

# Untargeted urine metabolomics reveals a biosignature for muscle respiratory chain deficiencies

Leonie Venter, Zander Lindeque, Peet Jansen van Rensburg, Francois van der Westhuizen, Izelle Smuts, Roan Louw

**Abstract** Mitochondrial diseases are a heterogeneous group of disorders characterised by impaired mitochondrial oxidative phosphorylation system. Most often for mitochondrial disease, where no metabolic diagnostic biomarkers exist, a deficiency is diagnosed after analysing the respiratory chain enzymes (complexes I-IV) in affected tissues or by identifying one of an ever expanding number of DNA mutations. This presents a great challenge to identify cases to undergo the invasive diagnostic procedures required. An untargeted liquid chromatography mass spectrometry metabolomics approach was used to search for a metabolic biosignature that can distinguish respiratory chain deficient (RCD) patients from clinical controls (CC). A cohort of 37 ethnically diverse cases was used. Sample preparation, liquid chromatography time-of-flight mass spectrometry methods and data processing methods were standardised. Furthermore the developed methodology used reverse phase chromatography in conjunction with positive electrospray ionisation and hydrophilic interaction chromatography with negative electrospray ionisation. Urine samples of 37 patients representing two different

experimental groups were analysed. The two experimental groups comprised of patients with confirmed RCDs and CC. After a variety of data mining steps and statistical analyses a list of 12 features were compiled with the ability to distinguish between patients with RCDs and CC. Although the features of the biosignature needs to be identified and the biosignature validated, this study demonstrates the value of untargeted metabolomics to identify a metabolic biosignature to possibly be applied in the selection criteria for RCDs.

**Keywords** Respiratory chain deficiency · Metabolomics · Urinary biomarker · LC-MS

## 1 Introduction

Mitochondrial diseases have a minimum prevalence of 1 in 5,000 live births (Schaefer et al. 2004) and are now considered the most prevalent group of inherited metabolic diseases in humans. This group of disorders are characterised by an impaired oxidative phosphorylation (OXPHOS) system. The OXPHOS system is primarily responsible for ATP generation and consists of five enzyme complexes (complex I-V), of which the first four form the respiratory chain (RC). Theoretically, respiratory chain deficiencies (RCDs), a subgroup of mitochondrial diseases, can give rise to any symptom, in any organ or tissue at any age, due to a range of nuclear-and mitochondrial DNA mutations (Munnich et al. 2011). Diagnosing a suspected mitochondrial disease is a multidisciplinary approach which generally refers to three levels of investigation which include: (1) clinical investigations; (2) biochemical and/or histochemical investigations and (3) molecular investigations (Rodenburg 2011).

---

L. Venter · Z. Lindeque · P. Jansen van Rensburg ·  
F. van der Westhuizen · R. Louw (✉)  
Centre for Human Metabonomics, North-West University  
(Potchefstroom Campus), Potchefstroom, South Africa  
e-mail: Roan.Louw@nwu.ac.za

I. Smuts  
Department of Paediatrics and Child Health, Steve Biko  
Academic Hospital, University of Pretoria, Pretoria,  
South Africa

Enzyme analyses on tissue samples (most often muscle) remain the gold standard to confirm the presence of an RCD. One of the major diagnostic challenges at hand is the selection of cases to undergo tissue biopsies. Different criteria has been published to guide clinicians in the diagnostic process of selecting patients with potential RCDs to undergo a muscle biopsy (Bernier et al. 2002; Wolf and Smeitink 2002; Phoenix et al. 2006; Schaefer et al. 2006). Apart from the hospitalization, anaesthetics and trauma associated with this procedure, logistic problems of patients living in remote areas also tend to complicate the process (Smuts et al. 2010). An adapted mitochondrial disease criteria (MDC) score was used to guide in the selection of patients qualifying for a muscle biopsy. It is important to note that the patients might have been underscored, because no muscle histology was available, but the intention was not to use the MDC as a diagnostic criteria. Biopsies were taken from patients if the MDC score was  $\geq 6$  or if the clinical phenotype was highly suggestive of a mitochondrial disease (e.g. two or more unrelated systems involved without any other identifiable aetiology). However, it was found that the sensitivity for the MDC  $\geq 6$  as screening tool was only 60 %, the specificity 66 %, the positive predictive value 74 % and the negative predictive value 51 % under these specific circumstances (unpublished data). It was realised that a non-invasive selection criteria should be developed to identify patients who should undergo a muscle biopsy to confirm a potential RCD and limit the number of patients that may have a negative enzyme results from undergoing a biopsy.

In a study conducted by Smuts et al. (2013) a putative urinary biosignature for RCD was proposed that consists of six amino acids (alanine, glycine, glutamic acid, serine, tyrosine and  $\alpha$ -amino adipic acid), six organic acids (3-hydroxy-3-methylglutaric acid, 3-hydroxyisovaleric acid, 3-hydroxyisobutyric acid, 2-hydroxyglutaric acid, succinic acid and lactic acid) and creatine. However, the control group used in this and related studies (Reinecke et al. 2012) consisted of clinically healthy patients. This signature can therefore not be applied to limit the inclusion of false positive patients (i.e. patients with clinical features compatible with a mitochondrial disease, but who will eventually have a negative enzyme result on the muscle biopsy) for enzyme analysis. For this reason, the aim of this study was to use an untargeted LC-MS metabolomics approach to investigate RCDs and generate a putative biosignature to use as secondary filter in the selection process of clinically suspected RCD patients to undergo muscle biopsy. This metabolic signature should thus be able to differentiate between patients with a positive enzyme result in muscle (true positive) and patients with normal enzyme results in muscle (true negative).

## 2 Materials and methods

### 2.1 Ethics statement

The study complied with all applicable institutional guidelines and terms of the Declaration of Helsinki of 1975 (as revised in 2008) for investigation of human participants and was approved by the relevant Ethics Committee of the University of Pretoria (No. 91/98 and amendments). Informed consent was obtained from the parents of the patients for the use of urine and muscle samples for research purposes.

### 2.2 Sample selection and experimental groups

All samples were obtained from patients referred to the Paediatric Neurology Unit at the Steve Biko Academic Hospital, Pretoria, South Africa. The collected urine samples were stored at  $-80^{\circ}\text{C}$  prior to metabolomic analyses. Enzyme analyses were performed on muscle biopsies from the *Vastus lateralis* muscle of all the patients who had a MDC score  $\geq 6$  or a clinical phenotype suggestive of one of the syndromic mitochondrial disorders to confirm the presence or absence of a RCD (Smuts et al. 2010).

Two different experimental groups formed part of this metabolomics investigation. The first group consisted of patients with RCDs which were confirmed on enzyme level. The second group consisted of clinical controls (CC) i.e. patients where a RCD was clinically suspected but had an RC enzyme activity within the reference range. Briefly, the clinical evaluation comprised of a detailed medical history, clinical examinations and baseline investigations including ammonia, creatine kinase, pyruvate and lactate assessments. Patient specific investigations e.g. chest X-ray, an electrocardiogram and cardiac sonography were performed only if it was clinically suspected. When hearing and visual impairments surfaced, brainstem auditory-evoked responses and visual evoked potentials were requested. Nerve conduction studies and electromyogram was requested if a neuropathy or myopathy was expected. In cases of suspected central nervous system involvement computed tomography scan or magnetic resonance imaging of the brain was performed. Ultimately an MDC score was calculated for all of the patients to guide if a muscle biopsy should be taken to confirm a possible RCD (Smuts et al. 2010).

Mitochondrial RC enzymes (CI-IV; EC 1.6.5.3, EC 1.3.5.1, EC 1.10.2.2, EC 1.9.3.1, respectively), pyruvate dehydrogenase complex (PDHc, EC 1.2.4.1), and citrate synthase (CS, EC 2.3.3.1) activities were measured in muscle (Smuts et al. 2010). Patients with confirmed enzyme deficiencies using the criteria similar to those

**Table 1** Clinical data on the 37 patients included in this study

| RC enzyme deficiency(ies) <sup>a</sup> | Group            | Ethnicity | Gender | Age (years) | Clinical profile <sup>b</sup>         | MDC score <sup>c</sup> |
|--|------------------|-----------|--------|-------------|---------------------------------------|------------------------|
| CI                                     | RCD              | A         | F      | 4           | DD, Dys, M                            | 4                      |
| CIII                                   | RCD              | A         | F      | 7           | DR, CNS, Eye, End                     | 6                      |
| CIII                                   | RCD              | A         | F      | 1           | M, CNS, Eye, L, DR                    | 7                      |
| CIII                                   | RCD              | A         | M      | 7           | M, AID                                | 4                      |
| CIII                                   | RCD              | A         | F      | 1           | DD, Dr, CNS, Eye, M                   | 8                      |
| CIII                                   | RCD              | A         | F      | 6           | CNS, E                                | 4                      |
| CIV                                    | RCD              | C         | M      | 43          | M                                     | 4                      |
| CI, CII+III                            | RCD              | A         | M      | 2           | DD, Eye, PNS, M                       | 7                      |
| CI, CIII                               | RCD              | C         | M      | 1           | M, CNS, Eye, R, DD, E                 | 8                      |
| CI, CIII                               | RCD              | C         | F      | 2           | M, End, ENT, G, DD                    | 8                      |
| CI, CIII, CII+III, CIV                 | RCD              | A         | M      | 5           | DD, DR, Be, CNS, M, G, End            | 8                      |
| CI, CIII, CIV                          | RCD              | A         | F      | 2           | M, CNS, Eye, ENT, S, DD               | 5                      |
| CI, CIII, CIV                          | RCD              | A         | M      | 5           | DD, DR, Dys, Be, CNS, E, M, End, L, S | 5                      |
| CI, CIV, CII+III                       | RCD              | A         | M      | 23          | DR, CNS, Eye, M, R, Car               | 8                      |
| CII+III                                | RCD              | A         | M      | 8           | M                                     | 3                      |
| CII+III                                | RCD              | C         | F      | 8           | DR, CNS, E, M                         | 6                      |
| CII+III                                | RCD              | C         | F      | 1           | M, CNS, Eye, Skin, DD, DR, E          | 8                      |
| CII+III, CIII                          | RCD              | A         | F      | 1           | M, CNS, DD                            | 3                      |
| CII+III, CIV                           | RCD              | C         | F      | 17          | DR, Be, CNS, M, G                     | 7                      |
| CIII, CII+III                          | RCD              | A         | M      | 1           | DR, CNS, M                            | 6                      |
| CIII, CIV                              | RCD              | A         | F      | 6           | M, CNS, End, S, DD, E                 | 8                      |
| CIII, CIV                              | RCD              | C         | M      | 6           | M, CNS, L, DD, E                      | 6                      |
| CIII, CIV                              | RCD              | C         | F      | 7           | M, G, Car, DD                         | 4                      |
| CIII, CIV                              | RCD              | A         | F      | 6           | CNS, Eye, S, DD, DR, Be               | 6                      |
| –                                      | Clinical control | C         | M      | 5           | DD, M                                 | 4                      |
| –                                      | Clinical control | I         | M      | 1           | DD, CNS, E, Eye                       | 5                      |
| –                                      | Clinical control | A         | M      | 3           | DD, CNS, Eye, M                       | 6                      |
| –                                      | Clinical control | C         | F      | 2           | DD, CNS                               | 2                      |
| –                                      | Clinical control | A         | M      | 1           | DD, Dys, Eye, M, End                  | 5                      |
| –                                      | Clinical control | A         | F      | 4           | DD, Be, CNS                           | 5                      |
| –                                      | Clinical control | C         | F      | 3           | DD, Be, CNS, E, Eye                   | 4                      |
| –                                      | Clinical control | A         | F      | 1           | DD, CNS, End                          | 6                      |
| –                                      | Clinical control | A         | F      | 9           | M                                     | 4                      |
| –                                      | Clinical control | A         | F      | 8           | DD, DR, CNS, E, G                     | 6                      |
| –                                      | Clinical control | A         | F      | 9           | G, E                                  | 2                      |
| –                                      | Clinical control | C         | M      | 1           | DD, CNS, E, End, Eye                  | 6                      |
| –                                      | Clinical control | A         | F      | 11          | DD, CNS, Eye, PNS, M                  | 4                      |

<sup>a</sup> An RC enzyme deficiency is defined as a combination of at least two values expressed against different markers (CS, CII and CIV) with at least one of them equal or lower than the 5<sup>th</sup> percentile and the second at least equal or lower than the lowest control value provided that the enzyme activity of the marker is normal. One of these two values must be CS. Any deficient marker (CII or CIV) was excluded for interpretation and if both CII and CIV were deficient, only CS was used for identification

<sup>b</sup> Clinical profile includes: *Aid* auto-immune disorder, *BE* behavior and emotional abnormalities, *Car* cardiac involvement, *CNS* central nervous system involvement, *DD* developmental delay, *DR* developmental regression, *Dys* dysmorphism (minor and major), *E* epilepsy, *End* endocrine abnormalities, *ENT* sensori-neural deafness, *Eye* vision involvement, *G* gastro-intestinal tract involvement, *L* liver involvement, *M* muscle involvement, *PNS* peripheral neuropathy, *R* renal involvement, *S* skeletal involvement, *Skin* Skin involvement

<sup>c</sup> *MDC score* Mitochondrial Disease Score (Wolf and Smeitink 2002)

described before (Smuts et al. 2010, Reinecke et al. 2012) were included in the RCD group and the patients with clinical features of a RCD but RC enzyme levels within the reference range were categorised to the CC group. Table 1 summarizes the characteristics of the selected experimental groups.

### 2.3 Sample preparation and quality control

Aliquots of the stored urine samples were used for creatinine determination. Urine samples were normalised for creatinine values so that each sample had a final creatinine content of 0.25  $\mu\text{mol}$ . Briefly, a predetermined volume of urine (containing 0.25  $\mu\text{mol}$  creatinine) was put in a microcentrifuge tube along with 10  $\mu\text{l}$  of internal standard solution (containing 1 mg/ml nor-leucine, 2-acetamidophenol and caffeine for positive ESI analysis and nor-leucine, 2-acetamidophenol and 3-phenylbutyric acid for negative ESI analysis). Twenty microliters acetonitrile was added, the samples frozen overnight in  $-80\text{ }^{\circ}\text{C}$  and then freeze dried. The dried samples were re-dissolved in 100  $\mu\text{l}$  water for reverse phase LC or 50  $\mu\text{l}$  acetonitrile with 50  $\mu\text{l}$  water for HILIC separation by incubating them for 30 min (hydration step) and vortexing for 10 s. Thereafter the samples were centrifuged  $25,055\times g$  for 10 min at  $4\text{ }^{\circ}\text{C}$  and the supernatant transferred to vials fitted with pointed inserts.

All samples were analysed within a week to limit week-to-week variation. The samples were randomised and divided into daily batches so that each batch contained more or less the same number of patient and control samples. Quality control (QC) samples (from pooled urine) were analysed after every 5th sample within each batch. A secondary QC was incorporated in all batches (after every tenth sample) which would be used to validate within- or between batch corrections (if needed). The secondary QC is identical to the first and was used as a test set that does not participate in “modelling” of a batch effect (like the training set).

### 2.4 Instrumentation and analysis

The samples were analysed using an Agilent 1290 series LC system coupled to a 6540 Q-TOF mass analyser (Agilent Technologies, Santa Clara, CA, USA) consisting of a Micro Vacuum Degasser (G1379B); Binary pump SL (G1312B); Preparative Autosampler HiP-ALS SL (G1367C); Thermostat ALS (G1330B) and Thermostatted Column Compartment SL (G1316B). The auto sampler’s temperature was set to  $4\text{ }^{\circ}\text{C}$ . An injection volume of 5  $\mu\text{l}$  was used for both reverse phase and HILIC chromatography.

#### 2.4.1 Reverse phase (C18)

Reverse phase chromatography was performed with a ZORBAX SB-Aq column (3.5  $\mu\text{m}$ ,  $2.1 \times 150\text{ mm}$ ) from Agilent. The chromatographic gradient started at 100 % solvent A (water with 0.1 % formic acid) and maintained for 5 min, before the gradient was increased to 35 % solvent B (acetonitrile with 0.1 % formic acid) over a period of 25 min. Next the gradient was increased to 70 % solvent B at 35 min and 100 % at 36 min. The gradient was kept at 100 % solvent B for 3 min and then decreased to 0 % within 3 min. A post run of 8 min was allowed to ensure equilibration of the column to give a total run time of 50 min (42 min gradient and 8 min post run) per sample. The dual ESI source was setup for positive ionisation. The drying gas temperature was set at  $280\text{ }^{\circ}\text{C}$  with a drying gas flow of 8 l/min and nebuliser pressure of 30 psi. The Q-TOF was set to scan from 50 to 1,000  $m/z$ . Both centroid and profile data were stored and the instrument set to extended dynamic range (2 GHz). A reference solution containing masses 121.050873 [M+H]<sup>+</sup> and 922.009798 [M+H]<sup>+</sup> were constantly infused as accurate mass reference.

#### 2.4.2 HILIC

HILIC chromatography was performed with a ZORBAX RRHD HILIC plus column (1.8  $\mu\text{m}$ ,  $2.1 \times 150\text{ mm}$ ) from Agilent. The chromatographic gradient started at 10 % solvent A (water with 5 mM acetic acid) and maintained for 2.5 min. After 2.5 min, the gradient was decreased to 50 % solvent B (acetonitrile with 5 mM acetic acid) over 10 min. The gradient was maintained at 50 % solvent B for 5 min and returned to 90 % at 20 min. The gradient was kept at 90 % for 5 min with a 10 min post run to give a total run time of 35 min (25 min gradient and 10 min post run) per sample. The dual ESI source was setup for negative ionisation. The drying gas temperature was set at  $280\text{ }^{\circ}\text{C}$ , the drying gas flow at 8 l/min and the nebuliser pressure at 30 psi. The Q-TOF was set to scan from 50 to 1,000  $m/z$ . Both centroid and profile data were stored. The instrument mode was set to extended dynamic range (2 GHz). A reference solution containing masses 119.036320 [M-H]<sup>-</sup> and 980.016375 [M+CH<sub>3</sub>COO]<sup>-</sup> were constantly infused as accurate mass reference.

### 2.5 Data extraction, pre-processing and normalisation

The data was extracted using Agilent’s Mass Hunter Qualitative software (Version B.05.00) and Mass Profiler Professional software (Version B.02.02). The molecular feature extraction and find by ion algorithm were used according to Agilent’s specifications. Feature annotation or

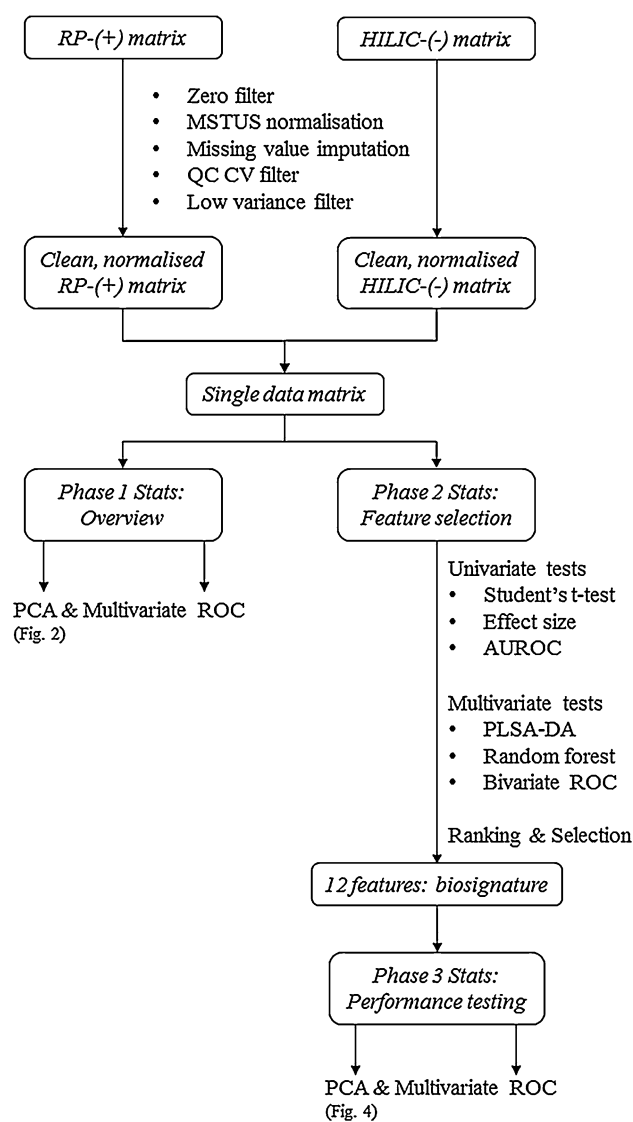
identification was not performed during data extraction as this step was performed only on the list of statistically important features. Zero filtering was implemented to eliminate features with extensive missing values, like compounds linked to diet or medication, which was seen in only a few urine samples but not in the rest. If a feature was detected in all but one sample in any experimental group, the feature remained in the data matrix for further analysis. Samples were normalised using the mass spectra total useful signal (MSTUS) normalization method described by Warrack et al. (2009) as this method corrects for difference in urine concentration or osmolality without using the creatinine concentration. It is debatable whether creatinine values in energy-related diseases (like RCDs) can be used to express a patient's urine concentration. Missing values were replaced after normalisation with half the detection limit, i.e. 50 % of the minimum value found in the dataset (Xia et al. 2009; Hrydziuszko and Viant 2012).

A QC coefficient of variation (CV) filter was applied to the dataset in order to remove compounds that were unstable during analysis or unreliably measured. The CV of all the features in the QCs was determined and all features with a CV >50 % was removed. Variables with low variance were further filtered using effect size as guidance. For features to be included in the data matrix, a specific (log transformed) feature needed to have an effect size (Ellis and Steyn 2003)  $d$  value of 0.3 or more. After the datasets from the positive and negative ESI were evaluated for any within- or between batch effects, the datasets was combined before any statistical analysis.

## 2.6 Data pre-treatment and statistical analysis

Statistical analyses were performed in three phases as indicated in Fig. 1. In the first phase principal component analyses (PCA) and support vector machine (SVM) multivariate ROC modelling were performed on the entire data matrix before feature selection in order to get an overview of the data. A scatter plot with a 90 % confidence ellipse was created to show sample grouping. In the second phase, various univariate and multivariate statistical tests were used for feature selection—to identify only a handful of important features that would allow for best discrimination between the RCD and CC group.

The three univariate tests used for feature selection were Student  $t$  test, effect size and area under (AU) the receiver operator characteristic curve (ROC). All features that differed significantly between the RCD and CC group with  $P < 0.05$  or  $d > 0.8$  or  $AUC > 0.7$  were considered important. The three multivariate tests used for feature selection were partial least square discriminant analysis (PLS-DA), random forest (RF) and bivariate AU-ROC. Features with a PLS-DA variable important in projection



**Fig. 1** Data pre-processing and statistical analysis workflow. The two data matrices obtained were separately cleaned before all the data was combined to form a single data matrix. In the first phase of statistical analyses an overview of the data was used to visualise any natural separation of the two experimental groups. In the second phase various statistical tests were used for feature selection that would allow for best discrimination between the RCD and clinical control groups. In the third phase, the discrimination power of the proposed biosignature was tested

(VIP) value (1st coeff) >1 were considered important. For RF, a combination of 10 features in each node was selected and a 1,000 trees grown. RF ranks the features according to their selection in the classification trees that gave good discrimination. Since this is merely a ranking and not values with common cut-off thresholds, the top 50 was considered important. With bivariate ROC, all binary feature pairs that gave an AU-ROC >0.7 were considered important.



All the important features identified with the above mentioned tests were listed and the average rank of each feature across all the tests determined. The top features that performed best overall based on a ranking plot were then selected as the potential biosignature that can discriminate between RCD and CC patients. It should be mentioned that it is often common practise to use Venn diagrams for final feature selection when multiple tests are performed. However, since many of the selected tests complement each other instead of “correcting” each other, the use of this average ranking method was preferred. For this reason, a feature that might have a high *P* value and do not differ significantly between the groups, but performed well in two of the multivariate tests could be included in the signature.

In the third phase of statistical analysis, the discrimination power of the selected biosignature was tested. PCA was performed to evaluate grouping of the samples based solely on the selected features. The performance of the biosignature was further evaluated with SVM and PLS-DA multivariate ROC modelling and associated prediction accuracy, cross-validation and permutation tests. MetaboAnalyst (Xia et al. 2009; [www.metaboanalyst.ca](http://www.metaboanalyst.ca)), ROC-CET web service (Xia et al. 2009; [www.roccet.ca](http://www.roccet.ca)) and Microsoft Excel were used for all statistical analyses. The data was pre-treated with a generalised logarithm (glog) or log 2 transformation before any statistical analysis.

## 2.7 Compound identification

METLIN (<http://metlin.scripps.edu>) and Human metabolome database ([www.hmdb.ca](http://www.hmdb.ca)) were used to identify the features of the biosignature using the accurate mass, isotope ratios and salt adduct patterns. Since the selected ionization conditions for untargeted analysis are not optimal for all compounds analysed, some fragmentation could occur and these fragments were used to confirm the identification where possible. Lower mass features that co-eluted with the target features were checked in the mentioned databases to identify possible fragments that could confirm the identity of the important feature. Furthermore, the target features were also regarded as possible fragments (instead of intact molecular ions) and used in the fragment search options of these databases. When the mass of the feature was linked as a fragment of a specific compound, other co-eluting masses were inspected for secondary fragments for confirmation as mentioned above.

## 3 Results and discussion

### 3.1 Data quality evaluation

The quality of the data was first evaluated before statistical analysis was performed. Batch effects and time drifts are

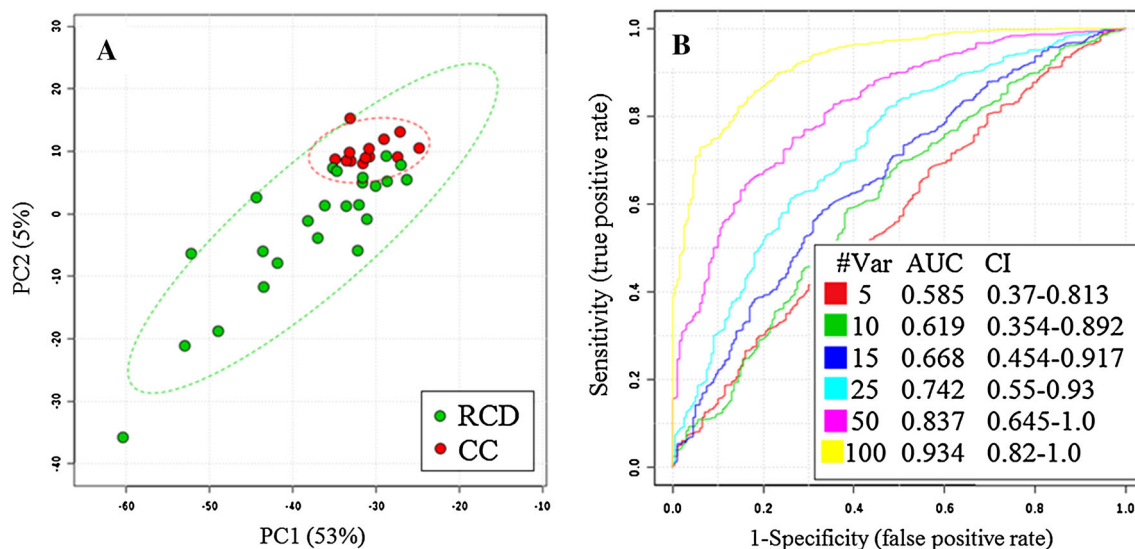
one of the main concerns in such studies and were visually checked by using PCA score plots and univariate scatter-plots of all the features after data normalization. The data matrices of the reverse phase-positive ionization analysis and HILIC-negative ionisation analysis were evaluated separately. The PCA score plots indicated that all QC samples, from all the batches, grouped close together which suggest no batch effect or time-related drifts (results not shown). Moreover, scatterplots of the feature intensities sorted by run-order were also clear of pertinent batch effects. Since the quality of the data was sufficient for further analysis, no batch- or time-drift corrections were made to the data. The QC samples were removed and the data from the positive- and negative ESI assays combined to yield a single data matrix for statistics (illustrated in Fig. 1). This was done to exploit potential covariance in metabolites detected by either method.

### 3.2 Overview of data before feature selection

The features that remained in the matrix after data clean-up were analysed to visualise any natural separation regarding the two relevant experimental groups. The PCA score plot displayed in Fig. 2a gives an overview of all the features found in the data matrix before any feature selection. The smaller cluster of samples represents the CC group. The RCD group is displayed by the scores that are spread out across the plot. It is expected that the RCD patient group would not cluster together tightly due to the diversity found between patients and because RCDs are a heterogeneous group of disorders—with any of the four complexes or combination thereof defective. Even though clear separation between the experimental groups is not prominent in this case, the tendency of the groups to cluster separate is clear. The inclusion of a few RCD samples in the CC cluster suggests that some of the RCD patients and CC have a very similar metabolite profile which could be expected since these controls showed clinical symptoms usually associated with RCDs.

From a clinical point of view, a “rule-in” approach is preferred for the development of a biosignature (Pewsner et al. 2004). This means that the developed signature must have a 0 % false negative rate but can have a certain percentage false positive rate. With this approach, the false positives (clinical referred controls in this study) with a good probability of having a RCD would be referred for a biopsy and enzyme analyses even if it turns out that they do not have RCDs. As mentioned in the introduction, the aim of this study was to find a biosignature that would lower the number of false positive patients being referred for a muscle biopsy.

Another prerequisite of the biosignature is that it must be the smallest set of metabolites that have the best



**Fig. 2** Overview PCA score plot (a) and multivariate ROC (b) before feature selection. Sample grouping are shown in the PCA with 90 % confidence ellipses. For ROC, the performance of the top number of

features (*#Var*) according to SVM with their *AUC* and confidence interval (*CI*) indicated. The discriminant model performed better the more features it contained

discrimination power between the groups of interest. More often than not, diagnosis of other inborn errors of metabolism is based on this principle where the clinician monitors only a few metabolites mostly related to the defective pathway (Xia et al. 2013). However, RCDs do not result in unique metabolites and influences the metabolism in a more global manner as the mitochondrion can be considered as the hub of metabolism. It is thus all the small changes collectively that result into the separation of the experimental groups (Fig. 2a). This was also confirmed when supervised statistical methods preferred to use >100 features in discrimination signatures or models (Fig. 2b). The discrimination power of the evaluated signatures by the supervised multivariate statistical test is shown in ROC curve format in Fig. 2b. The six curves displayed in Fig. 2b show the discriminant power of the six models tested by the supervised statistical test. The closer the ROC moves to the upper left corner, the higher the overall accuracy of the discrimination model (and the larger the *AUC*). Here the results indicate that the more features the signature contains, the better the modelling became.

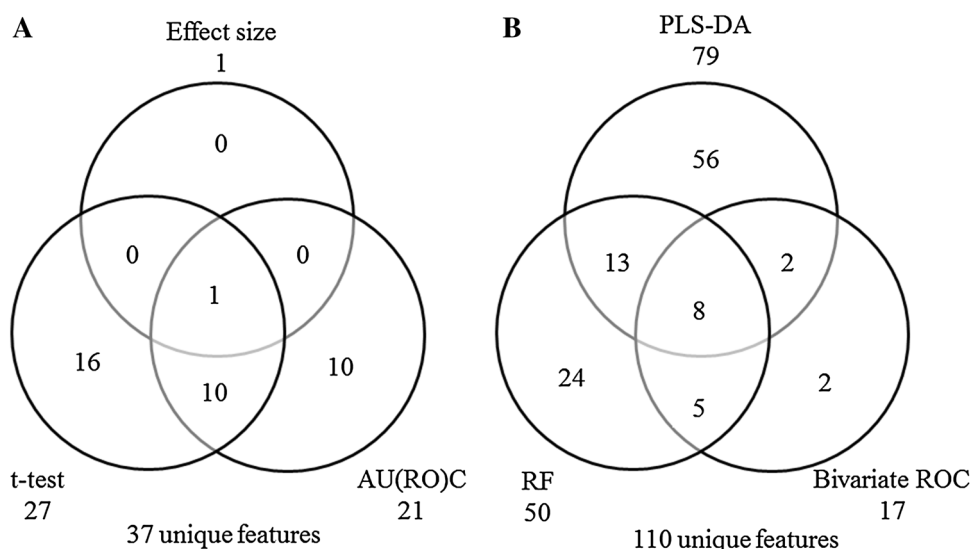
Considering the practical implication of measuring >100 features to best discriminate patients, a shortlist of about 2–10 features would be more practical for clinical testing (Xia et al. 2013). In order to find the smallest biosignature that can discriminate between the groups, feature selection was performed using several univariate and multivariate tests.

### 3.3 Univariate and multivariate feature selection

One of the most crucial steps in a metabolomics study is the accurate selection of biomarker candidates, since it determines the outcome of the discovery-phase study as well as the course of validation studies (Christin et al. 2013). The Venn diagrams in Fig. 3 indicate the number of features that were identified as important by univariate and multivariate tests, respectively. Univariate tests identified 37 significant features and multivariate tests indicated 110 features of worth. In total, 111 important features were found with both approaches. The goal was not to use the Venn diagrams to identify the common features found between the different tests, but merely to give an illustration of the amount of features found with the various methods. Many metabolomics studies consider the features found in the centre of a Venn diagram as most important. With this approach, one assumes that the selected tests is “correcting each other” by eliminating false positive features. Our approach differed from this common trend as we assume that the univariate and multivariate tests complement each other. Next all the features were ranked according to their importance in the tests. The average rank of each feature was determined, sorted accordingly and a graph was constructed (Fig. S1, online supplementary material) to identify the most suitable features for a biosignature.

As there is no measure to determine the number of features in the biosignature, the graph in Fig S1 was used as

**Fig. 3** Venn diagrams of important features found by univariate (a) and multivariate tests (b), respectively. Univariate tests identified 37 features of significance while 110 features were found by multivariate statistics



a guideline. By plotting the average ranking of all the features, a visible gap between the first 12 features and rest were found and used as preliminary cut-off. The list of the top 12 selected features is given in Table 2. The feature's annotated name, mass, retention time, separation method, effect size  $d$  value,  $t$  test  $P$  value, PLS-DA VIP value and AUC are included in the table. It should be noted that it was not possible to identify all the features and thus only putative identifications were given to some of the features. It can be mentioned that *N*-succinyl-L-2,6-diaminopimelate and a tetrapeptide (consisting of Asn, Ser, Ser and His, in any sequence) were annotated with a high degree of probability since both databases used for the identification only returned this match. However, AMP, N-acetylaspartate and oxoglutaric acid were annotated with less confidence since another match from the database was also possible. In these cases the more commonly detected and known name was then used in Table 2. The  $M = 222.1615$  feature appears to be a fragment of tetradecanedioic acid ( $M = 258.1831$ ). These two mass features were simultaneously detected in the mass analyser, along with another fragment ( $M = 240.1721$ ) and therefore the software group them as originating from one compound. After manual inspection, all three masses were found to be related to tetradecanedioic acid on the Metlin Spectral database. However, it was not within the scope of this study to interpret the origin or reason for the biological variance between the experimental groups. In such a study, all compounds that differed significantly should be included in pathway analysis for a better understanding of the biological responses. Furthermore, in order to give a clear biological interpretation of this feature list, validation of the compound identities are necessary. Possible validation steps will be discussed in the concluding section.

### 3.4 Evaluating the signature's discrimination power

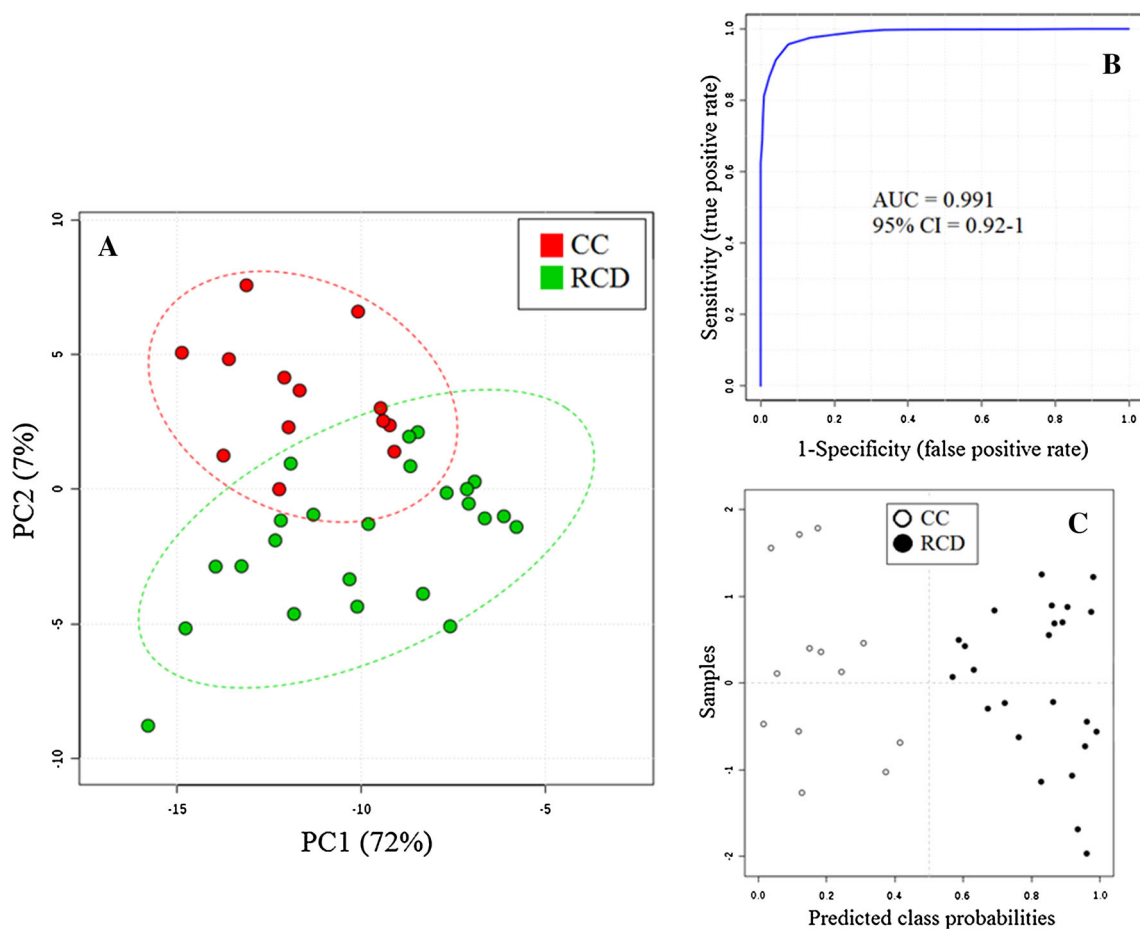
To verify if this proposed signature of 12 features can be used to distinguish RCD patients and CC from each other, PCA and multivariate ROC curves were used. Fig. 4a shows the PCA score plot of the RCD patients and CC group when using only the new biosignature. The smaller clustered group represents the CC group while the larger cluster represents the RCD patient group. With the use of a 90 % confidence ellipse, four of the CC samples are included within the RCD patient group classification while nine were not. As discussed previously, complete separation of the groups is not essential. The inclusion of these four patients for muscle biopsies and enzyme analyses is considerably more effective compared to sending all these CC for a biopsy.

The discrimination power of the biosignature using multivariate ROC is shown in Fig. 4b. The AUC of 0.991 with 95 % confidence interval that range between 0.92 and 1 indicates that this signature is both very specific and sensitive. It can also be noted that this proposed signature has 98 % sensitivity at 80 % specificity, making this a statistically valid method for biomarker performance. Hence, with the "rule-in" clinical approach, this test will basically include only 20 % false positive cases for biopsy and enzymatic analysis with basically 100 % of the true positive cases (similar to what the PCA score plot indicated). The average predicted class probability of each sample across 100 cross validations are shown in Fig. 4c. The Monte-Carlo cross validation algorithm used a balanced sub-sampling approach with the classification boundary at the centre (0.5). Ranging from 0 to 1, a probability score is calculated. If  $<0.5$ , the sample is classified as CC and  $>0.5$ , it is classified as part of the RCD



**Table 2** The biosignature metabolite markers identified using an LC-MS metabolomics approach

|    | Annotated name                            | Mass      | RT (min) | Separation method | Effect size (d value) | <i>t</i> test ( <i>P</i> value) | PLS-DA (VIP) | AUC    |
|----|---|-----------|----------|-------------------|-----------------------|---------------------------------|--------------|--------|
| 1  | AMP                                       | 347.0616  | 2.30     | RP                | 0.7450                | 0.0081                          | 2.0351       | 0.8782 |
| 2  | C9H19NO                                   | 157.1464  | 30.3     | RP                | 0.5857                | 0.0102                          | 2.0162       | 0.8750 |
| 3  | C23HNO8S4                                 | 546.856   | 12.7     | RP                | 0.6145                | 0.0194                          | 1.8464       | 0.8653 |
| 4  | N-Acetyl asparagine                       | 174.0641  | 8.7      | HILIC             | 0.9597                | 0.0031                          | 2.4327       | 0.9070 |
| 5  | Unknown 1                                 | 1043.4677 | 11.97    | RP                | 0.6062                | 0.0243                          | 1.6806       | 0.8429 |
| 6  | <i>N</i> -succinyl-L-L2.6 diaminopimelate | 290.1126  | 26.9     | RP                | 0.6152                | 0.0473                          | 1.4432       | 0.7307 |
| 7  | C16H26O4S                                 | 314.1547  | 24.9     | RP                | 0.4790                | 0.1097                          | <1           | <0.7   |
| 8  | Unknown 2                                 | 136.1248  | 27.9     | RP                | 0.7145                | 0.0034                          | 2.2290       | 0.8814 |
| 9  | Fragment of tetradecanedioic acid         | 222.1615  | 24.7     | RP                | 0.7012                | 0.0066                          | 2.0527       | 0.8814 |
| 10 | C14H24O2S                                 | 256.1491  | 27.67    | RP                | 0.4921                | 0.0588                          | <1           | <0.7   |
| 11 | Tetrapeptide                              | 443.1786  | 21.7     | RP                | 0.7352                | 0.0143                          | 1.9285       | 0.8717 |
| 12 | Oxoglutaric acid                          | 146.0215  | 2.35     | HILIC             | 0.7882                | 0.0213                          | 1.7953       | 0.8589 |



**Fig. 4** PCA score plot (a), multivariate ROC (b) and average predicted class probability (c) of the proposed signature. Sample grouping are shown in the PCA with 90 % confidence ellipses. The ROC performance of the biosignature according to SVM with *AUC*

and confidence interval (*CI*) indicated. Near 100 % predicted class probability of RCDs and clinical referred controls over 100 cross validations using the 12 features are shown

group. This result indicates that the biosignature was able to classify the samples  $\sim 100\%$  accurately (on average) in the 100 cross validation sub-samples of the original data. The formula for calculating the accuracy are given elsewhere (Xia et al. 2013).

Permutation testing was also used to validate the proposed signature performance as another form of cross-validation. This technique is based on the hypothesis that this proposed set of biomarkers could have been found again if every patient or control sample had been randomly assigned to a different group (Xia et al. 2013). To prove or disapprove this hypothesis, 1,000 models were permuted, by randomly assigning each sample to an experimental group and then using the proposed signature for the classification of the dummy groups. As a result a reference distribution (of the null hypothesis) is given to be compared to the correctly assigned model, which leads to a  $P$  value. With a  $P$  value  $< 0.05$  a randomly permuted outcome variable has less than 5% chance to produce a model of similar performance to the correctly assigned model (Xia et al. 2013). The proposed signature resulted in a permutation test value of  $P < 0.01$ , thus by using random guessing there is  $< 1\%$  chance to produce the same model indicating the specificity of the biosignature. The multivariate ROC and associated cross validation tests were performed using SVM as the underlying multivariate algorithm. To ensure that this test did not have any bias toward the specific data, PLS-DA was also tested in the multivariate ROC modelling. The use of PLS-DA produced the same results.

#### 4 Conclusion

An untargeted LC-MS metabolomics approach was used to produce a putative urinary biosignature to identify patients that will most likely have an identifiable RCD deficiency in muscle. The use of both reverse phase and HILIC chromatography (respectively) ensured that hydrophobic and small hydrophilic compounds (such as organic acids) were included in the screening and created the opportunity where covariance in seemingly unlinked metabolites could be exploited. Although RCD can be classified as an inborn error of metabolism, it is different in the sense that electrons (found in the ETC) built-up and not metabolites. Electron carriers such as NADH and FADH accumulate which leads to numerous other metabolic reactions being halted or slowed by the lack of sufficient NAD and FAD.

In contrast to previously published papers, this study focused on metabolic differences between true and false positive RCD patients instead of comparing true positives against healthy controls. The challenge was to develop a non-invasive method to distinguish between patients with

suspected RCD that will eventually have the diagnosis confirmed on enzyme analysis in muscle and those that will not have a confirmed RCD. A 12 feature biosignature was identified that was able to sufficiently distinguish between these patients groups. Thanks to the high-throughput nature of web statistical services like MetaboAnalyst, several complementary univariate and multivariate statistical tests were used to identify this signature. Follow-up studies are needed to identify these features (via structural elucidation and/or standards) after which the validity of these features as biosignature should be validated in the same sample set using targeted methods (such as LC-tandem MS). In the final phase, the biosignature should be validated on a new sample set and the practical use of the signature explored with the use of a user-friendly algorithm (for example logistic regression) where a numeric value would indicate whether a patient is likely to be diagnosed with a muscle RCD. However, this study clearly demonstrates the value of untargeted metabolomics to identify a metabolic biosignature to assist in the conformation process of a RCD.

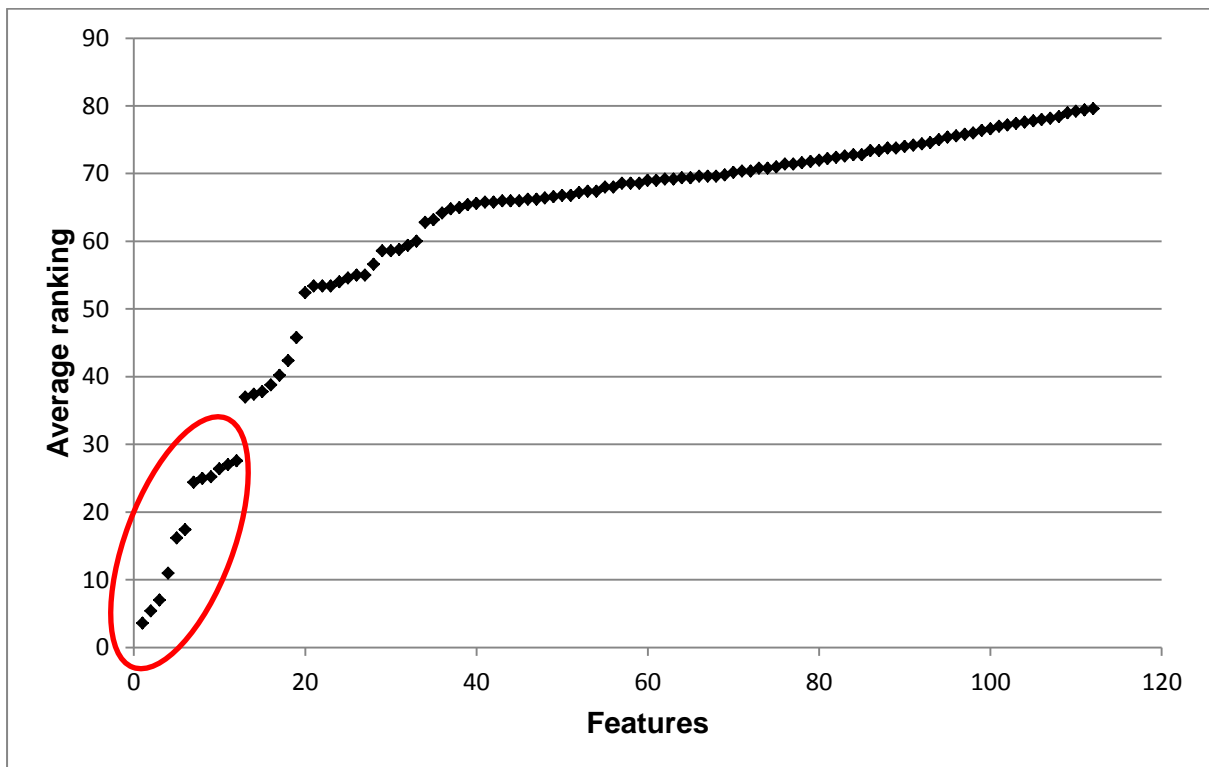
**Acknowledgments** Essential funding was obtained from the North-West University, Potchefstroom Campus.

#### References

- Bernier, F. P., Boneh, A., Dennett, X., Chow, C. W., Cleary, M. A., & Thorburn, D. R. (2002). Diagnostic criteria for respiratory chain disorders in adults and children. *Neurology*, 59(9), 1406–1411.
- Christin, C., Hoefsloot, H. C., Smilde, A. K., Hoekman, B., Suits, F., Bischoff, R., et al. (2013). A critical assessment of feature selection methods for biomarker discovery in clinical proteomics. *Molecular and Cellular Proteomics*, 12, 263–276.
- Ellis, S., & Steyn, H. (2003). Practical significance (effect sizes) versus or in combination with statistical significance (p values). *Management dynamics*, 12, 51–53.
- Hrydziusko, O., & Viant, M. R. (2012). Missing values in mass spectrometry based metabolomics: An undervalued step in the data processing pipeline. *Metabolomics*, 8, 161–174.
- Munnich, A., Rötig, A., Cormier-Daire, V. & Rustin, P. (2011). Clinical presentation of respiratory chain deficiency. In *The online metabolic and molecular base of inherited disease* 10.
- Pewsner, D., Battaglia, M., Minder, C., Marx, A., Bucher, H. C., & Egger, M. (2004). Ruling a diagnosis in or out with “SpPin” and “SnNOut”: A note of caution. *British Medical Journal*, 329, 209–213.
- Phoenix, C., Schaefer, A. M., Elson, J. L., Morava, E., Bugiani, M., Uziel, G., et al. (2006). A scale to monitor progression and treatment of mitochondrial disease in children. *Neuromuscular Disorders*, 16(12), 814–820.
- Reinecke, C. J., Koekemoer, G., van der Westhuizen, F. H., Louw, R., Lindeque, J. Z., Mienie, L. J., et al. (2012). Metabolomics of urinary organic acids in respiratory chain deficiencies in children. *Metabolomics*, 8, 264–283.
- Rodenburg, R. T. (2011). Biochemical diagnosis of mitochondrial disorders. *Journal of Inherited Metabolic Disease*, 34, 283–292.
- Schaefer, A. M., Phoenix, C., Elson, J. L., McFarland, R., Chinnery, P. F., & Turnbull, D. M. (2006). Mitochondrial disease in adults:

- a scale to monitor progression and treatment. *Neurology*, 66(12), 1932–1934.
- Schaefer, A. M., Taylor, R. W., Turnbull, D. M., & Chinnery, P. F. (2004). The epidemiology of mitochondrial disorders—past, present and future. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1659, 115–120.
- Smuts, I., Louw, R., Du Toit, H., Klopper, B., Mienie, L. J., & van der Westhuizen, F. H. (2010). An overview of a cohort of South African patients with mitochondrial disorders. *Journal of Inherited Metabolic Disease*, 33(3), 95–104.
- Smuts, I., van der Westhuizen, F. H., Louw, R., Mienie, L. J., Engelke, U. H., Wevers, R. A., et al. (2013). Disclosure of a putative biosignature for respiratory chain disorders through a metabolomics approach. *Metabolomics*, 9, 379–391.
- Warrack, B. M., Hnatyshyn, S., Ott, K. H., Reily, M. D., Sanders, M., Zhang, H., et al. (2009). Normalization strategies for metabolomic analysis of urine samples. *Journal of Chromatography B*, 877, 547–552.
- Wolf, N. I., & Smeitink, J. A. (2002). Mitochondrial disorders a proposal for consensus diagnostic criteria in infants and children. *Neurology*, 59, 1402–1405.
- Xia, J., Broadhurst, D. I., Wilson, M., & Wishart, D. S. (2013). Translational biomarker discovery in clinical metabolomics: an introductory tutorial. *Metabolomics*, 9(2), 280–299.
- Xia, J., Psychogios, N., Young, N., & Wishart, D. S. (2009). MetaboAnalyst: A web server for metabolomic data analysis and interpretation. *Nucleic Acids*, 37, W652–W660.

## Supplementary material



**Fig S-1: Scatter plot showing the average ranking of each feature.** The gap between the first 12 top ranked features and rest were used as (preliminary) cut-off.