haematologica
Journal of The Ferrata Storti Foundation

Untargeted metabolic profiling in dried blood spots identifies disease fingerprint for Pyruvate Kinase Deficiency

by Birgit van Dooijeweert, Melissa H. Broeks, Nanda M. Verhoeven-Duif, Eduard J. van Beers, Edward E.S. Nieuwenhuis, Wouter W. van Solinge, Marije Bartels, Judith J. Jans and Richard van Wijk

Haematologica 2020 [Epub ahead of print]

Citation: Birgit van Dooijeweert, Melissa H. Broeks, Nanda M. Verhoeven-Duif, Eduard J. van Beers, Edward E.S. Nieuwenhuis, Wouter W. van Solinge, Marije Bartels, Judith J. Jans and Richard van Wijk. Untargeted metabolic profiling in dried blood spots identifies disease fingerprint for Pyruvate Kinase Deficiency. Haematologica. 2020;105:xxx
doi:10.3324/haematol.2020.266957

Manuscript Title:

Untargeted metabolic profiling in dried blood spots identifies disease fingerprint for Pyruvate Kinase Deficiency.

Short running title:

Metabolic profiling in Pyruvate Kinase Deficiency

Authors:

Birgit van Dooijeweert[1,2‡], Melissa H. Broeks[3‡] Nanda M. Verhoeven-Duif[3], Eduard J. van Beers[4], Edward E.S. Nieuwenhuis[2], Wouter W. van Solinge[1], Marije Bartels[2,4], Judith J. Jans[3‡] and Richard van Wijk[1‡]

[‡]These authors contributed equally

Affiliations:

1. Central Diagnostic Laboratory-Research, University Medical Center Utrecht, Utrecht, The Netherlands.
2. Department of Pediatric Hematology, University Medical Center Utrecht, Utrecht, The Netherlands.
3. Section Metabolic Diagnostics, Department of Genetics, University Medical Center Utrecht, Utrecht, The Netherlands.
4. Van Creveldkliniek, University Medical Center Utrecht, Utrecht, The Netherlands.

Corresponding author:

Full name: Birgit van Dooijeweert

E-mail address: b.vandooijeweert-3@umcutrecht.nl

Address: Heidelberglaan 100, 3584 CX, Utrecht, The Netherlands

Phone number: +31 (0)88 75 55280; Fax: +31 (0)88 75 55418

Main text word count:	1752
Abstract word count:	174
Number of figures and tables:	2
Supplemental files:	1 (4 supplemental tables/figures)
Number of references:	22

Key points:

- We identified a metabolic fingerprint for patients with Pyruvate Kinase Deficiency using untargeted metabolomics in dried blood spots.
- This approach opens up a novel area of diagnosis and research in the field of red blood cell disorders.

Abstract:

The diagnostic evaluation and clinical characterization of rare hereditary anemia (RHA) is to date still challenging. In particular, there is little knowledge on the broad metabolic impact of many of the molecular defects underlying RHA. In this study we explored the potential of untargeted metabolomics to diagnose a relatively common type of RHA: Pyruvate Kinase Deficiency (PKD). In total, 1903 unique metabolite features were identified in dried blood spot samples from 16 PKD patients and 32 healthy controls. A metabolic fingerprint was identified using a machine learning algorithm, and subsequently a binary classification model was designed. The model showed high performance characteristics (AUC 0.990, 95%CI 0.981-0.999) and an accurate class assignment was achieved for all newly added control (13) and patient samples (6), with the exception of one patient (accuracy 94%). Important metabolites in the metabolic fingerprint included glycolytic intermediates, polyamines and several acyl carnitines. In general, the application of untargeted metabolomics in dried blood spots is a novel functional tool that holds promise for diagnostic stratification and studies on disease pathophysiology in RHA.

Introduction:

The group of rare hereditary anemias (RHA) includes a large variety of intrinsic defects of the red blood cell and erythropoiesis. Our knowledge of the pathophysiology of RHA has recently vastly improved, powered by genetic testing and subsequent increased knowledge of underlying molecular defects.[1-4] However, in a substantial number of patients, the clinical phenotype does not fit classical criteria of disease, response to therapy is unexpectedly poor, or a molecular defect cannot be identified.[5-7] In addition, in patients with well-described genetic defects, there is often no clear genotype-phenotype correlation.[7-9]

Pyruvate kinase deficiency (PKD; OMIM 266 200), the most common red cell glycolytic enzyme defect, is no exception in this respect. The clinical phenotype of PKD varies widely, from a well-compensated hemolytic anemia to severe hemolysis and neonatal mortality. Currently the diagnosis of PKD relies on measurement of PK activity and/or the identification of homozygous or compound heterozygous mutations in the *PKLR* gene.[10,11]

However, in a significant number of patients only one mutation is identified. In addition, the exact mechanisms leading to reduced lifespan of PK-deficient erythrocytes are still largely unknown. Thus, in order to improve diagnostic evaluation as well as our understanding of PKD pathophysiology and the genotype-to-phenotype correlation, novel functional tests are needed.

In this study we demonstrate the potential of untargeted metabolomics in dried blood spots (DBS) in the diagnostic evaluation of PKD and report for the first time a metabolic fingerprint for PKD.

Methods:

Samples:

16 patients diagnosed with PKD based on clinical phenotype, enzyme activity assays and molecular defect were included. Healthy controls (HC; institutional blood donor service) served as controls. All patients or their legal guardians approved the use of remnant samples for method development and validation, in agreement with institutional and national regulations. All procedures followed were in accordance with the ethical standards of the University Medical Center Utrecht and with the Helsinki Declaration of 1976, as revised in 2000. For DBS, 50 µL aliquots were spotted onto Guthrie card filter paper (Whatman no. 903 Protein Saver TM cards). All papers were left to dry for at least four hours at room temperature, and subsequently stored at -80^{0}C in a foil bag with a desiccant package pending further analysis.

Metabolic profiling

Sample preparation, direct infusion high resolution mass spectrometry (DI-HRMS) and data processing was performed as previously reported.[12, 13] Mass peak intensities for metabolite annotation were averaged over technical triplicates. In addition, as DI-HRMS is unable to separate isomers, mass peak intensities consisted of summed intensities of these isomers. Metabolite annotation was performed using a peak calling bioinformatics pipeline developed in R programming software, based on the human metabolome database (version 3.6) (https://github.com/UMCUGenetics/DIMS). This resulted in 3835 metabolite annotations corresponding to 1903 unique metabolite features.[14]

To compare the metabolic profiles between HC and PKD, mass peak intensities for each identified feature were converted to Z-scores. These scores, based on metabolic control samples that were added to each DI-HRMS run, were calculated by the following formula:

$$Z-score = \frac{(Mass\ peak\ intensity\ of\ Pt\ or\ HC\ sample - Mean\ mass\ peak\ intensities\ of\ metabolic\ control\ samples}{Standard\ deviation\ mass\ peak\ intensities\ of\ metabolic\ control\ samples\ ‡}$$

‡Metabolic controls exist of a batch of banked DBS samples from individuals in whom an inborn error of metabolism (IEM) was excluded after an extensive diagnostic workup.

Data analysis

T-test and multivariate analysis were conducted in MetaboAnalyst.[15] Classification of data was performed in R-software (Version 3.6.1) using the *caret* package, which contains a set of data processing functions that facilitate the generation of predictive models. Support vector machine (SVM) with linear kernel was used for the classification of HC and PKD samples. SVM algorithms use a set of mathematical functions that are defined as the kernel. The function of kernel is to take data as input and transform it into the required form, for example a linear or polynomial kernel. We applied SVM with a linear kernel, the simplest kernel function, to perform the classification of HC and PKD. SVM with linear kernel is a supervised machine learning model that uses a classification method, which is based on mapping the data into a high dimensional space. This allows the separation of two groups of samples into distinctive regions by the identification of a small fraction of samples that separates the groups, also referred to as 'support vectors'. Separation can be achieved by identifying a separating hyperplane, or decision boundary, between the support vectors.[16] Classification of the test set was determined by projecting each of the new samples into this space. Data and R code are available upon request.

Results

Explorative untargeted metabolomics analysis

A total of 1903 unique metabolite features (and their respective isomers) were analyzed for 16 PKD patients and 32 HC-samples. Clinical and laboratory characteristics, and baseline comparison are summarized in Table 1. The most significant differences between groups, identified by a t-test, included glycolytic intermediates like phosphoenolpyruvic acid and 2-/3-phosphoglyceric acid, polyamines (spermidine and spermine) and several acyl carnitines (methylmalonylcarnitine and propionylcarnitine) (Figure 1A). Broad data exploration to assess the variation between samples and

separation between groups was performed by unsupervised principal component analysis (PCA) and supervised partial least square discriminant analysis (PLS-DA), the latter taking group label into account as a response variable. Both analyses revealed close clustering of control samples and a more heterogeneous delineation for PKD patients (Supplemental Figure 1.).

Machine learning algorithm identifies metabolic profile for PKD

To explore the potential of this extensive metabolic fingerprint in predicting PKD a binary classification model was constructed using a support vector machine (SVM) with linear kernel. SVM has advantages over PLS-DA with regard to robustness to outliers, resistance to overfitting and predictive power.[16] An optimal hyperplane to separate classes based on all metabolomics data was determined by cross validation (4-fold, 5 repeats). The final model had high performance characteristics with an average accuracy of 96%.

In addition, receiver operator characteristic curves with area under the curve were used as performance indicator (Supplemental Figure 2A). Important features for classification in this model include the polyamines spermidine and spermine, as well as phosphoenolpyruvic acid, 2-/3-phosphoglyceric acid and glutathione (Figure 1B). Most of these features were increased in PKD, with the exception of glutathione and asparaginyl-proline/prolyl-asparagine (Figure 1C).

Metabolic profile predicts new samples with high accuracy

External model validation was performed by predicting new control (n=13) and PKD-samples (n=6). This resulted in accurate prediction for all controls, and all but one patient (Accuracy = 94%) (Figure 1D). To assess uncertainty of the model and its predictive ability, bootstrap resampling was applied to the complete dataset. By randomly generating training and validation (test) data from the original data, a similarly high prediction performance was achieved, supporting the validity of the presented model (Supplemental Figure 2B).

Metabolic profiles reflect PKD disease severity

To explore the heterogeneity of PKD metabolic profiles in relation to clinical phenotype, PCA and PLS-DA were performed for the entire group of patients and controls. Based on presence of spleen and transfusion frequency phenotypes were distinguished as mild, moderate and severe. Most resemblance to controls in metabolic profile was clear for mild phenotypes, followed by severely affected patients (Supplementary Figure 4).

Discussion

In this study we have performed untargeted metabolomics on dried blood spots and report for the first time a metabolic disease fingerprint for PKD. By establishing a predictive machine learning model, the diagnostic potential of this approach was demonstrated. This metabolic fingerprint has potential to mature into a powerful clinical tool, capable of confirming or ruling out the diagnosis of PKD. However, the limitations of machine learning models were also demonstrated by the incorrect classification of one PKD patient who was homozygous for the common p.(Arg510Gln) mutation.[17] Clinically, this patient exhibited very mild phenotypic features. As confirmed by the clinical severity PLS-DA, patients with a mild phenotype and controls overlap most in their DBS-metabolome (Supplementary Figure 4). Since ~30% of the initial cohort consists of such mildly affected patients, this could further explain why PCA and PLS-DA were unable to achieve separation between groups. Interestingly, severely affected patients who are heavily transfused (>6 erythrocyte transfusions in the past 12 months) despite having underwent a splenectomy, still showed a clearly distinctive metabolic profile compared to HC's and two of them were furthermore correctly assigned as patients (Figure 1D, Supplementary Table 1). Although numbers are modest and further studies are needed, this indicates that this approach is reliable even in the setting of transfusions.

Our approach using untargeted metabolomics provides novel insights regarding the broad metabolic impact of PKD that could be relevant to better understand the etiology of PKD-related symptoms. While glycolytic metabolites and their disturbance have been

characterized to some extent, little is known regarding the broad scale impact of PKD on metabolism. In this respect, the identification of novel distinctive metabolites, such as polyamines, which have been found to stabilize the red blood cell (RBC) plasma membrane,[18] and acyl carnitines, which are involved in turnover and repair of the RBC membrane,[19] are promising starting points for further study into PKD pathophysiology.

We here report for the first time a metabolic profile for PKD obtained from dried whole blood spots. This approaches the integrated disease specific metabolome to a greater extent compared to the exclusive investigation of the red blood cell metabolome.[20, 21] In addition, this analysis requires only 50 µL of whole blood and can be obtained in a minimally invasive manner by sampling a single blood drop, making it very attractive for (international) sample exchange. Further advantages of DI-HRMS include relatively uncomplicated sample extraction steps and a short run-time of 3 minutes per sample.

The rise of 'omic' approaches in the recent past has provided new opportunities for understanding and classifying a wide range of disorders. In contrast to conventional medical biology approaches, which focus on individual genes, proteins or metabolites, modern biology regards diseases as a complex, dynamic and especially integrated network.[22] Our study, demonstrates the potential diagnostic application of untargeted metabolomics for PKD. However, the current model was constructed for the binary classification of healthy controls and PKD patients. Future applications, including more samples from various types of RHA could enable development of an algorithm which is suited for the broader differential diagnosis of RHA in patients.

In conclusion, we demonstrate by proof of principle for PKD, that untargeted metabolomics in DBS is a novel functional tool to identify disease fingerprints and study pathophysiology in RHA.
This approach opens up a novel area of diagnosis and research in the field of red blood cell disorders and has the potential to improve diagnostic evaluation and clinical management of patients.

Acknowledgements:

This study was supported in part by research funding from Metakids (Grant No. 2017-075) to J.J.

The authors would like to thank Nienke van Unen for her technical support in Bio-informatics.

References:

1. Da Costa L, Narla A, Mohandas N. An update on the pathogenesis and diagnosis of Diamond-Blackfan anemia. F1000Res. 2018;7**:**F1000 Factulty Rev-1350.

2. Grace RF, Zanella A, Neufeld EJ, et al. Erythrocyte pyruvate kinase deficiency: 2015 status report. Am J Hematol. 2015;90(9):825-830.

3. Albuisson J, Murthy SE, Bandell M, et al. Dehydrated hereditary stomatocytosis linked to gain-of-function mutations in mechanically activated PIEZO1 ion channels. Nat Commun. 2013;4:1884.

4. Roy NBA, Babbs C. The pathogenesis, diagnosis and management of congenital dyserythropoietic anaemia type I. Br J Haematol. 2019;185(3):436-449.

5. Vercellati C, Marcello AP, Fermo E, Barcellini W, Zanella A, Bianchi P. A case of hereditary spherocytosis misdiagnosed as pyruvate kinase deficient hemolytic anemia. Clin Lab. 2013;59(3-4):421-424.

6. Steinberg-Shemer O, Keel S, Dgany O, et al. Diamond Blackfan Anemia: A Nonclassical Patient With Diagnosis Assisted by Genomic Analysis. J Pediatr Hematol Oncol. 2016;38(7):e260-262.

7. Russo R, Andolfo I, Manna F, et al. Multi-gene panel testing improves diagnosis and management of patients with hereditary anemias. Am J Hematol. 2018;93(5):672-682.

8. van Dooijeweert B, van Ommen CH, Smiers FJ, et al. Pediatric Diamond-Blackfan anemia in the Netherlands: An overview of clinical characteristics and underlying molecular defects. Eur J Haematol. 2018;100(2):163-170.

9. Zanella A, Fermo E, Bianchi P, Chiarelli LR, Valentini G. Pyruvate kinase deficiency: the genotype-phenotype association. Blood Rev. 2007;21(4):217-231.

10. Grace RF, Bianchi P, van Beers EJ, et al. Clinical spectrum of pyruvate kinase deficiency: data from the Pyruvate Kinase Deficiency Natural History Study. Blood. 2018;131(20):2183-2192.

11. Bianchi P, Fermo E, Glader B, et al. Addressing the diagnostic gaps in pyruvate kinase deficiency: Consensus recommendations on the diagnosis of pyruvate kinase deficiency. Am J Hematol. 2019;94(1):149-161.

12. Haijes HA, Willemsen M, Van der Ham M, et al. Direct Infusion Based Metabolomics Identifies Metabolic Disease in Patients' Dried Blood Spots and Plasma. Metabolites. 2019;9(1):12.

13. de Sain-van der Velden MGM, van der Ham M, Gerrits J, et al. Quantification of metabolites in dried blood spots by direct infusion high resolution mass spectrometry. Anal Chim Acta. 2017;979:45-50.

14. Wishart DS, Jewison T Fau - Guo AC, Guo Ac Fau - Wilson M, et al. HMDB 3.0--The Human Metabolome Database in 2013. (1362-4962 (Electronic)).

15. Chong J, Wishart DS, Xia J. Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. Curr Protoc Bioinformatics. 2019;68(1):e86.

16. Gromski PS, Muhamadali H, Ellis DI, et al. A tutorial review: Metabolomics and partial least squares-discriminant analysis--a marriage of convenience or a shotgun wedding. Anal Chim Acta. 2015;879:10-23.

17. Bianchi P, Fermo E, Lezon-Geyda K, et al. Genotype-phenotype correlation and molecular heterogeneity in pyruvate kinase deficiency. Am J Hematol. 2020;95(5):472-482.

18. Ballas SK, Mohandas N, Marton LJ, Shohet SB. Stabilization of erythrocyte membranes by polyamines. Proc Natl Acad Sci U S A. 1983;80(7):1942-1946.

19. Arduini A, Mancinelli G, Radatti GL, Dottori S, Molajoni F, Ramsay RR. Role of carnitine and carnitine palmitoyltransferase as integral components of the pathway for membrane phospholipid fatty acid turnover in intact human erythrocytes. J Biol Chem. 1992;267(18):12673-12681.

20. Darghouth D, Koehl B, Madalinski G, et al. Pathophysiology of sickle cell disease is mirrored by the red blood cell metabolome. Blood. 2011;117(6):e57-66.

21. Darghouth D, Koehl B, Heilier JF, et al. Alterations of red blood cell metabolome in overhydrated hereditary stomatocytosis. Haematologica. 2011;96(12):1861-1865.

22. Tebani A, Afonso C, Marret S, Bekri S. Omics-Based Strategies in Precision Medicine: Toward a Paradigm Shift in Inborn Errors of Metabolism Investigations. Int J Mol Sci. 2016;17(9):1555.

Tables & Figures

Table 1A. Clinical characteristics of PKD patients and baseline comparison to healthy controls

Age (yrs)	Gender	Hb (mmol/L)	RBC ($x10^{12}$/L)	Retics ($x10^9$/L)	WBC ($x10^9$/L)	Plts ($x10^9$/L)	Treatment	Allele 1	Allele 2
65	female	6.4	2.84	1014	7.3	576	splenectomy; no current treatment	c.1178A>G; p.(Asn393Ser)	not identified
2	female	4.8	2.59	343	13.4	345	regular transfusions	c.331G>A; p.(Gly111Arg)	c.331G>A; p.(Gly111Arg)
6	female	6.6	3.79	431	10.6	645	splenectomy; sporadic transfusion	c.331G>A; p.(Gly111Arg)	c.331G>A; p.(Gly111Arg)
51	female	9.1	5.26	37.8	4.97	188	no current treatment	c.1456C>T; p(Arg486Trp)	c.1529G>A; p.(Arg510Gln)
28	male	5.0	2.27	1011	21.1	ND	splenectomy; sporadic transfusion	c.1073G>A; p.(Gly358Glu)	c.1073G>A; p.(Gly358Glu)
29	female	6.3	2.99	180	7.1	239	no current treatment	c.142_159del; p.(Thr48_Thr53 del)	c.1269G>A; p.(?)
23	female	6.3	2.52	ND	10.0	696	splenectomy; no current treatment	c.1269G>A; p.([Met373_Ala423del;0])	c.1654G>A; p.(Val553Met)
35	male	6.2	3.45	198	5.5	179	no current treatment	c.194T>C; p.(Met65Thr)	c.721G>T; p.(Glu241*)
48	female	4.1	1.79	694	9.8	657	splenectomy; no current treatment	c.1462C>T; p(Arg488*)	c.1529G>A; p(Arg510Gln)
25	male	8.4	3.76	627	12.4	732	splenectomy; no current treatment	c.142_159del; p.(Thr48_Thr53 del)	c.494G>T(p.Gly165Val)
48	male	5.4	2.18	945	13.1	876	splenectomy; no current treatment	c.376-2A>C; p. (?)	c.1529G>A; p. (Arg510Gln)
21	male	7.2	3.60	181	5.5	245	no current treatment	c.390_392het_delCAT; p.(Ile131del)	c.1456C>T; p(Arg486Trp)
51	female	5.6	2.86	950	12.3	719	splenectomy; regular transfusions	c.507+1G>A; p.[=;0]	c.1436G>A; p.(Arg479His)
24	male	ND	ND	ND	ND	ND	splenectomy; no current treatment	c.694G>T; p.(Gly232Cys)	c.1529G>A; p. (Arg510Gln)
20	female	7.2	3.60	112	6.4	186	regular transfusions	c.1529G>A; p.(Arg510Gln)	c.1705C>T; p.(Arg569Trp)**
46	male	7.6	3.74	204	6.0	327	no current treatment	c.1121T>C; p.(Leu374Pro)	c.1706G>A; p.(Arg569Glu)
Normal range*		7.4-10.7	3.6-5.5	25-120	4.0-13.5	150-450			

*Age and gender dependent

** A third and rare mutation (c. 1639C>T; p(Arg547Cys)) with uncertain pathogenicity was identified in this patient

Table 1B. Baseline comparison to controls

	PKD	HC
Age (years)	32.6 ± 17.4	38.9 ± 12.8
Hb (mmol/L)	6.41 ± 1.35	9.12 ± 0.7
Retics ($x10^9$/L)	494.8 ± 367.5	58.6 ± 22.4
Median time to DBS (hours)	2.33	2.21

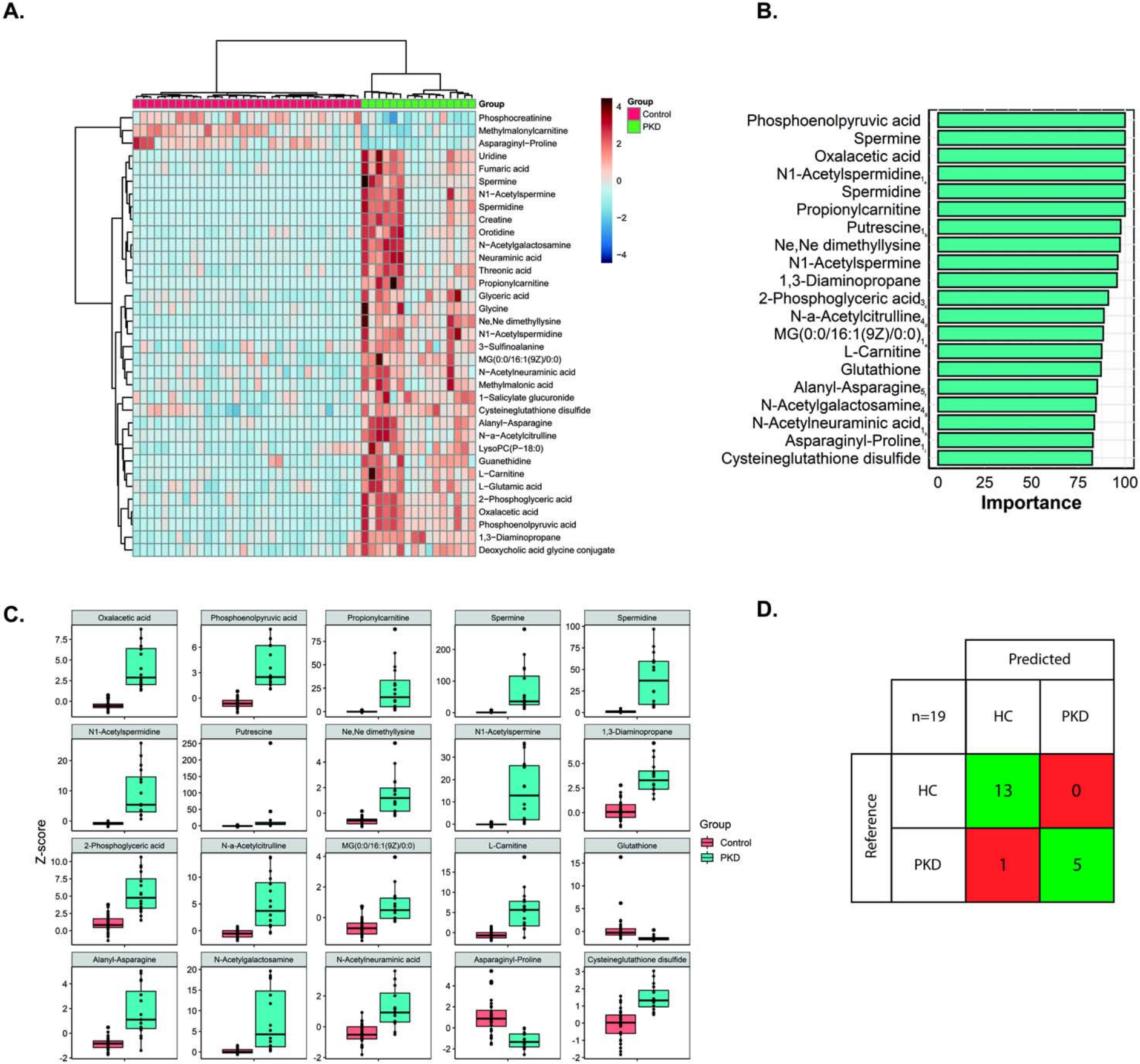
Table 1. Clinical characteristics of PKD patients and baseline comparison to healthy controls.

A. Clinical characteristics of PKD patients regarding age, gender, Hemoglobin (Hb), Red blood cell count (RBC) , Reticulocyte count (Retics), White blood cell count (WBC), platelets (Plts), treatment and genetic diagnostics. Regular transfusions are defined as ≥6 per 12 months. ND = not determined. **B**. Comparison of age, Hb, Retics and time between blood withdrawal and spotting (time to DBS) between healthy controls (HC) and PKD patients. Data are presented as mean ± SD, except for time to time to DBS which is presented as the median.

Figure Legends:

Figure 1. Univariate and multivariate analysis of untargeted metabolomics data from PKD patients and healthy controls

A. Heatmap of top 35 significant features identified by t-test (p-value cutoff =0.05). The heatmap was created using Euclidean ward clustering with autoscaling of features. **B.** Top 20 important features represented as percentage identified by support vector machine classification. As isomers could not be distinguished using DI-HRMS, the annotated numbers near the important features indicate the amount of isomers. In addition, letters in the footnote correspond to the following isomers: a) N8-Acetylspermidine, b) 1,4-Butanediammonium, c) 3-phosphoglyceric acid; 2-phospho-D-glyceric acid; (2R)-2-Hydroxy-3-(phosphonatooxy)propanoate, d) Alanyl-Glutamine; Alanyl-Gamma-glutamate; Glutaminyl-Alanine; Gamma-glutamyl-Alanine, e) MG(16:1(9Z)/0:0/0:0), f) Asparaginyl-Alanine; Glutaminyl-Glycine; Glycyl-Glutamine; Glycycl-Gamma-glutamate; Gamma-glutamyl-Glycine, g) N-Acetyl-D-glucosamine; Beta-N-Acetylglucosamine; N-Acetyl-b-D-galactosamine; N-Acetylmannosamine, h) N-Acetyl-a-neuraminic acid, i) Prolyl-Asparagine. **C.** Boxplots of each feature showing Z-scores for control and PKD groups, respectively. **D.** Confusion matrix for the prediction of additional samples by the SVM model.

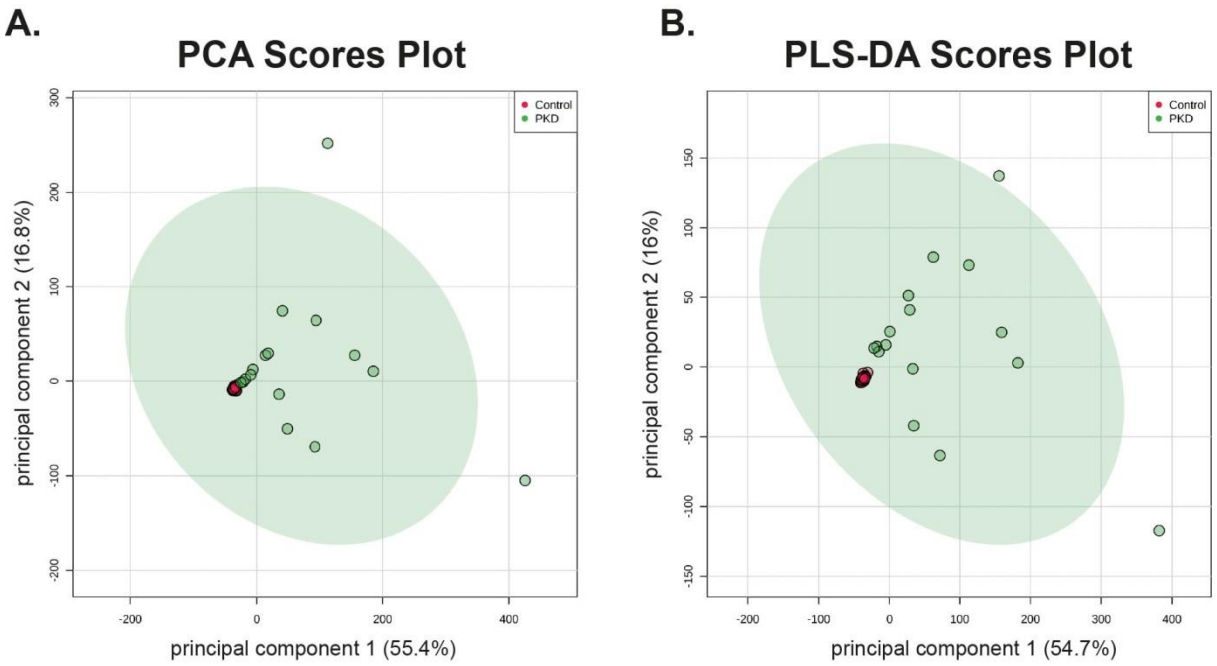

A.
Group
Control
PKD
4
2
0
−2
−4
Phosphocreatinine
Methylmalonylcarnitine
Asparaginyl−Proline
Uridine
Fumaric acid
Spermine
N1−Acetylspermine
Spermidine
Creatine
Orotidine
N−Acetylgalactosamine
Neuraminic acid
Threonic acid
Propionylcarnitine
Glyceric acid
Glycine
Ne,Ne dimethyllysine
N1−Acetylspermidine
3−Sulfinoalanine
MG(0:0/16:1(9Z)/0:0)
N−Acetylneuraminic acid
Methylmalonic acid
1−Salicylate glucuronide
Cysteineglutathione disulfide
Alanyl−Asparagine
N−a−Acetylcitrulline
LysoPC(P−18:0)
Guanethidine
L−Carnitine
L−Glutamic acid
2−Phosphoglyceric acid
Oxalacetic acid
Phosphoenolpyruvic acid
1,3−Diaminopropane
Deoxycholic acid glycine conjugate
B.
Phosphoenolpyruvic acid
Spermine
Oxalacetic acid
N1-Acetylspermidine
Spermidine
Propionylcarnitine
Putrescine
Ne,Ne dimethyllysine
N1-Acetylspermine
1,3-Diaminopropane
2-Phosphoglyceric acid
N-a-Acetylcitrulline
MG(0:0/16:1(9Z)/0:0)
L-Carnitine
Glutathione
Alanyl-Asparagine
N-Acetylgalactosamine
N-Acetylneuraminic acid
Asparaginyl-Proline
Cysteineglutathione disulfide
0
25
50
75
100
Importance
C.
Oxalacetic acid
Phosphoenolpyruvic acid
Propionylcarnitine
Spermine
Spermidine
N1-Acetylspermidine
Putrescine
Ne,Ne dimethyllysine
N1-Acetylspermine
1,3-Diaminopropane
2-Phosphoglyceric acid
N-a-Acetylcitrulline
MG(0:0/16:1(9Z)/0:0)
L-Carnitine
Glutathione
Alanyl-Asparagine
N-Acetylgalactosamine
N-Acetylneuraminic acid
Asparaginyl-Proline
Cysteineglutathione disulfide
Z-score
Group
Control
PKD
D.
Predicted
n=19
HC
PKD
Reference
HC
13
0
PKD
1
5

Supplemental information

Supplementary Table 1. Clinical characteristics of additional PKD patients (test-cohort)

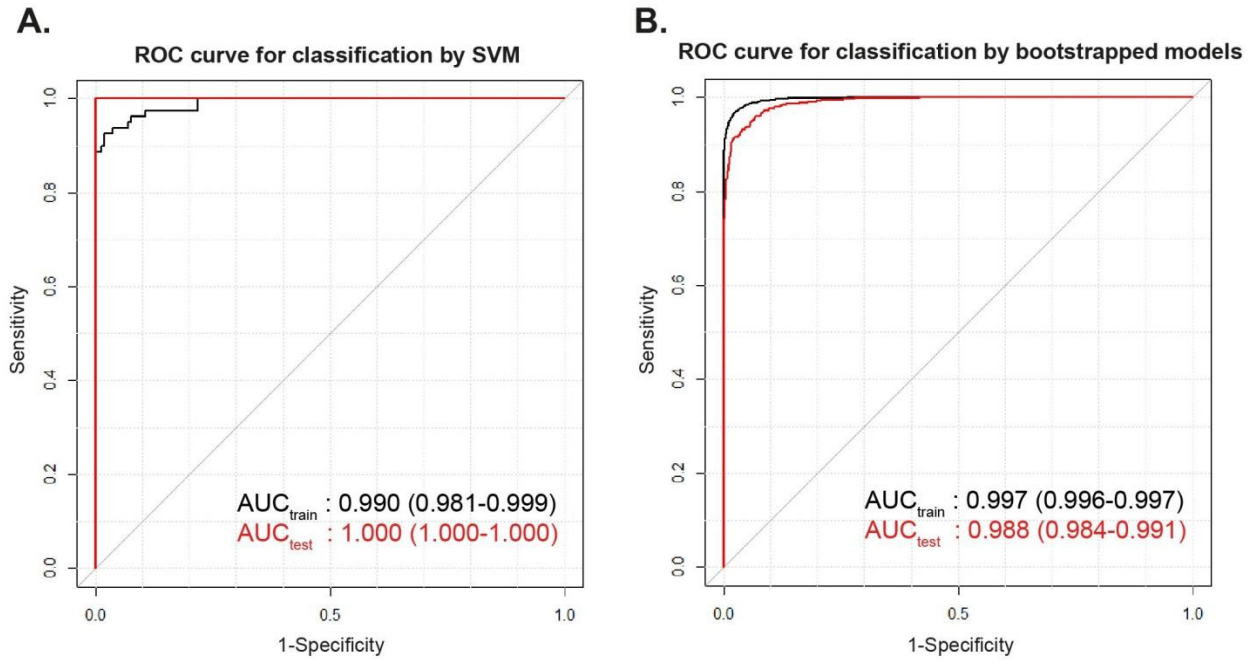
Age (yrs)	Gender	Hb (mmol/l)	RBC ($x10^{12}$/L)	Retics ($x10^{9}$/L)	WBC ($x10^{9}$/L)	Plts ($x10^{9}$/L)	Treatment	Allele 1	Allele 2
43	female	5.5	2.20	1045	11.6	714	splenectomy; no current treatment	c.401T>A; p.(Val134Asp)	c.1529G>A; p.(Arg510Gln)
44	female	5.7	2.75	756	9.3	841	splenectomy; regular transfusions	c.283G>A; p.(Gly95Arg)	c.401T>A; p.(Val134Asp)
32	female	6.5	3.49	338	9	464	splenectomy; regular transfusions	c.283G>A; p.(Gly95Arg)	c.401T>A; p.(Val134Asp)
26	male	5.0	2.33	930	15.7	780	splenectomy; sporadic transfusion	c.721C>T; p.(Glu241*)	c.1529G>A; p.(Arg510Gln)
4	*female*	*7.6*	*4.00*	*194*	*7.8*	*275*	*no current treatment*	*c.1529G>A; p.(Arg510Gln)*	*c.1529G>A; p.(Arg510Gln)*
54	male	6.2	3.08	166	5.3	150	no current treatment	c.142_159del; p.(Thr48_Thr53 del)	c.376-2A>C; p.(?)
Normal range*		7.4-10.7	3.6-5.5	25-120	4.0-13.5	150-450			

Abbrevations: Hb = hemoglobin, Retics = reticulocytes, Plts = platelets. Regular transfusions defined as ≥6 per 12 months.
Italic: patient who was predicted as control in predictive algorithm



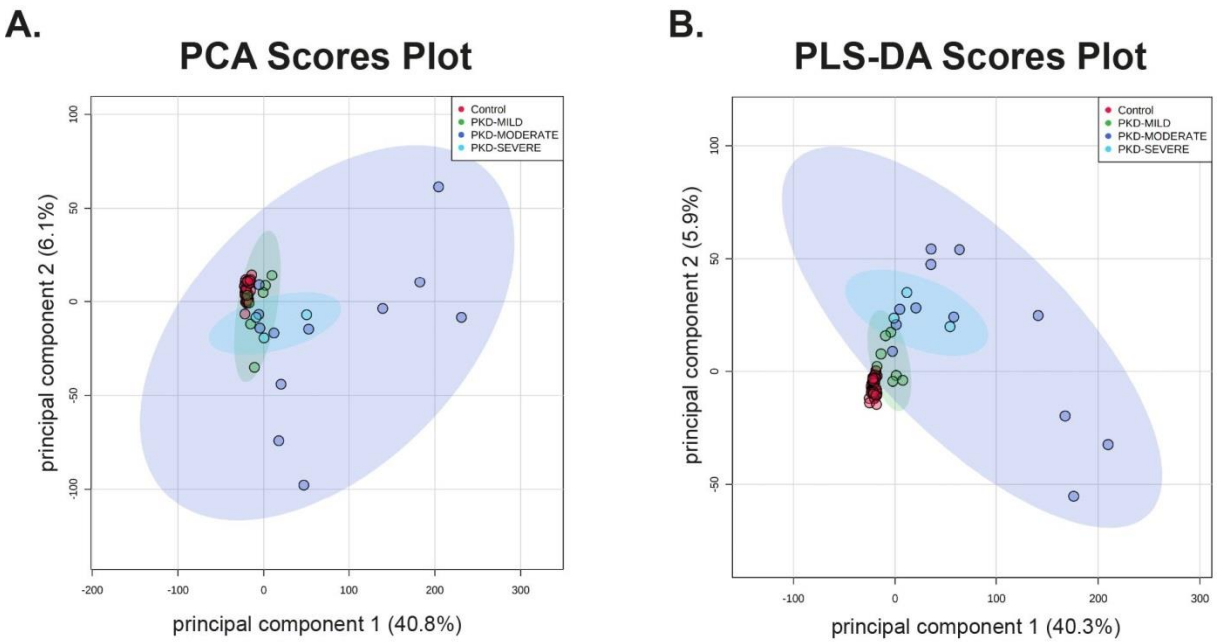
Supplementary Figure 1. Scores plots for PCA and PLS-DA

A) Principal component analysis (PCA) of PKD and control groups. B) Partial least square discriminant analysis (PLS-DA) of PKD and control groups.



Supplementary Figure 2. Receiver Operator Characteristic curves (ROC) for classification of training and test sets by SVM models.

A. Classification performance of samples in training and test set according to AUC. Note that AUC is a measure of the ability to rank samples according to the probability of class membership, meaning that even falsely classified samples can have a higher rank towards the correct class compared to other samples. B. Classification performance of samples in bootstrap models (n=100) of the complete data set.



Supplementary Figure 3. Multivariate analysis with distinction of phenotype severity

A. PCA plot, and B. PLS-DA plot distinguishing between disease phenotypes based on transfusion dependence and splenectomy. Most resemblance in metabolic profile is clear for mild phenotypes, followed by severely affected PKD patients (possibly related to interference of frequent transfusions).