Analyst

Rapid screening of classical galactosemia patients: a proof-of-concept study using high-throughput FTIR analysis of plasma

SCHOLARONE™
Manuscripts

Rapid screening of classical galactosemia patients: a proof-of-concept study using high-throughput FTIR analysis of plasma

Caroline Lacombe[1,2†], Valérie Untereiner[1,2†], Cyril Gobinet[1,2], Moktar Zater[3], Ganesh D. Sockalingum[1,2§], Roselyne Garnotel[1,2,4§*]

[1]Université de Reims Champagne-Ardenne, Equipe MéDIAN, Biophotonique et Technologies pour la Santé, UFR de Pharmacie, 51 rue Cognacq-Jay, 51096 Reims, France

[2]CNRS UMR 7369, Unité MEDyC (Matrice Extracellulaire et Dynamique Cellulaire), Reims, France

[3]Biochimie - Hôpital de Bicêtre, Hôpitaux Universitaires Paris-Sud, France

[4]CHU de Reims, Laboratoire de Biologie et Recherche Pédiatriques, 51092 Reims, France

caroline_lacombe@hotmail.fr

valerie.untereiner@univ-reims.fr

cyril.gobinet@univ-reims.fr

moktar.zater@bct.aphp.fr

ganesh.sockalingum@univ-reims.fr

rgarnotel@chu-reims.fr

[†] CL and VU have contributed equally to this study.

§ GDS and RG have equally managed this study.

*Corresponding author :

Dr Roselyne Garnotel

Equipe MéDIAN, Biophotonique et Technologies pour la Santé

Université de Reims Champagne-Ardenne, UFR de Pharmacie

CNRS UMR 7369, Unité MEDyC

51 rue Cognacq-Jay, 51096 Reims Cedex, France

rgarnotel@chu-reims.fr

Tél : +33.3.26.78.79.55; Fax : +33.3.26.78.84.56

Abstract

The classical galactosemia is an autosomal recessive metabolic disease involved in the galactose pathway, caused by deficiency of galactose-1-phosphate uridyltransferase. Galactose accumulation induces in newborns many symptoms, such as liver disease, cataracts, and sepsis leading to death if untreated. Neonatal screening is developed and applied in many countries using several methods to detect galactose or its derived products accumulation in blood or urine. High-throughput FTIR spectroscopy was investigated as a potential tool to the current screening methods. IR spectra were obtained from blood plasma from healthy, diabetic, and galactosemic patients. The major spectral differences were in the carbohydrate region, which was firstly analysed in an exploratory manner using principal component analysis (PCA). PCA score plots showed a clear discrimination between diabetic and galactosemic patients and this was more marked as a function of the glucose and galactose increased concentration in these patient plasmas respectively. Then, a support vector machine leave-one-out cross-validation (SVM-LOOCV) classifier was built with the PCA scores as input and the model was tested on median, mean and all spectra from the three population groups. This classifier was able to discriminate healthy/diabetic, healthy/galactosemic, and diabetic/galactosemic patients with sensitivity and specificity rates ranging from 80% to 94%. The total accuracy rate ranged from 87% to 96%. High-throughput FTIR spectroscopy combined with SVM-LOOCV classification procedure appears to be a promising tool in the screening of galactosemia patients, with good sensitivity and specificity. Further, this approach presents the advantages of being cost-effective, fast, and straight-forward in the screening galactosemic patients.

Keywords: galactosemia, FTIR spectroscopy, plasma, high-throughput screening, classifier

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Introduction

Classical galactosemia, a rare disease, also known as type I galactosemia, is a metabolic disease involved in the Leloir pathway [1]. This is an autosomal recessive disorder caused by a deficiency of galactose-1-phosphate uridyltransferase (GALT; EC 2.7.7.12). The Leloir pathway is responsible for the conversion of galactose into glucose, and GALT, more specifically of galactose-1-phosphate (Gal-1-P), into UDP-galactose (Fig. S1). Galactose is primarily derived from the lactose content of milk that newborns receive. In classical galactosemia, newborns cannot metabolise galactose leading to symptoms such as jaundice, failure to thrive, liver disease, cataracts, hepatosplenomegaly, and an accumulation of galactose and its derived products (Gal-1-P and galactitol) in blood and urines [2]. If the newborn stays untreated, death in early infancy occurs due to sepsis, especially caused by *Escherichia coli*. Currently, the only known treatment is a dietary regimen excluding galactose and/or lactose.

The Leloir pathway is also composed of two additional enzymes, themselves responsible for other types of galactosemia. Upstream the GALT, the galactokinase (GALK; EC 2.7.1.6) converts galactose into galactose-1-phosphate and is responsible for type II galactosemia. Downstream, the uridine diphosphate-galactose 4' epimerase (GALE; EC 5.1.3.2) converts UDP-galactose into UDP-glucose and is responsible for type III galactosemia. These two types of galactosemia are less severe than the type I but they nevertheless present some of the type I symptoms [3, 4].

Many methods have been used to detect classical galactosemia, the older using bacterial growth in presence of galactose [5]. Currently, two types of methods exist: those which quantify the concentration of galactose and/or its derived products, and those which assay GALT activity. For the first ones, a sensitive bioluminescent assay [6] and more recently a gas

3

chromatography/mass spectrometry (GC/MS) using stable isotope of galactose on blood [7, 8] and urine samples [9] as well as a high performance liquid chromatography (HPLC) coupled with a pulsed amperometric detection (or fluorescent detection) of galactose and Gal-1-P [10, 11] have been described. For GALT activity assays, radiometric, spectrophotometric, HPLC techniques, and fluorometric methods such as the old Beutler's [12] or Benedict's [13] tests are used. However, to date, the most commonly used method is a radiochemical assay in which conversion of ^{14}C-Gal-1-P to ^{14}C-UDP-Gal is measured using an anion-exchange chromatography or a thin layer chromatography with quantification of radioactivity by scintillation counting. But recently, new approaches have been described, such as a liquid chromatography-tandem mass spectroscopy (LC-MS/MS) [14], HPLC without using radioactive labels [15], and an ultra-performance liquid chromatography-tandem mass spectroscopy (UPLC-MS/MS) [16-18]. Moreover, analysis of mutations in *GALT* gene can also be used as a screening method, but it is rather a complementary analysis [19]. Besides these analytical techniques, biophotonic approaches appear as alternative methods of screening.

In this perspective, a new and promising methodological approach based on Fourier-transform infrared (FTIR) spectroscopy has proved its potential to detect disease *via* specific spectral biomarkers. FTIR spectroscopy measures molecular vibrations through which molecular composition and structure of macromolecules can be studied, either isolated [20, 21] or in complex biological systems like cells and tissues [22-25]. It has been recently applied to biofluids for screening diseases [26], such as Alzheimer's disease [27], vascular disorder Hereditary Hemorrhagic Telangiectasia (HHT) [28], hepatic fibrosis [29] or hepatocellular carcinoma [30]. It can in a single measurement detect spectral variations linked to various molecular constituents, such as nucleic acids, glucids, proteins or lipids, present in the sample, in a qualitative and quantitative way [31, 32]. Compared to biochemical methods, diagnostics by FTIR spectroscopy has proven to be reagent-free, simpler, cost-effective, and faster. A recent

review by Baker et al. highlights all aspects of FTIR spectroscopy in the analysis of biological materials [33].

The aim of this feasibility study is to assess, using plasma samples, the potential of FTIR spectroscopy as a screening method for differentiating on the one hand healthy from diabetic patients with accumulation of glucose and on the other hand classical galactosemic patients suffering of an accumulation of galactose and derived products. Spectral data obtained from the three populations were analyzed by multivariate statistical methods for sensitivity and specificity evaluation.

Methods

Galactosemic, diabetic, and healthy plasma samples

Plasma samples were collected from 3 sets of patients: healthy (n=47), diabetic (n=19), and galactosemic (n=30). Samples from diabetic and healthy patients were obtained from the Reims University Hospital whereas galactosemic plasmas were kindly given by Dr. A. Boutron from Bicêtre Hospital in Paris and by Dr. E. Jeannesson from Nancy University Hospital. Other patient information was also collected: sex and age for all patients, concentration of glucose for healthy patients (Table S1), concentration of glucose and hemoglobin Alc level for diabetic patients (Table S2), concentration of Gal-1-P, and mutation for galactosemic patients (Table 1). All plasma samples were collected in tubes containing heparin as anticoagulant (BD Biosciences, Heidelberg, Germany) and stored at -80°C until use.

Biochemical assays and genotyping

Biochemical parameters, glucose, and HbA1c, were measured with a Modular analyzer (Boehringer Mannheim, Meylan, France) and with a Variant II analyzer (Bio-Rad Laboratories, Marne-la-Coquette, France) respectively, according to the manufacturers' instructions. For galactose-1-phosphate quantification, after red cells deproteinization, galactose was produced from galactose-1-phosphate using alkaline phosphatase. Galactose was then oxidized using galactose dehydrogenase to β-galactonolactone with concomitant conversion of NAD^+ to NADH assayed by spectrophotometry [12]. Genotyping of galactosemic patients was performed according to Boutron *et al.*[34].

FTIR spectroscopic analysis

The workflow from sample preparation to results outcome is described in Fig. 1. After thawing, all plasmas were diluted threefold in sterile water, deposited to cover the spots of a 384-well silicon plate (Bruker Optics GmbH, Ettlingen, Germany), and then air-dried at room temperature. For each sample, 10 spots were realized each containing 5 µL. After drying, the plate was inserted in a high-throughput system (HTS-XT, Bruker Optics GmbH) coupled to a FTIR spectrometer (Tensor 27, Bruker Optics GmbH). FTIR spectra were acquired in the transmission mode using the OPUS v6.5 software (Bruker Optics GmbH) and using the following conditions: wavenumber range from 4000 to 400 cm^{-1}, spectral resolution of 4 cm^{-1}, and each spectrum was averaged over 32 scans (i.e., an acquisition time of 30 s/spectrum). Thus, for each sample, 10 replicate spectra were acquired. All spectra were then subjected to a quality test (OPUS v6.5) which takes into account the absorbance intensity threshold, the signal-to-noise ratio, and the presence of water content [35]. Spectra with a maximum absorbance less than 0.35 and more than 1.5 in arbitrary units (a.u.) were discarded. To calculate the signal-to-noise ratio, the signal was taken as the maximum absorbance of the amide I band between 1700 cm^{-1} and 1600 cm^{-1} (S1 value), and between 1260 cm^{-1} and 1170

cm^{-1} (S2 value). Noise intensity (N value) was calculated in the 2100-2000 cm^{-1} region, which is devoid of spectral signature. Water vapour content (W value) was evaluated in the 1847-1837 cm^{-1} range. The threshold values for the spectral quality test have been described in the pre-processing step of Fig. 1. Spectra that did not meet the quality test were discarded.

Data pre-processing and processing

All the spectra that passed the quality test were truncated so as to keep the spectral range between 4000-800 cm^{-1}. Spectra were baseline corrected with a second order polynomial function and normalized using Extended Multiplicative Scatter Correction (EMSC). Then, second derivative spectra were calculated using the Savitsky-Golay method [36] and a window length of 9 points. The mean of all second derivative spectra was computed and subtracted from each individual second derivative spectrum. In EMSC, a model is constructed in which the baseline correction and the SNV normalization are done simultaneously and the modelling error minimized. The reader can refer to the supplementary electronic information (Fig. S2) for the raw and pre-processed spectra of the three patient groups.

Spectral data are highly dimensional and it is often difficult to extract the pertinent information that can allow discriminating between groups. PCA is an unsupervised chemometrics method that is commonly employed to reduce spectral data dimensionality. Briefly, PCA replaces original and correlated variables by synthetic and uncorrelated variables called principal components (PCs), estimated by maximizing the projected data variance. These PCs contain the total of the information; they are orthogonal between them and are linear combinations of the original variables. The results are presented using the scores of the most explained PCs. PC loadings can also be useful to understand chemical variations. In this study, we used PCA score plots for a preliminary and exploratory analysis

of data in the spectral range 1200-900 cm^{-1} corresponding to the sugar absorption region (see insert of Fig. 2).

Then, a classification model based on the support vector machine-leave-one-out cross-validation (SVM-LOOCV) method has been implemented to compare the spectra of the three populations. SVM is a supervised classification method and here we used as input of the SVM the PC scores, ordered from the most to the less discriminant. Patient groups were compared pair-wise and the Mann-Whitney statistical test was applied to the scores of principal components to rank them according to their "p" values. In this procedure, (n-1) patients were used as the training set to build the model. The left-out-patient was then used for the model validation. This process is repeated "n" times until all patients were removed once. Sensitivity and specificity were computed for each model and the end result is the average percentage for the specificity and sensitivity. All computing was performed using in-house developed routines in the MatLab software (The MathWorks, Natick, MA., USA) on mean, median, and all spectra of each patient.

Results and discussion

The objective of this study was to evaluate the potential of high-throughput FTIR spectroscopy as a screening tool for classical galactosemia. The work was conducted on three patient populations, healthy, diabetic, and galactosemic, the last two suffering of an accumulation of carbohydrates (glucose and galactose isomers). To do so, we recorded by FTIR spectroscopy patient plasma samples collected in tubes containing heparin as anticoagulant. Spectral data obtained from the three populations were analysed in the first instance by an exploratory method then by using a trained classifier with sensitivity and specificity as end results.

Patient characteristics

Characteristics of healthy, diabetic, and galactosemic patients are summarized in Table S1, Table S2, and Table 1 with corresponding sex ratio of 18/29, 10/9, and 18/12 respectively. Diabetic patients were older (median age: 17 years), galactosemic patients were younger (median age: 6.5 years) and healthy ones were in-between (median age: 12 years). For healthy patients, glucose concentration was 5.1 ± 0.6 mmol/L (reference values: 3.3-6.1 mmol/L) *vs* 14.1 ± 6.9 mmol/L for diabetic patients correlated with HbA1c levels of $9.6 \pm 1.8\%$ (reference values: 4 to 6%). For galactosemic patients, red blood cells galactose-1-phosphate concentrations were very variable (1.0 to 27.0 µmol/L; no galactose-1-phosphate in healthy and diabetic patients) with a mean of 12.4 ± 8.7 µmol/L (median: 12.5 µmol/L). Genetic characteristics are classical with essentially mutations S135L and Q188R.

FTIR spectral analysis

After the quality test, twenty-one spectra from the whole dataset (960 spectra) were excluded namely 1.3%, 4.3% and 1% spectra of healthy, galactosemic and diabetic patients respectively. The mean infrared absorbance spectra obtained from plasma samples of healthy (green), diabetic (blue) and galactosemic (red) patients are shown in Fig. 2. The assignment of the main macromolecules (proteins, lipids, fatty acids, amino acids, nucleic acids, and carbohydrates) present in plasma samples are given in Table 2. The principal differences between the three mean spectra appear in the 1200-900 cm^{-1} range (enlarged in the insert of Fig. 2) corresponding to the carbohydrates absorption region. This spectral region therefore appears as an interesting discriminant region for characterising the three patient conditions.

In order to compare spectra from the three sets of patients, we applied to this spectral region a PCA, an unsupervised chemometric method, to find out if there is any global tendency in the data (Fig. 3). For clarity, only the median spectra are represented for each patient. The

comparison was performed in a pair-wise manner, i.e., healthy *vs* diabetic, healthy *vs* galactosemic, and galactosemic *vs* diabetic patients. The results show that by comparing PCs 1 and 2, there is a partial separation between healthy and diabetic patients (Fig. 3a). Similar observations can be made for healthy and galactosemic patients using PCs 1 and 2 (Fig. 3b). On the other hand, a very clear delineation can be observed between diabetic and galactosemic patients using PCs 1 and 2 (Fig. 3c). The scores of the most discriminant PCs were then used as input of a classifier based on SVM-LOOCV. Tables 3 and 4 summarise the results of SVM-LOOCV method applied to FTIR spectra of healthy, diabetic, and galactosemic patients. For each analysis, median, mean, and all individual spectra were tested. As shown in Table 3, this classification procedure allowed discriminating healthy from diabetic and from galactosemic patients with sensitivity between 80 and 95% and specificity between 87 and 94% respectively. The overall diagnostic total accuracy rate was between 87 and 94%. The best classification results were obtained between galactosemic and diabetic patients (Table 4) with sensitivity between 93 and 95% and specificity between 97 and 100%. The total accuracy rate was 96%. These results obtained for median, mean, and all individual spectra of the three patient groups appear very promising.

Screening newborns for galactosemia is done primarily to detect clinically devastating galactosemia due to defective function of GALT. Increases in blood galactose are also observed in other conditions, however, in the relatively rare galactokinase (GALK), partial GALT, and UDP-galactose-4-epimerase (GALE) deficiency, we have serious sequelae or no clinical consequences [2]. Additionally, there are other transient galactosemias of unknown causes and other known benign variants that are routinely flagged in newborn screening [37]. Most newborn screening programs for galactosemia monitor blood spot galactose concentration with a fluorescence assay as a first-line screen and follow up with a fluorometric blood spot enzyme assay for GALT. These two tests present a potential for false-

positive results, particularly when it is used to attempt to differentiate the variant forms of galactosemia.The direct fluorometric assays such as Beutler's or Benedict's tests also provide a high rate of false-positive [12, 13]. Other methods for the measurement of these metabolites have been developed with state-of-the-art technology, primarily radiometric, spectrophotometric, and fluorometric assays, and more recently a stable isotope-dilution selected ion-monitoring mass spectrometry [7, 9] and ultra-performance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS) [14]. However, in spite of these new technologies, the fractions of false-positive cases may be as high as 89% as reported recently in some screening programs [38].

Our pilot study based on an FTIR spectroscopic analysis of plasma showed the interest of this innovative approach based on the exploitation of a "marker region" reflecting the carbohydrates composition of the samples. It is important to note that the FTIR spectroscopic method presented here is not measuring the change of a single molecule or biomarker, but encompasses the overall metabolic changes caused by a disease, which is then mirrored by a specific IR spectrum or "biochemical fingerprint" of the plasma sample. Thus, FTIR spectroscopy presents the advantage of analysing in a single measurement and in a holistic manner the structural and molecular composition of the sample.

In this study, PCA was used as an unsupervised method to explore the structure of FTIR spectral data obtained from plasmas of healthy, diabetic, and galactosemic patients.

The results show that only a partial discrimination between healthy/diabetic and healthy/galactosemic patients could be reached. The diabetic and galactosemic patients with high concentrations of glucose (Fig. 3a) and galactose (Fig. 3b) respectively are well distinguished from healthy patients. Concerning the comparison between diabetic/galactosemic patients, the separation between the two populations is more visible (Fig. 3c). Further, the two populations are subdivided in two groups as a function of the

concentration of glucose and galactose. The subgroups corresponding to high concentrations of glucose (>20 mM) and galactose (>11 µM) are encircled in blue and red respectively. For the other concentrations, there is also a good separation between diabetic and galactosemic patients with a slight overlapping but this majorly concerns patients with low glucose or galactose concentrations. These results also indicate that the separation is mainly due to the sugar concentration levels and not at all related to the enzyme mutations (Table 1).

Although PCA seems to show a good tendency in the separation of the three populations, it is only an exploratory method. The development of FTIR spectroscopy as a screening method requires the implementation of a classifier.

Here, we have developed a SVM-LOOCV model and tested it on the PCA scores of median, mean, and all spectra of the three patient groups. This classification model was able to differentiate healthy/ diabetic, healthy/galactosemic, and diabetic/galactosemic patients with sensitivity and specificity rates ranging from 80-94%. The total accuracy rate ranges from 87-96%. For galactosemic patients, sensitivity and specificity were better with median spectra, respectively 93.3 and 93.6%. Median spectra have the advantage of being less influenced by outliers and they avoid (as for mean spectra) the redundant use of replicates from the same sample during the classification process.

This feasibility study demonstrates the potential of using FTIR spectroscopy of plasma to identify patients with galactosemia or diabetes. This method offers several advantages: the plasma samples can be used without pre-analytical manipulations; it is reagent-free, label-free, cost-effective, and rapid since it is possible to process 60 samples/hour. Further, blood sampling is easily available at a low cost, and the technique is adapted to newborns since a small volume of pure plasma (1.66 µL/spectrum) is sufficient.

Conclusion

High-throughput FTIR spectroscopy combined with SVM-LOOCV classification procedure appears to be a promising tool in the screening of galactosemia patients. Compared to procedures currently used, our results showed a good performance, in terms of sensitivity and specificity. This technique can be easily adapted to newborn screening. However, this is a proof-of-concept study which needs to be confirmed on a larger population and in parallel with further studies involving the Guthrie test (DBS: dried blood spot) currently used in neonatal screening.

Aknowledgements

Table S1 Characteristics of healthy patients

Patient number	Age	Sex	[Glucose] (mM)
1	11 y	M	4.3
2	5 y	F	5.3
3	3 y	M	5.2
4	7 y	F	5.6
5	2 m	F	6.1
6	48 y	F	4.7
7	6 y	M	4.4
8	30 y	F	5.4
9	8 y	F	5.6
10	4 y	F	5.3
11	22 y	M	4.8
12	17 y	M	6.0
13	0.5 m	M	5.4
13	25 y	F	5.3
15	29 y	M	5.3
16	45 y	F	4.9
17	49 y	M	5.0
18	50 y	M	5.5
19	57 y	M	5.4
20	15 y	F	4.9
21	19 y	F	4.6
22	20 y	F	4.6
23	21 y	F	4.6
24	26 y	F	4.8
25	10 y	F	5.3
26	11 y	M	4.7
27	5 y	F	5.4
28	6 y	F	4.3
29	3 y	M	4.3
30	2 y	F	3.7
31	3 m	F	5.8
32	4 m	M	5.5
33	5 m	M	6.0
34	7 m	M	5.0
35	10 m	F	6.1
36	1 y	F	5.5
37	16 y	M	4.8
38	24 y	F	5.7
39	52 y	F	5.9
40	18 y	F	4.6
41	31 y	F	5.4
42	13 y	F	4.8
43	12 y	F	4.5
44	12 y	M	4.1
45	9 y	M	4.1
46	13 y	F	4.8
47	13 y	F	4.7

Abbreviations: m, months; y, years.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table S2 Characteristics of diabetic patients

Patient number	Age (years)	Sex	[Glucose] (mM)	HbA1c (%)
1	16	M	8.2	7.8
2	60	F	14.0	8.2
3	89	F	12.6	9.1
4	13	M	7.9	12.7
5	45	M	35.1	9.9
6	48	M	8.8	8.3
7	32	F	12.7	9.0
8	37	M	19.3	11.0
9	26	M	14.7	12.6
10	15	M	9.3	7.3
11	4	F	10.3	8.5
12	30	F	9.4	7.9
13	5	F	27.3	11.1
14	12	F	9.8	13.2
15	6	M	13.2	9.1
16	5	M	14.4	8.2
17	17	F	9.5	7.9
18	27	F	11.4	10.8
19	2	M	20.4	10.6

Table 1 Characteristics of galactosemic patients

Patient number	Age	Sex	[Gal-1-P] (μM)	Mutation
1	2 m	F	3.5	S135L / F171S
2	3 m	M	25	S135L / K229N
3	4 m	M	27	S135L / K229N
4	4 y	F	12	L195P / L195P
5	12 y	F	1.1	K285N / IVS7+66t>a
6	45 y	M	1.1	S135L / S135L
7	7 m	M	15	Q188R / H319Q
8	13 y	M	4.1	Q188R / H132Q
9	10 m	M	16	Q188R / Q188R
10	15 y	M	18	Q188R / L226P
11	9 y	M	17	Q188R / E225G/N314D
12	4 y	F	20	Q188R / Q188R
13	10 y	M	18	Q188R / S143L
14	17 y	M	11	Q188R / A191D/N314D
15	10 y	M	1.0	S135L / R272H
16	6 y	M	1.6	S135L / R148W
17	3 y	F	1.0	S135L / S135L
18	5 y	M	11	Q188R / V168M
19	2 y	M	22	Q188R / Q188R
20	2 y	M	26	Q188R / Q188R
21	7 m	M	23	S135L / K229N
22	9 y	M	5.3	Q188R / G195D
23	5 y	F	15	L195P / L195P
24	7 y	F	8.1	Q188R / R333W
25	7 y	F	13	Q188R / R333W
26	8 y	M	13	K285N / G338G
27	25 y	F	2.7	Q188R / R328C
28	2 y	F	1.1	V128I / V128I
29	2 y	F	3.5	V128I / V128I
30	8 y	F	11	Q188R / Q188R

Abbreviations: m, months; y, years.

Table 2 Major assignment of FTIR absorption bands of plasma

Bands (cm^{-1})	Tentative assignment for plasma content
3300	ν(N-H) of proteins (amide A band)
3055-3090	ν(=CH) of lipids and proteins
2950-2960	ν_{as}(CH$_3$) of lipids and proteins
2920-2930	ν_{as}(CH$_2$) of lipids and proteins
2865-2880	ν_s(CH$_3$) of lipids and proteins
2840-2860	ν_s(CH$_2$) of lipids and proteins
1730-1760	ν(C=O) of fatty acids
1660	ν(C=O) of proteins (amide I band)
1550	δ(N-H) of proteins (amide II band)
1400	ν(COO$^-$) of amino acids
1240	ν_{as}(P=O) of nucleic acids
1170-1120	ν(C-O) and ν(C-O-C) of carbohydrates

Abbreviations: ν, stretching vibrations; δ, bending vibrations; s, symmetric; as, asymmetric.

Table 3 SVM-LOOCV classification results of healthy *vs* diabetic patients and of healthy *vs* galactosemic patients

	Healthy *vs* Diabetic			Healthy *vs* Galactosemic		
	Median	**Mean**	**All spectra**	**Median**	**Mean**	**All spectra**
Sick patients	19	19	188	30	30	287
Well classified	16	18	158	28	26	230
Wrongly classified	3	1	30	2	4	57
Healthy patients	47	47	464	47	47	464
Well classified	43	44	434	44	41	415
Wrongly classified	4	3	30	3	6	49
Sensitivity (%)	84.2	94.7	84.0	93.3	86.7	80.1
Specificity (%)	91.5	93.6	93.5	93.6	87.2	89.4
Total accuracy rate (%)	89.4	93.9	90.8	93.5	87.0	92.4

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 4 SVM-LOOCV classification results of diabetic *vs* galactosemic patients

| | Galactosemic *vs* Diabetic | | |
	Median	Mean	All spectra
Galactosemic patients	30	30	287
Well classified	28	28	272
Wrongly classified	2	2	15
Diabetic patients	19	19	188
Well classified	19	19	182
Wrongly classified	0	0	6
Sensitivity (%)	93.3	93.3	94.8
Specificity (%)	100	100	96.8
Total accuracy (%)	95.9	95.9	95.6

Figure captions

Fig. 1 Workflow of the experimental protocol: from sample preparation to diagnostic performance.

Fig. 2 Spectral comparisons of healthy, diabetic, and galactosemic patient plasmas.
Mean of 10 spectra of healthy (green line), diabetic (blue line), and galactosemic (red line) patient. Spectra are baseline-corrected and vector-normalized. Enlarged is the carbohydrates region.

Fig. 3 PCA score plots of median spectra.
PCA was performed on second derivative median spectra using the wavenumber range of 1200-900 cm^{-1}. Score plots are shown for (a) healthy (green circle) *vs* diabetic (blue triangle), (b) healthy (green circle) *vs* galactosemic (red square) patients, and (c) diabetic (blue triangle) *vs* galactosemic (red square) patients. The blue and red circles indicate patients with [glucose]>20 mM (see Table 2) and [galactose]>11 µM (see Table 3) respectively.

Fig. S1 Leloir pathway of congenital galactosemia.

Fig. S2 Raw spectra (10 replicates) and his mean spectrum with the standard deviation of the same patient. (a) raw spectra and (b) mean spectrum and standard deviation of healthy patient, (c) raw spectra and (d) mean spectrum and standard deviation of diabetic patient, (e) raw spectra and (f) mean spectrum and standard deviation of galactosemic patient.
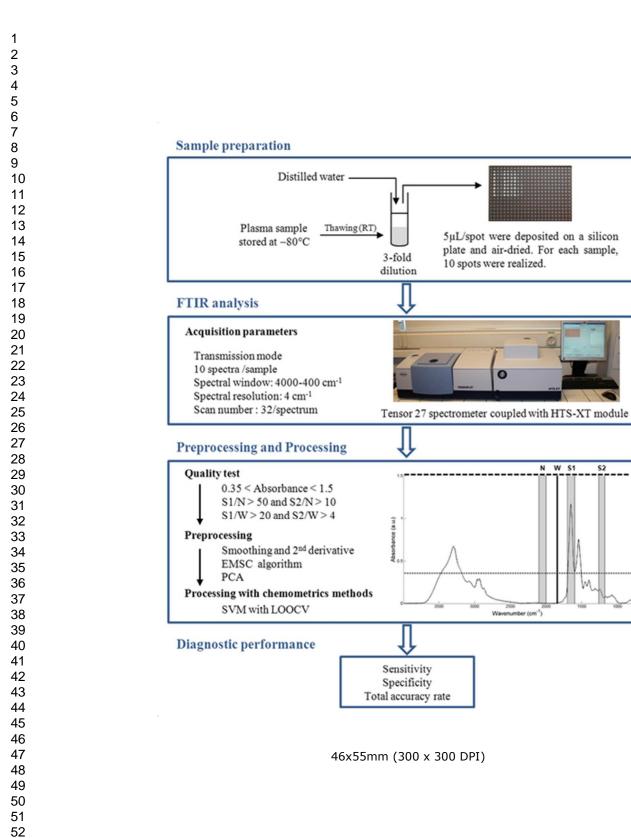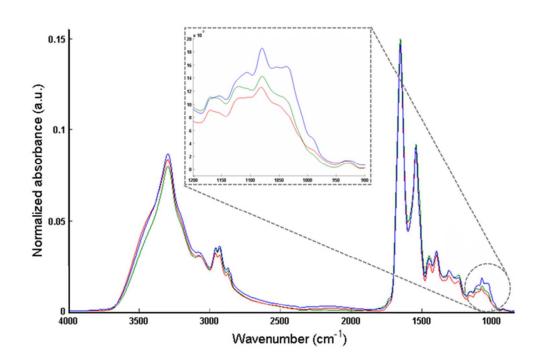
References:

1. H. M. Holden, I. Rayment and J. B. Thoden, *J Biol Chem.*, 2003, **278**, 43885-43888.
2. J. M. Saudubray, G. Berghe and J. H. Walter, *Inborn metabolic diseases*, Springer, New-York, 2012.
3. A. Bosch, H. Bakker, A. Van Gennip, J. Van Kempen, R. Wanders and F. Wijburg, *J Inherit Metab Dis*, 2003, **25**, 629-634.
4. K. K. Openo, J. M. Schulz, C. A. Vargas, C. S. Orton, M. P. Epstein, R. E. Schnur, F. Scaglia, G. T. Berry, G. S. Gottesman, C. Ficicioglu, A. E. Slonin, R. J. Schroer, C. Yu, V. E. Rangel, J. Keenan, K. Lamance and J. L. Fridovich-Keil, *Am J Hum Genet*, 2006, **78**, 89-102.
5. L. T. Kirby, M. G. Norman, D. A. Applegarth and D. F. Hardwick, *Can Med Assoc J*, 1985, **132**, 1033-1035.
6. R. Wilson, P. Riordan and P. Hartmann, *J Inherit Metab Dis*, 1990, **13**, 270-272.
7. J. Chen, C. Yager, R. Reynolds, M. Palmieri and S. Segal, *Clin Chem*, 2002, **48**, 604-612.
8. P. Schadewaldt, L. Kamalanathan, H. W. Hammen and U. Wendel, *Rapid Comm Mass Spectrom*, 2003, **17**, 2833-2838.
9. C. Yager, S. Wehrli and S. Segal, *Clin Chim Acta*, 2006, **366**, 216-224.
10. J. S. Jeong, H. J. Kwon, H. R. Yoon, Y. M. Lee, T. Y. Choi and S. P. Hong, *Anal Biochem*, 2008, **376**, 200-205.
11. J. S. Jeong, H. R. Yoon and S. P. Hong, *J Chromatogr A*, 2007, **1140**, 157-162.
12. E. Beutler and M. C. Baluda, *J Lab Clin Med*, 1966, **68**, 137-141.
13. D. Morell-Garcia, J. M. Bauça, A. Barceló, G. Perez-Esteban and M. Vila, *Clin Biochem*, 2014, **47**, 857-859.
14. Y. Li, A. S. Ptolemy, L. Harmonay, M. Kellogg and G. T. Berry, *Clin Chem*, 2010, **56**, 772-780.
15. M. Lindhout, M. E. Rubio-Gozalbo, J. A. Bakker and J. Bierau, *Clin Chim Acta*, 2010, **411**, 980-983.
16. D. H. Ko, S. H. Jun, H. D. Park, S. H. Song, K. U. Park, J. Q. Kim, Y. H. Song and J. Song, *Clin Chem*, 2010, **56**, 764-771.
17. D. H. Ko, S. H. Jun, K. U. Park, S. H. Song, J. Q. Kim and J. Song, *J Inherit Metab Dis*, 2011, **34**, 409-414.
18. Y. Li, A. S. Ptolemy, L. Harmonay, M. Kellogg and G. T. Berry, *Mol Genet Metab*, 2011, **102**, 33-40.
19. S. F. Dobrowolski, R. A. Banas, J. G. Suzow, M. Berkley and E. W. Naylor, *J Mol Diagn*, 2003, **5**, 42-47.
20. N. Mainreck, S. Brézillon, G. D. Sockalingum, F. X. Maquart, M. Manfait and Y. Wegrowski, *J Pharm Sci.*, 2011, **100**, 441-450.
21. N. Mainreck, S. Brézillon, G. D. Sockalingum, F. X. Maquart, M. Manfait and Y. Wegrowski, in *Proteoglycans*, Springer, New-York, 2012, pp. 117-130.
22. D. C. Fernandez, R. Bhargava, S. M. Hewitt and I. W. Levin, *Nat Biotechnol*, 2005, **23**, 469-474.
23. J. Nallala, C. Gobinet, M.-D. Diebold, V. Untereiner, O. Bouché, M. Manfait, G. D. Sockalingum and O. Piot, *J Biomed Opt*, 2012, **17**, 116013.
24. J. Pijanka, G. D. Sockalingum, A. Kohler, Y. Yang, F. Draux, G. Parkes, K.-P. Lam, D. Collins, P. Dumas, C. Sandt, D. Van Pittius, G. Douce, M. Manfait, V. Untereiner and J. Sulé-Suso, *Lab Invest*, 2010, **90**, 797-807.

25. J. K. Pijanka, D. Kumar, T. Dale, I. Yousef, G. Parkes, V. Untereiner, Y. Yang, P. Dumas, D. Collins, M. Manfait, G. D. Sockalingum, N. R. Forsyth and J. Sulé-Suso, *Analyst*, 2010, **135**, 3126-3132.

26. A. L. Mitchell, K. B. Gajjar, G. Theophilou, F. L. Martin and P. L. Martin-Hirsch, *J Biophotonics*, 2014, **7**, 153-165.

27. P. Carmona, M. Molina, M. Calero, F. Bermejo-Pareja, P. Martínez-Martín and A. Toledano, *J Alzheimers Dis*, 2013, **34**, 911-920.

28. A. Lux, R. Müller, M. Tulk, C. Olivieri, R. Zarrabeita, T. Salonikios and B. Wirnitzer, *Orphanet J Rare Dis*, 2013, **8**, 94-108.

29. E. Scaglia, G. D. Sockalingum, J. Schmitt, C. Gobinet, N. Schneider, M. Manfait and G. Thiéfin, *Anal Bioanal Chem*, 2011, **401**, 2919-2925.

30. X. Zhang, G. Thiéfin, C. Gobinet, V. Untereiner, I. Taleb, B. Bernard-Chabert, A. Heurgué, C. Truntzer, P. Ducoroy, P. Hillon and G. D. Sockalingum, *Transl Res*, 2013, **162**, 279-286.

31. C. Petibois and G. Déléris, *Arch Med Res*, 2004, **35**, 532-539.

32. C. Petibois, K. Gionnet, M. Gonçalves, A. Perromat, M. Moenner and G. Déléris, *Analyst*, 2006, **131**, 640-647.

33. M. J. Baker, J. Trevisan, P. Bassan, R. Bhargava, H. J. Butler, K. M. Dorling, P. R. Fielden, S. W. Fogarty, N. J. Fullwood, K. A. Heys, C. Hughes, P. Lasch, P. L. Martin-Hirsch, M. J. Obinaju, G. D. Sockalingum, J. Sulé-Suso, R. J. Strong, M. J. Walsh, B. R. Wood, P. Gardner and F. L. Martin, *Nat protoc*, 2014, **9**, 1771-1791.

34. A. Boutron, A. Marabotti, A. Facchiano, D. Cheillan, M. Zater, C. Oliveira, C. Costa, P. Labrune and M. Brivet, *Mol Genet Metab*, 2012, **107**, 438-447.

35. D. Helm, H. Labischinski and D. Naumann, *J Microbiol Methods*, 1991, **14**, 127-142.

36. A. Savitzky and M. J. Golay, *Anal Chem*, 1964, **36**, 1627-1639.

37. M. A. Michel, E. Raucourt, N. Bednarek and R. Garnotel, *Ann Biol Clin*, 2014, **72**, 193-196.

38. M. J. Bennett, *Clin Chem*, 2010, **56**, 690-692.

46x55mm (300 x 300 DPI)

55x36mm (300 x 300 DPI)

90x96mm (300 x 300 DPI)