Across Species Metabonomics: Identification of Common Spectral Changes in Mouse and Hamster Urine, Caused by Parasite Infection

R. Stoyanova[1], S. Du[2], Y. Wang[3], Q. Zhao[4], E. Holmes[3], J. Utzinger[5], P. Sajda[2], T. R. Brown[2,4]

[1]Radiation Oncology, Fox Chase Cancer Center, Philadelphia, PA, United States, [2]Biomedical Engineering, Columbia University, New York, NY, United States, [3]Biological Chemistry, Imperial College, London, United Kingdom, [4]Radiology, Columbia University, New York, NY, United States, [5]Public Health and Epidemiology, Swiss Tropical Institute, Basel, Switzerland

Introduction: The concept of metabonomics is to measure the small molecule concentrations via ^{1}H NMR spectroscopy or mass spectrometry and then apply pattern recognition techniques to identify clusterings in the data, related to the disease status or treatment end-points. The collection and interpretation of metabolic data is becoming ever more crucial as a key component in system biology, where the goal is to globally profile cell, tissue and organism. Common significant challenge for the 'omic' technologies, such as transcriptomics, proteomics and metabonomics is cross-referencing the data derived from different species in order to properly derive inferences regarding expression and disease state. Identification of such common biomarkers across species is instrumental for opening an entire sphere of investigations, where drug efficacy in experimental animals can be directly related to treatment outcome and toxicity in humans.

Here we present a metabonomic investigation of the spectral changes in the urine of two species – mice and hamsters – following parasite infection, where we identify a common spectral pattern associated with the presence of the disease.

Methods: Ten out of a total of twenty female mice (NMRI strain) were infected s.c. with 80 *S. mansoni* cercarie each. Between 0.7 and 1.5 ml urine was collected from each mouse at the 49th and 56th day postinfection. For a detailed description of the utilized host-parasite model, exact protocol of urine collection and animal dissection see [1]. Similarly, eighteen out of total of thirty-six male Syrian SLAC hamsters were infected with 100 *S. japonicum* cercarie and urine collected from infected and uninfected controls approximately 30 days postinfection. Sodium 3-(trimethylsilyl) propionate-2,2,3,3-d_4 (TMSO) was added (0.05%) to all urine samples for a chemical shift reference for 0 ppm. All ^{1}H NMR spectra were recorded on a Bruker DRX 600 NMR spectrometer. A standard 1D pulse sequence with water-suppression was used. 128 averages were collected in 16K complex data-points for each spectrum with a spectral width of 20 ppm. ^{1}H NMR spectra were processed with an in-house developed software tool - *HiRes*. The initial processing steps included FFT, phasing, and globally aligning the spectra in the dataset, using TMSO as a reference peak. Linear baseline correction was also applied. The total integral of the CH_2 peak of the creatinine (4.0 – 4.1 ppm) was used for normalization. The spectral region between 6.9 and 8.2 ppm was extracted for further analysis, resulting in a data-matrix D, where each row contains the frequency data for each sample. The frequency shifts of the individual peaks in D were locally adjusted [2]. Principal Component Analysis (PCA) was applied on the correlation matrix of D and the strengths of the eigenvalues were examined to determine the number (M) of significant Principal Components (PCs) [3]. Constrained Non-Negative Matrix Factorization (cNMF) [4] was applied to D, seeking to recover a representation of the original data as a linear mixture of M constituent spectral patterns (S) and their amplitudes (A), representing the mixing coefficients, corresponding to the magnitudes of each constituent spectral pattern contributing to the individual observed spectra. The two matrices, A and S, were recovered simultaneously such that their product reconstructs D, given the noise in the data. In the case of the metabonomic data, A and S are constrained to be non-negative. Given that the noise in the data can be reasonably modeled as Gaussian, one can formulate the problem as a maximum likelihood estimation. The cNMF algorithm constructs a gradient descent over a negative log-likelihood objective function thus optimizing A and S. The recovered spectral shapes in S were displayed on a ppm scale for metabolite identification. Their amplitudes (A) were divided into four subgroups, one for each cohort of control and infected mice and hamsters. A two-sided t-test for the means was applied to the amplitudes to determine significance of their association with each of the investigated subgroups.

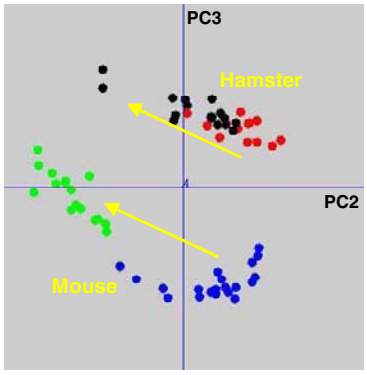

Figure 1. Representation of the aromatic region (6.9 - 8.2 ppm) of ^{1}H NMR urine spectra in a plane defined by the second and third Principal Components. Each point represents a urine spectrum in the dataset from control (blue) and infected (green) mice and control (red) and infected (black) hamsters. Separation between the infected and control animals can be observed, markedly stronger in the mouse data. The parallel yellow arrows indicate that differential metabonomic profiles of control and infected animals are driven in part by common spectral changes in the urine from the two species.

Results: The results of PCA analysis are represented in Figure 1. Firstly, the metabonomic data from mice and hamsters are grouped into two distinct clusters and this separation is most likely related to the phenotypical differences between the two species. Secondly, within each species a division between the infected and control animals can be observed, markedly stronger in the mouse data. The similar direction of the divergence of the infected from the controls animals in this representation indicates that the differential metabonomic profiles of control and infected animals are driven in part by common spectral changes in the urine from the two species. And lastly, it can be concluded that there are a variety of changes, related to the infection only in mice, which drives the larger separation between the two mouse groups.

Examination of the eigenvalues determined that there are 4 significant PCs, explaining more than 98% of the variance in the dataset. In Figure 2A the aromatic region in 600 MHz ^{1}H NMR urine spectrum, averaged over the entire dataset is presented together with some of the resonance assignments. cNMF was applied to the data in this spectral region seeking for M = 4 solutions. The spectral pattern, presented in Figure 2B was significantly associated with the infection in both species (Figure 2B). The pattern is almost entirely related to depletion (~ 40%) of hippurate. The reduction of hippurate excretion may suggest an alteration in phase I and II metabolism and may be also associated with the disturbance of microbial colonies [1]. The second spectral pattern in S was associated only with the urine from the infected mice and contained metabolites, upregulated in this particular group: phenylacetylglycine and p-cresol glucuronide. The elevated excretion of these metabolites can be caused by an increased supply from microbial agents due to the disturbance of the normal microbial ecosystem in the presence of S. *mansoni* worms. The third pattern was strongly associated with the discrimination between the urine of mice and hamsters. Finally the amplitude of the last pattern was randomly distributed among the four groups and it is most likely related to residual baseline variation in the region between 7.2 and 7.5 ppm.

Conclusion: The presented approach has promise for recovering complex metabolic responses across species and should find application to identification of relevant human biomarker combinations for disease diagnosis and monitoring and predicting the beneficial and adverse effects of pharmaceutical compounds.

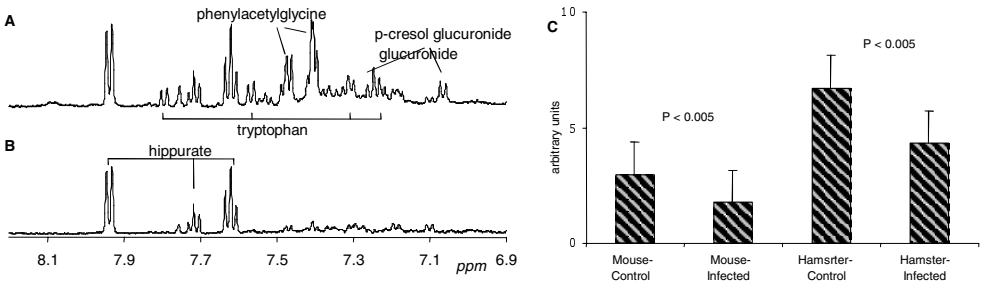

Figure 2. The aromatic region in 600 MHz ^{1}H-NMR urine spectrum, averaged over the entire dataset (A); One of the four spectral pattern recovered using cNMF and associated with common changes between infected and control animals in the two species (B). The bar graph indicates the average strength of the pattern in (B) in each of the four subgroups of data. There is significant depletion (~ 40%) of hippurate in the urine from the infected animals for both species.

Acknowledgements: NIH Metamobolic Roadmap (DK070301) and NSF BES-01-3380

Reference: [1] Wang, Y., Holmes, E., Nicholson, J.K., Cloarec, O., Chollet, J., Tanner, M., Singer, B.H., Utzinger, J., PNAS 101 (34): 12676-86 (2004). [2] Stoyanova, R., Nichols, A.W., Nicholson, J.K., Lindon, J.C., Brown, T.R., J. Magn. Res. 170(2):329-35 (2004). [3] Stoyanova, R., Brown, T.R., NMR in Biomed. 14: 271-7 (2001). [4] Sajda, P., Du, S., Brown, T., Stoyanova, R., Shungu. D., Mao, X., Parra, L., IEEE Transactions on Medical Imaging 23(12): 1453-65 (2004).