BMC Medicine

CrossMark

Urine metabolome profiling of immune-mediated inflammatory diseases

Arnald Alonso[1], Antonio Julià[1*] , Maria Vinaixa[2,3], Eugeni Domènech[4,5], Antonio Fernández-Nebro[6],
Juan D. Cañete[7], Carlos Ferrándiz[4], Jesús Tornero[8], Javier P. Gisbert[5,9], Pilar Nos[5,10], Ana Gutiérrez Casbas[5,11],
Lluís Puig[12], Isidoro González-Álvaro[9], José A. Pinto-Tasende[13], Ricardo Blanco[14], Miguel A. Rodríguez[2,3],
Antoni Beltran[2,3], Xavier Correig[2,3], Sara Marsal[1*] and for the IMID Consortium

Abstract

Background: Immune-mediated inflammatory diseases (IMIDs) are a group of complex and prevalent diseases where disease diagnostic and activity monitoring is highly challenging. The determination of the metabolite profiles of biological samples is becoming a powerful approach to identify new biomarkers of clinical utility. In order to identify new metabolite biomarkers of diagnosis and disease activity, we have performed the first large-scale profiling of the urine metabolome of the six most prevalent IMIDs: rheumatoid arthritis, psoriatic arthritis, psoriasis, systemic lupus erythematosus, Crohn's disease, and ulcerative colitis.

Methods: Using nuclear magnetic resonance, we analyzed the urine metabolome in a discovery cohort of 1210 patients and 100 controls. Within each IMID, two patient subgroups were recruited representing extreme disease activity (very high vs. very low). Metabolite association analysis with disease diagnosis and disease activity was performed using multivariate linear regression in order to control for the effects of clinical, epidemiological, or technical variability. After multiple test correction, the most significant metabolite biomarkers were validated in an independent cohort of 1200 patients and 200 controls.

Results: In the discovery cohort, we identified 28 significant associations between urine metabolite levels and disease diagnosis and three significant metabolite associations with disease activity ($P_{FDR} < 0.05$). Using the validation cohort, we validated 26 of the diagnostic associations and all three metabolite associations with disease activity ($P_{FDR} < 0.05$). Combining all diagnostic biomarkers using multivariate classifiers we obtained a good disease prediction accuracy in all IMIDs and particularly high in inflammatory bowel diseases. Several of the associated metabolites were found to be commonly altered in multiple IMIDs, some of which can be considered as hub biomarkers. The analysis of the metabolic reactions connecting the IMID-associated metabolites showed an over-representation of citric acid cycle, phenylalanine, and glycine-serine metabolism pathways.

Conclusions: This study shows that urine is a source of biomarkers of clinical utility in IMIDs. We have found that IMIDs show similar metabolic changes, particularly between clinically similar diseases and we have found, for the first time, the presence of hub metabolites. These findings represent an important step in the development of more efficient and less invasive diagnostic and disease monitoring methods in IMIDs.

Keywords: Metabolomics, Urine biomarkers, Disease activity, Autoimmune diseases, Inflammatory diseases

* Correspondence: toni.julia@vhir.org; sara.marsal@vhir.org
[1]Rheumatology Research Group, Vall d'Hebron Hospital Research Institute, Barcelona, Spain
Full list of author information is available at the end of the article

BioMed Central

Alonso *et al. BMC Medicine* (2016) 14:133

Page 2 of 12

Background

Rheumatoid arthritis (RA), psoriasis (Ps), psoriatic arthritis (PsA), systemic lupus erythematosus (SLE), Crohn's disease (CD), and ulcerative colitis (UC) are prevalent immune-mediated inflammatory diseases (IMIDs) [1–4]. This group of diseases is characterized by the aberrant and chronic activation of the immune system, affecting one or more tissues. IMIDs have a high socioeconomic impact [1, 4, 5] and are among the main causes of morbidity, disability, and mortality in developed countries [6–8]. Although each IMID targets different tissues and organs, they all share common molecular mechanisms like the activation of the Tumor Necrosis Factor cytokine pathway [9]. Recently, genome-wide association studies have demonstrated that IMIDs also share many genetic risk loci [10]. Consequently, the combined analysis of multiple IMIDs has the ability to leverage the identification of more relevant molecular features.

Improvements in the diagnosis of IMIDs would be of great benefit to the patient and would significantly reduce the socioeconomic burden of these diseases. There is increasing evidence that the administration of therapies, particularly biological treatments, at earlier stages of the disease results in a more effective control of the inflammatory process [11, 12]. In RA, for example, early diagnosis and treatment have been shown to increase the probability of entering disease remission [13–16], an accomplishment that was unthinkable only a decade ago. Similarly, the diagnosis of inflammatory bowel diseases CD and UC is often established too late, when severe complications have already occurred [17]. The identification of more accurate diagnostic biomarkers would therefore have a high impact on the improvement of disease outcomes in IMIDs.

Measuring disease activity is also a challenging problem in IMIDs. The lack of objective and highly informative markers of disease activity has a negative impact in key aspects of patient management, like the decision to initiate or terminate a specific therapy. Currently, different scores are available to measure disease activity in each IMID. These scores are based on clinical, laboratory, and/or imaging measures, and although they are frequently used in clinical practice, they have important limitations [18]. Disease activity scores are often based on unspecific and sometimes subjective variables that significantly increase their inter- and intra-observer variability, clearly reducing their accuracy and, consequently, affecting disease monitoring [19]. The dynamic nature and highly informative properties of biological molecules (i.e., biomarkers) could provide the level of objectivity and accuracy necessary for a better management of disease activity in IMIDs.

High-throughput analysis technologies are able to generate comprehensive profiles of different molecular species from multiple biological samples. Recent developments in these technologies could provide the level of precision that is required to improve disease management [20–22]. However, one limitation in the use of these approaches to study IMIDs is that the target tissue or organ cannot be easily sampled, resulting in a highly invasive procedure. Instead, the use of more accessible surrogate tissues or biofluids like blood, saliva and urine could help to circumvent this limitation. Urine, in particular, is a highly interesting sample source since its collection is very simple and is clearly non-invasive for the patient. The direct relationship with blood composition strongly supports the hypothesis that different molecular species that are present in both biological fluids like metabolites, nucleic acids, or proteins and whose variation is associated with pathological features could be highly informative biomarkers in IMIDs [23, 24].

The profiling of the metabolite composition of biological samples, metabolomics, is one of the most rapidly evolving high-throughput analysis approaches [25]. Metabolites could potentially serve as biomarkers in many diseases since they represent the biochemical end products of the genetic pathways, providing an accurate representation of the physiological state of an individual [26]. Nuclear magnetic resonance (NMR), together with mass spectrometry, is one of the most widely used metabolomic technologies [27]. NMR has been used in the determination of the metabolite profiles of tissue and biofluid samples of multiple diseases [28, 29]. To date, however, very few studies have analyzed the metabolomic profiles of IMIDs and most lack independent validation cohorts. Further, there is a lack of studies comparing the metabolomes of this group of inflammatory diseases in parallel.

In the present work, we have performed a large-scale high-throughput analysis of the urine metabolome of six of the most prevalent IMIDs (RA, PsA, Ps, SLE, CD, and UC) and a cohort of healthy control individuals in order to identify new biomarkers associated with disease diagnosis and disease activity. For this objective, we have used a two-stage study design consisting of a discovery stage where the urine metabolomes of 1210 IMID patients and 100 healthy controls were analyzed, and a validation stage where the most significant candidate metabolite biomarkers from the discovery stage were confirmed using an independent cohort of 1200 IMID patients and 200 healthy controls. To our knowledge, this study provides the first comprehensive characterization of urine metabolites associated with IMIDs.

Methods

Study design

A two-stage approach was used to characterize the urine metabolite profile associated with IMIDs. In the first

Alonso *et al. BMC Medicine* (2016) 14:133

Page 3 of 12

stage (discovery stage), candidate biomarkers for diagnosis and disease activity monitoring were identified using a cohort of 1310 individuals ($n = 1210$ IMID patients and $n = 100$ healthy controls). In the second stage (validation stage), the most significant candidate biomarkers where validated using a cohort of 1400 individuals ($n = 1200$ IMID patients and $n = 200$ healthy controls). In order to identify urine metabolites associated with disease activity, two similarly sized subgroups of patients showing extreme disease activity (i.e., very high and very low disease activity) were selected within each IMID disease (Table 1, Additional file 1: Figure S1). Previous metabolomic studies have shown that several epidemiological and technical variables can act as confounders and, therefore, particular care must be taken to avoid or minimize their effects. In the present study, two different measures were taken to reduce the impact of potential confounders. First, the patients and controls from the discovery and validation stages were selected so that they had similar distributions of epidemiological (gender, age and body mass index) and sample collection variables (fasting time of the individual before sample collection and the time of the day of sample collection). Second, in order to adjust for any additional confounding effect, all potential confounder variables were also included as covariates in the multivariate linear regression models testing for association with disease and with disease activity.

Ethics

The study was conducted according to the Declaration of Helsinki. Patients and controls included in the analysis were recruited by the Immune-Mediated Disease Consortium [29–32]. Informed consent was obtained from all participants, and protocols were reviewed and approved by local institutional review boards. All the patients included in the study met the corresponding consensus diagnostic criteria of each IMID (Additional file 1: Supplementary Methods).

Metabolomic analysis

Urine samples were collected, processed, and analyzed using ^{1}H-NMR as described in the Supplementary Methods (Additional file 1). Spectral processing of the urine NMR profiles was performed using FOCUS software [33], and reference metabolite databases [34] were used to identify the molecules corresponding to each spectral resonance. In order to confirm the identity of specific metabolites, two-dimensional ^{1}H-^{13}CHSQC (heteronuclear single quantum correlation) and ^{1}H-^{1}H COSY (correlation spectroscopy) was used in a selected group of samples.

Statistical analysis

Multivariate linear regression was carried out to test the association between metabolite levels and disease diagnosis as well as disease activity [35–37]. In each linear regression analysis, different epidemiological (i.e., sex, age, smoking habit, body mass index, lifestyle, and dietary habits) and technical variables (i.e., time at sample collection and fasting time) were included as covariates in order to control for confounding. To avoid the presence of false positives associated to drug treatment, we also tested the association between all metabolite levels and drug treatment at the time of sample collection. The drug treatments tested for association included antibody to tumor necrosis factor (anti-TNFα) therapy (i.e., infliximab and etanercept), disease-modifying drugs (i.e., methotrexate and leflunomide), corticoids, and non-steroidal anti-inflammatory drugs (i.e., ibuprofen). After removing known drug-specific metabolites (i.e., ibuprofen, acetaminophen, and 5-aminosalicylic acid) we found no significant association between urine metabolite levels and the presence of any particular therapy.

In the discovery phase, three types of analyses were performed: (1) diagnostic, comparing the metabolite levels between each IMID disease against the healthy control cohort, (2) differential, comparing the metabolite levels between IMIDs that have more similar clinical features, and (3) activity-related, comparing the metabolite

Table 1 Distribution of sample size and disease activity scores in the low and high activity groups of each immune-mediated inflammatory disease (IMID) after quality control

IMID	Disease activity score	Discovery cohort		Validation cohort	
		Low activity[a]	High activity[a]	Low activity[a]	High activity[a]
CD	Harvey-Bradshaw Index [71]	154 (0.0)	45 (7.0)	100 (0.0)	100 (9.0)
UC	Lichtiger Score [72]	124 (0.0)	81 (6.0)	99 (0.0)	98 (5.0)
RA	Disease Activity Score 28 [73]	114 (1.7)	127 (5.5)	98 (1.6)	95 (5.6)
PsA	Disease Activity Score 28 [73]	97 (1.5)	89 (3.9)	96 (1.5)	96 (4.2)
Ps	Psoriasis Area Severity Index [74]	101 (0.0)	84 (14.1)	100 (0.0)	92 (17.5)
SLE	Selena-Sledai [75] and BILAG [76]	123 (1.0[b])	41 (11.0[b])	90 (1.0[b])	88 (7.5[b])

Description of the sample sizes and median disease activity values in the discovery and validation stages for each IMID (CD, Crohn's disease; UC, ulcerative colitis; RA, rheumatoid arthritis; PsA, psoriatic arthritis; Ps, psoriasis; SLE, systemic lupus erythematosus)
[a]Number of samples (median disease activity values)
[b]Maximum of Selena-Sledai and BILAG indices of each patient

Alonso *et al. BMC Medicine* (2016) 14:133

Page 4 of 12

concentrations between patients with high and low disease activity within each IMID. Multiple test correction of the significance P values was performed using the discovery rate method (false discovery rate (FDR) < 0.05) both in the discovery and validation stages. The hierarchical clustering of urine IMID profiles was performed using the combined association ($-\log_{10}P$ values) for each disease obtained in the case-control analysis.

In order to evaluate the power of the urine metabolome for disease diagnosis, we built a classifier for each IMID using the partial least squares discriminant analysis method in the discovery dataset as described previously [38]. Once the optimal classifier was identified, it was subsequently tested using the independent validation dataset. The performance of the different disease classifiers was determined using the receiver operating characteristic (ROC) curve analysis as described previously [23, 38]. From each ROC, the area under the curve (AUC) statistic was estimated as a measure of the classifier's diagnostic performance.

In order to gain further biological insight of the associated metabolites, we used the MetaboNetworks software [39]. This method uses a set of predefined metabolic reactions in a single or multiple organisms to identify and define the shortest metabolic reaction chains linking a set of input metabolites. Here, we applied this network analysis approach to identify the shortest metabolic reaction chains linking all metabolites significantly associated with one or more IMIDs. For this analysis we used the set of KEGG reactions (Kyoto Encyclopedia of Genes and Genomes [40]) described for humans as well as the pathways associated with the most abundant endosymbionts from the gut microbiota (Firmicutes, Bacteroidetes, Alphaproteobacteria, Betaproteobacteria, Deltaproteobacteria, Gammaproteobacteria, and Actinobacteria phyla [41]).

Results

Sample characteristics and quality control
In the discovery dataset, 1210 IMID patients (203 CD, 213 UC, 250 RA, 167 SLE, 190 PsA, and 187 Ps) and 100 healthy subjects were included in the study. After quality control analysis of the resulting NMR urine spectra, the final discovery dataset consisted of 1180 IMID patient samples and 93 healthy control samples (Additional file 1: Supplementary Methods, Table S1).

The validation dataset used consisted of 1200 IMID patients ($n = 200$ patients per disease) and 200 healthy control subjects. After the quality control analysis of the urine NMR spectra, the final validation dataset consisted of 1152 patient and 196 control samples (Additional file 1: Table S1).

Within each IMID, patients were selected to represent two similarly sized groups of extreme disease activity (i.e., very low and very high disease activity). The average disease activity values for each subgroup are shown in Table 1 and Figure S1 (Additional file 1). The main clinical and epidemiological characteristics of the two cohorts as well as technical variables associated with the sample collection process are presented in Figure S2 (Additional file 1).

Metabolite panel
A total of 143 spectral peaks were identified in the urine NMR spectra from the discovery dataset. After quality control analysis and filtering of redundant peaks (i.e. peaks quantifying thee same metabolite), a final set of $n = 37$ unique metabolites was identified. To improve this metabolite identification stage, two-dimensional ^{1}H-^{13}CHSQC and ^{1}H-^{1}H COSY were performed to validate and resolve unclear metabolite assignments. From these, 37 metabolites identified, of which four metabolites (ibuprofen, acetaminophen, 5-aminosalicylic acid, and ethanol) were found to be either exogenous or drug-related molecules and were excluded from downstream analyses. From the final set of 33 urine metabolites, 25 could be confidently assigned to a known molecule, while the remaining 8 metabolites could not be associated to a known small molecule and therefore were defined using the prefix *Uknown* (Additional file 1: Table S2). According to the Human Metabolome Database [34] all the known metabolites are expected to be found in human urine, and most of them ($n = 23$, > 90 %) have been previously measured in human urine using NMR [42–44].

Assessment of urine diagnostic biomarkers for IMIDs
In the discovery stage, the comparison between the urine metabolite profiles between patients and controls identified a total of 28 significant associations (FDR < 0.05). In the validation stage, $n = 26$ of these metabolite associations (93 %) were significantly replicated (FDR < 0.05, Table 2). In a secondary analysis, we found $n = 13$ metabolite associations to be significant at the nominal level in both stages of study ($P < 0.05$, same direction of change, Table 2). Using MetaboNetworks to analyze the associated metabolite profiles [39] we found a overrepresentation of metabolites from the citric acid cycle, phenylalanine metabolism and glycine-serine metabolism pathways (Fig. 1).

Among the validated metabolites, six were found to be associated to three or more IMIDs (Fig. 2a). Since their patterns were very similar between diseases (i.e., significance of association and direction of change), they were considered as hub metabolites in IMIDs. From these, citrate showed the strongest hub properties, showing a significantly lower concentration in the urine of most IMIDs compared to controls (Fig. 2, $P_{CD} = 6.2 \times 10^{-16}$, $P_{SLE} = 2.3 \times 10^{-10}$, $P_{Ps} = 2.9 \times 10^{-8}$, $P_{RA} = 4.3 \times 10^{-7}$, $P_{PsA} = 3.5 \times 10^{-5}$). In UC, citrate levels were also lower than in

Alonso *et al. BMC Medicine* (2016) 14:133

Page 5 of 12

Table 2 Metabolites associated with each immune-mediated inflammatory disease (IMID) in the discovery and validation cohorts

IMID/Metabolite	log$_2$(IMID/Ctrl)$_{DIS}$[b]	P_{DIS}	log$_2$(IMID/Ctrl)$_{VAL}$[b]	P_{VAL}	P_{COMB}
CD/Citrate	−0.94 (−1.21 to −0.68)	1.4×10^{-8}	−0.80 (−1.01 to −0.59)	1.1×10^{-9}	6.2×10^{-16}
SLE/Citrate	−0.68 (−0.96 to −0.41)	6.1×10^{-5}	−0.85 (−1.11 to −0.59)	1.4×10^{-7}	2.3×10^{-10}
Ps/Citrate	−0.60 (−0.82 to −0.38)	1.0×10^{-5}	−0.45 (−0.64 to −0.26)	1.3×10^{-4}	2.9×10^{-8}
RA/Citrate	−0.49 (−0.72 to −0.26)	5.4×10^{-4}	−0.55 (−0.77 to −0.33)	4.3×10^{-5}	4.3×10^{-7}
PsA/Citrate	−0.39 (−0.63 to −0.16)	6.7×10^{-3}	−0.44 (−0.64 to −0.24)	3.7×10^{-4}	3.5×10^{-5}
UC/Citrate[a]	−0.34 (−0.57 to −0.11)	1.6×10^{-2}	−0.39 (−0.59 to −0.19)	1.6×10^{-3}	3.0×10^{-4}
UC/N-acetyl AAs	−0.57 (−0.85 to −0.30)	7.6×10^{-4}	−0.64 (−0.87 to −0.42)	2.4×10^{-6}	3.8×10^{-8}
RA/N-acetyl AAs[a]	−0.32 (−0.53 to −0.10)	1.7×10^{-2}	−0.61 (−0.86 to −0.37)	4.3×10^{-5}	1.1×10^{-5}
CD/N-acetyl AAs	−0.63 (−0.90 to −0.36)	1.4×10^{-4}	−0.26 (−0.47 to −0.06)	3.7×10^{-2}	6.7×10^{-5}
Ps/N-acetyl AAs	−0.27 (−0.43 to −0.11)	6.9×10^{-3}	−0.35 (−0.54 to −0.15)	3.3×10^{-3}	2.6×10^{-4}
PsA/N-acetyl AAs	−0.43 (−0.67 to −0.19)	3.9×10^{-3}	−0.28 (−0.46 to −0.10)	9.1×10^{-3}	3.9×10^{-4}
Ps/Trigonelline	−0.70 (−0.99 to −0.40)	1.0×10^{-4}	−0.73 (−0.94 to −0.51)	7.3×10^{-8}	2.0×10^{-10}
UC/Trigonelline	−0.56 (−0.88 to −0.25)	3.3×10^{-3}	−0.72 (−0.95 to −0.48)	7.3×10^{-7}	5.0×10^{-8}
CD/Trigonelline	−0.71 (−1.01 to −0.41)	1.3×10^{-4}	−0.54 (−0.75 to −0.32)	3.8×10^{-5}	1.0×10^{-7}
PsA/Trigonelline[a]	−0.46 (−0.74 to −0.18)	7.5×10^{-3}	−0.42 (−0.63 to −0.21)	1.1×10^{-3}	1.0×10^{-4}
SLE/Alanine	−0.29 (−0.44 to −0.14)	1.5×10^{-3}	−0.62 (−0.75 to −0.49)	4.8×10^{-14}	2.7×10^{-15}
Ps/Alanine	−0.31 (−0.45 to −0.16)	4.2×10^{-4}	−0.35 (−0.46 to −0.25)	6.5×10^{-8}	7.0×10^{-10}
PsA/Alanine[a]	−0.17 (−0.31 to −0.04)	4.0×10^{-2}	−0.37 (−0.47 to −0.26)	4.0×10^{-8}	3.4×10^{-8}
RA/Alanine	−0.24 (−0.38 to −0.10)	5.3×10^{-3}	−0.35 (−0.46 to −0.24)	4.4×10^{-7}	4.9×10^{-8}
CD/Alanine[a]	−0.23 (−0.38 to −0.08)	1.2×10^{-2}	−0.21 (−0.32 to −0.11)	6.7×10^{-4}	1.1×10^{-4}
SLE/Methylsuccinate[a]	−0.72 (−1.25 to −0.19)	2.5×10^{-2}	−2.10 (−2.61 to −1.59)	5.0×10^{-11}	3.6×10^{-11}
UC/Methylsuccinate	−1.09 (−1.58 to −0.60)	2.9×10^{-4}	−0.61 (−1.01 to −0.22)	1.1×10^{-2}	4.5×10^{-5}
CD/Methylsuccinate	−1.05 (−1.56 to −0.54)	7.4×10^{-4}	−0.67 (−1.08 to −0.26)	7.6×10^{-3}	7.4×10^{-5}
PsA/Methylsuccinate[a]	−0.74 (−1.25 to −0.23)	1.6×10^{-2}	−0.91 (−1.35 to −0.47)	6.7×10^{-4}	1.4×10^{-4}
Ps/Methylsuccinate	−0.95 (−1.49 to −0.40)	4.7×10^{-3}	−0.63 (−1.07 to −0.20)	1.7×10^{-2}	8.3×10^{-4}
UC/*Unknown 7*	5.49 (4.72 to 6.26)	2.2×10^{-26}	5.30 (4.68 to 5.93)	1.5×10^{-36}	4.7×10^{-60}
CD/*Unknown 7*	2.40 (1.61 to 3.20)	1.0×10^{-6}	2.75 (2.14 to 3.36)	5.1×10^{-13}	2.3×10^{-17}
RA/*Unknown 7*	1.68 (0.99 to 2.37)	7.6×10^{-5}	1.48 (0.94 to 2.02)	8.2×10^{-6}	1.4×10^{-8}
SLE/*Unknown 7*[a]	0.81 (0.18 to 1.44)	3.4×10^{-2}	1.33 (0.80 to 1.86)	4.8×10^{-5}	2.3×10^{-5}
CD/Hippurate	−1.74 (−2.08 to −1.40)	1.5×10^{-15}	−1.54 (−1.80 to −1.28)	4.5×10^{-20}	5.5×10^{-33}
UC/Hippurate	−1.03 (−1.32 to −0.73)	3.3×10^{-8}	−0.92 (−1.16 to −0.69)	3.8×10^{-10}	4.9×10^{-16}
Ps/Hippurate[a]	−0.40 (−0.66 to −0.13)	1.6×10^{-2}	−0.33 (−0.54 to −0.11)	1.4×10^{-2}	2.0×10^{-3}
RA/Carnitine[a]	−0.68 (−1.11 to −0.25)	9.2×10^{-3}	−1.05 (−1.35 to −0.75)	2.1×10^{-8}	4.4×10^{-9}
PsA/Carnitine[a]	−0.55 (−0.98 to −0.11)	3.9×10^{-2}	−0.80 (−1.11 to −0.50)	2.0×10^{-5}	1.2×10^{-5}
CD/3-Hydroxyisovalerate	−1.02 (−1.51 to −0.53)	6.7×10^{-4}	−1.88 (−2.27 to −1.48)	4.2×10^{-14}	1.1×10^{-15}
UC/Phenylacetylglycine	0.42 (0.19 to 0.65)	2.5×10^{-3}	0.48 (0.31 to 0.65)	5.5×10^{-6}	2.7×10^{-7}
CD/Free acetate[a]	−0.46 (−0.80 to −0.11)	3.0×10^{-2}	−0.51 (−0.71 to −0.30)	6.5×10^{-5}	2.8×10^{-5}
RA/Tyrosine	1.25 (0.60 to 1.90)	1.7×10^{-3}	0.57 (0.14 to 1.00)	3.0×10^{-2}	5.7×10^{-4}
CD/N,N-dimethylglycine[a]	−0.27 (−0.46 to −0.07)	2.3×10^{-2}	−0.25 (−0.44 to −0.06)	2.8×10^{-2}	5.5×10^{-3}

Association statistics for the discovery (DIS), the validation (VAL) and the combined (COMB) cohorts.
[a]Nominal association
[b]Logarithmic concentration ratios (median, 95 % CI) of the corresponding IMID cohort versus the control cohort
CD, Crohn's disease; N-acetyl AAs, N-acetyl amino acids; Ps, psoriasis; PsA, psoriatic arthritis; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus; UC, ulcerative colitis

Alonso *et al. BMC Medicine* (2016) 14:133

Page 6 of 12
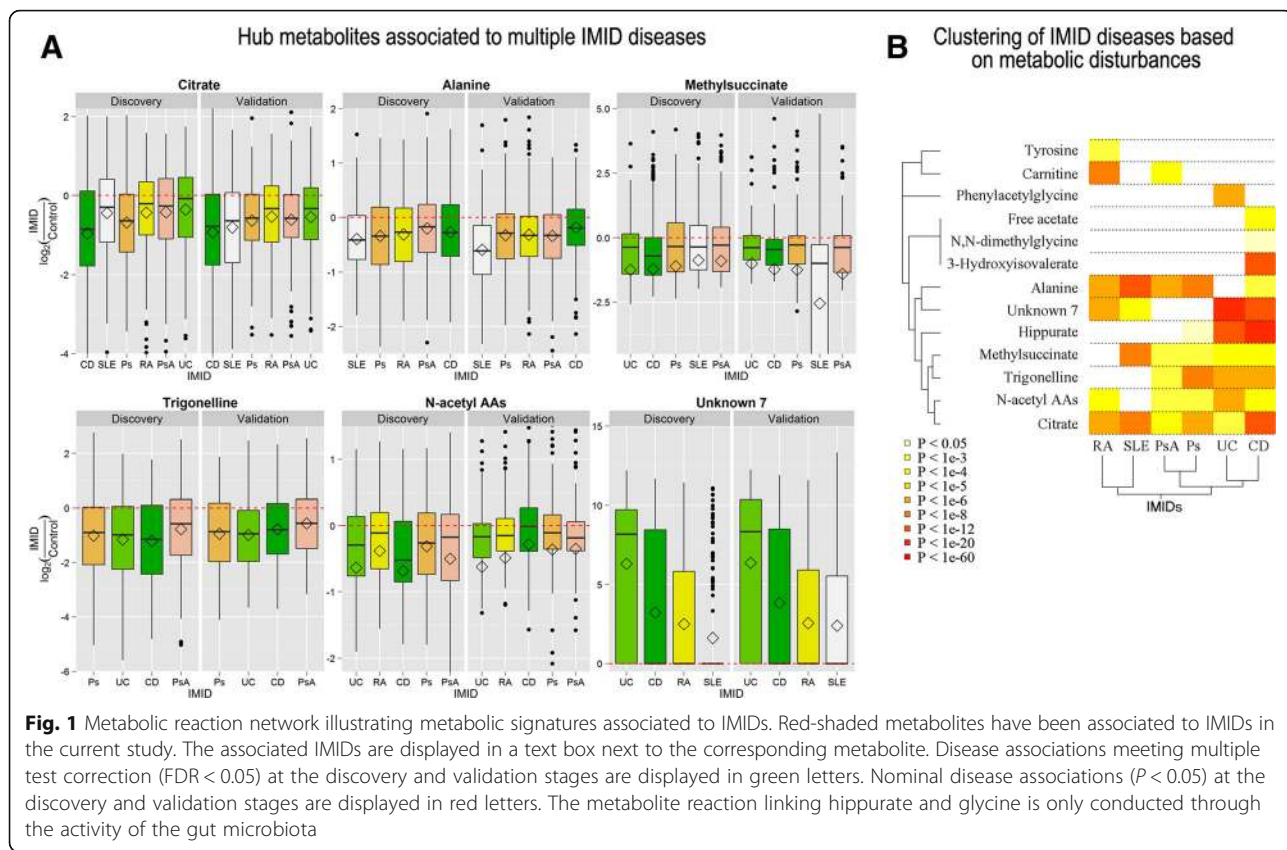
Fig. 1 Metabolic reaction network illustrating metabolic signatures associated to IMIDs. Red-shaded metabolites have been associated to IMIDs in the current study. The associated IMIDs are displayed in a text box next to the corresponding metabolite. Disease associations meeting multiple test correction (FDR < 0.05) at the discovery and validation stages are displayed in green letters. Nominal disease associations (*P* < 0.05) at the discovery and validation stages are displayed in red letters. The metabolite reaction linking hippurate and glycine is only conducted through the activity of the gut microbiota
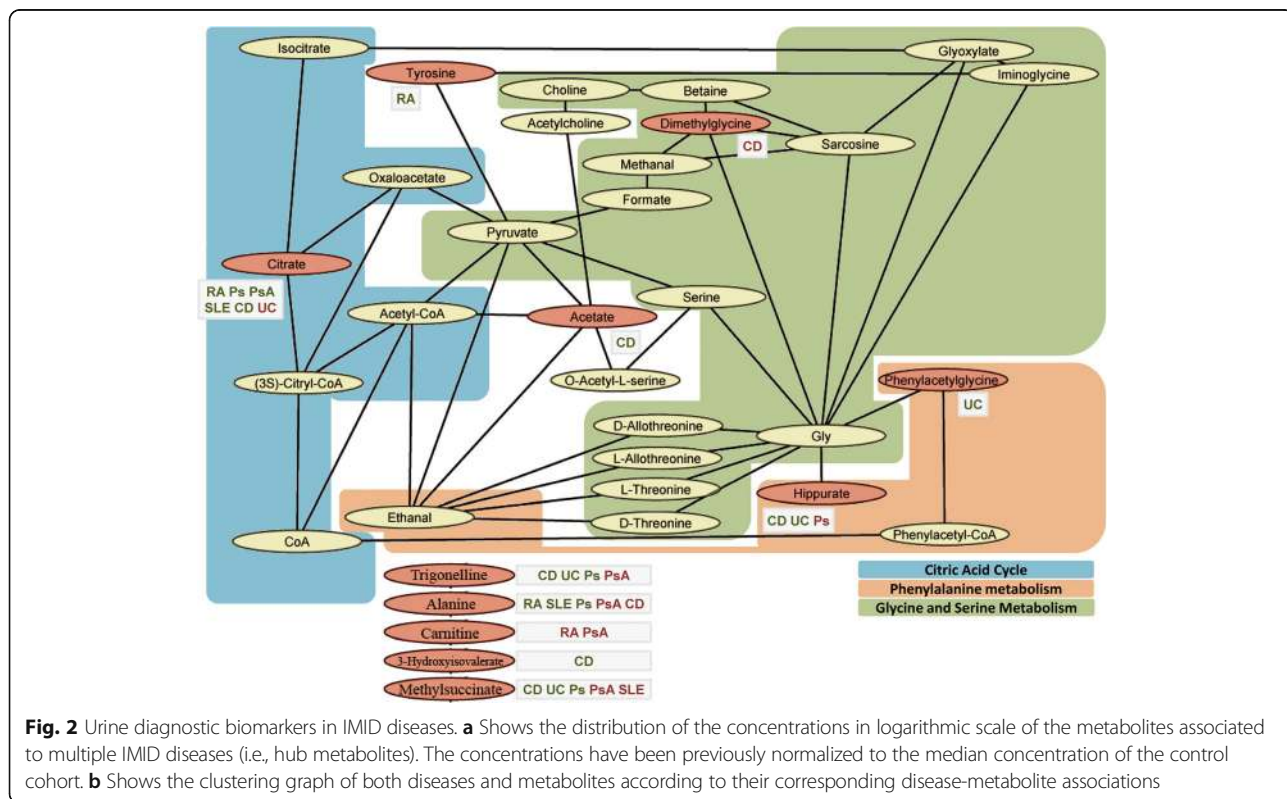
Fig. 2 Urine diagnostic biomarkers in IMID diseases. **a** Shows the distribution of the concentrations in logarithmic scale of the metabolites associated to multiple IMID diseases (i.e., hub metabolites). The concentrations have been previously normalized to the median concentration of the control cohort. **b** Shows the clustering graph of both diseases and metabolites according to their corresponding disease-metabolite associations

Alonso *et al. BMC Medicine* (2016) 14:133

Page 7 of 12

controls both in the discovery and validation cohorts, although the difference was only significant at the nominal ($P < 0.05$) level.

Similarly, five other hub metabolites were found to be significantly associated to multiple IMIDs. N-acetyl amino acids (N-acetyl AAs), alanine, methylsuccinate, and trigonelline showed lower concentrations in the urine of several different IMIDs compared to healthy normal controls (Table 2). From these, trigonelline has been previously shown to be associated to the consumption of coffee and tea. Our analysis shows that this metabolite remains significantly associated with different IMIDs even after adjusting for the daily consumption of coffee and/or tea, thereby discarding the possibility of a diet-based confounding ($P = 4.2 \times 10^{-6}$ and $r^2 = 0.47$ in the discovery cohort; Additional file 1: Figures S3 and S4). In addition to these metabolites, urine metabolite *Unknown 7* was found to be present at high levels in the urine metabolome of CD, UC, and RA patients compared to controls (Table 2).

A group of metabolites were found to have differential levels in urine only in IMIDs, with a more similar clinical phenotype. Hippurate levels were found to be significantly lower in the two inflammatory bowel diseases CD and UC compared to controls (Table 2). In the two chronic arthritis diseases, RA and PsA, low levels of carnitine were identified in the discovery stage and replicated in the validation stage (Table 2).

Finally, five metabolites were found to have a differential urine concentration in only one IMID. These disease-specific metabolites include phenylacetylglycine in UC ($P_{UC} = 2.7 \times 10^{-7}$), tyrosine in RA ($P_{RA} = 5.7 \times 10^{-4}$), and 3-hydroxyisovaleric ($P_{CD} = 1.1 \times 10^{-15}$), free acetate ($P_{CD} = 2.8 \times 10^{-5}$), and N,N-dimethylglycine in CD ($P_{CD} = 5.5 \times 10^{-3}$) (Table 2).

In order to assess the similarities between the urine metabolic profiles of the different IMIDs, we performed a clustering analysis (Fig. 2b). This analysis showed that the urine metabo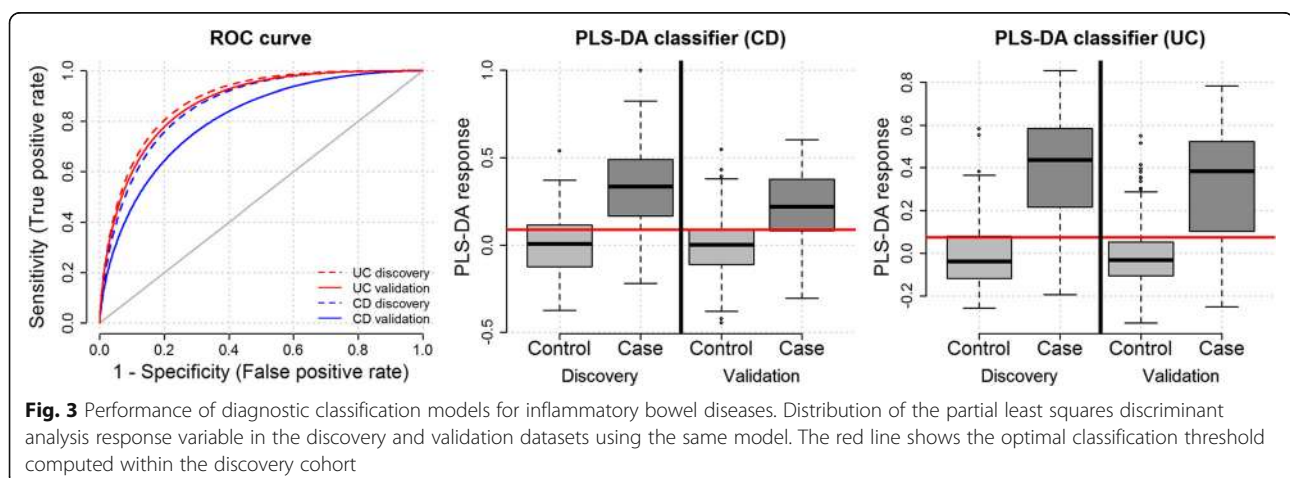lite profiles of IMIDs aggregate into three main clusters: (1) Ps and PsA (sharing $n = 5$ metabolite associations), (2) CD and UC (sharing $n = 6$ metabolite associations), and (3) RA and SLE (sharing $n = 3$ metabolite associations).

Urine metabolomic classifier for IMID diagnosis

In order to evaluate the power of the urine metabolome for disease diagnosis, a multivariate classification model was built for each IMID disease using the discovery cohort. In order to obtain an independent and non-biased assessment of the diagnostic accuracy of the metabolomic classifiers, these were tested in the validation cohort. Using this approach, the prediction accuracy was found to be high for SLE ($AUC_{SLE} = 0.73$, 95 % CI, 0.68–0.78), RA ($AUC_{RA} = 0.70$, 95 % CI, 0.65–0.75), Ps ($AUC_{PS} = 0.70$, 95 % CI, 0.64–0.75), and PsA ($AUC_{PSA} = 0.69$, 95 % CI, 0.63–0.74). The metabolomic classifiers from the two bowel inflammatory diseases, CD and UC, showed the strongest diagnostic performance (Fig. 3, Additional file 1: Figure S5). Using the metabolite levels in urine, both CD and UC could be predicted with an AUC higher than 0.80 ($AUC_{UC} = 0.87$, 95 % CI, 0.83–0.91 and $AUC_{CD} = 0.81$, 95 % CI, 0.76–0.86).

Urine biomarkers for differential diagnosis in IMIDs

The metabolite profiles of IMIDs showing a more similar clinical phenotype were directly compared, i.e., CD versus UC, RA versus PsA, Ps versus PsA, and RA versus SLE. In the discovery dataset, a total of 11 metabolites were found to be significantly different between similar IMIDs (FDR < 0.05, Additional file 1: Table S3). From these, three metabolite associations were replicated in the validation cohort (FDR < 0.05, Additional file 1: Table S3). These three validated differential diagnostic metabolites were all found when comparing the profiles of the two inflammatory bowel diseases UC and CD: hippurate ($P = 9.2 \times 10^{-8}$), citrate ($P = 1.6 \times 10^{-8}$), and *Unknown 7* ($P = 6.7 \times 10^{-18}$). All three metabolites

Fig. 3 Performance of diagnostic classification models for inflammatory bowel diseases. Distribution of the partial least squares discriminant analysis response variable in the discovery and validation datasets using the same model. The red line shows the optimal classification threshold computed within the discovery cohort

Alonso *et al. BMC Medicine* (2016) 14:133

Page 8 of 12

showed lower concentrations in the urine of CD patients compared to the urine of UC patients. At the nominal level, tyrosine amino acid ($P = 1.8 \times 10^{-4}$) and *Unknown 7* metabolite ($P = 7.9 \times 10^{-5}$) were also found to be lower in the urine of PsA patients compared to RA patients.
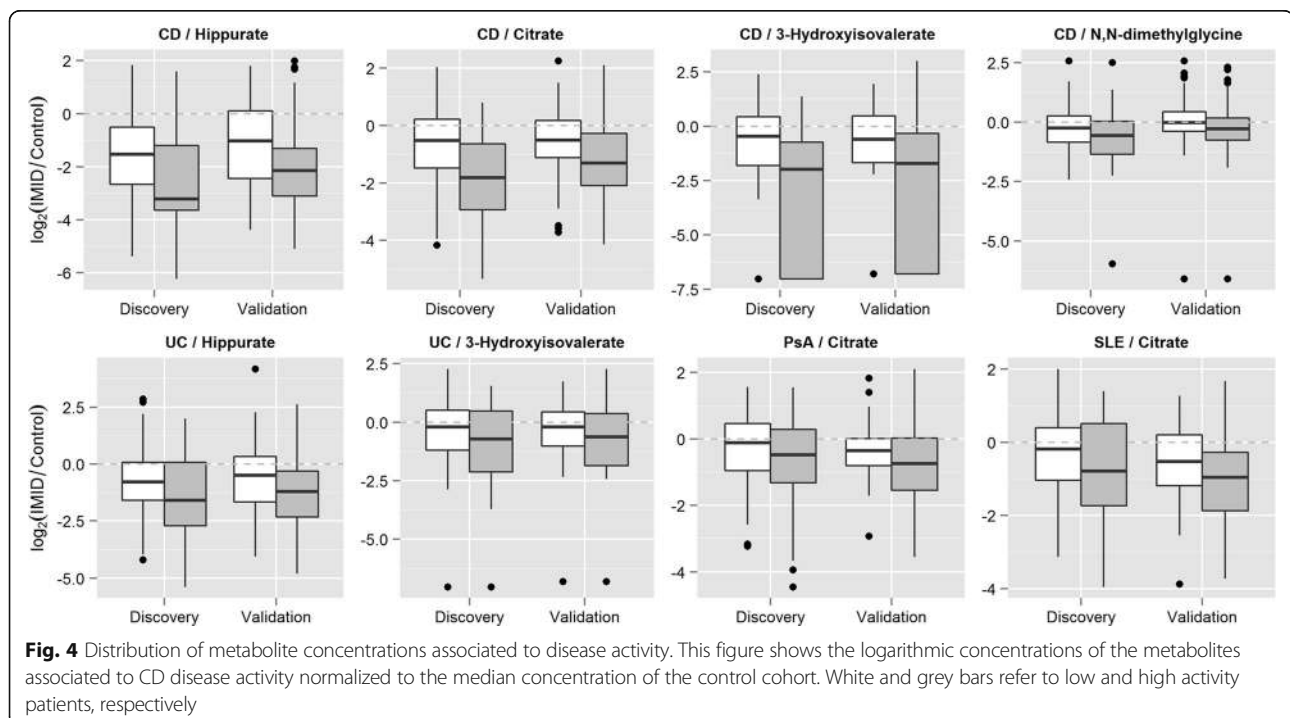
Urine biomarkers of disease activity in IMIDs

In the discovery cohort, three metabolites – citrate, hippurate, and 3-hydroxyisovalerate – were found to be significantly associated with disease activity in CD after multiple-test correction (Fig. 4, Additional file 1: Table S4). In particular, CD patients with high levels of disease activity were found to have much lower levels of these three metabolites compared to patients with low disease activity. Using the validation cohort, the association between the low levels of these three metabolites in urine and high disease activity in CD was replicated ($P_{citrate} = 4.4 \times 10^{-10}$, $P_{hippurate} = 6.0 \times 10^{-7}$, and $P_{3-hydroxyisovalerate} = 1.30 \times 10^{-5}$).

After multiple test correction, no other urine metabolite was significantly associated with disease activity. At the nominal level, however, five additional urine metabolites were associated with disease activity in both the discovery and validation cohorts ($P < 0.05$, Additional file 1: Table S4). The direction of the association was the same in both discovery and validation cohorts, which strongly supports the association of these biomarkers as candidates for disease activity monitoring. In UC, high disease activity was associated with low levels of urine hippurate and 3-hydroxyisovaleric acid ($P = 8.0 \times 10^{-5}$ and $P =$

1.4×10^{-3}, respectively). In PsA and SLE, patients with higher disease activity had lower levels of citrate ($P = 1.8 \times 10^{-5}$ and $P = 1.3 \times 10^{-3}$, respectively). Finally, low levels of N,N-dimethylglycine were also found to be associated with high disease activity in CD ($P = 9.0 \times 10^{-4}$).

Discussion

The metabolome represents the collection of small molecules produced by cells and, therefore, its analysis is providing a unique opportunity to identify biological perturbations associated with diseases [29, 45–47]. New technological advances are allowing the characterization of such biochemical variations, revealing unexpected metabolic changes associated with different human pathologies. From a translational perspective, the analysis of the metabolome is beginning to provide new and powerful biomarkers that are highly informative of specific disease processes and, therefore, could lead to more precise and efficient patient management. Despite their prevalence, there remain few studies analyzing the metabolome of IMIDs. In the present study, we report, for the first time, the results of a parallel analysis of the urine metabolome of six of the most prevalent IMIDs – RA, PsA, Ps, SLE, CD, and UC – for the search of clinically relevant biomarkers. Using a two-stage approach we have identified and validated multiple urine metabolites associated with disease diagnosis as well as disease activity. These results provide the most comprehensive analysis of the urine metabolome in IMIDs performed to date, leading to the identification of new biomarker

Fig. 4 Distribution of metabolite concentrations associated to disease activity. This figure shows the logarithmic concentrations of the metabolites associated to CD disease activity normalized to the median concentration of the control cohort. White and grey bars refer to low and high activity patients, respectively

Alonso *et al. BMC Medicine* (2016) 14:133

Page 9 of 12

metabolites, as well as providing strong evidence of shared metabolic pathways in this group of diseases.

The present large-scale profiling of the urine metabolome study has found unexpected strong similarities between IMIDs. Some of these metabolite variations were common across all or almost all diseases and, therefore, were considered as hub metabolites. To our knowledge, it is the first time that hub metabolites have been described in IMIDs. Among these metabolites, citrate, a central metabolite of the Krebs oxidative phosphorylation cycle, showed the strongest association to all IMIDs. Despite its essential role in cell energy production, citrate has been recently shown to have important immunologic properties [48], modulating, for example, the production of proinflammatory factors in macrophages or being a critical factor for dendritic cell antigen presentation. Previous studies have found that citrate is present at lower concentrations in the urine of inflammatory bowel disease (IBD) patients compared to controls [49, 50]. In RA and SLE, citrate has also been found to be in lower levels in the serum of patients compared to controls [51, 52]. Here, we show that the previously observed citrate variation in RA and SLE is also detected in urine, a much less invasive sample source than whole blood. Finally, we also demonstrate, for the first time, that Ps and PsA patients also have low concentrations of urine citrate compared to healthy controls. Together, the results of this study provide strong evidence of the presence of hub metabolites that could become "pan-IMID" biomarkers that could be easily measured in routine clinical settings.

The parallel analysis of this group of diseases has led to unique findings. The unsupervised analysis of the urine metabolite associations showed three strong and reproducible clusters of clinically similar IMIDs: (1) IMIDs involving skin affection (i.e., Ps and PsA), (2) inflammatory bowel diseases (i.e., CD and UC), and (3) RA and SLE, two diseases characterized by having a higher prevalence in women. These results correlate with the observed shared genetic risk components observed between different IMIDs using genome-wide association studies [53–56]. For example, CD and UC have shown to share more than 163 disease risk loci [57], Ps an PsA share up to 30 risk loci [58, 59], and SLE and RA have more than 80 common risk variants [60]. To our knowledge, it is the first time that metabolite patterns in urine have shown to etiologically group more similar IMIDs. This result confirms the validity of the urine metabolome in the characterization of biochemical pathways that are specifically associated with this group of diseases.

When assessing the metabolic context of the disease-associated metabolites by integrating the metabolic reactions that link them, the resulting network showed a high degree of overlap of three main metabolic pathways (Fig. 1). From these, the citric acid cycle is the predominant pathway identified, with citrate showing a common association to all the IMIDs. Previous studies have already shown that alterations within this metabolic pathway are related to immunity and inflammation, although the functional implications of the alterations of this pathway are still being investigated [61]. The second major metabolic pathway was the phenylalanine metabolism pathway. The metabolites included in this pathway have shown relevant and specific associations to IBDs in this study. This finding agrees with previous metabolomic studies that have shown the importance of this pathway in the etiology of IBDs [62]. Finally, network analysis also showed an important role for the glycine and serine metabolism pathway in IMIDs. Metabolites within this pathway act as major connectors between the two previous pathways and have been previously related with inflammatory processes. Glycine, the most connected metabolite in the resulting network, has been previously proposed to be an anti-inflammatory and immunomodulatory agent [63]. Although not directly detected by the NMR approach used in this study, our results strongly suggest that glycine could be a highly informative biomarker to the inflammatory processes that characterize IMIDs. Future studies using alternative analysis technologies like mass-spectrometry will help to determine the utility of this metabolite as a clinical biomarker of autoimmune diseases.

In this study, we also demonstrate that the urine metabolome has great potential for assessing disease activity. Citrate, the strongest hub metabolite for IMID diagnosis, was found to correlate with high disease activity in CD, PsA, and SLE. In IBDs, we also demonstrate that hippurate has a very strong correlation with disease activity. Therefore, this urine metabolite could be used not only for early disease diagnosis but also to monitor the level of disease activity in IBDs. This result further strengthens previously reported results that show how changes in the microbiome correlate with the level of inflammation in the gut and disease activity in IBD patients [64–67]. Future studies, aimed at characterizing the interrelation between bacterial species in the gut, tissue inflammation and the urine metabolites identified herein could therefore help to develop more objective and reproducible systems to monitor disease progression in IBDs.

The disease diagnostic models built in this study using the urine metabolites were found to have good performance in all IMIDs. In IBDs in particular, the classifiers were found to predict the disease with very high accuracy. These results are in agreement with previous studies [50, 68, 69] that suggested the use of urine metabolites for the diagnosis of IBDs. Compared to previous studies, we

Alonso *et al. BMC Medicine* (2016) 14:133

Page 10 of 12

here provide, for the first time, a validation analysis of the diagnostic predictor using an independent and large patient and control cohort. Providing an independent confirmatory analysis is an essential step for any new molecular diagnostic tool [70]. These findings support the analysis of the urine metabolome as a simple, cost-effective and non-invasive approach for the diagnosis of IBDs.

To our knowledge, there is no evidence that the metabolite patterns associated with IMIDs in this study have been previously associated to other diseases. While variations in single metabolites like citrate have been associated with other disease etiologies, the diagnostic ability generated by the combination of multiple metabolites clearly holds a much higher potential to be the approach finally used in the clinical setting. As shown in this study, it is the integration of variation in multiple metabolites that gives the best disease prediction accuracies. In order to further consolidate these diagnostic metabolite patterns as clinically useful tools, the next steps will include the study of the urine metabolome in individuals with pre-diagnostic symptoms as well as longitudinal studies to assess biomarker variability and correlation with specific features of disease progression. Further, future developments of the disease predictors could evaluate the inclusion of other molecular features like the presence of autoantibodies in sera or, even, the identification of additional metabolites in urine using mass-spectrometry approaches. For this latter objective, the results of this study will clearly be a highly valuable starting point.

Conclusion

We have performed, for the first time, a large-scale high-throughput profiling of the urine metabolome of six of the most prevalent IMIDs. Using a discovery and an independent validation cohort we have identified multiple urine metabolites associated with the diagnosis and the monitoring of disease activity. The parallel evaluation of all six IMIDs has allowed the identification of hub metabolites as well as the characterization of clusters of clinically similar diseases based exclusively on urine metabolite profiles. These common molecular features are in agreement with the shared genetic risk in IMIDs recently identified through genome-wide association studies [54]. Taken together, these results demonstrate the utility of urine metabolomics as a new source for clinically useful biomarkers for this prevalent group of chronic inflammatory diseases.

Additional file

Additional file 1: Supplementary note (Members of the Immune-Mediated Inflammatory Disease (IMID) Consortium), Supplementary methods, Supplementary Tables (**Table S1:** Sample quality control;

Table S2: Urine metabolite panel; **Table S3:** List of metabolic associations when comparing phenotypically closer IMID diseases; **Table S4:** List of metabolites with replicated associations to disease activity in Crohn's disease (CD) patients) and Supplementary Figures (**Figure S1:** Distribution of disease activity indices in the extreme low and high activity patient subgroups; **Figure S2:** Distribution of epidemiological and sample collection variables across the IMID and control groups; **Figure S3:** Distribution of trigonelline concentration according to daily coffee/tea consumption; **Figure S4:** Distribution of trigonelline concentration on each IMID cohort stratified by coffee/tea consumption; **Figure S5:** ROC curves of the diagnostic partial least squares discriminant analysis classification models and metabolite loadings of the CD and UC models). (PDF 2486 kb)

Abbreviations

AUC, area under the curve; CD, Crohn's disease; FDR, false discovery rate; IBD, inflammatory bowel disease; IMID, immune-mediated inflammatory disease; N-acetyl AAs, N-acetyl amino acids; NMR, nuclear magnetic resonance; Ps, psoriasis; PsA, psoriatic arthritis; RA, rheumatoid arthritis; ROC, receiver operating characteristic; SLE, systemic lupus erythematosus; UC, ulcerative colitis

Alonso *et al. BMC Medicine* (2016) 14:133

Page 11 of 12

Authors' contributions

All the authors made substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; and also contributed to drafting the article or revising it critically for important intellectual content. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Rheumatology Research Group, Vall d'Hebron Hospital Research Institute, Barcelona, Spain. [2]Centre for Omic Sciences, COS-DEEEA-URV-IISPV, Reus, Spain. [3]Metabolomics Platform, CIBERDEM, Reus, Spain. [4]Hospital Universitari Germans Trias i Pujol, Badalona, Spain. [5]CIBERehd, Madrid, Spain. [6]UGC Reumatología, Instituto de Investigación Biomédica (IBIMA), Hospital Regional Universitario de Málaga, Universidad de Málaga, Málaga, Spain. [7]Hospital Clínic de Barcelona and IDIBAPS, Barcelona, Spain. [8]Hospital Universitario Guadalajara, Guadalajara, Spain. [9]Hospital Universitario de la Princesa and IIS-IP, Madrid, Spain. [10]Hospital la Fe, Valencia, Spain. [11]Hospital General de Alicante, Alicante, Spain. [12]Hospital de la Santa Creu i Sant Pau, Barcelona, Spain. [13]Complejo Hospitalario Juan Canalejo, INIBIC, A Coruña, Spain. [14]Hospital Universitario Marqués de Valdecilla, Santander, Spain.

References

1. Burisch J, Jess T, Martinato M, Lakatos PL. The burden of inflammatory bowel disease in Europe. J Crohns Colitis. 2013;7(4):322–37.
2. Chandran V, Raychaudhuri SP. Geoepidemiology and environmental factors of psoriasis and psoriatic arthritis. J Autoimmun. 2010;34(3):J314–21.
3. Ferrándiz C. Bordas, García P, Puig S, Pujol R, Smandía A. Prevalence of psoriasis in Spain (Epiderma Project: phase I). J Eur Acad Dermatol Venereol. 2001;15(1):20–3.
4. Shapira Y, Agmon-Levin N, Shoenfeld Y. Geoepidemiology of autoimmune rheumatic diseases. Nat Rev Rheumatol. 2010;6(8):468–76.
5. Rosman Z, Shoenfeld Y, Zandman-Goddard G. Biologic therapy for autoimmune diseases: an update. BMC Med. 2013;11:88.
6. Cooper GS, Stroehla BC. The epidemiology of autoimmune diseases. Autoimmun Rev. 2003;2(3):119–25.
7. Eaton WW, Rose NR, Kalaydjian A, Pedersen MG, Mortensen PB. Epidemiology of autoimmune diseases in Denmark. J Autoimmun. 2007;29(1):1–9.
8. Youinou P, Pers J-O, Gershwin ME, Shoenfeld Y. Geo-epidemiology and autoimmunity. J Autoimmun. 2010;34(3):J163–7.
9. Hehlgans T, Pfeffer K. The intriguing biology of the tumour necrosis factor/tumour necrosis factor receptor superfamily: players, rules and the games. Immunology. 2005;115(1):1–20.
10. Anaya J-M. Common mechanisms of autoimmune diseases (the autoimmune tautology). Autoimmun Rev. 2012;11(11):781–4.
11. Hanauer SB. Positioning biologic agents in the treatment of Crohn's disease. Inflamm Bowel Dis. 2009;15(10):1570–82.
12. Scarpa R, Altomare G, Marchesoni A, Balato N, Matucci Cerinic M, Lotti T, Olivieri I, Vena GA, Salvarani C, Valesini G, et al. Psoriatic disease: concepts and implications. J Eur Acad Dermatol Venereol. 2010;24(6):627–30.
13. Emery P, Breedveld FC, Hall S, Durez P, Chang DJ, Robertson D, Singh A, Pedersen RD, Koenig AS, Freundlich B. Comparison of methotrexate monotherapy with a combination of methotrexate and etanercept in active, early, moderate to severe rheumatoid arthritis (COMET): a randomised, double-blind, parallel treatment trial. Lancet. 2008;372(9636):375–82.
14. Singh JA, Furst DE, Bharat A, Curtis JR, Kavanaugh AF, Kremer JM, Moreland LW, O'Dell J, Winthrop KL, Beukelman T, et al. 2012 Update of the 2008 American College of Rheumatology recommendations for the use of disease-modifying antirheumatic drugs and biologic agents in the treatment of rheumatoid arthritis. Arthritis Care Res. 2012;64(5):625–39.
15. Smolen JS, Han C, Van Der Heijde D, Emery P, Bathon JM, Keystone E, Kalden JR, Schiff M, Bala M, Baker D, et al. Infliximab treatment maintains employability in patients with early rheumatoid arthritis. Arthritis Rheum. 2006;54(3):716–22.
16. van der Kooij SM, le Cessie S, Goekoop-Ruiterman YPM, de Vries-Bouwstra JK, van Zeben D, Kerstens PJSM, Hazes JMW, van Schaardenburg D, Breedveld FC, Dijkmans BAC, et al. Clinical and radiological efficacy of initial

17. vs delayed treatment with infliximab plus methotrexate in patients with early rheumatoid arthritis. Ann Rheum Dis. 2009;68(7):1153–8.
17. Schoepfer AM, Dehlavi M-A, Fournier N, Safroneeva E, Straumann A, Pittet V, Peyrin-Biroulet L, Michetti P, Rogler G, Vavricka SR. Diagnostic delay in Crohn's disease is associated with a complicated disease course and increased operation rate. Am J Gastroenterol. 2013;108(11):1744–53.
18. Bakker MF, Cavet G, Jacobs JWG, Bijlsma JWJ, Haney DJ, Shen Y, Hesterberg LK, Smith DR, Centola M, van Roon JAG, et al. Performance of a multi-biomarker score measuring rheumatoid arthritis disease activity in the CAMERA tight control study. Ann Rheum Dis. 2012;71(10):1692–7.
19. Uhlig T, Kvien TK, Pincus T. Test–retest reliability of disease activity core set measures and indices in rheumatoid arthritis. Ann Rheum Dis. 2009;68(6):972–5.
20. Castro-Santos P, Laborde CM, Diaz-Pena R. Genomics, proteomics and metabolomics: their emerging roles in the discovery and validation of rheumatoid arthritis biomarkers. Clin Exp Rheumatol. 2015;8:8.
21. Huang H, Vangay P, McKinlay CE, Knights D. Multi-omics analysis of inflammatory bowel disease. Immunol Lett. 2014;162(2, Part A):62–8.
22. Villanova F, Di Meglio P, Nestle FO. Biomarkers in psoriasis and psoriatic arthritis. Ann Rheum Dis. 2013;72 Suppl 2:ii104–10.
23. Thongboonkerd V. Urinary proteomics: towards biomarker discovery, diagnostics and prognostics. Mol Biosyst. 2008;4(8):810–5.
24. Zhang A, Sun H, Wu X, Wang X. Urine metabolomics. Clin Chim Acta. 2012;414:65–9.
25. Patti GJ, Yanes O, Siuzdak G. Innovation: Metabolomics: the apogee of the omics trilogy. Nat Rev Mol Cell Biol. 2012;13(4):263–9.
26. Collino S, Martin F-PJ, Rezzi S. Clinical metabolomics paves the way towards future healthcare strategies. Br J Clin Pharmacol. 2013;75(3):619–29.
27. Zhang A, Sun H, Wang P, Han Y, Wang X. Modern analytical techniques in metabolomics analysis. Analyst. 2012;137(2):293–300.
28. De Preter V, Verbeke K. Metabolomics as a diagnostic tool in gastroenterology. World J Gastrointest Pharmacol Ther. 2013;4(4):97–107.
29. Julià A, Alonso A, Marsal S. Metabolomics in rheumatic diseases. Int J Clin Rheumatol. 2014;9(4):353–69.
30. Julià A, Domènech E, Ricart E, Tortosa R, García-Sánchez V, Gisbert JP, Nos Mateu P, Gutiérrez A, Gomollón F, Mendoza JL, et al. A genome-wide association study on a southern European population identifies a new Crohn's disease susceptibility locus at RBX1-EP300. Gut. 2013;62(10):1440–5.
31. Julià A, Tortosa R, Hernanz JM, Cañete JD, Fonseca E, Ferrándiz C, Unamuno P, Puig L, Fernández-Sueiro JL, Sanmartí R, et al. Risk variants for psoriasis vulgaris in a large case–control collection and association with clinical subphenotypes. Hum Mol Genet. 2012;21(20):4549–57.
32. Alonso A, Domènech E, Julià A, Panés J, García-Sánchez V, Mateu PN, Gutiérrez A, Gomollón F, Mendoza JL, Garcia-Planella E, et al. Identification of risk loci for Crohn's disease phenotypes using a genome-wide association study. Gastroenterology. 2015;148(4):794–805.
33. Alonso A, Rodríguez MA, Vinaixa M, Tortosa R, Correig X, Julià A, Marsal S. Focus: A Robust Workflow for One-Dimensional NMR Spectral Analysis. Anal Chem. 2013;86(2):1160–9.
34. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, Djoumbou Y, Mandal R, Aziat F, Dong E, et al. HMDB 3.0—The Human Metabolome Database in 2013. Nucleic Acids Res. 2013;41(Database issue):D801–7.
35. Menni C, Kastenmüller G, Petersen AK, Bell JT, Psatha M, Tsai P-C, Gieger C, Schulz H, Erte I, John S, et al. Metabolomic markers reveal novel pathways of ageing and early development in human populations. Int J Epidemiol. 2013;42(4):1111–9.
36. Shah SH, Hauser ER, Bain JR, Muehlbauer MJ, Haynes C, Stevens RD, Wenner BR, Dowdy ZE, Granger CB, Ginsburg G, et al. High heritability of metabolomic profiles in families burdened with premature cardiovascular disease. Mol Syst Biol. 2009;5:258.
37. Wang-Sattler R, Yu Z, Herder C, Messias AC, Floegel A, He Y, Heim K, Campillos M, Holzapfel C, Thorand B, et al. Novel biomarkers for pre-diabetes identified by metabolomics. Mol Syst Biol. 2012;8:615.
38. Alonso A, Marsal S, Julià A. Analytical methods in untargeted metabolomics: state of the art in 2015. Front Bioeng Biotechnol. 2015;3:23.
39. Posma JM, Robinette SL, Holmes E, Nicholson JK. MetaboNetworks, an interactive Matlab-based toolbox for creating, customizing and exploring sub-networks from KEGG. Bioinformatics. 2014;30(6):893–5.
40. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 2012;40(Database issue):D109–14.

Alonso *et al. BMC Medicine* (2016) 14:133

Page 12 of 12

41. Elliott P, Posma JM, Chan Q, Garcia-Perez I, Wijeyesekera A, Bictash M, Ebbels TM, Ueshima H, Zhao L, van Horn L. Urinary metabolic signatures of human adiposity. Sci Transl Med. 2015;7(285):285ra62.
42. Bouatra S, Aziat F, Mandal R, Guo AC, Wilson MR, Knox C, Bjorndahl TC, Krishnamurthy R, Saleem F, Liu P, et al. The human urine metabolome. PLoS One. 2013;8(9):e73076.
43. Gronwald W, Klein MS, Zeltner R, Schulze B-D, Reinhold SW, Deutschmann M, Immervoll A-K, Boger CA, Banas B, Eckardt K-U, et al. Detection of autosomal dominant polycystic kidney disease by NMR spectroscopic fingerprinting of urine. Kidney Int. 2011;79(11):1244–53.
44. Kang S-M, Park J-C, Shin M-J, Lee H, Oh J, Ryu DH, Hwang G-S, Chung JH. 1H nuclear magnetic resonance based metabolic urinary profiling of patients with ischemic heart failure. Clin Biochem. 2011;44(4):293–9.
45. Armitage EG, Barbas C. Metabolomics in cancer biomarker discovery: current trends and future perspectives. J Pharm Biomed Anal. 2014;87:1–11.
46. Duarte IF, Diaz SO, Gil AM. NMR metabolomics of human blood and urine in disease research. J Pharm Biomed Anal. 2014;93:17–26.
47. Mastrangelo A, Armitage EG, García A, Barbas C. Metabolomics as a tool for drug discovery and personalised medicine. A Review. Curr Top Med Chem. 2014;14(23):2627–36.
48. Infantino V, Iacobazzi V, Menga A, Avantaggiati ML, Palmieri F. A key role of the mitochondrial citrate carrier (SLC25A1) in TNFα- and IFNγ-triggered inflammation. Biochim Biophys Acta. 2014;1839(11):1217–25.
49. Dawiskiba T, Deja S, Mulak A, Ząbek A, Jawień E, Pawełka D, Banasik M, Mastalerz-Migas A, Balcerzak W, Kaliszewski K, et al. Serum and urine metabolomic fingerprinting in diagnostics of inflammatory bowel diseases. World J Gastroenterol. 2014;20(1):163–74.
50. Stephens NS, Siffledeen J, Su X, Murdoch TB, Fedorak RN, Slupsky CM. Urinary NMR metabolomic profiles discriminate inflammatory bowel disease from healthy. J Crohns Colitis. 2013;7(2):e42–8.
51. Jiang M, Chen T, Feng H, Zhang Y, Li L, Zhao A, Niu X, Liang F, Wang M, Zhan J, et al. Serum metabolic signatures of four types of human arthritis. J Proteome Res. 2013;12(8):3769–79.
52. Ouyang X, Dai Y, Wen JL, Wang LX. 1H NMR-based metabolomic study of metabolic profiling for systemic lupus erythematosus. Lupus. 2011;20(13):1411–20.
53. Cho JH, Gregersen PK. Genomics and the multifactorial nature of human autoimmune disease. N Engl J Med. 2011;365(17):1612–23.
54. Cotsapas C, Voight BF, Rossin E, Lage K, Neale BM, Wallace C, Abecasis GR, Barrett JC, Behrens T, Cho J, et al. Pervasive sharing of genetic effects in autoimmune disease. PLoS Genet. 2011;7(8):e1002254.
55. Li Y, Begovich AB. Unraveling the genetics of complex diseases: susceptibility genes for rheumatoid arthritis and psoriasis. Semin Immunol. 2009;21(6):318–27.
56. Sirota M, Schaub MA, Batzoglou S, Robinson WH, Butte AJ. Autoimmune disease classification by inverse association with SNP alleles. PLoS Genet. 2009;5(12):e1000792.
57. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, Lee JC, Philip Schumm L, Sharma Y, Anderson CA, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012;491(7422):119–24.
58. Liu Y, Helms C, Liao W, Zaba LC, Duan S, Gardner J, Wise C, Miner A, Malloy MJ, Pullinger CR, et al. A genome-wide association study of psoriasis and psoriatic arthritis identifies new disease loci. PLoS Genet. 2008;4(4):e1000041.
59. Hébert HL, Ali FR, Bowes J, Griffiths CEM, Barton A, Warren RB. Genetic susceptibility to psoriasis and psoriatic arthritis: implications for therapy. Br J Dermatol. 2012;166(3):474–82.
60. Márquez A, Vidal-Bralo L, Rodríguez-Rodríguez L, González-Gay MA, Balsa A, González-Álvaro I, Carreira P, Ortego-Centeno N, Ayala-Gutiérrez MM, García-Hernández FJ, et al. A combined large-scale meta-analysis identifies COG6 as a novel shared risk locus for rheumatoid arthritis and systemic lupus erythematosus. Ann Rheum Dis. 2016.doi: 10.1136/annrheumdis-2016-209436.
61. McGettrick AF, O'Neill LAJ. How metabolism generates signals during innate immunity and inflammation. J Biol Chem. 2013;288(32):22893–8.
62. Jansson J, Willing B, Lucio M, Fekete A, Dicksved J, Halfvarson J, Tysk C, Schmitt-Kopplin P. Metabolomics reveals metabolic biomarkers of Crohn's disease. PLoS One. 2009;4(7):e6386.
63. Wang W, Wu Z, Dai Z, Yang Y, Wang J, Wu G. Glycine metabolism in animals and humans: implications for nutrition and health. Amino Acids. 2013;45(3):463–77.
64. Baumgart M, Dogan B, Rishniw M, Weitzman G, Bosworth B, Yantiss R, Orsi RH, Wiedmann M, McDonough P, Kim SG, et al. Culture independent analysis of ileal mucosa reveals a selective increase in invasive Escherichia coli of novel phylogeny relative to depletion of Clostridiales in Crohn's disease involving the ileum. ISME J. 2007;1(5):403–18.
65. Frank DN, St. Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proc Natl Acad Sci. 2007;104(34):13780–5.
66. Li M, Wang B, Zhang M, Rantalainen M, Wang S, Zhou H, Zhang Y, Shen J, Pang X, Zhang M, et al. Symbiotic gut microbes modulate human metabolic phenotypes. Proc Natl Acad Sci. 2008;105(6):2117–22.
67. Machiels K, Joossens M, Sabino J, De Preter V, Arijs I, Eeckhaut V, Ballet V, Claes K, Van Immerseel F, Verbeke K, et al. A decrease of the butyrate-producing species Roseburia hominis and Faecalibacterium prausnitzii defines dysbiosis in patients with ulcerative colitis. Gut. 2014;63(8):1275–83. doi:10.1136/gutjnl-2013-304833.
68. Schicho R, Shaykhutdinov R, Ngo J, Nazyrova A, Schneider C, Panaccione R, Kaplan GG, Vogel HJ, Storr M. Quantitative metabolomic profiling of serum, plasma, and urine by 1H NMR spectroscopy discriminates between patients with inflammatory bowel disease and healthy individuals. J Proteome Res. 2012;11(6):3344–57.
69. Williams HRT, Cox IJ, Walker DG, North BV, Patel VM, Marshall SE, Jewell DP, Ghosh S, Thomas HJW, Teare JP, et al. Characterization of inflammatory bowel disease with urinary metabolic profiling. Am J Gastroenterol. 2009;104(6):1435–44.
70. Xia J, Broadhurst DI, Wilson M, Wishart DS. Translational biomarker discovery in clinical metabolomics: an introductory tutorial. Metabolomics. 2013;9(2):280–99.
71. Harvey RF, Bradshaw JM. A simple index of Crohn's-disease activity. Lancet. 1980;315(8167):514.
72. Lichtiger S, Present DH, Kornbluth A, Gelernt I, Bauer J, Galler G, Michelassi F, Hanauer S. Cyclosporine in severe ulcerative colitis refractory to steroid therapy. N Engl J Med. 1994;330(26):1841–5.
73. Prevoo MLL, Van'T Hof MA, Kuper HH, Van Leeuwen MA, Van De Putte LBA, Van Riel PLCM. Modified disease activity scores that include twenty-eight-joint counts development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. Arthritis Rheum. 1995;38(1):44–8.
74. Fredriksson T, Pettersson U. Severe psoriasis–oral therapy with a new retinoid. Dermatologica. 1978;157(4):238–44.
75. Petri M, Kim MY, Kalunian KC, Grossman J, Hahn BH, Sammaritano LR, Lockshin M, Merrill JT, Belmont HM, Askanase AD, et al. Combined oral contraceptives in women with systemic lupus erythematosus. N Engl J Med. 2005;353(24):2550–8.
76. Hay EM, Bacon PA, Gordon C, Isenberg DA, Maddison P, Snaith ML, Symmons DPM, Viner N, Zoma A. The BILAG index: a reliable and valid instrument for measuring clinical disease activity in systemic lupus erythematosus. QJM. 1993;86(7):447–58.