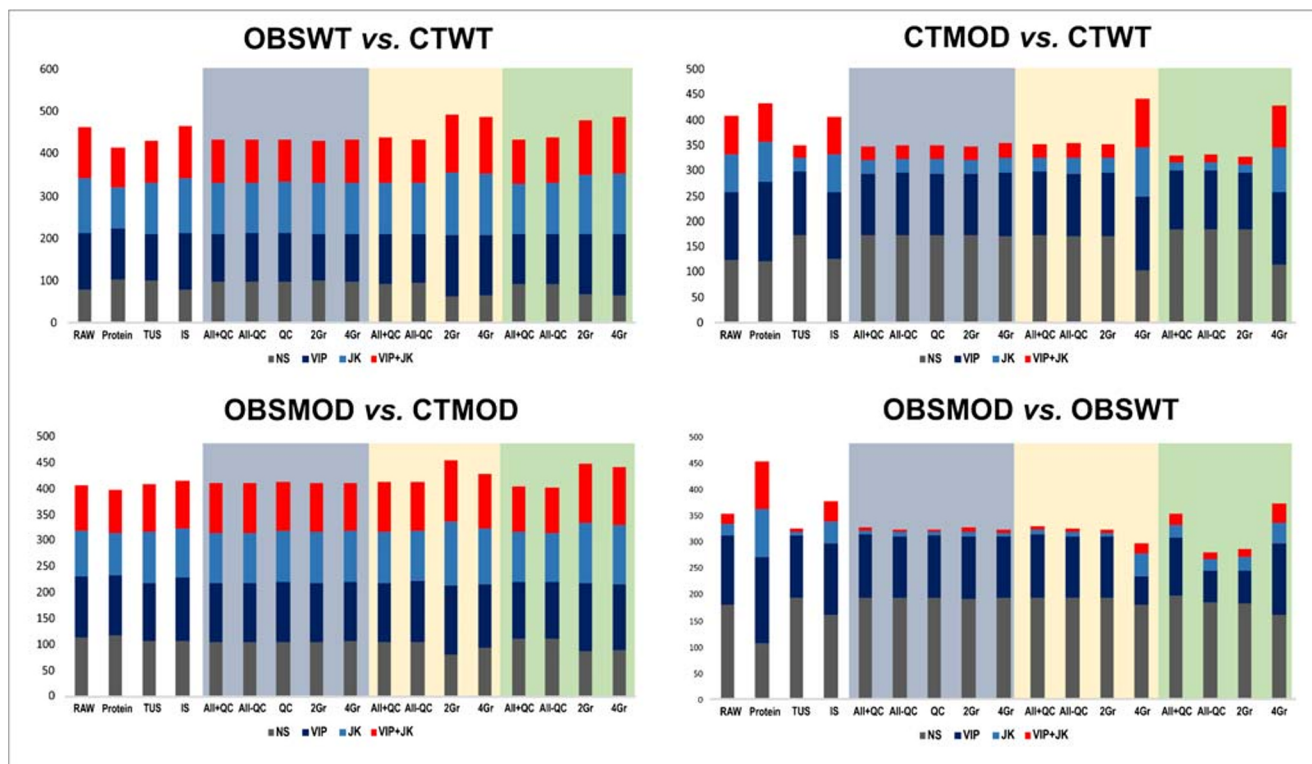
The questions arise on how to proceed with normalization. This approach in such cases is not straightforward and requires careful evaluation.

Furthermore, previous studies support the idea of group-dependent normalization. Those studies, e.g. Paulson et al. or Hicks et al., emphasize that most normalization methods generally make assumptions that are valid in consistent samples; however, those assumptions are violated in heterogeneous data sets, such as tissue samples. In those cases, conducting normalization is more challenging, even when comparing related samples as each of them may have different metabolite concentrations. The authors claim that in such cases global normalization methods have the potential to remove biologically driven variation [43, 44]. Therefore, Hicks et al. proposed an alternative strategy called smooth quantile normalization, based on the assumption that the statistical distribution of each sample should be the same within biological groups, but allowing that they could differ between groups. Moreover, the authors emphasize that the proposed method does not require any external information other than sample group assignment and it is not specific to one type of high-throughput data [44].

## Differential multivariate analysis

Orthogonal partial least-squares discriminant analysis (OPLS-DA) was used to assess the variance in each specified interpretation (OBSWT vs. CTWT, CTMOD vs. CTWT, OBSMOD vs. CTMOD, OBSMOD vs. OBSWT). Models were constructed with one predictive and one orthogonal component, and a detailed description of this analysis including quality parameters of the model ( $R^2$ ,  $Q^2$  and  $p$  value) is presented in ESM, Tab. S1. Estimation of levels from statistically significant data was measured by VIP ( $\geq 1$ ) and JK (95% confidence) uncertainties calculated for each data matrix and is presented in Fig. 6. It is not surprising that the most prominent differences were observed in the case of levels derived from metabolic changes provoked by the obstruction (OBSWT vs. CTWT). Both the quality of OPLS-DA models and the number of significant compounds were comparable in all normalization strategies.

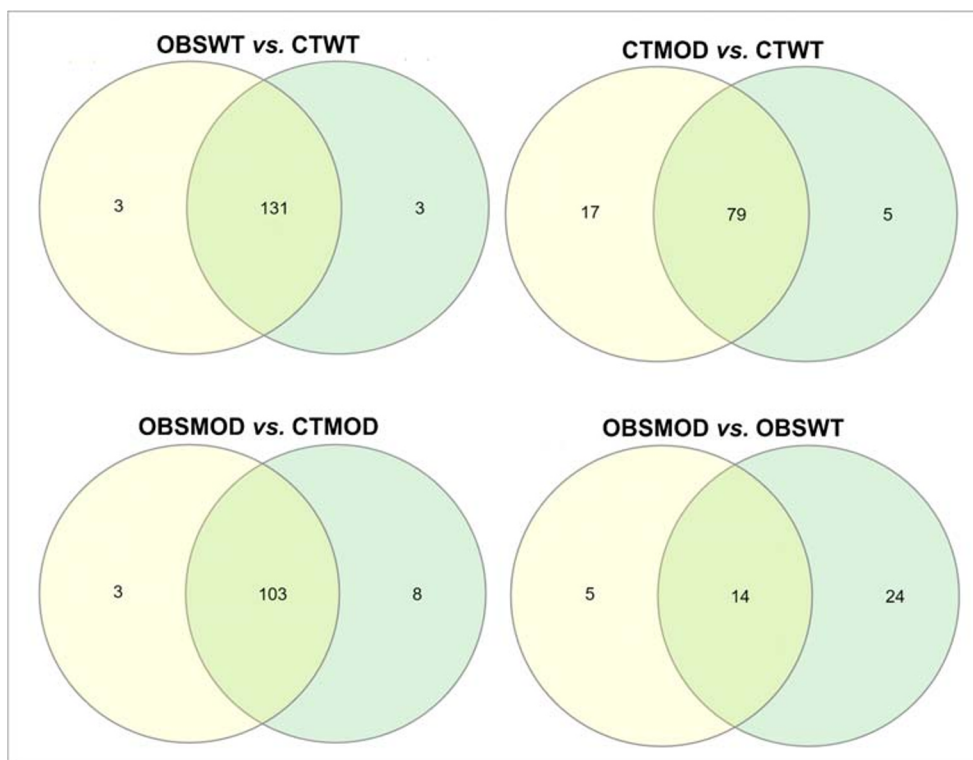
Addressing the effect triggered by the gene modification on the control group (CTMOD vs. CTWT), when normalization was performed by median FC and quantile with respect to the experimental group (4Gr based) showed better outcome. The model constructed for OBSMOD vs. CTMOD, in which both cases display gene modification, and when control cases are compared with their associated obstruction cases give comparable results. One of the most challenging tasks was to evaluate the effect of the gene modification on the metabolite profile in the obstruction cases, OBSMOD vs. OBSWT. Although all the OPLS-DA models present a large  $R^2$  value, it cannot be considered reliable due to the poor predictive



**Fig. 6** Cumulative chart displaying the total count of metabolic features reported as non-significant (NS), significant according to VIP, JK and the combination of VIP + JK with respect to the normalization strategy tested.

The grey colour indicates non-statistically significant data, dark blue data with  $VIP \geq 1$ , light blue data passing JK uncertainty level and finally, the level of statistically significant data ( $VIP \geq 1$  with JK) is marked on red colour

**Fig. 7** Venn diagram representing the number of statistically significant features, considering the median FC (beige) and quantile (green) normalization methods (4Gr based)



(negative) Q2 parameter and, what is important, due to the lack of modelling significance for the observed group separation according to CV-ANOVA.

However, a remarkable improvement of the overall model (OBSMOD vs. OBSWT) quality and the satisfactory performance with CV-ANOVA  $3.4E-02$  and  $9.4E-03$  was achieved with the data matrix normalized by median FC and quantile (4Gr based) with explained variation  $R^2$  0.87, 0.95 and variation predicted Q2 0.62, 0.72, respectively. Although there is a similarity in the quality of both models, we could observe the higher number of statistically significant features reported when the quantile normalization was applied. Moreover, the superior output of median FC and quantile normalization (4Gr based) over the other strategies could be observed in the case of the predictive component extracted from the OPLS-DA model for the interpretation OBSMOD vs. OBSWT, focusing into the variance which is critical for the defined group separation (ESM, Fig. S6). Both normalization strategies give a comparable number of statistically significant features which could be seen in Fig. 7. However, the fact of a better statistical performance was reported when the model related solely to the effect of gene modification in the cases of renal obstruction (OBSMOD vs. OBSWT), let us to conclude a slightly better overperformance of quantile (4Gr based) normalization over median FC.

Further, cross-validated OPLS-DA models performed for quantile normalized data (4Gr based) gave another evidence on data reliability. Therefore, we can conclude that this strategy is the best choice for this specific model-driven metabolomics data set adjustment (ESM, Fig. S7). Finally, it is worthy to comment that the description of compounds and their biological interpretation is beyond the scope of this paper.

## Conclusions

A biological model-driven normalization strategy can significantly decrease unwanted analytical and biological variation. It improves overall data quality and facilitates sample stratification due to the biological factor of interest. Our data proved that the performance of median fold change and quantile normalization was similar; however, it is a substantial matter to consider the data matrix composition. Data normalization is a critical step in the analytical workflow with a fundamental problem of data integrity and reliability behind. The normalization strategy should be compatible with experimental design and overall research purpose. Most normalization methods perform well in consistent samples, but highly heterogeneous data as it could be in case of tissue-derived samples pose a serious problem. Unwanted biological variations still seem to be underestimated and its evaluation in a given biological model or specific case should be considered

adequately. Data normalization is a concern not only in untargeted metabolomics studies but also in targeted strategies where protein normalization is frequently used. Therefore, minimum reporting criteria should include the normalization method chosen because its lack might contribute to incorrect data pre-treatment and misleading biological interpretations.

**Acknowledgements** This work was supported by Comunidad de Madrid (B-2017/BMD-3751 “NOVELREN-CM”), Ministerio de Ciencia, Innovación y Universidades (RTI 2018-095166-B-100) and Ministerio de Economía y Competitividad (MINECO) SAF2015-66107-R (SL), cofunded by the European Regional Development Fund and Instituto de Salud Carlos III REDinREN RD12/0021/0009 and RD16/0009/0016 (SL).

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflicts of interest.

## References

1. Gagnebin Y, Boccard J, Ponte B, Rudaz S. Metabolomics in chronic kidney disease: strategies for extended metabolome coverage. *J Pharm Biomed Anal.* 2018;161:313–25. <https://doi.org/10.1016/j.jpba.2018.08.046>.
2. Perales-Quintana MM, Saucedo AL, Lucio-Gutiérrez JR, Waksman N, Alarcon-Galvan G, Govea-Torres G, et al. Metabolomic and biochemical characterization of a new model of the transition of acute kidney injury to chronic kidney disease induced by folic acid. *PeerJ.* 2019;7:1–26. <https://doi.org/10.7717/peerj.7113>.
3. Kimura T, Yasuda K, Yamamoto R, Soga T, Rakugi H, Hayashi T, et al. Identification of biomarkers for development of end-stage kidney disease in chronic kidney disease by metabolomic profiling. *Sci Rep.* 2016. <https://doi.org/10.1038/srep26138>.
4. Kordalewska M, Macioszek S, Wawrzyniak R, Sikorska-Wiśniewska M, Śledziński T, Chmielewski M, et al. Multiplatform metabolomics provides insight into the molecular basis of chronic kidney disease. *J Chromatogr B Anal Technol Biomed Life Sci.* 2019;1117:49–57. <https://doi.org/10.1016/j.jchromb.2019.04.003>.
5. Zhang ZH, He JQ, Qin WW, Zhao YY, Tan NH. Biomarkers of obstructive nephropathy using a metabolomics approach in rat. *Chem Biol Interact.* 2018;229–39. <https://doi.org/10.1016/j.cbi.2018.10.004>.
6. Zhao Y-Y, Chen H, Tian T, Chen D-Q, Ba X, Wei F. A pharmacometabonomic study on chronic kidney disease and therapeutic effect of Ergone by UPLC-QTOF/HDMS. *PLoS One.* 2014;9:1–18. <https://doi.org/10.1371/journal.pone.0115467>.
7. Dudzik D, Barbas-Bernardos C, García A, Barbas C. Quality assurance procedures for mass spectrometry untargeted metabolomics. A review. *J Pharm Biomed Anal.* 2017;147:149–73. <https://doi.org/10.1016/j.jpba.2017.07.044>.
8. De Livera AM, Olshansky G, Simpson JA, Creek DJ. NormalizeMets: assessing, selecting and implementing statistical methods for normalizing metabolomics data. *Metabolomics.* 2018;14. <https://doi.org/10.1007/s11306-018-1347-7>.
9. Chen J, Zhang P, Lv M, Guo H, Huang Y, Zhang Z, et al. Influences of normalization method on biomarker discovery in gas chromatography-mass spectrometry-based untargeted

- metabolomics: what should be considered? *Anal Chem.* 2017;89:5342–8. <https://doi.org/10.1021/acs.analchem.6b05152>.
10. Li B, Tang J, Yang Q, Cui X, Li S, Chen S, et al. Performance evaluation and online realization of data-driven normalization methods used in LC/MS based untargeted metabolomics analysis. *Sci Rep.* 2016;6. <https://doi.org/10.1038/srep38881>.
  11. De Livera AM, Sysi-Aho M, Jacob L, Gagnon-Bartsch JA, Castillo S, Simpson JA, et al. Statistical methods for handling unwanted variation in metabolomics data. *Anal Chem.* 2015;87:3606–15. <https://doi.org/10.1021/ac502439y>.
  12. Sánchez-Illana Á, Pérez-Guaita D, Cuesta-García D, Sanjuan-Herráez JD, Vento M, Ruiz-Cerdá JL, et al. Model selection for within-batch effect correction in UPLC-MS metabolomics using quality control - support vector regression. *Anal Chim Acta.* 2018;1026:62–8. <https://doi.org/10.1016/j.aca.2018.04.055>.
  13. Thonusin C, IglayReger HB, Soni T, Rothberg AE, Burant CF, Evans CR. Evaluation of intensity drift correction strategies using MetaboDrift, a normalization tool for multi-batch metabolomics data. *J Chromatogr A.* 2017;1523:265–74. <https://doi.org/10.1016/j.chroma.2017.09.023>.
  14. Gagnebin Y, Tonoli D, Lescuyer P, Ponte B, de Seigneux S, Martin PY, et al. Metabolomic analysis of urine samples by UHPLC-QTOF-MS: impact of normalization strategies. *Anal Chim Acta.* 2017;955:27–35. <https://doi.org/10.1016/j.aca.2016.12.029>.
  15. Veselkov KA, Vingara LK, Masson P, Robinette SL, Want E, Li JV, et al. Optimized preprocessing of ultra-performance liquid chromatography/mass spectrometry urinary metabolic profiles for improved information recovery. *Anal Chem.* 2011;83:5864–72. <https://doi.org/10.1021/ac201065j>.
  16. De Livera AM, Dias DA, De Souza D, Rupasinghe T, Pyke J, Tull D, et al. Normalizing and integrating metabolomics data. *Anal Chem.* 2012;84:10768–76. <https://doi.org/10.1021/ac302748b>.
  17. DIRECTIVE 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. In: *Off. J. Eur. Union.*
  18. Lavoz C, Alique M, Rodrigues-Diez R, Pato J, Keri G, Mezzano S, et al. Gremlin regulates renal inflammation via the vascular endothelial growth factor receptor 2 pathway. *J Pathol.* 2015;236:407–20. <https://doi.org/10.1002/path.4537>.
  19. Chevalier RL, Forbes MS, Thornhill BA. Ureteral obstruction as a model of renal interstitial fibrosis and obstructive nephropathy. *Kidney Int.* 2009;75:1145–52.
  20. Naz S, García A, Barbas C. Multiplatform analytical methodology for metabolic fingerprinting of lung tissue. *Anal Chem.* 2013. <https://doi.org/10.1021/ac402411n>.
  21. González-Peña D, Dudzik D, García A, Ancos B, Barbas C, Sánchez-Moreno C. Metabolomic fingerprinting in the comprehensive study of liver changes associated with onion supplementation in hypercholesterolemic Wistar rats. *Int J Mol Sci.* 2017;18:267. <https://doi.org/10.3390/ijms18020267>.
  22. Kuligowski J, Sánchez-Illana Á, Sanjuán-Herráez D, Vento M, Quintás G. Intra-batch effect correction in liquid chromatography-mass spectrometry using quality control samples and support vector regression (QC-SVRC). *Analyst.* 2015;140:7810–7. <https://doi.org/10.1039/C5AN01638J>.
  23. Gil-de-la-Fuente A, Godzien J, Saugar S, Garcia-Carmona R, Badran H, Wishart DS, et al. CEU Mass Mediator 3.0: a metabolite annotation tool. *J Proteome Res.* 2019;18:797–802. <https://doi.org/10.1021/acs.jproteome.8b00720>.
  24. Silva AM, Cordeiro-da-Silva A, Coombs GH. Metabolic variation during development in culture of *Leishmania donovani* promastigotes. *PLoS Negl Trop Dis.* 2011;5. <https://doi.org/10.1371/journal.pntd.0001451>.
  25. Warrack BM, Hnatyshyn S, Ott K-H, Reily MD, Sanders M, Zhang H, et al. Normalization strategies for metabolomic analysis of urine samples. *J Chromatogr B.* 2009;877:547–52. <https://doi.org/10.1016/j.jchromb.2009.01.007>.
  26. Sysi-Aho M, Katajamaa M, Yetukuri L, Orešič M. Normalization method for metabolomics data using optimal selection of multiple internal standards. *BMC Bioinformatics.* 2007;8. <https://doi.org/10.1186/1471-2105-8-93>.
  27. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in <sup>1</sup>H NMR metabolomics. *Anal Chem.* 2006;78:4281–90. <https://doi.org/10.1021/ac051632c>.
  28. Lee J, Park J, Lim M, Seong SJ, Seo JJ, Park SM, et al. Quantile normalization approach for liquid chromatography-mass spectrometry-based metabolomic data from healthy human volunteers. *Anal Sci.* 2012;28:801–5. <https://doi.org/10.2116/analsci.28.801>.
  29. Wu Y, Li L. Sample normalization methods in quantitative metabolomics. *J Chromatogr A.* 2015;1430:80–95.
  30. Patterson AD, Li H, Eichler GS, Krausz KW, Weinstein JN, Fornace AJ, et al. UPC-ESI-TOFMS-based metabolomics and gene expression dynamics inspector self-organizing metabolomic maps as tools for understanding the cellular response to ionizing radiation. *Anal Chem.* 2008;80:665–74. <https://doi.org/10.1021/ac701807v>.
  31. Kapoore RV, Coyle R, Staton CA, Brown NJ, Vaidyanathan S. Influence of washing and quenching in profiling the metabolome of adherent mammalian cells: a case study with the metastatic breast cancer cell line MDA-MB-231. *Analyst.* 2017;142:2038–49. <https://doi.org/10.1039/c7an00207f>.
  32. Silva LP, Lorenzi PL, Purwaha P, Yong V, Hawke DH, Weinstein JN. Measurement of DNA concentration as a normalization strategy for metabolomic data from adherent cell lines. *Anal Chem.* 2013. <https://doi.org/10.1021/ac401559v>.
  33. Noonan MJ, Tinneland HV, Buesching CD. Normalizing gas-chromatography-mass spectrometry data: method choice can alter biological inference. *BioEssays.* 2018;40. <https://doi.org/10.1002/bies.201700210>.
  34. Li B, Tang J, Yang Q, Li S, Cui X, Li Y, et al. NOREVA: normalization and evaluation of MS-based metabolomics data. *Nucleic Acids Res.* 2017;45:W162–70. <https://doi.org/10.1093/nar/gkx449>.
  35. Cook T, Ma Y, Gamagedara S. Evaluation of statistical techniques to normalize mass spectrometry-based urinary metabolomics data. *J Pharm Biomed Anal.* 2020;177. <https://doi.org/10.1016/j.jpba.2019.112854>.
  36. Ejigu BA, Valkenburg D, Baggerman G, Vanaerschoot M, Witters E, Dujardin J-C, et al. Evaluation of normalization methods to pave the way towards large-scale LC-MS-based metabolomics profiling experiments. *Omi A J Integr Biol.* 2013;17:473–85. <https://doi.org/10.1089/omi.2013.0010>.
  37. Parsons HM, Ekman DR, Collette TW, Viant MR. Spectral relative standard deviation: a practical benchmark in metabolomics. *Analyst.* 2009;134:478–85. <https://doi.org/10.1039/b808986h>.
  38. Wang YN, Ma SX, Chen YY, Chen L, Liu BL, Liu QQ, et al. Chronic kidney disease: biomarker diagnosis to therapeutic targets. *Clin Chim Acta.* 2019;499:54–63.
  39. Nogueira A, Pires MJ, Oliveira PA. Pathophysiological mechanisms of renal fibrosis: a review of animal models and therapeutic strategies. *In Vivo (Brooklyn).* 2017;31:1–22.
  40. Gandolfo LC, Speed TP. RLE plots: visualizing unwanted variation in high dimensional data. *PLoS One.* 2018;13:1–9. <https://doi.org/10.1371/journal.pone.0191629>.
  41. Walach J, Filzmoser P, Hron K, Walczak B, Najdekr L. Robust biomarker identification in a two-class problem based on pairwise log-ratios. *Chemom Intell Lab Syst.* 2017;171:277–85. <https://doi.org/10.1016/j.chemolab.2017.09.003>.
  42. Filzmoser P, Walczak B. What can go wrong at the data normalization step for identification of biomarkers? *J Chromatogr A.* 2014;1362:194–205. <https://doi.org/10.1016/j.chroma.2014.08.050>.

43. Paulson JN, Chen CY, Lopes-Ramos CM, Kuijjer ML, Platig J, Sonawane AR, et al. Tissue-aware RNA-Seq processing and normalization for heterogeneous and sparse data. *BMC Bioinformatics*. 2017;18:1–10. <https://doi.org/10.1186/s12859-017-1847-x>.
44. Hicks SC, Okrah K, Paulson JN, Quackenbush J, Irizarry RA, Bravo HC. Smooth quantile normalization. *Biostatistics*. 2018;19:185–98. <https://doi.org/10.1093/biostatistics/kxx028>.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.