Mahapatra et al. BMC Infectious Diseases 2014, 14:53
http://www.biomedcentral.com/1471-2334/14/53

BMC Infectious Diseases

RESEARCH ARTICLE Open Access

A metabolic biosignature of early response to anti-tuberculosis treatment

Sebabrata Mahapatra[1], Ann M Hess[1], John L Johnson[2,3], Kathleen D Eisenach[4], Mary A DeGroote[1], Phineas Gitta[3], Moses L Joloba[3,5], Gilla Kaplan[2,6], Gerhard Walzl[7], W Henry Boom[2,3] and John T Belisle[1*]

Abstract

Background: The successful treatment of tuberculosis (TB) requires long-term multidrug chemotherapy. Clinical trials to evaluate new drugs and regimens for TB treatment are protracted due to the slow clearance of *Mycobacterium tuberculosis (Mtb)* infection and the lack of early biomarkers to predict treatment outcome. Advancements in the field of metabolomics make it possible to identify metabolic profiles that correlate with disease states or successful chemotherapy. However, proof-of-concept of this approach has not been provided for a TB-early treatment response biosignature (TB-ETRB).

Methods: Urine samples collected at baseline and during treatment from 48 Ugandan and 39 South African HIV-seronegative adults with pulmonary TB were divided into discovery and qualification sets, normalized to creatinine concentration, and analyzed by liquid chromatography-mass spectrometry to identify small molecule molecular features (MFs) in individual patient samples. A biosignature that distinguished baseline and 1 month treatment samples was selected by pairwise t-test using data from two discovery sample sets. Hierarchical clustering and repeated measures analysis were applied to additional sample data to down select molecular features that behaved consistently between the two clinical sites and these were evaluated by logistic regression analysis.

Results: Analysis of discovery samples identified 45 MFs that significantly changed in abundance at one month of treatment. Down selection using an extended set of discovery samples and qualification samples confirmed 23 MFs that consistently changed in abundance between baseline and 1, 2 and 6 months of therapy, with 12 MFs achieving statistical significance ($p < 0.05$). Six MFs classified the baseline and 1 month samples with an error rate of 11.8%.

Conclusions: These results define a urine based TB-early treatment response biosignature (TB-ETRB) applicable to different parts of Africa, and provide proof-of-concept for further evaluation of this technology in monitoring clinical responses to TB therapy.

Keywords: Tuberculosis, Metabolomics, Biomarker, Mass spectrometry, Small molecule biosignature, Anti-tuberculosis therapy, *Mycobacterium tuberculosis*, Urine

Background

The emergence of multidrug resistant strains of *Mycobacterium tuberculosis (Mtb)* underscores the need for new drugs and shorter regimens to treat tuberculosis (TB), one of the world's most prevalent infectious diseases. Considerable advancements have been realized in anti-TB drug development, but clinical evaluation of new drugs remains a protracted process [1,2]. Phase 2 clinical trials typically use patient conversion to negative-sputum culture after two months of treatment as a biomarker to assess therapeutic efficacy. However, this measure lacks the robustness needed to evaluate smaller numbers of patients in each arm of early clinical trials, or to evaluate treatment shortening regimens in Phase 3 trials [1,3]. Therefore, alternative, quantitative biomarkers that reliably measure a patient's response to anti-TB treatment after a short period of time and serve as surrogate endpoints are needed to accelerate clinical trials [4,5].

Disease and inflammatory states correlate with changes in the biochemistry of a system (host and pathogen), and

* Correspondence: jbelisle@colostate.edu
[1]Mycobacteria Research Laboratories, Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, CO 80523, USA
Full list of author information is available at the end of the article

BioMed Central

© 2014 Mahapatra et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited.

metabolomic approaches provide a direct measure of a systems biochemical profile [6-8]. Additionally, a diseased-state metabolic profile would be expected to revert to a normal-state (non-diseased) in response to successful treatment. Modern analytical platforms such as mass spectrometry (MS) provide accurate methods for assessing complex metabolic profiles, and when combined with multivariate statistical analyses, MS can elucidate time related metabolic changes that correlate with transition from a diseased- to a normal-state [7,9,10]. Metabolic flux has been exploited in the study of oncology, diabetes and cardiovascular disease, but has not been widely applied to infectious disease biomarker development [11-13]. We hypothesized that metabolomic analyses by liquid chromatography (LC)-MS of clinical samples collected at the time of TB diagnosis and at various treatment time points would reveal a metabolic flux that could be developed as a biosignature of treatment response.

Sputum and serum are conventional clinical specimens used for TB diagnosis and evaluation of TB treatment response [14,15]. However, urine is an alternative, non-invasive clinical sample, and has been successfully used to discover biomarkers for other diseases [11,16]. Urine contains a sizable fraction of the human metabolome and metabolites of microbial origin, and requires minimal processing for analysis by LC-MS [17-20]. Our efforts applied LC-MS to archived urine specimens collected as part of observational studies of TB patients undergoing standard therapy for pansusceptible, pulmonary TB. This resulted in the definition of a urine metabolite based TB-early treatment response biosignature (TB-ETRB) that measured a significant metabolic flux as early as one month of treatment.

Methods

Clinical samples

Anonymized archived urine samples used in the current studies were procured from the Tuberculosis Research Unit (TBRU) and Stellenbosch University. The samples provided by the TBRU originated from the NAA2m study (DMID 08–0023) conducted in Uganda (http://www.case.edu/affil/tbru/research_naa2m.html). This was a prospective observational cohort study of adults with newly-diagnosed sputum smear-positive, culture-confirmed pulmonary TB treated with supervised standard chemotherapy. Samples were collected from 48 patients at the time of TB diagnosis (D0) and month-1 (M1), month-2 (M2) and month-6 (M6) of treatment. The Stellenbosch University samples were from 39 patients from the Action TB Surrogate Marker study [21] and the D0, M1 and M6 time point samples were evaluated. All urine specimens were from adult pulmonary cavitary TB patients of both sexes without HIV co-infection. Urine specimens were stored at −80°C upon collection. All patients successfully responded to anti-TB therapy.

Study NAA2m (DMID 08–0023; A Pilot Study to Evaluate Nucleic Acid Amplification (NAA) and Other Tests to Predict Relapse of Tuberculosis and to Monitor the Effectiveness of Treatment) was conducted according to the Helsinki Declaration and International Conference on Harmonisation guidelines. There was no remuneration or financial incentive for participation. Individual written informed consent was obtained for study participation, HIV testing, and permission for blood and other samples to be stored and used in future TB studies. The study was approved by the institutional review boards of the U.S. Centers for Disease Control (IRB B), the University Hospitals Case Medical Center IRB (FWA00003937) and the Joint Clinical Research Centre Review Committee (IRB00002647), which is duly constituted under the supervision of the Uganda National Council for Science and Technology (FWA00001293; http://www.uncst.go.ug/) and the Office of the President of Uganda.

The Action TB study was conducted according to the Helsinki Declaration and International Conference on Harmonisation guidelines. There was no remuneration or financial incentive for participation. Individual written informed consent was obtained for study participation and HIV testing. The study "A prospective evaluation of surrogate markers for disease and relapse or reinfection in adult patients with pulmonary tuberculosis" was approved by the Stellenbosch University Health Research Ethics Committee (Ref no 99/039) and Cape Town City Health.

LC-MS analysis of urine samples

The D0, M1 and M6 samples from Stellenbosch University and the D0 and M1 samples from 14 patients of NAA2m study were used as discovery sets. Samples from all time points of the remaining 34 individuals of the NAA2m study were used as a qualification set. Samples used in the discovery phase (Figure 1) were sterilized by γ-irradiation. Those used in the down selection and qualification phase (Figure 1) were not γ-irradiated and additional biosafety measures were taken in handling these. The creatinine concentration of each sample was measured by an alkaline picrate based colorimetric assay (Oxford Biomedical Research, Oxford, MI). LC-MS analyses of urine samples were performed following the methods described in Mahapatra et al. [22]. Positive-ion MS data in both centroid and profile modes were collected using the Agilent MassHunter Data Acquisition software. Data for the discovery samples sets were obtained in the 4 GHz high-resolution modes and that used in down selection and qualification was obtained in the 2 GHz extended dynamic range mode.

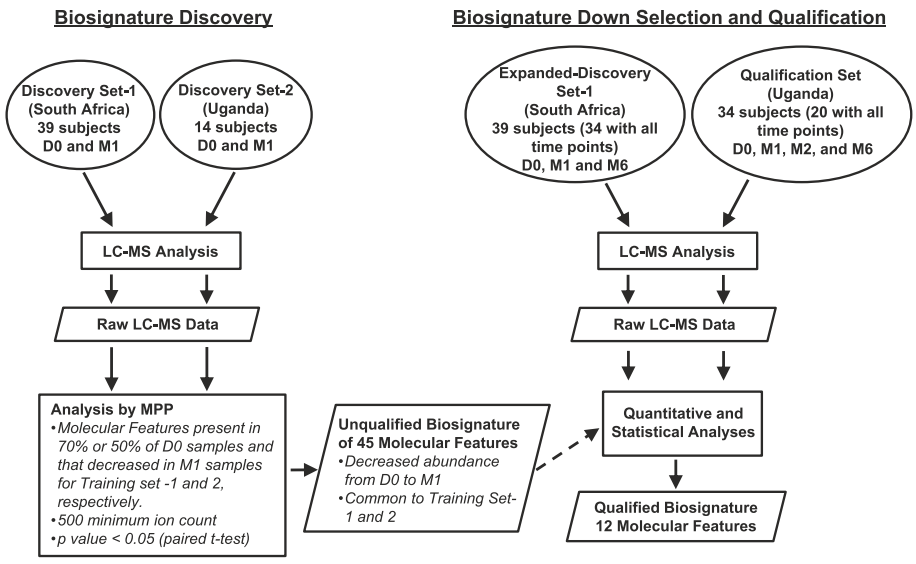

Figure 1 Work flow for the development and qualification of a urine metabolite biosignature of TB treatment response. Biosignature discovery (left) utilized D0 and M1 samples from two distinct geographical locations. This yielded 45 MFs that decreased in abundance between D0 and M1. Biosignature down selection and qualification (right) also used samples from two geographically distinct locations and expanded the analyses to additional time points after the start of anti-TB therapy.

Data analyses

LC-MS data from the discovery and qualification sample sets were processed with the Molecular Feature Extractor (MFE) algorithm tool of the Agilent MassHunter Qualitative Analysis software version B.04.00. The parameters used for MFE were minimum 500 counts, ion species H^+, charge state maximum 1, compound ion count threshold 2 or more ions, and all other parameters were default settings. Comparative analyses of the MFE data files from the training sample sets were performed with the Agilent Mass Profiler Pro (MPP) software version B. 02.01. The molecular features (MFs) were aligned with a 0.4 min retention time and 15 ppm mass tolerance. Aligned MFs were filtered based on their presence in 50%, 60%, or 70% of samples in at least one time point group. The relative abundance of MFs obtained from different treatment time points were compared and features that varied significantly by at least 2-fold based on pairwise t-test or ANOVA ($p < 0.05$) were selected. The MFs that decreased significantly in abundance between D0 and M1 for discovery set-1 and -2 (Figure 1) were selected by *k*-means clustering analysis. The abundances of selected MFs (area under the peak for the monoisotopic mass) in the extended discovery and qualification data sets were determined using the Agilent MassHunter quantitative analysis software and exported as a Microsoft Excel spreadsheet for statistical analyses.

Statistical analyses

Statistical analysis was performed using SAS version 9.2. All feature intensity (arbitrary numbers) values were transformed to the log2 scale prior to statistical analysis. To prevent undefined values, a value of one was added to all values prior to transformation. Separately for each MF of the Stellenbosch University and NAA2m data sets, a repeated measures analysis was performed using SAS proc mixed. The response variable was log2 transformed intensity (abundance) value from the Agilent MassHunter quantitative analysis software. The model included a fixed effect for time and a random subject effect to account for repeated measures. Comparisons across time points were based on contrasts of the model. After identifying a subset of features with consistent changes across time and sample sets, logistic regression was used to model treatment status (D0 versus M1).

Identification of the metabolites

The chemical formulas of the MFs were predicted from the accurate mass data using the molecular formula generator tool of Agilent MassHunter Qualitative Analysis software. These molecular formulas were searched against the Metlin compound database (Metlin_AMRT_PCDL.cdb) using the ID browser application of MPP. The same molecular formulas were also searched against publicly available Human Metabolome Database (HMDB, http://hmdb.ca) [23].

Results

Clinical sample evaluation

To allow development and assessment of a robust biosignature, urine samples from adult pulmonary cavitary TB patients of both sexes without HIV co-infection, collected at the initial time of TB diagnosis (D0) (i.e. before

start of therapy), and at different treatment time points during the course of therapy were divided into discovery and qualification sets (Figure 1). To minimize the impact of confounding variables (i.e. normal flora, diet, racial ethnicity, age, and sex), samples from two geographical regions (Uganda, 48 patients; and South Africa, 39 patients) were analyzed and the data normalized to urinary creatinine levels [24].

We initially applied LC-MS analyses and data processing as described in Materials and Methods to an expanded-discovery set of urine samples from South Africa (D0, M1 and M6) to assess the separability of these time-grouped samples based on metabolic signatures. Specifically, an ANOVA was performed to identify relevant MFs (ions with defined accurate mass and retention time) present in at least 60% of the samples for any given treatment time point and that differed in abundance between time points by at least 2 fold (log2 data). An unsupervised principal component analysis (PCA) demonstrated that D0 samples separated from those of the M1 and M6 treatment time points (Figure 2). However, the data also revealed a separation between the M1 and M6 samples. This indicated that metabolic profiling can be utilized to assess anti-TB treatment, but that the selection of a TB-ETRB based on samples from three time points would be difficult to deconvolute.

Biosignature discovery

To allow selection of a TB-ETRB, pairwise analyses of LC-MS data of D0 and M1 samples from Discovery Sets-1 and -2 were performed (Figure 1). This analysis was performed with the Agilent MPP software that provides a method to filter through large data sets for MFs that differ between groups (i.e. D0 and M1) based on the MF abundance as measured by peak height and normalized to the mean of all samples, as well as user provided parameters. This resulted in the identification of 822 and 2644 MFs, respectively, for Discovery Sets-1 and -2 that either decreased or increased in abundance (2 fold) following the onset of treatment. PCA based on the selected MFs demonstrated discrete clustering of the D0 and M1 samples of both discovery sets (Additional file 1). As a means of validating the overall approach, the predicted metabolites of anti-TB drugs [25-29] were identified based on accurate mass measurements. These compounds were included in the group of MFs that increased in abundance between the D0 and M1 samples (data not shown). To avoid biasing a TB-ETRB by inclusion of known or unknown drug metabolites, subsequent analyses were limited to MFs that decreased in abundance following the onset of treatment (Figure 3). These data demonstrated that the separation of D0 and M1 samples was more pronounced for Discovery Set-2 versus Discovery Set-1. Further, the inclusion of MFs that changed abundance in either direction (Additional file 1) versus those that only decreased in abundance (Figure 3) did not alter this trend between discovery sets. From the 183 and 659 MFs that decreased in abundance between D0 and M1 for Discovery Sets-1 and -2, respectively, 26 differentiating MFs were consistent between the two data sets. An additional 19 MFs were included for their high significance in a single discovery set to give the initial discovery-phase TB-ETRB of 45 MFs (Figure 1).

Biosignature down selection and qualification

Application of a MS defined metabolic biosignature to assess treatment response during clinical trials would be facilitated by the evaluation of LC-MS data in a manner that was independent of a specific software platform. With this in mind further downs selection and qualification of the 45-MF TB-ETRB was performed on quantitative data (area under the peak for the selected MFs) directly extracted from LC-MS files of urine specimens in Expanded Discovery Set-1 and the Qualification Set (Figure 1). It is noted that the Expanded Discovery Set-1 includes the D0 and M1 samples used in the initial TB-ETRB discovery phase, whereas the Qualification Set samples were analyzed for the first time during the down selection and qualification phase. The inclusion of the Expanded Discovery Set-1 was justified in these later evaluation steps since the more robust quantitative measure of MF peak area was used with the Agilent Quantitative Analysis

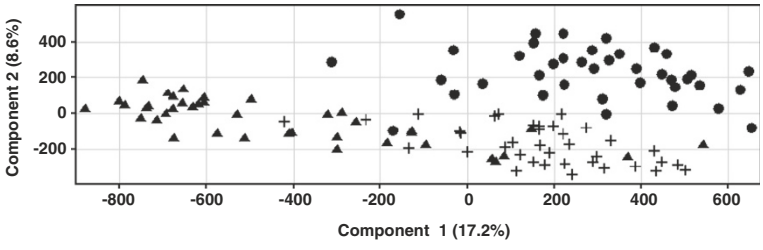

Figure 2 Unsupervised PCA of D0 (triangle), M1 (circle), and M6 (cross) samples from South African patients. The PCA was constructed based on MFs that were present in at least 60% of the samples for any given time point and differed in abundance between time points by at least 2 fold with a $p < 0.05$.

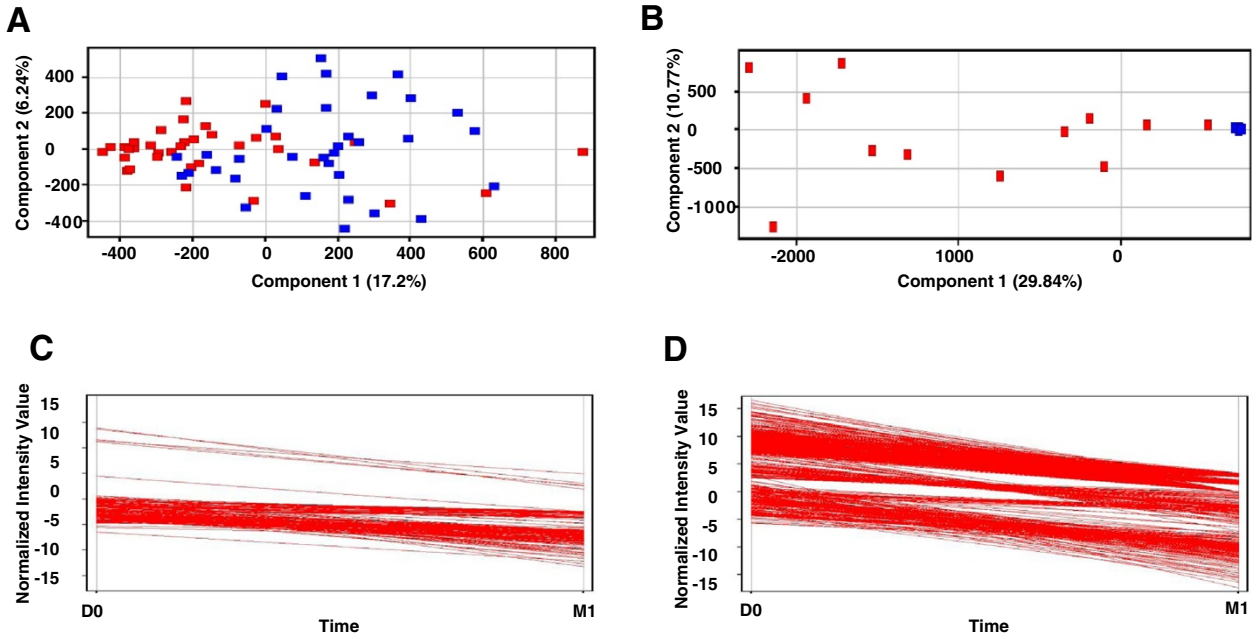

Figure 3 Unsupervised PCA of D0 (triangle) and M1 (circle) patient samples, and MFs demonstrating a decrease in abundance after the start of anti-TB treatment. The PCAs were constructed based on MFs present in at least 70% and 50% of the samples for any given time point of Discovery Set-1 **(A)** and Discovery Set-2 **(B)**, respectively, and that decreased in abundance between D0 and M1 by at least 2 fold with a $p < 0.05$. A *k*-means clustering analysis was performed to select MFs that decreased in abundance between D0 and M1 of Discovery Set-1 **(C)** and Discovery Set-2 **(D)**.

software. This was in contrast to the discovery phase that was based on the high-throughput filtering power of the MPP software with MF abundance based on peak height.

The quantitative data of each MF were evaluated for change in abundance across the D0, M1, and M6 time points for Expanded Discovery Set-1; and D0, M1, M2, and M6 samples for the Qualification Set. Ten of the 45 MFs failed to yield extractable quantitative data from the LC-MS files of the Qualification Set. Thus, the biosignature that could be further down selected based on robustness across multiple data sets was comprised of 35 MFs. The molecular formulas of these 35 MFs were predicted based on their accurate masses and interrogated against the Human Metabolome and METLIN databases to provide putative structure identification of the MFs (Additional file 2).

Hierarchical clustering based on the abundance of the 35 individual MFs across all time points of the Expanded Discovery Set-1 and Qualification Set was performed. This demonstrated nine clusters of MFs based on their kinetics of change in abundance, and two MF (174.0636, and 311.1239) with abundance changes that excluded them from any of the clusters (Figure 4). The MFs forming clusters C1, C3, C6 and C9 behaved inconsistently between the two sample sets, with the most relevant inconsistency occurring in abundance change from D0 to M1. A total of 23 MFs that typically displayed consistent abundance change trends across treatment time points of both the Expanded Discovery Set-1 and Qualification Set were encompassed by clusters C2, C4, C5, C7 and C8. As expected the MFs in these five clusters decreased in abundance between the D0 and M1 time points, and either continued to show decreased abundance at later time points or slightly increased in abundance after the initial D0 to M1 decrease.

The above analyses were performed with values averaged across patient samples of each data set. To further assess robustness, each MF was evaluated on a patient-by-patient basis by calculating the percentage of patients demonstrating a decrease in abundance between the D0 and later time points. The Expanded Discovery Set-1 demonstrated that 91% to 35% of the patients yielded a decreased abundance of individual MFs between D0 and M1 (Figure 5 and Additional file 3). A similar range of these same patients (94% to 24%) revealed a decrease in the abundance of individual MFs between D0 and M6. Likewise, the Qualification Set displayed a decrease in MF abundance between D0 and M1, or D0 and M6 that ranged between 95% and 20%, or 95% and 40% of the patients, respectively (Figure 5 and Additional file 3). A total of 17 MFs decreased in abundance between D0 and other time points in 70% or greater of the patients in at least one of the sample sets, and five of these 17 MFs decreased in abundance in at least 70% of the patients from both

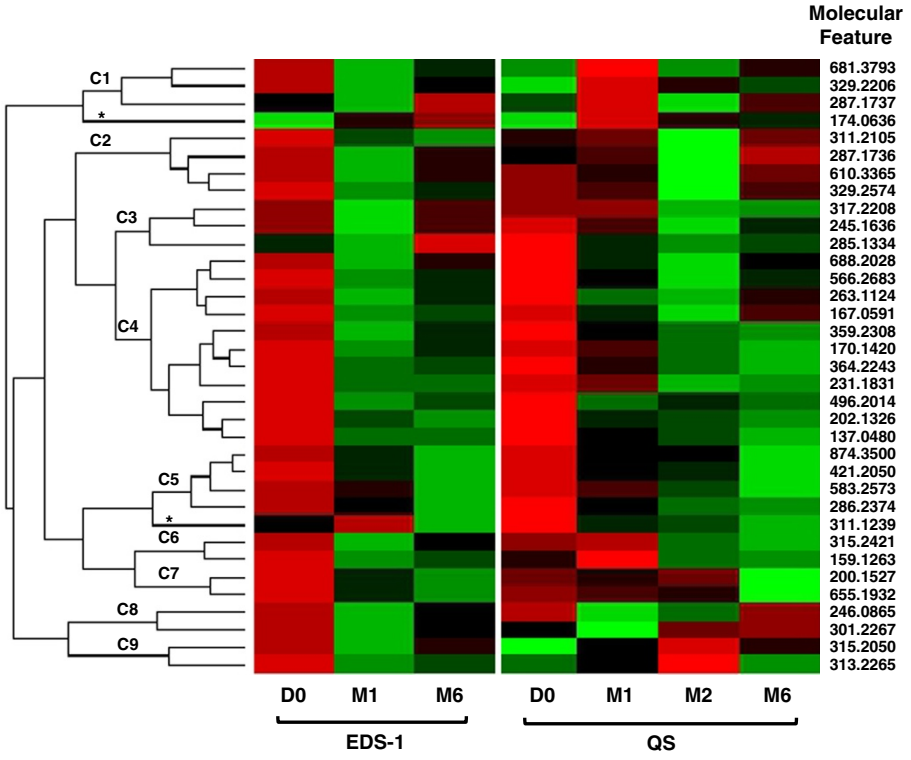

Figure 4 Heat map of 35 MFs showing the change in abundance across multiple time point. A cluster analysis was performed on log2 transformed averaged abundances of individual MFs at each time point for samples of Expanded Discovery Set-1 (EDS-1) and the Qualification Set (QS). MFs are displayed in rows and noted by the monoisotopic mass of the MF. Columns display time points. Visual analysis of the heat map revealed four clusters (C1, C3, C6, C9) of MFs that behaved dissimilarly between qualification sets; five clusters (C2, C4, C5, C7, C8) of MFs that behaved similarly between qualification sets; and two MFs (*) that did not fit into a specific cluster because of the low abundance at D0. Red indicates high abundance, green indicates low abundance.

sample sets. Three additional MFs from this group of 17 decreased in abundance between D0 and other time points in at least 60% of patients of both sample sets (Figure 5 and Additional file 3). All but three of the 17 MFs with the most consistent patient-to-patient abundance change were encompassed in cluster C2, C4, C5, C7 or C8 (Figure 4), further indicating that MFs in these clusters likely provide the most robust biosignature of treatment response. It was interesting to note that a large percentage of patients (at least 60%) from the Expanded Discovery Set-1 presented an increased abundance between D0 and other time points for MF 174.0636 and the majority of patients from the Qualification Set also showed this MF increasing in abundance over time. Thus, this outlier is truly a MF that tended to increase with treatment although the MPP software initially indicated it decreased in abundance. The MFs of C1, C3, C6 or C9 typically did not yield robust patient-to-patient results or produced dichotomous results between the two sample groups. This is consistent with the hierarchical cluster data (Figure 4) and the conclusion that these MFs would lack value in assessing treatment response.

To define a distinguishing TB-ETRB that could be applied at later anti-TB treatment time-points as well as M1, a repeated measures analysis was performed on the extracted quantitative data of the 35 MFs at all time points of the Expanded Discovery Set-1 and Qualification Set. This revealed 18 MFs with consistent and significant abundance changes between D0 and M1, and D0 and M6 for the Expanded Discovery Set-1 (Additional file 4). Of these 18 statistically significant MFs, 12 also showed consistent and significant changes in abundance between D0 and M1, and D0 and M2 or M6 of the Qualification Set (Table 1 and Figure 6). As expected, the repeated measures analyses showed that all 12 of the MFs except 174.0636 decreased in abundance between D0 and later time points. To further assess this 12 MF TB-ETRB, the D0 and M1 data of all 73 subjects used in the down selection and qualification phase were applied to a logistic regression model fitted to predict time status (D0 versus M1). Prior to model fitting, MF abundance values were log transformed and standardized to have mean zero and standard deviation of one for each MF and data set separately. Best subsets selection was used to identify

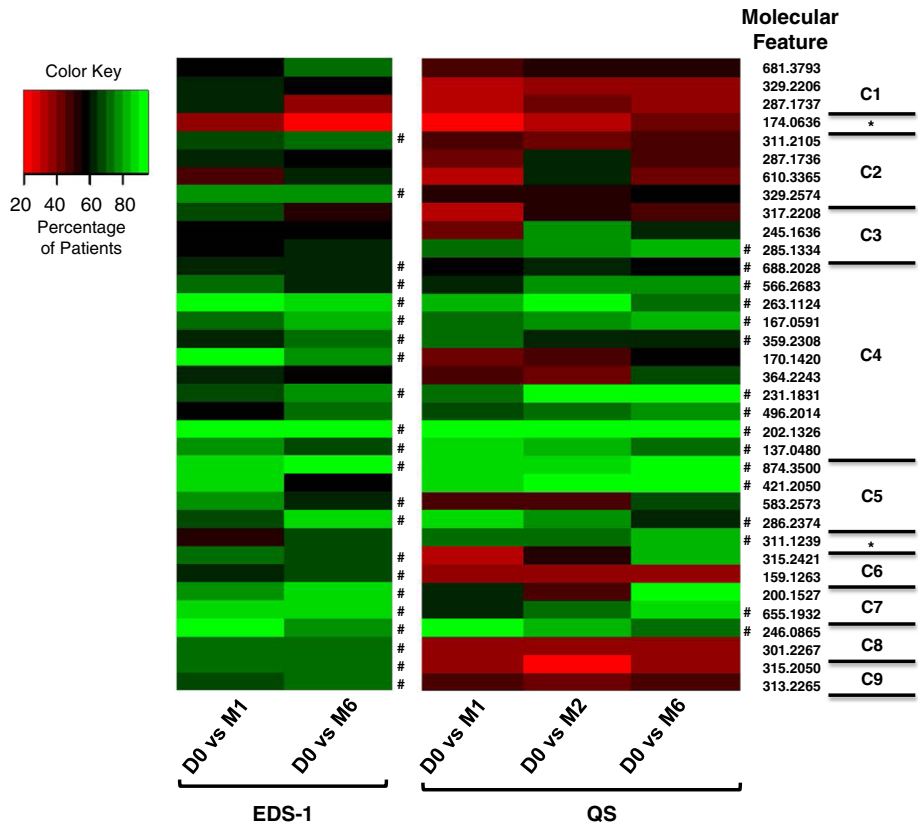

Figure 5 Heat map based on the percentage of patients that demonstrated a decreased abundance in each of the 35 MFs between D0 and other time points. The MFs are ordered based on the clusters identified in Figure 3, and the MF clusters are noted to the right of the individual MFs that are designated by their monoisotopic mass. The Expanded Discovery Set-1 (EDS-1) and the Qualification Set (QS) were evaluated separately for the percent of patients demonstrating a decrease in MF abundance. # indicates MFs that consistently decreased in abundance from D0 in at least 60% of the patients for all time points of a sample group. The exact patient percentages used to derive this graphical representation are provided in Additional file 3.

Table 1 Statistically significant MFs after down selection and qualification

MF[1]	Predicted formula[2]	Alternate formula ±10 ppm[3]	DB Match[4]
137.0484	$C_7H_7NO_2$	None	p-Aminobenzoic acid + 10
167.0591	$C_8H_9NO_3$	None	Pyridoxal/isopyridoxal + 10
174.0636	$C_6H_{10}N_2O_4$	$C_8H_{15}PS$	Formimino-L-glutamic acid + 1
202.1326	$C_9H_{18}N_2O_3$	None	Leu Ala + 8
231.1831	$C_{12}H_{25}NO_3$	None	None
246.0865	$C_9H_{14}N_2O_6$	Multiple	L-alpha-Aspartyl-L-hydroxyproline + 2
263.1124	$C_9H_{17}N_3O_6$	Multiple	Thr Gly Ser + 8
286.2374	$C_{14}H_{30}N_4O_2$	None	N1,N12-Diacetylspermine + 1
421.2051	$C_{19}H_{27}N_5O_6$	Multiple	None
496.2014	$C_{25}H_{28}N_4O_7$	Multiple	None
566.2683	$C_{29}H_{48}N_2OP_2S_2$	Multiple	None
874.3547	$C_{36}H_{65}N_2O_{16}P_3$	Multiple	None

[1]The MF is denoted by its monoisotopic mass.
[2]The chemical formula was predicted based on accurate mass by molecular formula generator algorithm of Agilent MassHunter software.
[3]The alternate chemical formulas are within 10 ppm of the observed monoisotopic mass.
[4]The structure is the product with the closest match to the observed monoisotopic mass. The number denotes the number of structures in METLIN or HMDB databases that share the same molecular formula.

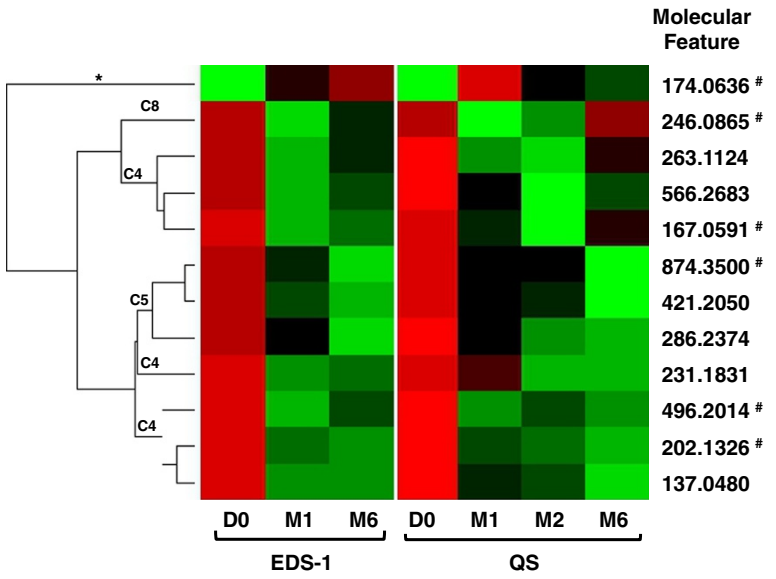

Figure 6 Heat map of 12 MFs determined to have the highest statistical significance in both the Expanded Discovery Set-1 (EDS-1) and Qualification Set (QS). The heat map displays change in abundance across multiple time point (columns) for each MF (rows). The MFs were arranged based on clustering analysis (left) and the associated clusters identified in Figure 4 are noted (C4, C5, C8 and *). #indicates the molecular features identified by logistic regression analysis that classify D0 and M1 samples with an error rate of 11.8%.

the logistic regression model with the smallest AIC value, and resulted in the selection of six MFs (167.0591, 174.0636, 202.1326, 246.0865, 496.2014, and 874.3500) (Figure 6). Evaluation of the model based on leave-one-out cross-validation revealed an error rate of 11.8%. It is noted that since all of the TB patients evaluated in this study responded to treatment, this model has not been tested for prediction of successful treatment response. Nevertheless, this analysis provides proof-of-concept that a metabolite-based biosignature can be established to monitor anti-TB treatment as early as one month.

Discussion

The present studies elucidated a small molecule metabolic biosignature that differentiates TB patients prior to the start of therapy from those successfully responding to treatment early (one month) after the start of therapy. Additionally, we demonstrated that a TB-ETRB can be developed using a non-invasive clinical specimen such as urine. Efforts to define diagnostic or prognostic biomarkers of TB have largely focused on microbiological and immunological measurements or identification of bacterial macromolecules [30-35]. Nevertheless, small molecule metabolites or metabolic processes such as adenosine deaminase activity in pleural fluid or hydroxyproline in urine are proposed as diagnostic markers of TB [36,37]. Moreover, several reports have evaluated the use of volatile organics in breath condensates as metabolic markers of TB [38-40]. Weiner et al. recently performed metabolomic analyses on sera from active and latent TB patients, as well as healthy controls to identify 20 small molecules as a diagnostic signature of TB [41]. Thus, as with non-infectious human diseases, [11-13] a growing body of evidence supports the use of small molecule metabolites as diagnostic or prognostic markers of TB.

Advanced computational tools greatly enhance the field of metabolomics and allow for the development of metabolic biosignatures [7]. Our metabolomic based analyses of urine allowed the separation of TB patients' samples based on treatment time points. However, it was noted that the profile of D0 versus M1 distinguishing metabolites likely differed from those allowing recognition of D0 versus M6 samples. Thus, a pairwise comparison of D0 and M1 samples and the subsequent filtering of the differentiating MFs such that they could be applied to other time points during therapy were adopted for the biosignature discovery phase. The first filtering step ensured patients from two distinct geographical regions shared the D0 versus M1 MFs. A second filtering step was used to remove potential bias introduced by drug metabolites and thus, the biosignature was limited to MFs that decreased in abundance after the start of therapy. This resulted in intermediate list of 45 MFs. The exclusion of MF that increased in abundance between D0 and M1 was a conservative approach for this proof-of-concept study. However, as we recently demonstrated with the discovery of an INH-NAD adduct in urine of TB patients undergoing treatment (22), there was a possibility that unknown drug metabolites might have biased biosignature data. Although this conservative approach can be justified for a proof-of-

concept study, it must be recognized that MFs increasing in abundance with successful anti-TB treatment are likely to be equally valuable in predicting treatment response as long as they are shown not to be direct products of drug metabolism.

The 45 MFs from the discovery phase were down selected with LC-MS data from an expanded set of discovery samples and previously untested patient samples representing a qualification set. Moreover, the down selection was performed on the quantitative data extracted from the raw LC-MS files and was extended to treatment time points beyond M1. This process yielded 23 MFs that correlated well between the sample sets, and over half of these 23 MFs were determined to be statistically significant ($p < 0.05$) by a univariant analysis. Multivariant statistical modeling further selected six MFs that yielded an 11.8% error rate in classifying D0 and M1 samples. Although down selection based on consistency of data across sample sets and statistical analyses define the most robust MFs to include in a TB-ETRB, it is also important to understand why some of the 45 MFs from the discovery phase behaved inconsistently in evaluation of data for down selection and qualification. Firstly, not all patients will be at the same state of disease clearance at a specific treatment time point, thus the kinetics of abundance change for individual metabolites could vary from patient-to-patient during the course of treatment. Another possibility is the difference in handling of samples between the discovery and down selection phase . To assess samples in a manner that more closely resembles what might occur during a clinical trial, the Qualification Set samples were not subjected to γ-irradiation; a process that was applied to the other samples and resulted in elevated temperatures and potential alteration of some metabolites. Additionally, significant time gaps existed between the analyses of each sample set, and some MFs may have lacked a consistency that would withstand minor technical variability between analyses. While such variability in sample handling could exclude potentially useful MFs, it also allowed for selection of the most robust MFs. A third variable was the data analysis used for discovery versus that used in down selection. This likely explains why one MF (174.0636) selected during the discovery phase for decreased abundance between D0 and M1 was subsequently shown to increase in abundance after the start of treatment when LC-MS data were analyzed for down selection and qualification. The MPP software analyzes MFs that are selected based on ^{12}C monoisotopic mass to ^{13}C isotopic mass ratios; a process vulnerable to technical issues such as detector saturation [9,42]. Analyses performed with the LC-MS data for down selection, however, only considered the abundance (peak area) of the monoisotopic mass of each MF and data that was collected in an extended dynamic range mode; thus, eliminating detector saturation artifacts. This observation underscores the need and utility of identifying and correcting misclassifications via qualification studies and the use of unmodified quantitative MS data.

Bacteriological analysis of sputa is the most widely accepted method for the diagnosis and assessing treatment response of pulmonary TB [1]. In fact sputum culture results at two months of treatment are applied in Phase II clinical trials as an early indicator of drug efficacy. Analysis of individual molecules in sputa, however, can be technically challenging owing to viscosity and complexity of the matrix, as well as the absence of markers against which analyte abundance can be normalized. In contrast urine allows for facile recovery of analytes and can be normalized based on creatinine concentrations [24]. Metabolic profiles of individual urine samples will vary due to differences in diet, normal flora and other lifestyle related factors. Racial origin, age and sex will also influence metabolite profiles in sample subpopulations [43]. The use of samples from two geographical regions and the implementation of stringent biosignature selection criteria allowed for the identification of multiple MFs in the urine that changed consistently with early treatment response in both geographical regions. The data and approach presented provide proof-of-principle that a small-molecule, urine based TB-ETRB could be applied as a tool in clinical trials of new TB drugs and regimens.

This study also demonstrated that a metabolic flux occurs early during the successful treatment of TB patients. Thus, the metabolites identified based on abundance change over the course of treatment are likely indicators of diseased individuals returning to a normal or healthy state. This is underscored by the observation that the TB-ETRB behaved similarly at M1 and later time points of treatment, including M6. Additional studies, however, are needed to validate these biomarkers in patients with TB who are treated and followed for longer term unfavorable clinical outcomes such as treatment failure, recurrent TB or death, and for which microbiological data can be correlated with metabolic flux. Future analyses should also include patients from other geographical regions and patients co-infected with HIV as well as variations in infecting *Mtb* strains [44,45]. Likewise, fully elucidating the structural identity and biological significance of each MF can enhance the utility of a biosignature. Further studies are required to confirm the putative chemical identities assigned to the MFs of the TB-ETRB. This can be a protracted process [22,46,47], and one that is currently being pursued.

Conclusions

Tuberculosis (TB) continues to have a major impact on global health, and the continued emergence of multiple drug resistant TB has enhanced the need for new anti-TB drugs. The need to accelerate anti-tuberculosis drug

development and the study of new compounds in clinical trials is hindered by a lack of biomarkers that serve as indictors of clinical outcome or response to treatment. The current measure of treatment response is dependent on sputum culture methodology that can take weeks to obtain results and for which a clinical specimen becomes more difficult to collect as patients respond favorably to treatment. LC-MS analyses of aliquots of human urine normalized to creatinine concentrations allowed for the elucidation of a qualified small molecule (< 1,000 Da) metabolite biosignature comprised of 23 MFs that consistently changed in abundance over the course of treatment in these individuals; with 12 of these MFs achieving statistical significance. These studies and data demonstrated for the first time the feasibility of using small molecule biosignatures to monitor the response of individuals to anti-TB therapy. The analysis of clinical samples by LC-MS is rapid in comparison to the currently used culture methods for monitoring of anti-TB therapy and the qualified biosignature developed had a level of robustness that allowed for it to be applied in separate geographical settings. Moreover, this methodology was applied to urine; a clinical sample that requires little processing and is collected in a non-invasive manner.

Additional files

Additional file 1: Unsupervised PCA of D0 (triangle) and M1 (circle) samples. The PCAs were constructed based on MFs present in at least 70% and 50% of the samples for any given time point of Discovery Set-1 (A) and Discovery Set-2 (B), respectively, and that differed in abundance between time points by at least 2 fold with a p<0.05.

Additional file 2: MFs of Biosignature Used for Qualification.

Additional file 3: Per Patient Evaluation of MF Robustness.

Additional file 4: Results of Repeated Measures Analysis.

Abbreviations

HMDB: Human metabolome database; LC-MS: Liquid chromatography-mass spectrometry; MPP: Mass profiler pro; MS: Mass spectrometry, MFs, Molecular features; MFE: Molecular feature extractor; M1: Month-1; M2: Month-2; M6: Month-6; *Mtb*: *Mycobacterium tuberculosis*; PCA: Principal component analysis; D0: Time of TB diagnosis; TB: Tuberculosis; TB-ETRB: TB-early treatment response biosignature; TBRU: Tuberculosis Research unit.

Competing interests

AMH, GK, GW, KDE, JLJ, MAD, MLJ, PG, and WHB have no competing interests. JTB and SM have filed an invention disclosure with Colorado State University covering the biomarkers described in this manuscript and Colorado State University is the employer of JTB and SM and will decide whether or not to file a patent on the biomarkers described.

Authors' contribution

SM performed the mass spectrometry; SM, AMH, MAD and JTB performed statistical and or data analyses; JIJ, GW, KDE, PG, MLJ and WHB provided coordination, collection and distribution of clinical samples and the associated data; SM, JTB, WHB, GK, KDE, JLJ, and GW all participated in and significantly contributed to the overall study design; JTB and SM were responsible for directing the science and writing of the manuscript; MAD, AH, GK, WHB, JLJ, and GW contributed significantly in the editing of the final manuscript. All authors read and approved the final manuscript.

Acknowledgements

The clinical samples from Stellenbosch University were part of the GSK-funded Action TB Surrogate Marker project, with co-investigators Prof. Nulda Beyers, Prof. Paul van Helden and Prof. Gerhard Walzl. The authors thank Mr. Brian Cranmer for technical help and comments. This project was funded by the Tuberculosis Research Unit (TBRU), established with Federal funds from the United States National Institutes of Allergy and Infectious Diseases & the United States National Institutes of Health and Human Services [Contract No. HHSN266200700022C/NO1-AI-70022]. United States Food and Drug Administration [Grant No. 1U18FD004038-01]; and the U.S. Centers for Disease Control and Prevention. Tuberculosis Trials Consortium (TBTC) [contract number 200-2009-32598].

Author details

[1]Mycobacteria Research Laboratories, Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, CO 80523, USA. [2]Department of Medicine, Tuberculosis Research Unit (TBRU), Case Western Reserve University, Cleveland, OH 44106, USA. [3]Uganda-Case Western Reserve University Research Collaboration, Kampala, Uganda. [4]Department of Pathology, University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA. [5]Department of Medical Microbiology, University School of Biomedical Sciences, College of Health Sciences, Makerere University, Kampala, Uganda. [6]Public Health Research Institute Center, University of Medicine and Dentistry of NJ, Newark, New Jersey 07103, USA. [7]DST/NRF Centre of Excellence for Biomedical Tuberculosis Research and MRC Centre for Molecular and Cellular Biology, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, PO Box 19063, Francie van Zijl Drive, Tygerberg 7505, South Africa.

Received: 29 July 2013 Accepted: 27 January 2014
Published: 31 January 2014

References

1. Wallis RS, Kim P, Cole S, Hanna D, Andrade BB, Maeurer M, Schito M, Zumla A: **Tuberculosis biomarkers discovery: developments, needs, and challenges.** *Lancet Infect Dis* 2013, **13**(4):362–372.
2. Wallis RS, Pai M, Menzies D, Doherty TM, Walzl G, Perkins MD, Zumla A: **Biomarkers and diagnostics for tuberculosis: progress, needs, and translation into practice.** *Lancet* 2010, **375**(9729):1920–1937.
3. Perrin FM, Lipman MC, McHugh TD, Gillespie SH: **Biomarkers of treatment response in clinical trials of novel antituberculosis agents.** *Lancet Infect Dis* 2007, **7**(7):481–490.
4. Nahid P, Saukkonen J, Mac Kenzie WR, Johnson JL, Phillips PP, Andersen J, Bliven-Sizemore E, Belisle JT, Boom WH, Luetkemeyer A, *et al*: **CDC/NIH workshop. Tuberculosis biomarker and surrogate endpoint research roadmap.** *Am J Respir Crit Care Med* 2011, **184**(8):972–979.
5. Walzl G, Ronacher K, Djoba Siawaya JF, Dockrell HM: **Biomarkers for TB treatment response: challenges and future strategies.** *J Infect* 2008, **57**(2):103–109.
6. Beisel WR: **Metabolic response to infection.** *Annu Rev Med* 1975, **26:**9–20.
7. Zhang A, Sun H, Wang P, Han Y, Wang X: **Modern analytical techniques in metabolomics analysis.** *Analyst* 2012, **137**(2):293–300.
8. Vinayavekhin N, Homan EA, Saghatelian A: **Exploring disease through metabolomics.** *ACS Chem Biol* 2010, **5**(1):91–103.
9. Theodoridis GA, Gika HG, Want EJ, Wilson ID: **Liquid chromatography-mass spectrometry based global metabolite profiling: a review.** *Anal Chim Acta* 2012, **711:**7–16.
10. Catchpole G, Platzer A, Weikert C, Kempkensteffen C, Johannsen M, Krause H, Jung K, Miller K, Willmitzer L, Selbig J, *et al*: **Metabolic profiling reveals key metabolic features of renal cell carcinoma.** *J Cell Mol Med* 2011, **15**(1):109–118.
11. Kim K, Aronov P, Zakharkin SO, Anderson D, Perroud B, Thompson IM, Weiss RH: **Urine metabolomics analysis for kidney cancer detection and biomarker discovery.** *Mol Cell Proteomics* 2009, **8**(3):558–570.
12. 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.
13. Rhee EP, Gerszten RE: **Metabolomics and cardiovascular biomarker discovery.** *Clin Chem* 2012, **58**(1):139–147.
14. Richeldi L: **An update on the diagnosis of tuberculosis infection.** *Am J Respir Crit Care Med* 2006, **174**(7):736–742.

15. McNerney R, Maeurer M, Abubakar I, Marais B, Mchugh TD, Ford N, Weyer K, Lawn S, Grobusch MP, Memish Z, *et al*: **Tuberculosis diagnostics and biomarkers: needs, challenges, recent advances, and opportunities.** *J Infect Dis* 2012, **205:**S147–S158.
16. Zhang H, Il'yasova D, Sztaray J, Young SP, Wang F, Millington DS: **Quantification of the oxidative damage biomarker 2,3-dinor-8-isoprostaglandin-F(2alpha) in human urine using liquid chromatography-tandem mass spectrometry.** *Anal Biochem* 2010, **399**(2):302–304.
17. Bain MD, Jones M, Borriello SP, Reed PJ, Tracey BM, Chalmers RA, Stacey TE: **Contribution of gut bacterial metabolism to human metabolic disease.** *Lancet* 1988, **1**(8594):1078–1079.
18. Nicholls AW, Mortishire-Smith RJ, Nicholson JK: **NMR spectroscopic-based metabonomic studies of urinary metabolite variation in acclimatizing germ-free rats.** *Chem Res Toxicol* 2003, **16**(11):1395–1404.
19. Dettmer K, Aronov PA, Hammock BD: **Mass spectrometry-based metabolomics.** *Mass Spectrom Rev* 2007, **26**(1):51–78.
20. Bollard ME, Stanley EG, Lindon JC, Nicholson JK, Holmes E: **NMR-based metabonomic approaches for evaluating physiological influences on biofluid composition.** *NMR Biomed* 2005, **18**(3):143–162.
21. Hesseling AC, Walzl G, Enarson DA, Carroll NM, Duncan K, Lukey PT, Lombard C, Donald PR, Lawrence KA, Gie RP, *et al*: **Baseline sputum time to detection predicts month two culture conversion and relapse in non-HIV-infected patients.** *Int J Tuberc Lung Dis* 2010, **14**(5):560–570.
22. Mahapatra S, Woolhiser LK, Lenaerts AJ, Johnson JL, Eisenach KD, Joloba ML, Boom WH, Belisle JT: **A novel metabolite of antituberculosis therapy demonstrates host activation of isoniazid and formation of the isoniazid-NAD + adduct.** *Antimicrob Agents Chemother* 2012, **56**(1):28–35.
23. Wishart DS, Knox C, Guo AC, Eisner R, Young N, Gautam B, Hau DD, Psychogios N, Dong E, Bouatra S, *et al*: **HMDB: a knowledgebase for the human metabolome.** *Nucleic Acids Res* 2009, **37**(Database issue):D603–D610.
24. Cone EJ, Caplan YH, Moser F, Robert T, Shelby MK, Black DL: **Normalization of urinary drug concentrations with specific gravity and creatinine.** *J Anal Toxicol* 2009, **33**(1):1–7.
25. Preziosi P: **Isoniazid: metabolic aspects and toxicological correlates.** *CurrDrug Metab* 2007, **8**(8):839–851.
26. Prasad B, Singh S: **In vitro and in vivo investigation of metabolic fate of rifampicin using an optimized sample preparation approach and modern tools of liquid chromatography-mass spectrometry.** *J Pharm Biomed Anal* 2009, **50**(3):475–490.
27. Yamamoto T, Moriwaki Y, Takahashi S, Hada T, Higashino K: **Study of the metabolism of pyrazinamide using a high-performance liquid chromatographic analysis of urine samples.** *Anal Biochem* 1987, **160**(2):346–349.
28. Zhu M, Burman WJ, Starke JR, Stambaugh JJ, Steiner P, Bulpitt AE, Ashkin D, Auclair B, Berning SE, Jelliffe RW, *et al*: **Pharmacokinetics of ethambutol in children and adults with tuberculosis.** *Int J Tuberc Lung Dis* 2004, **8**(11):1360–1367.
29. Sokolova GB, Ziia AV, Abramovich AG, Bessarabova TN, Ivleva A: **Ethambutol pharmacokinetics in pulmonary tuberculosis patients.** *Farmakol Toksikol* 1986, **49**(5):40–42.
30. Ehtesham NZ, Nasiruddin M, Alvi A, Kumar BK, Ahmed N, Peri S, Murthy KJ, Hasnain SE: **Treatment end point determinants for pulmonary tuberculosis: human resistin as a surrogate biomarker.** *Tuberculosis (Edinb)* 2011, **91**(4):293–299.
31. Walzl G, Ronacher K, Hanekom W, Scriba TJ, Zumla A: **Immunological biomarkers of tuberculosis.** *Nat Rev Immunol* 2011, **11**(5):343–354.
32. Alsleben N, Ruhwald M, Russmann H, Marx FM, Wahn U, Magdorf K: **Interferon-gamma inducible protein 10 as a biomarker for active tuberculosis and latent tuberculosis infection in children: a case–control study.** *Scand J Infect Dis* 2012, **44**(4):256–262.
33. Riou C, Perez Peixoto B, Roberts L, Ronacher K, Walzl G, Manca C, Rustomjee R, Mthiyane T, Fallows D, Gray CM, *et al*: **Effect of standard tuberculosis treatment on plasma cytokine levels in patients with active pulmonary tuberculosis.** *PLoS One* 2012, **7**(5):e36886.
34. Shui GH, Bendt AK, Jappar IA, Lim HM, Laneelle M, Herve M, Via LE, Chua GH, Bratschi MW, Rahim SZZ, *et al*: **Mycolic acids as diagnostic markers for tuberculosis case detection in humans and drug efficacy in mice.** *Embo Molecular Medicine* 2012, **4**(1):27–37.
35. Du Preez I, Loots DT: **New sputum metabolite markers implicating adaptations of the host to *Mycobacterium tuberculosis*, and vice versa.** *Tuberculosis (Edinb)* 2013, **93**(3):330–337.
36. Garcia-Zamalloa A, Taboada-Gomez J: **Diagnostic accuracy of adenosine deaminase and lymphocyte proportion in pleural fluid for tuberculous pleurisy in different prevalence scenarios.** *PLoS One* 2012, **7**(6):e38729.
37. Mach VJ, Reck R, Ziller R: **Diagnostic value of hydroxyproline excretion in bone tuberculosis.** *Beitr Orthop Traumatol* 1973, **20**(5):235–239.
38. Phillips M, Basa-Dalay V, Blais J, Bothamley G, Chaturvedi A, Modi KD, Pandya M, Natividad MP, Patel U, Ramraje NN, *et al*: **Point-of-care breath test for biomarkers of active pulmonary tuberculosis.** *Tuberculosis (Edinb)* 2012, **92**(4):314–320.
39. Banday KM, Pasikanti KK, Chan EC, Singla R, Rao KV, Chauhan VS, Nanda RK: **Use of urine volatile organic compounds to discriminate tuberculosis patients from healthy subjects.** *Anal Chem* 2011, **83**(14):5526–5534.
40. Syhre M, Manning L, Phuanukoonnon S, Harino P, Chambers ST: **The scent of mycobacterium tuberculosis–part II breath.** *Tuberculosis (Edinb)* 2009, **89**(4):263–266.
41. Weiner J 3rd, Parida SK, Maertzdorf J, Black GF, Repsilber D, Telaar A, Mohney RP, Arndt-Sullivan C, Ganoza CA, Fae KC, *et al*: **Biomarkers of inflammation, immunosuppression and stress with active disease are revealed by metabolomic profiling of tuberculosis patients.** *PLoS One* 2012, **7**(7):e40221.
42. Patti GJ: **Separation strategies for untargeted metabolomics.** *J Sep Sci* 2011, **34**(24):3460–3469.
43. Saude EJ, Adamko D, Rowe BH, Marrie T, Sykes BD: **Variation of metabolites in normal human urine.** *Metabolomics* 2007, **3**(4):439–451.
44. Gagneux S, Small PM: **Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development.** *Lancet Infect Dis* 2007, **7**(5):328–337.
45. Nicol MP, Wilkinson RJ: **The clinical consequences of strain diversity in *Mycobacterium tuberculosis*.** *Trans R Soc Trop Med Hyg* 2008, **102**(10):955–965.
46. Oppermann M, Damoc NE, Crone C, Moehring T, Muenster H, Hornshaw M: **High precision measurement and fragmentation analysis for metabolite identification.** *Methods Mol Biol* 2012, **860:**145–156.
47. Bowen BP, Northen TR: **Dealing with the unknown: metabolomics and metabolite atlases.** *J Am Soc Mass Spectrom* 2010, **21**(9):1471–1476.

doi:10.1186/1471-2334-14-53
Cite this article as: Mahapatra *et al.*: A metabolic biosignature of early response to anti-tuberculosis treatment. *BMC Infectious Diseases* 2014 **14**:53.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central