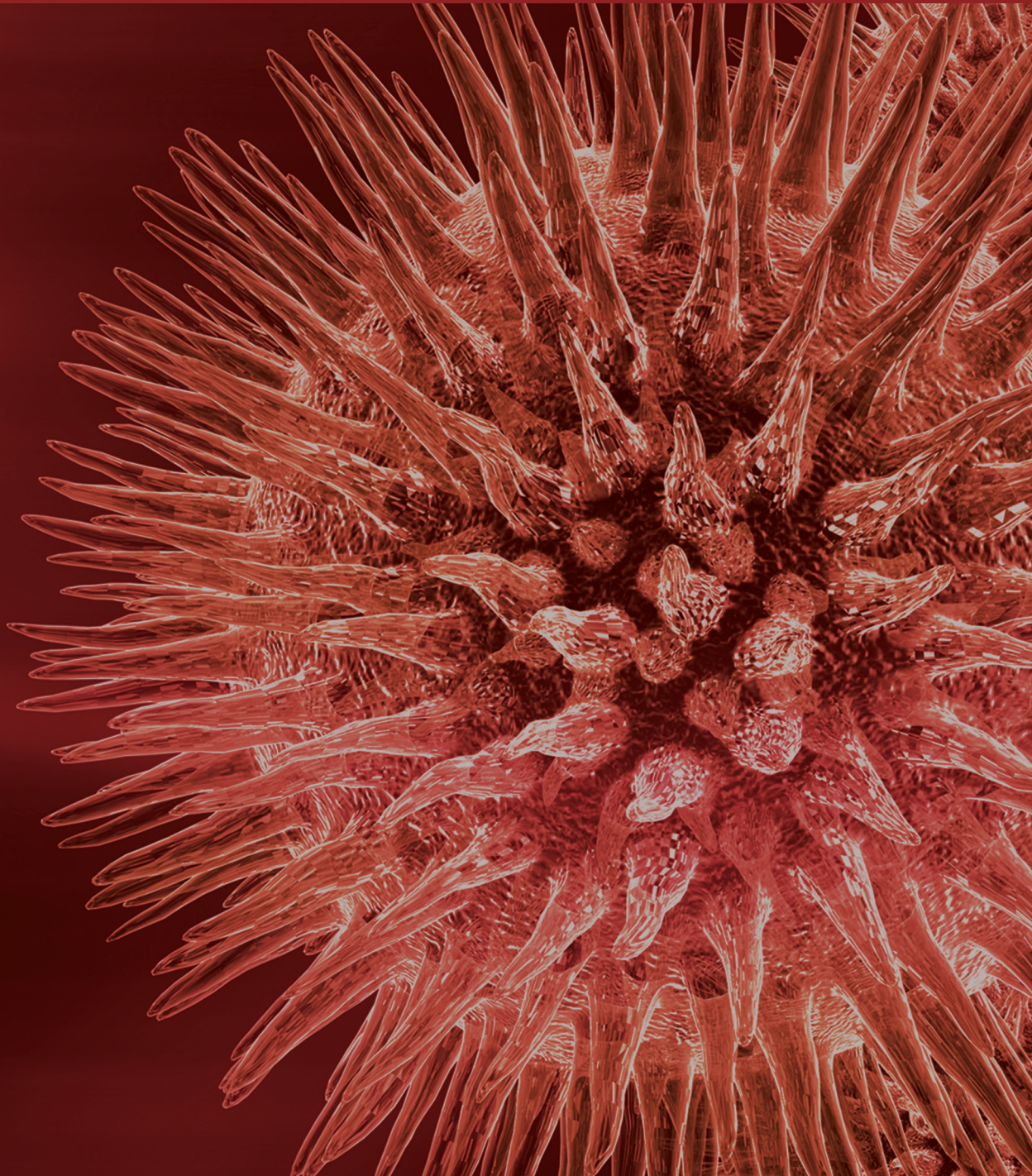


Journal of Biomedicine and Biotechnology

Clinical and Epidemiological Metabonomics

Guest Editors: Mika Ala-Korpela, Veikko Salomaa, and Olav M. Kvalheim





Clinical and Epidemiological Metabonomics

Journal of Biomedicine and Biotechnology

Clinical and Epidemiological Metabonomics

Guest Editors: Mika Ala-Korpela, Veikko Salomaa,
and Olav M. Kvalheim



Copyright © 2011 Hindawi Publishing Corporation. All rights reserved.

This is a special issue published in volume 2011 of "Journal of Biomedicine and Biotechnology." All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Editorial Board

The editorial board of the journal is organized into sections that correspond to the subject areas covered by the journal.

Agricultural Biotechnology

| | | |
|--------------------------------|-----------------------------|--------------------------|
| Guihua H. Bai, USA | Hari B. Krishnan, USA | B. C. Saha, USA |
| Christopher P. Chanway, Canada | Carol A. Mallory-Smith, USA | Mariam B. Sticklen, USA |
| Ravindra N. Chibbar, Canada | Dennis P. Murr, Canada | Chiu-Chung Young, Taiwan |
| Ian Godwin, Australia | Rodomiro Ortiz, Mexico | |

Animal Biotechnology

| | | |
|--------------------------|--------------------------|--------------------------------|
| E. S. Chang, USA | Thomas A. Hoagland, USA | Lawrence Reynolds, USA |
| Hans H. Cheng, USA | Tosso Leeb, Switzerland | Lawrence B. Schook, USA |
| Bhanu P. Chowdhary, USA | James D. Murray, USA | Mari A. Smits, The Netherlands |
| Noelle E. Cockett, USA | Anita M. Oberbauer, USA | Leon Spicer, USA |
| Peter Dovc, Slovenia | Jorge A. Piedrahita, USA | J. Verstegen, USA |
| Scott C. Fahrenkrug, USA | Daniel Pomp, USA | Matthew B. Wheeler, USA |
| Dorian J. Garrick, USA | Kent M. Reed, USA | Kenneth L. White, USA |

Biochemistry

| | | |
|-----------------------------|----------------------------|---------------------------|
| Robert Blumenthal, USA | Hicham Fenniri, Canada | Richard D. Ludescher, USA |
| David Ronald Brown, UK | Nick V. Grishin, USA | George Makhatadze, USA |
| Saulius Butenas, USA | J. Guy Guillemette, Canada | Leonid Medved, USA |
| Vittorio Calabrese, Italy | Paul W. Huber, USA | Susan A. Rotenberg, USA |
| F. Castellino, USA | Chen-Hsiung Hung, Taiwan | Jason Shearer, USA |
| Roberta Chiaraluce, Italy | Michael Kalafatis, USA | Andrei Surguchov, USA |
| D. M. Clarke, Canada | B. E. Kemp, Australia | John B. Vincent, USA |
| Francesca Cutruzzolà, Italy | Phillip E. Klebba, USA | Yujun George Zheng, USA |
| Paul W. Doetsch, USA | Wen-Hwa Lee, USA | |

Bioinformatics

| | | |
|----------------------------|------------------------------------|--------------------------|
| T. Akutsu, Japan | Stavros J. Hamodrakas, Greece | Florencio Pazos, Spain |
| Miguel A. Andrade, Germany | Paul Harrison, USA | Zhirong Sun, China |
| Mark Y. Borodovsky, USA | George Karypis, USA | Ying Xu, USA |
| Rita Casadio, Italy | Jack A. Leunissen, The Netherlands | A. Zelikovsky, USA |
| Artem Cherkasov, Canada | Guohui Lin, Canada | Albert Zomaya, Australia |
| David Corne, UK | Satoru Miyano, Japan | |
| Sorin Draghici, USA | Zoran Obradovic, USA | |

Biophysics

Miguel Castanho, Portugal
P. Bryant Chase, USA
Kuo-Chen Chou, USA
Rizwan Khan, India

Ali A. Khraibi, Saudi Arabia
Rumiana Koynova, USA
Serdar Kuyucak, Australia
Jianjie Ma, USA

S. B. Petersen, Denmark
Peter Schuck, USA
Claudio M. Soares, Portugal

Cell Biology

Omar Benzakour, France
Sanford I. Bernstein, USA
Phillip I. Bird, Australia
Eric Bouhassira, USA
Mohamed Boutjdir, USA
Chung-Liang Chien, Taiwan
Richard Gomer, USA
Paul J. Higgins, USA
Pavel Hozak, Czech Republic

Xudong Huang, USA
Anton M. Jetten, USA
Seamus J. Martin, Ireland
Manuela Martins-Green, USA
Shoichiro Ono, USA
George Perry, USA
M. Piacentini, Italy
George E. Plopper, USA
Lawrence Rothblum, USA

Michael Sheetz, USA
James L. Sherley, USA
G. S. Stein, USA
Richard Tucker, USA
Thomas van Groen, USA
Andre Van Wijnen, USA
Steve Winder, UK
Chuan Yue Wu, USA
Bin-Xian Zhang, USA

Genetics

Adewale Adeyinka, USA
Claude Bagnis, France
J. Birchler, USA
Susan Blanton, USA
Barry J. Byrne, USA
R. Chakraborty, USA
Domenico Coviello, Italy
Sarah H. Elsea, USA
Celina Janion, Poland

J. Spencer Johnston, USA
M. Ilyas Kamboh, USA
Feige Kaplan, Canada
Manfred Kayser, The Netherlands
Brynn Levy, USA
Xiao Jiang Li, USA
Thomas Liehr, Germany
James M. Mason, USA
Mohammed Rachidi, France

Raj S. Ramesar, South Africa
Elliot D. Rosen, USA
Dharambir K. Sanghera, USA
Michael Schmid, Germany
Markus Schuelke, Germany
Wolfgang Arthur Schulz, Germany
Jorge Sequeiros, Portugal
Mouldy Sioud, Norway
Rongjia Zhou, China

Genomics

Vladimir Bajic, Saudi Arabia
Margit Burmeister, USA
Settara Chandrasekharappa, USA
Yataro Daigo, Japan
J. Spencer Johnston, USA

Vladimir Larionov, USA
Thomas Lufkin, Singapore
Joakim Lundeberg, Sweden
John L. McGregor, France
John V. Moran, USA

Yasushi Okazaki, Japan
Gopi K. Podila, USA
Momiao Xiong, USA

Immunology

Hassan Alizadeh, USA
Peter Bretscher, Canada
Robert E. Cone, USA
Terry L. Delovitch, Canada
Anthony L. DeVico, USA
Nick Di Girolamo, Australia
Don Mark Estes, USA
Soldano Ferrone, USA
Jeffrey A. Frelinger, USA
John Robert Gordon, Canada

James D. Gorham, USA
Silvia Gregori, Italy
Thomas Griffith, USA
Young S. Hahn, USA
Dorothy E. Lewis, USA
Bradley W. McIntyre, USA
R. Mosley, USA
Marija Mostarica-Stojković, Serbia
Hans Konrad Muller, Australia
Ali Ouassii, France

Kanury V. S. Rao, India
Yair Reisner, Israel
Harry W. Schroeder, USA
Wilhelm Schwaeble, UK
Nilabh Shastri, USA
Yufang Shi, China
Piet Stinissen, Belgium
Hannes Stockinger, Austria
J. W. Tervaert, The Netherlands
Graham R. Wallace, UK

Microbial Biotechnology

Jozef Anné, Belgium
Yoav Bashan, Mexico
Marco Bazzicalupo, Italy
Nico Boon, Belgium
Luca Simone Cocolin, Italy

Peter Coloe, Australia
Daniele Daffonchio, Italy
Han de Winde, The Netherlands
Yanhe Ma, China
Bernd H. A. Rehm, New Zealand

Angela Sessitsch, Austria
Effie Tsakalidou, Greece
J. Wiegand, USA

Microbiology

D. Beighton, UK
Steven R. Blanke, USA
Stanley Brul, The Netherlands
Isaac K. O. Cann, USA
Peter Dimroth, Switzerland
Stephen K. Farrand, USA
Alain Filloux, UK

Gad Frankel, UK
Roy Gross, Germany
Hans-Peter Klenk, Germany
Tanya Parish, UK
Gopi K. Podila, USA
Frederick D. Quinn, USA
Didier A. Raoult, France

Isabel Sá-Correia, Portugal
P. L. C. Small, USA
Lori Snyder, UK
Michael Thomm, Germany
H. C. van der Mei, The Netherlands
Schwan William, USA

Molecular Biology

Rudi Beyaert, Belgium
Michael Bustin, USA
Douglas Cyr, USA
K. Iatrou, Greece
Lokesh Joshi, Ireland
David W. Litchfield, Canada

Wuyuan Lu, USA
Patrick Matthias, Switzerland
John L. McGregor, France
S. L. Mowbray, Sweden
Elena Orlova, UK
Yeon-Kyun Shin, USA

William S. Trimble, Canada
Lisa Wiesmuller, Germany
Masamitsu Yamaguchi, Japan

Oncology

| | | |
|------------------------------------|--------------------------------|---------------------------------------|
| Colin Cooper, UK | Steve B. Jiang, USA | P. J. Oefner, Germany |
| F. M. J. Debruyne, The Netherlands | Daehee Kang, Republic of Korea | Allal Ouhtit, USA |
| Nathan Ames Ellis, USA | Abdul R. Khokhar, USA | Frank Pajonk, USA |
| Dominic Fan, USA | Rakesh Kumar, USA | Waldemar Priebe, USA |
| Gary E. Gallick, USA | Macus Tien Kuo, USA | F. C. Schmitt, Portugal |
| Daila S. Gridley, USA | Eric W. Lam, UK | Sonshin Takao, Japan |
| Xin-yuan Guan, Hong Kong | Sue-Hwa Lin, USA | Ana Maria Tari, USA |
| Anne Hamburger, USA | Kapil Mehta, USA | Henk G. Van Der Poel, The Netherlands |
| Manoor Prakash Hande, Singapore | Orhan Nalcioğlu, USA | Haodong Xu, USA |
| Beric Henderson, Australia | Vincent C. O. Njar, USA | David J. Yang, USA |

Pharmacology

| | | |
|-----------------------------|------------------------------|--------------------------|
| Abdel A. Abdel-Rahman, USA | Ayman El-Kadi, Canada | Kennerly S. Patrick, USA |
| M. Badr, USA | Jeffrey Hughes, USA | Vickram Ramkumar, USA |
| Stelvio M. Bandiera, Canada | Kazim Husain, USA | Michael J. Spinella, USA |
| Ronald E. Baynes, USA | Farhad Kamali, UK | Quadiri Timour, France |
| R. Keith Campbell, USA | Michael Kassiou, Australia | Todd W. Vanderah, USA |
| Hak-Kim Chan, Australia | Joseph J. McArdle, USA | Val J. Watts, USA |
| Michael D. Coleman, UK | Mark J. McKeage, New Zealand | David J. Waxman, USA |
| J. Descotes, France | Daniel T. Monaghan, USA | |
| Dobromir Dobrev, Germany | T. Narahashi, USA | |

Plant Biotechnology

| | | |
|-------------------------------|------------------------------------|----------------------------|
| P. L. Bhalla, Australia | Liwen Jiang, Hong Kong | Ralf Reski, Germany |
| J. R. Botella, Australia | Pulugurtha Bharadwaja Kirti, India | Sudhir Kumar Sopory, India |
| Elvira Gonzalez De Mejia, USA | Yong Pyo Lim, Republic of Korea | |
| H. M. Häggman, Finland | Gopi K. Podila, USA | |

Toxicology

| | | |
|----------------------------|------------------------|-------------------------|
| Michael Aschner, USA | Youmin James Kang, USA | Kenneth Turteltaub, USA |
| Michael L. Cunningham, USA | M. Firoze Khan, USA | Brad Upham, USA |
| Laurence D. Fechter, USA | Pascal Kintz, France | |
| Hartmut Jaeschke, USA | R. S. Tjeerdema, USA | |

Virology

Nafees Ahmad, USA
Edouard Cantin, USA
Ellen Collisson, USA
Kevin M. Coombs, Canada
Norbert K. Herzog, USA
Tom Hobman, Canada
Shahid Jameel, India

Fred Kibenge, Canada
Fenyong Liu, USA
Éric Rassart, Canada
Gerald G. Schumann, Germany
Y.-C. Sung, Republic of Korea
Gregory Tannock, Australia

Ralf Wagner, Germany
Jianguo Wu, China
Decheng Yang, Canada
Jiing-Kuan Yee, USA
Xueping Zhou, China
Wen-Quan Zou, USA

Contents

Clinical and Epidemiological Metabonomics, Mika Ala-Korpela, Veikko Salomaa, and Olav M. Kvalheim
Volume 2011, Article ID 843150, 2 pages

Metabolomic Profiling for Identification of Novel Potential Biomarkers in Cardiovascular Diseases, Maria G. Bardenas, Carlos M. Laborde, Maria Posada, Fernando de la Cuesta, Irene Zubiri, Fernando Vivanco, and Gloria Alvarez-Llamas
Volume 2011, Article ID 790132, 9 pages

Studies of Complex Biological Systems with Applications to Molecular Medicine: The Need to Integrate Transcriptomic and Proteomic Approaches, Elena Silvestri, Assunta Lombardi, Pieter de Lange, Daniela Glinni, Rosalba Senese, Federica Cioffi, Antonia Lanni, Fernando Goglia, and Maria Moreno
Volume 2011, Article ID 810242, 19 pages

Navigating the Human Metabolome for Biomarker Identification and Design of Pharmaceutical Molecules, Irene Kouskoumvekaki and Gianni Panagiotou
Volume 2011, Article ID 525497, 19 pages

Bioactive Food Components and Cancer-Specific Metabonomic Profiles, Young S. Kim and John A. Milner
Volume 2011, Article ID 721213, 9 pages

Exploring Airway Diseases by NMR-Based Metabonomics: A Review of Application to Exhaled Breath Condensate, Matteo Sofia, Mauro Maniscalco, Guglielmo de Laurentiis, Debora Paris, Dominique Melck, and Andrea Motta
Volume 2011, Article ID 403260, 7 pages


metaP-Server: A Web-Based Metabolomics Data Analysis Tool, Gabi Kastenmüller, Werner Römisch-Margl, Brigitte Wägele, Elisabeth Altmaier, and Karsten Suhre
Volume 2011, Article ID 839862, 7 pages

Magnetic Resonance Microscopy Contribution to Interpret High-Resolution Magic Angle Spinning Metabolomic Data of Human Tumor Tissue, M. Carmen Martínez-Bisbal, Vicent Esteve, Beatriz Martínez-Granados, and Bernardo Celda
Volume 2011, Article ID 763684, 8 pages

Combined Reversed Phase HPLC, Mass Spectrometry, and NMR Spectroscopy for a Fast Separation and Efficient Identification of Phosphatidylcholines, Jan Willmann, Herbert Thiele, and Dieter Leibfritz
Volume 2011, Article ID 385786, 8 pages

Artificial Neural Networks for Classification in Metabolomic Studies of Whole Cells Using ^1H Nuclear Magnetic Resonance, D. F. Brougham, G. Ivanova, M. Gottschalk, D. M. Collins, A. J. Eustace, R. O'Connor, and J. Havel
Volume 2011, Article ID 158094, 8 pages

Identification of Urinary Biomarkers of Colon Inflammation in $\text{IL10}^{-/-}$ Mice Using Short-Column LCMS Metabolomics, Don Otter, Mingshu Cao, Hui-Ming Lin, Karl Fraser, Shelley Edmunds, Geoff Lane, and Daryl Rowan
Volume 2011, Article ID 974701, 12 pages



A Metabonomic Approach to Analyze the Dexamethasone-Induced Cleft Palate in Mice, Jinglin Zhou, Bin Xu, Bing Shi, Jing Huang, Wei He, Shengjun Lu, Junjun Lu, Liying Xiao, and Wei Li
Volume 2011, Article ID 509043, 8 pages

Metabolomics Reveals Relationship between Plasma Inositols and Birth Weight: Possible Markers for Fetal Programming of Type 2 Diabetes, Pia Marlene Nissen, Caroline Nebel, Niels Oksbjerg, and Hanne Christine Bertram
Volume 2011, Article ID 378268, 8 pages

Etiological Analysis of Neurodevelopmental Disabilities: Single-Center Eight-Year Clinical Experience in South China, Li Guo, Bing-Xiao Li, Mei Deng, Fang Wen, Jian-Hui Jiang, Yue-Qiu Tan, Yuan-Zong Song, Zhen-Huan Liu, Chun-Hua Zhang, Keiko Kobayashi, and Zi-Neng Wang
Volume 2011, Article ID 318616, 11 pages

Editorial

Clinical and Epidemiological Metabonomics

Mika Ala-Korpela,^{1,2,3,4} Veikko Salomaa,⁵ and Olav M. Kvalheim^{6,7}

¹ College of Chemistry and Chemical Engineering, Central South University, Changsha 410083, Hunan Province, China

² Computational Medicine Research Group, Institute of Clinical Medicine, Faculty of Medicine, University of Oulu and Biocenter Oulu, P.O. Box 5000, 90014 University of Oulu, Oulu, Finland

³ NMR Metabonomics Laboratory, Department of Biosciences, University of Eastern Finland, 70211 Kuopio, Finland

⁴ Department of Internal Medicine and Biocenter Oulu, Clinical Research Center, University of Oulu, 90014 University of Oulu, Oulu, Finland

⁵ Chronic Disease Epidemiology and Prevention Unit, Department of Chronic Disease Prevention, The National Institute for Health and Welfare, 00271 Helsinki, Finland

⁶ Department of Chemistry, University of Bergen, 5007 Bergen, Norway

⁷ Faculty of Health Studies, Sogn og Fjordane University College, 6800 Førde, Norway

Correspondence should be addressed to Mika Ala-Korpela, mika.ala-korpela@computationalmedicine.fi

Received 20 November 2010; Accepted 20 November 2010

Copyright © 2011 Mika Ala-Korpela et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Metabonomics is an omics approach to identify and monitor metabolic characteristics, changes, and phenotypes with respect to various synergetic factors such as environment, life style, diet, and potential pathophysiological processes. Recently, metabonomics has also opened up a possibility for functional genomics in large genome-wide studies. Mass spectrometry (MS) and proton nuclear magnetic resonance (NMR) spectroscopy are the two key experimental technologies in the field. Recent advancements are numerous, and methodologies currently exist that allow for automated high-throughput experimentation in a very cost-effective manner. Technical developments have evoked an increasing number of metabonomics applications in clinical and epidemiological disciplines to understand the complex molecular foundations of various diseases. Techniques with the aim of assessing large numbers of metabolites that are substrates, intermediates, or products in various metabolic pathways are particularly relevant in the risk assessment of metabolic conditions like diabetes and vascular diseases. It has also been envisioned that metabonomics approaches may overtake standard analytical measurements of individual metabolites and eventually lead to holistic multimetabolic risk phenotyping in the early detection of high-risk individuals for various metabolic diseases. Thus, metabonomics

offers potential means to move away from single biomarkers and thresholds in clinical medicine. And, furthermore, even though the current medical biomarkers and thresholds are necessary benchmarks for clinical practice, it will be crucial to let the future diagnostics to follow from new science. Good predictions are necessary for effective prevention; an important step towards this is expected to happen when complex traits will not anymore be limited by clinical diagnoses and definitions but will be handled as multivariate continuous dimensions. Consequently, clinical and epidemiological metabonomics is likely to be one of the key new omics areas to assist translational research in the near future.

Here, the Journal of Biomedicine and Biotechnology presents this special issue, 2011. Although only representing a small fraction of the contribution of metabonomics to various biomedical disciplines, the papers presented here show the increasing potential of metabonomics approaches to complement clinical and epidemiological research.

To start the special issue, we have five contributions that review several aspects of metabonomics research. M. G. Barderas et al. discuss the concepts of metabolic profiling, fingerprinting, and footprinting focusing on the area of cardiovascular diseases. E. Silvestri and coworkers discuss the timely and relevant issue of integrative analyses of related

omics data sets in order to furnish new insights not accessible through one-dimensional data. Further, I. Kouskoumvekaki and G. Panagiotou give an extensive up-to-date overview on the database resources on metabolic pathways and metabolomes and also summarize the statistical tools often used for the analysis of metabonomics data. Y. Kim and S. Milner deal with the challenges related to bioactive food components in cancer research and prevention. Moreover, M. Sofia and coauthors review the potential use of exhaled breath condensate as a suitable matrix for NMR-based metabonomics of airway diseases.

This special issue contains two methodology reports. G. Kastenmüller and coworkers have developed a freely accessible methodology to facilitate automated and standardized analysis for quantitative metabolic data, covering the steps from data acquisition to biological interpretation. M. C. Martínez-Bisbal and colleagues, using magnetic resonance microscopy, present a critical and valuable interpretation on the high-resolution magic angle spinning data of human tumor tissues.

The six original research papers in this special issue start with the paper by J. Willmann et al. who demonstrate the advantage of combining high-pressure liquid chromatography, MS, and NMR spectroscopy in analyzing three different types of extracts of a common membrane component phosphatidylcholine. Then, D. F. Brougham et al. focus on the use of artificial neural networks in classifying proton NMR spectroscopic data recorded on whole-cell culture samples of four different lung carcinoma cell lines, displaying different drug-resistance patterns. In a mouse study, D. Otter et al. apply MS-based nontargeted urinary metabolite profiling to identify biomarkers of colon inflammation. In another mouse study, J. Zhou et al. provide evidence that NMR metabonomics may be sensitive enough to detect small differences in the composition of maternal plasma and thus be helpful for identifying biomarkers of birth defects. Additionally, in low-birth weight and high-birth weight piglets P. M. Nissen and coworkers combine NMR and MS to reveal a relationship between birth weight and plasma inositols, suggesting their potential role in fetal programming of type 2 diabetes. L. Guo et al. end this special issue in providing the latest epidemiologic information on the etiology distribution of neurodevelopmental disabilities in Chinese, and by demonstrating, on the basis of 8-year clinical experience, that MS-based pediatric metabonomics is clinically helpful.

Overall, we hope that this special issue hints at the rising potential of metabonomics approaches in clinical and epidemiological research and translational medicine.

*Mika Ala-Korpela
Veikko Salomaa
Olav M. Kvalheim*

Review Article

Metabolomic Profiling for Identification of Novel Potential Biomarkers in Cardiovascular Diseases

Maria G. Barderas,¹ Carlos M. Laborde,¹ Maria Posada,² Fernando de la Cuesta,² Irene Zubiri,² Fernando Vivanco,^{2,3} and Gloria Alvarez-Llamas²

¹Department of Vascular Physiopathology, Hospital Nacional de Parapléjicos, SESCAM, 45071 Toledo, Spain

²Department of Immunology, IIS-Fundación Jiménez Díaz, 28040 Madrid, Spain

³Department of Biochemistry and Molecular Biology I, Universidad Complutense, 28040 Madrid, Spain

Correspondence should be addressed to Fernando Vivanco, fvivanco@fjd.es

Received 11 June 2010; Revised 11 August 2010; Accepted 12 November 2010

Academic Editor: Mika Ala-Korpela

Copyright © 2011 Maria G. Barderas et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Metabolomics involves the identification and quantification of metabolites present in a biological system. Three different approaches can be used: metabolomic fingerprinting, metabolic profiling, and metabolic footprinting, in order to evaluate the clinical course of a disease, patient recovery, changes in response to surgical intervention or pharmacological treatment, as well as other associated features. Characteristic patterns of metabolites can be revealed that broaden our understanding of a particular disorder. In the present paper, common strategies and analytical techniques used in metabolomic studies are reviewed, particularly with reference to the cardiovascular field.

1. Introduction

Cardiovascular disease (CVD) is the most prevalent cause of death in developed nations and it is increasing in prevalence in developing countries. While many factors contribute to the development of this disease in adults, such as, smoking, male gender, blood pressure, elevated cholesterol, diabetes, and renal failure, the mechanisms underlying CVD are still not fully understood [1–4]. One of the main problems in clinical practice is that the symptoms become evident late in the course of the disease. In fact, asymptomatic processes, that are associated with plaque formation, develop causing silent yet progressive tissue damage. If atheroma plaques finally rupture, highly thrombogenic material is released and an atherothrombotic event occurs. In this context, there is an urgent need to find out novel biomarkers of practical value for clinical intervention which, alone or combined with existing ones, allow cardiovascular risk prediction at individual level. Currently, controversy exists regarding contribution of biomarkers to the information derived from conventional risk factors. When novel markers utility for predicting CVD was investigated in a wide cohort

of more than 5000 individuals without CVD, the gains over conventional factors resulted to be minimal [5]. However, positive outcomes are expected when high-risk populations are investigated; thus, the risk level of selected patients, the chosen biomarkers to be investigated and other factors, such as, statistics, highly influence expected results. Combination of multiple biomarkers in assessment of individual responses adds only moderately to standard risk factors [6]. Therefore, there is a substantial interest in the discovery and use of newer biomarkers, to complement the best existing ones and to identify persons who are at risk for the development of cardiovascular disease and who could be targeted for preventive measures. In particular, finding biomarkers that predict the risk of rupture will provide the opportunity to institute a preventive life style and permit timely pharmacological treatment. Currently, the improvements in outpatient and inpatient care, diagnosis and biomarker discovery have reshaped the landscape of CVD. It is important to note that the new diagnostic methods currently available are based on noninvasive techniques that, although they present a number of benefits, may be limited in terms of specificity, sensitivity, availability and cost. The progress in “-omics”

technologies has provided sensitive, fast and robust tools to analyze biomarkers in CVD.

Metabolites are small molecules that participate in general metabolic reactions and that are required for the maintenance, growth and normal function of a cell. The term metabolome, derived from the word genome, refers to the complete set of metabolites in an organism and its organelles [7, 8] or the total complement of metabolites in a cell [9]. In this way, metabolomics and metabonomics refer to the use of analytical methods to identify and quantify all metabolites in a biological system, as well as the monitoring of changes in the metabolome of a biofluid, cell culture or tissue sample following perturbation [8, 10].

In parallel with genomics, transcriptomics and proteomics, application of metabolomic technologies to the study of CVD will increase our understanding of the pathophysiological processes involved and this should help us to identify potential biomarkers to develop new therapeutic strategies [11, 12]. Indeed, the identification and quantification of these low molecular weight molecules (e.g., lipids, amino acids, and sugars) will define the phenotype of these diseases [13]. From a clinical perspective, the study of metabolic changes that occur in response to different physiological processes will help establish the mechanisms underlying the disease. In terms of personalized medicine, pharmacometabonomic approaches can serve to predict the action of specific drugs in a particular individual based on the predose urinary metabolite profile. Furthermore, the gut bacterial fauna influences drug efficacy, which could be deliberately modified to optimize the benefits and minimize adverse effects of a given treatment [14]. In addition, this approach will help understanding how drugs act during patients' recovery or how they influence outcome.

2. Metabolomic Strategies, Analytical Approaches and Variability

There are several analytical strategies that can be used to analyse the metabolome [15], such as nuclear magnetic resonance (NMR) [16], Fourier transformation infrared spectroscopy (FT-IR) [17, 18], and mass spectrometry (MS) coupled to separation techniques such as high performance liquid chromatography (HPLC), gas chromatography (GC), or capillary electrophoresis (CE). The combination of these different analytical techniques offers important advantages when analyzing the complete metabolome. High field ^1H NMR is one of the preferred platforms for urine and plasma analysis [19, 20], as it is a nondestructive technique that does not require prior separation of the analytes and it provides detailed information on molecular structure. For example, the capacity to predict the occurrence of exercise-induced ischemia in patients with suspected CAD was investigated by NMR blood analysis, demonstrating lactate, glucose, lipids, and long-chain fatty acids to be the main metabolites involved [21]. Xanthine and ascorbate were proposed as possible markers of plaque formation in an atherosclerotic mouse model [22] and lipoprotein subclasses can now be analyzed by a commercial NMR-based protocol called NMR

LipoProfile [23, 24]. However, one of the main limitations of NMR is the poor sensitivity, although this can be improved enormously when it is combined with mass spectrometry.

Coupled to a separation technique, MS has recently been introduced into the metabolomics field and its use in such studies will constitute the main focus of this paper. Indeed, gas chromatography/mass spectrometry (GC-MS), liquid chromatography/mass spectrometry (LC-MS) and capillary electrophoresis/mass spectrometry (CE-MS) are the most powerful techniques for metabolite separation and analysis. GC-MS provides an extraordinary resolution, permitting the separation of structurally similar compounds that would otherwise be very difficult to separate by HPLC. However, this technique requires the analyte to be volatile and thermally stable. In some cases, a chemical derivatization step is required prior to the chromatographic separation in order to render polar metabolites volatile. Some of the metabolites best suited for GC-MS include fatty acids, organic acids, steroids, di-glycerides, sugars and sugar alcohols.

For those metabolites that are not volatile and which cannot be derivatized, LC is the separation technique of choice. Thus, LC-MS can analyze a much wider range of chemical species (polar and nonpolar metabolites) with ample selectivity and sensitivity. Apart from reversed phase chromatography (RP-LC), which is widely used in metabolomics applications, hydrophilic interaction chromatography (HILIC) is a complementary approach suitable for very polar metabolites (nonvolatile). Indeed, the metabolites suited to analysis by GC or HPLC can be represented according to their polarity (see scheme in Figure 1). Similarly, capillary electrophoresis (CE) can be coupled to a mass spectrometer (MS), with the particular advantage of improving the resolution of separation as narrower peaks than with LC are obtained. Accordingly, different approaches have been described in combination with ion trap (IT), triple quadrupole (QQQ), time of flight (TOF), and Q-TOF instruments. The main advantage of QQQ and Q-TOF instruments is that they provide the possibility of identifying the compounds by tandem MS/MS analyses. In order to obtain a full overview of the detectable molecules, electrospray ionization (ESI) should be performed in both positive and negative modes on the same sample.

Irrespective of the analytical approach used in metabolomics, particular care has to be taken in preparing the sample. Bearing in mind that the typical half-lives of metabolic reactions in an organism are less than 1 s, it is important not to monitor metabolic changes extrinsic to the pathology or drug effect under study, producing misleading interpretations of the situation. Evidence-based epidemiological studies have led to the discovery of well-established biomarkers. These studies now tend to be complemented by control-case investigation using a different methodology, based on two main stages: the discovery phase, resulting in a set of novel biomarkers candidates and the validation phase, where discovered potential biomarkers are further validated in a different cohort of samples. In this context, biological variation would be expected to be higher than the analytical variability and thus, it is essential to pay particular attention to: (a) the precise definition of

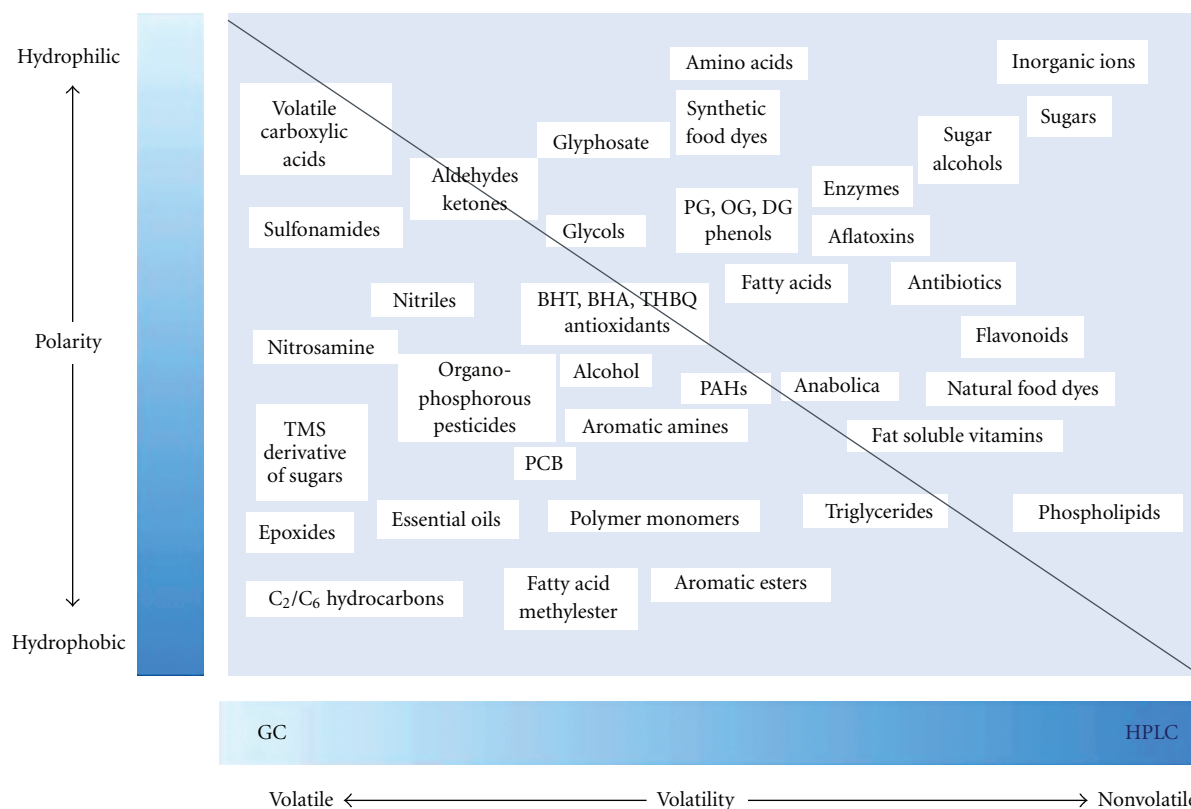


FIGURE 1: Suitability of gas and liquid chromatography for metabolomic analysis based on metabolite polarity. Courtesy of Agilent Technologies [53].

a clinical phenotype (in this sense, network-based analysis on associations among genes, proteins, metabolites, and environmental factors would be encouraged to increase sensitivity and selectivity of the diagnosis) [25] and (b) group matching in terms of sex, age, lifestyle, diet, or pharmacological treatment, otherwise attempts may fail in terms of disease prediction [26]. For instance, gender and statins treatment strongly influence the findings in studies of CVD and when individuals with normal coronary arteries were compared with CVD patients, a >99% confidence limit was only obtained for 6% of the predictions in the treated groups [27]. Technical reproducibility and sampling time are also critical to minimize external factors that will influence the results and their clinical relevance. Ideally, snapshots of different conditions should be taken so that they can be quantitatively compared. If all these considerations are kept in mind, metabolomic research can set out to identify characteristic patterns that can be used for diagnostic purposes and risk prediction, substituting traditional, more expensive clinical approaches (e.g., angiography).

In principle, metabolites can be measured in several body fluids or tissues, although plasma and urine are the most commonly used biological matrices in cardiovascular research due to their availability and clinical relevance as a source of potential biomarkers. Almost all cells in the body communicate with the plasma, either directly or through different tissues and biological fluids, releasing at least part of their intracellular content [28]. By contrast, urine is

produced by renal filtration of the plasma and it is widely considered as one of the most important samples for diagnosis as it contains not only many plasma components but also the catabolic products of different metabolic pathways. Sample pretreatment varies depending on the analytical platform chosen (see the common strategies employed in Figure 2). Metabolites from frozen tissue samples can be extracted and simultaneously fractionated by treating the ground tissue with mixtures of organic solvents, such that molecules are extracted in different fractions according to their polarity. If a biological fluid is the starting material (urine, serum, plasma), metabolite fractions are usually obtained after proteins are removed by precipitation. The crude or diluted sample can then be injected directly, although matrix effects causing ion suppression should be expected. If analyzed by LC-MS(/MS), it may be desirable to preconcentrate (e.g., by lyophilisation) or fractionate the sample prior to chromatographic separation. In case of GC-MS(/MS), preconcentration can be performed by solid phase microextraction (SPME) with or without head space (HS) procedures, which are particularly useful when analysing volatile organic compounds (VOCs). For CE analysis, the salt content should be minimized in the sample.

In general, three complementary approaches are used for metabolic research (see Figure 3): metabolic fingerprinting, metabolic profiling, and metabolic footprinting [29]. In the first case, and like proteomics strategies, an unbiased analysis is performed that is oriented towards defining

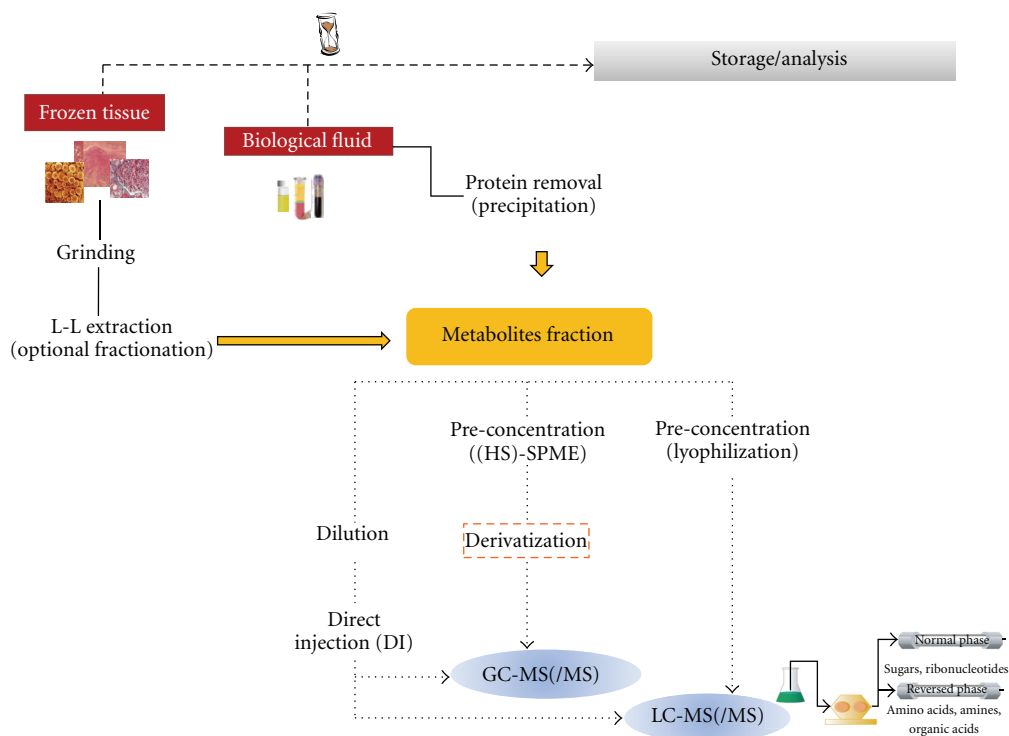


FIGURE 2: Schematic view of the sample pretreatment for metabolomic analysis of frozen tissue or biological fluid prior to GC-MS or LC-MS analysis.

clinically relevant differences rather than identifying all the molecules present in a sample [30]. Alternatively, metabolic profiling involves a preselection of a set of metabolites, or a specific class of compounds, that might participate in a targeted pathway. Metabolic footprinting does not rely on the measurement of intracellular metabolites but rather, on monitoring those that are secreted or fail to be taken up by a cell or tissue [31, 32]. Below, we will discuss relevant findings from these approaches in CVD (a compilation of the main studies is shown in Table 1).

3. Metabolic Fingerprinting in CVD

Metabolic fingerprinting does not aim to identify the entire set of metabolites but rather to compare patterns or fingerprints of metabolites that change in response to a disease state, pharmacological therapies or environmental alterations, for example. A wide variety of biological matrices can be used for metabolic fingerprinting, such as urine, plasma/serum, tissues/cells and saliva. This approach can be used as a diagnostic tool to evaluate the disease state by comparing healthy controls and disease subjects, or to assay the success of a particular treatment (prognosis/recovery) [33]. However, if we want to understand the mechanisms underlying a disease, qualitative and quantitative analyses are required. Once a differential pattern is discovered, which provides information that can be considered as the pathological phenotype [34], further steps to identify the participating compounds (qualitative) and to determine the

absolute amounts of metabolites that participate in the processes studied (quantitative) must be followed. This is not a trivial issue and prior to embarking on the task of discovering metabolic biomarkers, sufficiently sensitive and selective instruments and extensive compound libraries for metabolite identification should be available [35], while certain expertise in data analysis and interpretation will be necessary.

One of the few metabolomic studies in the field of CVD involved a comparison of the metabolomic fingerprint obtained by GC-MS of plasma samples from non-ST-segment elevation acute coronary syndrome (NSTEMACS) patients, stable atherosclerosis patients and healthy patients [36]. Citric acid, 4-hydroxyproline (4OH-Pro), aspartic acid and fructose were found to decrease in NSTEMACS patients, whereas lactate, urea, glucose, and valine increased. Both lactate and glucose are also involved in prediction of exercise-induced ischemia in patients with suspected CAD [21]. The decreased in 4OH-Pro was especially interesting because circulating 4OH-Pro is thought to prevent the binding of LDL to lipoprotein previously deposited in the vascular wall, as well as releasing already deposited LDL from the atherosclerotic lesions. It is also a component of collagen, which confers stability to the atherosclerotic plaque.

The high resolution of CE-MS makes it a powerful technique to separate and analyse charged metabolites, although only a few metabolomic applications have been published to date. The isolation of polypeptide fraction from urine or plasma was analyzed by CE-MS and used to discriminate between coronary artery disease (CAD)

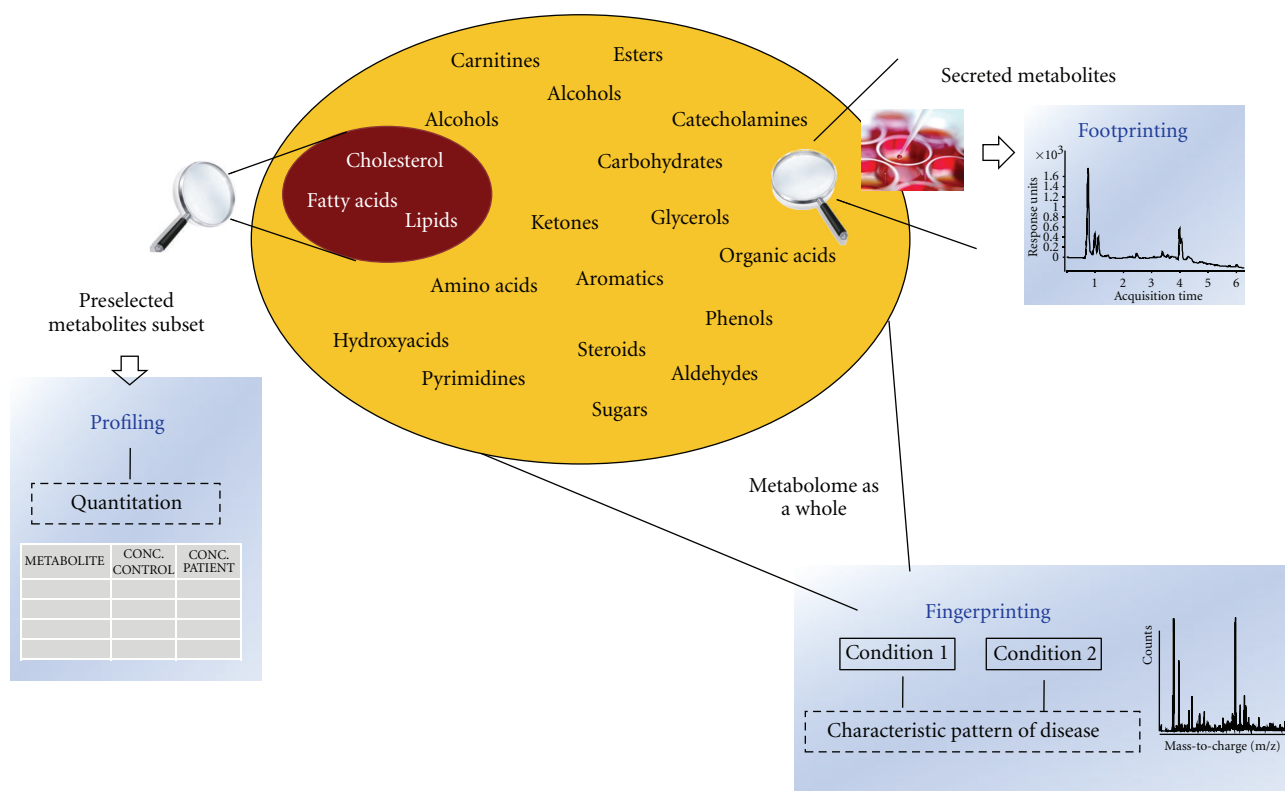


FIGURE 3: Common strategies in metabolomics: fingerprinting, profiling and footprinting.

and non-CAD patients with clinical symptoms and who had been subjected to coronary angiography [37]. The stability of urine samples and their resistance to oxidation or precipitation reflect the advantages of this biological fluid for proteomic analysis. Polypeptide profiling in urine is more reproducible than in plasma, with no significant loss of polypeptides over time when performing consecutive analyses over a 24-hour period, which also demonstrates the reproducibility of the CE-MS. In total, 200 of the most abundant polypeptides were detected and a set of 17 urinary polypeptides permitted CAD and non-CAD patients to be distinguished. Among them, collagen α -1 (I and III) was augmented in CAD samples, which was corroborated by their increased expression found in atherosclerotic plaques. This increase points to an important role for collagen in the development of atherosclerosis.

4. Metabolic Profiling in CVD

Metabolite profiling focuses on the analysis of a group of metabolites related to a specific metabolic pathway [38, 39]. In this approach, target metabolites are selected beforehand and they are assessed using specific analytical methods. Technological advances have increased the number of metabolites that can be quantified simultaneously. Moreover, the results of metabolic profiling are quantitatively independent of the technology used for data acquisition [40].

Metabolite profiling has been applied to CVD in order to identify and quantify metabolites that might serve as

new biomarkers. A metabolite profile of peripheral blood from individuals undergoing planned myocardial infarction (PMI) has been established [41]. Serial blood from 36 patients were obtained before and at various intervals after PMI, and the changes in circulating levels of metabolites were identified by mass spectrometry-based metabolite profiling. Most alterations produced by PMI were observed in the tricarboxylic acid cycle, in purine and pyrimidine catabolism, and in the pentose phosphate pathway. Indeed, 7 metabolites were significantly affected immediately 10 minutes after the onset of myocardial injury ($P < .005$): alanine, aminoisobutyric acid, hypoxanthine, isoleucine/leucine, malonic acid, threonine, and trimethylamine N-oxide (TMNO). All these alterations were especially interesting as they were observed before any significant rise in the clinically available biomarkers in plasma (CKMB and troponin T). After 60 minutes, six new metabolites had also changed significantly ($P < .005$): 1-methylhistamine, choline, inosine, serine, proline, and xanthine, with the later being a candidate of a marker for plaque formation in an atherosclerotic mouse model [22]. The anatomic origin of the early metabolic changes observed was further explored in a subgroup of 13 patients by simultaneously comparing the metabolite levels obtained in samples from peripheral blood and from a catheter placed in the coronary sinus. A further 8 metabolites were transmurally enriched at least 1.3-fold 10 minutes after PMI (taurine, ribose-5-phosphate, DCMP, lactic acid, AMP, malic acid, glutamine and glutamic acid) and once 60 minutes had passed, six additional metabolites augmented

TABLE 1: Compilation of the metabolomic studies in cardiovascular field, including candidate biomarkers.

| Pathology (Patients no.) | Metabolite | Body fluid/tissue | Analytical technique | Replication | Reference |
|--|--|-------------------|----------------------|---|-----------|
| NSTACS (9), stable atherosclerosis (10), healthy (10) | 4-hydroxyproline | Plasma | GC-MS | — | [36] |
| CAD (15), no CAD (14) | 17 polypeptides (CAD pattern) Collagen α 1 (I,III) | Urine, Plasma | CE-MS | Test set: CAD (26), no CAD (12) | [37] |
| PMI (20), control (16) SMI (12), control (9) | Aconitic acid, hypoxanthine, trimethylamine N-oxide, threonine | Plasma | LC-MS | Validation: PMI (16) | [41] |
| High-fat diet (9), common-diet (9) rats | 12 altered in plasma, 8 altered in urine (atherosclerotic rats) | Urine, Plasma | LC-MS | — | [42] |
| Atherosclerosis (9), healthy (10) | 24 altered metabolites (insulin resistance) | Plasma | NMR, GC-MS | — | [45] |
| CAD (12), LVD (10), control (17) | Acetylcarnitine, 3-hydroxybutyrylcarnitine | Plasma | FI-MS | — | [46] |
| Initial: CAD (174), control (174) Replication: CAD (140), control (140) | Dicarboxyl acylcarnitines | Plasma | GC-MS | Event replication: CAD (63), control (66) | [49] |
| Inducible ischemia (18), control (18) | Citric acid pathway | Plasma | LC-MS | — | [52] |
| Persistent AF: AF (8), SR (8) Post-operative AF: SR-AF (18), SR (19) | β -hydroxybutyrate, ketogenic amino acids, glycine | Atrial tissue | NMR | — | [50] |
| Apo E ^{-/-} mice captroil treated (8), untreated (8) | Allantoin (drug treatment) Xanthine, ascorbate (plaque formation) | Urine | NMR | — | [22] |

CAD: coronary artery disease, PMI: planned myocardial infarction, SMI: spontaneous myocardial infarction, LVD: left ventricular dysfunction, FI: flow injection, AF: atrial fibrillation, SR: sinus rhythm. Numbers in brackets correspond to number of assayed individuals (or animals if so specified).

(glycerol-3-phosphate, orotic acid, succinic acid, glycerate-2-phosphate, taurine and malic acid).

Plasma and urine samples from atherosclerotic and control rats have been compared by ultra fast liquid chromatography coupled to ion trap-time of flight (IT-TOF) mass spectrometry (UFLC/MS-IT-TOF) [42]. Accordingly, 12 metabolites were identified as potential biomarkers in rat plasma and 8 metabolites in rat urine. The concentration of leucine, phenylalanine, tryptophan, acetylcarnitine, butyrylcarnitine, propionylcarnitine and spermine decreased in plasma, and 3-O-methyl-dopa, ethyl N2-acetyl-L-argininate, leucylproline, glucuronate, N(6)-(N-threonylcarbonyl)-adenosine and methyl-hippuric acid were diminished in the urine of atherosclerosis rats. Conversely, ursodeoxycholic acid, chenodeoxycholic acid, LPC (C16:0), LPC (C18:0) and LPC (C18:1) increased in plasma and hippuric acid augmented in the urine from atherosclerosis rats. The alterations to these metabolites reflected the abnormal metabolism of phenylalanine, tryptophan, bile acids and amino acids. Lysophosphatidylcholine (LPC) plays an important role in inflammation and cell proliferation, highlighting the relationship between LPC with the progress of atherosclerosis and other inflammatory diseases.

The lipidomic profile of mice liver homogenates from cholesterol-free, low cholesterol and high cholesterol diets demonstrated the influence of dietary cholesterol intake and atherosclerosis [43]. To obtain individual metabolite fingerprints, nearly 300 metabolites were measured in plasma

samples by LC-MS/MS, including di- and tri-glycerides, phosphatidylcholines, lysophosphatidylcholines and cholesterol esters. When dietary cholesterol intake increased, the liver compensated for the elevation in plasma cholesterol by adjusting metabolic and transport processes related to lipid metabolism, which leads to an inflammatory, pro-atherosclerotic state. A cholesterol-free diet did not induce early atherosclerosis, while the low cholesterol diet only mildly induced early atherosclerosis. By contrast, intense early atherosclerosis was induced by the high cholesterol diet, in association with proinflammatory gene expression. Indeed, a relationship appears to exist between cholesterol intake (measured as cholesterol plasma levels) and atherosclerotic lesion size.

The lipidome of cell membranes and tissues has been studied by measuring the plasmalogens contained in rabbit and rat myocardial nuclei by ESI-MS [44]. Plasmalogen is an ether lipid where the first position of glycerol binds a vinyl residue with the double bond next to the ether bond. The second carbon has a typical ester-linked fatty acid and the third carbon usually has a phospholipid head group, which can protect cells against the damaging effect of singlet oxygen. This seems to be the reason for the strong enrichment of plasmalogens found in the membrane of myocardial cells.

Metabolic changes associated to atherosclerosis have also been investigated through NMR and GC-MS metabolite profiling [45]. There are clear biochemical explanations to

these findings, and the alterations to these metabolites, which cause the final atherosclerotic lesion, can be related to different disorders. For instance, insulin resistance in diabetic patients increases the activity of transaminases, which are critical enzymes in amino acid metabolic pathways. Hence, if the insulin response is deficient, these amino acid pathways will be altered, and many others metabolites will be affected such as glutamate, ketoglutarate, succinyl-CoA, 4-OH-L-proline (4OHPro), 2-hydroxybutyrate, creatinine, pyruvate, oxaloacetate, malate, glycolate and 2,3,4-trihydroxybutyrate. These effects could indicate damage to tissue at the intima artery walls.

The myocardial metabolic response has been investigated in CAD and left ventricular dysfunction (LVD) patients, both at baseline and following ischemia-reperfusion (I/R) [46]. Accordingly, glucose, lactate, free fatty acids, total ketones, 3-hydroxybutyrate, pyruvate, leucine/isoleucine and glutamate are present at lower concentrations in a preischemia state in the coronary sinus (CS) than in arterial samples (reflecting myocardial uptake). By contrast, the alanine concentration is higher (reflecting release). A principal components analysis (PCA) shows several potentially important postoperative metabolic changes during the clinical course of the disease. Ventricle dysfunction are associated with the global suppression of metabolic fuel uptake, and limited myocardial metabolic reserves and flexibility following global I/R stress is associated with cardiac surgery.

Some citric acid metabolites are depressed in acute ischemia and acute myocardial disease [47]. The citric acid cycle plays an important role in oxidative phosphorylation and ATP production in cardiomyocytes, and citric acid cycle intermediates are supplied by glycolysis and β -oxidation of fatty acids. Metabolomic profiling based on quantitative mass spectrometry was also used to study the heritability of premature coronary disease in 117 individuals unaffected by CAD but with a family member affected [48]. There was a strong heritability of amino acid levels such as arginine, ornithine, alanine, proline, leucine/isoleucine, valine, glutamate/glutamine, phenylalanine and glycine, free fatty acids such as arachidonic, palmitic, linoleic and acylcarnitines. Hence, it was concluded that metabolic changes associated with CAD can be inherited and they are strongly related to age. This would indicate that metabolic processes could be controlled genetically, implying a correlation between genotype and phenotype in families with CAD. More recently, a subset of 69 metabolites was shown to have diagnostic value, such that some derived factors showed discriminative capability for CAD after PCA. Moreover, a signature composed of dicarboxyacylcarnitines was predictive of further cardiovascular events in those patients and most significant differences persisted after adjustment for CAD risk factors [49].

Metabolic changes in human atrial fibrillation (AF) have been investigated by NMR, performing a quantitative analysis of 24 previously selected metabolites. Significant differences were found for beta-hydroxybutyrate, ketogenic amino acids and glycine, all of which augmented in AF patients when compared to control subjects, suggesting a pathological role for ketone bodies. Metabolic profiles enable

more than 80% of patients at risk of AF at the time of coronary artery bypass grafting to be classified, as a discordant regulation of energy metabolites was found to precede post-operative AF [50]. The effect of drug treatment on apoE^{-/-} mice was investigated by NMR analysis of metabolites in urine, showing allantoin to act as a marker for drug treatment, and xanthine and ascorbate as possible markers of plaque formation (both were elevated in untreated mice) [22].

5. Conclusion

The application of metabolic analysis to cardiovascular diseases is an emerging field [51], and at this incipient stage it is not possible to clearly define a metabolic picture which is responsible for CVD prediction and progression. Further metabolomic investigation promises to improve researchers and clinicians knowledge of these diseases in three critical ways. Firstly, a complete description of the metabolites altered in a disease will better define the pathophysiology of the disease. Secondly, metabolic profiling will enhance the feasibility of high-throughput patient screening to diagnose the disease state or risk evaluation [52]. Indeed, the identification of clinically relevant changes in circulating metabolites that may be considered as potential new biomarkers will also help with the evaluation of prognosis and will contribute to the development of new therapeutic strategies. Thirdly, metabolite profiling will enable the effects of pharmacological treatments to be monitored, in particular, assessing the individual's response to a particular drug. In contrast to genomics, metabolomics defines dynamic states that reflect the actual status of an organism, which requires the control of many variables (from an individual's status to metabolite degradation following sample collection). Failure to do so may lead to the production of erroneous results and misleading conclusions. Minimal protocol-specific differences can produce inconsistent findings, which must be clearly overcome prior to proposing the use of a biomarker to the scientific community. Similarly, the results must be confirmed in a validation cohort composed by a different set of samples than that used in the discovery phase. Adequate follow-up studies must corroborate earlier predictions, and adjustment for conventional risk factors to assess significant contribution of a discovered metabolite to current knowledge should be included. To date, there have been considerable efforts in improving instrumentation (e.g., mass spectrometry) and the analytical methods suitable to complement these approaches (e.g., based on NMR), resulting in an expansion of the metabolites with potential roles in the development of atherosclerosis that can be quantified. However, further research is still needed prior to proposing an ideal platform for metabolite analysis that can replace conventional CVD diagnosis in clinical practice. With the growth of public metabolite databases, further improvements in the sensitivity and selectivity of analytical techniques and the development and routine use of novel platforms of demonstrated potential, novel targets are expected to be discovered in the near future.

Acknowledgments

This work was supported by Ministerio de Educacion y Ciencia (BFU-2005-08838), CAM (Proteomarkers S2006/GEN-0247), Instituto de Salud Carlos III (FIS PI070537, FIS PI080970), Mutua Madrileña Automovilista (20174/004), and INDAS-Biotech. Fernando de la Cuesta and Gloria Alvarez-Llamas were supported by FIS (FI06/00583, CP09/00229). Maria Posada was supported by the Fundacion Conchita Rabago (IIS-Fundacion Jimenez Diaz). F. Vivanco and G. Alvarez-Llamas contributed equally to this work.

References

- [1] P. W. F. Wilson, R. B. D'Agostino, D. Levy, A. M. Belanger, H. Silbershatz, and W. B. Kannel, "Prediction of coronary heart disease using risk factor categories," *Circulation*, vol. 97, no. 18, pp. 1837–1847, 1998.
- [2] D. G. Hackam and S. S. Anand, "Emerging risk factors for atherosclerotic vascular disease: a critical review of the evidence," *Journal of the American Medical Association*, vol. 290, no. 7, pp. 932–940, 2003.
- [3] P. W. Wilson, "Progressing from risk factors to omics," *Circulation Cardiovascular Genetics*, vol. 1, no. 2, pp. 141–146, 2008.
- [4] P. Libby, P. M. Ridker, and A. Maseri, "Inflammation and atherosclerosis," *Circulation*, vol. 105, no. 9, pp. 1135–1143, 2002.
- [5] O. Melander, C. Newton-Cheh, P. Almgren et al., "Novel and conventional biomarkers for prediction of incident cardiovascular events in the community," *Journal of the American Medical Association*, vol. 302, no. 1, pp. 49–57, 2009.
- [6] T. J. Wang, P. Gona, M. G. Larson et al., "Multiple biomarkers for the prediction of first major cardiovascular events and death," *New England Journal of Medicine*, vol. 355, no. 25, pp. 2631–2639, 2006.
- [7] S. G. Oliver, M. K. Winson, D. B. Kell, and F. Baganz, "Systematic functional analysis of the yeast genome," *Trends in Biotechnology*, vol. 16, no. 9, pp. 373–378, 1998.
- [8] O. Fiehn, J. Kopka, P. Dörmann, T. Altmann, R. N. Trethewey, and L. Willmitzer, "Metabolite profiling for plant functional genomics," *Nature Biotechnology*, vol. 18, no. 11, pp. 1157–1161, 2000.
- [9] H. Tweeddale, L. Notley-McRobb, and T. Ferenci, "Effect of slow growth on metabolism of *Escherichia coli*, as revealed by global metabolite pool ('metabolome') analysis," *Journal of Bacteriology*, vol. 180, no. 19, pp. 5109–5116, 1998.
- [10] J. K. Nicholson, J. C. Lindon, and E. Holmes, "'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data," *Xenobiotica*, vol. 29, no. 11, pp. 1181–1189, 1999.
- [11] R. Kaddurah-Daouk, B. S. Kristal, and R. M. Weinshilboum, "Metabonomics: a global biochemical approach to drug response and disease," *Annual Review of Pharmacology and Toxicology*, vol. 48, pp. 653–683, 2008.
- [12] S. N. Goonewardena, L. E. Prevette, and A. A. Desai, "Metabonomics and atherosclerosis," *Current Atherosclerosis Reports*, vol. 12, no. 4, pp. 267–272, 2010.
- [13] C. L. Waterman, C. Kian-Kai, and J. L. Griffin, "Metabonomic strategies to study lipotoxicity in cardiovascular disease," *Biochimica et Biophysica Acta*, vol. 1801, no. 3, pp. 230–234, 2010.
- [14] T. A. Clayton, D. Baker, J. C. Lindon, J. R. Everett, and J. K. Nicholson, "Pharmacometabonomic identification of a significant host-microbiome metabolic interaction affecting human drug metabolism," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 34, pp. 14728–14733, 2009.
- [15] W. B. Dunn, N. J. C. Bailey, and H. E. Johnson, "Measuring the metabolome: current analytical technologies," *Analyst*, vol. 130, no. 5, pp. 606–625, 2005.
- [16] J. K. Nicholson and I. D. Wilson, "Understanding 'global' systems biology: metabonomics and the continuum of metabolism," *Nature Reviews Drug Discovery*, vol. 2, no. 8, pp. 668–676, 2003.
- [17] G. G. Harrigan, R. H. LaPlante, G. N. Cosma et al., "Application of high-throughput Fourier-transform infrared spectroscopy in toxicology studies: contribution to a study on the development of an animal model for idiosyncratic toxicity," *Toxicology Letters*, vol. 146, no. 3, pp. 197–205, 2004.
- [18] H. E. Johnson, D. Broadhurst, D. B. Kell, M. K. Theodorou, R. J. Merry, and G. W. Griffith, "High-throughput metabolic fingerprinting of legume silage fermentations via Fourier transform infrared spectroscopy and chemometrics," *Applied and Environmental Microbiology*, vol. 70, no. 3, pp. 1583–1592, 2004.
- [19] E. M. Lenz, J. Bright, I. D. Wilson, S. R. Morgan, and A. F. P. Nash, "A H NMR-based metabonomic study of urine and plasma samples obtained from healthy human subjects," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 33, no. 5, pp. 1103–1115, 2003.
- [20] M. Ala-Korpela, "Critical evaluation of H NMR metabonomics of serum as a methodology for disease risk assessment and diagnostics," *Clinical Chemistry and Laboratory Medicine*, vol. 46, no. 1, pp. 27–42, 2008.
- [21] I. Barba, G. De León, E. Martín et al., "Nuclear magnetic resonance-based metabolomics predicts exercise-induced ischemia in patients with suspected coronary artery disease," *Magnetic Resonance in Medicine*, vol. 60, no. 1, pp. 27–32, 2008.
- [22] G. C. Leo and A. L. Darrow, "NMR-based metabolomics of urine for the atherosclerotic mouse model using apolipoprotein-E deficient mice," *Magnetic Resonance in Chemistry*, vol. 47, supplement 1, pp. S20–S25, 2009.
- [23] E. J. Jeyarajah, W. C. Cromwell, and J. D. Otvos, "Lipoprotein particle analysis by nuclear magnetic resonance spectroscopy," *Clinics in Laboratory Medicine*, vol. 26, no. 4, pp. 847–870, 2006.
- [24] M. Ala-Korpela, "Potential role of body fluid H NMR metabonomics as a prognostic and diagnostic tool," *Expert Review of Molecular Diagnostics*, vol. 7, no. 6, pp. 761–773, 2007.
- [25] J. Loscalzo, I. Kohane, and A. L. Barabasi, "Human disease classification in the postgenomic era: a complex systems approach to human pathobiology," *Molecular Systems Biology*, vol. 3, article 124, 2007.
- [26] E. M. Lenz, J. Bright, I. D. Wilson et al., "Metabonomics, dietary influences and cultural differences: a H NMR-based study of urine samples obtained from healthy British and Swedish subjects," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 36, no. 4, pp. 841–849, 2004.
- [27] H. L. Kirschenlohr, J. L. Griffin, S. C. Clarke et al., "Proton NMR analysis of plasma is a weak predictor of coronary artery disease," *Nature Medicine*, vol. 12, no. 6, pp. 705–710, 2006.
- [28] V. M. Dardé, M. G. Barderas, and F. Vivanco, "Depletion of high-abundance proteins in plasma by immunoaffinity

- subtraction for two-dimensional difference gel electrophoresis analysis," *Methods in Molecular Biology*, vol. 357, pp. 351–364, 2007.
- [29] K. Dettmer and B. D. Hammock, "Metabolomics—a new exciting field within the "omics" sciences," *Environmental Health Perspectives*, vol. 112, no. 7, pp. A396–A397, 2004.
- [30] R. C. Deo, L. Hunter, G. D. Lewis et al., "Interpreting metabolomic profiles using unbiased pathway models," *PLoS Computational Biology*, vol. 6, no. 2, Article ID e1000692, 2010.
- [31] D. B. Kell, M. Brown, H. M. Davey, W. B. Dunn, I. Spasic, and S. G. Oliver, "Metabolic footprinting and systems biology: the medium is the message," *Nature Reviews Microbiology*, vol. 3, no. 7, pp. 557–565, 2005.
- [32] G. A. Pope, D. A. MacKenzie, M. Defernez et al., "Metabolic footprinting as a tool for discriminating between brewing yeasts," *Yeast*, vol. 24, no. 8, pp. 667–679, 2007.
- [33] J. L. Griffin and R. A. Kauppinen, "A metabolomics perspective of human brain tumours," *FEBS Journal*, vol. 274, no. 5, pp. 1132–1139, 2007.
- [34] S. M. Watkins and J. B. German, "Toward the implementation of metabolomic assessments of human health and nutrition," *Current Opinion in Biotechnology*, vol. 13, no. 5, pp. 512–516, 2002.
- [35] H. Horai, M. Arita, S. Kanaya et al., "MassBank: a public repository for sharing mass spectral data for life sciences," *Journal of Mass Spectrometry*, vol. 45, no. 7, pp. 703–714, 2010.
- [36] M. Vallejo, A. García, J. Tuñón et al., "Plasma fingerprinting with GC-MS in acute coronary syndrome," *Analytical and Bioanalytical Chemistry*, vol. 394, no. 6, pp. 1517–1524, 2009.
- [37] C. Z. Von Muhlen, E. Schiffer, P. Zuerblg et al., "Evaluation of urine proteome pattern analysis for its potential to reflect coronary artery atherosclerosis in symptomatic patients," *Journal of Proteome Research*, vol. 8, no. 1, pp. 335–345, 2009.
- [38] E. S. Musiek, H. Yin, G. L. Milne, and J. D. Morrow, "Recent advances in the biochemistry and clinical relevance of the isoprostane pathway," *Lipids*, vol. 40, no. 10, pp. 987–994, 2005.
- [39] H. J. Cho, J. D. Kim, W. Y. Lee, B. C. Chung, and M. H. Choi, "Quantitative metabolic profiling of 21 endogenous corticosteroids in urine by liquid chromatography-triple quadrupole-mass spectrometry," *Analytica Chimica Acta*, vol. 632, no. 1, pp. 101–108, 2009.
- [40] K. Dettmer, P. A. Aronov, and B. D. Hammock, "Mass spectrometry-based metabolomics," *Mass Spectrometry Reviews*, vol. 26, no. 1, pp. 51–78, 2007.
- [41] G. D. Lewis, R. U. Wei, E. Liu et al., "Metabolite profiling of blood from individuals undergoing planned myocardial infarction reveals early markers of myocardial injury," *Journal of Clinical Investigation*, vol. 118, no. 10, pp. 3503–3512, 2008.
- [42] F. Zhang, Z. Jia, P. Gao et al., "Metabonomics study of atherosclerosis rats by ultra fast liquid chromatography coupled with ion trap-time of flight mass spectrometry," *Talanta*, vol. 79, no. 3, pp. 836–844, 2009.
- [43] R. Kleemann, L. Verschuren, M. J. Van Erk et al., "Atherosclerosis and liver inflammation induced by increased dietary cholesterol intake: a combined transcriptomics and metabolomics analysis," *Genome Biology*, vol. 8, no. 9, article R200, 2007.
- [44] C. J. Albert, D. S. Anbukumar, J. K. Monda, J. T. Eckelkamp, and D. A. Ford, "Myocardial lipidomics. Developments in myocardial nuclear lipidomics," *Frontiers in Bioscience*, vol. 12, pp. 2750–2760, 2007.
- [45] J. Teul, F. J. Rupérez, A. Garcia et al., "Improving metabolite knowledge in stable atherosclerosis patients by association and correlation of GC-MS and ¹H NMR fingerprints," *Journal of Proteome Research*, vol. 8, no. 12, pp. 5580–5589, 2009.
- [46] A. T. Turer, R. D. Stevens, J. R. Bain et al., "Metabolomic profiling reveals distinct patterns of myocardial substrate use in humans with coronary artery disease or left ventricular dysfunction during surgical ischemia/reperfusion," *Circulation*, vol. 119, no. 13, pp. 1736–1746, 2009.
- [47] G. Zhao, H. J. Nam, S. C. Burgess et al., "Overexpression of pyruvate dehydrogenase kinase 4 in heart perturbs metabolism and exacerbates calcineurin-induced cardiomyopathy," *American Journal of Physiology*, vol. 294, no. 2, pp. H936–H943, 2008.
- [48] S. H. Shah, E. R. Hauser, J. R. Bain et al., "High heritability of metabolomic profiles in families burdened with premature cardiovascular disease," *Molecular Systems Biology*, vol. 5, article 258, 2009.
- [49] S. H. Shah, J. R. Bain, M. J. Muehlbauer et al., "Association of a peripheral blood metabolic profile with coronary artery disease and risk of subsequent cardiovascular events," *Circulation Cardiovascular Genetics*, vol. 3, no. 2, pp. 207–214, 2010.
- [50] M. Mayr, S. Yusuf, G. Weir et al., "Combined metabolomic and proteomic analysis of human atrial fibrillation," *Journal of the American College of Cardiology*, vol. 51, no. 5, pp. 585–594, 2008.
- [51] A. Giovane, A. Balestrieri, and C. Napoli, "New insights into cardiovascular and lipid metabolomics," *Journal of Cellular Biochemistry*, vol. 105, no. 3, pp. 648–654, 2008.
- [52] M. S. Sabatine, E. Liu, D. A. Morrow et al., "Metabolomic identification of novel biomarkers of myocardial ischemia," *Circulation*, vol. 112, no. 25, pp. 3868–3875, 2005.
- [53] A. Gratzfeld-Hüsgen and R. Schuster, *HPLC for Food Analysis: A Primer*, Hewlett-Packard, Palo Alto, Calif, USA, 1996.

Review Article

Studies of Complex Biological Systems with Applications to Molecular Medicine: The Need to Integrate Transcriptomic and Proteomic Approaches

Elena Silvestri,¹ Assunta Lombardi,² Pieter de Lange,³ Daniela Glinni,¹ Rosalba Senese,³ Federica Cioffi,³ Antonia Lanni,³ Fernando Goglia,¹ and Maria Moreno¹

¹ *Dipartimento di Scienze Biologiche ed Ambientali, Università degli Studi del Sannio, Via Port'Arsa 11, 82100 Benevento, Italy*

² *Dipartimento delle Scienze Biologiche, Sezione Fisiologia, Università degli Studi di Napoli "Federico II", Via Mezzocannone 8, 80134 Napoli, Italy*

³ *Dipartimento di Scienze della Vita, Seconda Università di Napoli, Via Vivaldi 43, 81100 Caserta, Italy*

Correspondence should be addressed to Maria Moreno, moreno@unisannio.it

Received 15 April 2010; Accepted 8 September 2010

Academic Editor: Mika Ala-Korpela

Copyright © 2011 Elena Silvestri et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Omic approaches to the study of complex biological systems with potential applications to molecular medicine are attracting great interest in clinical as well as in basic biological research. Genomics, transcriptomics and proteomics are characterized by the lack of an *a priori* definition of scope, and this gives sufficient leeway for investigators (a) to discern all at once a globally altered pattern of gene/protein expression and (b) to examine the complex interactions that regulate entire biological processes. Two popular platforms in "omics" are DNA microarrays, which measure messenger RNA transcript levels, and proteomic analyses, which identify and quantify proteins. Because of their intrinsic strengths and weaknesses, no single approach can fully unravel the complexities of fundamental biological events. However, an appropriate combination of different tools could lead to integrative analyses that would furnish new insights not accessible through one-dimensional datasets. In this review, we will outline some of the challenges associated with integrative analyses relating to the changes in metabolic pathways that occur in complex pathophysiological conditions (*viz.* ageing and altered thyroid state) in relevant metabolically active tissues. In addition, we discuss several new applications of proteomic analysis to the investigation of mitochondrial activity.

1. Introduction

Genomic and proteomic data analyses have proven to be essential for an understanding of the underlying factors involved in human disease and for the discovery of diagnostic biomarkers, as well as for the provision of further insights into the metabolic effects mediated by signaling molecules.

All classes of biological compounds, from genes through mRNA to proteins and metabolites, can be analyzed by the respective "omic" approaches, namely, genomics, transcriptomics, proteomics, or metabonomics. Such an "omic" approach leads to a broader view of the complex biological system, including the pathology of diseases. Indeed, while the data obtained from genomics may explain the disposition of diseases (*i.e.*, increased risk of acquiring a

certain disease), several other mechanisms that are not gene mediated may be involved in the onset of disease. Moreover, a single gene can be processed to result in several different mRNAs or proteins, which directly determine different cellular functions. Variations in metabolite fluxes, which may be taken as the downstream result of changes in gene expression and protein translation, may be expected to be amplified relative to changes in the transcriptome and proteome. However, time-dependent measurements and determinations of metabolite content at a single time-point can be misleading as these fluxes vary quickly. Therefore, while genomics/transcriptomics enables assessments of all potential information, proteomics enables us to assess the programs that are actually executed, and metabonomics will mostly display the results of such executions.

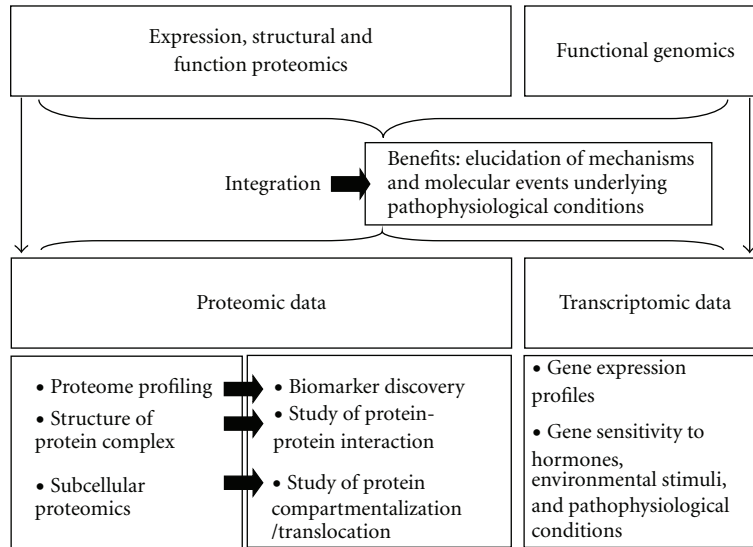


FIGURE 1: Categories and potential applications of proteomics and benefits of integration of proteomics and transcriptomics in the study of complex biological systems.

In the postgenomic era, functional analysis of genes and their products constitutes a novel and powerful approach since the expression levels of multiple genes and proteins can thereby be analyzed simultaneously, in both health and disease (Figure 1). Among the techniques used in functional genomics, both DNA microarrays [1–3] and classical and ongoing proteomic approaches (finalized to protein separation and identification) [4–6] hold great promise for the study of complex biological systems and have applications in molecular medicine. These technologies allow high-throughput analysis as they are complementary to each other, and they may lead to a better understanding of the regulatory events involved in physiological, and disease, processes. Proteins are excellent targets in disease diagnostics, prognostics, and therapeutics. Consequently, proteomic approaches (such as two-dimensional gel electrophoresis (2D-E), two-dimensional liquid chromatography (2-DL), and mass spectrometry (MS)), which allow the simultaneous measurement and comparison of the expression levels of hundreds of proteins, represent powerful tools for (a) the discovery of novel hormone/drug targets and biomarkers and (b) studies of cellular metabolism and protein expressions [7, 8]. Increasingly, proteomic techniques are being adopted—in particular, to avoid the limitations inherent in the more classical approaches—to solve analytical problems and obtain a more comprehensive identification and characterization of molecular events associated with pathophysiological conditions (Figure 1).

In this paper, we will discuss a variety of mainly recent transcriptomic- and proteomic-based studies that have provided a comprehensive mechanistic insight into two very complex biological phenomena, namely, age-associated muscle sarcopenia and thyroid-hormone signaling. Moreover, as mitochondria are severely affected during ageing and it is generally believed that dysfunctions of mitochondria also cause ageing and muscle sarcopenia, we will also discuss

proteomic analysis of the alterations in rat skeletal muscle mitochondria caused by ageing.

2. Ageing Sarcopenia

Ageing, one of the most complex biological phenomena, is a multifaceted process in which several physiological changes occur at both the tissue and the whole-organism level. Indeed, the age-associated decline in the healthy functioning of multiple organs/systems leads to an increased incidence of, and mortality from, diseases such as type II diabetes mellitus, neurodegenerative diseases, cancer, and cardiovascular disease [9].

One of the major engagements of gerontology is the understanding of the complex mechanisms involved in ageing at the molecular, cellular, and organ levels that would facilitate our understanding of age-related diseases. Research in this area has accelerated with the application of high-throughput technologies such as microarrays. To judge from such studies, several metabolic pathways are affected during ageing, and the picture becomes even more complex when we realize that most of them are interconnected.

Sarcopenia, the age-related decline in skeletal muscle mass and strength, is a major complication in the elderly [10, 11]. Since skeletal muscle represents the most abundant tissue in the body, fiber degeneration has severe consequences for posture, movement, the overall integration of metabolism, and heat homeostasis [12]. Although various metabolic and functional defects in ageing muscle have been described over the last decade, senescence-related muscle wasting is not well understood at the molecular and cellular levels. Consequently, no effective treatment has yet been developed to counteract age-related fiber degeneration [13].

Over the last decades, an attractive approach to the understanding of the molecular mechanisms involved in

TABLE 1: Summary of the models used and of the major findings obtained by applying microarray technologies to the study of ageing skeletal muscle.

| Authors | Experimental model | Number of analyzed genes | Identified affected pathways |
|-----------------------------|--|--------------------------|--|
| Mouse | | | |
| Lee et al., 1999 [25] | Studied tissue: aged skeletal muscle. | 6347 | Stress response, energy metabolism. |
| Rat | | | |
| Sreekumar et al., 2002 [28] | 12-months-old Sprague-Dawley rats. Studied tissue: gastrocnemius muscle. | 800 | Energy metabolism, signal transduction, stress response, glucose/lipid metabolism, and structural/contractile function. |
| Altun et al., 2007 [29] | 4- and 30-months-old rats. Studied tissue: gastrocnemius muscle. | 6240 | Redox homeostasis, iron load, regulation of contractile proteins, glycolysis, and oxidative phosphorylation. |
| Lombardi et al., 2009 [26] | 3- and 24-months-old rats. Studied tissue: gastrocnemius muscle. | 1176 | Energy metabolism, mitochondrial pathways, myofibrillar filaments, and detoxification. |
| Human | | | |
| Welle et al., 2003 [23] | 21–27 yr of age and 67–75 yr of age. Studied tissue: vastus lateralis muscle. | 44 000 | Cell cycle/cell growth, inflammation, signal transduction, protein metabolism, transcription, stress response/DNA repair, energy metabolism, and hormonal. |
| Welle et al., 2004 [19] | 20–29 yr of age and 65–71 yr of age, women. Studied tissue: vastus lateralis muscle. | 1000 | Stress response/DNA repair, energy metabolism. |
| Zahn et al., 2006 [21] | 16 and 89 yr of age. Studied tissue: skeletal muscle. | 54 675 | Electron transport chain, cell cycle/cell growth, extracellular matrix, chloride transport, complement activation, ribosomes. |

sarcopenia has been to screen all genetic pathways at one time, by the use of full-genome oligonucleotide chips, as well as the entire protein complement, by the use of using proteomic tools. These approaches, when applied together to the multifactorial muscle-wasting pathology observed in aged fibers, have allowed the identification of a variety of molecular and cellular changes. These include increased oxidative stress, mitochondrial abnormalities, disturbed microcirculation, hormonal imbalance, incomplete ion homeostasis, denervation, and impaired excitation-contraction coupling, as well as a decreased regenerative potential (see, Sections 2.1 and 2.2). In addition, altered posttranslational modifications, such as tyrosine nitration, glycosylation, and phosphorylation, were recently described as occurring in an age-related manner in numerous skeletal muscle proteins (see, Section 2.3).

2.1. Transcriptomic Analysis Pertaining to Ageing Skeletal Muscle. Knowledge of differential mRNA expressions (i.e., the transcriptome) constitutes the first essential level of information when studying integrated cell functions and cell-specific gene-expression profiles. Since the development of DNA microarray technology, it has been possible to survey thousands of genes in parallel, thereby obtaining information regarding transcriptional changes on a global scale. Such

an approach has been used to study the transcriptional alterations induced by ageing both in rodent models and in humans. Ageing-related transcriptomic studies have been performed both on home-spotted microarrays containing about 4000–6000 transcripts [14–16] and on commercial Affymetrix microarrays with from 12000 [17, 18] to about 54000 [19–26] transcripts on each array.

Concerning studies on humans, the design commonly used involved a cross-sectional comparison of young and elderly healthy individuals, with about eight individuals maximum per group, or an analysis of individuals across an age-range. In these studies, several pathways were highlighted by genes that were differentially expressed between young (19–29 years) and elderly (65–85 years) individuals [14], including genes involved in energy metabolism, the cell cycle, signal transduction, and DNA repair [19–22].

Biological pathways found to be changed with age in human skeletal muscle are listed in Table 1 and schematized in Figure 2. They include genes involved in the mitochondrial electron transport chain, cell cycle, and extracellular matrix. Zahn et al. [21], by comparing the transcriptional profile of ageing in muscle with previous transcriptional profiles of ageing in the kidney [22] and brain [17], found a common signature for ageing in these diverse human tissues. This common ageing signature consists of six genetic pathways; four display increased expression (genes in the extracellular

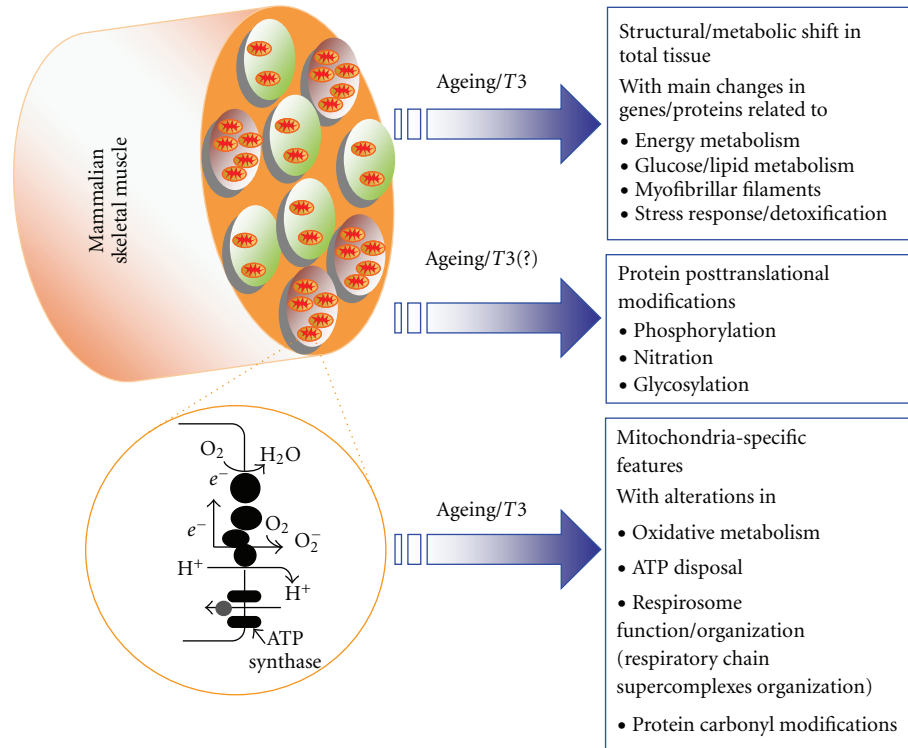


FIGURE 2: Integrated overview of the main ageing/T3-induced transcriptomic and proteomic alterations occurring in mammalian skeletal muscle. Schematic representation of the common events and mechanisms underlying the response of skeletal muscle to either ageing or T3 according to data obtained from cDNA microarray/proteomic-based studies in various mammalian models of ageing and thyroid state (mouse, rat, and human) (for details, see text and Tables 1, 2, 3, and 4).

matrix, genes involved in cell growth, genes encoding factors involved in complement activation, and genes encoding components of the cytosolic ribosome) and two display decreased expression in the aged muscle. These results indicate that those pathways, but not necessarily individual genes, are common elements in the age-related expression changes among different tissues [21]. This may imply that in addition to tissue-specific effects, a common ageing signature may be found in any tissue that reflects the age of the whole organism. This could have major implications for human epidemiological studies, for which frequently only blood is available.

Transcriptomic studies have been performed in laboratory animals using commercially available microarrays. As in humans, the main design used for measuring changes related to chronological age is a comparison between young and old individuals. These studies are usually performed on inbred strains, and so the variation between individual animals is smaller than among human individuals. The range of tissues studied includes liver, heart, skeletal muscle, aorta, and brain. Across all species, and in most experimental designs there is an influence of gender on ageing features and gene expressions [24]. Biological pathways found to be changed with age in rodent (mouse and rat) skeletal muscle are listed in Table 1.

Transcriptomic analysis of gastrocnemius muscle from 5- and 30-month-old male C57BL/6 mice revealed that

ageing resulted in a differential gene expression pattern [25]. Of the genes that increased in expression with age, 16% were mediators of stress responses, including heat shock-response genes, oxidative stress-inducible genes, and DNA damage inducible genes (Table 1). Genes involved in energy metabolism were downregulated with ageing, including genes associated with mitochondrial function and turnover. This suppression of metabolic activity was accompanied by a concerted decline in the expressions of genes involved in glycolysis, glycogen metabolism, and the glycerophosphate shunt (Table 1). Ageing was also characterized by the induction of genes involved in neuronal growth and large reductions in the expressions of biosynthetic enzymes.

We recently performed a transcriptomic study on gastrocnemius muscle from rats aged 3 months (young) and 24 months (old) via a DNA array [26]. Transcript levels for genes associated with cellular damage were elevated in the older muscle, while transcript levels for genes involved in energy metabolism were reduced with age. Among the biological classes of transcripts significantly decreased by ageing, there were transcription factors as well as ribosomal proteins, indicative of a lower transcription/translation activity in old than in young skeletal muscle (Table 1). In agreement with previous microarray studies on the skeletal muscles of humans and rodents [18, 25, 27], we found that ageing is accompanied by a decline in the expressions of genes associated with energy-metabolism functions [26] (Table 1).

Alterations in oxidative phosphorylation were revealed by decreased expression levels of cytochrome c oxidase, ATPase subunit, and carbonic anhydrase III [26]. The capacity of mitochondria to import and oxidize fatty acids would presumably be impaired since the mRNA levels for acylCoA synthase as well as carnitine palmitoyl transferase 1 (CPT 1 β) were reduced during ageing. Downregulation of the AK1 isoform of adenylate kinase [26] points toward decreased AMP production and hence decreased activity of AMP-activated protein kinase (AMPK), an inducer of glucose and fatty acid uptake and fatty acid oxidation. Gastrocnemius muscle from the old rats revealed increased expressions of various factors involved in muscle differentiation toward the “slow” phenotype (type I; oxidative fibers), including p27kip and muscle LIM protein (MLP) [26] (Table 1). As a whole, the above data support an ageing-induced shift towards moderate fat burning.

Ageing has been found to increase the mRNA levels of scavengers of free radicals such as phospholipids hydroperoxide glutathione peroxidase and the cytosolic superoxide dismutase Cu/Zn SOD (SOD1). In addition, 14-kDa ubiquitin-conjugating enzyme E2 mRNA (a component of the complex that adds ubiquitin to target proteins, making them capable of destruction by the proteasome machinery) and both proteasome subunit C5 and proteasome delta subunit precursor were downregulated in aged muscle. Since the proteasome is the major proteolytic complex responsible for the selective degradation of oxidized proteins, these data point toward a defective action of the proteasome. With regard to the lysosomal pathway of protein breakdown, cathepsin L (which acts upstream of the ubiquitin-proteasome system) was also downregulated in aged muscle [26], once again supporting a decline in proteolysis during ageing.

The above studies have been successful in elucidating some of the transcriptional changes that occur with age in muscle, as well as in other tissues, and in providing insights about age-related changes common to animals with different lifespans.

2.2. Proteomic Analysis Pertaining to Ageing Skeletal Muscle. Proteomic analysis has proved valuable in informing our understanding of the molecular mechanisms involved in the ageing process through the identification both of changes in protein levels and of various posttranslational modifications such as phosphorylation [30], nitration [31, 32], and glycosylation [33] that progress with age. In order to identify novel biomarkers of age-dependent skeletal muscle sarcopenia, mass spectrometry-based proteomics has been applied to the study of global muscle protein expression patterns. Mass spectrometric peptide fingerprinting, chemical peptide sequencing, electrophoretic mobility comparison using international 2-D gel databanks, and/or large-scale immunoblot analysis are among the most frequently utilized techniques.

Over the last years, several proteomic studies have catalogued the accessible skeletal muscle protein complement from various species and also investigated changes in protein expression levels in the sarcopenia of old age. The

data obtained so far have furnished databanks that form an important prerequisite for future large-scale proteomic investigations into muscle ageing.

Table 2 lists proteomic studies on age-related changes in skeletal muscle in rodent and human models of ageing. Although the lists of individual proteins found to be affected by the ageing process differ considerably between individual proteomic surveys, the main trend of the altered proteins involved in energy metabolism, cellular signaling, the stress response, cytoskeleton, and contraction shows agreement among the various studies. Gelfi et al. [34] performed a quantitative differential analysis of muscle protein expression in elderly and young subjects using a 2-D DIGE approach. The main difference observed in the elderly group included downregulation of regulatory myosin light chains, particularly the phosphorylated isoforms, a higher proportion of myosin heavy chain isoforms 1 and 2A, and enhanced oxidative and reduced glycolytic capacities.

Proteomic profiling of rodent muscle during ageing has been performed in several studies, resulting in the identification of a large cohort of sarcopenic biomarkers (for a schema, see Figure 2).

Age-dependent differential regulation in rodent muscle has been identified for several glycolytic and mitochondrial enzymes, which are important for energy metabolism. The glycolytic enzymes triosephosphate isomerase, glyoxalase I, and β -enolase were downregulated in aged muscle. Other features indicating perturbation of energy metabolism were downregulation of creatine kinase, of pyruvate kinase, and of the NADH-shuttle glycerol 3-phosphate dehydrogenase. At the mitochondrial level, key enzymes such as isocitrate dehydrogenase, cytochrome c oxidase, ATP synthase β subunit, and pyruvate dehydrogenase were all decreased in ageing muscle whereas there was an upregulation of aldehyde dehydrogenase [26, 29, 35–38].

Differential proteomic analyses have revealed that ageing is associated with perturbations of the myofibrillar network [26, 29, 35–38]. Notably, there is a downregulation of several isoforms of myosin long chain and of alpha-actin, as well as a differential expression of their major regulators. In contrast to the downregulation of myofibrillar components, old muscles display upregulation of many proteins of the intermediate filament, microtubules and microfilament cytoskeleton, among which are B-tubulin, desmin, and gelsolin. This suggests a mechanism affecting the cytoskeleton that compensates for perturbations in myofibrillar structure and so prevents extensive damage to the myofibers. Muscle ageing is also associated with the differential expression of enzymes implicated in the detoxification of cytotoxic products. The cytoplasmic Cu/Zn superoxide dismutase (SOD1) and H ferritin isoform, as well as the levels of glutathione transferase and mitochondrial aldehyde dehydrogenase, were found to be increased in older rats, while evidence of age-associated protein misfolding was provided by the upregulation of molecular chaperones (including HSP 27 and disulfide isomerase ER60) [26, 29, 35–38].

Most of the proteins identified by differential proteomics were previously unrecognized in ageing skeletal muscle.

TABLE 2: Summary of the models used and of the major findings obtained by applying proteomic approaches to the study of the ageing skeletal muscle.

| Authors | Experimental model | Proteomic analysis | Identified affected pathways and major findings |
|-----------------------------|--|--|--|
| Chang et al., 2003 [39] | 18-months-old C57B16 mice. Studied tissue: skeletal muscle. | Mouse Two-dimensional polyacrylamide gel electrophoresis. | Reproducibility of the 2-D PAGE system. |
| Cai et al., 2000 [35] | 12-, 18-, and 24-months-old rats. Studied tissue: extensor digitorum longus muscle and soleus muscle. | Rats Two-dimensional gel electrophoresis. | Analysis of aqueous proteins of skeletal muscle during aging. |
| Cai et al., 2001 [40] | 8-, 18-, and 24-months-old rats. Studied tissue: extensor digitorum longus muscle and soleus muscle. | Two-dimensional gel electrophoresis. | Analysis of parvalbumin expression in rat skeletal muscles. |
| Kanski et al., 2003 [31] | 4- and -24 months old Fisher 344 rats and -6 and -34 months old Fisher 344/BN F1 rats. Studied tissue: skeletal muscle. | 2-D gel electrophoresis, Western blot analysis, MALDI-TOF MS and ESI-MS/MS analysis. | Age-dependent nitration in muscle energy metabolism. |
| Piec et al., 2005 [36] | 7-, 18- and 30-months-old LOU/c/jall rats. Studied tissue: gastrocnemius muscle. | Two-dimensional gel electrophoresis, MALDI-ToF MS analyses, and immunoblotting. | Myofibrillar regulatory proteins, signal transduction, cytosolic and mitochondrial energy metabolisms, oxidative stress, detoxification, and RNA metabolism. |
| Kanski et al., 2005 [32] | 34-months-old Fisher 344/Brown Norway F1 hybrid rats. Studied tissue: skeletal muscle. | 2-D gel electrophoresis, Western Blot analysis, MALDI-TOF and NSI-MS/MS analysis. | Proteomic analysis of protein nitration. |
| Dencher et al., 2006 [41] | Studied tissue: skeletal muscle. | Blue-native/colorless-native gel electrophoresis, 2D-SDS-PAGE and MALDI MS. | Cellular dysfunction, ageing, and cellular death. |
| O'Connell et al., 2007 [37] | 3- and 30-months-old rat. Studied tissue: gastrocnemius muscle | Two-dimensional gel electrophoresis, MALDI-ToF, DALT-Twelve gel electrophoretic separation system, 2-D immunoblotting. | Proteomic profiling of senescent fibres: stress response, contractile apparatus, and metabolic regulation. |
| Altun et al., 2007 [29] | 4- and 30-months old rats. Studied tissue: gastrocnemius muscle. | Two-dimensional gel electrophoresis, MALDI-ToF/ToF, MALDI-MS/MS, ESI-IC-MS/MS and Western Blot analysis. | Redox homeostasis, iron load, regulation of contractile proteins, glycolysis, and oxidative phosphorylation. |
| O'Connell et al., 2008 [33] | 3- and 30-months old rats. Studied tissue: gastrocnemius muscle | Two-dimensional gel electrophoresis, MALDI-ToF MS analysis. | Proteomic profiling of senescent fibers. |
| Gannon et al., 2008 [30] | 3- and -30-months old rats. Studied tissue: gastrocnemius muscle. | Two-dimensional gel electrophoresis, MALDI-ToF MS analysis. | Phosphoproteomic analysis of aged skeletal muscle. |
| Feng et al., 2008 [42] | 12- and 26-months-old Fischer 344 rats. Studied tissue: soleus, semimembranosus, plantaris, extensor digitorum longus, and tibialis anterior muscles. | SDS-polyacrylamide gel electrophoresis, μ LC-ESI MS/MS analysis and Ingenuity Systems Analysis. | Carbonyl modifications, cellular function and maintenance, fatty acid metabolism, and citrate cycle. |
| Lombardi et al., 2009 [26] | 3- and 24-months-old rats. Studied tissue: gastrocnemius muscle. | Two-dimensional gel electrophoresis, Blue-Native PAGE, and MALDI-ToF MS analysis. | Energy metabolism, mitochondrial pathways, myofibrillar filaments, and detoxification. |
| Cobon et al., 2002 [43] | 56-79 yr of age. Studied tissue: vastus lateralis muscle. | Human Two-dimensional polyacrylamide gel electrophoresis and MALDI-TOF MS. | Human aged skeletal muscle protein profile. |
| Gelfi et al., 2006 [34] | Elderly and young subjects. Studied tissue: vastus lateralis muscle. | Two-dimensional difference gel electrophoresis, SDS-PAGE and ESI-MS/MS. | Elderly group: downregulation of regulatory myosin light chains, (phosphorylated isoforms), higher proportion of myosin heavy chain isoforms 1 and 2A, and enhanced oxidative and reduced glycolytic capacity. |

Their identification has not only provided further insight into the potential mechanisms of ageing, but may lead to the development of biomarkers of sarcopenia [26, 29, 35–38].

2.3. Proteomic Analysis Pertaining to Ageing Skeletal Muscle: Analyses of Protein Phosphorylation, Nitration, and Glycosylation. Since posttranslational modifications are key modulators of protein structure, function, signaling, and regulation, various subdisciplines of proteomics have emerged that focus on the cataloguing and functional characterization of proteins with extensively modified side chains [57]. In aged skeletal muscle, proteins undergo considerable changes in their posttranslational modifications [58]. These include, among others, phosphorylation, nitration, and glycosylation. Phosphorylation represents one of the most frequent peptide modifications [59], and abnormal phosphorylation is associated with various pathologies. A recent phosphoproteomic survey of aged muscle detected increased phosphorylation levels for myosin light chain 2, tropomyosin α , lactate dehydrogenase, desmin, actin, albumin, and aconitase [30]. In contrast, decreased phospho-specific dye binding was observed for cytochrome c oxidase, creatine kinase, and enolase. Thus, ageing-induced alterations in phosphoproteins appear to involve the contractile machinery and the cytoskeleton, as well as cytosolic and mitochondrial metabolism.

The nitration of protein tyrosine residues represents an oxidative and important posttranslational modification occurring under nitrative/oxidative stress during biological ageing. Comprehensive proteomic studies have identified an age-related increase in the nitration of numerous skeletal muscle proteins. These include enolase, aldolase, creatine kinase, tropomyosin, glyceraldehyde-3-phosphate dehydrogenase, myosin light chain, pyruvate kinase, actinin, actin, and the ryanodine receptor [31, 32]. The nitration of these essential muscle proteins may therefore be a significant causative factor in the age-related decline in muscle strength [31, 32].

Glycosylation is one of the most frequent posttranslational modifications found in proteins, and it plays a central role in cellular mechanisms in both health and disease [60]. Oligosaccharide attachment represents a common protein modification that influences the folding of the nascent peptide chain and the stability of glycoproteins, modifies enzymatic activity, controls protein-secretion events, presents critical information about the cellular targeting of a newly synthesized protein, and provides specific recognition motifs for other proteins in cell-cell interactions [61]. The identified muscle components belong mostly to the family of metabolic enzymes. They included glycolytic enzymes, such as pyruvate kinase, enolase, phosphoglycerate kinase, aldolase, glyceraldehyde-3-phosphate dehydrogenase, and phosphoglycerate-mutase, aconitase, carbonic anhydrase, and creatine kinase [33].

These data confirm that the sarcopenia of old age represents a complex neuromuscular pathology that is associated with drastic changes not only in the abundance, but also in the structure of key muscle proteins (Figure 2).

2.4. Proteomic Analysis Pertaining to Ageing Skeletal Muscle Mitochondria. Analysis of the protein profile of mitochondria, and of the changes in it that occur with age, represents a promising approach to the unraveling of the mechanisms involved in ageing. Although the role of mitochondria was long thought to be restricted to an influence on fuel metabolism, the importance of the activity of these organelles has recently been extended to the regulation of developmental/ageing processes [62]. Mitochondria have their own genome (mt-DNA) and specific mechanisms for replication, transcription, and protein synthesis. However, in terms of protein composition they are “hybrid” organelles resulting from the coordinated expression of the nuclear and their own genome. A bidirectional flow of information allows the two kinds of subcellular compartments to communicate with each other under the control of metabolic signals and several signal-transduction pathways that function across the cell. These pathways are differentially regulated by environmental and developmental signals, and under patho/physiological conditions, they allow tissues to adjust their energy production according to different energy demands possibly modulating/altering the mitochondrial phenotype. It is now beyond doubt that mitochondria are severely affected by ageing, and it is generally believed that dysfunctions of mitochondria trigger key steps in the ageing process [62].

Mitochondrial proteomes (mitoproteomes) are currently under vigorous investigation by way of both structural and comparative proteomics. In particular, we would like to emphasize the value of comparative proteomics as a tool capable of providing us with valuable information on mitochondrial physiology and on the role of these organelles in ageing muscle. First, mitochondria can be highly purified, leading to simplified 2D gels, which greatly facilitates the analysis and detection of less-abundant proteins. Second, mitochondrial proteins are generally distributed across wide ranges of both pH and molecular mass on 2D gels, leading to accurate protein resolution with only a few protein-spot overlaps. Third, most of the mitochondrial membrane-protein complexes exhibit soluble subunits that can be analyzed on 2D gels even though the hydrophobic subunits aggregate. Various detection methods are already available that allow us to monitor quantitative changes in the proteome. Of these, 2-DE-based methods appear quite promising, with isoelectric focusing (IEF), BN-SDS, and benzyldimethyl-*n*hexadecylammonium chloride (16-BAC)-PAGE at the forefront. However, application of IEF is restricted to proteins that are not highly hydrophobic or have no extreme isoelectric points. Indeed, by the use of classical 2D-E it is difficult to detect very acidic or very basic proteins or to distinguish small changes in the expressions of weakly expressed proteins. On the other hand, BN-SDS-PAGE deals efficiently with even hydrophobic membrane proteins, although some compromises in resolution have to be made [63]. Another advantage of the BN-PAGE system is the conservation of protein-protein interactions, enabling simultaneous elucidation of multimeric and multiprotein assemblies of soluble and membrane proteins [64]. Such a procedure might be a

viable alternative to other methods, such as yeast two-hybridization.

Comparative transcriptomic and proteomic studies have been initiated to determine global changes in mitochondria from young versus aged skeletal muscle [26, 39, 41, 62, 65–68]. Native-difference gel electrophoresis (DIGE) is an approach that facilitates sensitive quantitative assessment of changes in membrane and soluble proteins. Recently, O'Connell et al. [68] analyzed the mitochondria-rich fraction from aged rat skeletal muscle by DIGE. This proteomic analysis showed a clear age-related increase in key mitochondrial proteins, such as NADH dehydrogenase, the mitochondrial inner membrane protein mitofilin, peroxiredoxin isoform PRX-III, ATPase synthase, succinate dehydrogenase, mitochondrial fission protein Fis1, succinate-coenzyme A ligase, acyl-coenzyme A dehydrogenase, porin isoform VDAC2, ubiquinol-cytochrome c reductase core I protein, and prohibitin [68].

To gain deeper insights both into ageing mechanisms and into the resulting mitoproteome alterations, mitochondria have been studied by the blue-native gel approach, both with respect to protein abundance and the supramolecular organization of OXPHOS complexes [26].

The profiles obtained for muscle crude mitochondria from young and old rats—after detergent extraction with either dodecylmaltoside or digitonin, and subsequent BN-PAGE—have been reported by us within the past years [26]. The use of dodecylmaltoside allows individual resolution of the respiratory complexes. Our densitometric analysis revealed that gastrocnemius muscle mitochondria from old rats, versus those from young rats, contained significantly lower amounts of complex I (NADH:ubiquinone oxidoreductase), V (FoF1-ATP synthase), and III (ubiquinol:cytochrome c oxidoreductase) (–35%, –40%, –25%, resp.). The same mitochondria, on the other hand, contained a significantly larger amount of complex II (succinate: ubiquinone oxidoreductase) (+25%) and an unchanged amount of complex IV (cytochrome c oxidase, COX). The use of a combination of BN-PAGE and catalytic staining allowed us to detect reduced activity of all the complexes in ageing muscle. The observed reductions in the activities of respiratory complexes I, III, and V reflected lower protein levels, but the reduction in complex II activity was associated with an increase in the amount of the same complex. To elucidate whether the ageing process also alters the functional/structural organization of the respiratory chain in terms of the assembly of supercomplexes, mitochondria were extracted using the mild detergent digitonin since this extensively retains inner mitochondrial membrane supercomplexes [69]. In both young and old mitochondria, monomeric complex I and dimeric complex III were significantly reduced versus dodecylmaltoside solubilization. However, the missing amounts were found to be assembled in two major supercomplexes, a and b, and two minor ones, c and d, all within the molecular mass range of 1500–2100 kDa. The supercomplex profile of the older rats was significantly modified, band a being less represented in the profile than the heavier supercomplexes, such as bands c and d. A significant increase was detected in

the supramolecular assembly of respiratory chain complexes into respirosomes (each one being formed by complex I assembled with a dimeric complex III and a variable copy number of complex IV, represented by bands c and d). Possibly, this could be a compensatory mechanism that, in ageing muscle, is functionally directed towards substrate channeling and catalytic enhancement advantaging. Indeed, mitochondrial oxidative phosphorylation seems to be more efficient in aged than in young skeletal muscle, since old rats exhibited an increased respiratory control ratio that was attributed principally to a reduction in the reactions able to dissipate the proton motive force not associated with ATP synthesis. This could be interpreted as a compensation for the reduced level and activity of F1F0-ATP synthase.

The above data point up the ability of skeletal muscle to face the consequences of ageing in a metabolically economic way and highlight the occurrence of structural and metabolic adaptations. A comparison between these two studies [26, 68] each employing different proteomic approaches leads to the conclusion that beyond the expression/abundance changes in proteins, an insight can be obtained about the structural and functional heterogeneity in a given mitoproteome.

Another possible protein modification in skeletal muscle mitochondria, possibly contributing to its functional decline with age, is carbonylation, which can be considered an oxidative modification that may render a protein more prone to degradation. Feng et al. [42] recently identified mitochondrial proteins that were susceptible to carbonylation in a manner that was dependent on muscle type (slow-versus fast-twitch) and on age. Carbonylated mitochondrial proteins were more abundant in fast-twitch than in slow-twitch muscle. Twenty-two proteins displayed significant changes in carbonylation state with age, the majority of these increasing in their amount of carbonylation. Ingenuity pathway analysis revealed that these proteins belong to various functional classes and pathways, including cellular function and maintenance, fatty acid metabolism, and the citrate cycle. This study [42] provides a unique catalogue of promising protein targets deserving further investigation because of their potential role in the decline exhibited by ageing muscle. Since carbonylation is not repairable, this modification may, however, be of special importance in directing the affected protein to the path toward degradation.

Of note, in view of the importance of the functional mitochondrial membrane compartmentalization, together with proteomic approaches, lipidomic ones would be desirable to gain further insight into the understanding of the modification of lipids either as membrane components or energy store following aging processes.

3. Thyroid Hormones and Thyroid States

Thyroid hormones [THs; 3,5,3',5'-tetraiodo-L-thyronine, otherwise known as thyroxine (T4), and 3,5,3'-triiodo-L-thyronine (T3)] are essential for the regulation both of energy metabolism and of development and growth in all vertebrates. In humans, the early developmental role of THs

is vividly illustrated by the distinctive clinical features of cretinism, as observed in iodine-deficient areas. In adults, the primary effects of THs are manifested by alterations in metabolism. Even subclinical hyper- and hypothyroidism can have important consequences, such as atherosclerosis, obesity, and alterations in bone mineral density and heart rate [70, 71]. The effects induced by THs in the regulation of metabolism include changes in oxygen consumption and in protein, carbohydrate, lipid, and vitamin metabolism. Hyperthyroidism is associated with an increase (calorigenic effect), while hypothyroidism is associated with a decrease in metabolic rate. Of particular note, is that the number and the complexity of the clinical features of hyperthyroidism and hypothyroidism emphasize the pleiotropic effects of THs on many different pathways and target organs. Although great efforts have been made to elucidate the signaling pathways underlying the physiopathological effects of thyroid hormones, the network of factors and cellular events involved, as well as the possible role of derivatives of THs, is complicated and incompletely understood, as is the ultimate effect of THs on tissue transcriptomes and proteomes.

3.1. The Complexity of Action of Thyroid Hormones: An Overview. Most thyroid-related, direct genomic actions leading to protein changes appear to be attributable to T3. The mechanism of action that has gained general acceptance for this iodothyronine involves the binding of specific nuclear receptors (TRs) to thyroid hormone response elements (TREs) in target genes [72]. Within the nucleus, TRs dimers (hetero- or homodimers) bind to TREs and modulate gene activity by either silencing or activating transcription by recruitment of either corepressor or coactivator complexes, depending on the absence or presence of thyroid hormone [73–78]. In mammals, two genes encoding TRs have been characterized, *c-erb A α* and *c-erb A β* [79–81], and these encode several proteins (α and β isoforms) with different binding properties and patterns of tissue expression. For example, *c-Erb A β 1* is expressed across a wide range of tissues, while *c-ErbA β 2* is found almost exclusively in the pituitary, where it inhibits thyrotrophin (TSH) α - and β -subunit gene transcription [82] by binding to negative TREs present on these genes [83, 84]. New information on the mechanisms of action of THs have been obtained from TR gene knockout (KO) and knock-in studies [85].

In terms of cellular effects, theories proposed so far to explain the actions of THs on metabolic rate also include mechanisms such as: uncoupling of oxidative phosphorylation, stimulation of energy expenditure by activation of Na⁺-K⁺ ATPase activity, and direct modulation by THs of transporters and enzymes located within the plasma membrane and mitochondria [86–89]. Moreover, T3-mediated nuclear gene expression leads in turn to coordinated and synergistic effects on the mitochondrial genome [90]. Actually, it has been postulated that T3's actions on this genome are achieved through both an induction of nuclear-encoded mitochondrial factors and a direct binding of T3 to specific ligand-dependent mitochondrial transcription factors [90–94]. These last are nuclear-receptor homologs and are

thought to act on a number of mt-DNA response elements [95]. Indeed, T3 directly regulates the mitochondrial genes encoding ATPase subunit six [96], NADH dehydrogenase subunit three [97], and subunits of cytochrome-*c*-oxidase [98].

The complexity of action of T3 is broadened by the existence of nongenomic or TRE-independent actions, which have been extensively described and are now accepted [99]. Importantly, these can be either independent or dependent on the binding of T3 to TRs. They occur rapidly and are unaffected by inhibitors of transcription and protein synthesis [90, 93, 100–102]. Nongenomic actions of thyroid hormones have been described at the plasma membrane, in the cytoplasm, and within cellular organelles ([100] and references therein). These actions include modulations of Na⁺, K⁺, Ca²⁺, and glucose transport, activations of PKC, PKA, and ERK/MAPK, and regulation of phospholipid metabolism via activations of PLC and PLD [103], and they can be independent of the presence of nuclear TRs and mediated even by TH analogs [102]. For example, it has recently been shown that cytosolic TR β can interact with the p85 subunit of PI3K and thereby activate the PI3K-Akt/PKB signaling cascade [99, 104]. Moreover, it has been shown that THs activate the MAPK cascade and stimulate angiogenesis via their binding to integrin α V β 3 [100]. Importantly, it appears now to be well established that an interplay exists between the genomic and nongenomic actions when gene expression is regulated by the TR-T3 complex and the activity of the enzyme is modulated by a nongenomic process [100].

3.2. Transcriptomic Analysis Pertaining to the Actions of Thyroid Hormones. Although the molecular actions of THs have been thoroughly studied, their pleiotropic effects are not well understood and appear to be mediated by complex changes in the expressions of numerous, but still largely unknown, target genes in various tissues. DNA microarrays have been successfully used to identify T3-target genes in mouse, rat, and human tissues, cell lines, and tumors. Actually, pioneering systematic studies in the search for T3-target genes were performed by Seelig and coworkers as long ago as 1981 [105].

Feng et al. [44] first applied cDNA-microarray technology to the study of the *in vivo* T3 regulation of hepatic genes in the mouse. They identified new T3-target genes, the majority of which had not previously been reported to be regulated by the hormone. Surprisingly, many of these target genes were negatively regulated. The identity of the genes indicated that multiple cellular pathways are actually affected by T3, including glycogenolysis, gluconeogenesis, lipogenesis, cell proliferation, apoptosis, the action of insulin, immunogenicity, and protein glycosylation.

Weitzel et al. [49] detected novel T3-target genes and identified two T3-mediated gene-expression patterns after the administration of T3 to hypothyroid rats. In line with the long-known observation that T3 has profound influences over mitochondrial biogenesis and metabolic balance, the authors reported that numerous genes implicated in metabolic pathways (ANT2, apolipoprotein AIV, HMG-CoA

synthase, and ATP synthase β subunit) are affected by T3, as also are genes associated with a wide variety of cellular pathways (encompassing translation, protein turnover, cell structure, and apoptosis-associated proteins). These observations gave support to the idea that alongside the “classical” pathway of T3-mediated gene regulation (involving thyroid hormone-receptor binding to TREs), there appears to be an additional pathway mediated by transcription factors (such as NRF-1 and PPAR γ) and coactivators (such as the PGC-1 family of coactivators).

Flores-Morales et al. [45] verified the effect of T3 on liver in mice with a targeted mutation in the TR β gene. In accordance with the results of Weitzel [49], they reported that T3 regulates the expressions of functionally different sets of genes in temporally distinct ways. Importantly, using TR $\beta^{-/-}$ animals they also defined a number of T3-responsive genes that are dependent on TR β *in vivo*, thereby opening the way for the use of similar experimental strategies to identify the contributions made by specific transcription factors to the *in vivo* actions of multiple hormones and trophic factors.

Miller et al. [50] identified genes involved in glucose metabolism, biosynthesis, transcriptional regulation, protein degradation, and detoxification that were associated with T3-induced cell proliferation. Of particular significance were the findings that T3 rapidly suppresses the expressions of key regulators of the Wnt signaling pathway and that it suppresses the transcriptional downstream elements of the β -catenin-T-cell factor complex.

With the aim of defining the molecular basis of the target-tissue phenotype related to the hereditary TR β mutations causing resistance to thyroid hormone (RTH), Miller et al. [46] showed that in T3-target tissues such as cerebellum, heart, and WAT in animal models of both RTH and hyperthyroidism, T3 acts primarily to suppress gene expression, and that TR β has a greater modulating effect in the heart than originally thought. Moreover, their comprehensive multi-tissue gene-expression analysis uncovered complex multiple signaling pathways mediating the molecular actions of TR β mutants *in vivo*. It also revealed some T3-independent, but mutant-dependent, genomic responses contributing to those “changes-of-function” present in TR β mutants that are linked to the pathogenesis of RTH.

Dong et al. [48] studied the molecular mechanisms involved in the responses shown by developing mice to disruptions in maternal thyroid-hormone homeostasis. Among differentially expressed genes, Nr4a1 (nuclear receptor subfamily 4, group A, member 1), was upregulated by 3-fold in the hypothyroid juvenile mouse liver, while treatment of HepG2 cells with T3 resulted in its downregulation. A potential thyroid response element -1218 to -1188 bp upstream of the promoter region of Nr4a1 was identified and demonstrated to bind TR α and TR β receptors.

Notably, in recent years microarray approaches have been used to characterize the effects of T3 on gene expression profiles in the postnatal developing brain as well as in the adult mouse/rat brain [106–108].

The effects of THs on gene expression profiles have been studied less intensively in human tissues than in animal above all because of the poor availability and accessibility

of tissue. However, both *in vitro* [51, 52] and *in vivo* [53, 109] studies have been performed. Viguerie et al. [51], who showed that T3 regulates a large repertoire of genes in human adipocytes, provided support for the effect of T3 on catecholamine-induced lipolysis, and suggested downregulation of SREBP1c as a link between hyperthyroidism and insulin resistance. Moreover, in accordance with other array studies, the data showed that thyroid hormone can affect cellular processes such as signal transduction, apoptosis, and inflammatory responses. Moeller et al. [52] identified 91 T3-upregulated and 5 T3-downregulated genes in skin fibroblasts from normal humans. Some of these genes were not previously known to be induced by T3, namely aldolase reductase family 1 C1-3, collagen type VI alpha 3, member RAS oncogene family brain antigen RAB3B, platelet phosphofructokinase, hypoxia-inducible factor-1 alpha, and enolase 1 alpha. Importantly, these genes have a variety of regulatory functions in both development and metabolism.

Clément et al. [53] studied the effects of thyroid hormone on human skeletal muscle *in vivo*. Their data not only helped to explain the effects of T3 on protein turnover and energy metabolism, but also revealed new putative mechanisms extending beyond the classic metabolic effects of the hormone, and importantly, added to our understanding of the permissive effects of T3 on several cellular events (such as signal-transduction cascades, intracellular transport, and tissue remodeling).

Very recently, Visser et al. [54] examined the skeletal muscle transcriptome in thyroidectomized patients being treated for differentiated thyroid carcinoma, and compared it between those who were off or on L-thyroxine replacement. They reported for the first time that in humans as in animals, a large proportion of muscle genes (~43%) is significantly downregulated by L-thyroxine treatment. They also reported significant regulation of the primary transcripts of the noncoding RNAs miR-206 and miR-133b, which are key regulators in muscle differentiation and proliferation and may affect numerous target genes. The potential of T3 to regulate miRs may be of particular importance since this level of control would add an additional layer of complexity by which T3 may regulate cellular processes.

Collectively, these studies (summarized in Table 3) have provided a cornucopia of novel information (schematized in Figures 2 and 3) on the regulation of transcription by THs. However, the intrinsic nature of these studies means that they provide no information concerning the status of the corresponding encoded proteins, and this is particularly relevant because of the influence of thyroid hormone on protein half-life.

3.3. Proteomic Analysis Pertaining to the Actions of Thyroid Hormones. As stated above, overall T3 signaling can be modulated at many levels (i.e., the thyroid hormone-receptor isoforms present in the tissue, the DNA-response element in the regulated gene, the availability of receptor-binding partners, interactions with coactivators and corepressors, ligand availability, mRNA and protein stabilities, protein translocation, and metabolic interference) [72, 90–93]. Consequently, for a deeper investigation of the biological events

TABLE 3: Summary of the models used and of the major findings obtained by applying microarray technologies to the study of THs effects on relevant metabolically active tissues.

| Authors | Experimental model | Treatments | Number of genes | Number of genes affected by T3 | Identified affected pathways and major findings |
|----------------------------------|---|---|---------------------|---|---|
| Feng et al., 2000 [44] | Six-week-old mice. Studied tissue: liver. | Hypothyroidism induced by low-iodine diet supplemented with 0.15% PTU for 4 weeks, then hyperthyroidism induced by single i.p. injection of L-T3 or T4 100 µg/100 g body weight. | 2225 | Of 55 genes identified as target of T3, 41 were negatively regulated. | Glycogenolysis, gluconeogenesis, lipogenesis, proliferation, apoptosis, insulin signaling, immunogenicity, and protein glycosylation. |
| Flores-Morales et al., 2002 [45] | 2 to 3.5-months old (WT and TRβ ^{-/-}) male mice. Studied tissue: liver. | At the onset of the experiment, all groups of animals were provided a low-iodine diet for 14 d to accustom them to the synthetic chow. Hypothyroidism was then induced by inclusion of 0.05% methimazole and 1% potassium perchlorate in the drinking water for 21 d while still on the low-iodine diet. From d 35, one group of animals was injected daily with 5 µg T3 for an additional 5-d period to induce hyperthyroidism. Another group was injected with 5 µg T3 and 5 µg T4; on d 35, the animals in this group were killed 2 hours after the T3/T4 injection. | 4000 | T3 found to regulate more than 200 genes, more than 100 of which were not previously described. 60% of these genes showed dependence on the TRβ gene for T3 regulation. | Rapid or transient effects of T3 on lipogenic genes. Long-term effects of T3 on genes for the mitochondrial respiratory chain transcription factors and protein turnover. |
| Miller et al., 2004 [46] | 8- to 10-week-old (TRβ ^{+/+} and TRβ ^{PV/PV}) male siblings (mice). Studied tissues: cerebellum, heart, and white adipose. | TRβ ^{PV} mice contain a cytosine insertion in exon 10 of the TRβ1 gene at nucleotide position 1,642 of the TRβ1 cDNA that leads to a frameshift of the carboxy-terminal 14 amino acids of TRβ1. For 7 days, T3 (5 µg/mouse/day) was administered by i.p. injection. | 11500 | 163 genes responsive to T3 treatment and 187 genes differentially expressed between TRβ ^{PV/PV} mice and wild-type littermates. | T3 primarily acts to repress gene expression. TRβ has a powerful modulating effect in the heart. Novel physiologic candidates for T3 action are changes in immune-gene expression and in the induction of antiproliferative genes. The relative levels of TR isoforms lead to dramatic differential effects on gene expression. |
| Ventura-Holman et al., 2007 [47] | Murine non-transfected hepatocyte cell line AML 12, expressing endogenous TRs. | RNA obtained from cells incubated for 3 hours or 24 hours +/- 10 nM T3, in the presence of 10% stripped fetal bovine serum. Cells also incubated in the presence of cycloheximide (10 µg/ml) +/- 10 nM T3 for 3 hours to discriminate between primary and secondary responses. | 15000 | 12 genes upregulated and 5 genes downregulated in the presence of T3. | Novel T3 responsive genes were identified. Insights were obtained into the role of T3 in processes such as cholesterol metabolism, bile acid secretion, and oncogenesis. |
| Dong et al., 2007 [48] | Hypothyroid juvenile mice. Studied tissue: liver | Gene expression analyzed in livers of mice rendered hypothyroid by treating pregnant mice from gestational d 13 to postnatal d 15 with 6-propyl-2-thiouracil in drinking water. | approximately 20000 | 96 differentially expressed genes were identified. Of these, 72 genes encode proteins of known function, 15 of which had previously been identified as regulated by TH. | Metabolism, development, cell proliferation, apoptosis, and signal transduction. A potential thyroid response element -1218 to -1188 bp upstream of the promoter region of Nr4a1 was identified and demonstrated to bind TH receptor TRα and TRβ. |
| Weitzel et al., 2001 [49] | Adult male Wistar rats. Studied tissue: liver. | Hypothyroidism induced by i.p. injection of Na ¹³¹ I (250 µCi/100 g body weight) 28 days before the experiments. Hyperthyroidism provoked by i.p. injection of T3 (50 µg/100 g body weight); repeated after 24 hours. Rats killed at 0, 6, 24, and 48 hours after thyroid hormone. | 4608 | Sixty-two of the genes were reproducibly T3-responsive. | Beside the "classical" pathway of T3-mediated gene regulation by TRs binding to TREs, an additional pathway appears to be mediated by transcription factors like NRF-1 and PPARγ and coactivators (like the PGC-1 family of coactivators). |

TABLE 3: Continued.

| Authors | Experimental model | Treatments | Number of genes | Number of genes affected by T3 | Identified affected pathways and major findings |
|--------------------------------|--|---|------------------|---|---|
| Miller et al., 2001 [50] | GC cells (rat pituitary cell-line expressing functional TRs). | Cells were incubated without or with T3 (100 nM) for 1, 3, 6, 12, 24, or 72 hours. At each time-point, cells were harvested for total RNA preparation. | 4400 | 358 responsive genes were identified. 88% had not previously been reported to be modulated by T3. A few genes showed biphasic expression patterns. In total, 203 genes were upregulated and the remainder were downregulated by T3. | Glucose metabolism, biosynthesis, transcriptional regulation, protein degradation, and detoxification in T3-induced cell proliferation. |
| Viguerie and Langin, 2003 [51] | Human adipose tissue obtained from the s.c. abdominal fat depots of Caucasian women for cDNA array and RT competitive PCR experiments. | Surgical adipose tissue samples were dissected from skin and vessels, and cultured adipocytes were obtained. Cultures were treated with T3 (100 nm) or vehicle for 24 hours. Medium-free T3 concentration was measured at 1 and 24 hours after addition of T3 (using RIA kits). | 1 176 | Among the statistically significant changes in mRNA levels of more than 1.3-fold, 13 and 6 genes were positively or negatively regulated, respectively. | Signal transduction, lipid metabolism, apoptosis, and inflammatory responses. |
| Moeller et al., 2005 [52] | Skin fibroblasts of normal individuals. | Human skin was obtained by punch biopsy from three normal individuals and two patients with RTH. Fibroblasts were grown in supplemented with 10% bovine calf serum (BCS). At confluency, the medium was replaced with one containing TH-depleted BCS (TxBCS), obtained from a thyroidectomized calf. For microarrays, incubation with T ₃ was for 24 hours. | more than 15 000 | Microarray analysis identified 148 genes induced by 1.4-fold or more and five genes repressed to 0.7 or less 24 hours after treatment with 2 × 10 ⁻⁹ M T ₃ . Taking into account duplicate genes, these represented 91 up-regulated and five downregulated genes, respectively. | Aldo-keto reductase family 1 C1-3, collagen type VI α3, member RAS oncogene family brain antigen RAB3B, platelet phosphofruktokinase, hypoxia-inducible factor-1α, and enolase 1α genes, previously known to be induced by TH, were identified and validated. These genes have a variety of regulatory functions in development and metabolism. |
| Clément et al., 2002 [53] | Healthy male Caucasian volunteers (22–33 years old). The same investigations were performed on day 0 and day 14. Studied tissue: vastus lateralis muscle by percutaneous biopsies. | Participants took one tablet of 25 µg of T3 three times a day (75 µg per day) for 14 days. | 24 000 | A transcriptional profile of 383 genes regulated by T3 was obtained (381 were upregulated and only two downregulated). | Novel target genes for T3 were identified. They belong to functional classes including transcriptional control, mRNA maturation, protein turnover, signal transduction, cellular trafficking, and energy metabolism. |
| Visser et al., 2009 [54] | Thyroidectomized patients treated for differentiated thyroid carcinoma (DTC) off L-thyroxine replacement. Studied tissue: skeletal muscle | Included were patients who had been diagnosed with DTC and had received initial therapy consisting of near-total thyroidectomy and radioiodine ablation therapy. Four weeks after L-thyroxine withdrawal and 8 wk after subsequent L-thyroxine replacement, patients were admitted to the clinical research unit. A catheter was inserted in a dorsal hand vein to collect blood samples for measurement of serum TSH, free T4 (fT4), and T3. Muscle biopsies were taken from the quadriceps muscle (vastus lateralis). | 54 674 | 607 differentially expressed genes on L-thyroxine treatment were identified, of which approximately 60% were positively and approximately 40% were negatively regulated. | New genes associated with thyroid state and involved in energy and fuel metabolism were overrepresented among the up-regulated genes. L-thyroxine therapy induced a large downregulation of the primary transcripts of the noncoding microRNA pair miR-206/miR-133b. |

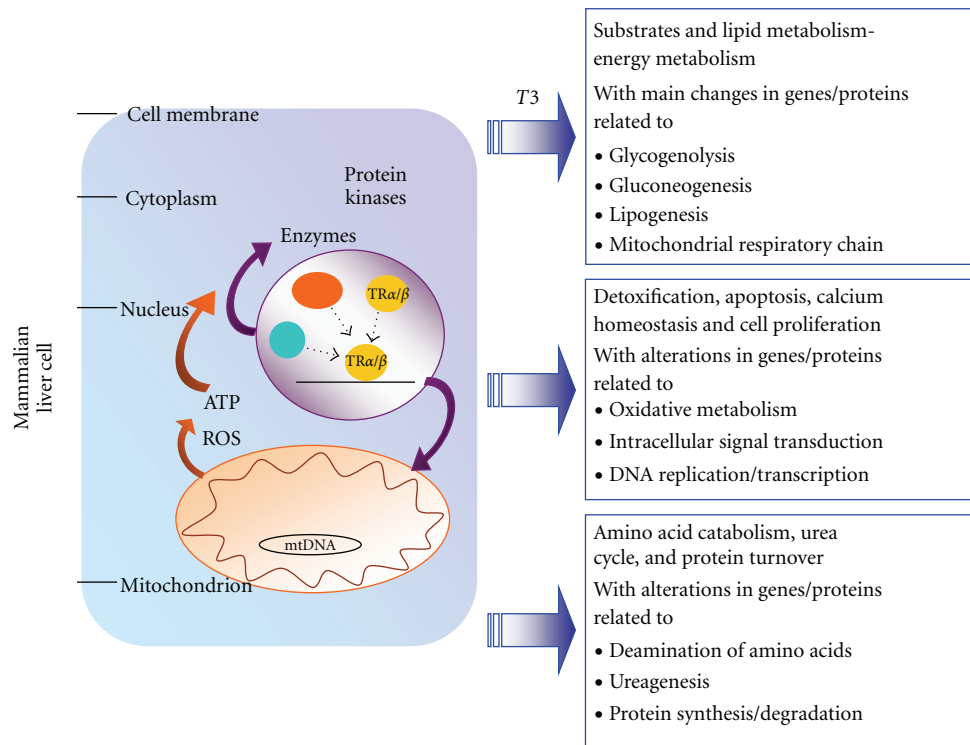


FIGURE 3: Overview of the main T3-induced transcriptomic and proteomic alterations occurring in mammalian liver. Schematic representation of the alterations in gene/protein expression underlying the response of liver to T3. Schematized are the main events and mechanisms underlying the actions of T3. Summarized are data obtained in cDNA microarray/proteomic-based studies in various mammalian models (mouse, rat, and human) (for details, see text and Tables 3 and 4).

modulated by T3 within target organs, a systematic analysis of the T3-induced changes in protein profile would appear to be appropriate.

We recently performed, on samples taken from rats in different thyroid states, high-resolution differential proteomic analysis, combining 2D-E and subsequent MALDI-ToF MS [55, 56]. These studies (summarized in Table 4) were the first application of proteomic technology to the study of the modulations that T3 exerts *in vivo* over tissue proteins, and they provided the first systematic identification of T3-induced changes in the protein expression profiles of rat liver and skeletal muscle. In the liver, we unambiguously identified 14 differentially expressed proteins involved in substrate and lipid metabolism, energy metabolism, detoxification of cytotoxic products, calcium homeostasis, amino acid catabolism, and the urea cycle [55]. We found that T3 treatment affected the expressions of enzymes such as mitochondrial aldehyde dehydrogenase, α -enolase, sorbitol dehydrogenase, acyl-CoA dehydrogenase, 3-ketoacyl-CoA thiolase, and 3-hydroxyanthranilate 3,4-dioxygenase. Interestingly, the first two enzymes were upregulated, while the others were downregulated.

Our data were in accordance with the reported role played by thyroid hormone in the stimulation of the rate of ethanol elimination [110], and they provided further insight into the mechanisms actuated by T3 in that pathway. T3 is known to stimulate gluconeogenesis and glucose production

in the liver, thereby opposing the action of insulin on hepatic glucose production [111]. Our results extended this knowledge by showing that T3 significantly enhances the level of α -enolase, thereby participating in glycolysis and gluconeogenesis. In addition, T3 administration induced a significant increase in the hepatic ATP synthase α -chain content (in accordance with the ability of T3 to stimulate ATP synthesis) and concomitantly reduced the expression level of electron transfer flavoprotein α -subunit (α -ETF), and also that of the acyl-CoA dehydrogenases [112]. T3 treatment is associated with significant reductions in the expression levels of both peroxisomal catalase and cytoplasmic glutathione-S-transferase [55], the former being important in the protection of cells against the toxic effects of hydrogen peroxide while the latter is implicated in the cellular detoxification of a number of xenobiotics by means of their conjugation to reduced glutathione. T3 treatment of hypothyroid rats is also associated with a selective upregulation of HSP60, a molecular chaperone [113]. SMP30, also known as regucalcin, which was previously not known to be affected by T3, has now been identified as a T3 target [55]. This opens new perspectives in our understanding of the molecular pathways related to intracellular T3-dependent signaling, raising the possibility that T3 may modulate a plethora of cellular events while also acting on multifunctional proteins such as SMP30, which in turn is able to modulate the levels of second messengers such as calcium.

TABLE 4: Summary of the models used and of the major findings obtained by applying 2D-E and MS to the study of THs effects.

| Authors | Experimental model | Treatments | Number of protein spots analyzed | Number of identified proteins affected by T3 | Identified affected pathways and major findings |
|-----------------------------|---|---|----------------------------------|--|--|
| Rat | | | | | |
| Silvestri et al., 2006 [55] | 3-months-old male Wistar rats. Studied tissue: liver. | Hypothyroidism was induced by i.p. administration of PTU (1 mg/100 g BW) for 4 weeks together with a weekly i.p. injection of IOP (6 mg/100 g BW). T3 was chronically administered by giving seven daily i.p. injections of 15 µg T3/100 g BW to hypothyroid rats, while the control euthyroid and hypothyroid rats received saline injections. | 600 | 14 | The whole cell protein content of rat liver was analyzed following T3 administration. Identified proteins were involved in substrates and lipid metabolism, energy metabolism, detoxification of cytotoxic products, calcium homeostasis, amino acid catabolism, and the urea cycle. |
| Silvestri et al., 2007 [56] | 3-months-old male Wistar rats. Studied tissue: skeletal muscle. | Hypothyroidism and hyperthyroidism were induced as above (Silvestri et al 2006 [55]). | 220 | 20 | The whole-cell protein content of gastrocnemius muscles was analyzed. The differentially expressed proteins unambiguously identified were involved in substrates and energy metabolism, stress response, cell structure, and gene expression. |

T3-treated rats exhibit significant reductions in the protein levels of both ornithine carbamoyltransferase and arginase 1 [55]. These data are in accordance with a previous report [114], and in line with the idea that in the hypothyroid state there are decreases in protein synthesis and turnover.

Concerning skeletal muscle, the whole-cell protein content of gastrocnemius muscle has been analyzed, and twenty differentially expressed proteins among euthyroid, hypothyroid, and hyperthyroid rats have been identified [56]. The largest group of affected proteins (50%) was involved in substrate and energy metabolism, another important group was represented by stress-induced proteins (HSPs) (21.4%), and the remainder were implicated in structural features or gene expression (transcription, translation), each of these two groups representing 14.3% of the identified proteins [56]. The thyroid state was found to induce structural shifts in gastrocnemius muscle, toward a slower phenotype in hypothyroidism and toward a faster phenotype in hyperthyroidism [56].

Among the proteins involved in substrate metabolism, three glycolytic enzymes have been identified, namely, beta-enolase, pyruvate kinase, and triosephosphate isomerase. Beta-enolase protein levels were increased following T3 treatment (hyperthyroidism), while pyruvate kinase and triosephosphate isomerase levels were decreased in hypothyroidism and elevated in hyperthyroidism [56]. This is in accordance with (a) a major T3-dependence on pyruvate kinase and triosephosphate isomerase and a generally decreased metabolic dependence on glycolysis in hypothyroidism, and (b) an increased reliance on glycolysis in hyperthyroidism [115]. Accordingly, hyperthyroidism was

found to be associated with an increased expression of cytoplasmic malate dehydrogenase. Moreover, phosphatidylethanolamine-binding protein, a basic protein that shows preferential affinity *in vitro* for phosphatidylethanolamine, was significantly increased in both the hypo- and hyperthyroid gastrocnemius (versus the euthyroid controls), most likely reflecting a thyroid state-associated cell-remodeling [56].

The expression level of the ATP synthase beta subunit was increased in both hypothyroid and hyperthyroid muscle (versus euthyroid controls), with a slight decrease in hyperthyroid animals versus hypothyroid ones. Cytosolic creatine kinase, on the other hand, was decreased in hypothyroidism versus both euthyroidism and hyperthyroidism, suggesting a decreased dependence of energy metabolism on the creatine kinase shuttle in hypothyroid muscle [56].

The expression level of HSP70 was found to be significantly increased in hypothyroid muscle (versus both euthyroid and hyperthyroid muscle), paralleling the changes in MHC1b [56]. A similar expression pattern was found for HSP20 which, despite not being a heat-inducible HSP, is biologically regulated by several cellular signalling pathways. Also identified was HSP27, which has been demonstrated to play important roles in smooth muscle cells (actin polymerization, remodeling, and even cross-bridge cycling), and which can, moreover, act as a chaperone in the regulation of contractile-protein activation [116] and also combat insulin resistance [117].

Concerning cell structure, in accordance with a predominant expression of MHC1b over MHCIIb in hypothyroidism and a reversal of the ratio between the two fiber-type

isoforms after T3 administration, the expression level of myosin regulatory light chain 2, typical of slow-twitch fibers, was strongly increased in hypothyroidism, with hyperthyroidism significantly reducing it (in each case, versus euthyroidism) [56].

Finally, both hypothyroidism and hyperthyroidism induced chromodomain-helicase-DNA-binding protein 1 (CHD 1), as well as eukaryotic translation initiation factor 3 subunit 10 (IF3A), two proteins that play important roles in different steps of gene expression: (1) initiation of transcription; and (2) initiation of translation [56].

4. Conclusions

In conclusion, although the biochemical and cellular mechanisms that underlie sarcopenia in ageing muscle and the effects elicited by thyroid hormones are only beginning to be elucidated, array-based transcriptomic studies, together with MS-based proteomic ones, are producing new insights into the pathophysiological mechanisms behind such complex phenomena.

As can be seen from the above discussion, the approaches used in the cited studies have allowed the identification of previously unrecognized proteins, thereby increasing our awareness of the large repertoire of proteins, and the multiple cell processes and signaling pathways that are affected by T3 and by ageing (for a schematic representation, see Figures 2 and 3). However, as the majority of the cited studies were performed *in vivo*, the possibility remains that certain hormones and/or other factors that are affected by such metabolic situations may have been partially responsible for the observed results.

On the basis of what has been achieved so far, the authors feel justified in championing the use of combined transcriptomic and proteomic approaches in living animals for the study of complex physiological, as well as pathophysiological, systems. Such approaches should also prove valuable for drug-design, enabling the agonist and/or antagonist properties of drugs (as well as their side effects) to be characterized on the basis of the changes they induce in protein-expression patterns.

Acknowledgments

This work was supported by Grant MIUR-COFIN 20089SRS2X and by Grant Regione Campania 2008.

References

- [1] M. Schena, D. Shalon, R. W. Davis, and P. O. Brown, "Quantitative monitoring of gene expression patterns with a complementary DNA microarray," *Science*, vol. 270, no. 5235, pp. 467–470, 1995.
- [2] D. Gerhold, T. Rushmore, and C. T. Caskey, "DNA chips: promising toys have become powerful tools," *Trends in Biochemical Sciences*, vol. 24, no. 5, pp. 168–173, 1999.
- [3] G. G. Lennon, "High-throughput gene expression analysis for drug discovery," *Drug Discovery Today*, vol. 5, no. 2, pp. 59–66, 2000.
- [4] J. E. Celis, M. Østergaard, N. A. Jensen, I. Gromova, H. H. Rasmussen, and P. Gromov, "Human and mouse proteomic databases: novel resources in the protein universe," *FEBS Letters*, vol. 430, no. 1-2, pp. 64–72, 1998.
- [5] R. D. Appel, C. Hoogland, A. Bairoch, and D. F. Hochstrasser, "Constructing a 2-D database for the World Wide Web," *Methods in Molecular Biology*, vol. 112, pp. 411–416, 1999.
- [6] M. J. Dunn, "Studying heart disease using the proteomic approach," *Drug Discovery Today*, vol. 5, no. 2, pp. 76–84, 2000.
- [7] P. G. Righetti, A. Castagna, F. Antonucci et al., "Critical survey of quantitative proteomics in two-dimensional electrophoretic approaches," *Journal of Chromatography A*, vol. 1051, no. 1-2, pp. 3–17, 2004.
- [8] A. Vlahou and M. Fountoulakis, "Proteomic approaches in the search for disease biomarkers," *Journal of Chromatography B*, vol. 814, no. 1, pp. 11–19, 2005.
- [9] A. A. Vandervoort, "Ageing of the human neuromuscular system," *Muscle and Nerve*, vol. 25, no. 1, pp. 17–25, 2002.
- [10] L. J. Melton III, S. Khosla, C. S. Crowson, M. K. O'Connor, W. M. O'Fallon, and B. L. Riggs, "Epidemiology of sarcopenia," *Journal of the American Geriatrics Society*, vol. 48, no. 6, pp. 625–630, 2000.
- [11] L. J. S. Greenlund and K. S. Nair, "Sarcopenia—consequences, mechanisms, and potential therapies," *Mechanisms of Ageing and Development*, vol. 124, no. 3, pp. 287–299, 2003.
- [12] E. Carmeli, R. Coleman, and A. Z. Reznick, "The biochemistry of aging muscle," *Experimental Gerontology*, vol. 37, no. 4, pp. 477–489, 2002.
- [13] L. Larsson, "The age-related motor disability: underlying mechanisms in skeletal muscle at the motor unit, cellular and molecular level," *Acta Physiologica Scandinavica*, vol. 163, no. 3, pp. S27–S29, 1998.
- [14] D. V. Rao, G. M. Boyle, P. G. Parsons, K. Watson, and G. L. Jones, "Influence of ageing, heat shock treatment and *in vivo* total antioxidant status on gene-expression profile and protein synthesis in human peripheral lymphocytes," *Mechanisms of Ageing and Development*, vol. 124, no. 1, pp. 55–69, 2003.
- [15] D. H. Ly, D. J. Lockhart, R. A. Lerner, and P. G. Schultz, "Mitotic misregulation and human aging," *Science*, vol. 287, no. 5462, pp. 2486–2492, 2000.
- [16] K. J. Kyng, A. May, S. Kølvråa, and V. A. Bohr, "Gene expression profiling in Werner syndrome closely resembles that of normal aging," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 21, pp. 12259–12264, 2003.
- [17] T. Lu, Y. Pan, S.-Y. Kao et al., "Gene regulation and DNA damage in the ageing human brain," *Nature*, vol. 429, no. 6994, pp. 883–891, 2004.
- [18] L. Erraji-Benchekroun, M. D. Underwood, V. Arango et al., "Molecular aging in human prefrontal cortex is selective and continuous throughout adult life," *Biological Psychiatry*, vol. 57, no. 5, pp. 549–558, 2005.
- [19] S. Welle, A. I. Brooks, J. M. Delehanty et al., "Skeletal muscle gene expression profiles in 20–29 year old and 65–71 year old women," *Experimental Gerontology*, vol. 39, no. 3, pp. 369–377, 2004.
- [20] A. B. Csoka, S. B. English, C. P. Simkevich et al., "Genome-scale expression profiling of Hutchinson-Gilford progeria syndrome reveals widespread transcriptional misregulation leading to mesodermal/mesenchymal defects and accelerated atherosclerosis," *Aging Cell*, vol. 3, no. 4, pp. 235–243, 2004.

- [21] J. M. Zahn, R. Sonu, H. Vogel et al., "Transcriptional profiling of aging in human muscle reveals a common aging signature," *PLoS genetics*, vol. 2, no. 7, article e115, 2006.
- [22] G. E. J. Rodwell, R. Sonu, J. M. Zahn et al., "A transcriptional profile of aging in the human kidney," *PLoS Biology*, vol. 2, no. 12, article e427, 2004.
- [23] S. Welle, A. I. Brooks, J. M. Delehanty, N. Needler, and C. A. Thornton, "Gene expression profile of aging in human muscle," *Physiological Genomics*, vol. 14, pp. 149–159, 2003.
- [24] J. Tower, "Sex-specific regulation of aging and apoptosis," *Mechanisms of Ageing and Development*, vol. 127, no. 9, pp. 705–718, 2006.
- [25] C.-K. Lee, R. G. Klopp, R. Weindruch, and T. A. Prolla, "Gene expression profile of aging and its retardation by caloric restriction," *Science*, vol. 285, no. 5432, pp. 1390–1393, 1999.
- [26] A. Lombardi, E. Silvestri, F. Cioffi et al., "Defining the transcriptomic and proteomic profiles of rat ageing skeletal muscle by the use of a cDNA array, 2D- and Blue native-PAGE approach," *Journal of Proteomics*, vol. 72, no. 4, pp. 708–721, 2009.
- [27] T. Kayo, D. B. Allison, R. Weindruch, and T. A. Prolla, "Influences of aging and caloric restriction on the transcriptional profile of skeletal muscle from rhesus monkeys," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 9, pp. 5093–5098, 2001.
- [28] R. Sreekumar, J. Unnikrishnan, A. Fu et al., "Effects of caloric restriction on mitochondrial function and gene transcripts in rat muscle," *American Journal of Physiology - Endocrinology and Metabolism*, vol. 283, no. 1, pp. E38–43, 2002.
- [29] M. Altun, E. Edström, E. Spooner et al., "Iron load and redox stress in skeletal muscle of aged rats," *Muscle and Nerve*, vol. 36, no. 2, pp. 223–233, 2007.
- [30] J. Gannon, L. Staunton, K. O'Connell, P. Doran, and K. Ohlndieck, "Phosphoproteomic analysis of aged skeletal muscle," *International Journal of Molecular Medicine*, vol. 22, no. 1, pp. 33–42, 2008.
- [31] J. Kanski, M. A. Alterman, and C. Schöneich, "Proteomic identification of age-dependent protein nitration in rat skeletal muscle," *Free Radical Biology and Medicine*, vol. 35, no. 10, pp. 1229–1239, 2003.
- [32] J. Kanski, S. J. Hong, and C. Schöneich, "Proteomic analysis of protein nitration in aging skeletal muscle and identification of nitrotyrosine-containing sequences in vivo by nanoelectrospray ionization tandem mass spectrometry," *The Journal of Biological Chemistry*, vol. 280, no. 25, pp. 24261–24266, 2005.
- [33] K. O'Connell, P. Doran, J. Gannon, and K. Ohlndieck, "Lectin-based proteomic profiling of aged skeletal muscle: decreased pyruvate kinase isozyme M1 exhibits drastically increased levels of N-glycosylation," *European Journal of Cell Biology*, vol. 87, no. 10, pp. 793–805, 2008.
- [34] C. Gelfi, A. Viganò, M. Ripamonti et al., "The human muscle proteome in aging," *Journal of Proteome Research*, vol. 5, no. 6, pp. 1344–1353, 2006.
- [35] D. Cai, M. Li, K. Lee, K. Lee, W. Wong, and K. Chan, "Age-related changes of aqueous protein profiles in rat fast and slow twitch skeletal muscles," *Electrophoresis*, vol. 21, no. 2, pp. 465–472, 2000.
- [36] I. Piec, A. Listrat, J. Alliot, C. Chambon, R. G. Taylor, and D. Bechet, "Differential proteome analysis of aging in rat skeletal muscle," *The FASEB Journal*, vol. 19, no. 9, pp. 1143–1145, 2005.
- [37] K. O'Connell, J. Gannon, P. Doran, and K. Ohlndieck, "Proteomic profiling reveals a severely perturbed protein expression pattern in aged skeletal muscle," *International Journal of Molecular Medicine*, vol. 20, no. 2, pp. 145–153, 2007.
- [38] P. Doran, K. O'Connell, J. Gannon, M. Kavanagh, and K. Ohlndieck, "Opposite pathobiochemical fate of pyruvate kinase and adenylate kinase in aged rat skeletal muscle as revealed by proteomic DIGE analysis," *Proteomics*, vol. 8, no. 2, pp. 364–377, 2008.
- [39] J. Chang, H. Van Remmen, J. Cornell, A. Richardson, and W. F. Ward, "Comparative proteomics: characterization of a two-dimensional gel electrophoresis system to study the effect of aging on mitochondrial proteins," *Mechanisms of Ageing and Development*, vol. 124, no. 1, pp. 33–41, 2003.
- [40] D. Q. Cai, M. Li, K. K. Lee et al., "Parvalbumin expression is downregulated in rat fast-twitch skeletal muscles during aging," *Archives of Biochemistry and Biophysics*, vol. 387, no. 2, pp. 202–208, 2001.
- [41] N. A. Dencher, S. Goto, N. H. Reifschneider, M. Sugawa, and F. Krause, "Unraveling age-dependent variation of the mitochondrial proteome," *Annals of the New York Academy of Sciences*, vol. 1067, no. 1, pp. 116–119, 2006.
- [42] J. Feng, H. Xie, D. L. Meany, L. V. Thompson, E. A. Arriaga, and T. J. Griffin, "Quantitative proteomic profiling of muscle type-dependent and age-dependent protein carbonylation in rat skeletal muscle mitochondria," *Journals of Gerontology A*, vol. 63, no. 11, pp. 1137–1152, 2008.
- [43] G. S. Cobon, N. Verrills, P. Papakostopoulos et al., "Proteomics of ageing," *Biogerontology*, vol. 3, no. 1-2, pp. 133–136, 2002.
- [44] X. Feng, Y. Jiang, P. Meltzer, and P. M. Yen, "Thyroid hormone regulation of hepatic genes in vivo detected by complementary DNA microarray," *Molecular Endocrinology*, vol. 14, no. 7, pp. 947–955, 2000.
- [45] A. Flores-Morales, H. Gullberg, L. Fernandez et al., "Patterns of liver gene expression governed by TR β ," *Molecular Endocrinology*, vol. 16, no. 6, pp. 1257–1268, 2002.
- [46] L. D. Miller, P. McPhie, H. Suzuki, Y. Kato, E. T. Liu, and S. Y. Cheng, "Multi-tissue gene-expression analysis in a mouse model of thyroid hormone resistance," *Genome Biology*, vol. 5, no. 5, article R31, 2004.
- [47] T. Ventura-Holman, A. Mamoon, J. F. Maher et al., "Thyroid hormone responsive genes in the murine hepatocyte cell line AML 12," *Gene*, vol. 396, no. 2, pp. 332–327, 2007.
- [48] H. Dong, C. L. Yauk, A. Williams, A. Lee, G. R. Douglas, and M. G. Wade, "Hepatic gene expression changes in hypothyroid juvenile mice: characterization of a novel negative thyroid-responsive element," *Endocrinology*, vol. 148, no. 8, pp. 3932–3940, 2007.
- [49] J. M. Weitzel, C. Radtke, and H. J. Seitz, "Two thyroid hormone-mediated gene expression patterns in vivo identified by cDNA expression arrays in rat," *Nucleic Acids Research*, vol. 29, no. 24, pp. 5148–5155, 2001.
- [50] L. D. Miller, K. S. Park, Q. M. Guo et al., "Silencing of Wnt signaling and activation of multiple metabolic pathways in response to thyroid hormone-stimulated cell proliferation," *Molecular and Cellular Biology*, vol. 21, no. 19, pp. 6626–6639, 2001.
- [51] N. Viguerie and D. Langin, "Effect of thyroid hormone on gene expression," *Current Opinion in Clinical Nutrition and Metabolic Care*, vol. 6, no. 4, pp. 377–381, 2003.

- [52] L. C. Moeller, A. M. Dumitrescu, R. L. Walker, P. S. Meltzer, and S. Refetoff, "Thyroid hormone responsive genes in cultured human fibroblasts," *Journal of Clinical Endocrinology and Metabolism*, vol. 90, no. 2, pp. 936–943, 2005.
- [53] K. Clément, N. Viguerie, M. Diehn et al., "In vivo regulation of human skeletal muscle gene expression by thyroid hormone," *Genome Research*, vol. 12, no. 2, pp. 281–291, 2002.
- [54] W. E. Visser, K. A. Heemstra, S. M. A. Swagemakers et al., "Physiological thyroid hormone levels regulate numerous skeletal muscle transcripts," *Journal of Clinical Endocrinology and Metabolism*, vol. 94, no. 9, pp. 3487–3496, 2009.
- [55] E. Silvestri, M. Moreno, L. Schiavo et al., "A proteomics approach to identify protein expression changes in rat liver following administration of 3,5,3'-triiodo-L-thyronine," *Journal of Proteome Research*, vol. 5, no. 9, pp. 2317–2327, 2006.
- [56] E. Silvestri, L. Burrone, P. de Lange et al., "Thyroid-state influence on protein-expression profile of rat skeletal muscle," *Journal of Proteome Research*, vol. 6, no. 8, pp. 3187–3196, 2007.
- [57] M. Mann and O. N. Jensen, "Proteomic analysis of post-translational modifications," *Nature Biotechnology*, vol. 21, no. 3, pp. 255–261, 2003.
- [58] C. Schöneich, "Protein modification in aging: an update," *Experimental Gerontology*, vol. 41, no. 9, pp. 807–812, 2006.
- [59] J. V. Olsen, B. Blagoev, F. Gnad et al., "Global, in vivo, and site-specific phosphorylation dynamics in signaling networks," *Cell*, vol. 127, no. 3, pp. 635–648, 2006.
- [60] K. Ohtsubo and J. D. Marth, "Glycosylation in cellular mechanisms of health and disease," *Cell*, vol. 126, no. 5, pp. 855–867, 2006.
- [61] H. H. Freeze, "Genetic defects in the human glycome," *Nature Reviews Genetics*, vol. 7, no. 8, pp. 660–674, 2006.
- [62] N. A. Dencher, M. Frenzel, N. H. Reifschneider, M. Sugawa, and F. Krause, "Proteome alterations in rat mitochondria caused by aging," *Annals of the New York Academy of Sciences*, vol. 1100, pp. 291–298, 2007.
- [63] S. Rexroth, J. M. W. Meyer zu Tittingdorf, F. Krause, N. A. Dencher, and H. Seelert, "Thylakoid membrane at altered metabolic state: challenging the forgotten realms of the proteome," *Electrophoresis*, vol. 24, no. 16, pp. 2814–2823, 2003.
- [64] F. Krause, "Detection and analysis of protein-protein interactions in organellar and prokaryotic proteomes by native gel electrophoresis: (Membrane) protein complexes and supercomplexes," *Electrophoresis*, vol. 27, no. 13, pp. 2759–2781, 2006.
- [65] V. Pesce, A. Cormio, F. Fracasso et al., "Age-related mitochondrial genotypic and phenotypic alterations in human skeletal muscle," *Free Radical Biology and Medicine*, vol. 30, no. 11, pp. 1223–1233, 2001.
- [66] D. Dani and N. A. Dencher, "Native-DIGE: a new look at the mitochondrial membrane proteome," *Biotechnology Journal*, vol. 3, no. 6, pp. 817–822, 2008.
- [67] J. Chang, J. E. Cornell, H. Van Remmen, K. Hakala, W. F. Ward, and A. Richardson, "Effect of aging and caloric restriction on the mitochondrial proteome," *Journals of Gerontology A*, vol. 62, no. 3, pp. 223–234, 2007.
- [68] K. O'Connell and K. Ohlendieck, "Proteomic DIGE analysis of the mitochondria-enriched fraction from aged rat skeletal muscle," *Proteomics*, vol. 9, no. 24, pp. 5509–5524, 2009.
- [69] H. Schägger and K. Pfeiffer, "Supercomplexes in the respiratory chains of yeast and mammalian mitochondria," *The EMBO Journal*, vol. 19, no. 8, pp. 1777–1783, 2000.
- [70] D. S. Cooper, "Subclinical hypothyroidism," *The New England Journal of Medicine*, vol. 345, no. 4, pp. 260–265, 2001.
- [71] N. Knudsen, P. Laurberg, L. B. Rasmussen et al., "Small differences in thyroid function may be important for body mass index and the occurrence of obesity in the population," *Journal of Clinical Endocrinology and Metabolism*, vol. 90, no. 7, pp. 4019–4024, 2005.
- [72] A. Oetting and P. M. Yen, "New insights into thyroid hormone action," *Best Practice and Research in Clinical Endocrinology and Metabolism*, vol. 21, no. 2, pp. 193–208, 2007.
- [73] D. Robyr, A. P. Wolffe, and W. Wahli, "Nuclear hormone receptor coregulators in action: diversity for shared tasks," *Molecular Endocrinology*, vol. 14, no. 3, pp. 329–347, 2000.
- [74] J. Zhang and M. A. Lazar, "The mechanism of action of thyroid hormones," *Annual Review of Physiology*, vol. 62, pp. 439–466, 2000.
- [75] F. Flamant, K. Gauthier, and J. Samarut, "Thyroid hormones signaling is getting more complex: STORMs are coming," *Molecular Endocrinology*, vol. 21, no. 2, pp. 321–333, 2007.
- [76] M. Busson, L. Daury, P. Seyer et al., "Avian MyoD and c-Jun coordinately induce transcriptional activity of the 3,5,3'-triiodothyronine nuclear receptor c-ErbA α 1 in proliferating myoblasts," *Endocrinology*, vol. 147, no. 7, pp. 3408–3418, 2006.
- [77] C. Desbois, D. Aubert, C. Legrand, B. Pain, and J. Samarut, "A novel mechanism of action for v-ErbA: abrogation of the inactivation of transcription factor AP-1 by retinoic acid and thyroid hormone receptors," *Cell*, vol. 67, no. 4, pp. 731–740, 1991.
- [78] L. Daury, M. Busson, F. Casas, I. Cassar-Malek, C. Wrutniak-Cabello, and G. Cabello, "The triiodothyronine nuclear receptor c-ErbA α 1 inhibits avian MyoD transcriptional activity in myoblasts," *FEBS Letters*, vol. 508, no. 2, pp. 236–240, 2001.
- [79] M. A. Lazar, "Thyroid hormone receptors: multiple forms, multiple possibilities," *Endocrine Reviews*, vol. 14, no. 2, pp. 184–193, 1993.
- [80] G. A. Brent, "Mechanisms of disease: the molecular basis of thyroid hormone action," *The New England Journal of Medicine*, vol. 331, no. 13, pp. 847–853, 1994.
- [81] J. Sap, A. Munoz, and K. Damm, "The c-erb-A protein is a high-affinity receptor for thyroid hormone," *Nature*, vol. 324, no. 6098, pp. 635–640, 1986.
- [82] E. D. Abel, E. G. Moura, R. S. Ahima et al., "Dominant inhibition of thyroid hormone action selectively in the pituitary of thyroid hormone-receptor- β null mice abolishes the regulation of thyrotropin by thyroid hormone," *Molecular Endocrinology*, vol. 17, no. 9, pp. 1767–1776, 2003.
- [83] J. Burnside, D. S. Sarling, F. E. Carr, and W. W. Chin, "Thyroid hormone regulation of the rat glycoprotein hormone α -subunit gene promoter activity," *The Journal of Biological Chemistry*, vol. 264, no. 12, pp. 6886–6891, 1989.
- [84] D. L. Bodenner, M. A. Mroczynski, B. D. Weintraub, S. Radovick, and F. E. Wondisford, "A detailed functional and structural analysis of a major thyroid hormone inhibitory element in the human thyrotropin β -subunit gene," *The Journal of Biological Chemistry*, vol. 266, no. 32, pp. 21666–21673, 1991.

- [85] D. Forrest and B. Vennström, "Functions of thyroid hormone receptors in mice," *Thyroid*, vol. 10, no. 1, pp. 41–52, 2000.
- [86] A. Lanni, M. Moreno, A. Lombardi, P. De Lange, and F. Goglia, "Control of energy metabolism by iodothyronines," *Journal of Endocrinological Investigation*, vol. 24, no. 11, pp. 897–913, 2001.
- [87] A. Lanni, M. Moreno, A. Lombardi, and F. Goglia, "Thyroid hormone and uncoupling proteins," *FEBS Letters*, vol. 543, no. 1*#8211;3, pp. 5–10, 2003.
- [88] E. Silvestri, L. Schiavo, A. Lombardi, and F. Goglia, "Thyroid hormones as molecular determinants of thermogenesis," *Acta Physiologica Scandinavica*, vol. 184, no. 4, pp. 265–283, 2005.
- [89] J. E. Silva, "Thermogenic mechanisms and their hormonal regulation," *Physiological Reviews*, vol. 86, no. 2, pp. 435–464, 2006.
- [90] F. Goglia, M. Moreno, and A. Lanni, "Action of thyroid hormones at the cellular level: the mitochondrial target," *FEBS Letters*, vol. 452, no. 3, pp. 115–120, 1999.
- [91] M. Gaspari, N.-G. Larsson, and C. M. Gustafsson, "The transcription machinery in mammalian mitochondria," *Biochimica et Biophysica Acta*, vol. 1659, no. 2-3, pp. 148–152, 2004.
- [92] R. C. Scarpulla, "Nuclear control of respiratory gene expression in mammalian cells," *Journal of Cellular Biochemistry*, vol. 97, no. 4, pp. 673–683, 2006.
- [93] C. Wrutniak-Cabello, F. Casas, and G. Cabello, "Thyroid hormone action in mitochondria," *Journal of Molecular Endocrinology*, vol. 26, no. 1, pp. 67–77, 2001.
- [94] A.-M. G. Psarra, S. Solakidi, and C. E. Sekeris, "The mitochondrion as a primary site of action of steroid and thyroid hormones: presence and action of steroid and thyroid hormone receptors in mitochondria of animal cells," *Molecular and Cellular Endocrinology*, vol. 246, no. 1-2, pp. 21–33, 2006.
- [95] C. Wrutniak, P. Rochard, F. Casas, A. Fraysse, J. Charrier, and G. Cabello, "Physiological importance of the T3 mitochondrial pathway," *Annals of the New York Academy of Sciences*, vol. 839, pp. 93–100, 1998.
- [96] C. H. Gouveia, J. J. Schultz, D. J. Jackson, G. R. Williams, and G. A. Brent, "Thyroid hormone gene targets in ROS 17/2.8 osteoblast like cells identified by differential display analysis," *Thyroid*, vol. 12, no. 8, pp. 663–671, 2002.
- [97] T. Iglesias, J. Caubín, A. Zaballos, J. Bernal, and A. Munoz, "Identification of the mitochondrial NADH dehydrogenase subunit 3 (ND3) as a thyroid hormone regulated gene by whole genome PCR analysis," *Biochemical and Biophysical Research Communications*, vol. 210, no. 3, pp. 995–1000, 1995.
- [98] R. J. Wiesner, T. T. Kurowski, and R. Zak, "Regulation by thyroid hormone of nuclear and mitochondrial genes encoding subunits of cytochrome-c oxidase in rat liver and skeletal muscle," *Molecular Endocrinology*, vol. 6, no. 9, pp. 1458–1467, 1992.
- [99] Y. Hiroi, H.-H. Kim, H. Ying et al., "Rapid nongenomic actions of thyroid hormone," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 38, pp. 14104–14109, 2006.
- [100] S.-Y. Cheng, J. L. Leonard, and P. J. Davis, "Molecular aspects of thyroid hormone actions," *Endocrine Reviews*, vol. 31, no. 2, pp. 139–170, 2010.
- [101] A. J. Hulbert, "Thyroid hormones and their effects: a new perspective," *Biological Reviews of the Cambridge Philosophical Society*, vol. 75, no. 4, pp. 519–631, 2000.
- [102] M. Moreno, P. de Lange, A. Lombardi, E. Silvestri, A. Lanni, and F. Goglia, "Metabolic effects of thyroid hormone derivatives," *Thyroid*, vol. 18, no. 2, pp. 239–253, 2008.
- [103] N. S. Kavok, O. A. Krasilnikova, and N. A. Babenko, "Thyroxine signal transduction in liver cells involves phospholipase C and phospholipase D activation. Genomic independent action of thyroid hormone," *BMC Cell Biology*, vol. 2, article 5, 2001.
- [104] X. Cao, F. Kambe, L. C. Moeller, S. Refetoff, and H. Seo, "Thyroid hormone induces rapid activation of Akt/protein kinase B-mammalian target of rapamycin-p70S6K cascade through phosphatidylinositol 3-kinase in human fibroblasts," *Molecular Endocrinology*, vol. 19, no. 1, pp. 102–112, 2005.
- [105] S. Seelig, C. Liaw, H. C. Towle, and J. H. Oppenheimer, "Thyroid hormone attenuates and augments hepatic gene expression at a pretranslational level," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 78, no. 8, pp. 4733–4737, 1981.
- [106] H. Dong, M. Wade, A. Williams, A. Lee, G. R. Douglas, and C. Yauk, "Molecular insight into the effects of hypothyroidism on the developing cerebellum," *Biochemical and Biophysical Research Communications*, vol. 330, no. 4, pp. 1182–1193, 2005.
- [107] H.-M. Zhang, Q. Su, and M. Luo, "Thyroid hormone regulates the expression of SNAP-25 during rat brain development," *Molecular and Cellular Biochemistry*, vol. 307, no. 1-2, pp. 169–175, 2008.
- [108] D. Diez, C. Grijota-Martinez, P. Agretti et al., "Thyroid hormone action in the adult brain: gene expression profiling of the effects of single and multiple doses of triiodo-L-thyronine in the rat striatum," *Endocrinology*, vol. 149, no. 8, pp. 3989–4000, 2008.
- [109] W. E. Visser, E. C. H. Friesema, J. Jansen, and T. J. Visser, "Thyroid hormone transport by monocarboxylate transporters," *Best Practice and Research in Clinical Endocrinology and Metabolism*, vol. 21, no. 2, pp. 223–236, 2007.
- [110] J. Li, V. Nguyen, B. A. French et al., "Mechanism of the alcohol cyclic pattern: role of the hypothalamic-pituitary-thyroid axis," *American Journal of Physiology*, vol. 279, no. 1, pp. G118–G125, 2000.
- [111] T. Merkulova, A. Keller, P. Oliviero et al., "Thyroid hormones differentially modulate enolase isozymes during rat skeletal and cardiac muscle development," *American Journal of Physiology*, vol. 278, no. 2, pp. E330–E339, 2000.
- [112] M. Nagao, B. Parimoo, and K. Tanaka, "Developmental, nutritional, and hormonal regulation of tissue-specific expression of the genes encoding various acyl-CoA dehydrogenases and α -subunit of electron transfer flavoprotein in rat," *The Journal of Biological Chemistry*, vol. 268, no. 32, pp. 24114–24124, 1993.
- [113] D. A. Hood and A.-M. Joseph, "Mitochondrial assembly: protein import," *Proceedings of the Nutrition Society*, vol. 63, no. 2, pp. 293–300, 2004.
- [114] M. Sochor, P. McLean, J. Brown, and A. L. Greenbaum, "Regulation of pathways of ornithine metabolism. Effects of thyroid hormone and diabetes on the activity of enzymes at the "ornithine crossroads" in rat liver," *Enzyme*, vol. 26, no. 1, pp. 15–23, 1981.
- [115] G. D. Dimitriadis, B. Leighton, M. Parry-Billings, D. West, and E. A. Newsholme, "Effect of hypothyroidism on the sensitivity of glycolysis and glycogen synthesis to insulin in the soleus muscle of the rat," *Biochemical Journal*, vol. 257, no. 2, pp. 369–373, 1989.

- [116] S. Somara and K. N. Bitar, "Tropomyosin interacts with phosphorylated HSP27 in agonist-induced contraction of smooth muscle," *American Journal of Physiology*, vol. 286, no. 6, pp. C1290–C1301, 2004.
- [117] M. F. McCarty, "Induction of heat shock proteins may combat insulin resistance," *Medical Hypotheses*, vol. 66, no. 3, pp. 527–534, 2006.

Review Article

Navigating the Human Metabolome for Biomarker Identification and Design of Pharmaceutical Molecules

Irene Kouskoumvekaki and Gianni Panagiotou

Department of Systems Biology, Center for Biological Sequence Analysis, Building 208, Technical University of Denmark, 2800, Lyngby, Denmark

Correspondence should be addressed to Irene Kouskoumvekaki, irene@cbs.dtu.dk and Gianni Panagiotou, gpa@bio.dtu.dk

Received 14 April 2010; Accepted 12 July 2010

Academic Editor: Mika Ala-Korpela

Copyright © 2011 I. Kouskoumvekaki and G. Panagiotou. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Metabolomics is a rapidly evolving discipline that involves the systematic study of endogenous small molecules that characterize the metabolic pathways of biological systems. The study of metabolism at a global level has the potential to contribute significantly to biomedical research, clinical medical practice, as well as drug discovery. In this paper, we present the most up-to-date metabolite and metabolic pathway resources, and we summarize the statistical, and machine-learning tools used for the analysis of data from clinical metabolomics. Through specific applications on cancer, diabetes, neurological and other diseases, we demonstrate how these tools can facilitate diagnosis and identification of potential biomarkers for use within disease diagnosis. Additionally, we discuss the increasing importance of the integration of metabolomics data in drug discovery. On a case-study based on the Human Metabolome Database (HMDB) and the Chinese Natural Product Database (CNPD), we demonstrate the close relatedness of the two data sets of compounds, and we further illustrate how structural similarity with human metabolites could assist in the design of novel pharmaceuticals and the elucidation of the molecular mechanisms of medicinal plants.

1. Introduction

Metabolomics is a new technology that applies advanced separation and detection methods to capture the collection of small molecules that characterize metabolic pathways. This rapidly developing discipline involves the study of the total repertoire of small molecules present in the biological samples, particularly urine, saliva, and blood plasma [1]. Metabolites are the byproducts of metabolism, which is itself the process of converting food energy to mechanical energy or heat. Experts believe there are at least 3,000 metabolites that are essential for normal growth and development (primary metabolites) and thousands more unidentified (around 20,000, compared to an estimated 30,000 genes and 100,000 proteins) that are not essential for growth and development (secondary metabolites) but could represent prognostic, diagnostic, and surrogate markers for a disease state and a deeper understanding of mechanisms of disease [2]. Of particular interest to metabolomics researchers are small,

low-molecular weight compounds that serve as substrates and products in various metabolic pathways [3].

Metabolomics, the study of metabolism at the global level, has the potential to contribute significantly to biomedical research, and ultimately to clinical medical practice [4, 5]. It is a close counterpart to the genome, the transcriptome and the proteome. Metabolomics, genomics, proteomics, and other “-omics” grew out of the Human Genome Project, a massive research effort that began in the mid-1990s and culminated in 2003 with a complete mapping of all the genes in the human body. When discussing the clinical advantages of metabolomics, scientists point to the “real-world” assessment of patient physiology that the metabolome provides since it can be regarded as the end-point of the “-omics” cascade [6]. Other functional genomics technologies do not necessarily predict drug effects, toxicological response, or disease states at the phenotype but merely indicate the potential cause for phenotypical response. Metabolomics can bridge this information gap since the identification

and measurement of metabolite profile dynamics of host changes provides the closest link to the various phenotypic responses [7–9]. Thus it is clear that the global mapping of metabolic signatures pre- and postdrug treatment is a promising approach to identify possible functional relationships between medication and medical phenotype [10–13].

At the center of metabolomics is the concept that an individual's metabolite state is a close representation of the individual's overall health status. This metabolic state reflects what has been encoded by the genome and modified by environmental factors. In this paper, we demonstrate the enormous potential of metabolomics in disease monitoring and identification of prognostic, diagnostic, and drug response markers (Figure 1 (i)–(iii)), as well as in drug discovery and development in combination with systems chemical biology and chemoinformatics (Figures 1(a)–1(c)).

2. Databases and Data Analysis Tools

Databases of metabolites and metabolic reactions offer a wealth of information regarding the interaction of small molecules with biological systems, notably in relation with their chemical reactivity. In Table 1, we summarize all such metabolite and metabolic pathway resources which contain hundreds of reactions, metabolites, and pathways for several organisms and are designed to facilitate the exploration of metabolism across many different species. For example, the BiGG database (<http://bigg.ucsd.edu/>) is a metabolic reconstruction of human metabolism designed for systems biology simulation and metabolic flux balance modelling. It is a comprehensive literature-based genome-scale metabolic reconstruction that accounts for the functions of 1,496 ORFs, 2,004 proteins, 2,766 metabolites, and 3,311 metabolic and transport reactions. MassBank (<http://www.massbank.jp/>) is a mass spectral database of experimentally acquired high resolution MS spectra of metabolites. Maintained and supported by the JST-BIRD project, it offers various query methods for standard spectra obtained from Keio University, RIKEN PSC, and other Japanese research institutions. It is officially sanctioned by the Mass Spectrometry Society of Japan. The database has very detailed MS data and excellent spectral/structure searching utilities. More than 13,000 spectra from 1900 different compounds are available. The METLIN Metabolite Database (<http://metlin.scripps.edu/index.jp>) is a repository for mass spectral metabolite data. All metabolites are neutral or free acids. It is a collaborative effort between the Siuzdak and Abagyan groups and Center for Mass Spectrometry at The Scripps Research Institute. METLIN is searchable by compound name, mass, formula, or structure. It contains 15,000 structures, including more than 8000 di- and tripeptides. METLIN contains MS/MS, LC/MS and FTMS data that can be searched by peak lists, mass range, biological source or disease. Below we describe in more detail three interconnected databases; the Human Metabolome Database (<http://www.hmdb.ca/>), the Small Molecule Pathway Database (<http://www.smpdb.ca/>) and the Toxin and Toxin-Target Database (<http://www.t3db.org/>) (Figure 2).

2.1. Human Metabolome Database (HMDB). Focusing on quantitative, analytic, or molecular scale information about metabolites, the enzymes and transporters associated with them, as well as disease related properties the HMDB represents the most complete bioinformatics and chemoinformatics medical information database. It contains records for thousands of endogenous metabolites identified by literature surveys (PubMed, OMIM, OMMBID, text books), data mining (KEGG, Metlin, BioCyc) or experimental analyses performed on urine, blood, and cerebrospinal fluid samples. The annotation effort is aided by chemical parameter calculators and protein annotation tools originally developed for DrugBank. The HMDB is fully searchable with many built-in tools for viewing, sorting, and extracting metabolites, biofluid concentrations, enzymes, genes, NMR or MS spectra and disease information. The HMDB currently contains 7,985 compounds that are linked to 69,295 different synonyms. These compounds are also connected to 908 C-NMR and 916 H-NMR spectra as well as 7,234 associated enzymes. All chemical structures in these pathway maps are hyperlinked to HMDB MetaboCards and all enzymes are hyperlinked to UniProt data cards for human enzymes. The majority of the compounds have been detected in blood (4,226) while 784 compounds were detected in urine, 363 in CSF (cerebrospinal fluid) and 315 in other biofluids. In order a compound to be included in the HMDB it must fulfil certain criteria; it should be of biological origin, the compound weight must be <1,500 Da, and it should be found at concentrations greater than 1 mM in one or more biofluids/tissues. Compounds that are not covered by the above description but are either biomedically important metabolites, like hormones, or certain very common drugs and some ubiquitous food additives, like vitamins, are some notable exceptions in the HMDB. For a large number of metabolites the concentration values in the biofluids are given with data for both normal and abnormal values.

A key feature that distinguishes the HMDB from other metabolic resources is its extensive support for higher level database searching and selecting functions. More than 175 hand-drawn-zoomable, fully hyperlinked human metabolic pathway maps can be found in HMDB and all these maps are quite specific to human metabolism and explicitly show the subcellular compartments where specific reactions are known to take place. As an equivalent to BLAST the HMDB contains a structure similarity search tool for chemical structures and users may sketch or paste a SMILES string of a query compound into the Chem-Query window. Submitting the query launches a structure similarity search tool that looks for common substructures from the query compound that match the HMDB's metabolite database. The wealth of information and especially the extensive linkage to metabolic diseases to normal and abnormal metabolite concentration ranges, to mutation/SNP data and to the genes, enzymes, reactions and pathways associated with many diseases of interest makes the HMDB one the most valuable tool in the hands of clinical chemists, nutritionists, physicians and medical geneticists.

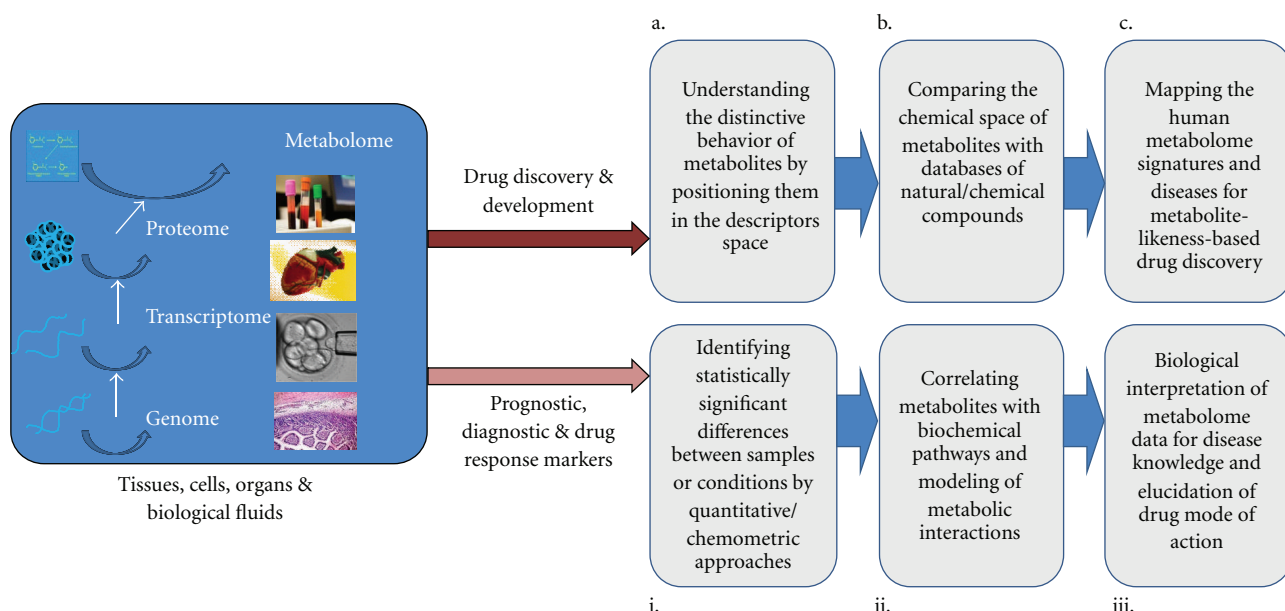


FIGURE 1: Metabolomics holds the promise to deliver valuable information about biochemical pathways perturbed in disease and upon treatment, to monitor healthy people to detect early signs of disease, to diagnose disease or predict the risk of a disease, to subclassify disease, to make safer drugs by predicting the potential for adverse drug reactions, and to speed the discovery and development of novel drug molecules.

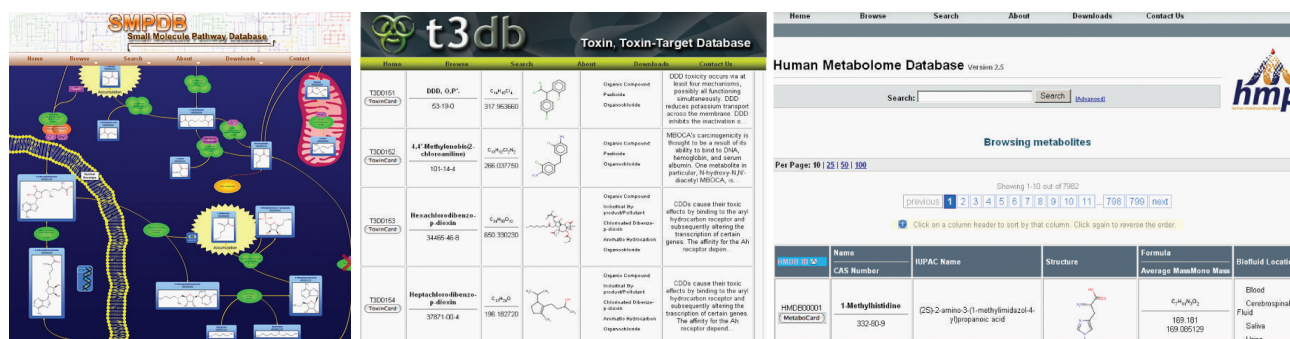


FIGURE 2: A screenshot montage of the HMDB, SMPDB and T3DB databases.

2.2. The Small Molecule Pathway Database (SMPDB).

SMPDB consists of approximately 350 hand-drawn pathways with more than 280 of them unique to SMPDB. These pathways describe small molecule metabolism or small-molecule processes that are specific to humans and fall into four different categories: (i) metabolic pathways; (ii) small-molecule disease pathways, (iii) small molecule drug pathways, and (iv) small molecule signalling pathways. In order for a metabolic pathway to be suitable for inclusion in SMPDB, it must be found in humans and it must contain at least five small molecules. If it is a human disease, drug or signalling pathway the determining factor for inclusion is its central feature being based on the action of at least one small molecule. More specifically, in SMPDB, disease pathways refer to those pathways describing human disease processes where small-molecule metabolite dysregulation is the primary hallmark of the disease. For qualifying a small molecule or set of small molecules to be included in SMPDB,

a significant concentration change, which is commonly used for the diagnosis, prognosis, or monitoring for a given disease, is required. The SMPDB interface is largely modelled after the interface used for DrugBank and the HMDB with a navigation panel for browsing, searching, and downloading the database. The users can choose between two browsing options, SMP-BROWSE, and SMP-TOC. The latter is basically a scrollable hyperlinked table of contents that lists all pathways by name and category. SMP-BROWSE is a more comprehensive browsing tool that provides a tabular synopsis of SMPDB's content using thumbnail images of the pathway diagrams, textual descriptions of the pathways, as well as lists of the corresponding chemical components and enzyme/protein components. All of the chemical structures and proteins/enzymes illustrated in SMPDB's diagrams are hyperlinked to other online databases or tables, but this is common in most pathway databases. Specifically, all metabolites, drugs or proteins shown in the SMP-BROWSE tables

TABLE 1: Machine-learning algorithms often used in metabolomics.

| Technique | Description |
|-----------|---|
| PCA | The Principal Component Analysis (PCA) is a frequently used method which is applied to extract the systematic variance in a data matrix. It helps to obtain an overview over dominant patterns and major trends in the data. The aim of PCA is to create a set of latent variables which is smaller than the set of original variables but still explains all the variance of the original variables. In mathematical terms, PCA transforms a number of correlated variables into a smaller number of uncorrelated variables, the so-called principal components. |
| PLS | Partial Least Squares (PLS), also called Projection to Latent Structures, is a linear regression method that can be applied to establish a predictive model, even if the objects are highly correlated. The X variables (the predictors) are reduced to principal components, as are the Y variables (the dependents). The components of X are used to predict the scores on the Y components, and the predicted Y component scores are used to predict the actual values of the Y variables. In constructing the principal components of X, the PLS algorithm iteratively maximizes the strength of the relation of successive pairs of X and Y component scores by maximizing the covariance of each X-score with the Y variables. This strategy means that while the original X variables may be multicollinear, the X components used to predict Y will be orthogonal. Also, the X variables may have missing values, but there will be a computed score for every case on every X component. Finally, since only a few components (often two or three) will be used in predictions, PLS coefficients may be computed even when there may have been more original X variables than observations. |
| O-PLS | The Orthogonal Projections to Latent Structures (O-PLS) is a linear regression method similar to PLS. However, the interpretation of the models is improved because the structured noise is modeled separately from the variation common to X and Y. Therefore, the O-PLS loading and regression coefficients allow for a more realistic interpretation than PLS, which models the structured noise together with the correlated variation between X and Y. Furthermore, the orthogonal loading matrices provide the opportunity to interpret the structured noise. |
| PLS-DA | PLS-Discriminant Analysis (PLS-DA) is a frequently used classification method that is based on the PLS approach, in which the dependent variable is chosen to represent the class membership. PLS-DA makes it possible to accomplish a rotation of the projection to give latent variables that focus on class separation. The objective of PLS-DA is to find a model that separates classes of objects on the basis of their X-variables. This model is developed from the training set of objects of known class membership. The X-matrix consists of the multivariate characterization data of the objects. To encode a class identity, one uses as Y-data a matrix of dummy variables, which describe the class membership. A dummy variable is an artificial variable that assumes a discrete numerical value in the class description. The dummy matrix Y has G columns (for G classes) with ones and zeros, such that the entry in the gth column is one and the entries in other columns are zero for observations of class g. |
| ANN | Artificial Neural Networks (ANN) is a method, or more precisely a set of methods, based on a system of simple identical mathematical functions, that working in parallel yield for each multivariate input X a single or multiresponse answer. ANN methods can only be used if a comparably large set of multivariate data is available which enables ANN training by example and work best if they are dealing with nonlinear relationships between complex inputs and outputs. The main component of a neural network is the neuron. Each neuron has an activation threshold, and a series of weighted connections to other neurons. If the aggregate activation a neuron receives from the neurons connected to it exceeds its activation threshold, the neuron fires and relays its activation to the neurons to which it is connected. The weights associated with these connections can be modified by training the network to perform a certain task. This modification accounts for learning. ANN are often organized into layers, with each layer receiving input from one adjacent layer, and sending it to another. Layers are categorized as input layers, output layers, and hidden layers. The input layer is initialized to a certain set of values, and the computations performed by the hidden layers update the values of the output layers, which comprise the output of the whole network. Learning is accomplished by updating the weights between connected neurons. The most common method for training neural networks is back propagation, a statistical method for updating weights based on how far their output is from the desired output. To search for the optimal set of weights, various algorithms can be used. The most common is gradient descent, which is an optimization method that, at each step, searches in the direction that appears to come nearest to the goal. |
| SOM | Self-Organizing Maps (SOM) or Kohonen network is an unsupervised neural network method which has both clustering and visualization properties. It can be used to classify a set of input vectors according to their similarity. The result of such a network is usually a two-dimensional map. Thus, SOM is a method for projecting objects from a high dimensional data space to a two-dimensional space. This projection enables the input data to be partitioned into "similar" clusters while preserving their topology, that is, points that are close to one another in the multidimensional space are neighbors in the two-dimensional space as well. |
| SVM | Support Vector Machines (SVM) perform classification by constructing an N-dimensional hyperplane that optimally separates the data into two categories. A SVM model using a sigmoid kernel function is equivalent to a two-layer, perceptron neural network. The task of choosing the most suitable representation is known as feature selection. A set of features that describes one object is called a vector. The goal of SVM modeling is to find the optimal hyperplane that separates clusters of vectors in such a way that objects with one category of the target variable are on one side of the plane and objects with the other category are on the other side of the plane. The vectors near the hyperplane are the support vectors. |

TABLE 1: Continued.

| Technique | Description |
|--------------------|--|
| K-means | K-means is a classic clustering technique that aims to partition objects into k clusters. First, you specify k, that is, how many clusters are being sought. Then, k points are chosen at random as cluster centers. All objects are assigned to their closest cluster center according to the ordinary Euclidean distance metric. Next, the centroid, or mean, of the objects in each cluster is calculated. These centroids are taken to be the new center values for their, respective clusters. Finally, the whole process is repeated with the new cluster centers. Iteration continues until the same points are assigned to each cluster in consecutive rounds, at which stage the cluster centers have stabilized. |
| Genetic Algorithms | Genetic algorithms are nondeterministic stochastic search/optimization methods that utilize the evolutionary concepts of selection, recombination or crossover, and mutation into data processing to solve a complex problem dynamically. Possible solutions to the problem as so-called artificial chromosomes, which are changed and adapted throughout the optimization process until an optimum solution is obtained. A set of chromosomes is called population and creation of a population from a parent population is called generation. In a first step, the original population is created. For each chromosome, the fitness is determined and a selection algorithm is applied to choose chromosomes for mating. These chromosomes are then subject to the crossover, and the mutation operators, which finally yield a new generation of chromosomes. |

or in a pathway diagram are linked to HMDB, DrugBank or UniProt, respectively. One of the most interesting search options in SMPDB is the SMP-MAP which offers both multiidentifier searches as well as transcriptomic, proteomic, or metabolomic mapping. SMP-MAP allows users to select the type of “-omic” data, then paste in a list of identifiers and have a table generated of appropriately highlighted pathways containing those components.

The content of SMPDB is not normally found in other pathway databases with 281 unique pathways (in total of 364). More specifically, 154/168 drug pathways, 11/13 metabolite signalling pathways, 4/70 metabolic pathways and 112/113 metabolic disease pathways of the SMPDB cannot be found in any of the known databases (KEGG, Reactome, EHMN, WikiPathways, HumanCyc, BioCarta, and PharmGKB). Especially in relation to metabolic disease pathways and drug pathways the SMPDB is currently the only pathway database that includes significant numbers of them. In addition SMPDB offers a significant amount of useful graphical content including the depiction of the relevant organs, cellular locations, organelles, cofactors and other cellular features. Because SMPDB is focused on small molecules, it does not include the key protein signalling pathway information which limits significantly its use in comparative metabolic studies, protein network analysis, metabolic engineering or metabolic evolution.

2.3. Toxin and Toxin-Target Database (T3DB). As the name indicates, T3DB is primarily intended to be a database that links toxins with their biological targets. However, the molecular interaction information is further supplemented with detailed descriptions of the toxin’s mechanism of action, its metabolism in the human body, its lethal or toxic dose levels, its potential carcinogenicity, exposure sources, symptoms or health effects and suggested treatment options. More than 2,900 toxin entries corresponding to more than 34,000 different synonyms are currently included in the T3DB. T3DB toxins were identified using a number of methods that include data mining, literature surveys, toxicology textbooks but also examining lists of controlled or banned substances. The toxic compounds that were identified were

subsequently used to derive additional substances that were toxic by relation. In order to ensure both completeness and correctness each toxin record entered in T3DB was reviewed by two different members of the team. Much of the annotation was done manually especially in areas such as route of delivery, mechanisms of action, health effects and target identification.

T3DB contains compounds that have been routinely identified as hazardous in relatively low concentrations (<1 mM for some, <1 μ M for others) and which appear on multiple toxin/poison lists provided by government agencies such as TOXNET or the toxicological and medical literature. In each case, the toxicity of each compound was assessed by examining the available toxicity measurements and health effects, such as minimum lethal dose, LD50, LC50 values and carcinogenicity. In addition these toxins are further connected to approximately 1,300 protein targets through almost 33,500 toxin and toxin-target bonds. All the above information is supported by more than 3,100 references. To facilitate browsing, the T3DB is divided into synoptic summary tables which, in turn, are linked to more detailed “Tox-Cards”-in analogy to the very successful “DrugCard” concept found in DrugBank. Each Tox-Card entry contains over 80 data fields, with ~50 data fields devoted to chemical and toxicological/medical data and ~30 data fields devoted to describing the toxin target(s). In addition to the data viewing and sorting features, the T3DB also offers a local BLAST search that supports both single and multiple sequence queries, a boolean text search based on KinoSearch (<http://www.rectangular.com/kinosearch/>), a chemical structure utility based on ChemAxon’s Marvin-View, and a relational data extraction tool similar to that found in DrugBank and the HMDB. The SeqSearch, a sequence searching utility of T3DB’s, provides the option to search through T3DB’s collection of 1,300 known human toxin targets. The SeqSearch makes possible the identification of both orthologous and paralogous targets for known toxins or toxin targets but facilitates also the identification of potential targets of other animal species. The T3DB’s data extraction utility employs a simple relational database system that allows users to select one or more data fields and

to search for ranges, occurrences or partial occurrences of words, strings or numbers.

In comparison to other databases that contain toxic substances T3DP probably has the smallest number of toxins or poisons in its collection since T3DB was designed as a database for common toxins and not for all known toxic substances. A key focus of the T3DB is on providing “depth” over “breadth” and with its unique emphasis on “common” substances should prove to be a valuable resource in toxicometabolomics and clinical toxicology research.

3. Identification of Disease Biomarkers

In clinical metabolomics one is almost always working with a biofluid or a fluidized tissue extract. The preference of working with biofluids over tissues is primarily dictated by the fact that fluids are far easier to process and analyze. Likewise the collection of biofluids is generally much less invasive than the collection of tissues. Biofluids analysis is always done with the assumption that the chemicals found in different biofluids are largely reflective of the biological state of the organ that produces or is bathed in this fluid. Metabolomics share many of the computational needs with genomics, proteomics, and transcriptomics. All four “-omics” techniques require electronically accessible and searchable databases, all of them require software to handle or process data from their own high-throughput instruments, all of them require laboratory information systems to manage their data and all require software tools to predict or model properties, pathways, relationships, and processes [14]. In terms of data analysis, metabolomics, like other functional genomics technologies, produces high-dimensional datasets, and so it is amenable to many of the analyses applied to microarray data. Statistical modelling (Table 2) range from univariate statistical testing to multivariate regression methods such as principal component analysis (PCA), partial least squares (PLS) or orthogonal projections to latent squares (OPLS), cluster analysis, machine-learning techniques and nonlinear methods, for example Kohonen’s self organizing maps (SOM), support vector machines (SVM), and neural networks (NN) [15–18]. In the following section we have chosen to focus on specific applications that demonstrate how the above statistical modelling tools can facilitate the diagnosis of diseases and the identification of potential biomarkers for use within disease diagnosis.

3.1. Cancer. The paper of Guan et al. [19] is the first application of SVMs and SVM-related feature selection methods (recursive feature elimination with linear and nonlinear kernel, L1SVM, and Weston’s method) for classifying LC/TOF MS data of serum samples from ovarian cancer patients and control. Sera from 37 ovarian cancer patients and 35 benign controls were studied and three evaluation processes (leave-one-out-cross-validation, 12-fold-cross-validation, and 52–20-split-validation) were used to examine the SVM models based on selected potential metabolic diagnostic biomarkers in terms of their ability for differentiating control versus disease serum samples. Classification of the serum sample

test set was over 90% accurate indicating promise that this approach may lead to the development of an accurate and reliable metabolomic-based protocol for detecting ovarian cancer.

The aim of another recent study [20] was to elucidate the predictability of breast cancer by means of urinary excreted nucleosides. The authors analyzed a balanced set of 170 urine samples, 85 breast cancer women and, respective healthy controls, and after identification of 51 nucleosides/ribosylated metabolites in the urine of breast cancer women a valid set of 35 candidates was selected for subsequent computational analysis. The bioinformatic tool of Oscillating Search Algorithm for Feature Selection (OSAF) was applied to iteratively improve features for training of SVMs to better predict breast cancer. The authors found a reasonable set of tumor-related metabolite pairs with SVM prediction performance of 83.5% sensitivity and 90.6% specificity, demonstrating that semiquantitative measurements are valuable for pattern detection using nonparametric machine-learning algorithms.

Arakaki et al. [21] described CoMet, a fully automated and general computational metabolomics method that uses a Systems Biology approach to predict the human metabolites which intracellular levels are more likely to be altered in cancer cells. The authors then prioritize the metabolites predicted to be lowered in cancer compared to normal cells as potential anticancer agents. They discovered eleven metabolites that either alone or in combination exhibit significant antiproliferative activity in Jurkat leukemia cells. Nine of these metabolites that were predicted to be lowered in Jurkat cells with respect to lymphoblasts (riboflavin, tryptamine, 3-sulfino-L-alanine, menaquinone, dehydroepiandrosterone, α -hydroxystearic acid, hydroxyacetone, seleno-L-methionine and 5,6-dimethylbenzimidazole) exhibited antiproliferative activity that has not been reported before. These results strongly suggest that many other metabolites with important roles in cellular growth control may be waiting to be discovered, opening up the possibility of novel approaches against cancer. CoMet adopts the viewpoint that the cell is an integrated machine and the author’s resulting simple hypothesis that inspired its creation can greatly assist in the understanding of the contribution of metabolism to this complex disease.

In a different approach using an animal model Southam et al. [22] applied NMR-based metabolomics to histopathologically well-characterized livers dissected from a wild-caught species of marine flatfish. The use of metabolic profiling and correlation networks enabled a more thorough interpretation of this dataset. Fingerprint analysis identified single metabolites that showed concentration changes between phenotypes, while network analysis highlighted alterations to the relationships of paired metabolites between phenotypes. Tumor tissues showed elevated anaerobic respiration and reduced TCA cycle activity, while alanine and proline were indicated to supplement pyruvate (and NAD⁺) production during anaerobic metabolism in the tumor tissue. Choline metabolism was altered in tumor including disruptions of the choline oxidation and CDP-choline pathways. The author’s hypothesis was that such disruption of the choline

TABLE 2: Freely available databases on metabolic pathways and the metabolome.

| Metabolic Pathways Databases | Webpage |
|--|---|
| BRENDA, the enzyme database, has comprehensive information on enzymes and enzymatic reactions. It is one of several databases nested within the metabolic pathway database set of the SRS5 sequence retrieval system at EBI. | http://www.brenda.uni-koeln.de/ |
| Reactome is an online bioinformatics database of biology described in molecular terms. The largest set of entries refers to human biology, but Reactome covers a number of other organisms as well. It is an on-line encyclopedia of core human pathways-DNA replication, transcription, translation, the cell cycle, metabolism, and signaling cascades. | http://www.reactome.org/ |
| KEGG Metabolic Pathways include graphical pathway maps for all known metabolic pathways from various organisms. Ortholog group tables, containing conserved, functional units in a molecular pathway or assembly, as well as comparative lists of genes for a given functional unit in different organisms, are also available. | http://www.genome.jp/kegg/metabolism.html |
| MetaCyc is a database of nonredundant, experimentally elucidated metabolic pathways. MetaCyc contains more than 1,400 pathways from more than 1,800 different organisms, and is curated from the scientific experimental literature. MetaCyc contains pathways involved in both primary and secondary metabolism, as well as associated compounds, enzymes, and genes. | http://metacyc.org/ |
| The WIT Metabolic Reconstruction project produces metabolic reconstructions for sequenced, or partially sequenced, genomes. It currently provides a set of over 25 such reconstructions in varying states of completion. Over 2900 pathway diagrams are available, associated with functional roles and linked to ORFs. | http://ergo.integratedgenomics.com/ |
| BioCarta website provides gene interactions in dynamic graphical models. The online maps depicts molecular relationships and it catalogs and summarizes important resources providing information for more than 12,000 genes from multiple species. It contains both classical pathways as well as suggestions for new pathways. | http://main.biocarta.com/genes/index.asp |
| EcoCyc describes the genome and the biochemical machinery of E. coli. It provides a molecular and functional catalog of the E. coli cell to facilitates system-level understanding. Its Pathway/Genome Navigator user interface visualizes the layout of genes, of individual biochemical reactions, or of complete pathways. It also supports computational studies of the metabolism, such as pathway design, evolutionary studies, and simulations. A related metabolic database is Metalgen. | http://ecocyc.org/ |
| BioSilico is a web-based database system that facilitates the search and analysis of metabolic pathways. Heterogeneous metabolic databases including LIGAND, ENZYME, EcoCyc and MetaCyc are integrated in a systematic way, thereby allowing users to efficiently retrieve the relevant information on enzymes, biochemical compounds and reactions. In addition, it provides well-designed view pages for more detailed summary information. | http://mbel.kaist.ac.kr/lab/index_ko.html |
| EXPASY - Biochemical Pathways is a searchable database of metabolic pathways, enzymes, substrates and products. Based on a given search, it produces a graphic representation of the relevant pathway(s) within the context of an enormous metabolic map. Neighboring metabolic reactions can then be viewed through links to adjacent maps. | http://www.expasy.ch/cgi-bin/search-biochem-index |
| BioPath is a database of biochemical pathways that provides access to metabolic transformations and cellular regulations derived from the Roche Applied Science "Biochemical Pathways" wall chart. BioPath provides access to biological transformations and regulations as described on the "Biochemical Pathways" chart. | http://www.molecular-networks.com/biopath/ |

TABLE 2: Continued.

| Metabolic Pathways Databases | Webpage |
|---|---|
| BioCyc is a collection of 505 Pathway/Genome Databases. Each database in the BioCyc collection describes the genome and metabolic pathways of a single organism. The BioCyc Web site contains many tools for navigating and analyzing these databases, and for analyzing omics data, including the following: Genome browser, Display of individual metabolic pathways, and of full metabolic maps, Visual analysis of user-supplied omics datasets by painting onto metabolic map, regulatory map, and genome map, Comparative analysis tools. | http://biocyc.org/ |
| Metabolome Databases | Webpage |
| The Biological Magnetic Resonance Data Bank (BMRB) focuses on quantitative data generated by spectroscopic investigations of biological macromolecules. It has links to search engines such as PubChem, that connect to recent articles and new data. It also links to projects and other databases that are all related to Metabolomics and Metabonomics. This database focuses on the NMR research aspect of metabolites discovery and their role in metabolism. | http://www.bmrwisc.edu/metabolomics/ |
| The Madison Metabolomics Consortium Database contains metabolites determined through NMR and MS. It contains information with the main focus on Arabidopsis thaliana, but also refers to many different species. The database also contains information on the presence of metabolites under several different physiological conditions, their structures in 2D and 3D, and links to related resource sources and other databases. | http://mmcd.nmrwisc.edu/ |
| The Human Metabolome Database is an extremely comprehensive, free electronic database that gives a detailed overview of human metabolites divided into chemical, clinical, and molecular biology/biochemistry data. | http://www.hmdb.ca/ |
| KNAPSAcK is a Java application that presents an interactive display of biochemical information that can be searched by organism or metabolite name. KNAPSAcK focuses primarily on the origin and mass spectra of particular metabolites. | http://kanaya.naist.jp/KNAPSAcK |
| The BiGG database is a metabolic reconstruction of human metabolism designed for systems biology simulation and metabolic flux balance modeling. It is a comprehensive literature-based genome-scale metabolic reconstruction that accounts for the functions of 1,496 ORFs, 2,004 proteins, 2,766 metabolites, and 3,311 metabolic and transport reactions. It was assembled from build 35 of the human genome. | http://bigg.ucsd.edu/ |
| SetupX, developed by the Fiehn laboratory at UC Davis, is a web-based metabolomics LIMS. It is XML compatible and built around a relational database management core. It is particularly oriented towards the capture and display of GC-MS metabolomic data through its metabolic annotation database called BinBase. | http://fiehnlab.ucdavis.edu:8080/m1/ |
| McGill-MD is a metabolome database containing metabolite mass spectra of organisms; with abiotic/biotic stress or in homeostasis. Users are able to obtain a table containing the metabolome of an organism, or download mass spectra of all the metabolites entered in the database. | http://metabolomics.mcgill.ca/ |
| SYSTEMONAS (SYSTEMs biology of pseudOMONAS) is a database for systems biology studies of Pseudomonas species. It contains extensive transcriptomic, proteomic and metabolomic data as well as metabolic reconstructions of this pathogen. Reconstruction of metabolic networks in SYSTEMONAS was achieved via comparative genomics. Broad data integration with well established databases BRENDA, KEGG and PRODORIC is also maintained. | http://www.systemonas.de/ |

TABLE 2: Continued.

| Metabolic Pathways Databases | Webpage |
|--|---|
| MassBank is a mass spectral database of experimentally acquired high resolution MS spectra of metabolites. Maintained and supported by the JST-BIRD project, it offers various query methods for standard spectra obtained from Keio University, RIKEN PSC, and other Japanese research institutions. It is officially sanctioned by the Mass Spectrometry Society of Japan. The database has very detailed MS data and excellent spectral/structure searching utilities. More than 13,000 spectra from 1900 different compounds are available. | http://www.massbank.jp/ |
| The Golm Metabolome Database provides public access to custom GC/MS libraries which are stored as Mass Spectral (MS) and Retention Time Index (RI) Libraries (MSRI). These libraries of mass spectral and retention time indices can be used with the NIST/AMDIS software to identify metabolites according their spectral tags and RI's. The libraries are both searchable and downloadable and have been carefully collected under defined conditions on several types of GC/MS instruments (quadrupole and TOF). | http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/gmd.html |
| The METLIN Metabolite Database is a repository for mass spectral metabolite data. All metabolites are neutral or free acids. It is a collaborative effort between the Siuzdak and Abagyan groups and Center for Mass Spectrometry at The Scripps Research Institute. METLIN is searchable by compound name, mass, formula or structure. It contains 15,000 structures, including more than 8000 di and tripeptides. METLIN contains MS/MS, LC/MS and FTMS data that can be searched by peak lists, mass range, biological source and or disease. | http://metlin.scripps.edu/index.php |

oxidation pathway could lead to reduced SAM production and potentially DNA hypomethylation of oncogenes.

3.2. Diabetes. The paper of Altmaier et al. [23] presents a bioinformatics analysis of what can be considered as a standard experimental setting of a preclinical drug testing experiment with two independent factors, “state” and “medication”. Targeted quantitative metabolomics covering a wide range of more than 800 relevant metabolites were measured in blood plasma samples from healthy and diabetic mice under rosiglitazone (a member of thiazolidinedione) treatment. The authors show that known and new metabolic phenotypes of diabetes and medication can be recovered in a statistically objective manner. Analyzing ratios between metabolite concentrations dramatically reduces the noise in the data set allowing the discovery of new potential biomarkers of diabetes, such as the N-hydroxyacyloylsphingosylphosphocholines SM(OH)28:0 and SM(OH)26:0. Using a hierarchical clustering technique on partial η^2 values the authors identified functionally related groups of metabolites, indicating a diabetes-related shift from lysophosphatidylcholine to phosphatidylcholine levels.

Coupled LC/MS technology to multivariate statistical analysis in order to study phospholipid metabolic profiling in diabetes mellitus and to discover the potential biomarkers was the approach of Wang et al. [24]. PCA and PLS-DA models were compared in class separation of type 2 diabetes mellitus (DM2) patients and healthy controls.

Uv (unit variance) scaling and OSC (orthogonal signal correction) data preprocessing methods were also developed to improve class separation. Using the supervised PLS-DA algorithm with Uv scaling and OSC technique on the data set, it was found that the separation of different classes was highly improved (compared to PCA analysis) particularly with OSC. The application of LC/MS coupled to PLS-DA of data with OSC scaling made it possible to classify DM2 and control and further to discover potential biomarkers that can be identified by MS/MS.

NMR-based metabolomics coupled with sophisticated bioinformatics was shown capable of identifying rapid changes in global metabolite profiles in urine and plasma (treatment “fingerprints”) which may be linked to the well-documented early changes in hepatic insulin sensitivity following thiazolidinedione intervention in Type 2 diabetes mellitus [12]. Several endogenous metabolites in urine and plasma of T2DM patients that responded to rosiglitazone treatment were identified. In urine these changes were related to a gender-independent relative reduction of hippurate and a further increase of aromatic acids. The gender-dependent changes observed in plasma samples included an increase in branched chain amino acids, alanine, glutamine/glutamate and citrate, coinciding with a decrease in lactate, acetate, tyrosine, and phenylalanine in the female T2DM group, where changes in the male T2DM group included an increase in branched chain amino acids, alanine, glutamine, and threonine. A good distinction between diabetic patients

and healthy volunteers as well as separation by gender was accomplished when Supervised Principal Component Discriminant Analysis (PC-DA) of plasma or urine samples was applied which comprises an important new addition to the early clinical development “proof of concept” toolbox for thiazolidinediones.

Diabetes is associated with increased incidence of vascular complications, and premature aging. In the study of Makinen et al. [25], the emphasis was on the metabolic continuum that underlies the slow and often elusive development of chronic complications. The authors obtained serum samples to measure two molecular windows, —the lipoprotein lipids (LIPO) window and the low molecular weight molecules (LMWM)—for 613 patients with type I diabetes, and diverse spread of complications. The H-NMR analyses combined with SOM instead of linear decomposition methods allowed the authors to transform the spectral data into an accessible form of information. The work of Makinen et al. demonstrated the limitations of single diagnostic biomarkers and illustrated a fundamental diagnostic challenge. Even though there is a common biochemical basis of diabetic kidney disease, diabetic retinal disease, the metabolic syndrome, and macrovascular diseases however they do not conclusively define each other.

Salek et al. [26] describe the application of ¹H-NMR spectroscopy-based metabolomics, combined with multivariate and univariate statistics, to investigate the urinary metabolic profiles in two animal models (mice and rat) of T2DM, and they compared these metabolic changes with perturbations observed in a human population. This study demonstrated metabolic similarities between the three species examined. Along with the expected changes in hepatic glycolysis/ gluconeogenesis changes in the excretion of TCA cycle intermediates, polyols, amines, and amino acids were detected. Furthermore significant changes in pyruvate and fatty acid metabolism as well as hepatic amino acid metabolism were observed including tryptophan metabolism. A profound perturbation in nucleotide metabolism, previously linked with peroxisome proliferation, was also observed and may indicate a metabolic consequence of substrate excess in many tissues, especially the liver.

In the study of Connor et al. [27], the authors have generated NMR-based metabolomic and transcriptomic data from the db/db diabetic mouse, one of the most extensively studied animal models of T2D. Db/db mice lack a functioning leptin receptor resulting in defective leptin-mediated signal transduction. Metabolomics data identified 24 distinct pathways that were altered in the diabetic mice when compared to their euglycaemic littermates. Several of these pathways were related to known disease effects, but in addition novel effects on branched chain amino acid metabolism, nicotinamide metabolites, pantothenic acid, and gut microflora metabolism were also observed. Integrative pathway analysis of the metabolite-centric networks and the cross-platform transcriptomics and metabolomics results effectively linked many of the metabolite changes to pathways involved in gluconeogenesis, and those generating substrates for gluconeogenesis, mitochondrial dysfunction

and oxidative stress, and altered protein turnover. Overall, these metabolites are likely reflective of additional underlying pathophysiology that is present in T2D.

The objective of Lanza et al. [28] was to illustrate the utility of a combination of analytical methods and multivariate statistical analysis for detecting a metabolic fingerprint that reflects known pathways that are altered with insulin deficiency. The authors analyzed plasma from type 1 diabetic (T1D) humans during insulin treatment (I+) and acute insulin deprivation (I-) and nondiabetic participants (ND) and they generated correlation matrices for the plasma metabolites measured by both MS and NMR to create a compendium metabolic profile that integrates the complementary information from the two analytical methods. Multivariate statistics differentiated proton spectra from I- and I+ based on several derived plasma metabolites that were elevated during insulin deprivation (lactate, acetate, allantoin, and ketones) as well as several underlying physiological processes that are known to be altered by short-term insulin deprivation in type 1 diabetic people (e.g., mitochondrial dysfunction, oxidative stress, protein synthesis, degradation, and oxidation, gluconeogenesis, and ketogenesis).

Bao et al. [29] performed a metabolomic study to determine metabolic variations associated with T2DM and the drug treatments on 74 patients who were newly diagnosed with T2DM and received a 48-week treatment of a single drug, repaglinide, metformin, or rosiglitazone. A total of 212 individual metabolites were consistently detected in at least 90% of the serum samples and orthogonal projections to latent structures discriminant analysis, a newly developed supervised pattern recognition method, was used to capture the subtle intergroup variations and establish a prediction model to assess the physiological impact by drug treatment. As compared to healthy controls, the altered serum metabolites in diabetic subjects, include the significantly increased valine, maltose, glutamate, urate, butanoate, and long-chain fatty acid (C16:0, C18:1, C18:0, octadecanoate, and arachidonate), and decreased glucuronolactone, lysine, and lactate suggesting a hypercatabolic state in T2DM patients. Rosiglitazone treatment was able to reverse more abnormally expressed metabolites, such as valine, lysine, glucuronolactone, C16:0, C18:1, urate, and octadecanoate, than the other two drugs.

Čuperlović-Culf [30] presented an application of fuzzy *K*-means (F-KM) method for the classification of metabolic profiles of urine samples in diabetic patients. F-KM is a fuzzy version of standard *K*-means clustering. In F-KM clustering, each sample has an overall membership, that is, sum of membership values for all clusters, of 1. This overall membership is appointed to clusters based on the similarity between the sample's metabolic fingerprint and the profile of cluster's centroid. From the membership values, it is then possible to determine different levels of coclustering between samples-based on the top membership, second highest membership, and so forth. In their work different clustering methods were compared with F-KM. For human type II diabetes and healthy phenotypes membership values, F-KM lead to better sample separation while it was the only

method that allowed distinction on both major groups and sample subtypes.

3.3. Neurological and Other Diseases. The study of Rozen et al. [4] was designed to assess whether there are systematic differences between redox-active metabolites in the blood of patients with motor neuron disease (MND) and healthy controls by analyzing the blood plasma of 30 healthy controls and 28 individuals with MND. To determine which metabolites were significantly elevated or reduced in MND the authors used three measures of class association, the *t*-statistic, Pearson's correlation coefficient, and the "relative class association" measure. All three measures produced similar rankings of their metabolites by their level of association with MND versus control. The authors assessed statistical significance by permutation testing and all measures showed similar numbers of metabolites to have significantly higher or lower concentrations in MND compared to controls. Subsequently they analyzed these data to determine if the metabolites were capable of distinguishing four subgroups (normal controls MND patients taking riluzole medication, MND without riluzole medication, and the subgroup enriched for LMN-lower motor neuron disease) using the 317 metabolite concentrations. Using PLS-DA, a supervised projection technique, the authors found a three-dimensional projection in which these four subgroups were significantly separated.

¹H nuclear magnetic resonance spectroscopy in conjunction with computerized pattern recognition analysis were employed to investigate metabolic profiles of a total of 152 cerebrospinal fluid (CSF) samples from drug-naïve or minimally treated patients with first-onset paranoid schizophrenia and healthy controls [7]. Plots of PLS-DA scores showed a clear differentiation between healthy volunteers and drug-naïve patients. The PLS-DA score plots show that atypical antipsychotic drug treatment results in a shift of approximately 50% of patients with schizophrenia towards the cluster of healthy controls. A striking finding of this study is the effect of the number of psychotic episodes prior to commencing antipsychotic treatment on the CSF metabolite profile in patients with schizophrenia. Of 21 patients who commenced antipsychotic medication during their first psychotic episode, 57% clustered with healthy controls whereas six out of the seven patients who had several psychotic episodes prior to treatment clustered with the group of drug-naïve patients with first-onset schizophrenia. These results suggest that the initiation of antipsychotic treatment during a first psychotic episode may influence treatment response and/or indeed outcome.

Pre-eclampsia is an important cause of maternal morbidity and mortality while the World Health Organization estimates that worldwide over 100,000 women die from pre-eclampsia each year. By using GC-tof-MS the authors [31] were able to separate and detect several hundred metabolites from both control (87) and diseased (87) samples. The application of genetic algorithms on these data indicated that the pre-eclamptic plasma could be discriminated from the matched controls on the basis of just three metabolite peaks (two of which tended to be lower and one tended to

be higher in the samples from women with pre-eclampsia, and to a certain extent this correlated with the severity of the disease). In this context it is worth commenting that genetic algorithms is advantageous over other machine-learning methods such as neural networks and support vector machines, as it allows one to understand the problem in terms of small subsets of input variables that it combines into rules. In the case of Kenny and colleagues [31], only 10 of each the disease and control samples were taken at a gestational age of under 30 weeks, and a clear task for the future is to establish the extent to which these diagnostic rules apply earlier in pregnancy and thus are of greater prognostic value.

A metabolic "bioprofile" consisting of predictive serum metabolite features from ¹H NMR spectral data of the murine K/BxN model of arthritis were presented in the study of Weljie et al. [32]. A unique method was developed by combining technologies such as quantitative targeted profiling, O-PLS-DA pattern recognition analysis and metabolic-pathway-based network analysis for interpretation of results. In total, 88 spectral features were profiled (59 metabolites and 28 unknown resonances). A highly significant subset of 18 spectral features (15 known compounds and 3 unknown resonances) was identified and in this metabolic bioprofile, metabolites relating to nucleic acid, amino acid, and fatty acid metabolism, as well as lipolysis, reactive oxygen species generation, and methylation were among them. Pathway analysis suggested a shift from metabolites involved in numerous reactions (hub metabolites) toward intermediates and metabolic endpoints associated with arthritis.

4. Metabolomics in Drug Discovery and Polypharmacology Studies

Drug molecules generally act on specific targets at the cellular level, and upon binding to the receptors, they exert a desirable alteration of the cellular activities, regarded as the pharmaceutical effect. Current drug discovery depends largely on random screening, either high-throughput screening (HTS) in vitro, or virtual screening (VS) in silico. Because the number of available compounds is huge, several drug-likeness filters are proposed to reduce the number of compounds that need to be evaluated. The ability to effectively predict if a chemical compound is "drug-like" or "non-drug-like" is, thus, a valuable tool in the design, optimization, and selection of drug candidates for development [33]. Drug-likeness is a general descriptor of the potential of a small-molecule to become a drug. It is not a unified descriptor but a global property of a compound processing many specific characteristics such as good solubility, membrane permeability, half-life, and having a pharmacophore pattern to interact specifically with a target protein. These characteristics can be reflected as molecular descriptors such as molecular weight, log *P*, the number of hydrogen-bond donors, the number of hydrogen-bond acceptors, the number of rotatable bonds, the number of rigid bonds, the number of rings in a molecule, and so forth [34]. Lipinski's widely used rule of 5 defines drug-like "as those compounds

that have sufficiently acceptable absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties to survive through the completion of Human Phase I clinical trials” [35]. It has been observed that metabolites tend to obey in their majority the Lipinski “Rule of 5”, which hints to the fact that drugs are indirectly synthesized to mimic the original endogenous substrates [36]. Based on this, metabolite-likeness and biological relevance filters have recently been developed, which consider that chemical compounds from virtual screens of large pharmaceutical libraries that are similar to endogenous metabolites stand more chances for being successful drug candidates [37, 38]. The approach leverages the “chemical similarity principle”, which states that molecules with similar structure likely have similar biological properties.

Drug developers have long-mined small-molecule metabolism for the design of enzyme inhibitors chemically similar to their endogenous substrates. The approach has yielded many successes, including antimetabolites such as folate derivatives used in cancer therapy [39] and the nucleoside analog prodrugs used for antiviral therapy [40]. With the recent availability of databases of metabolites and metabolic reactions, we have gained a wealth of information regarding the interaction of small molecules with biological systems. At the same time, the notion of chemical space and the advance of chemoinformatics tools have paved the way to link the metabolome with structural and physicochemical properties of endogenous metabolites and to predict links between synthetic molecules and human metabolism.

Recent developments in the area of systems biology have lead scientists to realize the limitations of reductionism and begin to lay emphasis on more holistic research patterns, such as systems biology and network pharmacology [41–44]. Most diseases are not caused by changes in a single causal gene but by an unbalanced regulating network resulting from the dysfunctions of multiple genes or their products. At the same time, drug molecules commonly participate in biological networks and both their intended effect and side effect are rather systemic than specific to a single biological target.

On this direction, Corey Adams and coworkers have recently demonstrated a new method to predict what enzymes drugs might affect based on the chemical similarity between classes of drugs and the natural chemicals used by enzymes. The authors have applied the method to 246 known drug classes and a collection of 385 organisms to create maps of potential drug action on metabolism. Moreover, they show how the predicted connections can be used to find new ways to kill pathogens and to avoid unintentionally interfering with human enzymes [45].

In the work of Macchiarulo and coworkers, human metabolic pathways are projected and clustered on the chemical space based on similarity of the involved metabolites translated in a set of selected physicochemical and topological descriptors. Further to this, the authors develop a classifier that estimates the proximity of marketed drugs to any given pathway, with the aim to elucidate the extend of overlap and to uncover cross-interactions between drugs and the major human pathways. The model performs well

for tightly clustered, isolated pathways, but it loses its predictive ability when it comes to overlapping pathways [46].

5. Metabolomics for the Study of Polypharmacology of Natural Compounds.

Internationally, there is a growing and sustained interest from both pharmaceutical companies and public in medicine from natural sources. For the public, natural medicine represent a holistic approach to disease treatment, with potentially less side effects than conventional medicine. For the pharmaceutical companies, bioactive natural products constitute attractive drug leads, as they have been optimized in a long-term natural selection process for optimal interaction with biomolecules. To promote the ecological survival of plants, structures of secondary products have evolved to interact with molecular targets affecting the cells, tissues and physiological functions in competing microorganisms, plants, and animals. In this, respect, some plant secondary products may exert their action by resembling endogenous metabolites, ligands, hormones, signal transduction molecules, or neurotransmitters and thus have beneficial effects on humans due to similarities in their potential target sites [47].

Complementary to the above studies on drug polypharmacology and in order to elucidate the extend of overlap and similarity between natural compounds from plants used in ethnomedicine and human metabolites, we created chemical networks between natural compounds from the Chinese Natural Products Database (CNPD v.2004.1) and human metabolites from HMDB. CNPD is a compilation of 57,346 compounds found in plants largely used in TCM (Traditional Chinese Medicine). These compounds come from 2,611 plant species belonging to 457 different plant genera. After removal of salts, inorganic compounds, and duplicates, we extracted 53,180 unique, organic compounds in SDF format, which we imported into a Molecular Operating Environment (MOE, v.2008.10) [48] database. 1417 of these compounds are annotated with experimentally derived bioactivity information. HMDB v. 2.5 was used as source of human metabolites and 7,985 compounds were extracted in SDF format. All structures were washed, that is all ionizable groups were coordinated with neutral pH conditions, and energy minimized using the MMFF946 force field.

To get a first overview, we compared the two databases considering common descriptors for drug-like molecules, namely molecular weight (MW), number of hydrogen-bond donors (HB donors), number of hydrogen-bond acceptors (HB acceptors), number of rings and number of rotatable bonds. As seen in the violin plots of Figure 3, the human metabolites have higher average molecular weight (MW = 661.2) and broader distribution (std dev = 403.4), which is obviously due to the presence of many lipids (3800 out of 7985 compounds are lipids in the newest version of the HMDB) [49]. The number of HB-donors is almost the same in both CNPD and HMDB sets, with an average value of 2.4 and 2.5, respectively and 90% of the compounds in each data

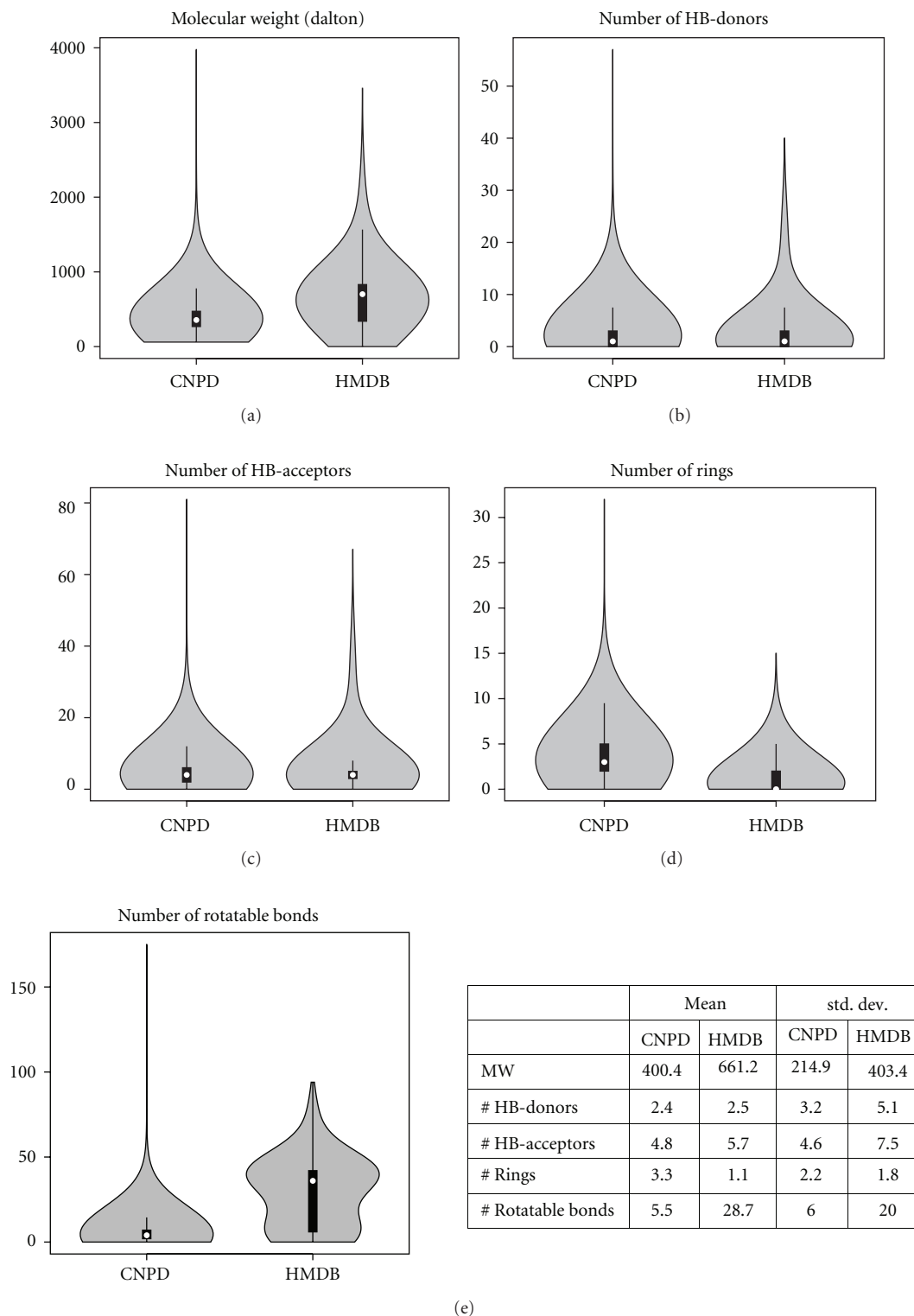


FIGURE 3: Comparison of the distribution of selected druglike molecular properties for natural compounds from CNPD and human metabolites from HMDB. Violin plots for (a) molecular weight, (b) hydrogen-bond donors, (c) hydrogen-bond acceptors, (d) number of rings and (e) number of rotatable bonds, along with table with mean values and standard deviations. A violin plot is a combination of a box plot and a kernel density plot and offers a more detailed view of a dataset's variability than a box plot alone. The white marker indicates the median of the data and the black box the interquartile range (the difference between the third and first quartiles that contain 50% of the distribution). The black lines extend to one and a half times the width of the box. Violin plots were made in R.

set having ≤ 5 HB-donors. When it comes to HB-acceptors, the two profiles differ slightly. 90% of the compounds in each data set have ≤ 10 HB-acceptors, but there is higher percentage of human metabolites with 9-10 HB-acceptors (6.9%) than are natural compounds (4.7%). Due to this, CNPD has a lower mean value and std deviation (4.8 and 4.6, respectively). The number of rings is lower in HMDB, again because of the presence of lipids that are acyclic. As a consequence, compounds from HMDB have on average many more rotatable bonds than their counterparts from CNPD. While 95% of compounds from CNPD have up to 15 rotatable bonds, half of the compounds from HMDB have between 30 and 50 rotatable bonds.

Despite the differences noted above, there is a significant room for overlap between the chemical spaces of the two datasets, which we attempt to elucidate via a more thorough structural similarity analysis that follows. First we investigate how many compounds are present in both data sets, by comparing their SMILES (Simplified Molecular Input Line Entry Specification) strings. Following that, we analyze the extend of structural similarity of the two data sets. For this, all pairs of molecules between the two sets are compared using a pairwise similarity metric, which consists of a descriptor and a similarity criterion. For the descriptor, MACCS (Molecular ACCess System) keys were calculated in MOE. The MACCS keys represent each molecule as a vector of 166 bits, each indicating the presence or absence of a predefined substructure or functional group (e.g., aromatic rings, oxygens, amine groups, etc.). The similarity criterion is the widely used Tanimoto coefficient (T_c) [50]. T_c is calculated as shown in (1). If two molecules have a and b bits in their fragment bit-strings, respectively, with c of these bits being present in both their fingerprints, then T_c corresponds to the ratio of the number of bits the two molecules have in common to the total number of occupied bins by both molecules

$$T_c = \frac{c}{a + b - c}. \quad (1)$$

T_c gives values in the range of zero (no bits in common, 0% similarity) to unity (all bits the same, 100% similarity). The T_c threshold for two compounds being similar was set to 0.85 and the similarity networks were visualized using the Organic Layout of Cytoscape v. 2.6.3 [51].

5.1. Overlap between Human Metabolites and Natural Compounds. There are 383 compounds shared between the two databases, which denotes that, apart from participating in the human metabolism, these natural compounds are secondary metabolites of plants used in ethnopharmacology. For example, 2-pyrocatechuic acid (HMDB00397) is a normal human benzoic acid metabolite found in plasma that is an intermediate of the phenyl propanoid biosynthesis. It has been isolated from black currant [52], which has long been used in European and Chinese folk medicine as diuretic, treating diarrhea, arthritic pain, and so forth. Recently, 2-pyrocatechuic acid was found to be weak inhibitor of Selectin E [53] and potent inhibitor of 15-lipoxygenase-catalysed oxygenation of arachidonic acid

that is involved in many aspects of inflammatory disease and in particular in the development of colorectal cancer [54].

Another example, indole (HMDB00738), is an aromatic heterocyclic organic compound that occurs naturally in human feces and has an intense fecal smell. At very low concentrations, however, it has a flowery smell and is a constituent of many flower scents. Natural jasmine oil that contains around 2.5% of indole is used traditionally for healing the female reproducing system, to treat headaches and insomnia. In human metabolism, indole participates in the tryptophan metabolic pathway, which is a highly regulated biological process. There has been significant research on the medical implications involved in dysregulation of tryptophan metabolism. Abnormalities in it may play a role in central nervous system diseases such as acquired immunodeficiency syndrome- (AIDS-) related dementia [55], Huntington's disease [56] and psychopathological disorders [57]. In addition, data from the literature suggest that a mechanism dependent on tryptophan catabolism might regulate the immune responses to a number of diseases [58–60].

5.2. Similarity Networks of Human Metabolites and Natural Compounds. There are 15,523 natural compounds in CNPD (29% of the total data set) that have a Tanimoto similarity coefficient of 0.85 or higher with at least one human metabolite. In total, there are formed 233,211 similarity pairs between the two datasets, which indicates that each natural compound is similar—on average—with 15 human metabolites.

As an illustrative example, Figures 4 and 5 below show the similarity networks of 2-pyrocatechuic acid and indole that were discussed in the previous section. As seen in Figure 4, 2-pyrocatechuic acid is linked with $T_c \geq 0.9$ to seven other human metabolites and 28 natural compounds from CNPD. Interestingly, the human metabolites of this similarity network belong to two main metabolic processes. HMDB01866, HMDB06242, HMDA00152, and HMDB01856 are involved in tyrosine metabolism/biosynthesis, while HMDB00397, HMDB03501, and HMDB01964 are intermediates of the phenyl propanoid biosynthesis. Recent research on tyrosine metabolism suggests strong correlation with chronic kidney failure [61], eating disorders and migraine [62]. The natural compounds from CNPD that are met in the network are primarily benzoic acid derivatives from diverse sources of plants (e.g., *picea maximowiczii*, *grevillea robusta*), fungi (e.g., *polyporus tumulosus*, *boletus scaber*) and flowers (e.g., *centaurium erythraea*, *anthemis nobilis*), many of which are known as folk medicine.

Indole, shown in Figure 5, is linked to one other human metabolite, and four natural compounds from CNPD. HMDB00466 is the compound 3-methyl indole that is involved in tryptophan metabolism as well. Three natural compounds from CNPD have high similarity to the two human metabolites. 1-methyl-9H-carbazole (cas: 6510-65-2), 3-methyl-9H-carbazole (cas: 4630-20-0) and 2,4-dimethyl-1H-indole (cas: 10299-61-3) are alkaloids found in *Tedania ignis* (a sponge species) [63], glycosmis pentaphylla

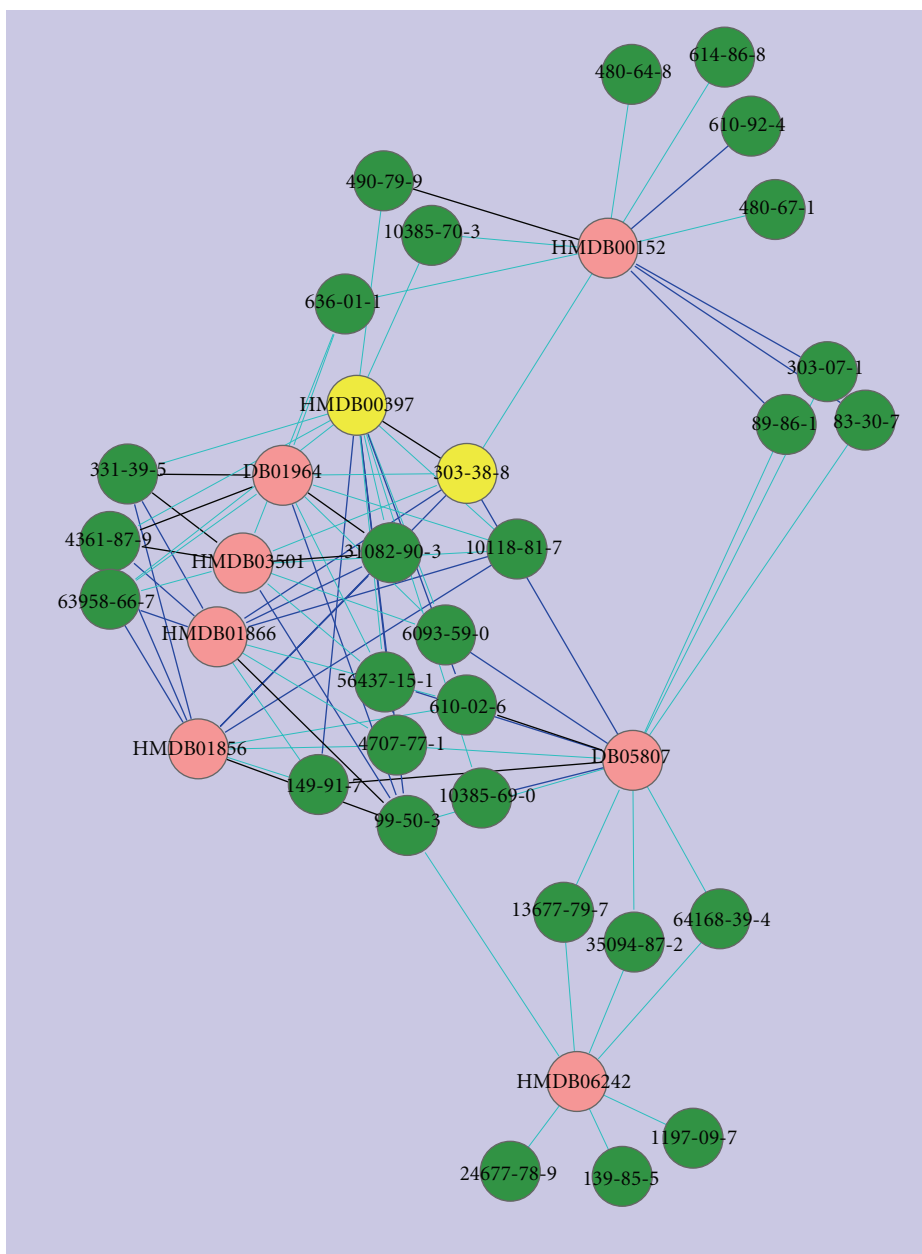


FIGURE 4: Similarity network of 2-pyrocatechuic acid. Pink nodes indicate human metabolites from HMDB and green nodes indicate natural compounds from CNPD. Node labels denote the respective ID codes of the compounds. The nodes are linked when the two compounds have $T_c \geq 0.85$. Due to the high number of pairs with similarity between 0.85 and 0.90, we included in the figure only connections of $T_c \geq 0.90$ to allow better visualization of the network. The width and color of the edges are analogous to the value of T_c : Cyan: $0.90 \leq T_c < 0.95$, Blue: $0.95 \leq T_c < 1.0$, Black: $T_c = 1$. The two nodes in yellow denote 2-pyrocatechuic acid with HMDB ID and CAS registry number, respectively.

(orangeberry) [64], and *Tricholoma virgatum* (a mushroom species) [65], respectively. These natural compounds from CNPD that are found similar to well-studied human metabolites are potentially interesting leads with druglike and metabolite-like properties that would be worth investigating further for their medicinal properties and their impact on human health.

In order to evaluate how the 29% similarity of CNPD to HMDB compares with other types of data sets, we performed the same analysis for 4,567 approved and experimental drugs

from DrugBank v.2, as well as for a randomly selected subset of 59,025 compounds from ChemDiv, a commercial provider of small compounds for drug discovery HTS. Quite remarkably, the compounds from DrugBank showed the same extend of similarity to HMDB as natural compounds from CNPD. 1,331 drug compounds (29%) were found to be similar to human metabolites, forming 35,635 similarity pairs. On the other hand, only 182 compounds from the subset of ChemDiv were found similar to any human metabolite, forming just 1,563 similarity pairs in total.

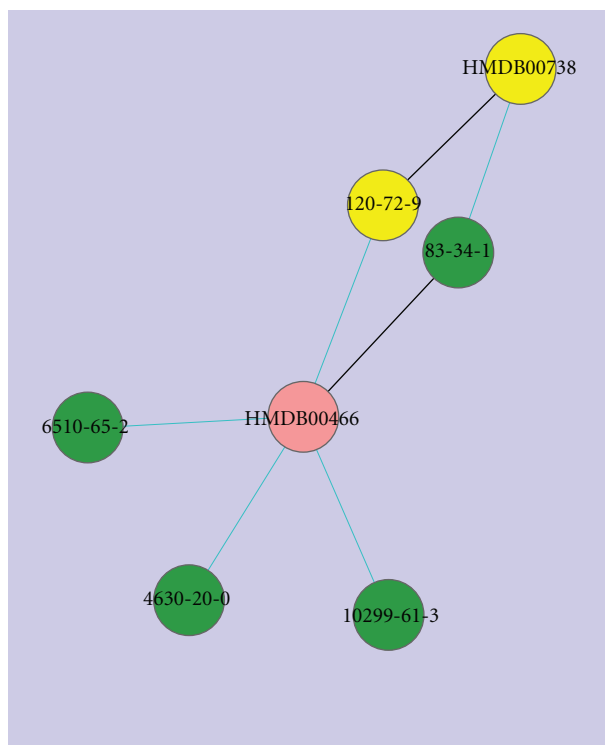


FIGURE 5: Similarity network of indole. Pink nodes indicate human metabolites from HMDB and green nodes indicate natural compounds from CNPD. Node labels denote the respective ID codes of the compounds. The nodes are linked when the two compounds have $T_c \geq 0.85$. The width and color of the edges are analogous to the value of T_c : Cyan: $0.85 \leq T_c < 0.95$, Black: $T_c = 1$. The node in yellow denotes indole with HMDB ID and CAS registry number, respectively.

This low similarity of ChemDiv could be attributed to the fact that HTS databases contain small molecules with simple structures that can be easily modified further to more potent drug candidates. These findings support the hypothesis that as drug candidates move forward on the drug optimization platform, there is favorable selection towards those that mimic the endogenous substrates. The fact that natural compounds also resemble the latter may indicate that plants with medicinal properties may exert their action via their molecular components resembling human endogenous metabolites.

6. Future Perspectives

Metabolomics, the study of metabolism at the global level, is moving to exciting directions. With the development of more sensitive and advanced instrumentation and computational tools for data interpretation in the physiological context, metabolomics have the potential to impact our understanding of molecular mechanisms of diseases. A state-of-the-art metabolomics study requires knowledge in many areas and especially at the interface of chemistry, biology, and computer science. High-quality samples, improvements in automated metabolite identification, complete coverage of the human metabolome, establishment of spectral databases of metabolites and associated biochemical identities, innovative experimental designs to best address a hypothesis, as well as novel computational tools to handle metabolomics

data are critical hurdles that must be overcome to drive the inclusion of metabolomics in all steps of drug discovery and drug development. The examples presented above demonstrated that metabolite profiles reflect both environmental and genetic influences in patients and reveal new links between metabolites and diseases providing needed prognostic, diagnostic, and surrogate biomarkers. The integration of these signatures with other omic technologies is of utmost importance to characterize the entire spectrum of malignant phenotype.

Systems chemical biology networks that assemble and integrate known and predicted links between small compounds of biological relevance, including human metabolites, can have a great potential in pharmaceutical research that could be used in a variety of ways. Novel ligands can be selected on the premise of being similar to endogenous metabolites with the desired bioactivity profile. Pathways for orphan metabolites could be predicted, based on their similarity with compounds of known biological target and mode of action. New ways to kill pathogens and to avoid unintentionally interfering with human enzymes can be investigated and cross-interactions between drugs and the major human pathways can be unravelled. Last but not least, one could predict the biological targets of bioactive natural compounds from medicinal plants, by looking at their similarity networks with human metabolites with known biological targets. By adding information about

the metabolic pathways that these metabolites are involved, one could also extract hypotheses regarding the mode of action and therapeutic mechanism of the medicinal plant at the molecular level, which is at the moment the missing link for the coupling of WM with TCM and ethnomedicine in general.

Acknowledgments

The authors acknowledge the funding from the Danish Research Council for Technology and Production Sciences. They thank Sonny Kim Nielsen for the Tanimoto coefficient calculations.

References

- [1] K. Dettmer and B. D. Hammock, "Metabolomics—a new exciting field within the 'omics' sciences," *Environmental Health Perspectives*, vol. 112, no. 7, pp. 396–397, 2004.
- [2] C. W. Schmidt, "What's happening downstream of DNA," *Environmental Health Perspectives*, vol. 112, no. 7, pp. 410–415, 2004.
- [3] J. Van der Greef, S. Martin, P. Juhasz et al., "The art and practice of systems biology in medicine: mapping patterns of relationships," *Journal of Proteome Research*, vol. 6, no. 4, pp. 1540–1559, 2007.
- [4] S. Rozen, M. E. Cudkowicz, M. Bogdanov et al., "Metabolomic analysis and signatures in motor neuron disease," *Metabolomics*, vol. 1, no. 2, pp. 101–108, 2005.
- [5] L. A. Paige, M. W. Mitchell, K. R. P. Krishnan, R. Kaddurah-Daouk, and D. C. Steffens, "A preliminary metabolomic analysis of older adults with and without depression," *International Journal of Geriatric Psychiatry*, vol. 22, no. 5, pp. 418–423, 2007.
- [6] R. Kaddurah-Daouk, B. S. Kristal, and R. M. Weinshilboum, "Metabolomics: a global biochemical approach to drug response and disease," *Annual Review of Pharmacology and Toxicology*, vol. 48, pp. 653–683, 2008.
- [7] E. Holmes, T. M. Tsang, J. T. J. Huang et al., "Metabolic profiling of CSF: evidence that early intervention may impact on disease progression and outcome in schizophrenia," *PLoS Medicine*, vol. 3, no. 8, article e327, 2006.
- [8] R. Kaddurah-Daouk, "Metabolic profiling of patients with schizophrenia," *PLoS Medicine*, vol. 3, no. 8, article e363, 2006.
- [9] D. Morvan and A. Demidem, "Metabolomics by proton nuclear magnetic resonance spectroscopy of the response to chloroethylnitrosourea reveals drug efficacy and tumor adaptive metabolic pathways," *Cancer Research*, vol. 67, no. 5, pp. 2150–2159, 2007.
- [10] J. Yang, G. Xu, Y. Zheng et al., "Diagnosis of liver cancer using HPLC-based metabolomics avoiding false-positive result from hepatitis and hepatocirrhosis diseases," *Journal of Chromatography B*, vol. 813, no. 1–2, pp. 59–65, 2004.
- [11] X. Fan, J. Bai, and P. Shen, "Diagnosis of breast cancer using HPLC metabolomics fingerprints coupled with computational methods," in *Proceedings of the 27th Annual International Conference of the Engineering in Medicine and Biology Society (EMBS '05)*, pp. 6081–6084, Shanghai, China, September 2005.
- [12] M. Van Doorn, J. Vogels, A. Tas et al., "Evaluation of metabolite profiles as biomarkers for the pharmacological effects of thiazolidinediones in type 2 diabetes mellitus patients and healthy volunteers," *British Journal of Clinical Pharmacology*, vol. 63, no. 5, pp. 562–574, 2007.
- [13] T. A. Clayton, J. C. Lindon, O. Cloarec et al., "Pharmacometabonomic phenotyping and personalized drug treatment," *Nature*, vol. 440, no. 7087, pp. 1073–1077, 2006.
- [14] R. Madsen, T. Lundstedt, and J. Trygg, "Chemometrics in metabolomics—a review in human disease diagnosis," *Analytica Chimica Acta*, vol. 659, no. 1–2, pp. 23–33, 2010.
- [15] R. Kramer, *Chemometric Techniques for Quantitative Analysis*, Marcel Dekker, New York, NY, USA, 1998.
- [16] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological Cybernetics*, vol. 43, no. 1, pp. 59–69, 1982.
- [17] D. Meyer, F. Leisch, and K. Hornik, "The support vector machine under test," *Neurocomputing*, vol. 55, no. 1–2, pp. 169–186, 2003.
- [18] P. Müller and D. R. Insua, "Issues in Bayesian analysis of neural network models," *Neural Computation*, vol. 10, no. 3, pp. 749–770, 1998.
- [19] W. Guan, M. Zhou, C. Y. Hampton et al., "Ovarian cancer detection from metabolomic Liquid chromatography/mass spectrometry data by support vector machines," *BMC Bioinformatics*, vol. 10, article 259, 2009.
- [20] C. Henneges, D. Bullinger, R. Fux et al., "Prediction of breast cancer by profiling of urinary RNA metabolites using Support Vector Machine-based feature selection," *BMC Cancer*, vol. 9, article 104, 2009.
- [21] A. K. Arakaki, R. Mezencev, N. J. Bowen, Y. Huang, J. F. McDonald, and J. Skolnick, "Identification of metabolites with anticancer properties by computational metabolomics," *Molecular Cancer*, vol. 7, article 57, 2008.
- [22] A. D. Southam, J. M. Easton, G. D. Stentiford, C. Ludwig, T. N. Arvanitis, and M. R. Viant, "Metabolic changes in flatfish hepatic tumours revealed by NMR-based metabolomics and metabolic correlation networks," *Journal of Proteome Research*, vol. 7, no. 12, pp. 5277–5285, 2008.
- [23] E. Altmaier, S. L. Ramsay, A. Graber, H.-W. Mewes, K. M. Weinberger, and K. Suhre, "Bioinformatics analysis of targeted metabolomics—uncovering old and new tales of diabetic mice under medication," *Endocrinology*, vol. 149, no. 7, pp. 3478–3489, 2008.
- [24] C. Wang, H. Kong, Y. Guan et al., "Plasma phospholipid metabolic profiling and biomarkers of type 2 diabetes mellitus based on high-performance Liquid chromatography/electrospray mass spectrometry and multivariate statistical analysis," *Analytical Chemistry*, vol. 77, no. 13, pp. 4108–4116, 2005.
- [25] V.-P. Mäkinen, P. Soininen, C. Forsblom et al., "¹H NMR metabolomics approach to the disease continuum of diabetic complications and premature death," *Molecular Systems Biology*, vol. 4, article 167, 2008.
- [26] R. M. Salek, M. L. Maguire, E. Bentley et al., "A metabolomic comparison of urinary changes in type 2 diabetes in mouse, rat, and human," *Physiological Genomics*, vol. 29, no. 2, pp. 99–108, 2007.
- [27] S. C. Connor, M. K. Hansen, A. Corner, R. F. Smith, and T. E. Ryan, "Integration of metabolomics and transcriptomics data to aid biomarker discovery in type 2 diabetes," *Molecular Biosystems*, vol. 6, pp. 909–921, 2010.
- [28] I. R. Lanza, S. Zhang, L. E. Ward, H. Karakelides, D. Raftery, and K. S. Nair, "Quantitative metabolomics by H-NMR and LC-MS/MS confirms altered metabolic pathways in diabetes," *PLoS One*, vol. 5, e10538 pages, 2010.

- [29] Y. Bao, T. Zhao, X. Wang et al., "Metabonomic variations in the drug-treated type 2 diabetes mellitus patients and healthy volunteers," *Journal of Proteome Research*, vol. 8, pp. 1623–1630, 2009.
- [30] M. Čuperlović-Culf, N. Belacel, A. S. Culf et al., "NMRmetabolic analysis of samples using fuzzy K-means clustering," *Magnetic Resonance in Chemistry*, vol. 47, supplement 1, pp. 96–104, 2009.
- [31] L. C. Kenny, W. B. Dunn, D. I. Ellis, J. Myers, P. N. Baker, and D. B. Kell, "Novel biomarkers for pre-eclampsia detected using metabolomics and machine learning," *Metabolomics*, vol. 1, no. 3, pp. 227–234, 2005.
- [32] A. M. Weljie, R. Dowlatabadi, B. J. Miller, H. J. Vogel, and F. R. Jirik, "An inflammatory arthritis-associated metabolite biomarker pattern revealed by ¹H NMR spectroscopy," *Journal of Proteome Research*, vol. 6, no. 9, pp. 3456–3464, 2007.
- [33] M.-Q. Zhang and B. Wilkinson, "Drug discovery beyond the 'rule-of-five'," *Current Opinion in Biotechnology*, vol. 18, no. 6, pp. 478–488, 2007.
- [34] P. Willett, "Similarity-based virtual screening using 2D fingerprints," *Drug Discovery Today*, vol. 11, no. 23-24, pp. 1046–1053, 2006.
- [35] C. A. Lipinski, "Drug-like properties and the causes of poor solubility and poor permeability," *Journal of Pharmacological and Toxicological Methods*, vol. 44, no. 1, pp. 235–249, 2000.
- [36] I. Nobeli, H. Pongstingl, E. B. Krissinel, and J. M. Thornton, "A structure-based anatomy of the *E. coli* metabolome," *Journal of Molecular Biology*, vol. 334, no. 4, pp. 697–719, 2003.
- [37] S. Gupta and J. Aires-de-Sousa, "Comparing the chemical spaces of metabolites and available chemicals: models of metabolite-likeness," *Molecular Diversity*, vol. 11, no. 1, pp. 23–36, 2007.
- [38] D.-X. Kong, W. Ren, W. Lu, and H.-Y. Zhang, "Do biologically relevant compounds have more chance to be drugs?" *Journal of Chemical Information and Modeling*, vol. 49, no. 10, pp. 2376–2381, 2009.
- [39] G. V. Scagliotti and G. Selvaggi, "Antimetabolites and cancer: emerging data with a focus on antifolates," *Expert Opinion on Therapeutic Patents*, vol. 16, no. 2, pp. 189–200, 2006.
- [40] A. Meerbach, C. Meier, A. Sauerbrei, H.-M. Meckel, and P. Wutzler, "Antiviral activity of cyclosigenyl prodrugs of the nucleoside analogue bromovinyldeoxyuridine against herpes viruses," *International Journal of Antimicrobial Agents*, vol. 27, no. 5, pp. 423–430, 2006.
- [41] T. I. Oprea, A. Tropsha, J.-L. Faulon, and M. D. Rintoul, "Systems chemical biology," *Nature Chemical Biology*, vol. 3, no. 8, pp. 447–450, 2007.
- [42] S. I. Berger and R. Iyengar, "Network analyses in systems pharmacology," *Bioinformatics*, vol. 25, no. 19, pp. 2466–2472, 2009.
- [43] E. E. Schadt, S. H. Friend, and D. A. Shaywitz, "A network view of disease and compound screening," *Nature Reviews Drug Discovery*, vol. 8, no. 4, pp. 286–295, 2009.
- [44] J. Zhao, P. Jiang, and W. Zhang, "Molecular networks for the study of TCM pharmacology," *Briefings in Bioinformatics*, vol. 11, no. 4, pp. 417–430, 2010.
- [45] J. C. Adams, M. J. Keiser, L. Basuino et al., "A mapping of drug space from the viewpoint of small molecule metabolism," *PLoS Computational Biology*, vol. 5, no. 8, Article ID e1000474, 2009.
- [46] A. Macchiarulo, J. M. Thornton, and I. Nobeli, "Mapping human metabolic pathways in the small molecule chemical space," *Journal of Chemical Information and Modeling*, vol. 49, no. 10, pp. 2272–2289, 2009.
- [47] D. P. Briskin, "Medicinal plants and phytomedicines. Linking plant biochemistry and physiology to human health," *Plant Physiology*, vol. 124, no. 2, pp. 507–514, 2000.
- [48] Chemical Computing Group, *Molecular Operating Environment*, Montreal, Canada, 2008.
- [49] D. S. Wishart, C. Knox, A. C. Guo et al., "HMDB: a knowledgebase for the human metabolome," *Nucleic Acids Research*, vol. 37, no. 1, pp. 603–610, 2009.
- [50] T. Tanimoto, "IBM internal report," 1957.
- [51] P. Shannon, A. Markiel, O. Ozier et al., "Cytoscape: a software Environment for integrated models of biomolecular interaction networks," *Genome Research*, vol. 13, no. 11, pp. 2498–2504, 2003.
- [52] R. Zadernowski, M. Naczka, and J. Nesterowicz, "Phenolic acid profiles in some small berries," *Journal of Agricultural and Food Chemistry*, vol. 53, no. 6, pp. 2118–2124, 2005.
- [53] R. Kranich, A. S. Busemann, D. Bock et al., "Rational design of novel, potent small molecule pan-selectin antagonists," *Journal of Medicinal Chemistry*, vol. 50, no. 6, pp. 1101–1115, 2007.
- [54] W. R. Russell, L. Scobbie, G. G. Duthie, and A. Chesson, "Inhibition of 15-lipoxygenase-catalysed oxygenation of arachidonic acid by substituted benzoic acids," *Bioorganic and Medicinal Chemistry*, vol. 16, no. 8, pp. 4589–4593, 2008.
- [55] J. P. Ruddick, A. K. Evans, D. J. Nutt, S. L. Lightman, G. A. W. Rook, and C. A. Lowry, "Tryptophan metabolism in the central nervous system: medical implications," *Expert Reviews in Molecular Medicine*, vol. 8, no. 20, pp. 1–27, 2006.
- [56] N. Stoy, G. M. Mackay, C. M. Forrest et al., "Tryptophan metabolism and oxidative stress in patients with Huntington's disease," *Journal of Neurochemistry*, vol. 93, no. 3, pp. 611–623, 2005.
- [57] A. L. Zignego, A. Cozzi, R. Carpenedo et al., "HCV patients, psychopathology and tryptophan metabolism: analysis of the effects of pegylated interferon plus ribavirin treatment," *Digestive and Liver Disease*, vol. 39, no. 1, pp. 107–111, 2007.
- [58] M. I. Torres, M. A. Lopez-Casado, P. Lorite, and A. Rios, "Tryptophan metabolism and indoleamine 2,3-dioxygenase expression in coeliac disease," *Clinical and Experimental Immunology*, vol. 148, no. 3, pp. 419–424, 2007.
- [59] K. C. Meyer, R. A. Arend, M. V. Kalayoglu, N. S. Rosenthal, G. I. Byrne, and R. R. Brown, "Tryptophan metabolism in chronic inflammatory lung disease," *Journal of Laboratory and Clinical Medicine*, vol. 126, no. 6, pp. 530–540, 1995.
- [60] K. Schrocksnadel, B. Wirleitner, C. Winkler, and D. Fuchs, "Monitoring tryptophan metabolism in chronic immune activation," *Clinica Chimica Acta*, vol. 364, no. 1-2, pp. 82–90, 2006.
- [61] J. D. Kopple, "Phenylalanine and tyrosine metabolism in chronic kidney failure," *Journal of Nutrition*, vol. 137, no. 6, pp. 1586–1598, 2007.
- [62] G. D'Andrea, R. Ostuzzi, A. Bolner et al., "Study of tyrosine metabolism in eating disorders. Possible correlation with migraine," *Journal of the Neurological Sciences*, vol. 29, no. 1, pp. 88–92, 2008.
- [63] R. L. Dillman and J. H. Cardellina II, "Aromatic secondary metabolites from the sponge *Tedania ignis*," *Journal of Natural Products*, vol. 54, no. 4, pp. 1056–1061, 1991.

- [64] L. Garlaschelli, Z. Pang, O. Sterner, and G. Vidari, "New indole derivatives from the fruit bodies of *Tricholoma sciodes* and *T. virgatum*," *Tetrahedron*, vol. 50, no. 11, pp. 3571–3574, 1994.
- [65] B. K. Chowdhury, A. Mustapha, M. Garba, and P. Bhattacharyya, "Carbazole and 3-methylcarbazole from *Glycosmis pentaphylla*," *Phytochemistry*, vol. 26, no. 7, pp. 2138–2139, 1987.

Review Article

Bioactive Food Components and Cancer-Specific Metabonomic Profiles

Young S. Kim and John A. Milner

Nutritional Science Research Group, Division of Cancer Prevention, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA

Correspondence should be addressed to Young S. Kim, yk47s@nih.gov

Received 16 June 2010; Revised 29 September 2010; Accepted 5 October 2010

Academic Editor: Olav Kvalheim

Copyright © 2011 Y. S. Kim and J. A. Milner. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Cancer cells possess unique metabolic signatures compared to normal cells, including shifts in aerobic glycolysis, glutaminolysis, and *de novo* biosynthesis of macromolecules. Targeting these changes with agents (drugs and dietary components) has been employed as strategies to reduce the complications associated with tumorigenesis. This paper highlights the ability of several food components to suppress tumor-specific metabolic pathways, including increased expression of glucose transporters, oncogenic tyrosine kinase, tumor-specific M2-type pyruvate kinase, and fatty acid synthase, and the detection of such effects using various metabonomic technologies, including liquid chromatography/mass spectrometry (LC/MS) and stable isotope-labeled MS. Stable isotope-mediated tracing technologies offer exciting opportunities for defining specific target(s) for food components. Exposures, especially during the early transition phase from normal to cancer, are critical for the translation of knowledge about food components into effective prevention strategies. Although appropriate dietary exposures needed to alter cellular metabolism remain inconsistent and/or ill-defined, validated metabonomic biomarkers for dietary components hold promise for establishing effective strategies for cancer prevention.

1. Introduction

Cancer cells exhibit unique metabolic signatures that are required for their aberrant proliferation [1]. Thus, monitoring changes in small-molecular-weight compounds (metabonomics) may represent an approach for detecting subtle shifts in tumor cell behavior [2]. Consequently, metabonomics holds promise as an effective tool for diagnosing disease progression, for identifying potential molecular targets, and for developing preventive and therapeutic agents (drugs and bioactive food components) [3, 4].

The two most effective approaches for metabonomic measures include nuclear magnetic resonance (NMR) and mass spectrometry (MS). Urine [5, 6], saliva [7], and blood plasma [8] have all been used successfully to monitor metabonomic patterns prior to and following intervention. Currently, a wealth of information about metabonomics is accumulating, thanks to highly sensitive Fourier transform ion cyclotron resonance MS (FT-ICR MS) [9, 10] and reliable

candidate technologies such as stable isotope monitoring [11].

Scientists have long recognized that the discovery of novel and noninvasive biomarkers is fundamental to designing effective clinical studies, forecasting disease risk, understanding how molecular pathways are interconnected within cells and organisms, and ultimately predicting health. For example, a change in serum lactate concentration, as determined by ¹H-NMR spectroscopy predicts weight gain during chemotherapy and ultimately breast cancer recurrence risk [12]. Likewise, the elevated urinary prostaglandin E(2) metabolite (11- α -hydroxy-9, 15-dioxo-2,3,4,5-tetranorpropane-1, 20-dioic acid) has been reported to predict a poor prognosis in head and neck squamous cell carcinoma patients and thus the need for more aggressive intervention approaches [13]. Unfortunately, the relationship between a given set of biomarkers and a particular pathological condition is often complicated by adjustments and adaptations, making the predictive value of such measures difficult to

establish. Metabonomic profiles of blood, urine, or tissues reflecting *in situ* status of metabolites may provide a means of identifying such biomarkers.

The annotation of metabolites identified by NMR- and MS-based experiments remains a challenge due to the diversity of study conditions. Thus, it is important to develop a standardized format that allows the scientific community to understand results from studies with widely different conditions. To facilitate this process, a working group of the Metabolomics Standards Initiative was formed in 2004 [14]. This group has established guidelines for the exchange and/or report of metabonomics data. Currently available metabonomic databases, including NIST08, PubChem, and KEGG, also help identify metabolite spectral data and structural information. An extensive review of metabonomic databases is found in a published paper [15]. Well-annotated and user-friendly databases should be able to assist in resolving unknown dietary metabolites that play a role in cancer initiation and development.

Considerable evidence points to diet as a variable that can influence cancer risk and/or tumor behavior [16]. This dietary effect may arise from its interactions with key regulatory molecules in various cancer processes, including carcinogen metabolism, hormonal balance, cell signaling, cell cycle control, apoptosis, and differentiation. The goal of this paper is to provide some insights into the metabonomic assessment of dietary effects on cancer. This is critical to evaluate whether a single or multiple dietary component(s) modulate the early transition phase from normal to cancer phenotypes, including changes in specific enzymatic activities and carbon flows.

2. Metabonomic Shifts Caused by Bioactive Food Components in Cancer Cells

Metabolic profiling of cancer cells represents global phenotypic changes, including those in glucose metabolism, amino acid metabolism, and fatty acid metabolism. This section discusses diet-induced alterations in basic metabolism that are associated with changes in cancer cell behavior and the monitoring of those changes using global metabonomic profiles.

2.1. Glucose Metabolism. A distinct difference between normal and cancer cells lies in the way these cells utilize glucose for their survival. The energy that sustains cancer cells is preferentially derived from aerobic glycolysis, which produces significantly higher amounts of lactate from pyruvate, whereas normal cells use the aerobic tricarboxylic acid (TCA) cycle, which oxidizes pyruvate to CO₂ and water. Furthermore, proliferating cancer cells also utilize glucose for the synthesis of nucleotides through the nonoxidative pentose phosphate pathway (PPP) and *de novo* fatty acid synthesis mediated by fatty acid synthase (FAS) [1]. These differences in metabolism between normal and cancer cells are particularly interesting because they are detectable by non-invasive profiling in blood or urine and thereby can serve as clinically useful biomarkers for predicting diet-induced shifts in the metabolic switch. Potential metabonomic targets for bioactive food components are discussed

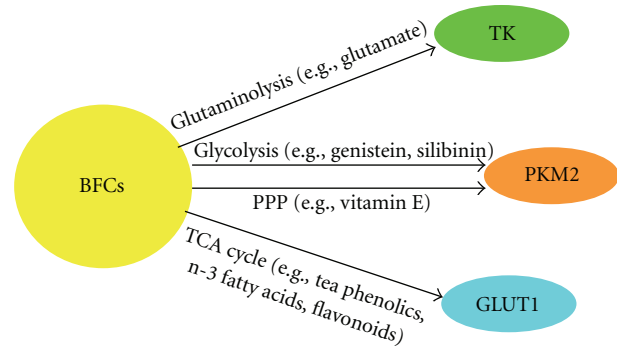


FIGURE 1: Potential metabonomic targets for bioactive food components during glucose metabolism in cancer cells. Cancer cells metabolize glucose and glutamine more than normal cells to support the *de novo* biosynthesis of nucleotides and energy required for the high rate of cell proliferation. Tumor-specific PKM2 determines the ratio of glucose metabolism between glycolysis and PPP, whereas GLUT1 and TK play critical roles in energy production in neoplastic cells via the TCA cycle and glutaminolysis, respectively. Each of these pathways is modulated by specific bioactive food components (see text). BFCs: bioactive food components; PKM2: pyruvate kinase M2 isoform; PPP: pentose phosphate pathway; TK: tyrosine kinase; GLUT1: glucose transporter 1; TCA: tricarboxylic acid.

in detail in the following sections and are summarized in Figure 1.

2.1.1. Increased Glucose Uptake. Increased glucose uptake is recognized as a hallmark for cancer cell malignancy and serves as a basis for using ¹⁸F-deoxyglucose positron emission tomography for cancer detection. Recently, a *KRAS* mutation with aberrant cell proliferation was found to relate to the upregulation of the glucose transporter (GLUT1) gene, suggesting that genetic mutations acquired in tumor cells account for enhanced glucose uptake [17]. This change is critical for cell proliferation [18], and thus GLUT1 expression may represent a promising target for cancer prevention. A number of dietary components, including green tea polyphenolics, cinnamon polyphenol extracts, fish n-3 fatty acids, and vegetable flavonoids such as myricetin and quercetin, have been documented as modifiers of glucose uptake by altering GLUT1 expression [19–22]. Impressively, these effects occur in several cell lines, suggesting widespread utility of their benefits. The concentrations used, at least for some, appear physiologically relevant and therefore may occur after a simple dietary change. Although these findings provide promise for the diet-mediated prevention of cancer cell proliferation, unfortunately none of these dietary effects was examined with high-throughput metabonomic profiles. Considering the ability of metabonomics to detect changes in glucose metabolism in cancer cells [23], it is likely that these dietary effects on glucose uptake, including GLUT1 expression, can be detected using the same approach.

2.1.2. Aerobic Glycolysis. The conversion of glucose to lactate in the presence of oxygen is a critical aerobic pathway that allows cancer cells to proliferate rapidly. *In vivo* expression

of the M2 isoform of the glycolytic enzyme pyruvate kinase (PK) is required for this process [24]. PK has four isozymes that are tissue specific: the L and R isoforms are expressed in liver and red blood cells, respectively the M1 in most adult tissues and M2 in lung and embryonic cells. In normal proliferating cells, PK exists in a tetrameric form that transfers one phosphate group from phosphoenolpyruvate (PEP) to adenosyl diphosphate, yielding one molecule of adenosine triphosphate (ATP). In cancer cells, however, this enzyme is exclusively expressed as an M2 type of pyruvate kinase (PKM2), which results in a conformational change from a tetrameric to a dimeric form that possesses a low affinity for PEP and thus limits enzymatic function. This structural change was shown to be mediated through phosphotyrosine peptides or *src* in cancer cells [25], suggesting that oncogenic tyrosine kinases (TKs) might be promising targets for cancer prevention. Support for this conclusion comes from observations that PKM2 is released from tumors into the blood and can be monitored non-invasively. Plasma tumor PKM2 is reported to increase significantly as tumor progresses in lung cancer, qualifying it as a valuable biomarker [26]. Recently, delphinidin, an anthocyanidin that is abundant in grapes, cranberries, and pomegranates, has been reported to inhibit TK activities, including receptor TKs [27]. Additionally, isothiocyanates in broccoli have been reported to inhibit TK [28]. Overall, these findings suggest that some of the antiproliferative effects of bioactive food components may arise from their ability to convert tetrameric PKM2 to dimeric PKM2 and thereby bring about a shift in glucose flux from a nonoxidative to an oxidative pathway. Although some evidence exists that changes in glucose carbon flux from glycolysis to gluconeogenesis can be measured by the combination of LC/MS metabolomic profiling with stable isotope techniques using ^{13}C -labeled glucose [29], these approaches have not been used for the study examining dietary effects. Thus, it is warranted to use metabolomic profiling to monitor the effects of diet or dietary components on PKM2-mediated shifts in glucose flux.

Although limited numbers of metabolomic studies have examined the cancer preventive effects of dietary components, the efficacy of the dietary supplement silibinin provides some proof of principle. Silibinin, extracted from a milk thistle, has been reported to significantly influence the metabolism of prostate cancer cells [30]. Quantitative high-resolution proton NMR spectroscopy was used to monitor the metabolomic profile of blood and tissue extracts in non-human animals receiving or not receiving silibinin. The metabolomic profile suggested that feeding TRAMP/C57BL/6 mice with a 1% silibinin-supplemented diet reduced lactate formation, increased glucose oxidation through the TCA cycle, reversed the increase in citrate use, and decreased cholesterol and phosphatidylcholine levels in prostatic tumors, which parallels earlier findings regarding the prostate cancer preventive potential of silibinin in this animal model. These results suggest that non-invasive metabolomic studies can help monitor the effectiveness of cancer preventives against malignancy.

2.1.3. Increased De Novo Biosynthesis of Nucleotides via the Pentose Phosphate Pathway. Since PKM2 in cancer cells does not convert PEP to pyruvate efficiently, the upstream glucose metabolites are redirected towards a biosynthetic pathway such as PPP. Any rapidly proliferating cells, including cancer cells, utilize PPP to generate ribose-5-phosphate and the reduced form of nicotinamide adenine dinucleotide phosphate (NADPH) from glucose, which is essential for *de novo* DNA formation. Evidence exists that vitamin E, which is abundant in wheat germ extract and various other fat sources, inhibits nonoxidative tumor cell PPP and thereby suppresses cancer cell proliferation [31]. In leukemia patients, vitamin E was found to decrease pentose cycle substrate flow into RNA ribose, which was measured using stable isotope-labeled [1,2- ^{13}C] glucose as the single tracer and biological MS. These changes in glucose carbon flux correlated with the cell growth-controlling and apoptosis-inducing effects of fermented wheat germ [31].

Another dietary component, genistein, which is found in soy, has also been shown to exhibit a similar effect on PPP without modulating fatty acid synthesis [32]. When MIA pancreatic cancer cells were treated with a physiological concentration of genistein (20 μM) for 3 days, the carbon flux from ^{13}C -labeled glucose to nucleic acid ribose through the nonoxidative PPP decreased significantly [32]. These changes in PPP coincided with the cancer protective effects of either vitamin E or genistein, implying that a shift in a metabolic signature may be a useful biomarker for monitoring the efficacy of dietary components.

2.2. Amino Acid and Fatty Acid Metabolism

2.2.1. Glutaminolysis. Glucose is the major energy source in normal cells. This nutrient undergoes the metabolic switch in tumor cells from energy production to *de novo* biosynthesis of macromolecules, including nucleic acids. This shift increases the demand for carbon sources in cancer cells to generate ATP. It has been reported that cancer cells accumulate glutamine faster than noncancerous cells [33] and utilize it as a substrate for the TCA cycle (glutaminolysis). Ultimately, the cancer cell catabolizes glutamine to ATP and lactate. This metabolic shift is thought to be mediated by the *myc* oncogene [34]. It is well established that the activity of this oncogene can be modulated by various dietary components, including betaine, folic acid, genistein, and fat [35–38]. For example, supplementing the diet with betaine (1%) fed to rats developing liver tumors caused by diethylnitrosamine significantly reduced liver *c-myc* expression [35]. ^1H -NMR spectroscopy-based metabolic profiles of plasma and urine have revealed an interaction between betaine exposure and low-density lipoprotein receptor gene expression [39]. Metabolomic profiles that explore the influence of dietary fat metabolism regulators to detect variation in *c-myc* oncogene expression are warranted.

2.2.2. Fatty Acid Synthase. Treatment of mouse mammary epithelial cells (HC11) in culture with 40 nM of genistein for 6 hours has been reported to significantly suppress

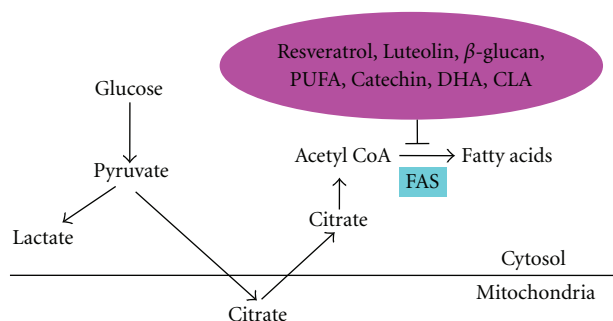


FIGURE 2: Fatty acid synthase as a metabolomic target for bioactive food components in cancer cells. Tumor cells exhibit the increased activity of FAS, which converts citrate-originated acetyl-CoA to fatty acids, mainly palmitate. The citrate is generated in mitochondria and, instead of further oxidation, is exported to cytosol as citrate. Upon exit, citrate forms acetyl CoA and is converted to fatty acids in cytosol, which is suppressed by a variety of dietary components, including tea catechin, DHA in fish oil, β -glucan in barley and mushrooms, resveratrol in red grapes, the vegetable flavonoid luteolin, and CLA in milk. FAS: fatty acid synthase; CoA: coenzyme A; PUFA: polyunsaturated fatty acid; DHA: docosahexaenoic acid; CLA: conjugated linoleic acid.

Wnt-1-induced, but not basal, expression of *c-Myc* [37]. It is logical that several dietary components may affect *myc* gene expression *in vivo* to regulate glutaminolysis, and this issue should be clarified using metabolomics. In tumor cells, the limited affinity of tumor-specific PKM2 for PEP forces some, but not all, pyruvates to be oxidized to citrate in the mitochondria. This citrate is then exported to the cytosol, where it is used as a substrate for fatty acid synthesis (Figure 2). This shift from oxidation to lipogenesis in citrate metabolism is evident by the increased activity of FAS, which converts glucose-originated malonyl-CoA to long-chain fatty acids, mainly palmitate, in tumor cells. The synthesized fatty acids are incorporated into phospholipids, the accumulation of which may cause metabolic diseases, including cancer. For example, the *de novo* biosynthesis of fatty acids from glucose as determined by proton NMR spectroscopy ($^1\text{H-MRS}$) is frequently observed with the lipid droplet formation in brain tumor glioma cells [40]. The possibility that these synthesized fatty acids may modulate the oncogenic PI3K/AKT pathway and thereby influence glucose uptake and metabolism cannot be ignored.

A variety of dietary components, including tea catechin, docosahexaenoic acid (DHA) in fish oil, β -glucan in barley and mushrooms, resveratrol in red grapes, the vegetable flavonoid luteolin, and conjugated linoleic acid in milk, have been shown to suppress FAS expression and activity [41–46]. Treatment of the human breast cancer cell SK Br-3, which constitutively expresses FAS, with $150\ \mu\text{M}$ epigallocatechin-3-gallate (EGCG) for 24 hours resulted in a marked reduction in this enzyme activity ($59\% \pm 13\%$), but not its expression, compared to controls [41]. In the same study, the suppressive effect of EGCG on FAS activity was confirmed *in vivo* using C57BL/6J male mice. Likewise, the oral feeding of C57BL/KsJ-*db/db* mice with DHA (1 g DHA/100 g

diet) for 4 weeks significantly reduced hepatic FAS activity accompanied by lowered levels of triglyceride [42]. Finally, feeding C57BL mice a high-fat diet supplemented with 4% barley containing a high amount of β -glucan for 12 weeks significantly reduced the hepatic FAS activity [43]. It should be possible to detect diet-induced changes in FAS with metabolomic profiling, which allows the monitoring of changes in physiologic enzyme activities using LC/MS [47].

The responsiveness of FAS to dietary components, the differential expression of this enzyme in various stages of cancer development, and its critical role in *de novo* lipid synthesis made it an ideal target for the dietary prevention of cancer. It has been shown that altered FAS expression brings about the signature metabolic changes in human prostate cancer. Human prostatic epithelial cells are highly unique in accumulating zinc that blocks citrate degradation. In cancer cells, however, the prostate tissue contains low levels of citrate, because it does not accumulate zinc and because most of the citrate is used for fatty acid synthesis. $^1\text{H-MRS}$ of prostate tissues confirmed the dramatic decrease in citrate levels in prostate gland during malignancy [30]. Likewise, silibinin in milk thistle was reported to significantly decrease the expression of FAS and its transcription factor, sterol response element-binding protein 1c, as examined with $^1\text{H-MRS}$ -based metabolomic approaches in neoplastic prostate cells [30]. The metabolic profile of prostatic tumors obtained from TRAMP mice fed a silibinin diet for 20 weeks revealed decreases in lactate and increases in citrate. Thus, a shift in the metabolic profile may help explain silibinin's antitumorigenic effects. Fatty acid synthesis is also known to be influenced by several confounding factors, including peroxisome proliferator-activated receptors and choline availability [48]. Metabolomic profiles of tissue extracts may provide important clues about the fatty acid homeostasis that is likely to be influenced by multiple variables, including dietary habits.

3. Stable Isotope Technologies

The activity of bioactive food components in the body is dynamic, and thus a single point in time measurement may not lead to appropriate conclusions. Fortunately, these subtle shifts in metabolites can be traced using labels, including stable isotopes. Stable isotopes such as ^{13}C or ^{15}N that generate mass difference and thus can be detected by NMR or imaging technologies are widely used to trace dynamic movements, including nutrient flux, bioavailability, or kinetics *in vivo* [49]. Other major advantages of stable isotopes include their stability and independency, which allows long-term studies as well as the simultaneous labeling of more than one molecule on the same compound.

Although stable isotope technologies combined with NMR, especially 2D ^1H total correlation spectroscopy, can be used for correlating a single isotopomer to its parental compound, the resolving power and sensitivity of this analytical technique are much less than those of MS. Stable isotope labeling, combined with various types of MS equipped with versatile ion sources such as FT-ICR, electrospray, and matrix-assisted laser desorption/ionization,

TABLE 1: Preclinical and clinical nutrition studies using stable-isotope metabolomics.

| Specimen | Analysis | Stable isotope | Results | Reference |
|---------------|------------|-----------------|--|-------------------|
| Blood, urine | NMR, GC-MS | ^{13}C | Differential glucose metabolic pathways between normal and cancer cells in lung: ^{13}C glucose was infused to lung cancer patients and showed enhanced production of Asp and Glu via glycolysis in lung tumor tissues. | Fan et al. [50] |
| Cell extracts | GC/MS | ^{13}C | The treatment of MIA pancreatic adenocarcinoma cells with 200 μM genistein for 3 days reduced the glucose (labeled with [1,2- ^{13}C] glucose) metabolism via the nonoxidative pentose pathway, which coincided with its antiproliferative effects. | Boros et al. [32] |
| Cell extracts | GC-MS | ^{13}C | The altered flux in response to gluconeogenic substrates in fasting rat hepatocytes was measured with [1,2- $^{13}\text{C}_2$] glucose. | Marin et al. [51] |
| Cell extracts | GC-MS | ^{13}C | The treatment of butyrate-sensitive HT29 human colon adenocarcinoma cells with 5 mM butyrate resulted in the inhibition of glucose uptake, oxidation, and nucleic acid ribose synthesis. | Boren et al. [52] |
| Plasma | GC/MS | ^{13}C | The hypoglycemia seen in the fasting PPAR α null mouse is due to the reduced recycling of glucose carbon from lactate back to glucose. | Xu et al. [53] |

TABLE 2: Metabonomic profiles of metabolites from selected bioactive food components in human blood and urine.

| Food source | Component | Metabonomic analysis | Blood | Urine | Reference |
|------------------------|---|-------------------------|--|---|---|
| Green tea | EGCG | HPLC-MS/MS | EGCG | Sulfated and glucuronidated conjugates of EGCG, hippuric acids | Del Rio et al. [58] Wang JS et al. [59]. |
| Soy | Genistein | LC-ESI-MS/MS | Monoglucuronides of genistein | Glucuronidated genistein | Guy et al. [60] Kano et al. [61]. |
| Cruciferous vegetables | Glucobrassicin | HPLC-MS/MS | DIM, LTr1, H1-1M | DIM | Reed et al. [62] Anderton et al. [63]. |
| Red grapes | Piceid | LC-MS/MS, NMR, HPLC/DAD | Sulfate and glucuronide conjugates of transresveratrol | Sulfate and glucuronide conjugates of trans-resveratrol | Burkon and Somoza [64]., Boocock et al. [65]. |
| Cruciferous vegetables | Glucoraphanin | HPLC-MS, LC-(ESI)MS/MS | Sulforaphane | Mercapto-conjugates of sulforaphane (N-acetyl cysteine conjugates are major metabolites.) | Egner et al. [66], Al Janobi et al. [67]. |
| Fish, Mushroom | Vitamin D ₂ , D ₃ | LC-MS/MS | 25-OH vitamin D | 24,25(OH) ₂ D ₃ , 1 α ,24,25(OH) ₂ D ₃ | van den Ouweland et al. [68]. |

can isolate and detect the molecule of interest—even when several compounds may be crowded at one location on the profile—and thus serves as a powerful tracing technique. Representative preclinical and clinical studies using stable isotope technologies involving nutrition and cancer are provided in Table 1 [50–53]. Several excellent reviews about the merits of stable isotopes have been recently published [11, 54].

4. Metabonomics Can Detect Cancer Preventive Dietary Metabolites Generated by Microorganisms

It seems obvious that the biological functions of small-molecular-weight compounds (metabolites) are modulated

by microbiota accounting for more than 100 trillion living organisms in the human gut. Gastrointestinal microbes are quick responders to dietary changes in terms of their metabolism and/or gene expression. Evidence exists that these changes may affect the utility of nutrients obtained from the diet and thereby influence human health and disease outcome [55]. The precise role of microbes in physiological metabolic pathways was investigated using a well-defined animal model transplanted with human fecal microbial communities [55]. When these animals were fed a high-fat and high-sugar diet (Harlan-Teklad TD96132), they developed increased adiposity, which can be characterized metabolically when switched to a low-fat and plant-polysaccharide-rich diet (B&K 7378000) [55]. These results suggest that microbes may either provide additional

metabolic pathways to the host or directly modify the existing processes, resulting in altered phenotypes. Specific examples of how microbes metabolize dietary components and release cancer preventive compounds are described in detail in an earlier review [56]. With the current advances in “omics” technologies, these changes caused by human gut microbes can be directly detected by metabolomic profiles [57]. Although the application of metabolomics to microbiome projects is in its infant stage, the effects of diet on the human gut microbiome and the utilization of metabolomic analysis for its detection hold promise for future investigation.

5. Metabonomics Involving Bioactive Food Components in Clinical Studies

The current status of metabonomics that has been used for identifying and characterizing the physiological metabolites of food ingredients in humans is summarized in Table 2 [58–68]. The analysis of active food metabolites usually has been conducted with compounds known to be efficacious to human diseases. The molecules that interact with cellular components of target tissues are not food constituents but rather their metabolites that are absorbed and transported to each site of action [69–71]. For example, EGCG, which occurs in green tea, is considered to be a tumor preventive agent. Nevertheless, the prostate tissue of men who had consumed 6 cups of green tea daily for 6 weeks contained both EGCG and its methylated metabolite 4'-O-methyl EGCG. The latter has been reported to be less active than EGCG in terms of its cancer preventive action [69]. On the other hand, urolithin, a microbially generated metabolite of an active component in pomegranate juice, ellagic acid, has been found to exhibit cancer protection in various tissues, including colon, breast, and prostate [70, 71]. Thus, findings from metabonomic studies should be able to assist with clarifying how food or dietary components are metabolized in the body, possibly modulate human cell behavior, and ultimately and possibly influence human health. Currently, limited numbers of investigations exist in this area, but the science is poised to provide important new insights.

6. Future Directions

Metabonomics is the systematic study of small-molecular-weight substances that are the final products of genes and proteins. Although extensive research about genomics and proteomics is beginning to unravel the role of genes and proteins, the metabolites that dictate their characteristics for a given phenotype remain largely ill-defined. Metabonomic profiling using readily available biofluids, tissue extracts, or intact tissues may provide early clues about cellular events that influence phenotypic characteristics. Recently, there has been increased interest in cancer metabonomics and its potential for use in clinical and/or epidemiological investigations. The enriched understanding of basic mechanisms that account for shifts in the metabolome should become reality for metabonomic investigations so that effective interventions can be employed.

Although nutrition is known to modulate a number of regulatory networks involved with pathways leading to and promoting cancer, there is a dearth of studies that have examined diet as an important variable. Admittedly, there are critical issues in understanding metabonomic profiling information from nutritional studies, including the quantity and the temporal relationship of the bioactive food component consumed. Regardless, the fundamental nature of nutrition in harnessing overall cellular metabolism deserves greater attention. Controlled human intervention studies that incorporate physiologically relevant concentrations and exposures for various times are fundamental to the understanding of the utility of metabonomics as a tool for predicting phenotypic change caused by diet. Although there will be many challenges to the interpretation of these studies, their societal implications are enormous given the central role that diet has in health promotion and disease prevention, especially that related to cancer.

References

- [1] X. Tong, F. Zhao, and C. B. Thompson, “The molecular determinants of de novo nucleotide biosynthesis in cancer cells,” *Current Opinion in Genetics and Development*, vol. 19, no. 1, pp. 32–37, 2009.
- [2] E. S. Ong, L. Zou, S. Li, P. Y. Cheah, K. W. Eu, and C. N. Ong, “Metabolic profiling in colorectal cancer reveals signature metabolic shifts during tumorigenesis,” *Molecular & Cellular Proteomics*. In press.
- [3] Y. S. Kim, P. Maruvada, and J. A. Milner, “Metabonomics in biomarker discovery: future uses for cancer prevention,” *Future Oncology*, vol. 4, no. 1, pp. 93–102, 2008.
- [4] A. M. Zivkovic and J. B. German, “Metabonomics for assessment of nutritional status,” *Current Opinion in Clinical Nutrition and Metabolic Care*, vol. 12, no. 5, pp. 501–507, 2009.
- [5] T. Kind, V. Tolstikov, O. Fiehn, and R. H. Weiss, “A comprehensive urinary metabolomic approach for identifying kidney cancer,” *Analytical Biochemistry*, vol. 363, no. 2, pp. 185–195, 2007.
- [6] R. M. Salek, M. L. Maguire, E. Bentley et al., “A metabolomic comparison of urinary changes in type 2 diabetes in mouse, rat, and human,” *Physiological Genomics*, vol. 29, no. 2, pp. 99–108, 2007.
- [7] M. C. Walsh, L. Brennan, J. P. G. Malthouse, H. M. Roche, and M. J. Gibney, “Effect of acute dietary standardization on the urinary, plasma, and salivary metabolomic profiles of healthy humans,” *American Journal of Clinical Nutrition*, vol. 84, no. 3, pp. 531–539, 2006.
- [8] K. O. Boernsen, S. Gatzek, and G. Imbert, “Controlled protein precipitation, in combination with chip-based nanospray infusion mass spectrometry. An approach for metabonomics profiling of plasma,” *Analytical Chemistry*, vol. 77, no. 22, pp. 7255–7264, 2005.
- [9] S. C. Brown, G. Kruppa, and J.-L. Dasseux, “Metabonomics applications of FT-ICR mass spectrometry,” *Mass Spectrometry Reviews*, vol. 24, no. 2, pp. 223–231, 2005.
- [10] J. Han, R. M. Danell, J. R. Patel et al., “Towards high-throughput metabolomics using ultrahigh-field Fourier transform ion cyclotron resonance mass spectrometry,” *Metabonomics*, vol. 4, no. 2, pp. 128–140, 2008.

- [11] A. N. Lane, T. W. Fan, and R. M. Higashi, "Stable isotope-assisted metabolomics in cancer research," *IUBMB life*, vol. 60, no. 2, pp. 124–129, 2008.
- [12] H. C. Keun, J. Sidhu, D. Pchejetski et al., "Serum molecular signatures of weight change during early breast cancer chemotherapy," *Clinical Cancer Research*, vol. 15, no. 21, pp. 6716–6723, 2009.
- [13] V. D. Kekatpure, J. O. Boyle, X. K. Zhou et al., "Elevated levels of urinary prostaglandin E metabolite indicate a poor prognosis in ever smoker head and neck squamous cell carcinoma patients," *Cancer Prevention Research*, vol. 2, no. 11, pp. 957–965, 2009.
- [14] S.-A. Sansone, T. Fan, R. Goodacre et al., "The metabolomics standards initiative," *Nature Biotechnology*, vol. 25, no. 8, pp. 846–848, 2007.
- [15] E. P. Go, "Database resources in metabolomics: an overview," *Journal of Neuroimmune Pharmacology*, vol. 5, pp. 18–30, 2010.
- [16] WCRF/AICR Report, *Food, Nutrition, Physical Activity, and the Prevention of Cancer: A Global Perspective*, 2009, <http://www.dietandcancerreport.org>.
- [17] J. Yun, C. Rago, I. Cheong et al., "Glucose deprivation contributes to the development of KRAS pathway mutations in tumor cells," *Science*, vol. 325, no. 5947, pp. 1555–1559, 2009.
- [18] T. Amann, U. Maegdefrau, A. Hartmann et al., "GLUT1 expression is increased in hepatocellular carcinoma and promotes tumorigenesis," *American Journal of Pathology*, vol. 174, no. 4, pp. 1544–1552, 2009.
- [19] H. Cao, I. Hininger-Favier, M. A. Kelly et al., "Green tea polyphenol extract regulates the expression of genes involved in glucose uptake and insulin signaling in rats fed a high fructose diet," *Journal of Agricultural and Food Chemistry*, vol. 55, no. 15, pp. 6372–6378, 2007.
- [20] H. Cao, J. F. Urban Jr., and R. A. Anderson, "Cinnamon polyphenol extract affects immune responses by regulating anti- and proinflammatory and glucose transporter gene expression in mouse macrophages," *Journal of Nutrition*, vol. 138, no. 5, pp. 833–840, 2008.
- [21] F. Pifferi, F. Roux, B. Langelier et al., "(n-3) polyunsaturated fatty acid deficiency reduces the expression of both isoforms of the brain glucose transporter GLUT1 in rats," *Journal of Nutrition*, vol. 135, no. 9, pp. 2241–2246, 2005.
- [22] P. Strobel, C. Allard, T. Perez-Acle, R. Calderon, R. Aldunate, and F. Leighton, "Myricetin, quercetin and catechin-gallate inhibit glucose uptake in isolated rat adipocytes," *Biochemical Journal*, vol. 386, part 3, pp. 471–478, 2005.
- [23] C. Denkert, J. Budczies, W. Weichert et al., "Metabolite profiling of human colon carcinoma—deregulation of TCA cycle and amino acid turnover," *Molecular Cancer*, vol. 7, article 72, 2008.
- [24] S. Mazurek, "Pyruvate kinase type M2: a key regulator of the metabolic budget system in tumor cells," *International Journal of Biochemistry and Cell Biology*. In press.
- [25] T. Hitosugi, S. Kang, M. G. Vander Heiden et al., "Tyrosine phosphorylation inhibits PKM2 to promote the Warburg effect and tumor growth," *Science Signaling*, vol. 2, no. 97, p. ra73, 2009.
- [26] J. Schneider, K. Neu, H.-G. Velcovsky, H. Morr, and E. Eigenbrodt, "Tumor M2-pyruvate kinase in the follow-up of inoperable lung cancer patients: a pilot study," *Cancer Letters*, vol. 193, no. 1, pp. 91–98, 2003.
- [27] N. Teller, W. Thiele, U. Boettler, J. Sleeman, and D. Marko, "Delphinidin inhibits a broad spectrum of receptor tyrosine kinases of the ErbB and VEGFR family," *Molecular Nutrition and Food Research*, vol. 53, no. 9, pp. 1075–1083, 2009.
- [28] S. Mukherjee, H. Gangopadhyay, and D. K. Das, "Broccoli: a unique vegetable that protects mammalian hearts through the redox cycling of the thioredoxin superfamily," *Journal of Agricultural and Food Chemistry*, vol. 56, no. 2, pp. 609–617, 2008.
- [29] J. Xu, G. Xiao, C. Tirujillo et al., "Peroxisome proliferator-activated receptor α (PPAR α) influences: substrate utilization for hepatic glucose production," *Journal of Biological Chemistry*, vol. 277, no. 52, pp. 50237–50244, 2002.
- [30] K. Raina, N. J. Serkova, and R. Agarwal, "Silibinin feeding alters the metabolic profile in TRAMP prostatic tumors: 1H-nmrs-based metabolomics study," *Cancer Research*, vol. 69, no. 9, pp. 3731–3735, 2009.
- [31] B. Comín-Anduix, L. G. Boros, S. Marin et al., "Fermented wheat germ extract inhibits glycolysis/pentose cycle enzymes and induces apoptosis through poly(ADP-ribose) polymerase activation in Jurkat T-cell leukemia tumor cells," *Journal of Biological Chemistry*, vol. 277, no. 48, pp. 46408–46414, 2002.
- [32] L. G. Boros, S. Bassilian, S. Lim, and W.-N. P. Lee, "Genistein inhibits nonoxidative ribose synthesis in MIA pancreatic adenocarcinoma cells: a new mechanism of controlling tumor growth," *Pancreas*, vol. 22, no. 1, pp. 1–7, 2001.
- [33] B. P. Bode, B. C. Fuchs, B. P. Hurley et al., "Molecular and functional analysis of glutamine uptake in human hepatoma and liver-derived cells," *American Journal of Physiology*, vol. 283, no. 5, pp. G1062–G1073, 2002.
- [34] D. R. Wise, R. J. DeBerardinis, A. Mancuso et al., "Myc regulates a transcriptional program that stimulates mitochondrial glutaminolysis and leads to glutamine addiction," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 48, pp. 18782–18787, 2008.
- [35] Y.-P. Du, J.-S. Peng, A. Sun, Z.-H. Tang, W.-H. Ling, and H.-L. Zhu, "Assessment of the effect of betaine on p16 and c-myc DNA methylation and mRNA expression in a chemical induced rat liver cancer model," *BMC Cancer*, vol. 9, article 261, 2009.
- [36] R. Lu, X. Wang, D. F. Sun et al., "Folic acid and sodium butyrate prevent tumorigenesis in a mouse model of colorectal cancer," *Epigenetics*, vol. 3, no. 6, pp. 330–335, 2008.
- [37] Y. Su and R. C. M. Simmen, "Soy isoflavone genistein upregulates epithelial adhesion molecule E-cadherin expression and attenuates β -catenin signaling in mammary epithelial cells," *Carcinogenesis*, vol. 30, no. 2, pp. 331–339, 2009.
- [38] N. Kobayashi, R. J. Barnard, J. Said et al., "Effect of low-fat diet on development of prostate cancer and Akt phosphorylation in the Hi-Myc transgenic mouse model," *Cancer Research*, vol. 68, no. 8, pp. 3066–3073, 2008.
- [39] K.-K. Cheng, G. M. Benson, D. C. Grimsditch, D. G. Reid, S. C. Connor, and J. L. Griffin, "Metabolomic study of the LDL receptor null mouse fed a high-fat diet reveals profound perturbations in choline metabolism that are shared with ApoE null mice," *Physiological Genomics*, vol. 41, no. 3, pp. 224–231, 2010.
- [40] J. L. Griffin and R. A. Kauppinen, "A metabolomics perspective of human brain tumours," *FEBS Journal*, vol. 274, no. 5, pp. 1132–1139, 2007.
- [41] T. Puig Miquel, J. Relat, P. F. Marrero, D. Haro, J. Brunet, and R. Colomer, "Green tea catechin inhibits fatty acid synthase without stimulating carnitine palmitoyltransferase-1 or inducing weight loss in experimental animals," *Anticancer Research*, vol. 28, no. 6, pp. 3671–3676, 2008.

- [42] N. Gotoh, K. Nagao, S. Onoda et al., "Effects of three different highly purified n-3 series highly unsaturated fatty acids on lipid metabolism in C57BL/KsJ-dbl db mice," *Journal of Agricultural and Food Chemistry*, vol. 57, no. 22, pp. 11047–11054, 2009.
- [43] J. S. Choi, H. Kim, M. H. Jung, S. Hong, and J. Song, "Consumption of barley β -glucan ameliorates fatty liver and insulin resistance in mice fed a high-fat diet," *Molecular Nutrition and Food Research*, vol. 54, no. 7, pp. 1004–1013, 2010.
- [44] G. V. Gnoni and G. Paglialonga, "Resveratrol inhibits fatty acid and triacylglycerol synthesis in rat hepatocytes," *European Journal of Clinical Investigation*, vol. 39, no. 3, pp. 211–218, 2009.
- [45] K. Brusselmans, R. Vrolix, G. Verhoeven, and J. V. Swinnen, "Induction of cancer cell apoptosis by flavonoids is associated with their ability to inhibit fatty acid synthase activity," *Journal of Biological Chemistry*, vol. 280, no. 7, pp. 5636–5645, 2005.
- [46] D. S. Y. Lau and M. C. Archer, "The 10t,12c isomer of conjugated linoleic acid inhibits fatty acid synthase expression and enzyme activity in human breast, colon, and prostate cancer cells," *Nutrition and Cancer*, vol. 62, no. 1, pp. 116–121, 2010.
- [47] L. P. S. de Carvalho, H. Zhao, C. E. Dickinson et al., "Activity-based metabolomic profiling of enzymatic function: identification of Rv1248c as a mycobacterial 2-hydroxy-3-oxoadipate synthase," *Chemistry and Biology*, vol. 17, no. 4, pp. 323–332, 2010.
- [48] L. D. Roberts, D. G. Hassall, D. A. Winegar, J. N. Haselden, A. W. Nicholls, and J. L. Griffin, "Increased hepatic oxidative metabolism distinguishes the action of peroxisome proliferator-activated receptor delta from Peroxisome proliferator-activated receptor gamma in the ob/ob mouse," *Genome Medicine*, vol. 1, p. 115, 2009.
- [49] J. Boren, W.-N. P. Lee, S. Bassilian et al., "The stable isotope-based dynamic metabolic profile of butyrate-induced HT29 cell differentiation," *Journal of Biological Chemistry*, vol. 278, no. 31, pp. 28395–28402, 2003.
- [50] T. W. M. Fan, A. N. Lane, R. M. Higashi et al., "Altered regulation of metabolic pathways in human lung cancer discerned by ^{13}C stable isotope-resolved metabolomics (SIRM)," *Molecular Cancer*, vol. 8, article 41, 2009.
- [51] S. Marin, W.-N. P. Lee, S. Bassilian et al., "Dynamic profiling of the glucose metabolic network in fasted rat hepatocytes using $[1,2-^{13}\text{C}_2]\text{glucose}$," *Biochemical Journal*, vol. 381, part 1, pp. 287–294, 2004.
- [52] J. Boren, W.-N. P. Lee, S. Bassilian et al., "The stable isotope-based dynamic metabolic profile of butyrate-induced HT29 cell differentiation," *Journal of Biological Chemistry*, vol. 278, no. 31, pp. 28395–28402, 2003.
- [53] J. Xu, G. Xiao, C. Tirujillo et al., "Peroxisome proliferator-activated receptor α (PPAR α) influences: substrate utilization for hepatic glucose production," *Journal of Biological Chemistry*, vol. 277, no. 52, pp. 50237–50244, 2002.
- [54] S. Ando and Y. Tanaka, "Mass spectrometric studies on brain metabolism, using stable isotopes," *Mass Spectrometry Reviews*, vol. 24, no. 6, pp. 865–886, 2005.
- [55] P. J. Turnbaugh, V. K. Ridaura, J. J. Faith, F. E. Rey, R. Knight, and J. I. Gordon, "The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice," *Science Translational Medicine*, vol. 1, no. 6, p. 6ra14, 2009.
- [56] C. D. Davis and J. A. Milner, "Gastrointestinal microflora, food components and colon cancer prevention," *Journal of Nutritional Biochemistry*, vol. 20, no. 10, pp. 743–752, 2009.
- [57] M. Li, B. Wang, M. Zhang et al., "Symbiotic gut microbes modulate human metabolic phenotypes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 6, pp. 2117–2122, 2008.
- [58] D. Del Rio, L. Calani, C. Cordero, S. Salvatore, N. Pellegrini, and F. Brighenti, "Bioavailability and catabolism of green tea flavan-3-ols in humans," *Nutrition*, vol. 26, no. 11-12, pp. 1110–1116, 2010.
- [59] J.-S. Wang, H. Luo, P. Wang et al., "Validation of green tea polyphenol biomarkers in a phase II human intervention trial," *Food and Chemical Toxicology*, vol. 46, no. 1, pp. 232–240, 2008.
- [60] L. Guy, N. Védrine, M. Urpi-Sarda et al., "Orally administered isoflavones are present as glucuronides in the human prostate," *Nutrition and Cancer*, vol. 60, no. 4, pp. 461–468, 2008.
- [61] M. Kano, T. Takayanagi, K. Harada, S. Sawada, and F. Ishikawa, "Bioavailability of isoflavones after ingestion of soy beverages in healthy adults," *Journal of Nutrition*, vol. 136, no. 9, pp. 2291–2296, 2006.
- [62] G. A. Reed, D. W. Arneson, W. C. Putnam et al., "Single-dose and multiple-dose administration of indole-3-carbinol to women: pharmacokinetics based on 3,3'-diindolylmethane," *Cancer Epidemiology Biomarkers and Prevention*, vol. 15, no. 12, pp. 2477–2481, 2006.
- [63] M. J. Anderton, M. M. Manson, R. D. Verschoyle et al., "Pharmacokinetics and tissue disposition of indole-3-carbinol and its acid condensation products after oral administration to mice," *Clinical Cancer Research*, vol. 10, no. 15, pp. 5233–5241, 2004.
- [64] A. Burkon and V. Somoza, "Quantification of free and protein-bound trans-resveratrol metabolites and identification of trans-resveratrol-C/O-conjugated diglucuronides—two novel resveratrol metabolites in human plasma," *Molecular Nutrition and Food Research*, vol. 52, no. 5, pp. 549–557, 2008.
- [65] D. J. Boocock, K. R. Patel, G. E. S. Faust et al., "Quantitation of trans-resveratrol and detection of its metabolites in human plasma and urine by high performance liquid chromatography," *Journal of Chromatography B*, vol. 848, no. 2, pp. 182–187, 2007.
- [66] P. A. Egner, T. W. Kensler, J.-G. Chen, S. J. Gange, J. D. Groopman, and M. D. Friesen, "Quantification of sulforaphane mercapturic acid pathway conjugates in human urine by high-performance liquid chromatography and isotope-dilution tandem mass spectrometry," *Chemical Research in Toxicology*, vol. 21, no. 10, pp. 1991–1996, 2008.
- [67] A. A. Al Janobi, R. F. Mithen, A. V. Gasper et al., "Quantitative measurement of sulforaphane, iiberin and their mercapturic acid pathway metabolites in human plasma and urine using liquid chromatography-tandem electrospray ionisation mass spectrometry," *Journal of Chromatography B*, vol. 844, no. 2, pp. 223–234, 2006.
- [68] J. M. W. van den Ouweland, A. M. Beijers, P. N. M. Demacker, and H. van Daal, "Measurement of 25-OH-vitamin D in human serum using liquid chromatography tandem-mass spectrometry with comparison to radioimmunoassay and automated immunoassay," *Journal of Chromatography B*, vol. 878, no. 15-16, pp. 1163–1168, 2010.
- [69] P. Wang, W. J. Aronson, M. Huang et al., "Green tea polyphenols and metabolites in prostatectomy tissue: implications for

cancer prevention,” *Cancer Prevention Research*, vol. 3, no. 8, pp. 985–993, 2010.

- [70] S. G. Kasimsetty, D. Blalonska, M. K. Reddy, G. Ma, S. I. Khan, and D. Ferreira, “Colon cancer chemopreventive activities of pomegranate ellagitannins and urolithins,” *Journal of Agricultural and Food Chemistry*, vol. 58, no. 4, pp. 2180–2187, 2010.
- [71] M. Larrosa, A. González-Sarriás, M. T. García-Conesa, F. A. Tomás-Barberán, and J. C. Espín, “Urolithins, ellagic acid-derived metabolites produced by human colonic microflora, exhibit estrogenic and antiestrogenic activities,” *Journal of Agricultural and Food Chemistry*, vol. 54, no. 5, pp. 1611–1620, 2006.

Review Article

Exploring Airway Diseases by NMR-Based Metabonomics: A Review of Application to Exhaled Breath Condensate

Matteo Sofia,¹ Mauro Maniscalco,¹ Guglielmo de Laurentiis,¹ Debora Paris,² Dominique Melck,² and Andrea Motta²

¹Department of Respiratory Medicine, A.O. Monaldi, University of Naples Federico II, 80131 Naples, Italy

²Institute of Biomolecular Chemistry, National Research Council, 80078 Pozzuoli (Naples), Italy

Correspondence should be addressed to Mauro Maniscalco, mauromaniscalco@hotmail.com

Received 16 May 2010; Revised 27 December 2010; Accepted 16 January 2011

Academic Editor: Mika Ala-Korpela

Copyright © 2011 Matteo Sofia et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

There is increasing evidence that biomarkers of exhaled gases or exhaled breath condensate (EBC) may help in detecting abnormalities in respiratory diseases mirroring increased, oxidative stress, airways inflammation and endothelial dysfunction. Beside the traditional techniques to investigate biomarker profiles, “omics” sciences have raised interest in the clinical field as potentially improving disease phenotyping. In particular, metabonomics appears to be an important tool to gain qualitative and quantitative information on low-molecular weight metabolites present in cells, tissues, and fluids. Here, we review the potential use of EBC as a suitable matrix for metabonomic studies using nuclear magnetic resonance (NMR) spectroscopy. By using this approach in airway diseases, it is now possible to separate specific EBC profiles, with implication in disease phenotyping and personalized therapy.

1. Introduction

Metabonomics is “the quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification” [1] due to any exposure (including drug administration), lifestyle and environmental stress. It, therefore, appears to be a powerful tool to monitor possible changes in metabolic pathways, and measure the levels of biochemical molecules generated in a living system. Metabolites are small molecules with molecular mass ≤ 1 kD [2] and are the end products of cellular activity. Observation of changes in metabolite concentrations may reveal the range of biochemical effects induced by a disease condition or its therapeutic intervention. The metabonomic analysis has two major potential applications, with implications in early diagnosis and disease phenotyping. It may also allow the recognition of unexpected or even unknown metabolites to formulate new pathophysiological hypotheses [3]. Moreover, the identification of individual metabolic characteristics could predict personal drug effectiveness and/or toxicity [4, 5].

The application of metabonomic analysis in chronic airway diseases has not been fully explored, but it holds a valid background. Several airway diseases, such as asthma or chronic obstructive pulmonary disease (COPD), which are largely spread in the population, cannot be qualified by a single biomarker and need a system biology analysis. Furthermore, other airway diseases such as cystic fibrosis (CF), although characterized by genetic abnormality, might be fruitfully investigated. Finally, the respiratory tract offers a natural matrix, the exhaled breath, which appears to be noteworthy for metabonomic analysis. Exhaled breath contains many different molecular species such as small inorganic molecules like nitric oxide (NO) or carbon monoxide (CO), volatile organic compounds (VOCs), and so forth, [6], which can be assayed in both the liquid and gaseous phases.

2. NMR-Metabonomics

The principal techniques used in metabonomics of breath (“breathomics”) are mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy, since they can

handle complex biological samples with a high sensitivity, selectivity, and high throughput [7].

MS, usually combined with chromatographic separation methods, separates the molecules of a sample on the basis of their retention time and mass-to-charge ratio m/z , and their representation in a spectrum [8, 9].

Real-time measurements of breath are also possible using direct breathing ports and techniques such as proton transfer reaction-mass spectrometry (PTR-MS), selected ion flow tube-mass spectrometry (SIFT-MS), and ion mobility spectrometry (IMS), as well as other analytical techniques including chemical sensors and various forms of laser spectrometers [6]. MS metabolomics has recently been applied to CF, where airway inflammation brings about an increased production of reactive oxygen species, responsible for degradation of cell membranes and causing the formation of VOCs. Robroeks et al. [10] have evaluated if VOCs metabolomics, analyzed by gas chromatography-time of flight-mass spectrometry to assess VOC profiles, could discriminate CF and controls, and CF patients with and without *Pseudomonas* colonization. By using 22 VOCs, a 100% correct separation of CF patients and controls was possible, while with 10 VOCs, 92% of the subjects were correctly classified. The reproducibility of VOC measurements with a one-hour interval was very good. The authors concluded that metabolomics of VOCs in exhaled breath was possible in a reproducible way, and it was not only able to discriminate between CF patients and controls, but also between CF patients with or without *Pseudomonas* colonization.

NMR spectroscopy studies molecules by recording the interaction of a radiofrequency electromagnetic radiation with the nuclei (e.g., ^1H , ^{13}C , ^{15}N , etc.) placed in a strong magnetic field. A single nucleus in a molecule can be “observed” by monitoring the corresponding line (a “resonance”) in an NMR spectrum, and the various parameters of that line (frequency, splitting, linewidth and amplitude) can be used to determine the molecular structure, conformation and dynamics of the molecule. In principle, assignment (i.e., identification) of NMR resonances for common metabolites could be possible by comparing the observed chemical shifts (i.e., the position of the line in a spectrum) with published reference data. When dealing with metabolites of unknown structure, chemical procedures for the separation of each molecule and use of two-dimensional NMR experiments (that spread signals in two dimensions) are required. Since NMR spectra show hundred of resonances, the presence of a discriminating element (e.g., a signal characteristic of a specific metabolite) in a series of spectra is often undetectable by visual inspection due to the inherent spectral complexity generated by line overlapping, and it is better highlighted by multivariate analysis (principal component analysis, PCA), which carefully identifies hidden phenomena and trends in ensembles of spectra [11]. The application of PCA to a group of spectra can immediately show whether all spectra behave similarly grouping in a single class, or fall apart into different groups. The main advantage of using NMR spectroscopy is its ability to provide a rapid and accurate metabolic picture of the sample with minimal sample pretreatment [12]. Furthermore, since the technique

is nondestructive, the samples can be investigated several times as long as some preventative measures are taken to avoid metabolite degradation.

3. Use of NMR Metabonomics for the Study of Airways

Metabonomics has been employed to investigate several body fluids such as urine, plasma, serum, and tissue extracts as well as *in-vivo* cells and their extracts [13], but only few applications to airway diseases characterization have been reported.

Airway hyperreactivity (AHR), an important characteristic of airway pathophysiology in human asthma, has recently been evaluated in an animal model of asthma exacerbation by urine NMR-based metabonomics [14]. The authors assumed that airway dysfunction and inflammation would produce unique patterns of urine metabolites observed by high-resolution proton (^1H) NMR spectroscopy, and the data analyzed by multivariate statistical analysis. In this model, challenged (ovalbumin, administered intraperitoneally, plus ovalbumin aerosol) guinea pigs developed AHR and increased inflammation compared with sensitized or control animals. Partial least-squares discriminant analysis using concentration differences in metabolites could discriminate challenged animals with 90% accuracy. Noteworthy, urine metabonomic profiles were able to separate not only sensitized from challenged and from naïve animals, but also from animals treated with dexamethasone which improves AHR. Recently, Slupsky et al. demonstrated specific changes in NMR metabonomic urinary profiles during episodes of pneumonia caused by *Streptococcus pneumoniae* or *Staphylococcus aureus* [14].

NMR metabonomics was also used to study the mechanism behind the formation of airway biofilm caused by *Pseudomonas aeruginosa*, an infection particularly prevalent in patients with CF [15]. In this kind of patients, the sessile lifestyle, referred to as a biofilm, allows the antibiotic resistance and makes easier the process of colonization through the synthesis of sticky, polymeric compounds. In contrast, the planktonic, free-floating cells are more easily eradicated with antibiotics. In this study, chemical differences between planktonic and biofilm cells, based on ^1H -NMR, have been reported. In this study, NMR techniques have highlighted the metabolic differences between the two modes of growth in *P. aeruginosa*, and PCA, and spectral comparisons revealed that the overall metabolism of planktonic and biofilm cells displayed marked differences, which require more extensive NMR investigations.

More recently [16], metabolite profiles of bronchoalveolar lavage fluid (BALF) from pediatric patients with CF were correlated to the degree of airway inflammation using NMR-based metabonomics. BALF was collected from 11 children with CF during clinically indicated bronchoscopy. The BALF spectra with high levels of neutrophilic airway inflammation displayed signals from numerous metabolites whereas the spectra from subjects with low levels of inflammation were very sparse. The metabolites identified in samples taken from subjects with high inflammation include known markers of

inflammation such as amino acids and lactate, as well as many novel signals. Statistical analysis highlighted the most important metabolites that distinguished the high- from the low-inflammation groups. This first demonstration of metabonomics of human BALF shows that clear distinctions in the metabolic profiles can be observed between subjects experiencing high versus low inflammation. However, the bronchoalveolar lavage has the important limitation of being invasive, requiring the introduction of exogenous fluid into alveolar space.

4. EBC

EBC is a noninvasive method of sampling the airways; it can be easily repeated and is acceptable to patients. Currently, EBC is used to measure biomarkers of airway inflammation and oxidative stress, and guidelines for its use have been recently published [17]. EBC can also be considered a matrix for analysis of environmental toxicants.

EBC collection requires the cooling of the exhaled breath (Figure 1(a)), resulting in a fluid sample that contains evaporated and condensed particles (water, ammonia, etc.) plus some droplets from the airway lining fluid [17, 18]. These droplets are released by turbulent airflow and can be added to the water vapor from anywhere between the alveoli and the mouth. Therefore, not only volatiles, but also several other mediators with nonvolatile characteristics can be found in EBC samples, including adenosine, different interleukins (-4, -5, -8), interferon- γ [17]. EBC is mainly (>99%) formed by water vapor, but also contains aerosol particles in which several other biomolecules including leukotrienes, 8-isoprostane, prostaglandins, hydrogen peroxide, nitric oxide-derived products, and hydrogen ions, can be detected [17]. EBC markers of oxidative stress such as hydrogen peroxide, isoprostanes, nitrogen oxides, pH, ammonia, prostanoids and leukotrienes are increased in bronchial asthma [19]. EBC pH is lower in asthmatics and correlates well with sputum eosinophilia, total nitrate/nitrite, and oxidative stress [20], but did not reflect the clinical status of the patients. EBC markers that correlate with disease severity, response to treatment, or both are hydrogen peroxide, leukotrienes, 8-isoprostane, nitrate, and nitrite [10]. It is assumed that airway surface liquid becomes aerosolized during turbulent airflow so that the content of the condensate reflects the composition of airway surface liquid, although large molecules may not aerosolize as well as small soluble molecules.

The major advantage of EBC is represented by the possibility to analyze both volatile and nonvolatile compounds [21]. There are some recent approaches to compare traditional blood test (glucose and urea) with the EBC in metabolic diseases. Accordingly, glucose in EBC from healthy volunteers was reproducible, unaffected by changes in salivary glucose, and increased during experimental hyperglycaemia [22].

Notably, EBC parameters are influenced by smoking, alcohol consumption, equipment, exercise, mode and rate of breathing, nasal contamination, environmental temperature and humidity, and assays used [23, 24], leading to undesirable variability. Exogenous contamination may also originate

from the oral cavity. Ammonia and sulfur-containing compounds like H₂S, methyl sulfide or mercaptans are released from the oral cavity, being produced by bacteria from different oral niches. However, oral sterilization before EBC collection or continuous saliva deglutition have been suggested to limit the effects of such contaminations [14]. The influence of age, sex, circadian rhythm, and infection remains unknown. Thus the analysis of EBC currently has important limitations.

Reference analytical techniques are required to provide definitive evidence for the presence of some inflammatory mediators in EBC and for their accurate quantitative assessment in this biological fluid. Finally, the physiological meaning and biochemical origin of most of volatile compounds are still not known, and biochemical pathways of their generation, origin, and distribution are only partly understood. Unfortunately, the concentrations of various mediators studied are very low, requiring highly sensitive assays.

5. Metabonomics of EBC in Respiratory Diseases

NMR-based metabonomics can be used to analyze EBC samples from adults, allowing a clear-cut separation between healthy subjects and patients with airway disease [11]. Although less sensitive than ELISA and MS, NMR spectroscopy requires minimal sample preparation with a rapid acquisition time of spectra (10–15 min). Furthermore, it shows a high degree of sensitivity ($\leq \mu\text{mol/L}$), and is nondestructive, allowing complete detection of metabolites present in the sample (“sample metabolic fingerprint”) at a reasonable cost. NMR is also able to detect potential contamination of EBC from saliva, and examine the interfering effect of residual external contaminants, which is crucial for a correct EBC analysis of the variability of some biomarkers [11, 25, 26].

To date there are several recommendations on the methodological approach to EBC collection, but its standardization is not completely defined, as EBC can be contaminated by metabolites originating from saliva as well as microbes present in the mouth [27, 28]. We have recently proposed a possible protocol for EBC collection for NMR purposes (Figure 1) [11]. It requires that subjects breath through a mouthpiece and a two-way nonbreathing valve, which also serve as a saliva trap, at normal frequency and tidal volume, while sitting comfortably and wearing a nose-clip, for a period of 15 minutes (Figure 1(a)) [29]. They maintained a dry mouth during collection by periodically swallowing excess saliva. Condensate samples (1–2 mL) are immediately transferred into glass vials, closed with 20-mm butyl rubber lined with PTFE septa, and crimped with perforated aluminum seals. Volatile substances, possibly deriving from extrapulmonary sources are removed by applying a gentle stream of nitrogen for 3 minutes before sealing [30, 31]. Nitrogen was used because concentration of volatile solutes in EBC is dependent on their distribution between the saliva, exhaled air, and droplets, and the condensate, which can be altered by multiple factors including minute

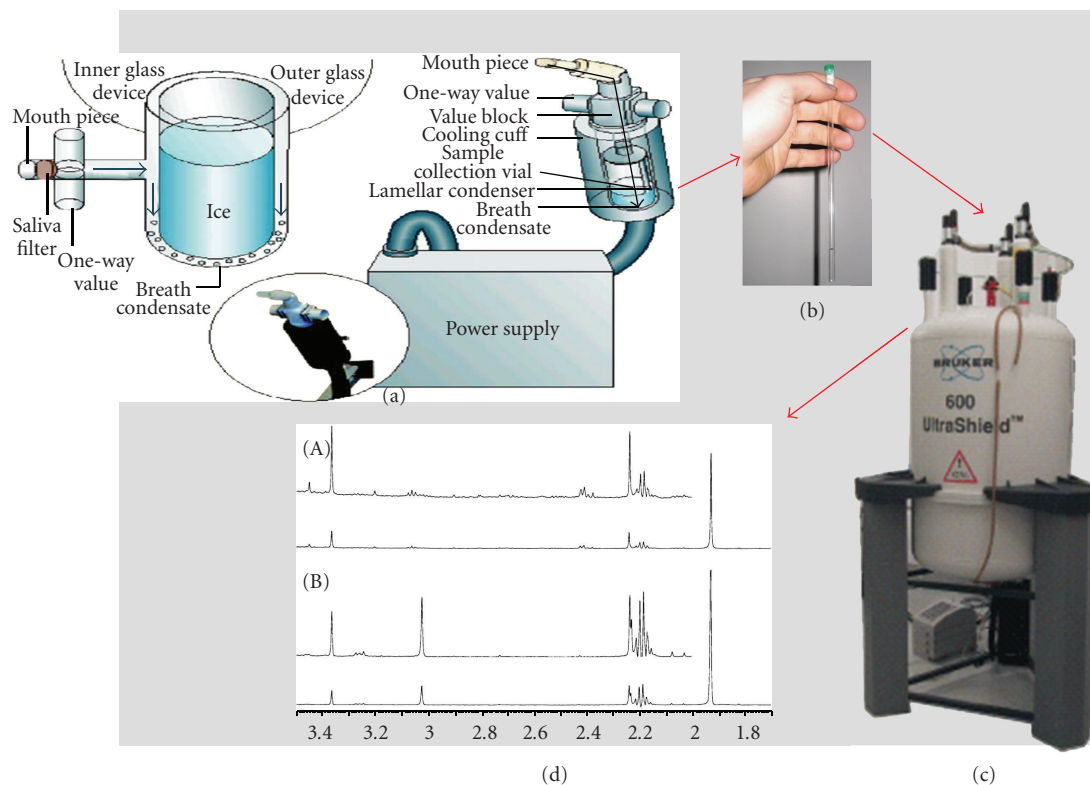


FIGURE 1: Metabonomics of EBC using NMR. The exhaled breath is cooled in (a), then transferred into the NMR tube (0.5–0.7 mL) (b) and put in the spectrometer (c) to collect the spectra (d).

ventilation, salivary pH, solubility, temperature, and sample preparation [29]. Therefore, spectral differences may depend upon uncontrollable variables that prevent reliable quantification. The nitrogen stream also removes oxygen from solutions, which, together with freezing of sealed samples in liquid nitrogen, immediately “quenches” metabolism at the collection time, and prevents any metabolic decay [32, 33]. Samples are then stored at -80°C until NMR analysis. Drying of the samples should be avoided to circumvent irreversible solute precipitation, and/or formation of insoluble aggregates, which we observed upon dissolving the dried condensate for NMR measurements.

Before NMR acquisition, EBC samples should be rapidly defrosted and transferred into the NMR tube (Figure 1(b)). To provide a field frequency lock for NMR acquisition, $70\ \mu\text{L}$ of a D_2O solution [containing 1 mM sodium 3-trimethylsilyl [2,2,3,3- $^2\text{H}_4$] propionate (TSP) as a chemical shift and concentration reference for ^1H spectra, and sodium azide at 3 mM] are added to $630\ \mu\text{L}$ of condensate reaching $700\ \mu\text{L}$ of total volume.

Following acquisition (Figures 1(c) and 1(d)), ^1H -NMR spectra are automatically data reduced to 200–250 integral segments (“buckets”) using dedicated software packages (e.g., AMIX, Bruker Biospin, Germany). The resulting integrated regions are then used for statistical analysis and pattern recognition. To avoid possible errors in signal intensity due to difference in the volume of collected EBC samples, before pattern recognition analysis each integral

region is normalized to the sum of all integral region of each spectrum. In the presence of contaminant peaks (e.g., those originating from the condenser disinfectant), which randomly alter the total area of the spectrum, each bucket has to be normalized to the TSP peak of known concentration, referring to a standard region comprised, for example, between 0.014 and -0.014 ppm.

Figure 2 represents spectra of saliva (left traces) and EBC samples (right traces) from a healthy subject (HS, lower spectra), a laryngectomized (middle spectra) and a COPD (top spectra) subjects. A visual examination establishes a striking correspondence between EBC spectra of HS and laryngectomized, suggesting that potential oral contamination (e.g., bacteria and/or saliva) is undetectable and, if present, beyond the sensitivity limit of NMR. By resorting to literature data [34] and two-dimensional NMR experiments, we identified all resonances present in EBC spectra (Figure 3). In saliva, signals between 3.3 and 6.0 ppm originate from carbohydrates [35] and these are virtually absent in the EBC spectra (Figure 2). Compared to saliva, EBC spectra present fewer signals, but both saliva and EBC spectra appear to be different among the HS, laryngectomized and COPD classes considered (Figure 2); this is the basis for class separation in PCA based on NMR data.

A recent study [11] has evaluated the capability of NMR to separate EBC subjects with airway diseases (COPD) from subjects without respiratory diseases. Based on qualitative and quantitative spectral differences, five NMR signals

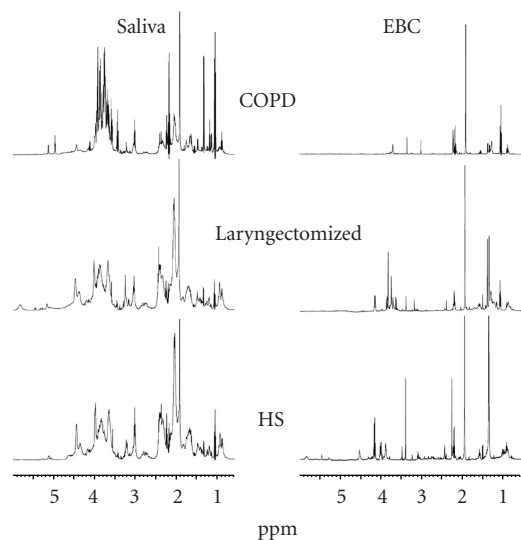


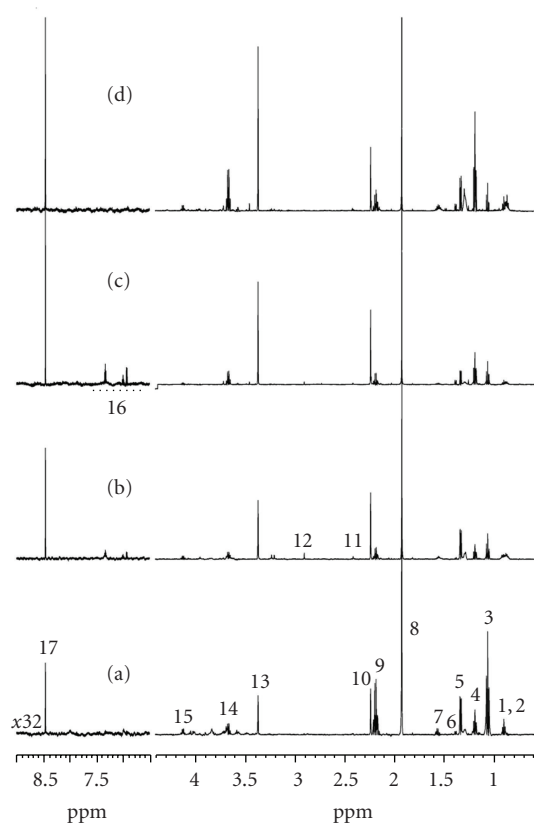
FIGURE 2: Representative one-dimensional ^1H -NMR spectra from different patients. Spectra of saliva (left traces) and EBC samples (right traces) from a healthy (HS, lower traces), a laryngectomized (middle spectra), and a COPD (top spectra) subject are reported. Spectra were recorded on a Bruker Avance spectrometer operating at a frequency of 600.13 MHz (^1H) and equipped with a TCI CryoProbe. The water resonance was suppressed by using a specific pulse sequence designed to avoid intensity alteration of signals. The total acquisition time was *ca.* 10 minutes per sample. Spectra were referred to TSP assumed to resonate at $\delta = 0.00$ ppm. In saliva spectra, the group of signals centered at 3.8 ppm originates from carbohydrates, and is not visible in the corresponding EBC spectra.

appear to differentiate “respiratory” (COPD) from “non-respiratory” (healthy and laryngectomized) subjects. It was also clearly proved that saliva and condensate have different profiles, with saliva contamination showing little influence on the interpretation of EBC by NMR-based metabonomics [11].

Likewise, Carraro et al. [36] reported the acetate signal variation as distinctive in asthmatic children with respect to controls, concluding that acetate increase might be related to increased acetylation of proinflammatory proteins in the extracellular space in the airway environment. Whether the metabonomic of exhaled breath condensate changes during systemic or metabolic disease is currently unknown.

NMR-based metabonomic analyses of EBC could clearly discriminate between asthmatic and healthy children, with 95% success rate in their classification. Many authors believe that asthma should no longer be considered a single disease, and that efforts should be made to identify the different biochemical and inflammatory profiles behind asthma symptoms in order to treat them with specifically-targeted therapies [37]. Montuschi et al. [38] recently applied NMR-based metabonomics to discriminate between healthy individuals, patients with stable CF, and cases of unstable CF, showing that NMR is a powerful technique to monitor EBC in CF.

In addition, we are currently applying NMR-based EBC metabonomics in other genetic airway diseases such as



| | |
|-----------------------|-------------------|
| 1, 2 Leu + n-Butyrate | 10 Pyruvate |
| 3 Propionate | 11 Succinate |
| 4 Thr | 12 Trimethylamine |
| 5 Lactate | 13 Methanol |
| 6 Ala | 14 Ethanol |
| 7 n-Butyrate | 15 Lactate |
| 8 Acetate | 16 Phe |
| 9 n-Butyrate | 17 Formate |

FIGURE 3: Resonance assignments of representative ^1H -NMR spectra of EBC samples. The spectra of healthy (a), asthmatic (b), steroid-treated asthmatic (c), and COPD (d) patients are depicted. Peaks are labeled with progressive numbers, and assignments are listed below.

primary ciliary dyskinesia, in light of the diffusion of fast screening methods based on the exhaled NO on nasal or oral breath.

6. Conclusions

The power of NMR-based metabonomics has been shown for several biofluids, including blood, urine, and saliva. We believe that NMR metabonomics could also be applied to EBC, which has the advantage of being noninvasive and reproducible; furthermore it shows a distinctive profile in comparison to saliva, thus supporting its origin from lower airways. Moreover, EBC metabonomic analysis is applied to a living matrix in the absence of external induced perturbations that may represent important preanalytical variable with conventional assay measuring single compound.

There is only limited experience with metabonomics on EBC in humans, but reproducibility of method has been successfully assessed, and useful protocols to differentiate metabolic profile of patients with asthma, COPD, or cystic fibrosis have been reported. However, more studies are needed to show, if true, that the holistic approach of EBC metabonomics may be a progress over the traditional reductionistic approach in chronic airway disease.

References

- [1] J. K. Nicholson, J. C. Lindon, and E. Holmes, “Metabonomics”: understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data,” *Xenobiotica*, vol. 29, no. 11, pp. 1181–1189, 1999.
- [2] M. E. Bollard, E. G. Stanley, J. C. Lindon, J. K. Nicholson, and E. Holmes, “NMR-based metabonomic approaches for evaluating physiological influences on biofluid composition,” *NMR in Biomedicine*, vol. 18, no. 3, pp. 143–162, 2005.
- [3] U. Sauer, M. Heinemann, and N. Zamboni, “Getting closer to the whole picture,” *Science*, vol. 316, no. 5824, pp. 550–551, 2007.
- [4] J. M. Meyer and G. S. Ginsburg, “The path to personalized medicine,” *Current Opinion in Chemical Biology*, vol. 6, no. 4, pp. 434–438, 2002.
- [5] T. A. Clayton, J. C. Lindon, O. Cloarec et al., “Pharmacometabonomic phenotyping and personalized drug treatment,” *Nature*, vol. 440, no. 7087, pp. 1073–1077, 2006.
- [6] W. M. Amann, J. Pleil, T. Risby, and J. Schubert, “Methodological issues of sample collection and analysis of exhaled breath,” *European Respiratory Monthly*, vol. 49, pp. 96–114, 2010.
- [7] M. Coen, E. Holmes, J. C. Lindon, and J. K. Nicholson, “NMR-based metabolic profiling and metabonomic approaches to problems in molecular toxicology,” *Chemical Research in Toxicology*, vol. 21, no. 1, pp. 9–27, 2008.
- [8] K. Dettmer, P. A. Aronov, and B. D. Hammock, “Mass spectrometry-based metabolomics,” *Mass Spectrometry Reviews*, vol. 26, no. 1, pp. 51–78, 2007.
- [9] X. Feng, X. Liu, Q. Luo, and B. I. F. Liu, “Mass spectrometry in systems biology: an overview,” *Mass Spectrometry Reviews*, vol. 27, no. 6, pp. 635–660, 2008.
- [10] C. M. H. H. T. Robroeks, J. J. B. N. Van Berkel, J. W. Dallinga et al., “Metabolomics of volatile organic compounds in cystic fibrosis patients and controls,” *Pediatric Research*, vol. 68, no. 1, pp. 75–80, 2010.
- [11] G. de Laurentiis, D. Paris, D. Melck et al., “Metabonomic analysis of exhaled breath condensate in adults by nuclear magnetic resonance spectroscopy,” *European Respiratory Journal*, vol. 32, no. 5, pp. 1175–1183, 2008.
- [12] E. J. Saude, I. P. Obiefuna, R. L. Somorjai et al., “Metabolomic biomarkers in a model of asthma exacerbation: urine nuclear magnetic resonance,” *American Journal of Respiratory and Critical Care Medicine*, vol. 179, no. 1, pp. 25–34, 2009.
- [13] O. Beckonert, H. C. Keun, T. M. Ebbels et al., “Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts,” *Nature protocols*, vol. 2, no. 11, pp. 2692–2703, 2007.
- [14] C. M. Slupsky, A. Chyepesh, D. V. Chao et al., “Streptococcus pneumoniae and Staphylococcus aureus pneumonia induce distinct metabolic responses,” *Journal of Proteome Research*, vol. 8, no. 6, pp. 3029–3036, 2009.
- [15] E. L. Gjersing, J. L. Herberg, J. Horn, C. M. Schaldach, and R. S. Maxwell, “NMR metabolomics of planktonic and biofilm modes of growth in *Pseudomonas aeruginosa*,” *Analytical Chemistry*, vol. 79, no. 21, pp. 8037–8045, 2007.
- [16] J. E. Wolak, C. R. Esther, and T. M. O’Connell, “Metabolomic analysis of bronchoalveolar lavage fluid from cystic fibrosis patients,” *Biomarkers*, vol. 14, no. 1, pp. 55–60, 2009.
- [17] I. Horváth, J. Hunt, P. J. Barnes et al., “Exhaled breath condensate: methodological recommendations and unresolved questions,” *European Respiratory Journal*, vol. 26, no. 3, pp. 523–548, 2005.
- [18] S. A. Kharitonov and P. J. Barnes, “Exhaled markers of pulmonary disease,” *American Journal of Respiratory and Critical Care Medicine*, vol. 163, no. 7, pp. 1693–1722, 2001.
- [19] R. Accordino, A. Visentin, A. Bordin et al., “Long-term repeatability of exhaled breath condensate pH in asthma,” *Respiratory Medicine*, vol. 102, no. 3, pp. 377–381, 2008.
- [20] K. Ganas, S. Loukides, G. Papatheodorou, P. Panagou, and N. Kalogeropoulos, “Total nitrite/nitrate in expired breath condensate of patients with asthma,” *Respiratory Medicine*, vol. 95, no. 8, pp. 649–654, 2001.
- [21] M. Maniscalco, G. De Laurentiis, C. Pentella et al., “Exhaled breath condensate as matrix for toluene detection: a preliminary study,” *Biomarkers*, vol. 11, no. 3, pp. 233–240, 2006.
- [22] E. H. Baker, N. Clark, A. L. Brennan et al., “Hyperglycemia and cystic fibrosis alter respiratory fluid glucose concentrations estimated by breath condensate analysis,” *Journal of Applied Physiology*, vol. 102, no. 5, pp. 1969–1975, 2007.
- [23] K. Czebe, I. Barta, B. Antus, M. Valyon, I. Horváth, and T. Kullmann, “Influence of condensing equipment and temperature on exhaled breath condensate pH, total protein and leukotriene concentrations,” *Respiratory Medicine*, vol. 102, no. 5, pp. 720–725, 2008.
- [24] T. Kullmann, I. Barta, B. Antus, M. Valyon, and I. Horváth, “Environmental temperature and relative humidity influence exhaled breath condensate pH,” *European Respiratory Journal*, vol. 31, no. 2, pp. 474–475, 2008.
- [25] P. Montuschi, S. A. Kharitonov, G. Ciabattini, and P. J. Barnes, “Exhaled leukotrienes and prostaglandins in COPD,” *Thorax*, vol. 58, no. 7, pp. 585–588, 2003.
- [26] P. Montuschi, “Exhaled breath condensate analysis in patients with COPD,” *Clinica Chimica Acta*, vol. 356, no. 1–2, pp. 22–34, 2005.
- [27] P. Latzin, J. Beck, A. Bartenstein, and M. Griese, “Comparison of exhaled breath condensate from nasal and oral collection,” *European Journal of Medical Research*, vol. 8, pp. 505–510, 2003.
- [28] J. Chladkova, I. Krcmova, J. Chladek, P. Cap, S. Micuda, and Y. Hanzalkova, “Validation of nitrite and nitrate measurements in exhaled breath condensate,” *Respiration*, vol. 73, no. 2, pp. 173–179, 2006.
- [29] R. M. Effros, “Exhaled breath condensate pH,” *American Journal of Respiratory and Critical Care Medicine*, vol. 173, no. 9, pp. 1047–1048, 2006.
- [30] R. M. Effros, K. W. Hoagland, M. Bosbous et al., “Dilution of respiratory solutes in exhaled condensates,” *American Journal of Respiratory and Critical Care Medicine*, vol. 165, no. 5, pp. 663–669, 2002.
- [31] C. L. Whittle, S. Fakharzadeh, J. Eades, and G. Preti, “Human breath odors and their use in diagnosis,” *Annals of the New York Academy of Sciences*, vol. 1098, pp. 252–266, 2007.
- [32] J. D. Bell, J. C. Brown, and P. J. Sadler, “NMR studies of body fluids,” *NMR in Biomedicine*, vol. 2, no. 5–6, pp. 246–256, 1989.

- [33] O. Beckonert, H. C. Keun, T. M. Ebbels et al., "Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts," *Nature protocols*, vol. 2, no. 11, pp. 2692–2703, 2007.
- [34] T. W. M. Fan, "Metabolite profiling by one- and two-dimensional NMR analysis of complex mixtures," *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 28, no. 2, pp. 161–219, 1996.
- [35] M. Grootveld and C. J. L. Silwood, "¹H-NMR analysis as a diagnostic probe for human saliva," *Biochemical and Biophysical Research Communications*, vol. 329, no. 1, pp. 1–5, 2005.
- [36] S. Carraro, S. Rezzi, F. Reniero et al., "Metabolomics applied to exhaled breath condensate in childhood asthma," *American Journal of Respiratory and Critical Care Medicine*, vol. 175, no. 10, pp. 986–990, 2007.
- [37] C. Auffray, I. M. Adcock, K. F. Chung, R. Djukanovic, C. Pison, and P. J. Sterk, "An integrative systems biology approach to understanding pulmonary diseases," *Chest*, vol. 137, no. 6, pp. 1410–1416, 2010.
- [38] P. Montuschi, D. Paris, D. Melck, V. Lucidi, and A. Motta, "Metabolomic analysis by nuclear magnetic resonance spectroscopy of exhaled breath condensate in patient with cystic fibrosis," *European Respiratory Journal*, vol. 34, supplement 53, p. 63s, 2009.

Methodology Report

metaP-Server: A Web-Based Metabolomics Data Analysis Tool

Gabi Kastenmüller,¹ Werner Römisch-Margl,¹ Brigitte Wägele,^{1,2}
Elisabeth Altmaier,¹ and Karsten Suhre^{1,3}

¹Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany

²Chair of Genome Oriented Bioinformatics, Life and Food Science Center of Weihenstephan, Technische Universität München, Maximus-von-Imhof-Forum 3, 85354 Freising-Weihenstephan, Germany

³Department of Biology, Ludwig-Maximilians-Universität München, Großhaderner Straße 2, 82152 Planegg-Martinsried, Germany

Correspondence should be addressed to Karsten Suhre, karsten.suhre@helmholtz-muenchen.de

Received 15 May 2010; Accepted 6 August 2010

Academic Editor: Olav Kvalheim

Copyright © 2011 Gabi Kastenmüller et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Metabolomics is an emerging field that is based on the quantitative measurement of as many small organic molecules occurring in a biological sample as possible. Due to recent technical advances, metabolomics can now be used widely as an analytical high-throughput technology in drug testing and epidemiological metabolome and genome wide association studies. Analogous to chip-based gene expression analyses, the enormous amount of data produced by modern kit-based metabolomics experiments poses new challenges regarding their biological interpretation in the context of various sample phenotypes. We developed *metaP-server* to facilitate data interpretation. *metaP-server* provides automated and standardized data analysis for quantitative metabolomics data, covering the following steps from data acquisition to biological interpretation: (i) data quality checks, (ii) estimation of reproducibility and batch effects, (iii) hypothesis tests for multiple categorical phenotypes, (iv) correlation tests for metric phenotypes, (v) optionally including all possible pairs of metabolite concentration ratios, (vi) principal component analysis (PCA), and (vii) mapping of metabolites onto colored KEGG pathway maps. Graphical output is clickable and cross-linked to sample and metabolite identifiers. Interactive coloring of PCA and bar plots by phenotype facilitates on-line data exploration. For users of commercial metabolomics kits, cross-references to the HMDB, LipidMaps, KEGG, PubChem, and CAS databases are provided. *metaP-server* is freely accessible at <http://metabolomics.helmholtz-muenchen.de/metap2/>.

1. Introduction

Metabolomics is an emerging “omics” technology that focuses on the identification and quantification of all or, in practice, the largest possible set of low-molecular-weight metabolites in a biological sample. In the series of the “omics” technologies genomics-transcriptomics-proteomics-metabolomics, metabolomics describes the physiological endpoint arising from the interplay of all regulatory and enzymatic processes in the biological system under consideration at a given time [1–4]. In other words, metabolomics analyses show the net effect of environmental and genomic factors influencing the status of a biological system.

In the recent years, advances in nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS)

have allowed for quantitating hundreds of metabolites in blood and urine samples in a high-throughput manner. Due to the development of modern MS/MS-based analytical pipelines and metabolomics kits, application of metabolomics analyses is no longer restricted to specialized laboratories but can be used widely in biological and pharmaceutical research. As an example, metabolomics kits from Biocrates (AbsoluteIDQ) have been used to monitor effects of specific drugs on the metabolism of diabetic and nondiabetic mice [5] and in epidemiological studies to gain new insights into the effects of nutrition or genotypes on the human metabolism [6–8]. Other possible applications could be to data acquired on a set of different commercial platforms, such as MS data provided by Metabolon [9], Phenomenome [10], or NMR data processed using the Chenomx software suite [11].

Analogous to chip-based gene expression analyses, the enormous amount of data produced in high-throughput metabolomics experiments poses new challenges for automated data analysis. Various commercial and free stand-alone analysis tools dedicated to metabolomics data support experimentalists in aligning and binning peaks in MS and NMR spectra, provide functionality for annotating peaks with metabolites, or offer statistical analysis [12–14]. Recently, two web-based metabolomics data analysis tools have been published: (i) *MetaboAnalyst* (<http://www.metaboanalyst.ca>) [15], a tool that provides data analysis focusing on biomarker discovery and classification with respect to a single two-class phenotype (e.g., the sample phenotype treatment with the two classes treated and untreated), and (ii) *MeltDB* (<http://www.cebitec.uni-bielefeld.de/groups/brf/software/meltdb-info/index.html>) [16], which provides a data analysis pipeline for raw GC- and LC-MS data sets including metabolite identification. For preprocessed metabolite quantities, *MeltDB* offers statistical data analyses (e.g., *t*-test and PCA) with respect to the classes of a single phenotype. In contrast to *MetaboAnalyst* and *metaP-server*, *MeltDB* requires login and password to get access. While *MetaboAnalyst* and *MeltDB* are valuable tools for estimating the associations of a single phenotype with metabolite quantities, many experiments involve more than one phenotype with often more than two classes per phenotype. As an example, a metabolomics experiment for drug testing in mice can comprise phenotypes such as sex, strain (e.g., wild type/knock out strain), drug dose (e.g., 0/20/40 mg), and days of treatment (days 1–5). In this case, each sample measured is linked to the classes of multiple different phenotypes (e.g., Sample01: female; wild type; 20 mg; day 2). For such experiments, new tools are needed in order to get an overview over the observed trends in the metabolomics data across all phenotypes involved.

Here, we present *metaP-server*, a web-based, easy-to-use analysis tool for the statistical analysis of metabolomics data. In contrast to the existing web-servers, *metaP-server* mainly focuses on the interactive exploration and interpretation of metabolomics data (whether metabolite concentrations or peak lists) in the context of multiple multiclass (e.g., “treated with drug A, B, C”) and metric (e.g., weight, age) sample phenotypes. Thus, the *metaP-server* supports experimentalist in gaining first insights into how the different sample phenotypes affect the metabolite quantities observed. These insights facilitate choosing the data subsets and phenotypes that should be analyzed by using further statistical methods for classification and biomarker discovery (as, e.g., provided by *MetaboAnalyst*). *metaP-server* provides hypothesis tests and correlation tests for nonmetric and metric phenotypes, optionally including all possible pairs of metabolite concentration ratios as quantitative traits. As shown in previous metabolomics studies, using ratios can reduce noise caused by individual differences in absolute metabolite concentrations and, thus, strengthen the association [5, 7, 8]. Furthermore, the server offers principal component analysis (PCA). PCA plots and barplots showing the concentration of a particular metabolite in the samples

can be colored interactively by phenotypes. Moreover, the graphical output is clickable and cross-linked to the respective sample or metabolite pages. Concentrations of metabolites in samples relative to the mean are mapped onto colored KEGG pathway maps. Interactive coloring, cross-linking, and pathway mapping particularly aim to facilitate on-line data exploration in the context of multiple phenotypes. For the special needs of kit-based high-throughput experiments, we implemented functions for the estimation of reproducibility and batch effects, as well as outlier detection. For users of the Biocrates *AbsoluteIDQ* kit, original cross references to the HMDB [17], LipidMaps [18], KEGG [19], PubChem [20], and CAS databases have been derived and are freely provided. The *metaP-server* is freely accessible at <http://metabolomics.helmholtz-muenchen.de/metap2/>.

2. Data Input

To start a new analysis in *metaP-server* (link: “Start a new run” on the main page), users are asked to provide the metabolomics quantitation data table in semicolon separated format, which can be exported from most spreadsheets. Users can choose between three different input formats: (i) “quant. data”, (ii) “quant. data with KEGG ids”, and (iii) “*AbsoluteIDQ* kit”. The server expects samples to be in rows and quantitated metabolites or peak intensities in columns. While for *AbsoluteIDQ*, *metaP-server* directly accepts the export file (extension. csv) as produced by the *MetIQ* software (shipped with the kit), data from other experiments can be provided as a table where the first row contains identifiers for the metabolites and the first column contains unique sample identifiers such as barcodes. Optionally, a second column entitled “Sample Description” can contain user sample identifiers that are not necessarily unique. The server also accepts files with data starting at another column than column two or three, when the users specifies an additional parameter input frame for auxiliary data. Optionally, the user can provide KEGG identifiers with the data. For this purpose, the user must choose “quant. data with KEGG ids” and must add one or more rows with the keyword “KEGG” in the first column after the header row of the table.

In principle, the processing of the data can be started immediately after providing the metabolomics quantitation data. However, for using the full functionality of the server, phenotypes or experimental conditions related to the samples measured can be specified in a separate file. The first column of this table must contain the unique sample identifiers of the respective samples. Following columns may include any categorical (e.g., sex, strain) or metric (e.g., weight, age, drug dose) sample phenotype or experimental condition (e.g., batch number). Categorical phenotypes are not restricted in the number of categories. If phenotypes are described by numeric values but are categorical rather than continuous regarding the actual values (e.g., drug dose: 10 mg, 20 mg, 40 mg corresponding to low, medium, and high dose), these phenotypes are analyzed using both the hypothesis test for categorical phenotypes and the correlation test for metric phenotypes. Special key words for column

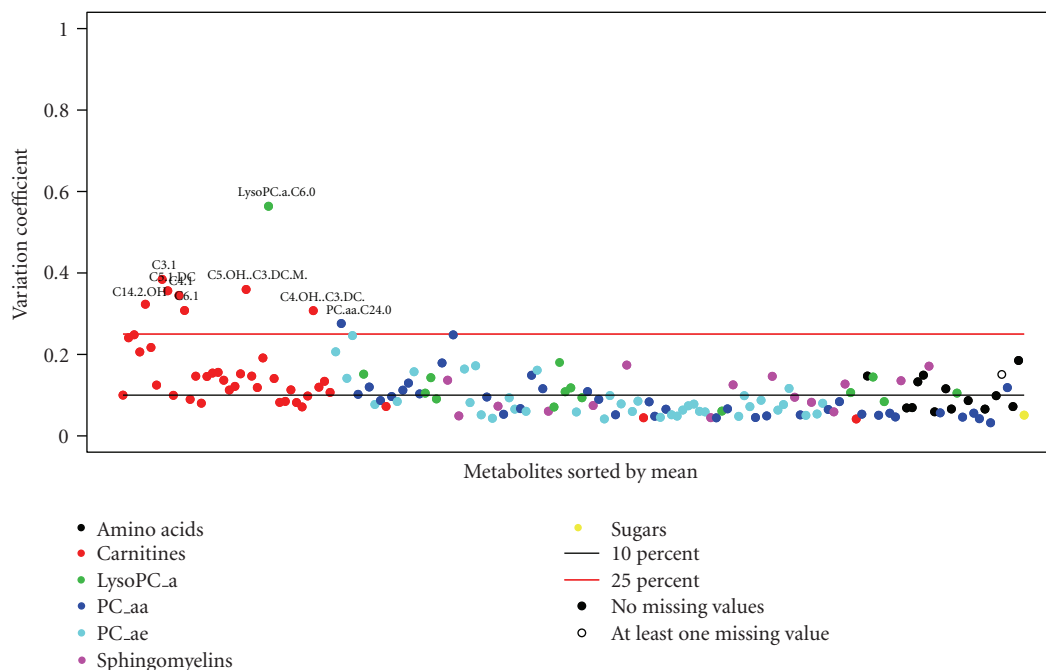


FIGURE 1: Example for output generated by *metaP-server*: plot of coefficient of variation for replicated measurement of reference samples (controls) as part of data quality checks.

headers such as “Replicates” and “Batches” can be used for defining control samples (replicated reference samples) and batches in the data set. By using these keywords in the uploaded phenotype file, basically any subset of samples can be specified as controls for the calculation of coefficients of variation (*cv*) and the estimation of batch effects. For this purpose, all control samples must be denoted by the same word (e.g., “control”) in the “Replicates” column. For data in *AbsoluteIDQ* format, replicated samples and batch information is extracted automatically if available.

Before starting the processing, the user can specify a job description and optionally provide an email address for notification regarding the job status. By default jobs are kept private. In this case, the data is only accessible via the unique job id created by the server. Moreover, the users can change the settings for several parameters (e.g., for forcing deletion of reference samples, outliers, and/or noisy metabolites before statistical data analysis, and forcing the calculation of metabolite ratios). All job results remain on the server for at least four weeks. Analysis results can be downloaded as an archive (zip) file.

3. Processing and Methods

After submitting the job, the web-server first tests the compliance of uploaded data with the format specified. The server then starts several analyses that are related to data quality control. Depending on the options chosen on the submission page, the server either deletes outliers, noisy metabolites, and reference samples before further data analysis or the server continues analysis based on all data points disregarding quality.

Though *metaP-server* does not explicitly restrict the number of samples, quantitated metabolites/peak intensities, or phenotypes that can be uploaded, the time required for the complete analysis including the generation of clickable images largely depends on these numbers. As an example, data analysis for a typical data set from a kit-based experiment with 96 samples and 163 metabolites took 2 minutes for two phenotypes. For 1000 samples and 200 metabolites the analysis of two phenotypes was finished after 27 minutes, while the analysis of 100 samples and 5000 peak intensities took 204 minutes. When the option for the calculation and analysis of all-against-all metabolite ratios was chosen for the first example with 96 samples and 163 metabolites, the run time increased to 24 minutes.

3.1. Data Quality Control. If replicated measurements of reference samples (controls) are provided with the data, *metaP-server* calculates the corresponding coefficient of variation (*cv*) for each metabolite and tags all metabolites with a *cv* above a given threshold. The *cv* for each metabolite is visualized in a diagram (Figure 1). For estimating batch effects in large metabolomics experiments, *metaP-server* provides *p*-values for the association of metabolite concentrations with batches. Boxplots showing the metabolomics data and, if available, the corresponding reference data depending on the batches help to immediately capture potential batch effects. *metaP-server* also reports outliers among the samples. Samples are considered as outliers if the metabolite quantities measured for this sample lie 1.5 times the inter quartile range (IQR) below or above the corresponding median for 30% of the data columns.

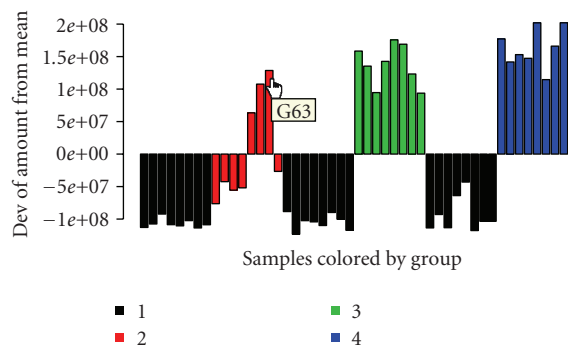


FIGURE 2: Examples for output generated by *metaP-server*: metabolite barplot colored by the phenotype “groups” with the classes 1–4.

The uploaded phenotypes are matched with the samples provided in the quantitation data according to the unique sample identifiers. Empty columns and columns containing nonnumeric values that have different values for all samples are ignored for further analysis.

3.2. Data Analysis in the Context of Sample Phenotypes. The main objective of *metaP-server* is allowing for the analysis of metabolomics data in the context of sample phenotypes. The following types of analyses are provided.

3.2.1. General Statistical Measures. *metaP-server* calculates general statistical measures for the metabolite quantifications including mean, median, and standard deviation in relation to the mean. The server also provides histograms for estimating the distribution of metabolite concentrations in the samples. QQ-Plots plotting the actual distribution versus the corresponding theoretical values for normal (red) and log-normal (black) distributions allow for deciding which of the theoretical distributions fits best. Metabolite barplots show the concentrations of a particular metabolite in all samples measured (Figure 2) and, vice versa, sample barplots visualize the concentrations of all metabolite concentrations measured for a particular sample (Figure 3). Metabolite barplots can be easily colored by the phenotypes.

3.2.2. Principal Component Analysis. In general, principal component analysis (PCA) transforms the original data into a new system of orthogonal axes (components) with the first components covering the major variance in the data. Thus, looking at the projections of the data onto the first principal components often reveals intrinsic groups in the data. PCA is an unsupervised method and, thus, does not use any prior phenotypic knowledge for calculating the principal components. Principal components represent combinations of the original dimensions (metabolites), whose contributions to the component can give hints which metabolites separate intrinsic groups (if any) best. Please note that groups becoming apparent on a PCA plot do not necessarily correspond to phenotype classes, since PCA—as an unsupervised method—is only based on the metabolomics data matrix without using any phenotypic information on groups. The so-called loadings of the components are

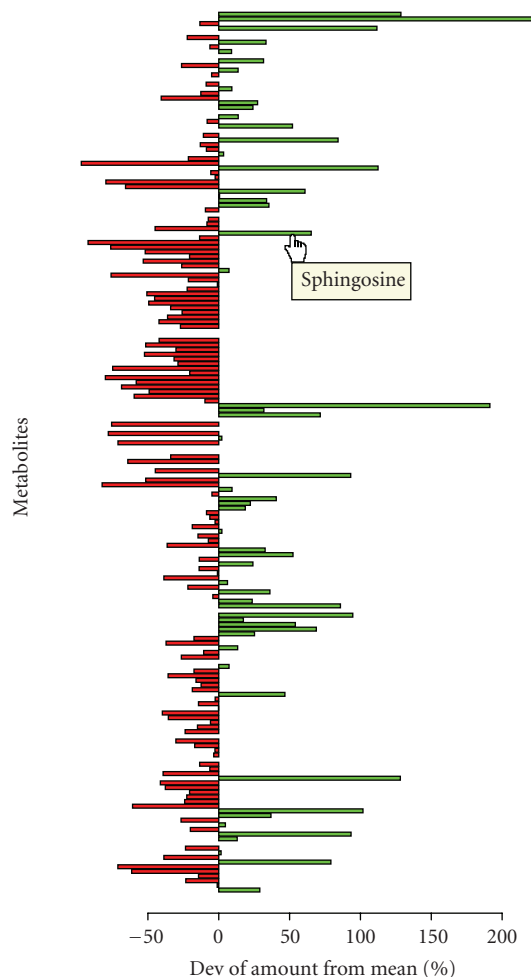


FIGURE 3: Examples for output generated by *metaP-server*: sample barplot with green color denoting high and red color denoting low concentration of the respective metabolite relative to the mean value.

provided for download in table format (semicolon separated values). Before PCA analysis, *metaP-server* scales the original data to mean 0 and standard deviation 1 in order to make the concentrations of the metabolites comparable. The server shows the proportion of variance covered by the first ten principal components and plots for the projections of the data to the first three components. The user can color the data points (each representing a specific sample) by the categorical phenotypes uploaded (Figure 4). Moreover, each data point is cross-linked to a sample page describing the details for the respective sample and showing the sample barplot described previously. Typical representatives of specific phenotypic groups as well as extreme samples can thus be picked easily.

3.2.3. Hypothesis Tests and Correlation Analysis. For testing the association of metabolite concentrations with categorical phenotypes, we use the Mann-Whitney nonparametric hypothesis test for two-class phenotypes and the non-parametric Kruskal-Wallis test for multi-class categorical

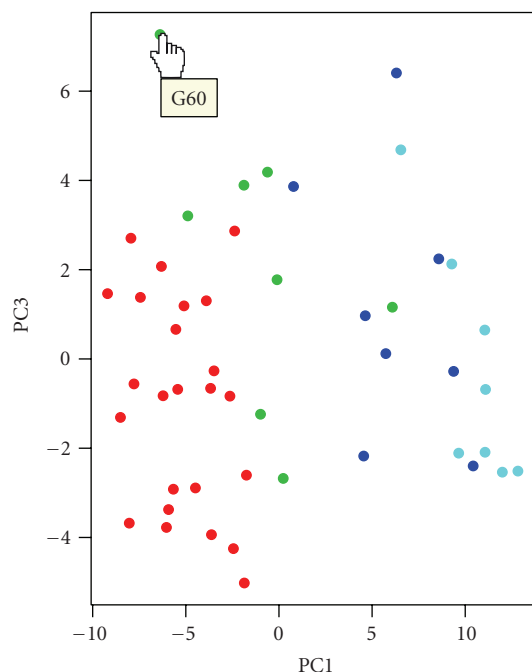


FIGURE 4: Examples for output generated by *metaP-server*: PCA plot colored by phenotype.

phenotypes. For visualizing potential association, the server creates boxplots for the metabolite concentrations depending on the classes of the respective phenotype. The calculated P -values are given within the boxplots. With respect to the problem of multiple testing, only those metabolite-phenotype associations are marked as significant that show a P -value below the significance level after Bonferroni correction. Thereby “*” denotes a significance level of 5% (after correction) and “**” denotes a significant level of 1% (after correction). If the user provides only two phenotypes with categorical values, *metaP-server* additionally performs hypothesis tests for the first phenotype depending on the different classes of the second phenotype and vice versa (see Section 5 and Figure 5). For each phenotype column containing numeric values, *metaP-server* tests the correlation of the metabolite concentrations with the phenotypes using the non-parametric Kendall method. The resulting correlation coefficients are visualized in a heatmap showing negative association in red and positive correlations in green. P -values and correlation coefficients are provided for download in table format (semicolon separated values).

3.2.4. Ratios. If the user has chosen the respective option, *metaP-server* calculates all-against-all metabolite concentration ratios (with logarithmic scaling). In this case, the server automatically tests for associations between all ratios and the phenotypes as described above. Using ratios instead of single metabolites can bring up new associations if the underlying metabolites are, for example, closely linked by occurring in the same pathway [5, 8].

3.2.5. Mapping Metabolites on KEGG Pathways. In the sample barplots, the metabolite quantities measured for a sample are shown relative (up/down) to the metabolites’ mean derived from the complete uploaded data set. These relative concentrations can be mapped onto KEGG pathway maps by coloring the corresponding KEGG compounds red in case of metabolites with low concentration and green in case of metabolites with high concentration relative to their mean values.

3.2.6. Cross-References for Kit Metabolites. For AbsoluteIDQ data, detailed information on the kit metabolites including cross references to HMDB, LipidMaps, KEGG, PubChem, and CAS numbers are provided.

3.3. Implementation. The web-server is mainly based on Perl CGI scripts. For statistical analyses, we rely on the open source R-project (<http://www.R-project.org>). For coloring metabolites on KEGG pathway maps, we use the forms provided by KEGG.

4. Interpreting the Results

In order to illustrate how the results of *metaP-server* server analysis can be interpreted, we provide two walk-through examples from typical applications. (i) LC-MS/MS data (raw area counts) from a drug dosing study in liver tissue (Metabolon Inc., 2006). In this case, the phenotypes “(drug) dose”, “day”, “group”, and “weight” are provided. (ii) Human, mouse, and bovine plasma samples are measured using AbsoluteIDQ. Users can easily upload the example data files (via hyperlinks) onto the job submission page and rerun the examples at any time.

After completion of the processing, there are several starting points for exploring the data. The walk-through examples contain detailed descriptions of these starting points and of the analysis results produced by *metaP-server*. Here, we only highlight a few specific possibilities how the server can be used for data exploration in the first example. The user can, for instance, immediately check whether the concentrations of a specific metabolite shows the expected difference between the control group and the treated groups. For this purpose, the user can click on that metabolite in the metabolite overview and color the appearing barplot by the phenotype “groups” (Figure 2). Analogously, using PCA as a starting point, the user can check whether the grouping of samples in the PCA reflects the grouping given by the various uploaded phenotypes such as “(drug) dose” or “day” (Figure 4). The user can also pick a representative sample for a specific group and check immediately on the sample page which of the metabolites are up or down compared to other samples (Figure 3). These relative concentrations can also be automatically mapped onto KEGG pathway maps. The user can then screen the colored maps for situations where all metabolites downstream of a certain metabolite are red whereas the metabolites upstream are green. Hypothesis tests provided by the *metaP-server* server are another starting point for data exploration. In particular, tests performed for

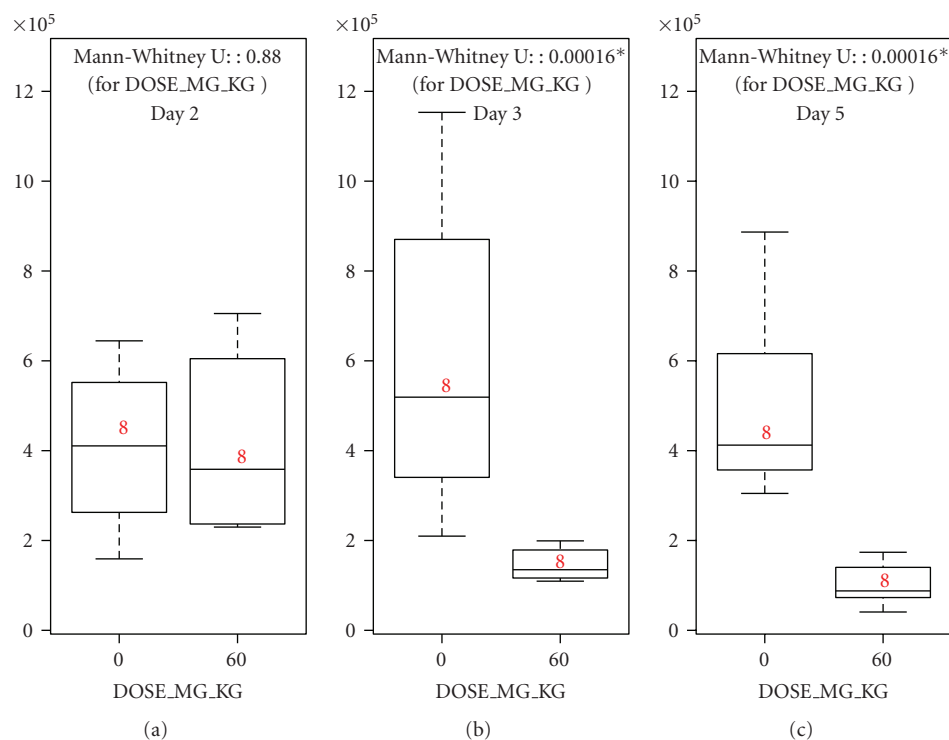


FIGURE 5: Examples for output generated by *metaP-server*: boxplots produced for groupwise hypothesis tests applied to the data set from the walk-through example provided with *metaP-server*; the association between ophthalmate and drug dose is tested for each day of treatment separately producing three separated box plots.

a phenotype separately for the different classes of a second phenotype can easily highlight effects that are otherwise only seen by using more sophisticated statistical methods. In the drug testing example, for instance, the metabolite ophthalmate is not significantly associated with the phenotype day when hypothesis testing is performed on complete data. However, in case of separated analysis as shown in Figure 5 the dependency of ophthalmate concentrations on the day of treatment for the group taking the drug becomes apparent. The concentration of ophthalmate is significantly increased at day three and day five for drug intake versus control whereas it is not significantly increased on day one.

5. Conclusion

The main objective of *metaP-server* is responding to the raising need for interpretation of high-throughput metabolomics data with respect to multiple sample phenotypes on an easy-to-use web-server-based platform, especially in the context of identifying metabolic biomarker for drug testing, therapy and diagnosis, and in epidemiological and metabolome wide association studies. *metaP-server* is mainly adapted to quantitative metabolomics data from kits and commercial platforms, such as Biocrates, Chenomx, Metabolon, and Phenomenome, but may also be used with any other metabolomics data set that is available in tabular format. The server has been developed in close cooperation with experimentalists and, as a result, focuses more on interactive and intuitive data exploration in the context of

multiple multiclass phenotypes rather than on providing a large set of different statistical methods. Nonetheless, the spectrum of analysis tools implemented ranging from estimation of reproducibility and batch effects, hypothesis and correlation tests, PCA analysis, to pathway mapping still covers various types of typical approaches used in data analysis. Of particular interest for the community is probably the ability of *metaP-server* to directly analyze metabolite concentration ratios, which corresponds to a relatively high computational effort. Moreover, special emphasis has been put on the mapping of metabolite identifiers of the Biocrates AbsoluteIDQ kit to the major metabolomics databases. As the metabolomics community and the number of kits will increase, we intend to implement new features in close cooperation with the relevant parties.

Acknowledgments

The authors thank Jerzy Adamski and Cornelia Prehn from the Genome Analysis Center (GAC) at the Helmholtz Zentrum München and John Ryals and Mike Milburn from Metabolon Inc. (Durham, NC, USA) for providing metabolomics data. They also thank Jerzy Adamski, Julia Henrichs, Anna Artati, Katja Vohouk, Cornelia Prehn, Susanne Neschen, and Melanie Kahle for intensive testing the server and for many valuable suggestions and comments. This work was supported in part by a Grant from the German Federal Ministry of Education and Research (BMBF) to the German Center Diabetes Research (DZD e.V.), by BMBF

Grant no. 03IS2061B (project Gani_Med), by BMBF Grant no. 0315494A (project SysMBo), and by Era-Net grant no. 0315442A (project PathoGenoMics). G. Kastenmüller and W. Römisch-Margl contributed equally to this work.

References

- [1] J. K. Nicholson, J. Connelly, J. C. Lindon, and E. Holmes, "Metabonomics: a platform for studying drug toxicity and gene function," *Nature Reviews Drug Discovery*, vol. 1, no. 2, pp. 153–161, 2002.
- [2] O. Fiehn, "Metabolomics—the link between genotypes and phenotypes," *Plant Molecular Biology*, vol. 48, no. 1–2, pp. 155–171, 2002.
- [3] R. J. Bino, R. D. Hall, O. Fiehn et al., "Potential of metabolomics as a functional genomics tool," *Trends in Plant Science*, vol. 9, no. 9, pp. 418–425, 2004.
- [4] M. Y. Hirai, M. Yano, D. B. Goodenowe et al., "Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 27, pp. 10205–10210, 2004.
- [5] E. Altmaier, S. L. Ramsay, A. Graber, H.-W. Mewes, K. M. Weinberger, and K. Suhre, "Bioinformatics analysis of targeted metabolomics—uncovering old and new tales of diabetic mice under medication," *Endocrinology*, vol. 149, no. 7, pp. 3478–3489, 2008.
- [6] E. Altmaier, G. Kastenmüller, W. Römisch-Margl et al., "Variation in the human lipidome associated with coffee consumption as revealed by quantitative targeted metabolomics," *Molecular Nutrition and Food Research*, vol. 53, no. 11, pp. 1357–1365, 2009.
- [7] C. Gieger, L. Geistlinger, E. Altmaier, et al., "Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum," *PLoS Genet*, vol. 4, no. 11, Article ID e1000282, 2008.
- [8] T. Illig, C. Gieger, G. Zhai, et al., "A genome-wide perspective of genetic variation in human metabolism," *Nature Genetics*, vol. 42, no. 2, pp. 137–141, 2010.
- [9] A. Sreekumar, L. M. Poisson, T. M. Rajendiran et al., "Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression," *Nature*, vol. 457, no. 7231, pp. 910–914, 2009.
- [10] P. L. Wood, R. Mankidy, S. Ritchie, et al., "Circulating plasmalogen levels and Alzheimer disease assessment scale-cognitive scores in Alzheimer patients," *Journal of Psychiatry & Neuroscience*, vol. 35, no. 1, pp. 59–62, 2010.
- [11] J. Shearer, G. Duggan, A. Weljie, D. S. Hittel, D. H. Wasserman, and H. J. Vogel, "Metabolomic profiling of dietary-induced insulin resistance in the high fat-fed C57BL/6J mouse," *Diabetes, Obesity and Metabolism*, vol. 10, no. 10, pp. 950–958, 2008.
- [12] M. Katajamaa, J. Miettinen, and M. Orešič, "MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data," *Bioinformatics*, vol. 22, no. 5, pp. 634–636, 2006.
- [13] K. Suhre and P. Schmitt-Kopplin, "MassTRIX: mass translator into pathways," *Nucleic Acids Research*, vol. 36, Web Server issue, pp. W481–W484, 2008.
- [14] J. Xia, T. C. Bjorndahl, P. Tang, and D. S. Wishart, "MetaboMiner—semi-automated identification of metabolites from 2D NMR spectra of complex biofluids," *BMC Bioinformatics*, vol. 9, article 507, 2008.
- [15] J. Xia, N. Psychogios, N. Young, and D. S. Wishart, "MetaboAnalyst: a web server for metabolomic data analysis and interpretation," *Nucleic Acids Research*, vol. 37, Web Server issue, pp. W652–W660, 2009.
- [16] H. Neuweger, S. P. Albaum, M. Dondrup et al., "MeltDB: a software platform for the analysis and integration of metabolomics experiment data," *Bioinformatics*, vol. 24, no. 23, pp. 2726–2732, 2008.
- [17] D. S. Wishart, C. Knox, A. C. Guo et al., "HMDB: a knowledgebase for the human metabolome," *Nucleic Acids Research*, vol. 37, Database issue, pp. D603–D610, 2009.
- [18] M. Sud, E. Fahy, D. Cotter et al., "LMSD: LIPID MAPS structure database," *Nucleic Acids Research*, vol. 35, Database issue, pp. D527–D532, 2007.
- [19] M. Kanehisa, S. Goto, M. Furumichi, M. Tanabe, and M. Hirakawa, "KEGG for representation and analysis of molecular networks involving diseases and drugs," *Nucleic Acids Research*, vol. 38, Database issue, pp. D355–D360, 2010.
- [20] Y. Wang, J. Xiao, T. O. Suzek, J. Zhang, J. Wang, and S. H. Bryant, "PubChem: a public information system for analyzing bioactivities of small molecules," *Nucleic Acids Research*, vol. 37, Web Server issue, pp. W623–W633, 2009.

Methodology Report

Magnetic Resonance Microscopy Contribution to Interpret High-Resolution Magic Angle Spinning Metabolomic Data of Human Tumor Tissue

M. Carmen Martínez-Bisbal,^{1,2} Vicent Esteve,^{1,2} Beatriz Martínez-Granados,² and Bernardo Celda^{1,2}

¹Networking Research Center on Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), ISC-III, C/Doctor Moliner 50, 46100 Burjassot, Valencia, Spain

²Department of Physical Chemistry, University of Valencia, C/Doctor Moliner 50, 46100 Burjassot, Valencia, Spain

Correspondence should be addressed to M. Carmen Martínez-Bisbal, carmen.martinez-bisbal@uv.es

Received 20 April 2010; Revised 16 July 2010; Accepted 3 August 2010

Academic Editor: Mika Ala-Korpela

Copyright © 2011 M. Carmen Martínez-Bisbal et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

HRMAS NMR is considered a valuable technique to obtain detailed metabolic profile of unprocessed tissues. To properly interpret the HRMAS metabolomic results, detailed information of the actual state of the sample inside the rotor is needed. MRM (Magnetic Resonance Microscopy) was applied for obtaining structural and spatially localized metabolic information of the samples inside the HRMAS rotors. The tissue was observed stuck to the rotor wall under the effect of HRMAS spinning. MRM spectroscopy showed a transference of metabolites from the tissue to the medium. The sample shape and the metabolite transfer after HRMAS indicated that tissue had undergone alterations and it can not be strictly considered as intact. This must be considered when HRMAS is used for metabolic tissue characterization, and it is expected to be highly dependent on the manipulation of the sample. The localized spectroscopic information of MRM reveals the biochemical compartmentalization on tissue samples hidden in the HRMAS spectrum.

1. Introduction

NMR is a very valuable tool for the study of healthy or diseased brain given the possible translation of that knowledge to the clinical practice. Brain tumors are among the most studied pathologies by NMR in its different modalities. In vivo NMR spectroscopy (MRS) can support tumor presurgical diagnosis (tumor identification, border delineation and infiltration detection) as well as treatment monitoring [1–5]. In in vivo MRS the spatial and spectral resolution are limited. With the purpose to obtain a better metabolite identification and quantification, high-resolution NMR has been applied to the tumor tissue ex vivo. These studies were initially performed in solution, and required a previous chemical extraction with Perchloric acid or organic solvents [6, 7]. High-Resolution Magic Angle Spinning (HRMAS) NMR has become in recent years a valuable

technique to obtain detailed metabolic profiles of brain tumors [8–13]. This technique requires low tissue amount (ca. 20 mg) and minimum sample handling compared to chemical extracts. Probably because this limited processing of the sample [9–11], HRMAS is considered a metabolomic technique that provides “intact” tissue spectra. Nevertheless, some researchers have shown some reluctance to assume the integrity of the tissue due to the sample handling and measurement conditions. An issue of interest is the postmortem interval for tissue obtained from autopsies, or in the case of pieces from resection, the time that the sample remains in the surgery room before being preserved. The results so far obtained show that this time period between tissue death and tissue preservation causes changes in the tissue metabolite concentration [8, 14] as might be expected. Knowing this effect, the temperature to run the experiments needs to be low to preserve as much as

possible the biochemical profile. In fact, comparing the biochemical profiles of experiments at 20°C and 2°C, there were 4-fold reduction in the content of some metabolites at 20°C compared to 2°C [8], which indicates the tissue degradation at 20°C. However, it has also been found that for tissue maintained at 4°C and spinning at 5 kHz there were more significant changes in the metabolite concentrations over 4 hours of measurement than the changes due to the delayed initial freezing [14]. The cellular damage by high spinning speed of HRMAS seems to be promoting some release of metabolites that previously were NMR-invisible [14]. Finally, the effect of freezing-thawing the tissue samples has been of interest for some researchers [15, 16]. Freezing and thawing of any tissue is likely to cause unpredictable amounts of cell damage and lysis [16]. In some tissues there have been reported changes in the molecular composition of the samples frozen, compared to samples studied without previous freezing. According to these researchers, tissue samples that are subjected to freezing will undergo some physical disruption of cellular compartmentalization and changes in molecular composition [15]. Moreover, the tissues thawed after freezing show that a great amount of metabolites (40%–50% for human prostate and 20% for rat brain) leak into the storage buffer at the time that thawing is completed, and the amount of a given metabolite observed in the medium varies from one biopsy to another and is not proportional to the amount originally present in the tissue [16].

Therefore, a proper biochemical interpretation of the HRMAS results would require detailed and localized molecular information on the tissue that can be correlated with the histopathological analysis and other molecular characterization techniques. It would be advisable as well to know the actual state of the tissue inside the rotor and a direct visualization could be of help, providing, as additional information, the overall shape of the sample. This study shows the potential role of NMR microscopy, MRM (Magnetic Resonance Microscopy), in obtaining structural and spatially localized metabolic information of the tissue samples inside the HRMAS rotors and to study the real integrity of the tissue in HRMAS. With this objective MRM imaging, MRM spectroscopy, and HRMAS has been applied to the study of human Glioblastoma Multiforme (GBM) biopsies.

Gliomas are the most common primary brain tumors. GBM is the highest glioma grade and the most common malignant central nervous system tumor in adults [17, 18].

2. Materials and Methods

This study was reviewed and approved by the ethics committee of the Hospital de la Ribera, Alzira, Spain. Informed consent was obtained from all patients as a part of eTUMOUR Project (FP6-2002-LSH 503094). Eight GBM samples (A–H) were studied by MRM and HRMAS inside 4 mm/50 μ L ZrO₂ rotors. Sample preparation was done following already published procedures [9]. The amount of tumoral tissue ranged from 9.8 to 27.3 mg (see Table 1). This quantity was

split from the whole tumoral mass submerged in liquid nitrogen. Samples were introduced in the rotor without flushing, thus avoiding the leakage of metabolites reported in the washing of tissues [16]. The volume of the rotor was fitted with cool D₂O. Then, the rotor was transferred into the NMR probe precooled to 6°C. In both probes, MRM and HRMAS, the real temperature was internally assessed using a 100% MeOH sample in a 4 mm rotor in the same conditions than the samples.

Several settings were tested: MRM and HRMAS experiments were performed on two tissue samples extracted from the same tumor (samples A–E), MRM before HRMAS (sample F), and MRM after HRMAS (samples G and H). Moreover, in sample G the solution part was separated from the tissue at the end of the HRMAS and MRM studies and introduced in a 4 mm/12 μ L rotor to acquire an additional HRMAS CPMG spectrum. Details of the experiments performed on each sample are given in Table 1.

MRM studies were conducted in a 14 T magnet (Bruker Avance DRX 600 spectrometer operating at a ¹H frequency of 600.13 MHz) connected to a microimage console with a gradient system of 300 G/cm at 60 A current (Bruker Biospin). The instrument was equipped with a standard bore Micro 5 imaging probehead with an rf insert of 5 mm. A Bruker Cooling Unit Extreme controlled the temperature during the acquisition. HRMAS experiments were conducted in an 11.7 T magnet (Bruker Avance DRX 500 spectrometer operating at a ¹H frequency of 500.13 MHz). The instrument was equipped with a 4 mm triple resonance ¹H/³¹P/¹³C HRMAS probe. A Bruker cooling unit was used to control the temperature by cooling down the bearing air flowing into the probe. Sample temperature was 6°C for MRM and HRMAS experiments.

In the MRM study of each sample, a global shimming was performed first, including the whole content of the rotor. The tissue pieces were then localized inside the rotors using imaging multislice sequences RARE (Rapid Acquisition with Relaxation Enhancement) with TR/TE 2000/80 ms and MDEFT (Modified Driven-Equilibrium Fourier Transform) with TR/TE 5000/4.5 ms adapted to MRM. In these RARE and MDEFT images, the FOV was 10 mm and the slice thickness was 500 μ m. The matrix dimensions were 256 \times 256 elements, which gave an in-plane spatial resolution of 39 \times 39 μ m. MRM-localized spectroscopy was performed with PRESS (Point resolved spectroscopy) single voxel (TR/TE 2100/12 ms). Before acquiring each single voxel spectrum, a specific shimming was performed, shimming only the volume determined by each single voxel and optimizing the 1st-order shims. This shimming procedure was repeated in each location, tissue and solution, for each sample. VAPOR (Variable pulse power and optimized relaxation delays) was used for water suppression. The spectral width was 10 ppm/6000 Hz and the number of points was 4k. Depending on the volume selected, the number of repetitions was from 512 to 4096. Single-voxel PRESS spectroscopy was performed in the tissue, and the solution parts in samples A–G. Single-voxel in sample H was acquired only in the solution part, since the tissue piece was a thin slice stuck to the rotor wall.

TABLE 1: MRM and HRMAS experiments for each tumor and sample.

| GBM tumor | Sample ID | Weight (mg) MRM | Weight (mg) HRMAS | Medium | MRM single-voxel dimensions (mm) | HRMAS performance |
|-----------|-----------|-----------------|-------------------|-------------------------|----------------------------------|--|
| 1 | A | 20.6 | 21.5 | | 1 × 1 × 1 | |
| 2 | B | 16.7 | 17.3 | D ₂ O | 1 × 1 × 1 | Parallel: other piece from the same tumors |
| 3 | C | 15.2 | 20.0 | | 1 × 1 × 1 | |
| 4 | D | 9.8 | 22.7 | | 1.5 × 1.5 × 1.5 | |
| 5 | E | 13.0 | 22.1 | | 1 × 1.8 × 1.8 | |
| 5 | F | | 24.7 | PBS in D ₂ O | 1.2 × 1.2 × 1.2 | after MRM |
| 6 | G | | 13.9 | | 1 × 1 × 1 | before MRM |
| 4 | H | | 11.1 | | 1.5 × 1.5 × 1.5 | |

In HRMAS experiments samples were spun at 4000 Hz. The shimming was optimized checking the quality of the signals and the resolution of the alanine doublet, usually present in this kind of samples. HRMAS proton CPMG spectra (Carr-Purcell-Meiboom-Gill) was performed with a 2D CPMG sequence with presaturation in the water signal. 2D CPMG spectra were acquired with a spectral width of 12 ppm/6000 Hz and the number of points in F2 was 16k. The delay was fixed to 2 ms and the scan with 4 loops gave an effective TE of 16 ms. This was the nearest TE for comparison with MRM PRESS single-voxel TE 12 ms.

3. Results and Discussion

MRM images of the samples without previous HRMAS rotation (A–F) showed the tissue suspended in the solution and without any relevant deformation of the piece. Examples of these images (samples D and E) are shown in Figure 1. On the contrary, MRM images in samples G and H clearly showed the disposition of the tissue close to the rotor wall, due to the effect of high spinning speed on the previous HRMAS study (see MRM axial images in Figure 1). In the samples G and H, in which a previous HRMAS study was acquired, the solution part that remained in the central axis of the rotor (PBS in D₂O) contained a considerable amount of metabolites. This fact is clearly seen in the MRM single-voxel spectrum of sample G (Figure 2(g), middle spectrum); the main resonances were assigned to lactate, alanine, lipids, creatine, choline, and phosphocholine, glycine and mannitol (peaks are assigned in Figure 2(g), top spectrum) [9]. The single-voxel spectrum in the tissue of sample G (Figure 2(g), bottom spectrum) showed lower resolution than in solution (Figure 2(g) middle spectrum), but still could be observed lipids, creatine, choline, and mannitol resonances. In this later single voxel, the tissue location close to the rotor wall seriously precluded an adequate field homogeneity.

The HRMAS study of the separated solution part from sample G (after the HRMAS and MRM study) confirmed the metabolical content of the solution using the same technique (Figure 3(d)). The resulting metabolic profile of the solution was similar to the previous HRMAS CPMG spectrum in the sample and to the MRM single-voxel spectrum in the solution. The observation of this solution with a magnifying

glass (50x) assessed the absence of any tissue or cell in this solution.

The samples either with pure D₂O (A–C) or PBS in D₂O (D–F), but without previous rotation, also showed transference of metabolites from the tissue to the medium (Figure 2, middle spectra). A similar leakage of metabolites had been reported in previously frozen samples [16]. The metabolites appeared in the buffer solution (PBS in D₂O) immediately after thawing, and the amount of a given metabolite observed in the medium varied from one biopsy to another and was not proportional to the amount originally present in the tissue [16]. Figure 2 gathers the MRM single-voxel spectra from the tissue and the solution for all the samples in this study. In all the cases the single-voxel spectra in the solution showed the presence of metabolites.

There were substantial differences between the MRM single-voxel spectra in solution and tissue parts. The most striking differences were observed on the lipids, which almost preponderantly appeared in the tissue single-voxel spectra (it is very evident in samples A and C, see Figures 2(a) and 2(c), bottom and middle spectra). In fact, the lipids are commonly observed on GBM in vivo and ex vivo spectra and are usually related to their aggressiveness. Moreover, quantitative differences could be observed between the solution and the tissue parts in other metabolites: in sample B the choline derivatives proportion in the tissue and the solution was different, and in sample E the alanine was undetectable in the tissue but it was present in the solution part.

Noteworthy, in the samples here studied the global metabolic profiles observed either by MRM or by HRMAS showed some significant differences, which is in agreement with the heterogeneous nature of GBM tumor (areas of actively growing tumor, necrosis, border infiltration,...) [17]. The metabolite transfer from the tissue to the solution was also different, which agrees with previously published studies on tissues [16], despite being preserved and studied by the same procedures. An example of these differences can be observed in the samples C and D (Figures 2(c) and 2(d)). According to this, the percentage of the contribution of the solution and tissue spectra to the whole HRMAS spectrum might not be the same for each sample. In the sample C the lipid content is high (as seen in HRMAS and MRM tissue spectra) and a large part of these molecules remains

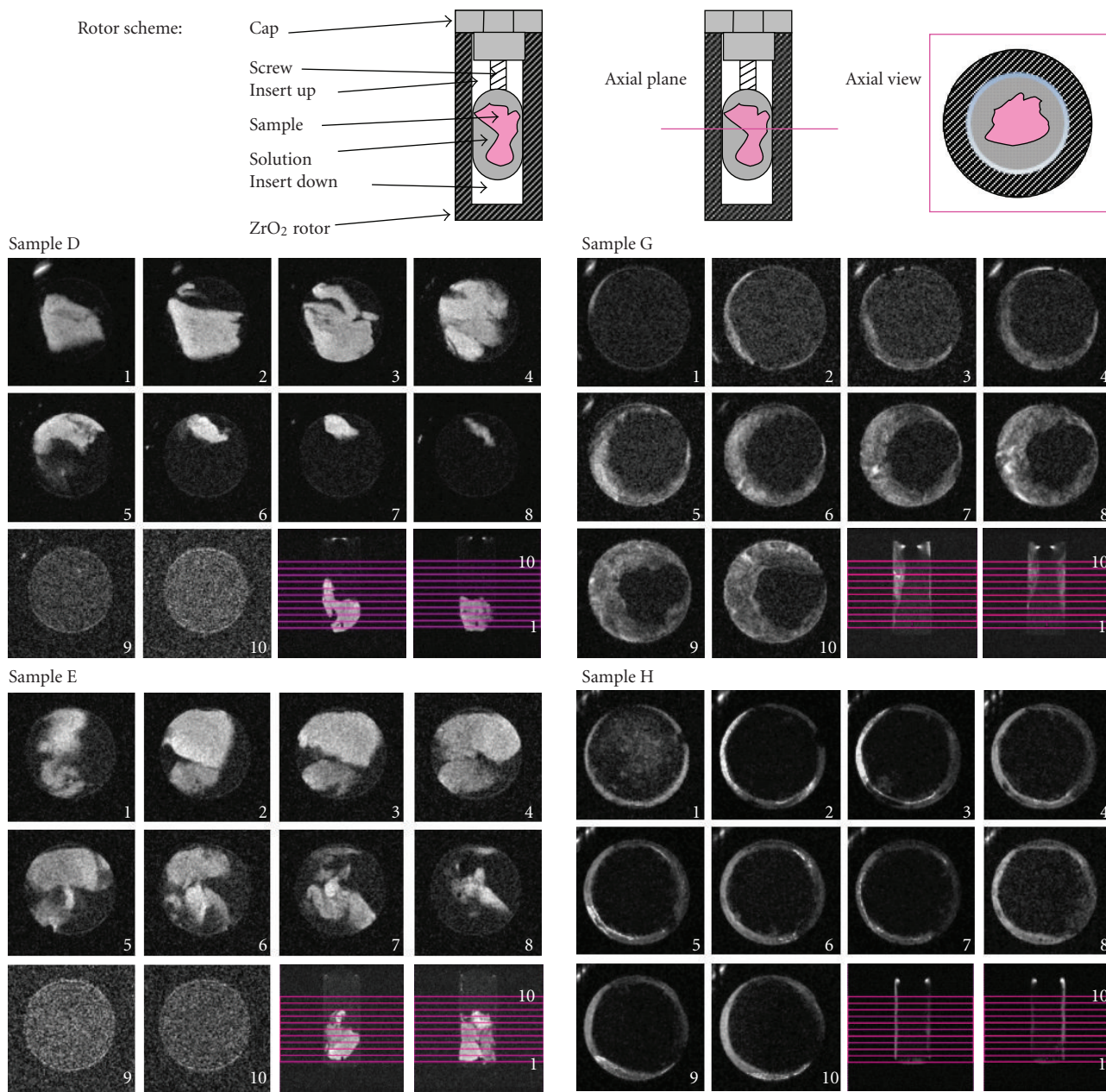


FIGURE 1: The upper part shows the scheme of a rotor containing the sample and the solution and the axial MRM images of four samples. The scheme of the HRMAS rotor shows the different parts with the sample and the solution in the inner cavity. The position of the axial plane is drawn with a magenta line and the axial view of the rotor on the right (upper part) shows the axial section, which is perpendicular to the longitudinal axis of the rotor. Below, the set of MRM images shows 10 RARE axial slices and the central sagittal and coronal planes for each of the samples showing the shape of tissue pieces inside the rotor for samples D, E, G, and H. The position of the 10 axial images is detailed by the 10 magenta horizontal lines in the sagittal and coronal planes in each sample. Axial images are zoomed to achieve the optimal view of the sample. In these MRM RARE images the solution, the rotor, and the air outside the rotor are darker than the tissue.

in the tissue whereas a small amount of low molecular weight metabolites are transferred to the liquid environment (Figure 2(c)). In this case, the main contribution to the final global spectrum will come from the tissue. On the contrary, sample D shows a low amount of lipids and hence, an important part of low molecular weight metabolites is observed in the solution, and the linear combination of

solution and tissue spectra will be different to the one for the sample C.

In the case of sample H, that underwent HRMAS previous to the MRM study, the effect of metabolite transference is such that even some lipids signals appear clearly in the MRM spectrum in the solution part (Figure 2(h), bottom spectrum). As it has been demonstrated in previous studies

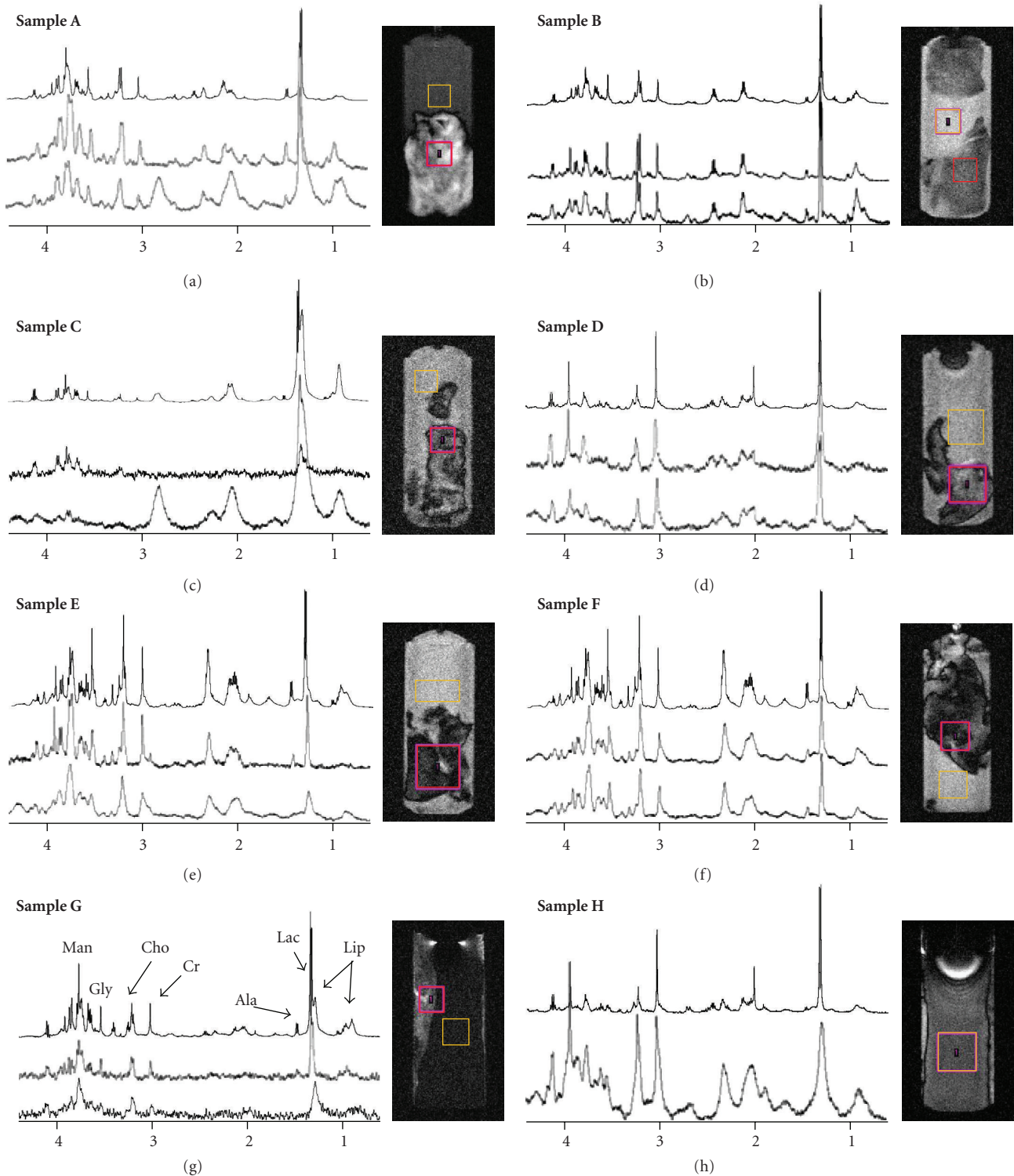


FIGURE 2: HRMAS CPMG (top) and MRM spectroscopy PRESS single-voxel spectra from solution (middle) and tissue (bottom) in each of the samples (A–H). MRM single-voxel location in tissue (red box) and solution (yellow box) is illustrated in the MRM images. Samples G and H were the only samples which underwent HRMAS experiments before MRM. Sample F underwent HRMAS study after MRM. The samples A–E underwent parallel study by HRMAS in other piece from the same biopsies. The resonances assigned in the CPMG HRMAS spectrum of sample G are: mannitol (Man), glycine (Gly), choline and phosphocholine (Cho), creatine (Cr), alanine (Ala), lactate (Lac), and lipids (Lip). For each sample, the MRM images shown are those that better illustrate the tissue shape and position inside the rotor. The MRM images shown for samples A and G belong to RARE sequences and the MRM images for samples B–F and H belong to MDEFT images. Moreover, the contrast in these MRM image was balanced to enhance the differences between the tissue and the solution part.

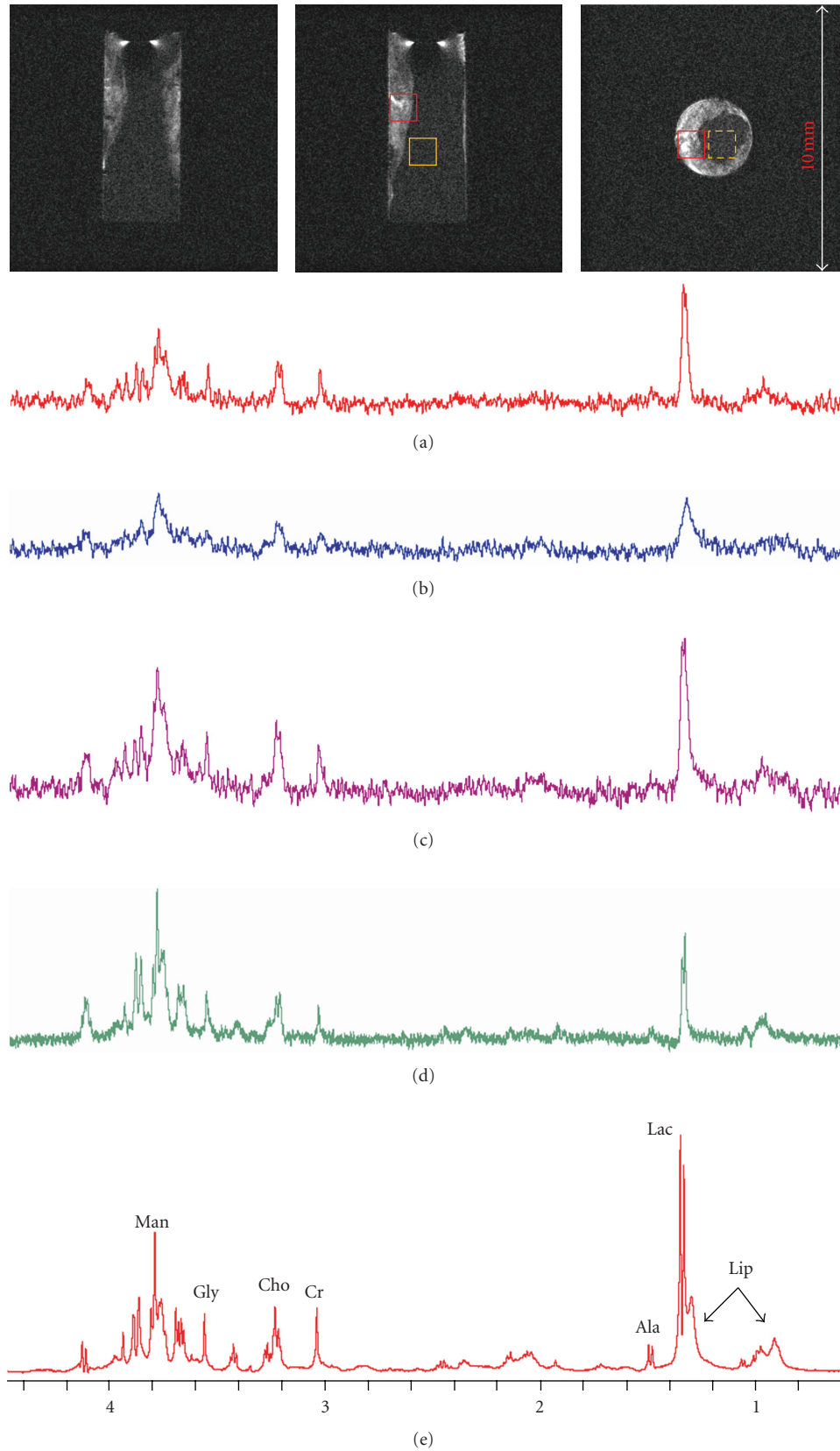


FIGURE 3: Three orthogonal planes allowed the location of single-voxel in the tissue part (red box) and in the solution (yellow box) for sample G. (a) MRM spectroscopy single-voxel spectra in liquid part, (b) MRM spectroscopy single-voxel spectra in tissue part, (c) the linear combination of spectra (a) and (b), (d) shows the HRMAS CPMG only on the solution collected after the MRM study, and (e) shows the HRMAS CPMG on the whole sample before the MRM study.

[14], the freezing-thawing and the ischemia processes are not the only ones that can promote changes in the spectra but also the high spinning speed of HRMAS study even at low temperatures. In this type of samples, as G and H, the expected effect is still higher since the biopsies underwent both freezing-thawing and HRMAS effects, as usually in the metabolomic studies involving HRMAS of biological tissues.

It is also remarkable the different peak width of these resonances coming from the tissue and the solution parts (Figure 2(e), bottom and middle, resp.), with two different contributions to the HRMAS overall spectrum. The effect is a broadening in the base of the peaks (Figure 2(e), top spectrum) that can be a drawback when automated analyses strategies are used. Like the linewidths, also the chemical shifts of metabolites within the tissue and outside in the solution may be different, and this may reduce the accuracy of quantification [16].

4. Conclusion

In conclusion, the MRM approach here presented allowed to visualize the great distortion of the brain tumor tissue produced by the HRMAS spinning and to detect a selective transference of metabolites from the tissue to the medium, even in nonstressing osmotic conditions. Our analysis implies that in some cases HRMAS spectra could correspond mainly to a solution and not so to an “intact” tissue. This effect must be accounted for in the subsequent quantification and interpretation of the results of HRMAS spectra in tissues, since the transference from the tissue to the solution is not equivalent for all the metabolites and for all the samples, and can also be influenced by the way in which the sample has been obtained and preserved. It might be expected a different contribution to the HRMAS spectrum of those metabolites transferred to the solution and those which preferentially remain in the tissue (in fact, the relaxation conditions are characteristically different in the tissue and in the solution, and therefore the shape of the peaks can experiment considerable variations). The MRM analysis provides a connection between the microscopic image and the metabolic composition, yielding spatially localized molecular information that can be relevant in the correlation with other molecular imaging techniques and in the understanding of their biochemical significance.

Acknowledgments

The authors are grateful for the financial support of Spanish Government SAF2007-6547 and VI Framework E.U. FP6-2002-LSH 503094' projects.

References

- [1] M. Castillo and L. Kwock, “Clinical applications of proton magnetic resonance spectroscopy in the evaluation of common intracranial tumors,” *Topics in Magnetic Resonance Imaging*, vol. 10, no. 2, pp. 104–113, 1999.
- [2] I. M. Burtscher and S. Holtås, “Proton magnetic resonance spectroscopy in brain tumours: clinical applications,” *Neuroradiology*, vol. 43, no. 5, pp. 345–352, 2001.
- [3] C. Majós, M. Julià-Sapé, J. Alonso et al., “Brain tumor classification by proton MR spectroscopy: comparison of diagnostic accuracy at short and long TE,” *American Journal of Neuroradiology*, vol. 25, no. 10, pp. 1696–1704, 2004.
- [4] A. Stadlbauer, S. Gruber, C. Nimsy et al., “Preoperative grading of gliomas by using metabolite quantification with high-spatial-resolution proton MR spectroscopic imaging,” *Radiology*, vol. 238, no. 3, pp. 958–969, 2006.
- [5] B. Celda, D. Monleón, M. C. Martínez-Bisbal et al., “MRS as endogenous molecular imaging for brain and prostate tumors: FP6 project “eTUMOR”,” *Advances in Experimental Medicine and Biology*, vol. 587, pp. 285–302, 2006.
- [6] J. Peeling and G. Sutherland, “High-resolution 1H NMR spectroscopy studies of extracts of human cerebral neoplasms,” *Magnetic Resonance in Medicine*, vol. 24, no. 1, pp. 123–136, 1992.
- [7] R. J. Maxwell, I. Martínez-Pérez, S. Cerdán et al., “Pattern recognition analysis of 1H NMR spectra from perchloric acid extracts of human brain tumor biopsies,” *Magnetic Resonance in Medicine*, vol. 39, no. 6, pp. 869–877, 1998.
- [8] L. L. Cheng, M. J. Ma, L. Becerra et al., “Quantitative neuropathology by high resolution magic angle spinning proton magnetic resonance spectroscopy,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 12, pp. 6408–6413, 1997.
- [9] M. C. Martínez-Bisbal, L. Martí-Bonmatí, J. Piquer et al., “1H and 13C HR-MAS spectroscopy of intact biopsy samples ex vivo and in vivo 1H MRS study of human high grade gliomas,” *NMR in Biomedicine*, vol. 17, no. 4, pp. 191–205, 2004.
- [10] K. S. Opstad, B. A. Bell, J. R. Griffiths, and F. A. Howe, “Toward accurate quantification of metabolites, lipids, and macromolecules in HRMAS spectra of human brain tumor biopsies using LCModel,” *Magnetic Resonance in Medicine*, vol. 60, no. 5, pp. 1237–1242, 2008.
- [11] T. E. Sjøbakk, R. Johansen, T. F. Bathen et al., “Characterization of brain metastases using high-resolution magic angle spinning MRS,” *NMR in Biomedicine*, vol. 21, no. 2, pp. 175–185, 2008.
- [12] M. C. Martínez-Bisbal, D. Monleón, O. Assemat et al., “Determination of metabolite concentrations in human brain tumour biopsy samples using HR-MAS and ERETIC measurements,” *NMR in Biomedicine*, vol. 22, no. 2, pp. 199–206, 2009.
- [13] M. Wilson, N. P. Davies, M.-A. Brundler, C. McConville, R. G. Grundy, and A. C. Peet, “High resolution magic angle spinning 1H NMR of childhood brain and nervous system tumours,” *Molecular Cancer*, vol. 8, article no. 6, 2009.
- [14] K. S. Opstad, B. A. Bell, J. R. Griffiths, and F. A. Howe, “An assessment of the effects of sample ischaemia and spinning time on the metabolic profile of brain tumour biopsy specimens as determined by high-resolution magic angle spinning (1)H NMR,” *NMR in Biomedicine*, vol. 21, no. 10, pp. 1138–1147, 2008.
- [15] D. A. Middleton, D. P. Bradley, S. C. Connor, P. G. Mullins, and D. G. Reid, “The effect of sample freezing on proton magic-angle spinning NMR spectra of biological tissue,” *Magnetic Resonance in Medicine*, vol. 40, no. 1, pp. 166–169, 1998.
- [16] R. Bourne, T. Dzendrowskyj, and C. Mountford, “Leakage of metabolites from tissue biopsies can result in large errors in quantitation by MRS,” *NMR in Biomedicine*, vol. 16, no. 2, pp. 96–101, 2003.

- [17] D. Russell and L. J. Rubinstein, *Russel and Rubinstein's Pathology of Tumors of the Nervous System*, Arnold, London, UK, 1988.
- [18] V. A. Levin, S. A. Leibel, and P. H. Gutin, *Cancer Principles and Practice of Oncology*, eddited by De Vita V. T. Jr., Devita Hellman S., Rosenberg S. A., Lippincott-Raven, Philadelphia, Pa, USA, 5th edition, 1997.

Research Article

Combined Reversed Phase HPLC, Mass Spectrometry, and NMR Spectroscopy for a Fast Separation and Efficient Identification of Phosphatidylcholines

Jan Willmann,¹ Herbert Thiele,² and Dieter Leibfritz¹

¹Institute of Organic Chemistry, University of Bremen, NW2C, 28359 Bremen, Germany

²Bruker Daltonik GmbH, Fahrenheitstr. 4, 28359 Bremen, Germany

Correspondence should be addressed to Dieter Leibfritz, dieter.leibfritz@uni-bremen.de

Received 14 May 2010; Accepted 21 July 2010

Academic Editor: Olav Kvalheim

Copyright © 2011 Jan Willmann et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In respect of the manifold involvement of lipids in biochemical processes, the analysis of intact and underivatized lipids of body fluids as well as cell and tissue extracts is still a challenging task, if detailed molecular information is required. Therefore, the advantage of combined use of high-pressure liquid chromatography (HPLC), mass spectrometry (MS), and nuclear magnetic resonance (NMR) spectroscopy will be shown analyzing three different types of extracts of the ubiquitous membrane component phosphatidylcholine. At first, different reversed phase modifications were tested on phosphatidylcholines (PC) with the same effective carbon number (ECN) for their applicability in lipid analysis. The results were taken to improve the separation of three natural PC extract types and a new reversed phase (RP)-HPLC method was developed. The individual species were characterized by one- and two-dimensional NMR and positive or negative ion mode quadrupole time of flight (q-TOF)-MS as well as MS/MS techniques. Furthermore, ion suppression effects during electrospray ionisation (ESI), difficulties, limits, and advantages of the individual analytical techniques are addressed.

1. Introduction

The analysis of native and underivatized lipids within body fluids as well as cell and tissue extracts is still a challenging task, in particular, if the molecular structure of individual components needs to be identified in decently short time. The lipid composition consists of different main classes such as fatty acids, neutral lipids, and lipids with positively or negatively charged head groups with manifold subclasses of structural diversity. Variations within the lipid composition were attributed to different pathologies such as neoplastic and neurodegenerative diseases, diabetes mellitus, and many others. Furthermore, some lipid classes are involved in cell death (apoptosis, necrosis), cellular signaling and are precursors for lysophospholipids (i.e., lysophosphatidylcholine), diacylglycerols, and phosphatic and arachidonic acid [1–25].

1,2-Diacyl-sn-glycero-3-phosphatidylcholine (PC) represents a major constituent of cell membranes. It consists of

the polar head group phosphorylcholine attached to the sn-3 position of glycerol and differing saturated and unsaturated fatty acids esterified to the sn-1 and sn-2 position, whereby fatty acids in position sn-1 are preferentially saturated as a rule. Numerous studies dealt with PCs in the past because of their utmost biochemical and clinical importance and many different analytical techniques have been proposed to get an insight into metabolic turnover or to characterize pathophysiological deviations of the native lipid composition. Most of these techniques suffer from various drawbacks as being time-consuming, insensitive, destructive, or not related to individual substructures. Gas chromatography (GC) [26–28], thin layer chromatography (TLC) [29–31], and high-performance liquid chromatography (HPLC) [32–37] are commonly used for lipid analysis. GC-based techniques are quantitative but require time-consuming sample preparation techniques. GC is often used in combination with TLC for the lipid class separation. The spots on a TLC plate are

scratched out and their fatty acid residues are analyzed upon derivatization into a volatile substrate and recorded by GC. However, the precise molecular structure of an individual lipid is lost because of the preceding hydrolysis of the lipids. Enzymatic cleavage of the ester bond using phospholipases allows a successive hydrolysis of the sn-2 and sn-1 fatty acid, but it is rather time-consuming due to intense laboratory work and already minor contamination of the enzyme leads to false results. HPLC offers the separation of lipid classes using the normal phase mode (NP) and additionally the separation according to the different fatty acid residues of an individual lipid in the reversed phase mode (RP). In this case, a successful separation depends distinctly on the appropriate selection of the stationary phase. Alternatively, MS-based techniques are widely used, as they are fast, sensitive and require only minor sample preparation [38]. The use of high resolution MS systems give access to the molecular formula. In addition, characteristic fragmentations identify the lipid class and molecular structure. When coupled with a HPLC-system their selectivity is much higher and benefits from both techniques. NMR spectroscopy is capable to measure intact biomaterials nondestructively without any preceding derivatization. Especially ^{31}P -NMR is well-suited to quantify phospholipid class analysis and needs only less sample preparation [39–44]. Again only minor information is obtained with respect to the fatty acid residues. ^1H -NMR measurements are also widely used, as they contain more information about the fatty acids in general, but the connection to the glycerol backbone is missing due to massive signal overlap. 2D-NMR involving the ^{13}C nucleus provides a lot more resolution and more information about individual species, but the low NMR sensitivity of the ^{13}C isotope prevents a fast and wide application of this technique in a routine analysis [44–48].

This paper presents an efficient RP-HPLC setup to separate phosphatidylcholines, which ultimately will be extendable to separate other polar phospholipids. Subsequently the HPLC tool is combined with the highly informative molecular assignment potentials of MS and NMR [49]. Five different types of silica-based reversed phase modifications were tested with respect to their capability to separate lipids containing fatty acids with an equivalent carbon number (ECN), which is the number of carbon atoms within a fatty acid chain minus twice the number of double bonds. The extension of a lipid by a C=C double bond will not change the hydrophobicity. The performance of all columns was tested on a mixture of five PCs with the same ECN whereas two of them are even constitutional isomers concerning the 1,2-positions of glycerol, which hampers the separation even more.

Then, the HPLC column with best performance was used to achieve an efficient baseline separation of three native PC extracts (soy bean, bovine brain, and egg yolk). Furthermore, the MS fragmentation behavior in the positive and negative ion mode is investigated for individual PCs to identify characteristic fragmentation patterns for this lipid class and its fatty acid residues. 1D and 2D high-resolution NMR spectra were also acquired to confirm the molecular structure.

2. Material and Methods

2.1. Chemicals. Methanol- d_4 and deuterated chloroform, methanol (LC-MS grade), all fatty acids, the test mixture compounds dipalmitoyl-phosphatidylcholine (DPPC), palmitoyl-oleoyl-phosphatidylcholine (POPC), oleoyl-palmitoyl-phosphatidylcholine (OPPC), dioleoyl-phosphatidylcholine (DOPC), stearoyl-linoleoyl-phosphatidylcholine (SLPC), and also the soy bean, bovine brain, and egg yolk extracts were purchased from Sigma-Aldrich Chemie GmbH (Taufkirchen, Germany). The double distilled water was taken from the in-house system.

2.2. High-Performance Liquid Chromatography. A HP 1100 series HPLC system (Agilent Technologies, Waldbronn, Germany) was used. The injection volume was $3\ \mu\text{L}$ of the standard prepared in methanol. Five columns with different stationary phases were tested with respect to their separation performance for lipid analysis:

- (1) type A silica-based endcapped C_{18} (Nucleosil 100-5 C_{18} , $250 \times 3\ \text{mm}$),
- (2) type A silica-based phenyl (Nucleosil 100-5 C_6H_5 , $250 \times 4\ \text{mm}$),
- (3) type B silica-based high density C_{18} (Nucleodur C_{18} Gravity, $5\ \mu\text{m}$, $250 \times 3\ \text{mm}$),
- (4) type B silica-based polymer/cross linked C_{18} (Nucleodur C_{18} Isis, $5\ \mu\text{m}$, $250 \times 3\ \text{mm}$),
- (5) type B silica-based mixed mode phenyl/ C_{18} (Nucleodur Sphinx RP, $5\ \mu\text{m}$, $250 \times 3\ \text{mm}$).

All HPLC columns and materials were a kind gift of Macherey-Nagel (Düren, Germany).

The 3 mm columns were operated at flow rate of 0.6 mL/min and the 4 mm column at 1 mL/min. The mobile phase was optimized by adapting the methanol content in different runs between 90% and 100% for the alkyl phases and between 80% and 100% for the phenyl phase with respect to the hydrophobic interaction of the analytes with the RP packing.

An 8 mm Nucleodur Sphinx RP was operated under isocratic conditions at 4.1 mL/min flow with a mobile phase consisting of methanol and water (90:10) for the semi preparative approach. To collect the individual species for NMR measurements a Gilson 215 liquid handler (Gilson International B.V., Bad Camberg, Germany) was used. The column temperature was kept at 40°C in all runs.

2.3. Mass Spectrometry. An esquire LC iontrap system (Bruker Daltonik GmbH, Bremen, Germany) was used for mass spectrometric detection for positive and negative ion mode mass and MS/MS spectra of each PC compound were recorded. The capillary voltage was set to $-3800\ \text{V}$ and the end plate offset to 500 V in positive ion mode. For the HPLC the nebulizer gas was set to 40 psi, dry gas and dry heat were set to 10 L/min and 300°C , respectively. In case of direct infusion via a syringe pump, the dry and nebulizer gases were reduced to 5 L/min and 5 psi, respectively. The collision

energy for MS/MS experiments was optimized with respect to the precursor ion stability.

A micrOTOF-Q equipped with the Apollo ESI ion source (Bruker Daltonik GmbH, Bremen, Germany) was used for precision mass detection. The capillary voltage was set to 4500 V and the end plate offset to -500 V in negative ion mode. The nebulizer gas was set to 0.4 bar, dry gas and dry heat were set to 4 L/min and 200°C , respectively. For MS/MS experiments the collision energy of the quadrupole was -42 eV/z. The molecular formula was generated by matching high mass accuracy and isotopic pattern (SigmaFit, Bruker Daltonik GmbH, Bremen, Germany).

2.4. Nuclear Magnetic Resonance. All samples were stored at -80°C before the measurements. In case of dissolved samples, the solvents were evaporated by a gentle stream of nitrogen and redissolved in $\text{CDCl}_3/\text{CD}_3\text{OD}$ (2:1). 1D (^1H , ^{13}C) and high-resolution 2D (HSQC, HSQC-TOCSY, HMBC) NMR spectra with a digital resolution of 1k data points in F1 and 4k data point in F2 dimension of each PC species were acquired on a Bruker DRX 600 MHz NMR spectrometer equipped with 5 mm TXI probe (Bruker BioSpin GmbH, Rheinstetten/Karlsruhe, Germany).

3. Results

3.1. High-Performance Liquid Chromatography. A comparison of five different reversed phase columns revealed the following behavior: The separation of the test mixture on type A silica-based materials showed only poor results for all PC compounds. Although, it seems that the Nucleosil material separates all peaks very well (see Table 1), the extreme peak broadening and a distinct tailing spoils the pretended peak separation. In contrast, the type B silica based materials separated DPPC, DOPC, SLPC, and POPC or OPPC very well. However, the two lipid isomers POPC and OPPC were only well separated (Table 1) on the polymer cross link RP packing (ISIS). With all RP materials it was possible to separate lipids containing two monounsaturated fatty acids from lipids with one or two saturated or one polyunsaturated fatty acid. The shortest separation times with sharp chromatographic peaks were achieved by the mixed mode stationary phase (Sphinx). Therefore, this stationary phase was selected to separate the individual compounds within the lipid extracts of natural sources.

3.1.1. Separation of the Phosphatidylcholine Extracts. The mixed mode stationary phase (Sphinx) and mobile phase of 90% methanol and 10% water allows the baseline separation of all species within the three different extracts (i.e., extraction residues). The separation of the PC species is not influenced by the type or contaminations of the extract. The results are listed in Table 2.

Altogether, twelve PC components and one plasmalogene (bovine brain extract) were identified. All species contained a saturated fatty acid in position sn-1 (i.e., myristic (1 species), palmitic (7 species) or stearic acid (4 species). Palmitic or stearic acid were esterified to the sn-2 position within three

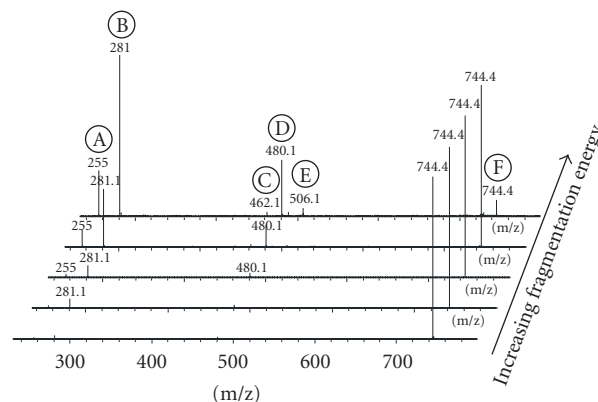


FIGURE 1: Variation of the fragmentation energy of negative ion mode ESI-MS² of POPC (744.4).

compounds. The other nine compounds contained a mono- (four times) or polyunsaturated fatty acid (five times) in position sn-2. The plasmalogen consists of octadecanol (sn-1) and oleic acid (sn-2). Most species were identified in bovine brain extract (11 + 1) and the fewest in the egg yolk extract (8).

3.1.2. Preparation of the Individual Compounds for NMR Measurement. The same mobile phase (90% methanol and 10% water) and column type (MN Sphinx) were used for the semipreparative HPLC runs as for the analytical measurements. However, the HPLC column dimension was upscaled (8 mm instead of 3 mm). The flow was splitted after the column and a small amount was used for peak detection and identification into the micro-TOF-Q. The residual eluent was collected for NMR measurements.

3.2. Mass Spectrometry. The ionization efficiencies of all equimolar concentrated saturated fatty acids were recorded relative to the internal standard undecanoic acid in different measurements using the negative ESI ion mode. The ion counts of undecanoic acid were set to 100 percent in all cases and the ion counts of all other fatty acids were recalculated with respect to this value (Table 3). All fatty acids with a shorter chain length as the internal standard show lower ionization efficiencies and all fatty acids with 12 or more carbon atoms show higher ionization efficiencies. The ionization efficiency of the fatty acids increases not linearly.

3.2.1. Fragmentation. The positive ion mode MS spectra showed better signal to noise ratios than in the negative ion mode. In positive ion mode, the base peak results from the adsorption of a sodium ion. In negative ion mode, the base peak results from the demethylation of the parent ion during the transfer into the ion trap mass spectrometer. $[\text{M} + \text{A}]^-$, whereby A is chloride or formate, was observed to a small extent only. In positive ion mode, MS/MS the polar headgroup of phosphorylcholine was cleaved off. In negative ion mode, MS/MS spectra the fatty acids of each

TABLE 1: Retention times of different PC-components with the same ECN on different reversed phase columns.

| Lipid | Column type | | | | |
|-------|---|---------------------|-----------------------|---------------------|--------------------|
| | Nucleosil 100-5 C ₆ H ₅ | Nucleosil 100-5 C18 | Nucleodur C18 Gravity | Nucleodur Sphinx RP | Nucleodur C18 Isis |
| DPPC | 1.46 (5.76)* | 56.82 | 78.00 | 14.31 | 49.19 |
| OPPC | 1.52 (6.49)* | 59.27 | 86.00 | 16.26 | 51.81 |
| POPC | 1.52 (6.48)* | 57.36 | 86.00 | 16.26 | 52.19 |
| DOPC | 1.64 (7.30)* | 57.18 | 98.67 | 19.11 | 56.40 |
| SLPC | 1.64 (7.58)* | 69.95 | 107.00 | 19.66 | 60.81 |

*mobile phase methanol/water (80 : 20).

TABLE 2: Identified PC compounds within extracts of bovine brain, egg yolk, and soy bean.

| Fatty acid position | | ECN | Bovine brain | Egg yolk | Soy bean |
|----------------------|-------------------------|-----|--------------|----------|----------|
| sn-1 position | sn-2 position | | | | |
| Myristic acid (14:0) | Palmitic acid (16:0) | 30 | + | - | + |
| Palmitic acid (16:0) | Palmitoleic acid (16:1) | 30 | + | + | + |
| Palmitic acid (16:0) | Linoleic acid (18:2) | 30 | + | + | + |
| Palmitic acid (16:0) | Linolenic acid (18:3) | 28 | - | + | - |
| Palmitic acid (16:0) | Arachidonic acid (20:4) | 28 | + | + | + |
| Palmitic acid (16:0) | Palmitic acid (16:0) | 32 | + | + | + |
| Palmitic acid (16:0) | Oleic acid (18:1) | 32 | + | + | + |
| Stearic acid (18:0) | Linoleic acid (18:2) | 32 | + | + | + |
| Stearic acid (18:0) | Arachidonic acid (20:4) | 30 | + | - | + |
| C18:0 Plasmalogen | Oleic acid (18:1) | 34 | + | - | - |
| Palmitic acid (16:0) | Stearic acid (18:0) | 34 | + | - | - |
| Stearic acid (18:0) | Oleic acid (18:1) | 34 | + | + | + |
| Stearic acid (18:0) | Eicosenoic acid (20:1) | 36 | + | - | - |

TABLE 3: Electrospray ionization efficiency of different fatty acid compared to undecanoic acid ($pK_s = 4,69$ for all acids).

| Fatty acid | Number of carbons | Empirical formula | Ionisation efficiency* (%) |
|-----------------|-------------------|--|----------------------------|
| caprylic acid | 8 | C ₈ H ₁₆ O ₂ | 52 |
| capric acid | 10 | C ₁₀ H ₂₀ O ₂ | 64 |
| undecanoic acid | 11 | C ₁₁ H ₂₂ O ₂ | 100 |
| lauric acid | 12 | C ₁₂ H ₂₄ O ₂ | 131 |
| myristic acid | 14 | C ₁₄ H ₂₈ O ₂ | 194 |
| palmitic acid | 16 | C ₁₆ H ₃₂ O ₂ | 285 |
| stearic acid | 18 | C ₁₈ H ₃₆ O ₂ | 609 |
| eicosanoic acid | 20 | C ₂₀ H ₄₀ O ₂ | 2300 |

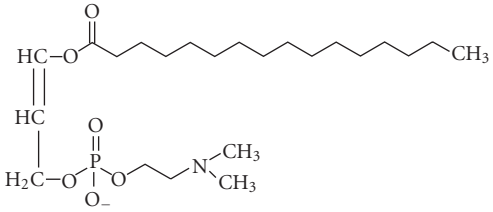
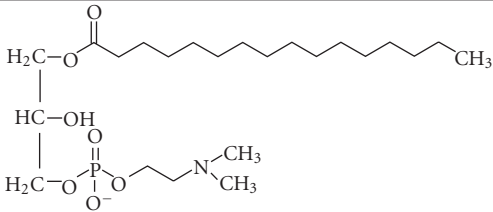
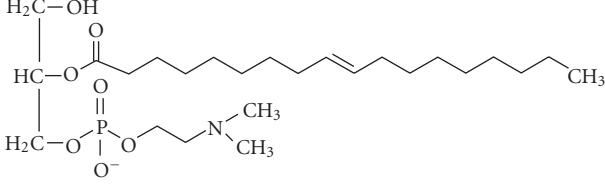
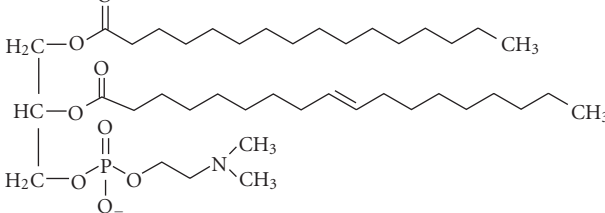
*compared to undecanoic acid.

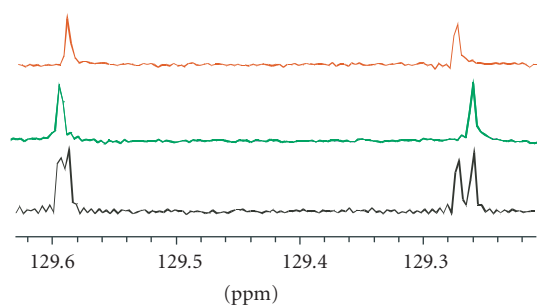
PC were assigned by detection of their $[M - H]^-$ ion accompanied by a neutral loss of the fatty acid ketene. The sn-2 fatty acid of every PC species shows always the more intense signal compared to the sn-1 fatty acid signal. Figure 1 shows several overlaid MS/MS iontrap spectra of POPC, but with increasing fragmentation energy to show the energy dependence of the different fragment ions. Table 4 lists the individual observed fragments and their corresponding m/z values.

3.2.2. *Identification of the Phosphatidylcholines in the Different Extracts.* Negative ion mode MS and auto-MS/MS q-TOF spectra were recorded during the HPLC runs. The fatty acid residues of each individual PC component was assigned by its relative fragment ion intensities. Furthermore, the molecular formula of each detected ion was generated by matching high mass accuracy and isotopic pattern to confirm the results. The lipid class was confirmed by the molecular formula and reconstructing the precursor ion of the fragment ions. Table 2 shows the identified species within the three different extracts.

3.3. *Nuclear Magnetic Resonance.* 1D-(¹H and ¹³C) and 2D-(HSQC, HSQC-TOCSY, and HMBC) NMR spectra of the five reference PCs were recorded for peak assignment. The NMR parameters of all reference compounds are listed in the supplements. Lipids with saturated (DPPC), mono unsaturated fatty acids (MOFA) as in POPC, OPPC, DOPC, or polyunsaturated fatty acids (PUFA) as in SLPC show zero, two or four carbon signals between 120 and 130 ppm. The MOFA and PUFA show unambiguously different chemical shifts for the olefinic carbons; however, the three lipids with MOFAs have nearly identical olefinic carbon shifts respectively the distance between the double bond signals depends on the MOFA location (sn-1 or sn-2). Figure 2

TABLE 4: Assignment of the individual observed fragments of POPC.

| Fragment | Assignment | m/z |
|----------|--|-------|
| A | palmitic acid $[M - H]^-$ | 255.0 |
| B | oleic acid $[M - H]^-$ | 281.0 |
| C |  | 462.1 |
| D |  | 480.1 |
| E |  | 506.1 |
| F |  | 744.4 |

FIGURE 2: Overlaid ^{13}C -NMR spectral sections of olefinic carbons of PCs with MOFAs (top OPPC, middle POPC, and bottom DOPC).

shows the overlaid ^{13}C NMR spectra of POPC, OPPC, and DOPC. The chemical shift difference between these two signals is about 50 Hz for POPC and about 47 Hz for OPPC. The corresponding ^1H -NMR spectra show identical signals for the olefinic protons (not shown). Nonetheless, a lipid with two MOFAs can be deduced from the intensity ratio of

the olefinic protons to the glycerol protons in the ^1H -NMR spectra.

3.3.1. NMR of the PC Extracts. HPLC-MS structural results were confirmed for all molecular species using 1D- (^1H) and 2D-(HSQC, HSQC-TOCSY, and HMBC) NMR techniques. The results are shown in Table 2. No branched chain or oxidized fatty acids were observed. The NMR parameter of unsaturated fatty acids have been measured separately (see Supplement) and are identical within the mixtures.

4. Discussion

Lipids with equal ECN should have the same hydrophobicity, which is the discriminating factor in reversed phase chromatography. This holds for PCs with the same ECN within a mixture and even more, if they are stereoisomers such as POPC and OPPC. The successful reversed phase separation of hydrophobic and zwitterionic molecules like phosphatidylcholines depends on very well endcapped silica materials, as Coulomb interactions of the choline group

with the silica gel lead to peak broadening, retention time shifts (data not shown), and peak tailing. Packing materials with hydrophobic van der Waals interactions only (i.e., high density C18 materials) show no separation of isomeric lipids. Additional interactions like steric effects by polymer cross-linked RP packings can overcome this problem. Stationary phases with only aromatic modifications are not hydrophobic enough to achieve a good phospholipid separation. The mixed mode stationary phase contains additionally alkyl ligands and offers therefore more hydrophobic interactions and good silica gel coverage. This column showed the shortest phospholipid retention times for all alkyl stationary phases and a good separation with narrow chromatographic peaks. Saturated lipids have the lowest and polyunsaturated lipids the highest retention times. Comparing lipids of the same ECN with two MOFAs or one PUFA and one saturated fatty acid the retention of the lipid with the PUFA is higher, as their spatial demand is larger and the π - π interactions with a PUFA residue is not affected by the other fatty acid.

4.1. Separation of the Phosphatidylcholine Extracts. By means of the HPLC separation it was possible to get semiquantitative information of the individual compounds in the extracts. Referring to the results of the separation of the lipid standard mixture it was possible to separate all species within the extracts. The elution order of lipids with the same ECN is in the same line as for the reference mixture (lipids with saturated fatty acids, one MOFA, two MOFAs, one PUFA). However, in case of PUFAs with four double bonds (i.e., arachidonic acid; 20:4) the π - π interactions become so strong that this compound elutes as last compound in the next higher ECN group (see Table 2). Nonetheless, they are unambiguously identified by means of their MS fragmentation or NMR spectra.

4.2. Preparation of the Individual Compounds for NMR Measurement. The upscaling with the same type of RP packing of the newest generation is no problem, although a peak broadening can occur because of the higher sample load.

4.3. Ion Suppression Effects. The ionization efficiency of the different fatty acids during ESI increases nonlinearly. The results can be correlated with the molecules hydrophobicity, which shows the same progression. The hydrophobicity is obtained from the octanol-water partitioning coefficient ($\log p_{ow}$) (data not shown). There are no differences between the pK_a values of the individual fatty acids (Table 3), so that the ionization efficiency of the fatty acids depends only on the molecules hydrophobicity, which increases by the number of carbon atoms.

As POPC and OPPC or SLPC and DOPC, respectively, have the same molecular weight/precision mass, they cannot be assigned based on their mol peak only. However, they are distinguished by MS/MS spectra, as the sn-2 fatty acid is always cleaved off at lower fragmentation energies due to sterical effects. In addition, the intensity ratio of the sn-2 fatty acid anion to the sn-1 fatty acid anion is constant

for a particular fragmentation energy. A semiquantitative or even quantitative analysis of stereoisomeric PCs within a mixture (f.e., POPC and OPPC) still is difficult, because the anion signal of an individual fatty acid will be the same regardless the position on the glycerol backbone and no other distinguishing signal is observed for one of the isomers. However, using the intensity ratio of the sn-2 fatty acid anion to the sn-1 fatty acid anion for the pure reference compound one can calculate the approximate POPC to OPPC ratio.

The MS detection limits also benefits from the isocratic HPLC method with high-organic solvent concentration in the mobile phase. In case of gradient HPLC, methods the ionization efficiency varies with the varying organic percentage of the mobile phase. Furthermore, the risk of ion suppression during the ionization process is minimized by the sample introduction after HPLC separation.

The identification of the individual species in the different extracts was achieved by MS/MS and comparing the individual fragment intensities as described earlier. The use of high-resolution spectra acquired by the micro-TOF-Q allows the generation of the compound's molecular formula by matching high mass accuracy and isotopic pattern. Abnormalities like oxidation of the double bond, and so forth, were not observed.

It may be noted, that not only diacyl PLs can be identified within a mixture, but also alkyl/alkenyl, acyl PLs because of the different fragmentation pattern of the fatty acid residue compared to an ether link.

The HPLC and MS results were confirmed by NMR spectroscopy, especially the configuration and location of double bonds in the fatty acid residues. Only NMR having the highest qualitative and quantitative structure elucidation potential allows a complete structure elucidation. 1H -, ^{13}C -, or ^{31}P -NMR spectra are capable to identify the various lipid classes. Beyond this, the degree of unsaturation is obtained from the proton signal intensity ratio of the double bond signal versus the choline group signal. MOFAs and PUFAs are differentiated by the number of carbon signals within the double bond region. The location (sn-1 or sn-2) of the MOFA follows from the ^{13}C -NMR spectra or from ESI-MS. The risk of peak overlapping in the NMR spectra was avoided by recording 2D-NMR spectra by preceding separation of individual phosphatidylcholines by HPLC.

5. Conclusion

The separation of lipids with equivalent chain lengths in complex mixtures can be improved using RP-HPLC packings of the newer generation containing additional discriminators. Already simple PC mixture cannot be assigned by a single analytical technique, while the combination of HPLC separation power, MS sensitivity with accurate mass measurement of molecular and fragment ions and NMR structure elucidation power will meet most suitably the challenge. They overcome the limits of any single technique and also proof the potential of their combination ultimately to analyze native (lipid) mixtures. The molecular structure of a novel compound may not be evaluated by NMR alone,

if the native concentration is very low. Conversely, MS data may give molecular weight, fragmentation and molecular formulae that may be insufficient to assign ambiguously the molecular structure of an unknown compound. However, online NMR and MS detections in parallel provide complementary data and minimize ambiguities between LC-MS and LC-NMR systems very efficiently.

Combination of these data allows the identification of a lipid class, reconstruction of the lipid structure, and both the location of an individual fatty acid to the sn-1 or sn-2 position in the glycerol moiety or the location of double bond within the fatty acid chains.

References

- [1] F. Hullin, M.-J. Bossant, and N. Salem Jr., "Aminophospholipid molecular species asymmetry in the human erythrocyte plasma membrane," *Biochimica et Biophysica Acta*, vol. 1061, no. 1, pp. 15–25, 1991.
- [2] T. Matsura, A. Togawa, M. Kai et al., "The presence of oxidized phosphatidylserine on Fas-mediated apoptotic cell surface," *Biochimica et Biophysica Acta*, vol. 1736, no. 3, pp. 181–188, 2005.
- [3] T. R. Pettitt, S. K. Dove, A. Lubben, S. D. J. Calaminus, and M. J. O. Wakelam, "Analysis of intact phosphoinositides in biological samples," *Journal of Lipid Research*, vol. 47, no. 7, pp. 1588–1596, 2006.
- [4] H.-P. Ma, C.-F. Chou, S.-P. Wei, and D. C. Eaton, "Regulation of the epithelial sodium channel by phosphatidylinositides: experiments, implications, and speculations," *Pflügers Archiv European Journal of Physiology*, vol. 455, no. 1, pp. 169–180, 2007.
- [5] M. J. Berridge, "Inositol trisphosphate and calcium signalling," *Nature*, vol. 361, no. 6410, pp. 315–325, 1993.
- [6] O. Kafrawy, M. Zerouga, W. Stillwell, and L. J. Janski, "Docosahexaenoic acid in phosphatidylcholine mediates cytotoxicity more effectively than other ω -3 and ω -6 fatty acids," *Cancer Letters*, vol. 132, no. 1-2, pp. 23–29, 1998.
- [7] I. Kan, E. Melamed, D. Offen, and P. Green, "Docosahexaenoic acid and arachidonic acid are fundamental supplements for the induction of neuronal differentiation," *Journal of Lipid Research*, vol. 48, no. 3, pp. 513–517, 2007.
- [8] T. Kolter and K. Sandhoff, "Sphingolipide—ihre stoffwechselwege und die pathobiochemie neurodegenerativer erkrankungen," *Angewandte Chemie*, vol. 111, pp. 1632–1670, 1999.
- [9] R. N. Kolesnick, "Sphingomyelin and derivatives as cellular signals," *Progress in Lipid Research*, vol. 30, no. 1, pp. 1–38, 1991.
- [10] M. Garnier, E. J. Dufourc, and B. Larijani, "Characterisation of lipids in cell signalling and membrane dynamics by nuclear magnetic resonance spectroscopy and mass spectrometry," *Signal Transduction*, vol. 6, no. 2, pp. 133–143, 2006.
- [11] J. H. Exton, "Phosphatidylcholine breakdown and signal transduction," *Biochimica et Biophysica Acta*, vol. 1212, no. 1, pp. 26–42, 1994.
- [12] C. D. Stubbs and A. D. Smith, "The modification of mammalian membrane polyunsaturated fatty acid composition in relation to membrane fluidity and function," *Biochimica et Biophysica Acta*, vol. 779, no. 1, pp. 89–137, 1984.
- [13] J. Adachi, N. Yoshioka, M. Sato, K. Nakagawa, Y. Yamamoto, and Y. Ueno, "Detection of phosphatidylcholine oxidation products in rat heart using quadrupole time-of-flight mass spectrometry," *Journal of Chromatography B*, vol. 823, no. 1, pp. 37–43, 2005.
- [14] F. M. Corrigan, D. F. Horrobin, E. R. Skinner, J. A. O. Besson, and M. B. Cooper, "Abnormal content of $n-6$ and $n-3$ long-chain unsaturated fatty acids in the phosphoglycerides and cholesterol esters of parahippocampal cortex from Alzheimer's disease patients and its relationship to acetyl CoA content," *International Journal of Biochemistry and Cell Biology*, vol. 30, no. 2, pp. 197–207, 1998.
- [15] Y. L. Ching, A. Lesimple, Å. Larsen, O. Mamer, and J. Genest, "ESI-MS quantitation of increased sphingomyelin in Niemann-Pick disease type B HDL," *Journal of Lipid Research*, vol. 46, no. 6, pp. 1213–1228, 2005.
- [16] J. Silva, S. Dasgupta, G. Wang, K. Krishnamurthy, E. Ritter, and E. Bieberich, "Lipids isolated from bone induce the migration of human breast cancer cells," *Journal of Lipid Research*, vol. 47, no. 4, pp. 724–733, 2006.
- [17] M. Okita, D. C. Gaudette, G. B. Mills, and B. J. Holub, "Elevated levels and altered fatty acid composition of plasma lysophosphatidylcholine (LysoPC) in ovarian cancer patients," *International Journal of Cancer*, vol. 71, no. 1, pp. 31–34, 1997.
- [18] E. Iorio, D. Mezzanzanica, P. Alberti et al., "Alterations of choline phospholipid metabolism in ovarian tumor progression," *Cancer Research*, vol. 65, no. 20, pp. 9369–9376, 2005.
- [19] P. Bounoux, V. Chajes, M. Lanson et al., "Prognostic significance of tumor phosphatidylcholine stearic acid level in breast carcinoma," *Breast Cancer Research and Treatment*, vol. 20, no. 3, pp. 185–194, 1992.
- [20] T. E. Merchant, J. N. Kasimos, T. Vroom et al., "Malignant breast tumor phospholipid profiles using ^{31}P magnetic resonance," *Cancer Letters*, vol. 176, no. 2, pp. 159–167, 2002.
- [21] C. Wang, H. Kong, Y. Guan et al., "Plasma phospholipid metabolic profiling and biomarkers of type 2 diabetes mellitus based on high-performance liquid chromatography/electrospray mass spectrometry and multivariate statistical analysis," *Analytical Chemistry*, vol. 77, no. 13, pp. 4108–4116, 2005.
- [22] Y. Wang, D. Botolin, J. Xu et al., "Regulation of hepatic fatty acid elongase and desaturase expression in diabetes and obesity," *Journal of Lipid Research*, vol. 47, no. 9, pp. 2028–2041, 2006.
- [23] J.-C. Hogue, B. Lamarche, A. J. Tremblay, J. Bergeron, C. Gagné, and P. Couture, "Evidence of increased secretion of apolipoprotein B-48-containing lipoproteins in subjects with type 2 diabetes," *Journal of Lipid Research*, vol. 48, no. 6, pp. 1336–1342, 2007.
- [24] S. J. Eun, Y. S. Hae, R. K. Mi et al., "Sphingosylphosphorylcholine induces proliferation of human adipose tissue-derived mesenchymal stem cells via activation of JNK," *Journal of Lipid Research*, vol. 47, no. 3, pp. 653–664, 2006.
- [25] J. Laurencikienė, V. Van Harmelen, E. A. Nordström et al., "NF- κ B is important for TNF- α -induced lipolysis in human adipocytes," *Journal of Lipid Research*, vol. 48, no. 5, pp. 1069–1077, 2007.
- [26] P. Q. Tranchida, P. Donato, G. Dugo, L. Mondello, and P. Dugo, "Comprehensive chromatographic methods for the analysis of lipids," *Trends in Analytical Chemistry*, vol. 26, no. 3, pp. 191–205, 2007.
- [27] S. Meier, S. A. Mjøs, H. Joensen, and O. Grahl-Nielsen, "Validation of a one-step extraction/methylation method for determination of fatty acids and cholesterol in marine tissues," *Journal of Chromatography A*, vol. 1104, no. 1-2, pp. 291–298, 2006.

- [28] M. P. Styczynski, J. F. Moxley, L. V. Tong, J. L. Walther, K. L. Jensen, and G. N. Stephanopoulos, "Systematic identification of conserved metabolites in GC/MS data for metabolomics and biomarker discovery," *Analytical Chemistry*, vol. 79, no. 3, pp. 966–973, 2007.
- [29] J. C. Touchstone, "Thin-layer chromatographic procedures for lipid separation," *Journal of Chromatography B*, vol. 671, no. 1-2, pp. 169–195, 1995.
- [30] J. K. Yao and G. M. Rastetter, "Microanalysis of complex tissue lipids by high-performance thin-layer chromatography," *Analytical Biochemistry*, vol. 150, no. 1, pp. 111–116, 1985.
- [31] J. J. Myher and A. Kuksis, "General strategies in chromatographic analysis of lipids," *Journal of Chromatography B*, vol. 671, no. 1-2, pp. 3–33, 1995.
- [32] C. Silversand and C. Haux, "Improved high-performance liquid chromatographic method for the separation and quantification of lipid classes: application to fish lipids," *Journal of Chromatography B*, vol. 703, no. 1-2, pp. 7–14, 1997.
- [33] L. L. Dugan, P. Demediuk, C. E. Pendley II, and L. A. Horrocks, "Separation of phospholipids by high-performance liquid chromatography: all major classes, including ethanolamine and choline plasmalogens, and most minor classes, including lysophosphatidylethanolamine," *Journal of Chromatography*, vol. 378, no. 2, pp. 317–327, 1986.
- [34] N. U. Olsson and N. Salem Jr., "Molecular species analysis of phospholipids," *Journal of Chromatography B*, vol. 692, no. 2, pp. 245–256, 1997.
- [35] J. Becart, C. Chevalier, and J. P. Biesse, "Quantitative analysis of phospholipids by HPLC with light scattering evaporating detector—application to raw materials for cosmetic use," *Journal of High Resolution Chromatography*, vol. 13, pp. 126–129, 1990.
- [36] H.-Y. Kim, T.-C. L. Wang, and Y.-C. Ma, "Liquid chromatography/mass spectrometry of phospholipids using electrospray ionization," *Analytical Chemistry*, vol. 66, no. 22, pp. 3977–3993, 1994.
- [37] P. J. Kalo, V. Ollilainen, J. M. Rocha, and F. X. Malcata, "Identification of molecular species of simple lipids by normal phase liquid chromatography-positive electrospray tandem mass spectrometry, and application of developed methods in comprehensive analysis of low erucic acid rapeseed oil lipids," *International Journal of Mass Spectrometry*, vol. 254, no. 1-2, pp. 106–121, 2006.
- [38] B. Brügger, G. Erben, R. Sandhoff, F. T. Wieland, and W. D. Lehmann, "Quantitative analysis of biological membrane lipids at the low picomole level by nano-electrospray ionization tandem mass spectrometry," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 6, pp. 2339–2344, 1997.
- [39] F. Süllentrop, D. Moka, S. Neubauer et al., "³¹P NMR spectroscopy of blood plasma: determination and quantification of phospholipid classes in patients with renal cell carcinoma," *NMR in Biomedicine*, vol. 15, no. 1, pp. 60–68, 2002.
- [40] J. M. Pearce and R. A. Komoroski, "Analysis of phospholipid molecular species in brain by ³¹P NMR spectroscopy," *Magnetic Resonance in Medicine*, vol. 44, no. 2, pp. 215–223, 2000.
- [41] J. Schille and K. Arnold, "Application of high resolution ³¹P NMR spectroscopy to the characterization of the phospholipid composition of tissues and body fluids—a methodological review," *Medical Science Monitor*, vol. 8, no. 11, pp. 205–222, 2002.
- [42] N. M. Loening, A. M. Chamberlin, A. G. Zepeda, R. G. Gonzalez, and L. L. Cheng, "Quantification of phosphocholine and glycerophosphocoline with ³¹P edited ¹H NMR spectroscopy," *NMR in Biomedicine*, vol. 18, no. 7, pp. 413–420, 2005.
- [43] J. Schiller, M. Müller, B. Fuchs, K. Arnold, and D. Huster, "³¹P NMR spectroscopy of phospholipids: from micelles to membranes," *Current Analytical Chemistry*, vol. 3, pp. 283–301, 2007.
- [44] D. Leibfritz, "An introduction to the potential of ¹H-, ³¹P- and ¹³C-NMR-spectroscopy," *Anticancer Research*, vol. 16, pp. 1317–1324, 1996.
- [45] J. Henke, J. Engelmann, B. Kutscher et al., "Changes of intracellular calcium, fatty acids and phospholipids during Miltefosine-induced apoptosis monitored by fluorescence- and ¹³C NMR-spectroscopy," *Anticancer Research*, vol. 19, no. 5, pp. 4027–4032, 1999.
- [46] J. Henke, W. Willker, J. Engelmann, and D. Leibfritz, "Combined extraction techniques of tumour cells and lipid-phospholipid assignment by two dimensional NMR spectroscopy," *Anticancer Research*, vol. 16, no. 3, pp. 1417–1427, 1996.
- [47] W. Willker and D. Leibfritz, "Assignment of mono- and polyunsaturated fatty acids in lipids of tissues and body fluids," *Magnetic Resonance in Chemistry*, vol. 36, no. 998, pp. S79–S84, 1998.
- [48] T. F. Bathen, J. Krane, T. Engan, K. S. Bjerve, and D. Axelson, "Quantification of plasma lipids and apolipoproteins by use of proton NMR spectroscopy, multivariate and neural network analysis," *NMR in Biomedicine*, vol. 13, no. 5, pp. 271–288, 2000.
- [49] J. Willmann, K. Mahlstedt, D. Leibfritz, M. Spraul, and H. Thiele, "Characterization of sphingomyelins in lipid extracts using a HPLC-MS-offline-NMR method," *Analytical Chemistry*, vol. 79, no. 11, pp. 4188–4191, 2007.

Research Article

Artificial Neural Networks for Classification in Metabolomic Studies of Whole Cells Using ^1H Nuclear Magnetic Resonance

D. F. Brougham,^{1,2} G. Ivanova,^{1,3} M. Gottschalk,¹ D. M. Collins,¹ A. J. Eustace,¹
R. O'Connor,^{1,4} and J. Havel⁵

¹National Institute for Cellular Biotechnology, Dublin City University, Dublin 9, Ireland

²School of Chemical Sciences, Dublin City University, Dublin 9, Ireland

³REQUIMTE, Department of Chemistry, Faculty of Sciences, University of Porto, 4169-007 Porto, Portugal

⁴School of Nursing, Dublin City University, Dublin 9, Ireland

⁵Department of Chemistry, Masaryk University, 611 37 Brno, Czech Republic

Correspondence should be addressed to D. F. Brougham, dermot.brougham@dcu.ie

Received 9 April 2010; Revised 14 June 2010; Accepted 23 July 2010

Academic Editor: Mika Ala-Korpela

Copyright © 2011 D. F. Brougham et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We report the successful classification, by artificial neural networks (ANNs), of ^1H NMR spectroscopic data recorded on whole-cell culture samples of four different lung carcinoma cell lines, which display different drug resistance patterns. The robustness of the approach was demonstrated by its ability to classify the cell line correctly in 100% of cases, despite the demonstrated presence of operator-induced sources of variation, and irrespective of which spectra are used for training and for validation. The study demonstrates the potential of ANN for lung carcinoma classification in realistic situations.

1. Introduction

Nuclear magnetic resonance spectroscopy (NMR, or MRS) has enormous potential for the study of biochemical and physiological changes in cancer tissues, due to its noninvasive nature and the large quantity of specific molecular information it can generate. Despite the sensitivity limitations of the technique, the inherent complexity of the spectra, and inevitable presence of overlapping resonances, there have been several successful NMR-metabonomics studies of cell tissue culture and culture extracts. The focus has been on elucidating the physiopathology of tumors and tumor cells, their drug toxicology and drug resistance, often with a view to identifying diagnostic markers [1–8]. A further significant complication in such studies arises from variability in the metabolite profile from sample to sample. This reflects many factors [9] including minor variations in growing conditions, the biochemical heterogeneity of the growing cells, the effect of different batches of sera (if used), and variations in cell and sample preparation. These additional factors may mask the

inherent metabolite distribution, which may be diagnostic of the pathophysiological state of interest.

Experimental complications and difficulties also compromise the extraction of critical information from *in vivo* MRS experiments. In this case, the problems arise from the use of different MR-protocols, which affect the quality of the water suppression, differences in echo time and in the baseline, and so forth. While the causes are different in origin, they have a similar effect on the application. For both forms of magnetic resonance, many of these issues can, in principle, be addressed by improved experimental design, however, it is common for additional sources of variance to be identifiable only after extensive experimentation. In addition to technical issues are the natural physiological variability and the individual treatment history of the subject. As a result, there is an ongoing requirement for the development of magnetic resonance-based diagnostics using advanced statistical-, or other data-, analysis techniques which can reduce or compensate for additional sources of variability.

^1H NMR spectra of intact tissues or whole-cell samples are inherently complex due to the large number of contributing species which results in significantly overlapping resonance signals. Cell membranes also produce magnetic field inhomogeneity, further broadening the spectra [10]. In the case of cancer cells, a significant proportion of the lipids reside in a fluid environment and hence appear in the liquid-state ^1H spectra as strong “mobile-lipid” resonances [7, 8, 11]. Although the identification of the major resonances in ^1H NMR spectra can be used to characterise the metabolite profile, the complexity of the data sets usually necessitates the use of data reduction and pattern recognition techniques. These can provide information on the biochemical and physiological changes in cancer tissues, related to their pathophysiology, drug toxicology, and drug resistance [12, 13]. Prominent amongst such techniques is principal component analysis (PCA), [14, 15] which involves diagonalisation of the spectral correlation or covariance matrix to identify independent sources of variance (principal components) across the set of spectra, and ranking of the components by their contribution to the overall variance. Thus, PCA is an unsupervised approach to data reprojection that can reveal the presence of classes, it has been applied to a variety of problems in biological science [16, 17].

Artificial Neural Networks (ANNs) belong to the so-called Artificial Intelligence group of methods, which were inspired by neurobiology and by the architecture of the human brain [18]. In recent times, these approaches have found applications in many branches of science. For example, they have been used in chemotaxonomy to classify limpets [19] from HPLC mass spectrometric data and in the identification of insect species from morphological measurements [20]. ANNs can be used to model data where the relations, or functions, are not known.

There have been some reports of the use of artificial intelligence and network methods in medical diagnostics which have involved analysis of magnetic resonance spectroscopic data. El-Deredy et al. [21] used ANNs to achieve reasonable prediction of the measured *in vitro* chemotherapeutic response from ^1H NMR of glioma biopsy extracts. More recently, Suna et al. [22] demonstrated the diagnostic potential of unsupervised approaches to classification by successfully analysing simulated ^1H NMR spectra using self-organising maps. This approach allowed the identification of stages along a metabolic pathway ranging from “normolipidaemic” to “metabolic syndrome”. Tate and coworkers [23] reported the trial of an automated decision support system for classification of brain tumors from *in vivo* MRS, which showed a small but significant improvement in diagnostic accuracy over spectroscopy used and interpreted on its own.

In recent work [24], we reported PCA of ^1H NMR spectra recorded for a group of human lung carcinoma cell lines in culture and ^1H NMR analysis of extracts from the same samples. The samples studied were cells of lung tumor origin with differing chemotherapy drug resistance patterns. For whole-cell samples, it was found that the statistically significant causes of spectral variation were an increase in the choline and a decrease in the methylene and mobile lipid ^1H resonance intensities, which were correlated with

our knowledge of the level of resistance displayed by the different cell lines. In this paper, we investigate the use of artificial neural network (ANN), a supervised method, to classify lung carcinoma. Two sets of whole-cell ^1H NMR spectra will be examined. These were recorded for two groups of human lung carcinoma cell lines, these were grown in culture and characterised over two different periods by two different groups of researchers (each consisting of a biologist and a spectroscopist), who both adhered to the same experimental protocol and used the same spectrometer. The cell lines studied include (i) the parent cell line DLKP, a human squamous non-small cell lung carcinoma; (ii) DLKP-A; (iii) DLKP-A5F, two resistant daughter lines; (iv) A549, a human lung adenocarcinoma cell line. The study also examines the capability of supervised techniques to compensate for experimental sources of variance, which may include operator bias and the cell culture growth process and in particular provide a test case for the application of ANN architectures in the identification and monitoring of resistance states in cancer tissue by MRS.

2. Experimental

2.1. Cell Samples. The cell lines DLKP [25, 26], DLKP-A [27], DLKP-A5F [28], and A549 were grown in culture to approximately 70–80% confluency in 175 cm² tissue culture flasks. Culture conditions were as follows: DLKP, DLKP-A, and DLKP-A5F and were cultured in minimal essential medium/Hams F12 (1 : 1, v/v) supplemented with 5% fetal calf serum and 2 mM L-glutamine. A549 was cultured in Dulbecco’s modified Eagle’s medium/Hams F12 (1 : 1, v/v) supplemented with 5% fetal calf serum. Cells were cultured as monolayers in tissue culture flasks and incubated at 37°C. A cell count was performed and c. 5×10^7 cells were separated and pelleted. These were then resuspended in deuterated PBS buffer and were kept in a container at 37°C before the start of the NMR measurements. The methods used were described in detail previously [24]. DLKP cells express a small amount of the multidrug resistance protein-1 (MRP-1) MDR drug efflux pump [25, 26]. DLKP-A [27] is a highly resistant clone of DLKP, which overexpresses the P-gp drug efflux pump. DLKP-A5F [28] was derived from DLKP by a different drug exposure profile, it is also highly drug resistant. A549 is an unrelated human lung adenocarcinoma cell line which was obtained from the American Type Culture Collection.

The first group of 13 samples, G1_13_21, were grown by a biologist during a six-month period, they were analysed by a first NMR spectroscopist. G1_13_21 contained 21 spectra and so was relatively sparse, it comprised DLKP [4 samples, 6 spectra], DLKP-A [4, 6], DLKP-A5F [3, 5], and A549 [2, 4]. The second group of 17 samples, G2_17_33, was grown independently, by a second biologist during a later six-month period and was analysed by a second spectroscopist [24]. G2_17_33 contained 33 spectra, it comprised DLKP [3, 6], DLKP-A [5, 10], DLKP-A5F [5, 9], and A549 [4, 8]. Thus for the integrated study presented here, a total of 30 samples were prepared and 54 ^1H spectra was recorded. The same protocols and methods were used by all the researchers

for cell growth and NMR spectroscopy. The biologist and spectroscopist who produced G1_13_21 will be collectively referred to as R1, and the biologist and spectroscopist who produced G2_17_33 will be referred to as R2. Due to the significant work involved in producing the large number of cells required for each spectrum, the number of samples in the study is inevitably somewhat limited. However, the total data set is larger than those usually reported in the analysis of NMR data by pattern recognition methods [16, 17, 29].

2.2. ^1H NMR Spectroscopy of Intact Cells. NMR spectra of the intact cell samples were recorded in deuterated PBS buffer on a Bruker DPX 400 spectrometer operating at 400.13 MHz for ^1H . Before all NMR experiments, the sample temperature was calibrated and controlled at $36.4 \pm 0.2^\circ\text{C}$ using an internal ethylene glycol thermometer (80% solution of ethane-1,2-diol in dimethyl sulfoxide- d_6). ^1H NMR spectra were acquired, without spinning, using WET [30] solvent suppression, with two Carr-Purcell-Meiboom-Gill (CPMG) echoes appended, using an echo delay of 1 ms [10]. Chemical shifts were referenced to an external 0.1% solution of sodium trimethylsilyl-[2,2,3,3- d_4]-propionate (TSP) in D_2O . All experiments were performed with a spectral width of 5200 Hz, an acquisition time of 3.15 s, and relaxation delay of 2 s. Three acquisition schemes were used to record the one-dimensional ^1H NMR spectra, all amounting to 128 scans. The first scheme (I) employed cycles of 16 dummy scans followed by four acquisition scans, $(16,4)_{32}$, giving an acquisition time of 3/4 hour. In the second scheme (II), 16 dummy scans were applied once prior to acquisition $16((0,16)_8)$, giving an acquisition time of 13 minutes. In the third scheme (III), 16 dummy scans and 128 acquisition scans were collected into 32 K data points, giving an acquisition time of 15 minutes. The time taken from resuspension to the start of data acquisition was typically less than 3/4 hour, and never more than 1 hour. All the data presented were recorded within 1 hour.

For the first group of 13 samples (G1_13_21) in the study, the acquisition schemes (I) and (II) were used for each sample. For the second group of 17 samples (G2_17_33), all three schemes were tested for each sample. Hence, the greater number of repeat spectra is for the second group. The inclusion of multiple spectra in the analysis from the same sample tests the stability of the samples over the time of the analysis. The insensitivity of the spectra to the sampling scheme used demonstrates that the samples do not change, for example, due to sedimentation, over the timescale that a single spectrum is acquired.

2.3. PCA Analysis. In the spectral region from 1.08 to 1.20 ppm, ethanol was observed, which was probably the result of endogenous processes. However, its intensity was highly variable, even within the same cell line, so this region was excluded from the analysis. The region containing the residual water resonance signal (3.56–6.05 ppm) was also excluded. The region above 6.05 ppm contained no features of sufficient intensity for reliable quantification, given the linewidth. For this study, we chose, as descriptors, the integrals over chemical shift regions (bins) of size 0.04 ppm

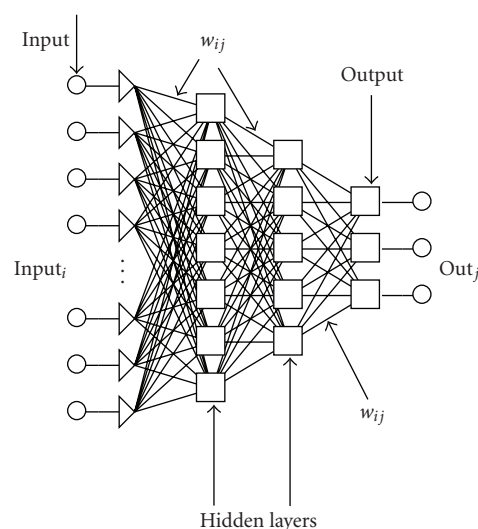


FIGURE 1: Schematic representation of a four-layer ANN architecture.

[12] which was found to produce the clearest separation of the cell types in the scores plots and the least noise in the corresponding loadings plots. Thus, the NMR spectra were reduced to 71 descriptors, with bin centres in the range 0.60–1.04, 1.24–3.56 ppm. We adopted the conventional approach [31] of normalisation relative to the total sum of the bin intensities in the region of interest. All the measures were implemented through writing an MATLAB (version 6.5.1, The Mathworks Inc.) code making use of the built in eigensolver.

2.4. ANN Analysis. ANNs are a sophisticated computational modelling tool, which can be used to solve a wide variety of complex problems. The attractiveness of ANNs comes from their capability to “learn” and/or model very complex systems and from the possibility of using them in classification. An ANN is a computational model formed from a certain number of single units, artificial neurons, or nodes, connected with coefficients (weights), w_{ij} , which constitute the neural structure. Many different neural network architectures can be used. One of the most common is the feed forward neural network of multilayer perceptions. The network is conventionally constructed with three or more layers, that is, input, output, and hidden layers, Figure 1.

Each layer has a different number of nodes. The input layer receives the information about the system (the nodes of this layer are simple distributive nodes, which do not alter the input value at all). The hidden layer processes the information initiated at the input, while the output layer is the observable response or behaviour. The inputs, $input_i$, multiplied by connection weights w_{ij} are first summed and then passed through a transfer function to produce the output, out_j . The determination of the appropriate number of hidden layers and number of hidden nodes in each layer is one of the most critical tasks in ANN design. Unlike the input and output layers, one starts with no prior knowledge of the number and size of hidden layers.

The use of ANN consists of two steps: “Training” and “Prediction”. The “Training” consists first of selecting input and output data for the network. This data is referred to as the training set. In the training phase, where actual data must be used, the optimum structure, weight coefficients and biases of the network are identified. Training is considered complete when the neural networks achieve the desired statistical accuracy, that is, when they produce the required outputs for a given sequence of inputs. A good criterion to find the correct network structure and therefore to stop the learning process is to minimise the root mean square (RMS) error as follows:

$$\text{RMS} = \sqrt{\frac{\sum_{i=1}^N \sum_{j=1}^M (y_{ij} - \text{out}_{ij})^2}{N \times M}}, \quad (1)$$

where y_{ij} is the element of the matrix ($N \times M$) for the training set or test set, and out_{ij} is the element of the output matrix ($N \times M$) of the neural network, where N is the number of variables in the pattern, and M is the number of samples. RMS gives a single number, which summarises the overall error.

After a supervised network performs well on training data, it is important to check its performance with data that has not been used in training. This process is called *verification*. This testing is critical to insure that the network has not simply memorised the training set but has learned the general patterns involved within an application. At this stage, other input data are submitted to the network in order to evaluate if it can predict the outputs. In this case, the outputs are already known, but they are not shown to the network. The predicted value is compared to the experimental one to see how well the network is performing. If the system does not give reasonable outputs for this test set, the training period is not over or the network is able to model the data but cannot predict them.

In this work, ANN was used as a supervised method where a training data set was created from the library of NMR spectra, and the lung carcinoma classification of this training data set was known. The backpropagation method was used throughout. Firstly, the optimal ANN architecture was searched for and when the correct classification in the training phase was obtained, the usefulness of the created database and the prediction power of the networks were validated using an independent verification set. For the ANN analysis, we used 72 inputs; the 71 binned NMR intensities and the identity of the pairs of researchers (R1 and R2) as numbers 1 and 2. For output 4, nominal values were used, these identify the four cell lines, DLKP, DLKPA, DLKP-A5F, and A549, for which there were 12, 16, 14, and 12 spectra, respectively. All calculations were performed using the software Trajan Neural Network Simulator, Release 3.0 D. (Trajan Software Ltd 1996–1998, UK), on a standard PC computer running Microsoft Windows Professional XP 2000.

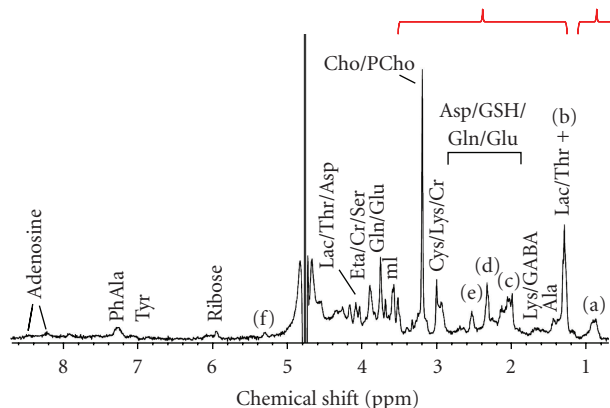


FIGURE 2: Typical 400 MHz ^1H NMR spectra of DLKP lung carcinoma whole cells. (a) CH_3 , (b) CH_2 , (c) $\text{CH}_2\text{CH}=\text{CH}$, (d) CH_2COO , (e) $=\text{CHCH}_2\text{CH}=\text{CH}$, (f) $\text{HC}=\text{CH}/\text{CHOCOR}$. The spectral regions used for statistical analysis (0.60–1.04 and 1.24–3.56 ppm) are indicated.

3. Results

3.1. ^1H NMR Spectroscopy of Whole Cells. A typical ^1H NMR spectrum of intact DLKP cells is shown in Figure 2. The appearance of the spectra and the assignment suggested below are broadly similar for all the cell samples analysed. A tentative assignment which is consistent with the literature [2, 4, 32, 33] is included in the figure [24]. Direct quantitative analysis of the whole-cell spectra is hampered by the potential multiple contributions from different metabolites to any given resonance line by the nonlorentzian lineshapes and by the broadness of the resonance lines. The resonances in the downfield region arise from species that are at low concentration, so quantification is precluded by the sensitivity limitations of the NMR measurement.

3.2. PCA Visualization of Whole-Cell Spectra. The binned NMR spectra of the intact cells were analysed using PCA. The scores plots are shown in Figure 3. Separation of the four cell types, within each of the two data sets, is apparent using the first two PCs, demonstrating that resistance type can be classified by PCA. It also demonstrates that the samples were stable over the course of the experiment and that the spectra are insensitive to the NMR sampling scheme. Loadings analysis shows that, for each data set, the spectral regions that contribute significantly to the first two principal components are from 1.24 to 1.50 ppm, corresponding to overlapped resonances from lipid methylenes and lactate methyls, and from 2.90 to 3.40 ppm, corresponding to overlapped resonances from N-methyl signals in the choline moieties of phosphatidylcholine, phosphocholine, and glycerophosphocholine. The contribution from other spectral regions to these two principal components is marginal.

Despite the fact that the same spectral regions allow separation within each data set, separation using PCA fails when the two sets of spectra are combined into one; see Supplementary Material available at doi:10.1155/2011/158094. It is apparent that, in addition to the metabolite differences

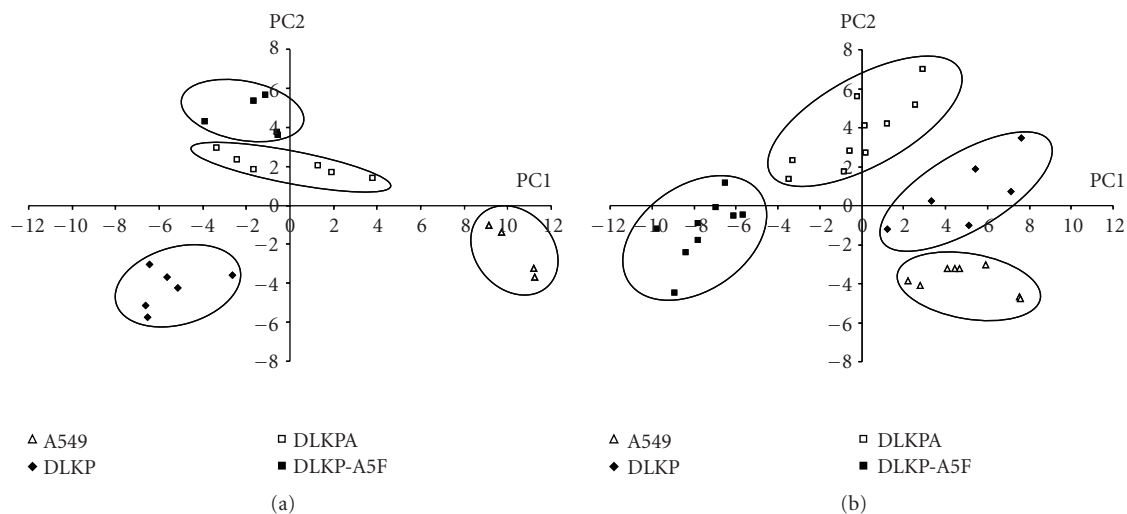


FIGURE 3: PCA scores plots for A549, DLKP, DLKPA, and DLKP-A5F, whole-cell data. Analysis is shown for G1_13_21 (a), G2_17_33 (b). The right hand panel is reproduced from [24] with permission.

of biological interest, there are subtle differences between G1_13_21 and G2_17_33 in the distribution of metabolites, which prevent classification of the entire (54 spectra) data set. The loadings analysis indicates contributions from across the spectral range, which may suggest variations in more than one metabolite. These spectral differences arise despite stringent efforts of the second group of researchers to adhere to the original experimental protocols and are reflected in the fact that there is not a simple correspondence between the orientation of the first two principal components between the two sets of spectra, Figure 3.

3.3. ANN Analysis of Whole-Cell Spectra. ANN analysis consists of separate training and verification steps. For this study, we adopted the strategy of choosing multiple verification sets of spectra at random from the 54 spectra available. In training, the first aim is to find an optimal ANN architecture to enable classification of the training data set. Several architectures of three up to four layered structures were examined for this purpose.

3.4. 3-Layers Architecture. Initially we adopted the simplest 3 layers architecture, in which case the search of the optimal architecture consists of optimising the number of nodes in the single hidden layer, effectively determining the corresponding weights, w_{ij} , to minimize the RMS (root mean square error) value according to (1). For our analysis, the RMS value ceases to decrease significantly above 5 to 6 nodes, Figure 4, we therefore used networks with 6 hidden nodes for verification. This optimal architecture will be labelled (72, 6, 4), with it we obtained an $RMS = 1.38 \times 10^{-3}$. Figure 4 illustrates the process of searching for the optimal network architecture.

In spite of the fact that very low values for the residual mean squares were achieved using the (72, 6, 4) architecture, the appropriateness of the architecture and of the training set

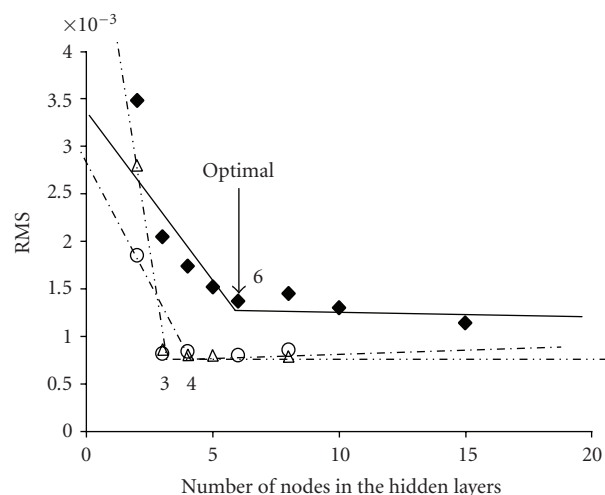


FIGURE 4: Plot of residual mean squares as a function of the number of nodes in the hidden layers, in the three-layer network (\blacklozenge), and in the second (\triangle) and third (\circ) layers of the four-layer network. For the networks labelled (\triangle), 3 nodes were used in the third layer; and for the networks labelled (\circ), 4 nodes were used in the second layer. The lines have no physical meaning; they are included to better illustrate the optimal number of nodes.

was then tested with various verification sets, that is, a “cross-validation” procedure was undertaken. Initially, five spectra were randomly chosen and excluded from the training set and used then as the verification set. From 10 combinations and 10 independent networks trained, in only two cases were any of the 5 spectra classified as unknown, Table 1. These results are encouraging; two cases represent $\sim 4\%$ of the total, so for (72, 6, 4) the classification was verified as 96% successful. The failures may have arisen due to an insufficient number of spectra in the training set or because

TABLE 1: Results of cross-validation verification process for the three- and four-layer ANN networks.

| Architecture (72, 6, 4)* | | |
|----------------------------|---------------------------------------|--------------------------------|
| Verification set no. | Spectra used in verification set | Results of Classification |
| 1 | 2, 13, 17, 27, 38 | all correct |
| 2 | 21, 24, 31, 35, 51 | all correct |
| 3 | 4, 12, 22, 35, 44 | spec. 35 classified as unknown |
| 4 | 16, 17, 22, 25, 52 | all correct |
| 5 | 15, 16, 17, 23, 54 | all correct |
| 6 | 9, 15, 20, 24, 43 | spec. 9 classified as unknown |
| 7 | 3, 12, 15, 25, 51 | all correct |
| 8 | 19, 21, 43, 47, 54 | all correct |
| 9 | 16, 36, 37, 47, 48 | all correct |
| 10 | 12, 42, 44, 48, 50 | all correct |
| Architecture (72, 4, 3, 4) | | |
| 1 | 5, 13, 20, 21, 22, 23, 24, 31, 51, 54 | all correct |
| 2 | 5, 8, 12, 15, 16, 29, 35, 36, 42, 49 | all correct |
| 3 | 8, 10, 13, 18, 23, 28, 33, 39, 40, 53 | all correct |
| 4 | 3, 5, 7, 9, 17, 27, 41, 45, 50, 52 | all correct |
| 5 | 5, 11, 16, 14, 20, 22, 24, 26, 44, 50 | all correct |

* where (72, 6, 4) refers to (the no. of inputs, the number of nodes in the hidden layer(s), the number of outputs).

networks with three layers have insufficient complexity for 100% prediction accuracy, in this case.

3.5. 4-Layers Architecture. We then examined networks with four layers (2 hidden). From several cases examined, it was found that four-layer ANN architectures performed similarly to simpler three layers architectures. Networks of the form (72, 4, 3, 4) or (72, 5, 4, 4) were investigated, note that the numbers in brackets refer to the number of inputs, the number of nodes in the first and in the second hidden layers, and the number of outputs. Acceptable RMS values, of 1.22×10^{-3} and 1.41×10^{-3} were obtained for (72, 4, 3, 4) and (72, 5, 4, 4), respectively, which are similar to the values obtained using the optimal three-layer architecture. Networks with the architecture (72, 4, 3, 4) performed very similarly to (72, 5, 4, 4) and require fewer unknowns (or weights, w_{ij}), 312 as opposed to 396. As a result, (72, 4, 3, 4) was found to converge faster and to be less sensitive to the number of spectra excluded from training to form the verification set. In fact, we found that 5 to 10 samples could be used for verification with 100% correct classification of the spectra, see Table 1. So in summary, the optimal 3- and 4-layer architectures were found to be (72, 6, 4) and (72, 4, 3, 4), respectively, Figure 5.

4. Discussion

The ^1H NMR spectra of intact cells for both G1_13_21 and G2_17_33 have similar general appearance with severe signal overlap and line broadening. Reprojection of either data set, using PCA, demonstrates that separation by cell types is possible due to systematic differences in the lipid methylene and lactate methyl resonances and the overlapped N-methyl ^1H nuclei of the choline-containing species [24].

Alterations in signal intensity and chemical shift from such cellular metabolites and biochemical intermediates have been described by other researchers in the area [6, 11]. However, because of the complex biochemical role played by these substances, we cannot ascribe a particular functional role to the findings, what is more the alterations appear to correlate and associate with particular phenotypic changes, for example, drug resistance. On the basis of the principal component analysis of either group, one could speculate that metabolite profiling by *in vivo* MRS has potential applications in monitoring the development of resistance in a given cancerous tissue. However, for the full data set such a possibility is effectively prevented by other influences on the metabolite distribution, which are comparable to, and nonorthogonal with, the “relevant” biochemical variation. We have shown that this significant obstacle can be eliminated, at least for *in vitro* studies of cell culture, by using a suitable ANN architecture. The most successful network was a four-layer structure with two hidden layers. After appropriate training, the (72, 4, 3, 4) architecture enabled 100% successful classification. Our approach may, in time, be expanded to the classification of larger data sets of spectra which have been recorded with less stringent control over sources of variance unrelated to the classification of interest. This result is encouraging and it is, to our knowledge, the first reported application of the use of ANNs specifically to correctly classify ^1H NMR spectra in a data set when additional “nonrelevant” sources of variance are included.

Other related examples of the combination of supervised and unsupervised methods include a report by Griffiths and coworkers [34], who obtained 85% accurate classification of meningiomas from nonmeningiomas, by initially using PCA to reduce the dimensionality of ^1H NMR spectra recorded for tumor biopsy extracts. The first thirty PCs

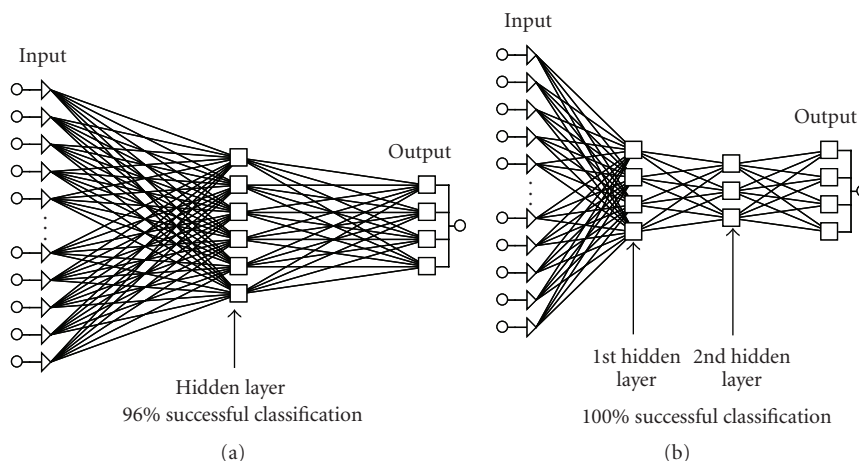


FIGURE 5: (a) Structure of the optimal 3-layer ANN architecture (72, 6, 4). (b) Structure of the optimal 4-layer ANN architecture (72, 3, 4).

from this first stage of analysis were then classified using a network. More recently, the performance of lineshape fitting and quantitative ANN analyses were compared by Hiltunen et al. [35] for both *in vivo* and simulated ^1H spectra. The good correlation obtained with these two approaches, for simulated data at least, suggested that ANNs have potential for quantification of *in vivo* MRS long echo time spectra. A further advantage of ANNs in the development of analysis methods for *in vivo* MRS is that they require less processing time than line fitting or other computational approaches [36]. Thus, our study adds to the growing number of applications of supervised techniques for exploiting the diagnostic potential of ^1H NMR spectra for biomedical purposes.

5. Conclusions

We have found that NMR data recorded for human lung carcinoma whole-cell culture samples can be used for analysis and classification. When sources of variation not directly related to the biological state of interest (drug resistance) are minimised or kept constant, visual separation of the cell type can be achieved using unsupervised pattern recognition techniques, such as PCA. On the other hand, when this condition is not met, in our case when different researchers were responsible for cell culture and spectroscopy, successful classification of the cell type could be achieved using artificial neural networks. The experimental and ANN methodology developed are a step towards the goal of robust and reliable diagnostics based on magnetic resonance spectral data. Furthermore, as similar experimental problems may be encountered in metabolomics applications using other spectroscopic techniques, biological classification using ANNs of data sets that include “nonbiological” sources of variance may be generally possible.

Acknowledgments

The authors acknowledge the support of the Higher Education Authority of Ireland, under the Programme for

Research in Third Level Institutions (PRTL13). D. Brougham, M. Gottschalk, and G. Ivanova acknowledge the financial support of the National Institute for Cellular Biotechnology, at DCU. They would like to thank the School of Chemical Sciences for its provision of spectrometer time. J. Havel would like to acknowledge the support of the EU Erasmus/Socrates exchange program between DCU and Masaryk University and to thank the Ministry of Education and Sports of the Czech Republic, Project LC 0635.

References

- [1] U. Sharma, A. Mehta, V. Seenu, and N. R. Jagannathan, “Biochemical characterization of metastatic lymph nodes of breast cancer patients by *in vitro* ^1H magnetic resonance spectroscopy: a pilot study,” *Magnetic Resonance Imaging*, vol. 22, no. 5, pp. 697–706, 2004.
- [2] B. Martínez-Granados, D. Monleón, M. C. Martínez-Bisbal et al., “Metabolite identification in human liver needle biopsies by high-resolution magic angle spinning ^1H NMR spectroscopy,” *NMR in Biomedicine*, vol. 19, no. 1, pp. 90–100, 2006.
- [3] F.-G. Lehnhardt, C. Bock, G. Röhn, R.-I. Ernestus, and M. Hoehn, “Metabolic differences between primary and recurrent human brain tumors: a ^1H NMR Spectroscopic Investigation,” *NMR in Biomedicine*, vol. 18, no. 6, pp. 371–382, 2005.
- [4] M. C. Martínez-Bisbal, L. Martí-Bonmatí, J. Piquer et al., “ ^1H and ^{13}C HR-MAS spectroscopy of intact biopsy samples *ex vivo* and *in vivo* ^1H MRS study of human high grade gliomas,” *NMR in Biomedicine*, vol. 17, no. 4, pp. 191–205, 2004.
- [5] F.-G. Lehnhardt, G. Rhn, R.-I. Ernestus, M. Grne, and M. Hoehn, “ ^1H - and ^{31}P -MR spectroscopy of primary and recurrent human brain tumors *in vitro*: malignancy-characteristic profiles of water soluble and lipophilic spectral components,” *NMR in Biomedicine*, vol. 14, no. 5, pp. 307–317, 2001.
- [6] L. Le Moyec, O. Legrand, V. Larue et al., “Magnetic resonance spectroscopy of cellular lipid extracts from sensitive, resistant, and reverting K562 cells and flow cytometry for investigating the P-glycoprotein function in resistance reversion,” *NMR in Biomedicine*, vol. 13, no. 2, pp. 92–101, 2000.
- [7] M. T. Santini, R. Romano, G. Rainaldi et al., “The relationship between ^1H -NMR mobile lipid intensity and cholesterol in

- two human tumor multidrug resistant cell lines (MCF-7 and LoVo)," *Biochimica et Biophysica Acta*, vol. 1531, no. 1-2, pp. 111–131, 2001.
- [8] A. Mannechez, B. Collet, L. Payen et al., "Differentiation of the P-gp and MRP1 multidrug resistance systems by mobile lipid ^1H -NMR spectroscopy and phosphatidylserine externalization," *Anticancer Research*, vol. 21, no. 6, pp. 3915–3919, 2001.
- [9] J. L. Griffin, M. Bollard, J. K. Nicholson, and K. Bhakoo, "Spectral profiles of cultured neuronal and glial cells derived from HRMAS ^1H NMR spectroscopy," *NMR in Biomedicine*, vol. 15, no. 6, pp. 375–384, 2002.
- [10] M. Bloom, K. T. Holmes, C. E. Mountford, and P. G. Williams, "Complete proton magnetic resonance in whole cells," *Journal of Magnetic Resonance*, vol. 69, no. 1, pp. 73–91, 1986.
- [11] L. Le Moyec, R. Tatoud, M. Eugene et al., "Cell and membrane lipid analysis by proton magnetic resonance spectroscopy in five breast cancer cell lines," *British Journal of Cancer*, vol. 66, no. 4, pp. 623–628, 1992.
- [12] M. Spraul, P. Neidig, U. Klauk et al., "Automatic reduction of NMR spectroscopic data for statistical and pattern recognition classification of samples," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 12, no. 10, pp. 1215–1225, 1994.
- [13] J. K. Nicholson, J. C. Lindon, and E. Holmes, "Metabonomics: understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data," *Xenobiotica*, vol. 29, no. 11, pp. 1181–1189, 1999.
- [14] J. C. Lindon, E. Holmes, and J. K. Nicholson, "Pattern recognition methods and applications in biomedical magnetic resonance," *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 39, no. 1, pp. 1–40, 2001.
- [15] E. Holmes, A. W. Nicholls, J. C. Lindon et al., "Development of a model for classification of toxin-induced lesions using ^1H NMR spectroscopy of urine combined with pattern recognition," *NMR in Biomedicine*, vol. 11, no. 4-5, pp. 235–244, 1998.
- [16] J. L. Griffin, K. K. Lehtimäki, P. K. Valonen et al., "Assignment of ^1H nuclear magnetic resonance visible polyunsaturated fatty acids in BT4C gliomas undergoing ganciclovir-thymidine kinase gene therapy-induced programmed cell death," *Cancer Research*, vol. 63, no. 12, pp. 3195–3201, 2003.
- [17] Y. Wang, E. Holmes, J. K. Nicholson et al., "Metabonomic investigations in mice infected with *Schistosoma mansoni*: an approach for biomarker identification," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 34, pp. 12676–12681, 2004.
- [18] D. Zipser, "Identification models of the nervous system," *Neuroscience*, vol. 47, no. 4, pp. 853–862, 1992.
- [19] J. Hernández-Borges, R. Corbella-Tena, M. A. Rodríguez-Delgado, F. J. García-Montelongo, and J. Havel, "Content of aliphatic hydrocarbons in limpets as a new way for classification of species using artificial neural networks," *Chemosphere*, vol. 54, no. 8, pp. 1059–1069, 2004.
- [20] P. Fedor, I. Malenovský, J. Vaňhara, W. Sierka, and J. Havel, "Thrips (Thysanoptera) identification using artificial neural networks," *Bulletin of Entomological Research*, vol. 98, no. 5, pp. 437–447, 2008.
- [21] W. El-Dereby, S. M. Ashmore, N. M. Branston, J. L. Darling, S. R. Williams, and D. G. T. Thomas, "Pretreatment prediction of the chemotherapeutic response of human glioma cell cultures using nuclear magnetic resonance spectroscopy and artificial neural networks," *Cancer Research*, vol. 57, no. 19, pp. 4196–4199, 1997.
- [22] T. Suna, A. Salminen, P. Soinen et al., " ^1H NMR metabonomics of plasma lipoprotein subclasses: elucidation of metabolic clustering by self-organising maps," *NMR in Biomedicine*, vol. 20, no. 7, pp. 658–672, 2007.
- [23] A. R. Tate, J. Underwood, D. M. Acosta et al., "Development of a decision support system for diagnosis and grading of brain tumours using in vivo magnetic resonance single voxel spectra," *NMR in Biomedicine*, vol. 19, no. 4, pp. 411–434, 2006.
- [24] M. Gottschalk, G. Ivanova, D. M. Collins, A. Eustace, R. O'Connor, and D. F. Brougham, "Metabolomic studies of human lung carcinoma cell lines using in vitro ^1H NMR of whole cells and cellular extracts," *NMR in Biomedicine*, vol. 21, no. 8, pp. 809–819, 2008.
- [25] E. Law, U. Gilvarry, V. Lynch, B. Gregory, G. Grant, and M. Clynes, "Cytogenetic comparison of two poorly differentiated human lung squamous cell carcinoma lines," *Cancer Genetics and Cytogenetics*, vol. 59, no. 2, pp. 111–118, 1992.
- [26] C. P. Duffy, C. J. Elliott, R. A. O'Connor et al., "Enhancement of chemotherapeutic drug toxicity to human tumour cells in vitro by a subset of non-steroidal anti-inflammatory drugs (NSAIDs)," *European Journal of Cancer*, vol. 34, no. 8, pp. 1250–1259, 1998.
- [27] M. Heenan, L. O'Driscoll, I. Cleary, L. Connolly, and M. Clynes, "Isolation from a human MDR lung cell line of multiple clonal subpopulations which exhibit significantly different drug resistance," *International Journal of Cancer*, vol. 71, no. 5, pp. 907–915, 1997.
- [28] C. O'Loughlin, M. Heenan, S. Coyle, and M. Clynes, "Altered cell cycle response of drug-resistant lung carcinoma cells to doxorubicin," *European Journal of Cancer*, vol. 36, no. 9, pp. 1149–1160, 2000.
- [29] H. Hanaoka, Y. Yoshioka, I. Ito, K. Niitu, and N. Yasuda, "In vitro characterization of lung cancers by the use of ^1H nuclear magnetic resonance spectroscopy of tissue extracts and discriminant factor analysis," *Magnetic Resonance in Medicine*, vol. 29, no. 4, pp. 436–440, 1993.
- [30] R. J. Ogg, R. B. Kingsley, and J. S. Taylor, "WET, a T_1 - and B_1 -insensitive water-suppression method for in vivo localized ^1H NMR spectroscopy," *Journal of Magnetic Resonance. Series B*, vol. 104, no. 1, pp. 1–10, 1994.
- [31] J. E. Jackson, *A User's Guide to Principal Components*, Wiley-Interscience, New Jersey, NJ, USA, 2003.
- [32] N. J. Waters, E. Holmes, C. J. Waterfield, R. D. Farrant, and J. K. Nicholson, "NMR and pattern recognition studies on liver extracts and intact livers from rats treated with α -naphthylisothiocyanate," *Biochemical Pharmacology*, vol. 64, no. 1, pp. 67–77, 2002.
- [33] V. Govindaraju, K. Young, and A. A. Maudsley, "Proton NMR chemical shifts and coupling constants for brain metabolites," *NMR in Biomedicine*, vol. 13, no. 3, pp. 129–153, 2000.
- [34] R. J. Maxwell, I. Martínez-Pérez, S. Cerdán et al., "Pattern recognition analysis of ^1H NMR spectra from perchloric acid extracts of human brain tumor biopsies," *Magnetic Resonance in Medicine*, vol. 39, no. 6, pp. 869–877, 1998.
- [35] Y. Hiltunen, J. Kaartinen, J. Pulkkinen, A.-M. Häkkinen, N. Lundbom, and R. A. Kauppinen, "Quantification of human brain metabolites from in vivo ^1H NMR magnitude spectra using automated artificial neural network analysis," *Journal of Magnetic Resonance*, vol. 154, no. 1, pp. 1–5, 2002.
- [36] H. Bhat, B. R. Sajja, and P. A. Narayana, "Fast quantification of proton magnetic resonance spectroscopic imaging with artificial neural networks," *Journal of Magnetic Resonance*, vol. 183, no. 1, pp. 110–122, 2006.

Research Article

Identification of Urinary Biomarkers of Colon Inflammation in IL10^{-/-} Mice Using Short-Column LCMS Metabolomics

Don Otter,¹ Mingshu Cao,¹ Hui-Ming Lin,² Karl Fraser,¹ Shelley Edmunds,²
Geoff Lane,¹ and Daryl Rowan²

¹AgResearch Grasslands, Private Bag 11008, Palmerston North 4442, New Zealand

²The New Zealand Institute for Plant and Food Research Limited (Plant & Food Research), Private Bag 11600, Palmerston North 4442, New Zealand

Correspondence should be addressed to Daryl Rowan, daryl.rowan@plantandfood.co.nz

Received 17 May 2010; Revised 27 September 2010; Accepted 20 October 2010

Academic Editor: Olav Kvalheim

Copyright © 2011 Don Otter et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The interleukin-10-deficient (IL10^{-/-}) mouse develops colon inflammation in response to normal intestinal microflora and has been used as a model of Crohn's disease. Short-Column LCMS metabolite profiling of urine from IL10^{-/-} and wild-type (WT) mice was used, in two independent experiments, to identify mass spectral ions differing in intensity between these two genotypes. Three differential metabolites were identified as xanthurenic acid and as the glucuronides of xanthurenic acid and of α -CEHC (2,5,7,8-tetramethyl-2-(2'-carboxyethyl)-6-hydroxychroman). The significance of several differential metabolites as potential biomarkers of colon inflammation was evaluated in an experiment which compared metabolite concentrations in IL10^{-/-} and WT mice housed, either under conventional conditions and dosed with intestinal microflora, or maintained under specific pathogen-free (SPF) conditions. Concentrations of xanthurenic acid, α -CEHC glucuronide, and an unidentified metabolite m/z 495⁻/497⁺ were associated with the degree of inflammation in IL10^{-/-} mice and may prove useful as biomarkers of colon inflammation.

1. Introduction

Crohn's disease is a chronic relapsing inflammatory disorder of the bowel which is characterized by recurring episodes of inflammation especially in the small and large intestine accompanied by abdominal pain and persistent diarrhea [1]. Crohn's disease has been attributed to a dysregulation of the intestinal mucosal immune response towards normal intestinal microflora and results from a complex interaction of genetic [2, 3] and environmental factors [4]. The complex aetiology and absence of a cure means that the management of Crohn's disease requires the continuing assessment of the inflammation status of the patient so that the efficacy of treatments can be determined. Current methods of assessing bowel inflammation require the sampling of intestinal tissue for histological evaluation or blood for measuring plasma biomarkers of inflammation. Metabolomic characterization, largely by NMR, of urine [5], faeces [6–8], and colon mucosal tissues [9] from Crohn's disease patients has revealed a range

of metabolites whose concentrations are perturbed in association with established inflammation. Urinary biomarkers predictive of inflammation status would be preferable to sampling of intestinal tissue or blood as the collection of urine samples is relatively noninvasive and multiple samples can more readily be obtained.

A number of metabolomic studies of mouse models of Crohn's disease have been used to study the biochemical changes associated with inflammation and to identify possible biomarkers. LCMS metabolite profiling of serum from mice treated with dextran sulphate sodium (DSS) to induce colitis showed that colon inflammation was increased by DSS-inhibition of stearoyl-CoA desaturase 1 (SCD1)-mediated oleic acid biogenesis [10]. Examination by NMR of colonic mucosal samples from Sprague Dawley rats treated with 2% carrageenan for 1 or 2 weeks to induce inflammation showed increases in concentrations of creatinine, phosphatidylcholine and unsaturated lipids [11]. The interleukin-10-deficient (IL10^{-/-}) mouse shows

physiological, and biochemical similarities to Crohn's disease [3] including the development of intestinal inflammation in response to the presence of intestinal microflora and has been widely used as a model of Crohn's disease [12]. Interleukin-10 (IL10) is an immunosuppressive cytokine that down-regulates cell-mediated immune responses and has an important role in maintaining intestinal mucosal immunity [13]. Metabolic profiling of plasma of IL10^{-/-} mice by NMR indicated the inflamed state was characterised by impaired metabolism of glyco- and lipoproteins and loss of energy homeostasis with higher levels of fatty acid oxidation and glycolysis and interconversion of amino acids to produce energy [14]. NMR metabolite profiling of 24-hour pooled urine [15] and GCMS analysis of spot urine samples [16] from IL10^{-/-} mice identified changes in urinary metabolites associated with colon inflammation providing insights into biochemical changes associated with disease progression. In particular, fucose and xanthurenic acid were identified as early biomarkers of inflammation in the urine of IL10^{-/-} mice [17].

The use of multiple metabolomic platforms extends the range of metabolites and hence the possible perturbations in biochemical pathways that may be observed. Recently, in a cross-platform comparison of metabolomics methods, LCMS was recommended as providing the best combination of versatility and robustness on a single platform [18]. Direct infusion (DI) MS/MS [19, 20] is a recently developed metabolomic analysis tool developed for the rapid screening of complex biological samples. DI-MS/MS combines the analytical power of MS/MS for metabolite characterisation with the advantages of minimal sample handling and high sample throughput.

For nontargeted urinary metabolite profiling, we have modified the DI-MS/MS procedure using a short 20-mm reversed-phase column (Short-Column LCMS) to separate and discard to waste highly ionic material and urea which would otherwise dominate the MS data [21]. In this study, we demonstrate the use of this methodology to rapidly screen sample sets and to identify metabolite differences between the urine from IL10^{-/-} and wild-type mice that were not detected with other techniques. Structural identification of three of these differential metabolites using MS/MS, high resolution MS, and chemical synthesis has provided further insight into the metabolic changes associated with the IL10^{-/-} deficiency. Finally the usefulness of these metabolites as biomarkers of colon inflammation was tested in an experiment where IL10^{-/-} mice were either dosed with intestinal microflora to initiate inflammation or maintained under specific pathogen-free (SPF) conditions to limit the development of inflammation.

2. Materials and Methods

2.1. Animal Experiments. All animal studies were reviewed and approved by the Crown Research Institute Animal Ethics Committee in Hamilton, New Zealand, according to the New Zealand Animal Welfare Act (1999). Male IL10^{-/-} mice of C57BL/6 background strain (B6.129P2-*Il10*^{tm1Cgn}) and wild-type C57BL/6 mice (The Jackson Laboratory, Bar Harbor,

Maine, USA) of average age upon arrival of 4, 5.3, and 4.4 weeks old were used for experiments 1, 2 [16], and 3 [17], respectively. For experiments 1 and 2 (Figure 1), each mouse was housed individually in a shoebox-style cage under conventional conditions and fed powdered AIN76A diet with food intake adjusted to equal the mean amount of food consumed by IL10^{-/-} mice the previous week. Four days after arrival, mice were orally dosed with a mixture of *Enterococcus faecalis*, *E. Faecium*, and complex intestinal flora to ensure the same initial microbial exposure and the development of consistent intestinal inflammation [22]. In experiment 1, yellow-fleshed kiwifruit (*Actinidia chinensis*) and in experiment 2, green-fleshed kiwifruit (*A. delicoisa* "Hayward"), fruit extracts were added to the diets of some mice [16] to test their possible anti-inflammatory effects [23]. Transcriptomic and proteomic results of diet treatments will be reported elsewhere. In experiment 3, all mice were fed powdered AIN76A diet only. Approximately half the IL10^{-/-} and wild-type mice were maintained under conventional conditions as above and the other half were maintained under SPF conditions and not dosed with intestinal microflora to maintain a lower level of colon inflammation in the IL10^{-/-} mice throughout the experiment [17]. Spot urine samples were collected from each experiment on four collection days when the average age of mice was 5.5, 7, 8.5, and 10.5 weeks old (experiment 1), 7, 9, 11.5 and 12 weeks old (experiment 2) [16], and 6.3, 8.0, 9.4 and 10.4 weeks old (experiment 3) [17]. Final urine samples for experiments 1 and 2 were collected immediately before mice were euthanized and after a cycle of feeding, fasting, and feeding designed to standardize the timing of the final food intake. Mice from each treatment of experiment 3 were euthanized after three and six weeks for histology and measurement of serum amyloid-A (SAA) protein concentrations [17].

2.2. Sample Preparation of Urine for LCMS. Urine (4 μ L) was diluted with 0.1% formic acid (200 μ L) in an Eppendorf tube, centrifuged at 12100 g for 2 minutes to precipitate particulates and an aliquot transferred to a vial before injection into the LCMS. Randomised batches of samples for each time point were analyzed sequentially.

2.3. Short-Column LCMS for Biomarker Identification. Samples were analyzed by a rapid LC-MS procedure [21] using a Thermo Surveyor pump and auto-sampler connected to a Thermo LTQ linear ion-trap mass spectrometer (Thermo Electron Corporation, San Jose, CA) using negative electrospray ionization. The capillary temperature was 275°C and source ionization voltage was -4000 V. A 20 μ L aliquot of diluted urine was loaded onto a Strata-X on-line extraction cartridge (20 \times 2.0 mm, Phenomenex, Torrance, CA) with water (300 μ L min⁻¹). For the first 0.5 min. the flow from the cartridge was diverted away from the mass spectrometer to waste, after which all flow entered the electrospray source. Samples were eluted with a gradient comprising water containing 0.1% formic acid (A) and acetonitrile containing 0.1% formic acid (B). The cartridge was eluted with solvent A

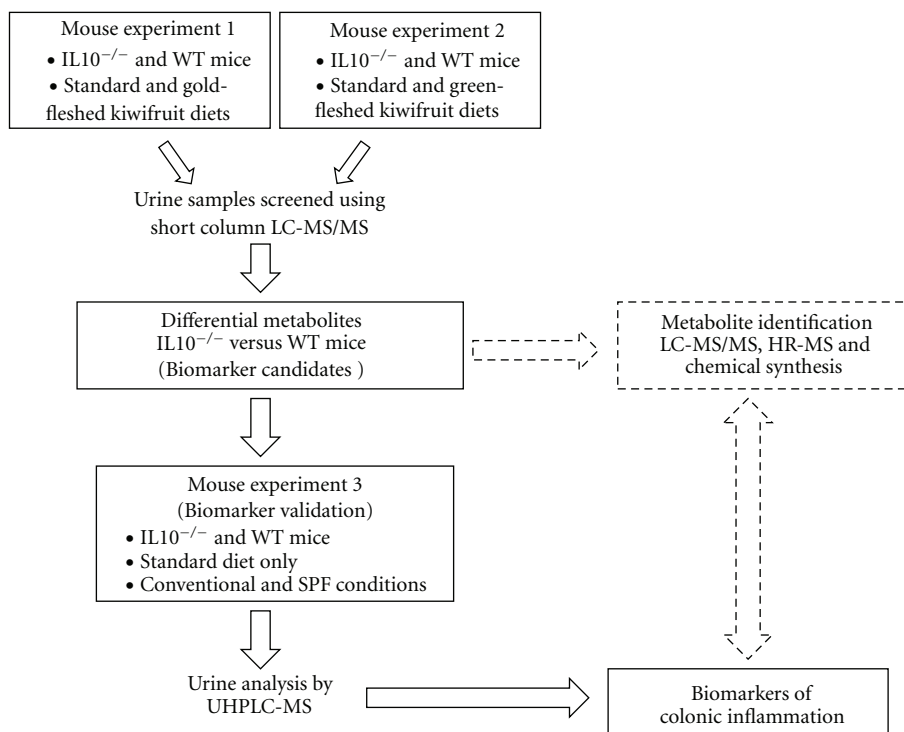


FIGURE 1: Overall design of mouse experiments 1, 2, and 3 and process used to identify and validate urinary biomarkers of colonic inflammation in the $IL10^{-/-}$ mouse model of Crohn's disease.

for 1.5 min, then with a linear gradient to 100% B over 3 min and held at 100% B for 3 min. The initial solvent (100% A) was restored over a 0.5-minute linear gradient and eluted for 1.5 min at $600 \mu\text{L min}^{-1}$ and for 0.5 min at $300 \mu\text{L min}^{-1}$. The ion trap was programmed to collect an MS^1 spectrum from 100 to 1000 m/z followed by MS^2 and MS^3 spectra from collision-induced dissociation of the most intense ions in the MS^1 spectrum, using a 12-second exclusion list to cycle through coeluting ions.

2.4. LCMS/MS for Metabolite Characterization. Samples selected for further metabolite characterisation were analyzed by LCMS/MS in negative and positive electrospray ionization mode using a Thermo LTQ ion-trap mass spectrometer with temperature and voltage settings as above. The HPLC system consisted of two binary Jasco X-LC 3185PU pumps (Jasco Corporation, Tokyo, Japan) connected to a HTS-PAL auto-sampler (CTC Analytics, Zwingen, Switzerland). A $15 \mu\text{L}$ aliquot of diluted urine was loaded onto a Luna C18 column ($150 \times 2.0 \text{ mm}$, $5 \mu\text{m}$ particle size, Phenomenex, Torrance, CA) with a mobile phase flow rate of $200 \mu\text{L min}^{-1}$. For the first 3 min, the flow from the column was diverted away from the mass spectrometer to waste, after which all flow entered the electrospray source. Samples were eluted with a gradient comprising water containing 0.1% formic acid (A) and acetonitrile containing 0.1% formic acid (B). The gradient was 95% solvent A for the first 3 min, then a linear gradient to 98% B over 27 min. This was held at 98% B for 3 min and then the initial solvent (95% A) was restored over a 1 min linear gradient and re-equilibrated

for 6 min. The ion trap was programmed to collect a MS^1 spectrum from 150 to 850 m/z followed by MS^2 , MS^3 , and MS^4 spectra from collision-induced dissociation of the most intense target ion from a parent ion list comprising candidate negative ions as in Table 1 or the appropriate product ion spectrum. MS data was analysed with XCalibur software (Thermo Electron Corporation, San Jose, CA). Ion trees were compiled from extracted product ion spectra.

2.5. UHPLC-MS. Samples were analyzed using a Jasco X-LC UHPLC and HTS-PAL autosampler connected to the LTQ with instrument settings as described above. A $20 \mu\text{L}$ aliquot of diluted urine was loaded onto a Zorbax SB-C18 UPHLC column ($50 \times 3.0 \text{ mm}$, $1.8 \mu\text{m}$ particle size, Agilent Technologies, Santa Clara, CA, USA) with a mobile phase flow rate of $900 \mu\text{L min}^{-1}$. For the first 0.7 min, the flow from the column was diverted away from the mass spectrometer to waste, after which all flow entered the electrospray source. Samples were eluted with a gradient comprising water containing 0.1% formic acid (A) and acetonitrile containing 0.1% formic acid (B). The gradient was 95% solvent A for the first 0.5 min, then with a linear gradient to 100% B over 4.5 min and then restored to the initial solvent (95% A) over a 0.5 min linear gradient and re-equilibrated for 2 min. The ion trap was programmed to collect an MS^1 spectrum from 100–1000 m/z followed by MS^2 and MS^3 spectra from collision-induced dissociation of the most intense ions in the MS^1 spectrum, using a 4-second exclusion list to cycle through coeluting ions.

TABLE 1: Biomarker Candidate ions (m/z) selected from Short-Column LCMS and LC-MS/MS analysis of urine samples from IL10^{-/-} and wild-type (WT) mice.

| Short-Column LCMS Biomarker ion (m/z) | Average fold change Log ₂ (IL10 ^{-/-} /WT) | P value | Early difference observed | Identification, |
|---|--|------------------|---------------------------|------------------------------|
| Negative ions | | | | |
| 204 | 1.8 (expt. 1) | <.01 (FDR <0.05) | Y | xanthurenic acid |
| | 2.5 (expt. 2) | <.01 (FDR <0.05) | | |
| 308 | -1.5 (expt. 1) | <.01 (FDR >0.05) | ? (expt. 1) | unknown |
| | -1.8 (expt. 2) | <.01 (FDR <0.05) | Y (expt. 2) | |
| 380 | -1.1 (expt. 1) | >.01 | N (expt. 1) | xanthurenic acid glucuronide |
| | -1.6 (expt. 2) | >.01 | Y (expt. 2) | |
| 453 | -1.5 (expt. 1) | <.01 (FDR >0.05) | N (expt. 1) | α -CEHC glucuronide |
| | -2.0 (expt. 2) | <.01 (FDR <0.05) | Y (expt. 2) | |
| 495 | -1.5 (expt. 1) | <.01 (FDR >0.05) | Y (expt. 1) | unknown |
| | -2.2 (expt. 2) | <.01 (FDR <0.05) | Y (expt. 2) | |
| Positive ions | | | | |
| 206 | 1.7 (expt. 1) | <.01 (FDR <0.05) | Y (expt. 1) | xanthurenic acid |
| | 1.2 (expt. 2) | >.01 | N (expt. 2) | |
| 382 | -1.2 (expt. 1) | >.01 | N (expt. 1) | xanthurenic acid glucuronide |
| | -1.5 (expt. 2) | <.01 (FDR <0.05) | Y (expt. 2) | |
| 497 | -1.2 (expt. 1) | >.01 | Y (expt. 1) | unknown |
| | -1.4 (expt. 2) | <.01 (FDR <0.05) | Y (expt. 2) | |

2.6. *High Resolution (HR) Mass Spectrometry.* Salts and urea were removed from a pooled IL10^{-/-} urine sample by elution of the sample and an aqueous wash through a C-18 SPE cartridge, and the 80% methanol eluate was collected for MS analysis. The eluate was infused into an LTQ-FTMS (Thermo Electron Corporation, San Jose, CA, USA) for MS/MS and high resolution MS analysis under both positive and negative electrospray ionisation conditions. HR ions observed for xanthurenic acid were negative ions, MS² product ion (m/z 204 \rightarrow 160) m/z 160.0405 ([M-H-CO₂]⁻, C₉H₅O₂N⁻ requires 160.0398) and positive ions MS¹ m/z 206.0450 ([M+H]⁺, C₁₀H₇O₄N⁺ requires 206.0448), and MS² product ion (m/z 206 \rightarrow 188) m/z 188.0344 (C₁₀H₅O₃N⁺ requires 188.0347). HR ions observed for α -CEHC glucuronide were MS¹ m/z 453.1773 ([M-H]⁻, C₂₂H₂₉O₁₀⁻ requires 453.1761) and m/z 435.1664 ([M-H-H₂O]⁻, source fragment), C₂₂H₂₇O₉⁻ requires 435.1655). HR LCMS ions observed for xanthurenic acid glucuronide were MS¹ m/z 382.0762 ([M+H]⁺, C₁₆H₁₆NO₁₆ requires 382.0774) and MS² product ion m/z 206.0450 (C₁₀H₈NO₄ requires 206.0453).

2.7. *Bioinformatic Analysis of Short-Column LCMS/MS Data.* Raw data (positive and negative ionization) were converted to mzXML format. Nominal mass binning (integer $m/z \pm 0.5$) was carried out for the full m/z range of 100 to 1000 resulting in 901 nominal bins for each individual sample. Signal intensity at nominal mass resolution was retrieved using continuous wavelet transform (CWT) algorithms [24]

for peaks within each bin with S/N > 3 (noise defined as 95% quantile of absolute CWT coefficients of scale one). The median peak intensity of the identified peaks in each bin was used to represent the intensity of that bin. Peak intensity was then normalised by a linear regression of log ion intensity against run sequence. Empirical Bayes moderated t statistics [25] (R Limma package) were applied to identify MS¹ ions that were significantly different (FDR adjusted $P < .05$) between samples from IL10^{-/-} and wild-type mice in each experiment. Linear modelling of the full factorial design with three factors (genotype, diet (experiments 1 and 2), and day of collection of urine samples) was also carried out for each experiment. No significant diet effects were found for the metabolites discussed below.

2.8. *Metabolite Identification.* Possible metabolite structures were inferred from MS/MS ion trees and high resolution FTMS data using web-based metabolite databases [26] and by comparison with authentic compounds and published mass spectral fragmentation pathways where available. Xanthurenic acid was obtained from Aldrich Chemical Co. Xanthurenic acid sulphate dipotassium salt was prepared by chemical synthesis [27]. α -CEHC glucuronide was synthesised following the method of Pope et al. [28].

3. Results

3.1. *Short-Column LCMS Data Analysis.* To identify metabolic differences between IL10^{-/-} and wild-type mice,

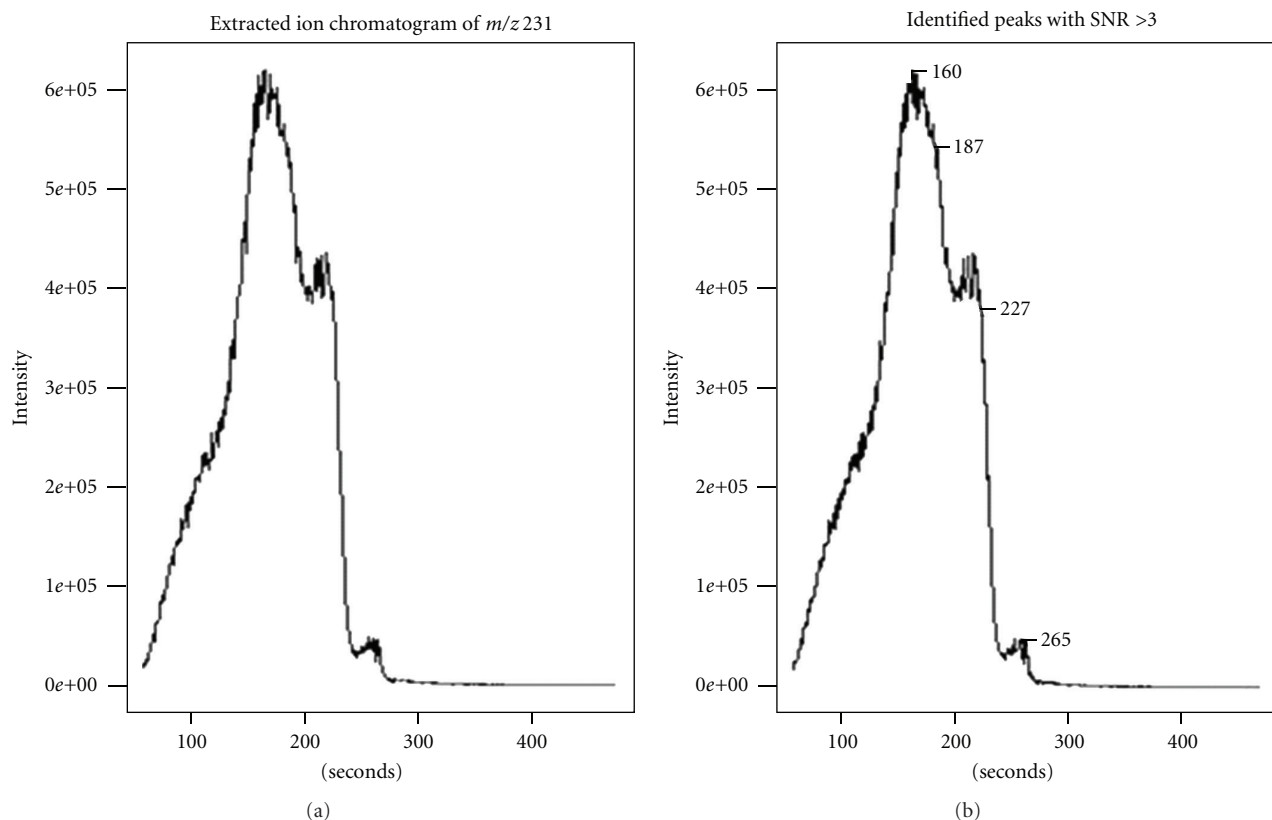


FIGURE 2: Short-Column LCMS ion chromatogram for m/z 231 showing broad partially resolved peaks typically obtained on desorption of analytes from the 20 mm Strata-X on-line extraction cartridge. Wavelet algorithms were used to identify local peaks, and the median intensity of these peaks was taken to represent the intensity of m/z 231.

urine samples were collected at four time points from each of two independent experiments (experiments 1 and 2) [16] (Figure 1). Short-Column LCMS using a short (20 mm) solid phase extraction column enabled extremely rapid sample analysis (8 minutes per sample) and avoided blocking the source with salt build-up but provided only minimal chromatographic separation of analytes. The LCMS data for each sample consisted of broad peaks of varying signal intensity on each of the unit mass channels of the mass spectrometer (Figure 2). LCMS data files were processed based on methods developed for the direct-infusion LCMS data analysis [29] but with notable modifications. With the short solid phase extraction column used in these experiments, the signal distributions within mass bins varied with time with variable and non-Gaussian chromatographic peaks. Robust statistics were needed to summarise the signal intensities within each mass bin. Wavelet-based signal identification was used to identify any m/z peaks within each bin with $S/N > 3$ [24] with the median average of the identified peaks in each bin being used to represent the intensity of that bin.

A progressive decline in ion intensity, as observed by Koulman et al. [20], was observed during the infusion of the large number of urine samples. Several normalization approaches [30] including linear regression, local regression loess and quantile were tested to correct for this decline.

The loss of ion intensity was best explained by a linear relationship between the log of ion intensity and the position of samples in the run sequence. Normalization based on this model was carried out for all MS ions and the normalized data were subjected to statistical analysis.

3.2. Selection of Biomarker Candidate Ions from Short-Column LCMS Data. MS^1 ions that were significantly different (FDR adjusted P -value $P < .05$) between samples from $IL10^{-/-}$ and wild-type mice in each experiment were identified using empirical Bayes moderated t -statistics [25] (R Limma package) (Supplementary Materials available online at doi: 10.1155/2011/974701, Table S1). More differential metabolites were found by negative mode Short-Column LCMS, however, there was also a pleasing correspondence between the top candidate ions identified in both experiments (1 and 2) and to a lesser extent between candidate ions identified in both positive and negative ionisation modes. From this initial list, biomarker candidates were selected for further investigation based on additional criteria: the detection and level of significant of ions in both experiments, detection early in the experiment, detection as both positive and negative ions, and the presence of sufficient peak intensity to allow detailed MS/MS analysis. A more complex analysis using linear mixed modelling to incorporate effects of diet and the metabolic variation of individual mice was used to evaluate changes

in the metabolic responses over time but did not produce additional candidate ions discriminating the IL10^{-/-} mice. LC-MS/MS analysis of selected samples was then carried out to identify which candidate ions (and ion trees) identified from Short-Column LCMS analysis corresponded to single or to multiple metabolites, to eliminate source fragment ions and adducts, and to obtain expanded MS/MS ion trees of chromatographically resolved metabolites of interest. The mass spectral ions selected by these criteria are listed in Table 1.

3.3. Structural Characterization of Biomarker Candidates

3.3.1. Xanthurenic Acid (*m/z* 160, 204 [M-H]⁻ and *m/z* 206 [M+H]⁺). Three major metabolites with *m/z* 204⁻ were detected in mouse urine by LCMS of which only the major, and earliest eluting metabolite, was elevated in the urine from IL10^{-/-} mice. This metabolite co-eluted with the only significant *m/z* 160⁻ ion in the samples (also elevated in IL10^{-/-} mice), a likely source fragment which was shown to arise from the *m/z* 204⁻ ion, (loss of 44 amu) by MS/MS. High resolution FTMS and LCMS comparison with authentic material identified this metabolite as xanthurenic acid.

3.3.2. Xanthurenic Acid Glucuronide (*m/z* 380 [M-H]⁻ and *m/z* 382 [M+H]⁺). LCMS identified the negative ions at *m/z* 380 and the positive ion at *m/z* 382 as resulting from the same metabolite that occurred at lower concentrations in the urine from IL10^{-/-} mice (Table 1). LC-MS/MS showed loss of a glucuronide-derived fragment (176 amu) in both negative and positive ionization modes to form a product ion (*m/z* 204 or 206 resp.) which underwent further fragmentations as observed for the corresponding xanthurenic acid ions (Figure 3). High resolution MS data is consistent with xanthurenic acid glucuronide, although this remains to be confirmed by chemical synthesis. Xanthurenic acid glucuronide was identified as a differentiating metabolite (FDR adjusted *P* < .05) in both positive and negative ion LCMS and in both experiments 1 and 2, but with concentrations changing in the opposite direction to that of its aglycone (Table 1).

3.3.3. α -CEHC Glucuronide (2,5,7,8-Tetramethyl-2-(2'-Carboxyethyl)-6-Hydroxychroman Glucuronide, *m/z* 453 [M-H]⁻). The ion (*m/z* 453⁻) occurred at lower concentrations in urine from IL10^{-/-} mice. A metabolite (M⁺ 454) not containing nitrogen was assumed. LC-MS/MS analysis identified this metabolite as a second glucuronide (MS² loss of 176 amu to *m/z* 277) with a subsequent loss of 44 amu, implying a glucuronide acid conjugate. The molecular formula therefore requires a minimum of nine oxygen atoms. FTMS analysis indicated that the major *m/z* 453⁻ ion in the sample showed the required ion tree and had an accurate mass of 453.1773. Constraining any molecular formula to a minimum of 9 oxygen atoms gave C₂₂H₂₉O₁₀⁻ as the most likely formula for this ion (theoretical value 453.1761). Similarly, the major fragment ion at *m/z* 435.1664 (loss of

18 amu) is assigned a formula of C₂₂H₂₇O₉ (theoretical value 435.1655). The molecular formula is then C₂₂H₃₀O₁₀ with C₁₆H₂₂O₄ being the formula of the aglycone (a carboxylic acid). A search for such aglycones in the Human Metabolome Database [31] suggested that this metabolite was the glucuronide of α -CEHC. The MS/MS fragmentation appeared consistent with this structure (Figure 4). α -CEHC glucuronide was prepared by synthesis [28] and the MS/MS and LCMS matched that of the urine metabolite. α -CEHC is a major water-soluble metabolite of vitamin E, which circulates in the blood and is excreted in the urine [28].

3.3.4. Unidentified Metabolites. The *m/z* 284⁻ ion was detected amongst the most significant differentiating species between IL10^{-/-} and wild-type mice (Table S1), and xanthurenic acid sulphate, a known urinary metabolite of xanthurenic acid [27], seemed a likely candidate. UHPLC-MS analysis, however, showed that this ion was an isotopologue of *m/z* 283⁻, also a highly significant differentiating species (Table S1), and eluted at a later retention time than authentic xanthurenic acid sulphate [27]. Xanthurenic acid sulphate (*m/z* 284 [M-H]⁻) was not detected by UHPLC-MS (Experiment 3) using an authentic standard.

Two additional ions (*m/z* 308⁻ and *m/z* 497⁺/495⁻) were significantly reduced in urine from IL10^{-/-} mice (Table 1). The first metabolite (*m/z* 308) was found only as a negative ion, and LC-MS/MS analysis (MS² *m/z* 124, MS³ *m/z* 107, 80) suggested a taurine conjugate. The second pair of ions (*m/z* 497⁺/495⁻) would seem to be derived from the same metabolite as they appear at the same retention time on LCMS/MS with the same losses of 147 and 75 (glycine) amu from the corresponding pseudomolecular ions. The structures of these metabolites remain unknown.

3.4. Evaluation of Candidates as Biomarkers of Colon Inflammation. The biological significance of these metabolic differences as biomarkers of colon inflammation was evaluated in experiment 3 that compared IL10^{-/-} and wild-type mice housed either under conventional conditions and dosed with intestinal microflora or maintained under SPF conditions without dosing with intestinal microflora [17]. The rationale for this experiment was that biochemical differences arising from colon inflammation in IL10^{-/-} mice should be reduced when these mice were maintained under SPF conditions [3]. Biochemical differences arising from other genetic differences between the mouse strains [32] or as a consequence of microbial dosing should not be similarly affected. IL10^{-/-} mice housed under these SPF conditions showed decreased colon inflammation as measured by histology scores and SAA levels in serum [17]. Urine samples were analyzed using UHPLC-MS in both positive and negative ionisation modes with both targeted analysis of biomarker candidates and untargeted metabolic profiling of all metabolites. The top scoring metabolites identified as differing between IL10^{-/-} and wild-type mice by untargeted analysis included xanthurenic acid, xanthurenic acid glucuronide, and α -CEHC glucuronide among the top 20 candidates (Supplementary Table S2: ion labels below

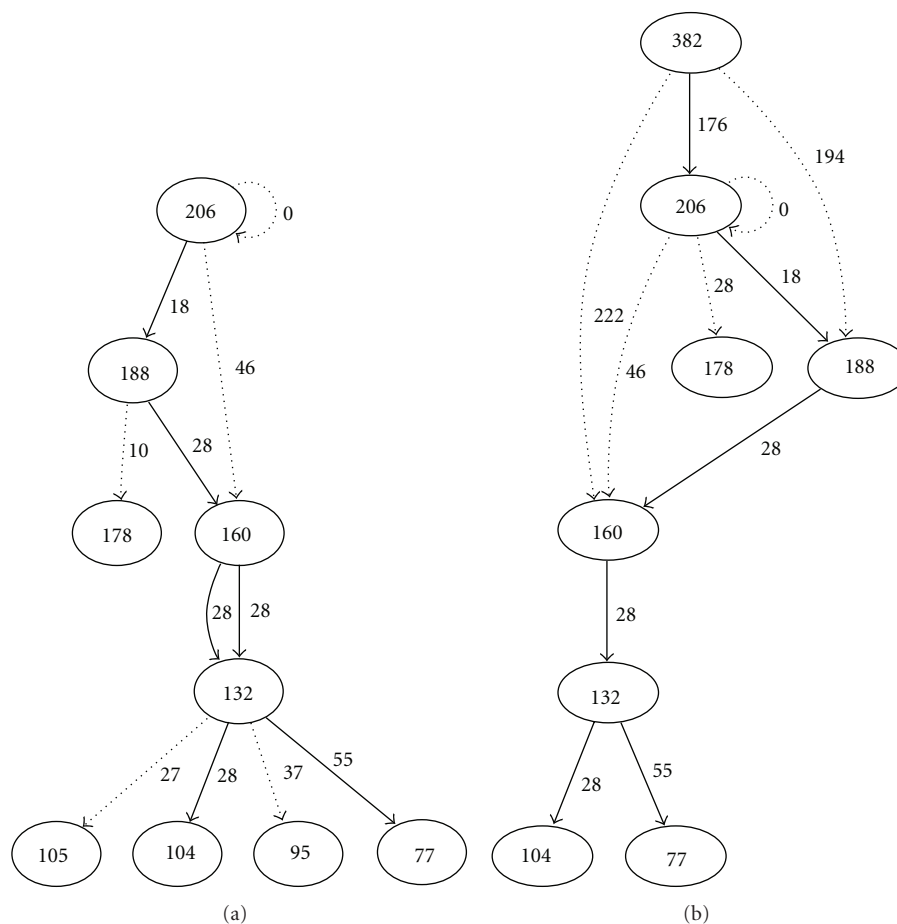


FIGURE 3: LC-MS/MS ion trees (positive ions) for A, xanthurenic acid (m/z 206 $[M-H]^-$) and B, xanthurenic acid glucuronide (m/z 382 $[M-H]^-$) showing major ($>80\%$ relative intensity), intermediate ($\geq 30 \leq 80\%$), and minor ($>1 < 30\%$) MS/MS fragmentations using bold, normal and dotted arrows, respectively.

refer to this table). Amongst the positive ions m/z 206 (M206T94), with isotopologues m/z 207 and 208, was the top ranking candidate showing an average 2.4-fold increase ($P < .0001$) in urine from IL $IL10^{-/-}$ mice. However, the corresponding negative ion (M204T91) was only the 21st ranking candidate despite the high ranking for this ion in previous experiments. Targeted analysis of xanthurenic acid monitoring the m/z 204⁻ MS¹ ion in negative mode showed interference from a closely eluting metabolite.

Xanthurenic acid concentrations were significantly higher in dosed mice ($P = .002$) with this increase being more apparent between the dosed and SPF $IL10^{-/-}$ animals ($P < .001$). In addition, SPF $IL10^{-/-}$ mice showed an increase in xanthurenic acid concentrations from the earliest timepoint ($P < .001$) supporting the hypothesis that xanthurenic acid is an early biomarker of inflammation and consistent with the increasing histology scores and SAA concentrations measured for these mice [17].

UHPLC-MS analysis confirmed that xanthurenic acid glucuronide (M382T85 and M380T82) occurred at significantly higher concentrations in urine from WT mice ($P < .0001$) (Figure 5) and showed that, while concentrations of this metabolite increased throughout the experiment in both types of mice ($P < .001$), concentrations in

$IL10^{-/-}$ mice were not affected by housing conditions (SPF or conventional) ($P = .5$). We therefore conclude that the difference in concentrations of this metabolite between $IL10^{-/-}$ and WT mice represents a metabolic difference between the two mouse strains unrelated to the state of inflammation [17, 32]. Concentrations of this metabolite increased with time in $IL10^{-/-}$ mice under SPF conditions. However, similar increases also occurred in WT mice and would seem to be related to the age of the mice rather than to inflammation.

UHPLC-MS analysis showed that α -CEHC glucuronide (M453T145) occurred at significantly higher concentrations in urine from WT mice (Figure 6) and at higher concentrations when either $IL10^{-/-}$ or WT mice were housed under SPF conditions ($P < .0001$). The concentration of this metabolite in urine from WT mice was unaltered over the course of the experiment ($P = .5$) but decreased ($P = .02$) in the $IL10^{-/-}$ mice under both SPF and conventional conditions. Overall these results suggest that reduced concentrations of α -CEHC glucuronide may serve as a marker of inflammation status in the $IL10^{-/-}$ mice

The metabolite with ions M495T117/M497T116 appeared as the 29th and 41st ranking ions that were significantly different between $IL10^{-/-}$ and wild-type

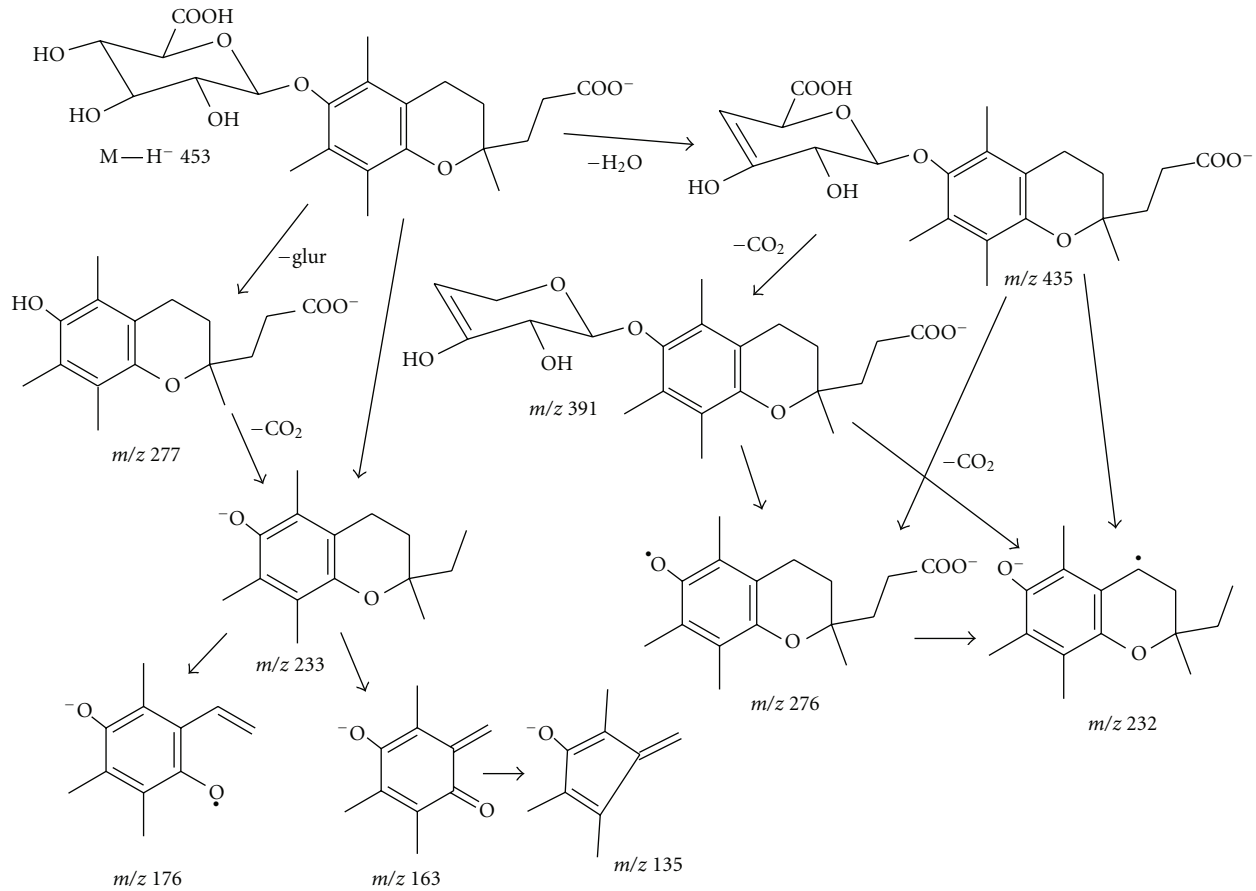


FIGURE 4: LC-MS/MS fragmentations for α -CEHC (2,5,7,8-tetramethyl-2'-(2'-carboxyethyl)-6-hydroxychroman) glucuronide.

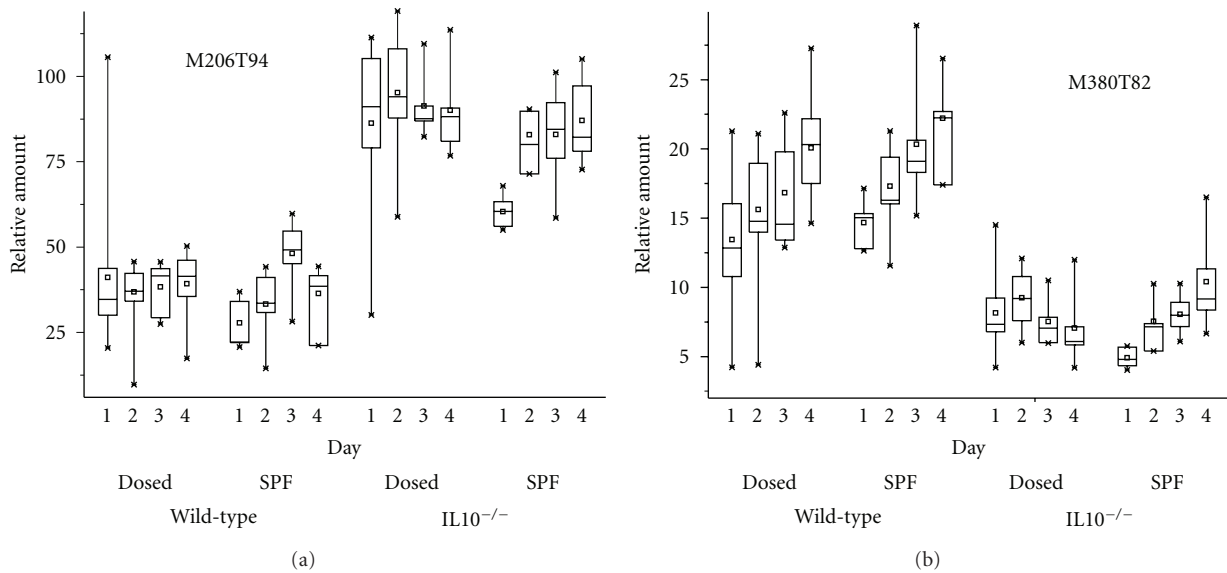


FIGURE 5: Relative concentrations of xanthurenic acid (M206T94) and xanthurenic acid glucuronide (M380T82, m/z 380 ($[M-H]^-$), identified by LCMS/MS) measured by UHPLC-MS in urine samples from $IL10^{-/-}$ and wild-type (WT) mice housed either under conventional conditions and dosed with intestinal microflora or maintained under specific pathogen free (SPF) throughout the experiments. Urines samples were collected 7, 19, 29, and 36 days after dosing [17].

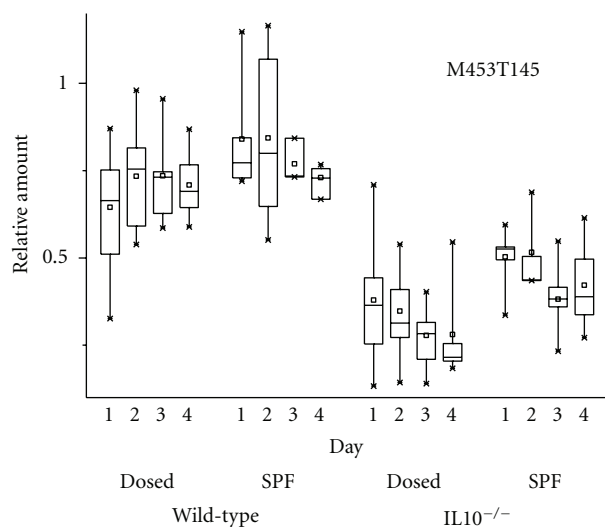


FIGURE 6: Relative concentrations of α -CEHG glucuronide measured by UHPLC-MS in urine samples from $IL10^{-/-}$ and wild-type (WT) mice housed either under conventional conditions and dosed with intestinal microflora or maintained under specific pathogen free (SPF) throughout the experiments. Urines samples were collected 7, 19, 29, and 36 days after dosing [17].

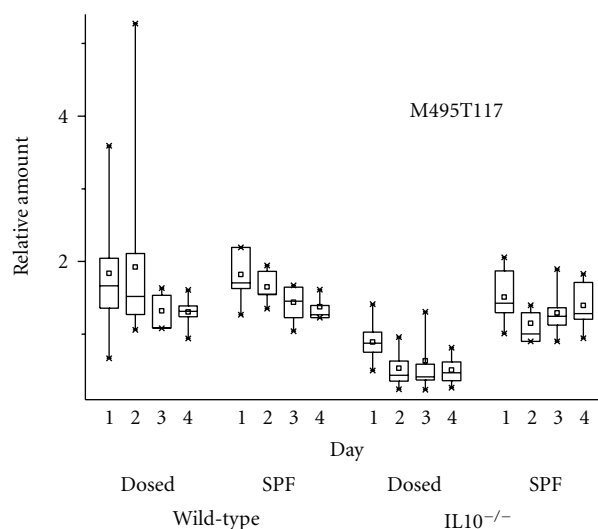


FIGURE 7: Relative concentrations of biomarker M495T117 (m/z 495⁻) measured by UHPLC-MS in urine samples from $IL10^{-/-}$ and wild-type (WT) mice housed either under conventional conditions and dosed with intestinal microflora or maintained under specific pathogen free (SPF) throughout the experiments. Urines samples were collected 7, 19, 29, and 36 days after dosing [17].

samples. Both ions occurred at significantly lower concentrations ($P < .001$) in urine from $IL10^{-/-}$ mice (Figure 7) confirming results obtained by Short-Column LCMS; however, this suppression was partially removed by SPF treatment (Treatment effect $P < .001$ for m/z 495 and $P = .03$ for the less intense m/z 497 ion). Metabolite concentrations in urine from $IL10^{-/-}$ mice were significantly reduced as a result of microbial dosing ($P < .0001$). Concentrations of the more intense m/z 495⁻ ion were also significantly reduced with time ($P = .014$ overall; $P < .001$ for $IL10^{-/-}$ mice). These results suggest that reduced concentrations of this metabolite were related to the extent of colon inflammation in the $IL10^{-/-}$ mouse.

UHPLC-MS analysis confirmed that the metabolite with m/z 308 (M308T85) occurred at slightly higher concentrations in urine from $IL10^{-/-}$ mice ($P < .0001$) and showed that concentrations were reduced in urine from $IL10^{-/-}$ mice under SPF conditions ($P = .003$) (Figure 8). However, the downward trend in urinary concentrations of this metabolite with time for all treatments ($P < .0001$) suggested that its elevated concentrations in $IL10^{-/-}$ mice was not associated with colon inflammation.

4. Discussion

Xanthurenic acid was previously identified by GCMS as an early biomarker of inflammation in urine from $IL10^{-/-}$ mice [16, 17] and is a product of tryptophan catabolism through the kynurenine pathway. This pathway is activated by pro-inflammatory stimuli such as bacterial lipopolysaccharides and interferon- γ cytokine [33]; however, metabolites of this pathway are also associated with the induction of immune tolerance [33–35]. Thus the elevated levels of xanthurenic

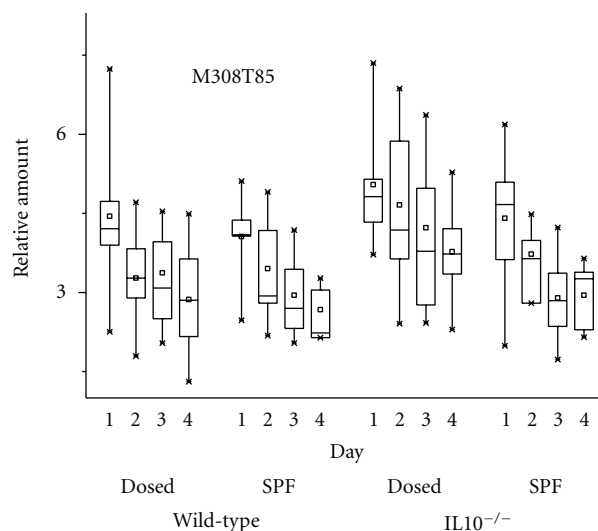


FIGURE 8: Relative concentrations of biomarker M308T85 (m/z 308⁻) measured by UHPLC-MS in urine samples from $IL10^{-/-}$ and wild-type (WT) mice housed either under conventional conditions and dosed with intestinal microflora or maintained under specific pathogen free (SPF) throughout the experiments. Urines samples were collected 7, 19, 29, and 36 days after dosing [17].

acid in urine from $IL10^{-/-}$ mice may result from the absence of negative feedback control by IL10 on the production of kynurenine metabolites.

Xanthurenic acid glucuronide, identified on the basis of its high resolution LC-MS/MS fragmentation pattern, was also a differential metabolite. Concentrations of the glucuronide were lower in urine from $IL10^{-/-}$ mice suggesting a shift to higher free xanthurenic concentrations associated

with colon inflammation. However, the low urinary concentration of the glucuronide in IL10^{-/-} mice kept under SPF conditions suggested its concentration was not related to inflammation status. Similar observations have been made for short chain dicarboxylic acids that occurred at reduced concentrations in the urine of IL10^{-/-} mice regardless of inflammation status [17]. The origin of such metabolic differences was attributed to residual 129P2 embryonic stem cell-derived genetic material flanking the IL10 gene in the IL10^{-/-} C57BL/6 mouse [17, 32]. The UDP glucuronosyl-transferase (UGT1) gene family (<http://www.ensembl.org/>) also lies within this region of embryonic stem cell-derived genetic material and may contribute to this metabolic difference between mouse strains. Xanthurenic acid 8-*O*-sulphate and 8-*O*-glucoside have also been identified in human urine as natriuretic hormones that regulate sodium excretion by the kidney [27]. Xanthurenic acid sulphate was not detected in the present study while xanthurenic acid glucoside (*m/z* 366 [M-H]⁻) was only tentatively detected by UHPLC as the differential metabolite (M366T84) (Table S2); higher in the urine from IL10^{-/-} mice.

α -CEHC is a degradation product of vitamin E normally present in urine [28] where its concentration may reflect that of α -tocopherol and of α -CEHC in the liver [36]. Vitamin E is reduced in serum from Crohn's disease patients possibly as a result of reduced concentrations of serum lipids and of reduced absorption of lipids from the diet [37]. The reduced concentrations of α -CEHC glucuronide observed in urine of IL10^{-/-} mice, and their recovery under SPF conditions, is consistent with reduced availability, degradation, and excretion of vitamin E in mice with impaired intestinal function. Reduced concentrations of α -CEHC glucuronide, and of the related metabolite γ -CEHC β -*D*-glucoside, have also been identified in urine of mice treated with a pregnane X receptor (PXR) activator [38]. It was suggested that these conjugates in urine may be useful urinary biomarkers of PXR activation. PXR is highly expressed in the intestines where it functions as a xenosensor to regulate the expression of metabolic enzymes and the excretion of xenobiotics as well as of toxic endogenous metabolites such as bile acids [39]. PXR activation was shown to be protective in the DSS-induced colitis mouse model of inflammatory bowel disease [40]. In IL10^{-/-} mice, increased intestinal permeability may likewise result in an increased ingress of xenobiotics and toxic endogenous metabolites such as bile acids resulting in activation of PXR and reduced urinary vitamin E metabolites. Urinary α -CEHC glucuronide may thus serve as a good biomarker of intestinal inflammation.

The discovery of α -CEHC glucuronide and other unknown metabolites as potential inflammatory biomarkers illustrates both the power and the limitations of current analytical techniques in metabolomic analysis. Short-Column LCMS analysis enables extremely rapid metabolic profiling of multiple samples with simultaneous collection of characterising MS/MS data. However, the limited separation of individual metabolites hinders the structural and biochemical characterisation of candidate ions and their biological validation. UHPLC-MS allows rapid measurement of chromatographically individual metabolites but

with reduced mass spectral data, so that additional MS experimentation is required to characterise metabolites of interest. The challenge of metabolite identification remains a major constraint but is essential for biological validation of biomarker candidates. Comparison of our results with those obtained from other metabolomics studies of urine from Crohn's disease patients or from animal models indicates how the different analytical tools each brings their own selectivity [18] so that the application of new analytical technology, as here, will reveal new metabolic biomarkers. NMR profiling of urine from Crohn's and ulcerative colitis patients has identified a number of metabolites affected by gut microbes, such as formate, hippurate, and *p*-cresol sulphate, which were able to discriminate between the patient cohorts [5]. Methylhistidine, glycine, guanidoacetate, and citrate were further metabolites important in distinguishing between cohorts. Methylhistamine has previously been reported as a urinary marker of disease activity in Inflammatory Bowel Disease [41]. However, hippurate, *p*-cresol sulphate, methylhistidine, and methylhistamine were not detected as differentiating metabolites in the present study. More recently, Lin et al. using GCMS [16, 17] have proposed xanthurenic acid and fucose as early discriminating metabolites of colon inflammation in the urine of the IL10^{-/-} mouse. This paper extends those observations raising wider questions about the role of tryptophan catabolism and xanthurenic acid conjugates in the regulation of urinary salt balance [27] and inflammation.

Acknowledgments

The authors thank M. Barnett, N. Roy, and R. Broadhurst for their help with mouse experiments and D. Greenwood for FTMS. They also acknowledge funding from the Foundation for Research, Science and Technology (Contract C06X0702) and Nutrigenomics New Zealand (a collaboration between AgResearch Limited, Plant & Food Research, and The University of Auckland).

References

- [1] R. B. Sartor, "Mechanisms of disease: pathogenesis of Crohn's disease and ulcerative colitis," *Nature Clinical Practice Gastroenterology and Hepatology*, vol. 3, no. 7, pp. 390–407, 2006.
- [2] J. C. Barrett, S. Hansoul, D. L. Nicolae et al., "Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease," *Nature Genetics*, vol. 40, no. 8, pp. 955–962, 2008.
- [3] E.-O. Glocker, D. Kotlarz, K. Boztug et al., "Inflammatory bowel disease and mutations affecting the interleukin-10 receptor," *The New England Journal of Medicine*, vol. 361, no. 21, pp. 2033–2045, 2009.
- [4] J. M. E. Fell, M. Paintin, F. Arnaud-Battandieri et al., "Mucosal healing and a fall in mucosal pro-inflammatory cytokine mRNA induced by a specific oral polymeric diet in paediatric Crohn's disease," *Alimentary Pharmacology and Therapeutics*, vol. 14, no. 3, pp. 281–289, 2000.
- [5] H. R.T. Williams, I. J. Cox, D. G. Walker et al., "Characterization of inflammatory bowel disease with urinary metabolic

- profiling," *American Journal of Gastroenterology*, vol. 104, no. 6, pp. 1435–1444, 2009.
- [6] T. Bezabeh, R. L. Somorjai, and I. C.P. Smith, "MR metabolomics of fecal extracts: applications in the study of bowel diseases," *Magnetic Resonance in Chemistry*, vol. 47, supplement 1, pp. S54–S61, 2009.
- [7] J. R. Marchesi, E. Holmes, F. Khan et al., "Rapid and noninvasive metabolomic characterization of inflammatory bowel disease," *Journal of Proteome Research*, vol. 6, no. 2, pp. 546–551, 2007.
- [8] J. Jansson, B. Willing, M. Lucio et al., "Metabolomics reveals metabolic biomarkers of Crohn's disease," *PLoS ONE*, vol. 4, no. 7, Article ID e6386, pp. 1–10, 2009.
- [9] K. Balasubramanian, S. Kumar, R. R. Singh et al., "Metabolism of the colonic mucosa in patients with inflammatory bowel diseases: an in vitro proton magnetic resonance spectroscopy study," *Magnetic Resonance Imaging*, vol. 27, no. 1, pp. 79–86, 2009.
- [10] C. Chen, Y. M. Shah, K. Morimura et al., "Metabolomics reveals that hepatic stearyl-CoA desaturase 1 downregulation exacerbates inflammation and acute colitis," *Cell Metabolism*, vol. 7, no. 2, pp. 135–147, 2008.
- [11] S. Varma, R. Bird, M. Eskin, B. Dolenko, J. Raju, and T. Bezabeh, "Detection of inflammatory bowel disease by proton magnetic resonance spectroscopy (1H MRS) using an animal model," *Journal of Inflammation*, vol. 4, article 24, 2007.
- [12] W. Strober, I. J. Fuss, and R. S. Blumberg, "The immunology of mucosal models of inflammation," *Annual Review of Immunology*, vol. 20, pp. 495–549, 2002.
- [13] K. W. Moore, R. de Waal Malefyt, R. L. Coffman, and A. O'Garra, "Interleukin-10 and the interleukin-10 receptor," *Annual Review of Immunology*, vol. 19, pp. 683–765, 2001.
- [14] F.-P. J. Martin, S. Rezzi, I. Montoliu et al., "Metabolic assessment of gradual development of moderate experimental colitis in IL-10 deficient mice," *Journal of Proteome Research*, vol. 8, no. 5, pp. 2376–2387, 2009.
- [15] T. B. Murdoch, H. Fu, S. MacFarlane, B. C. Sydora, R. N. Fedorak, and C. M. Slupsky, "Urinary metabolic profiles of inflammatory bowel disease in interleukin-10 gene-deficient mice," *Analytical Chemistry*, vol. 80, no. 14, pp. 5524–5531, 2008.
- [16] H.-M. Lin, S. J. Edmunds, N. A. Helsby, L. R. Ferguson, and D. D. Rowan, "Nontargeted urinary metabolite profiling of a mouse model of crohn's disease," *Journal of Proteome Research*, vol. 8, no. 4, pp. 2045–2057, 2009.
- [17] H.-M. Lin, M. Barnett, N. Roy, N. Joyce, and W. S. Zhu, "Metabolomic analysis identifies inflammatory and non-inflammatory metabolic effects of genetic modification in a mouse model of Crohn's disease," *Journal of Proteome Research*, vol. 9, pp. 1965–1975, 2010.
- [18] J. M. Büscher, D. Czernik, J. C. Ewald, U. Sauer, and N. Zamboni, "Cross-platform comparison of methods for quantitative metabolomics of primary metabolism," *Analytical Chemistry*, vol. 81, no. 6, pp. 2135–2143, 2009.
- [19] A. Koulman, M. Cao, M. Faville, G. Lane, W. Mace, and S. Rasmussen, "Semi-quantitative and structural metabolic phenotyping by direct infusion ion trap mass spectrometry and its application in genetical metabolomics," *Rapid Communications in Mass Spectrometry*, vol. 23, no. 15, pp. 2253–2263, 2009.
- [20] A. Koulman, B. A. Tapper, K. Fraser, M. Cao, G. A. Lane, and S. Rasmussen, "High-throughput direct-infusion ion trap mass spectrometry: a new method for metabolomics," *Rapid Communications in Mass Spectrometry*, vol. 21, no. 3, pp. 421–428, 2007.
- [21] G. A. Lane, K. Fraser, M. Cao et al., "Urinary biomarkers of forage feeding from ESI-MS/MS fingerprinting: exploratory studies," *New Zealand Society of Animal Production*, vol. 66, pp. 230–235, 2006.
- [22] N. Roy, M. Barnett, B. Knoch, Y. Dommels, and W. McNabb, "Nutrigenomics applied to an animal model of Inflammatory Bowel Diseases: transcriptomic analysis of the effects of eicosapentaenoic acid- and arachidonic acid-enriched diets," *Mutation Research*, vol. 622, no. 1–2, pp. 103–116, 2007.
- [23] M. Philpott, L. Mackay, L. R. Ferguson, D. Forbes, and M. Skinner, "Cell culture models in developing nutrigenomics foods for inflammatory bowel disease," *Mutation Research*, vol. 622, no. 1–2, pp. 94–102, 2007.
- [24] P. Du, W. A. Kibbe, and S. M. Lin, "Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching," *Bioinformatics*, vol. 22, no. 17, pp. 2059–2065, 2006.
- [25] G. K. Smyth, "Linear models and empirical bayes methods for assessing differential expression in microarray experiments," *Statistical Applications in Genetics and Molecular Biology*, vol. 3, no. 1, article 3, 2004.
- [26] T. Tohge and A. R. Fernie, "Web-based resources for mass-spectrometry-based metabolomics: a user's guide," *Phytochemistry*, vol. 70, no. 4, pp. 450–456, 2009.
- [27] C. D. Cain, F. C. Schroeder, S. W. Shankel, M. Mitchnick, M. Schmertzler, and N. S. Bricker, "Identification of xanthurenic acid 8-O- β -D-glucoside and xanthurenic acid 8-O-sulfate as human natriuretic hormones," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 45, pp. 17873–17878, 2007.
- [28] S. A. S. Pope, G. E. Burtin, P. T. Clayton, D. J. Madge, and D. P. R. Muller, "Synthesis and analysis of conjugates of the major vitamin E metabolite, α -CEHC," *Free Radical Biology and Medicine*, vol. 33, no. 6, pp. 807–817, 2002.
- [29] M. Cao, A. Koulman, L. J. Johnson, G. A. Lane, and S. Rasmussen, "Advanced data-mining strategies for the analysis of direct-infusion ion trap mass spectrometry data from the association of perennial ryegrass with its endophytic fungus, *Neotyphodium lolii*," *Plant Physiology*, vol. 146, no. 4, pp. 1501–1514, 2008.
- [30] S. J. Callister, R. C. Barry, J. N. Adkins et al., "Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics," *Journal of Proteome Research*, vol. 5, no. 2, pp. 277–286, 2006.
- [31] D. S. Wishart, D. Tzur, C. Knox et al., "HMDB: the human metabolome database," *Nucleic Acids Research*, vol. 35, no. 1, pp. D521–D526, 2007.
- [32] A. F. Eisener-Dorman, D. A. Lawrence, and V. J. Bolivar, "Cautionary insights on knockout mouse studies: the gene or not the gene?" *Brain, Behavior, and Immunity*, vol. 23, no. 3, pp. 318–324, 2009.
- [33] A. L. Mellor and D. H. Munn, "IDO expression by dendritic cells: tolerance and tryptophan catabolism," *Nature Reviews Immunology*, vol. 4, no. 10, pp. 762–774, 2004.
- [34] M. L. Belladonna, C. Orabona, U. Grohmann, and P. Puccetti, "TGF- β and kynurenines as the key to infectious tolerance," *Trends in Molecular Medicine*, vol. 15, no. 2, pp. 41–49, 2009.
- [35] J. R. Moffett and M. A. Nambodiri, "Tryptophan and the immune response," *Immunology and Cell Biology*, vol. 81, no. 4, pp. 247–265, 2003.

- [36] S. W. Leonard, E. Gumpricht, M. W. Devereaux, R. J. Sokol, and M. G. Traber, "Quantitation of rat liver vitamin E metabolites by LC-MS during high-dose vitamin E administration," *Journal of Lipid Research*, vol. 46, no. 5, pp. 1068–1075, 2005.
- [37] F. Kuroki, M. Iida, M. Tominaga, T. Matsumoto, K. Kanamoto, and M. Fujishima, "Is vitamin E depleted in Crohn's disease at initial diagnosis?" *Digestive Diseases*, vol. 12, no. 4, pp. 248–254, 1994.
- [38] J.-Y. Cho, D. W. Kang, X. Ma et al., "Metabolomics reveals a novel vitamin E metabolite and attenuated vitamin E metabolism upon PXR activation," *Journal of Lipid Research*, vol. 50, no. 5, pp. 924–937, 2009.
- [39] S. A. Kliewer, B. Goodwin, and T. M. Willson, "The nuclear pregnane X receptor: a key regulator of xenobiotic metabolism," *Endocrine Reviews*, vol. 23, no. 5, pp. 687–702, 2002.
- [40] Y. M. Shah, X. Ma, K. Morimura, I. Kim, and F. J. Gonzalez, "Pregnane X receptor activation ameliorates DSS-induced inflammatory bowel disease via inhibition of NF- κ B target gene expression," *American Journal of Physiology*, vol. 292, no. 4, pp. G1114–G1122, 2007.
- [41] S. Winterkamp, M. Weidenhiller, P. Otte et al., "Urinary excretion of N-methylhistamine as a marker of disease activity in inflammatory bowel disease," *American Journal of Gastroenterology*, vol. 97, no. 12, pp. 3071–3077, 2002.

Research Article

A Metabonomic Approach to Analyze the Dexamethasone-Induced Cleft Palate in Mice

Jinglin Zhou,¹ Bin Xu,¹ Bing Shi,¹ Jing Huang,² Wei He,¹ Shengjun Lu,¹ Junjun Lu,¹ Liying Xiao,¹ and Wei Li^{1,3}

¹State Key Laboratory of Oral Diseases, Sichuan University, Chengdu 610041, China

²West China College of Pharmacy, Sichuan University, Chengdu 610041, China

³State Key Laboratory of Stomatology, Sichuan University, No.14 Sec 3 Renminnan Road, Chengdu 610041, China

Correspondence should be addressed to Wei Li, leewei2000@sina.com

Received 6 February 2010; Revised 3 April 2010; Accepted 8 June 2010

Academic Editor: Mika Ala-Korpela

Copyright © 2011 Jinglin Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Mice models are an important way to understand the relation between the fetus with cleft palate and changes of maternal biofluid. This paper aims to develop a metabonomics approach to analyze dexamethasone-induced cleft palate in pregnant C57BL/6J mice and to study the relationship between the change of endogenous small molecular metabolites in maternal plasma and the incidence of cleft palate. To do so, pregnant mice were randomly divided into two groups. The one group was injected with dexamethasone. On E17.5th day, the incident rates of cleft palate from embryos in two groups were calculated. The ¹H-NMR spectra from the metabolites in plasma in two groups was collected at same time. Then the data were analyzed using metabonomics methods (PCA and SIMCA). The results showed that the data from the two groups displayed distinctive characters, and the incidence of cleft palate were significantly different ($P < .005$). To conclude, this study demonstrates that the metabonomics approach is a powerful and effective method in detecting the abnormal metabolites from mother in the earlier period of embryos, and supports the idea that a change from dexamethasone induced in maternal metabolites plays an important role in the incidence of cleft palate.

1. Introduction

Congenital cleft lip and palate (CLP) is the most common birth defect in humans. The etiology is complicated, and it involves genetic and environmental factors [1, 2]. Maternal condition during pregnancy also appears to play an important role [3]. Smoking [4–6], over intake of vitamin A [7, 8], and deficiency in folic acid and Bs may increase risk for oral clefts [9–12]. In addition, drug-induced (corticosteroid) teratogenesis has also received some attention [13–16]. Most studies above are based on epidemiological investigation. As a useful model, experiments of mouse development are used to assess the mechanism of palate defects in fetuses resulting from exposure to the risk factors. Current strategies to study orofacial defects focus on the related genotype and transcription factors [17, 18]. However, the capability of maternal detoxification during orofacial development is important for normal palate formation. As a result, a fetus lacking genes associated with cleft palates still has

high risk of defect if maternal detoxification is insufficient. [2, 19]

Metabonomics, based on the NMR spectroscopic and multivariate statistics, can be useful for the description and recognition of the dynamic multivariate metabolic response of an organism to a pathological event or genetic modification. The ¹H-NMR spectra of the biofluids from an organism contain a significant amount of useful metabolic information. Application of automated data reduction algorithms and chemometric analysis, which is called pattern recognition analysis (PR), can be competent for the description and recognition of the dynamic multivariate metabolism. It is included two major approaches, one is termed “unsupervised”, which could be used to not only reduce the complex and volume data to a suitable level but also screen for the outlier and examine condition of the clusters, including Principal Component Analysis (PCA), Nonlinear Mapping (NLM) and Hierarchical Cluster Analysis (HCA). The other is termed “supervised” that

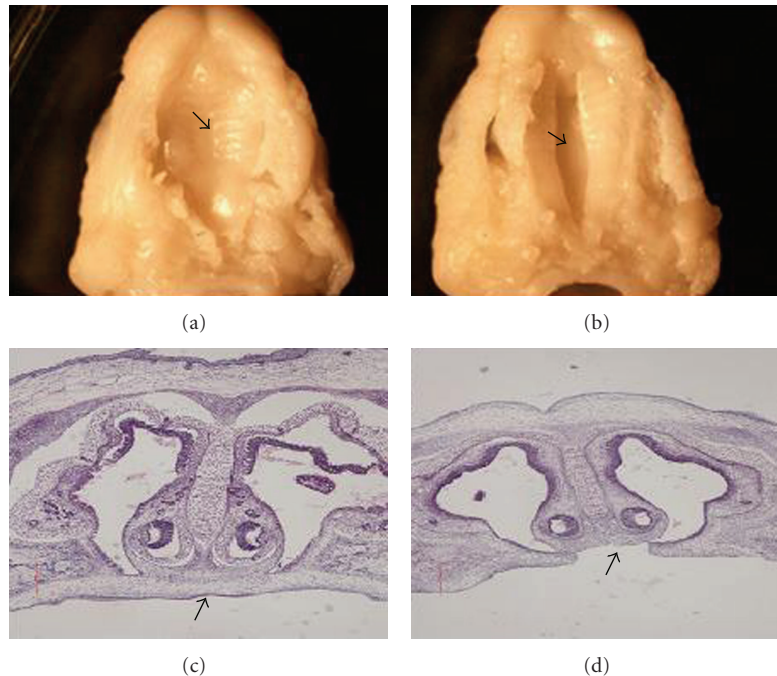


FIGURE 1: Photos of fusion or lack of fusion by embryonic palatal shelves. The fusion situation of embryonic palatal shelves was observed and photographed. (a) Fused embryonic palatal shelf. (b) Nonfused embryonic palatal shelf. (c) Fused embryonic palatal shelf in histological section. (d) Nonfused embryonic palatal shelf in histological section.

means, make a mathematical model by a training set of known class from the samples and then draw the estimated or predictive model by independent validation set. There are many methods such as Soft Independent Modeling of Class Analogy (SIMCA), Partial Least Squares (PLS), Linear discriminant analysis (LDA), K-nearest neighbor analysis (KNN), and Bayesian methods. These approaches have been successfully applied to study diseases and toxic processes [20–27]. Principal component analysis (PCA) as a bilinear decomposition method is one of the most easy and efficient approaches to analyze the NMR data, A PCA model can trustily display a summary of all samples in the data table [27]. Combination with using a prediction model, such as SIMCA, can be constructed to analyze the unknown samples [28–30].

Development of the mammalian palate involves a number of critical steps: growth, elevation, contact, and fusion (the medial edge epithelium disappearance). Any problem arising from these steps can lead to the incidence of cleft palate. DEX, one of the glucocorticoids (GC), can penetrate the blood-placental barrier and bind to GC receptor (GR) in the cytoplasm, and can depress the ability of palatal mesenchymal proliferation. Since the palate is smaller, development of bilateral palates is interrupted, which results in the cleft [31–34]. We hypothesized that the metabonomic method could provide the framework for the studies on the consequences of maternal environmental changes during pregnancy and for illustrating the relationship between the changes of maternal environment and development of palates. To test this hypothesis, we studied the plasma samples from pregnant C57BL/6J mice, which were induced

with dexamethasone (DEX) to trigger cleft palate formation in the fetuses. The analyses demonstrated the relationship between the change in endogenous small molecular metabolites of maternal plasma and the incidence of cleft palate (CP) in fetuses.

2. Materials and Methods

2.1. Animal Handling and Dosing. All animal experimentation was approved by the Animal Research Committee of the West China College of Stomatology (Sichuan University). C57BL/6J (C57) mice, about 8 weeks of age, were obtained from the Laboratory Animal Center of Sichuan University. All mice were reared in plastic cages (28 cm × 16 cm × 12 cm), under a 12/12-hour light/dark cycle. The room was maintained at a controlled ambient temperature of 20–27°C with 40%–70% relative humidity. A commercial diet (Laboratory Animal Center of Sichuan University, China) and tap water were fed ad libitum. Mice were allowed to acclimatize for 48 hours prior to mating. Two virgin females and one male were placed overnight in a cage and checked on following morning for the copulatory plug of female mice. Females with copulation plugs were weighed immediately and the date was designated to Embryonic day 0 (E_0).

Pregnant mice (42) were randomly divided into two groups (the experimental group and the control group each group 21 mice). From the 10th day to the 12th day of pregnancy (From E_{10} to E_{12}), the mice in the experimental group were intraperitoneally injected daily with Dexamethasone (DEX dexamethasone sodium phosphate injection, Tianjing China) at 6 mg/kg and while the others in the control group

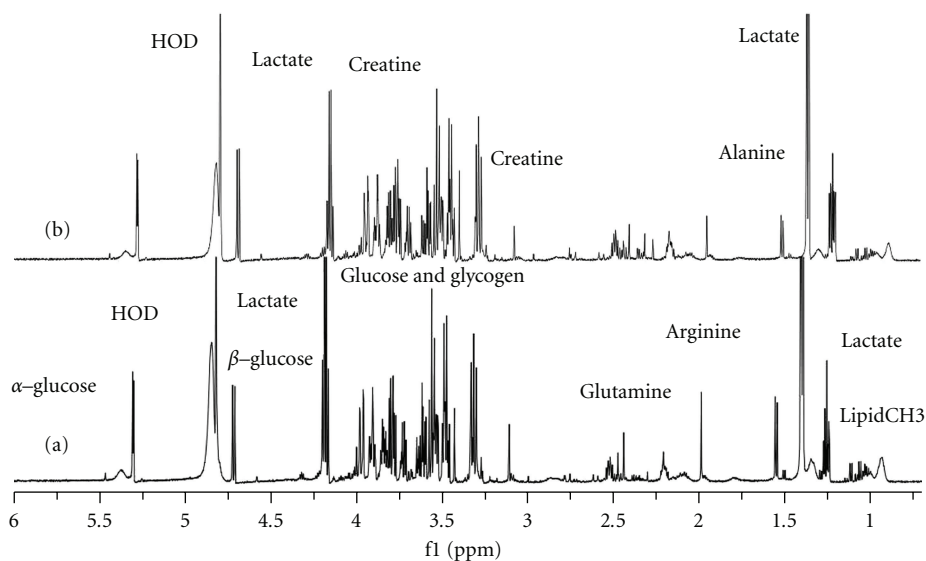


FIGURE 2: 600 MHz $^1\text{H-NMR}$ spectra of plasma samples from the pregnant mice in two groups: “a” showed the spectrum from one sample in the control group; “b” showed the spectrum from one sample in the DEX group.

with isotonic sodium chloride (0.9% NaCl). The dose levels were based on the literature report [35].

2.2. Sample Collection. On the day of $E_{17.5}$, all the pregnant mice were sacrificed and weighed. The peripheral blood was placed in a lithium heparin tube (Vacuette Austria) immediately. Within 2 hours of the collection, the erythrocytes were separated by centrifugation ($3000 \times g$ at 4°C or 10 min) to pick up the biofluid top of the plasma for analysis. All the plasma samples were stored at -80°C before the $^1\text{H-NMR}$ analysis. At the same time, a stereomicroscope was used to determine the number of live fetuses and gross malformations.

2.3. Statistical Analysis of the Incidence of Cleft Palate. After removal of the mandibles, embryonic palates were calculated using stereomicroscope ($10\times$) by morphologic in two groups. Target palates were processed into paraffin wax. Histological sections of embryonic palates were cut and stained with haematoxylin and eosin (HE). The results were photographed by a Leica photographic system (Germany). The numbers of CP in two groups were analyzed to detect the different incidence rates by using chi-square test and variance-based statistical algorithm (SPSS 10.0).

2.4. Preparation of Plasma Samples for $^1\text{H-NMR}$ Spectroscopy. The $^1\text{H-NMR}$ spectra were obtained immediately after each sample was thawed at 300K and then kept at the same temperature in a Bruker Avance II 600 spectrometer (Bruker Biospin, Germany), which was operating at 600.13 MHz with a 5-mm PATXI probe. In order to obtain a deuterium lock signal for the NMR spectrometry, $200\ \mu\text{l}$ aliquots of all the plasma were diluted up to $500\ \mu\text{l}$ with $300\ \mu\text{l}$ of Deuterium oxide (D_2O).

2.5. Acquisition of $^1\text{H-NMR}$ Spectra of Plasma. The spectrum was obtained by using two different pulse sequences: one is selective presaturation pulse sequence (Bruker Biospin, Germany) for water suppression (located in δ 4.8 ppm) and other is CPMGPR1D pulse sequence (Bruker Biospin, Germany). The later one, which was used to attenuate the broad protein signal in the plasma, is a modification of the Carr-Purcell-Meiboom-Gill pulse sequence (CPMG-pulse, Bruker Biospin Germany) to suppress the residual water signal. For each sample, 1D $^1\text{H-NMR}$ spectrum was collected with 64 K data points, 64 scans and 15 ppm spectral width. Other acquisition parameters were 5 s relaxation delay, 8 dummy scans, $400\ \mu\text{s}$ fixed echo time to allow the elimination of J-mod, and 400 CPMG loops for T_2 filter. The phase and the baseline of all acquired NMR spectra were manually calibrated, and the chemical shifts corrected by the reference of the lactate doublet at δ 1.32 [36] with TopSpin 1.3 (Bruker Biospin, Germany).

2.6. Reduction of the NMR Spectral Data. All signals of the 42 samples' were located in the range of δ 0–7 ppm in spectral region and have no visible resonance peak after δ 7 ppm. Using MestReC (version 4.8.1.1, Spain), each $^1\text{H-NMR}$ spectrum was divided with δ 0.04 ppm width into 162 contiguous segments and integrated from the range of δ 0–7 ppm [37, 38]. The region of the spectrum (δ 5.0–4.5 ppm) was removed to exclude the influence of the water signal. The result was set up to a kind of 2D matrix ($n \times d$), “ n ” representing the 42 samples, “ d ” meaning 162 contiguous segments. The matrix data were normalized to the unit area with the appropriate weighting coefficients in Excel (Microsoft USA) and then exported into the SIMCA-P software package (version 11.0, Umetrics AB, Umeå, Sweden). The average value of each variable was calculated and subtracted from the data.

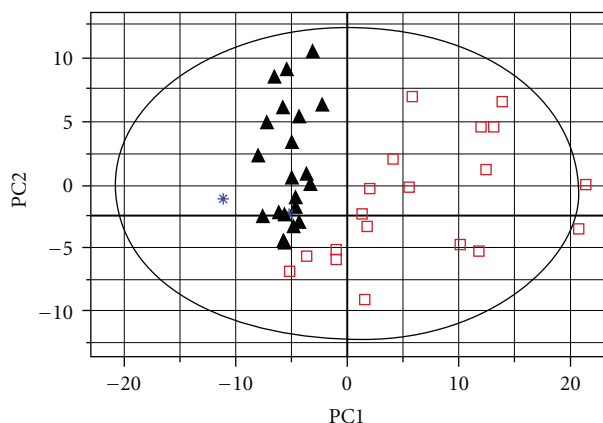


FIGURE 3: The plot of PCA analysis from two groups. “▲” showed the data from pregnant mice with CP fetuses in DEX group; “□” showed the data from pregnant mice with fusion fetuses in control group. “*” showed the pregnant mice including CP fetuses in the control group.

TABLE 1: Comparison of the incidence of cleft palate in the embryos between two groups.

| | Non-Fusion | Fusion | Total | Incidence of non-fusion (%) |
|---------------|------------|--------|-------|-----------------------------|
| Control group | 3 | 103 | 106 | 2.83 |
| DEX group | 27 | 96 | 123 | 21.95* |
| Total | 30 | 199 | 229 | 13.10 |

*Significant difference with respect to control ($\chi^2 = 18.285$, $P < .005$).

2.7. Statistical Analysis of the Data from NMR Spectra. After the average value being calculated from each variable and subtracted from the data, PCA was applied to the mean-centered data to detect the differences among the samples. SIMCA was performed to validate the results. Firstly, separate PC models were built for the training sets of 10 samples from DEX group with CP fetuses and 10 samples including only normal fetus of control group and were defined two distances against boundaries (cross-validation) by the 95% confidence interval. As a typically visualizing method, the Cooman’s plot was often used to display the result [39]. SIMCA formulated a PC model for each distinct class and estimated its performance. Secondly, the procedure was repeated for all samples (the test sets) in turn, to predict and assess whether samples were separated into two classes. Finally, the result of the Cooman’s plot illustrated the visible separation of two groups. Each step was repeated five times in order to keep the validation set.

3. Results

3.1. The Statistics of the Incidence of Cleft Palate in Two Groups. Forty-two C57BL/6J pregnant mice were examined, and 229 fetuses were collected, including 123 fetuses from DEX treatment group and 106 fetuses from the control group (Table 1). Each pregnant mouse had multiple fetuses. There was at least one CP fetus in each pregnant mouse in DEX treatment group (named DEX group), while there were 3 CP fetuses in two pregnant mice of the control group (named the control group with CP). The incidence of CP in the DEX treatment group was 21.95%, significantly different from the control group ($P < .005$). The character of morphology and

histology were recorded (Figure 1). “a” and “c” show fusion of palate shelves, while “b” and “d” show a failure of the palate shelves to fuse.

3.2. $^1\text{H-NMR}$ Spectra from Two Groups. The $^1\text{H-NMR}$ spectrum of the plasma from pregnant mice revealed great complexity and significant information about the biofluid (Figure 2). The chemical shift of δ 4.8 was the water signal. The regions of the significant metabolite signals usually ranged δ 0–4.5 ppm, including alkene (δ 4–5.5 ppm), alkyne (δ 1.8–3.5 ppm), and the aliphatic series (δ 0–2.8 ppm). Articles of the locations of different chemical groups and some low-molecular-weight metabolites had already been published [36, 40]. To detect the difference between the two groups, special software was used in the chemometric analysis.

3.3. PCA of the Plasma in the DEX Group and the Control Group. The matrix data from the $^1\text{H-NMR}$ spectra were exported into the SIMCA-P software and processed by PCA. The new principal component (PC) variables were created (6 new contributories of the PCs), which accounted for 84.8% of the original data and each PC was orthogonal with all the other PCs. The result of PCA could be shown (Figure 3) as the clustering of the individual dataset defined by the 95% Hotelling’s T2 Limit. Except for two control samples (*: the blue star), all the 21 DEX samples (▲: the black triangles) and 19 control samples with fusion fetuses (□: the open squares), obviously, were separated well along the first principal component. It indicated that the major difference between the two groups occurred in the first two principal

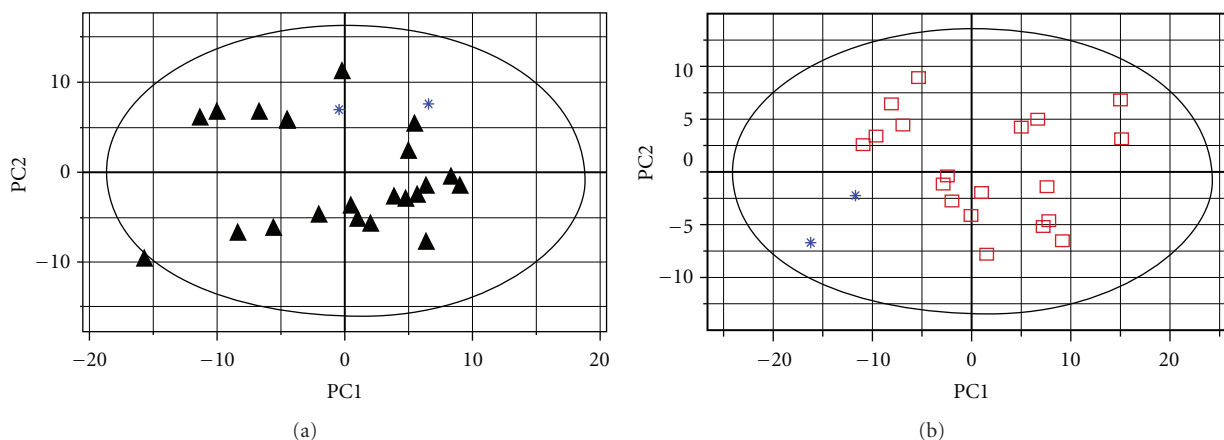


FIGURE 4: The plots of PCA analysis for two control pregnant mice including CP fetuses in different groups “a” showed the PCA analysis between two outliers from the control group and the data of pregnant mice with CP fetuses in DEX group; “b” showed the PCA analysis between two outliers from the control group and the data of pregnant mice with no-CP fetuses in control group.

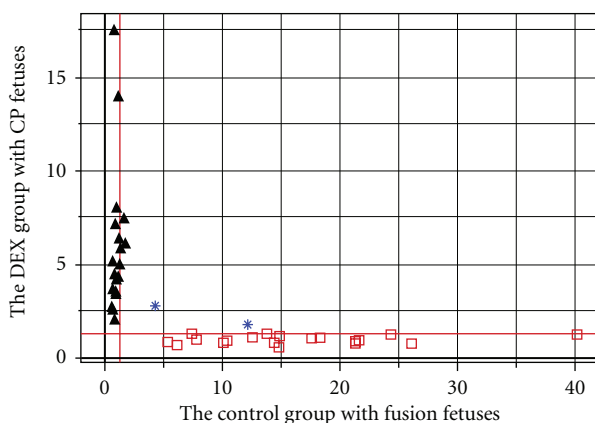


FIGURE 5: The Cooman's plot of two groups in test set from SIMCA. “▲” showed the DEX group with CP fetuses; “□” showed the control group with fusion fetuses; “*” showed two outlier samples (pregnant mice with CP fetuses in control group); the red lines indicate the D-Crit (0.05) level; two groups were located in two different spaces; the outliers not belong to either group.

components. The detailed different biochemical changes were informed by the loadings (Table 2). In order to detect the outliers, the two outliers as one group to be detected with Dex group and control group respectively. PCAs were repeated twice again using different groups (Figure 4). The result indicated that the two outliers with CP fetuses have some extent different from the control group, while mixed in the Dex-injected group because of CP fetuses.

3.4. SIMCA of the Plasma in the DEX Group and the Control Group. The result of SIMCA was shown in the Cooman's plot (Figure 5). All the samples had been tested with an independent model made of train set. The condition of samples can be classified in a pair of distances against boundaries (red lines) for non-fusion data in DEX group and fusion data in control group. The plot contained four regions separated by 95% confidence limits in a 2D graph in which samples below the horizontal line belonged to the control group with fusion fetuses (95%), while samples in the left region of the vertical line belonged to DEX group with CP

fetuses (95%). The data from control group with CP fetuses (the outliers) neither belongs to DEX group nor belongs to control group. It demonstrated that plasma classes from pregnant mice with CP fetuses and normal control did not share the same space. Therefore, this model should be able to predict whether maternal environment has been affected by DEX or not.

4. Discussion

In the normal mouse embryo, palate shelves grow and elevate into a horizontal position by embryonic day 14 (E14). By the day of E17, the process of fusion has completely finished. We chose to collect samples on E17.5 for the following reasons. It was convenient to use morphology for embryo analysis, and it was easy to display the change of maternal metabolism without the influence of DEX [41]. Furthermore, another study also proved that the spectra of plasma from Dex-injected and control male mice showed no significant difference during different time frames

TABLE 2: the main metabolites changes that were partially responsible for discriminating Dex group from control using PCA.

| ¹ H bucket region (δ) | metabolites | Changes in Dex group |
|---|--|----------------------|
| 1.32, 4.12 | Lactate | ↑* |
| 1.36 | unknown | ↑ |
| 1.48 | Alanine CH ₃ | ↑ |
| 1.72 | Arginine(γ -H) | ↓ |
| 1.92 | Arginine (β -H) | ↓ |
| 2.02 | Acetate | ↑ |
| 2.04 | Lipid CH ₂ CH ₂ CH=CH | ↑ |
| 2.12 | Glutamine | ↑ |
| 2.28 | Valine (β -H) | ↓ |
| 2.36 | Glutamate | ↑ |
| 2.4801 | unknown | ↑ |
| 2.5201 | Lipid -C=C-CH ₂ -C=C | ↓ |
| 2.8 | Aspartic acid (β -H) | ↓ |
| 3.24 | Choline N (CH ₃) ₃ ⁺ | ↑ |
| 3.64 | Valine (α -H) | ↓ |
| 3.92 | Creatine | ↑ |
| 4.04 | myo-inositol | ↓ |

* ↑, relative increase in signal by loading plot.

(2 hours, 1 day, 2 days, 3 days, and 7 days). The PCAs also show that all data from different time frames were a mixture together, which accounted for 89.9% from the original data (supplement).

Different murine strains also have different susceptibilities to the DEX-induced cleft palate. The incidence in our model was in accord with previous work [42]. Some studies indicate [43–45] that the incidence of cleft palate may be closely related to high maternal concentration of plasma homocysteine or lower activity of the glucocorticoids prereceptor metabolizing enzyme 11 β -hydroxysteroid dehydrogenase type 2 (11 β HSD2) in placental trophoblastic cells. However, the exact mechanisms by which interaction between maternal environment and DEX causes teratogenic effects are poorly understood. NMR spectroscopy-based metabonomic approach offers a unique opportunity to focus on the relationship between genotype and phenotype and is especially suited to uncover changes in drug-induced metabolites [46, 47]. In this study, each mother at least has one CP fetus in DEX-injected group. Some changes of maternal metabolites informed us about the abnormality of the embryo. Therefore, the result from two groups displayed the different metabolites between CP fetus and fusion one. From the spectra of ¹H-NMR (Figure 2), we can see clearly some differences between the samples of the two groups within the range of chemical shift from δ 1.8 to δ 3.5 ppm and the range of glucose and amino acid CH. The PCA plots showed good separation of the two groups except for two control samples (Figure 3), which matched the results from morphologic analysis identifying three cleft palate fetuses from two different individual mice in the control group. In order to analyze the outlier from the control group, the two outliers as a group were analyzed again. The result also show

these two samples were closely blended in the DEX group (Figure 4(a)), while far away from the controls' (Figure 4(b)). Some biochemical changes such as decrease of Arginine (1.72, 1.92 ppm), increase of Alanine (1.48 ppm), N-acetyl glycoprotein (2.12 ppm), Choline (3.24 ppm), and Creatine (3.92 ppm) may response to the change of methyl group metabolism in maternal environment. The key donor of methylation is S-adenosyl methionine (SAM) which comes from methionine. Arginine and Creatine may increase ATP to supply the circle from methionine to homocysteine [48]. Normally, the incidence of cleft palate in C57BL/6J natural mice is zero. The reason may be that we injected isotonic sodium chloride (0.9% NaCl) in control samples in order to keep the same stimulating circumstance as in the DEX group. After SIMCA analysis, the result of the Cooman's plot revealed a good fit model to detect the NMR data and separate the CP group from the controls by using a test set of *F*-statistic. Both plots clearly showed a relationship between the fetus' cleft palate and the plasma metabolic profile of the mother. In conclusion, this study provides evidence that metabonomic method is sufficiently sensitive to detect small differences in the composition of maternal biofluids and may be helpful for identifying biomarkers of teratogenesis.

Conflict of Interest Statement

The authors declare no conflict of interests.

Abbreviations

| | |
|--------|--|
| PRA: | Pattern recognition analysis |
| DEX: | Dexamethasone |
| HE: | Haematoxylin and eosin |
| CPMG: | Carr-Purcell-Meiboom-Gill pulse sequence |
| PCA: | The principal component analysis |
| SIMCA: | Soft independent modeling of class analogy |
| GC: | Glucocorticoids |
| GR: | GC receptor |
| CP: | Cleft palate. |

Acknowledgments

This study was supported by Grants from the Key Research Project of the Ministry of Education in China (no. 307022), the National Natural Science Foundation of China (no. 30530730) and the Youth Teacher Science Foundation of Sichuan University (no. 2009SCU11162). The authors are grateful to Dr. Shan Lu, PhD and Chapin Rodriguez, PhD for their useful comments.

References

- [1] A. Jugessur and J. C. Murray, "Orofacial clefting: recent insights into a complex trait," *Current Opinion in Genetics and Development*, vol. 15, no. 3, pp. 270–278, 2005.
- [2] A. C. Lidral, L. M. Moreno, and S. A. Bullard, "Genetic factors and orofacial clefting," *Seminars in Orthodontics*, vol. 14, no. 2, pp. 103–114, 2008.

- [3] J. S. Zeiger and T. H. Beaty, "Is there a relationship between risk factors for oral clefts?" *Teratology*, vol. 66, no. 5, pp. 205–208, 2002.
- [4] K. Källén, "Maternal smoking and orofacial clefts," *Cleft Palate-Craniofacial Journal*, vol. 34, no. 1, pp. 11–16, 1997.
- [5] K. C. Chung, C. P. Kowalski, H. M. Kim, and S. R. Buchman, "Maternal cigarette smoking during pregnancy and the risk of having a child with cleft lip/palate," *Plastic and Reconstructive Surgery*, vol. 105, no. 2, pp. 485–491, 2000.
- [6] J. Little, A. Cardy, and R. G. Munger, "Tobacco smoking and oral clefts: a meta-analysis," *Bulletin of the World Health Organization*, vol. 82, no. 3, pp. 213–218, 2004.
- [7] R. H. Finnell, G. M. Shaw, E. J. Lammer, K. L. Brandl, S. L. Carmichael, and T. H. Rosenquist, "Gene-nutrient interactions: importance of folates and retinoids during early embryogenesis," *Toxicology and Applied Pharmacology*, vol. 198, no. 2, pp. 75–85, 2004.
- [8] R. G. Munger, "Maternal nutrition and oral clefts," in *Cleft Lip and Palate: From Origin to Treatment*, D. F. E. Wyszynski, Ed., pp. 170–192, Oxford University Press, Oxford, UK, 2002.
- [9] G. M. Shaw, S. L. Carmichael, C. Laurent, and S. A. Rasmussen, "Maternal nutrient intakes and risk of orofacial clefts," *Epidemiology*, vol. 17, no. 3, pp. 285–291, 2006.
- [10] G. M. Shaw, E. J. Lammer, C. R. Wasserman, C. D. O'Malley, and M. M. Tolarova, "Risks of orofacial clefts in children born to women using multivitamins containing folic acid preconceptionally," *Lancet*, vol. 346, no. 8972, pp. 393–396, 1995.
- [11] L. D. Botto, R. S. Olney, and J. D. Erickson, "Vitamin supplements and the risk for congenital anomalies other than neural tube defects," *American Journal of Medical Genetics*, vol. 125, no. 1, pp. 12–21, 2004.
- [12] I. P. C. Krapels, I. A. L. M. Van Rooij, M. C. Ocké, C. E. West, C. M. A. M. Van Der Horst, and R. P. M. Steegers-Theunissen, "Maternal nutritional status and the risk for orofacial cleft offspring in humans," *Journal of Nutrition*, vol. 134, no. 11, pp. 3106–3113, 2004.
- [13] S. L. Carmichael, G. M. Shaw, C. Ma, M. M. Werler, S. A. Rasmussen, and E. J. Lammer, "Maternal corticosteroid use and orofacial clefts," *American Journal of Obstetrics and Gynecology*, vol. 197, no. 6, pp. 585.e1–585.e7, 2007.
- [14] M. Melnick, T. Jaskoll, and H. C. Slavkin, "Corticosteroid-induced cleft palate in mice and H-2 haplotype: maternal and embryonic effects," *Immunogenetics*, vol. 13, no. 5, pp. 443–450, 1981.
- [15] R. Azziz and R. L. Ladda, "Dexamethasone receptor levels in palatal and lung fibroblasts of adult A/J and C57BL/6J mice: relationship to glucocorticoid-induced cleft palate," *Cleft Palate Journal*, vol. 27, no. 4, pp. 388–391, 1990.
- [16] L. Pinsky and A. M. DiGeorge, "Cleft palate in the mouse: a teratogenic index of glucocorticoid potency," *Science*, vol. 147, no. 3656, pp. 402–403, 1965.
- [17] T. L. Thoma, E. Glover, and C. A. Bradfield, "A maternal Ahr null genotype sensitizes embryos to chemical teratogenesis," *Journal of Biological Chemistry*, vol. 279, no. 29, pp. 30189–30194, 2004.
- [18] T. Hamachi, Y. Sasaki, K. Hidaka, and M. Nakata, "Association between palatal morphogenesis and Pax9 expression pattern in CL/Fr embryos with clefting during palatal development," *Archives of Oral Biology*, vol. 48, no. 8, pp. 581–587, 2003.
- [19] T. Senda, N. Natsume, J. Kuno, T. Toyoda, and K. Shimozato, "Rate of occurrence of dexamethasone-induced cleft palate affected by uterine environment in the mouse," *Plastic and Reconstructive Surgery*, vol. 115, no. 4, pp. 1208–1210, 2005.
- [20] O. Fiehn, "Metabolomics—the link between genotypes and phenotypes," *Plant Molecular Biology*, vol. 48, no. 1–2, pp. 155–171, 2002.
- [21] B. C. M. Potts, A. J. Deese, G. J. Stevens, M. D. Reily, D. G. Robertson, and J. Theiss, "NMR of biofluids and pattern recognition: assessing the impact of NMR parameters on the principal component analysis of urine from rat and mouse," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 26, no. 3, pp. 463–476, 2001.
- [22] J. L. Griffin, H. J. Williams, E. Sang, K. Clarke, C. Rae, and J. K. Nicholson, "Metabolic profiling of genetic disorders: a multitissue ^1H nuclear magnetic resonance spectroscopic and pattern recognition study into dystrophic tissue," *Analytical Biochemistry*, vol. 293, no. 1, pp. 16–21, 2001.
- [23] J. T. Brindle, H. Antti, E. Holmes et al., "Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using ^1H -NMR-based metabolomics," *Nature Medicine*, vol. 8, no. 12, pp. 1439–1444, 2002.
- [24] M. E. Bollard, E. Holmes, J. C. Lindon et al., "Investigations into biochemical changes due to diurnal variation and estrus cycle in female rats using high-resolution ^1H -NMR spectroscopy of urine and pattern recognition," *Analytical Biochemistry*, vol. 295, no. 2, pp. 194–202, 2001.
- [25] N. J. Waters, E. Holmes, C. J. Waterfield, R. D. Farrant, and J. K. Nicholson, "NMR and pattern recognition studies on liver extracts and intact livers from rats treated with α -naphthylisothiocyanate," *Biochemical Pharmacology*, vol. 64, no. 1, pp. 67–77, 2002.
- [26] E. Holmes and H. Antti, "Chemometric contributions to the evolution of metabolomics: mathematical solutions to characterising and interpreting complex biological NMR spectra," *Analyst*, vol. 127, no. 12, pp. 1549–1557, 2002.
- [27] J. Taylor, R. D. King, T. Altmann, and O. Fiehn, "Application of metabolomics to plant genotype discrimination using statistics and machine learning," *Bioinformatics*, vol. 18, supplement 2, pp. S241–S248, 2002.
- [28] J. K. Nicholson, J. C. Lindon, and E. Holmes, "Metabolomics: understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data," *Xenobiotica*, vol. 29, no. 11, pp. 1181–1189, 1999.
- [29] J. K. Nicholson, J. Connelly, J. C. Lindon, and E. Holmes, "Metabolomics: a platform for studying drug toxicity and gene function," *Nature Reviews Drug Discovery*, vol. 1, no. 2, pp. 153–161, 2002.
- [30] J. C. Lindon, E. Holmes, and J. K. Nicholson, "Pattern recognition methods and applications in biomedical magnetic resonance," *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 39, no. 1, pp. 1–40, 2001.
- [31] R. M. Greene and D. M. Kochhar, "Some aspects of corticosteroid induced cleft palate: a review," *Teratology*, vol. 11, no. 1, pp. 47–55, 1975.
- [32] J. F. Hackney, "A glucocorticoid receptor in fetal mouse: its relationship to cleft palate formation," *Teratology*, vol. 21, no. 1, pp. 39–51, 1980.
- [33] R. M. Pratt, E. L. Perry, L. M. Chapman, and E. H. Goulding, "Glucocorticoid teratogenesis in mouse whole embryo culture," *Teratology*, vol. 30, no. 1, pp. 71–81, 1984.
- [34] M. Melnick, T. Jaskoll, and H. C. Slavkin, "Corticosteroid-induced cleft palate in mice and H-2 haplotype: maternal and embryonic effects," *Immunogenetics*, vol. 13, no. 5, pp. 443–450, 1981.
- [35] Y. Eishi, K. Hirokawa, and S. Hatakeyama, "Long-lasting impairment of immune and endocrine systems of offspring

- induced by injection of dexamethasone into pregnant mice," *Clinical Immunology and Immunopathology*, vol. 26, no. 3, pp. 335–349, 1983.
- [36] H. Tang, Y. Wang, J. K. Nicholson, and J. C. Lindon, "Use of relaxation-edited one-dimensional and two dimensional nuclear magnetic resonance spectroscopy to improve detection of small metabolites in blood plasma," *Analytical Biochemistry*, vol. 325, no. 2, pp. 260–272, 2004.
- [37] E. Holmes, A. W. Nicholls, J. C. Lindon et al., "Development of a model for classification of toxin-induced lesions using ¹H-NMR spectroscopy of urine combined with pattern recognition," *NMR in Biomedicine*, vol. 11, no. 4-5, pp. 235–244, 1998.
- [38] E. Holmes, A. W. Nicholls, J. C. Lindon et al., "Chemometric models for toxicity classification based on NMR spectra of biofluids," *Chemical Research in Toxicology*, vol. 13, no. 6, pp. 471–478, 2000.
- [39] D. Coomans, I. Broeckaert, M. P. Derde, A. Tassin, D. L. Massart, and S. Wold, "Use of microcomputer for the definition of multivariate confidence regions in medical diagnosis based on clinical laboratory profiles," *Computers and Biomedical Research*, vol. 17, no. 1, pp. 1–14, 1984.
- [40] M. E. Bollard, E. G. Stanley, J. C. Lindon, J. K. Nicholson, and E. Holmes, "NMR-based metabonomic approaches for evaluating physiological influences on biofluid composition," *NMR in Biomedicine*, vol. 18, no. 3, pp. 143–162, 2005.
- [41] F. X. R. van Leeuwen, Dexamethasone, IPCS INCHEM, <http://www.inchem.org/documents/jecfa/jecmono/v33je09.htm>.
- [42] R. Azziz and R. L. Ladda, "Dexamethasone receptor levels in palatal and lung fibroblasts of adult A/J and C57BL/6J mice: relationship to glucocorticoid-induced cleft palate," *Cleft Palate Journal*, vol. 27, no. 4, pp. 388–392, 1990.
- [43] I. A. L. M. Van Rooij, D. W. Swinkels, H. J. Blom, H. M. W. M. Merkus, and R. P. M. Steegers-Theunissen, "Vitamin and homocysteine status of mothers and infants and the risk of nonsyndromic orofacial clefts," *American Journal of Obstetrics and Gynecology*, vol. 189, no. 4, pp. 1155–1160, 2003.
- [44] M. Vujkovic, M. C. Ocke, P. J. Van Der Spek, N. Yazdanpanah, E. A. Steegers, and R. P. Steegers-Theunissen, "Maternal western dietary patterns and the risk of developing a cleft lip with or without a cleft palate," *Obstetrics and Gynecology*, vol. 110, no. 2, pp. 378–384, 2007.
- [45] L. S. Kerzner, B. S. Stonestreet, K.-Y. Wu, G. Sadowska, and M. P. Malee, "Antenatal dexamethasone: effect on ovine placental 11 β -hydroxysteroid dehydrogenase type 2 expression and fetal growth," *Pediatric Research*, vol. 52, no. 5, pp. 706–712, 2002.
- [46] M. E. Bollard, H. C. Keun, O. Beckonert et al., "Comparative metabonomics of differential hydrazine toxicity in the rat and mouse," *Toxicology and Applied Pharmacology*, vol. 204, no. 2, pp. 135–151, 2005.
- [47] S. Vangala and A. Tonelli, "Biomarkers, metabonomics, and drug development: can inborn errors of metabolism help in understanding drug toxicity?" *AAPS Journal*, vol. 9, no. 3, pp. E284–E297, 2007.
- [48] D. H. Baker and G. L. Czarnecki, "Transmethylation of homocysteine to methionine: efficiency in the rat and chick," *Journal of Nutrition*, vol. 115, no. 10, pp. 1291–1299, 1985.

Research Article

Metabolomics Reveals Relationship between Plasma Inositols and Birth Weight: Possible Markers for Fetal Programming of Type 2 Diabetes

Pia Marlene Nissen, Caroline Nebel, Niels Oksbjerg, and Hanne Christine Bertram

Department of Food Science, Faculty of Agricultural Sciences, Research Centre Aarslev, Aarhus University, Kirstinebjergvej 10, 5792 Aarslev, Denmark

Correspondence should be addressed to Hanne Christine Bertram, hannec.bertram@agrsci.dk

Received 19 March 2010; Revised 20 April 2010; Accepted 3 June 2010

Academic Editor: Mika Ala-Korpela

Copyright © 2011 Pia Marlene Nissen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Epidemiological studies in man and with experimental animal models have shown that intrauterine growth restriction (IUGR) resulting in low birth weight is associated with higher risk of programming welfare diseases in later life. In the pig, severe IUGR occurs naturally and contribute substantially to a large intralitter variation in birth weight and may therefore be a good model for man. In the present paper the natural form of IUGR in pigs was studied close to term by nuclear magnetic resonance (NMR)-based metabolomics. The NMR-based investigations revealed different metabolic profiles of plasma samples from low-birth weight (LW) and high-birth weight (HW) piglets, respectively, and differences were assigned to levels of glucose and myo-inositol. Further studies by GC-MS revealed that LW piglets had a significant higher concentration of myoinositol and D-chiro-inositol in plasma compared to larger littermates. Myo-inositol and D-chiro-inositol have been coupled with glucose intolerance and insulin resistance in adults, and the present paper therefore suggests that IUGR is related to impaired glucose metabolism during fetal development, which may cause type 2 diabetes in adulthood.

1. Introduction

It is well established that decreased growth during fetal development, leading to intrauterine growth retardation (IUGR) and consequently low birth weight, has crucial influence on health later in life, and is documented in population studies [1]. However, the relationship between birth weight and health later on in life cannot be described by a simple linear relationship but seems to be U-shaped [2]. The hypothesis that poor fetal growth increases the risk of developing metabolic disorder, like type 2 diabetes, coronary heart disease, elevated blood pressure, and obesity, in adult life, was first put forward by Hales and Barker[3]. Long-term epidemiological studies in humans have shown a relationship between birth weight and adult health, and later experimental studies mainly in animal models of IUGR have documented the original hypothesis[4]. The term fetal

metabolic programming is generally accepted to describe the phenomenon of the long-term effects of a stimulus or insult during fetal development [4].

Several different experimental animal models have been used to study fetal metabolic programming. Accordingly, maternal metabolism during pregnancy has been manipulated, and the effects on the offspring have been investigated [5]. Maternal calorie restriction throughout gestation in guinea pigs has a negative effect on fetal growth, and postnatal glucose tolerance tests demonstrated a decreased glucose tolerance and increased fasting plasma insulin levels, suggesting insulin resistance [6]. Also maternal protein deprivation in rats showed an alteration in the glucose metabolism in the liver of the offspring [7]. Structural changes in the liver of offspring from protein deprived mothers were observed in the same study. Placental restriction induced by surgery caused reduced fetal growth, increased

adiposity postnatal, and impaired glucose-stimulated insulin production in young sheep offspring [8]. Hyperinsulinemia produced in the fetal rhesus monkey during the last third of gestation indicated that insulin is important for fetal weight characteristics [9]. The cause of the naturally occurring IUGR is not fully understood, but decreased placental growth and efficiency for nutrient transfer seems to be important [10]. In litter-bearing species, like rat, mice, and pig, a naturally occurring form of IUGR are present. In these species, low birth weight animals have been coupled with retarded postnatal growth, hypertension and glucose intolerance [11]. Thus, studies in pigs showed that low birth weight is associated with glucose intolerance at 1 year of age [12]. It has been suggested that impairments in early cell development result in fetal malnutrition and predispose individuals to development of type 2 diabetes later in life [13]. An alternative hypothesis suggests that genetic variants predisposing the type 2 diabetes phenotype might also reduce birth weight by altering intrauterine insulin secretion or action [14]. Other factors which likely explain fetal programming of adult health comprise changes in DNA methylation, increased apoptosis in the developing kidney, alterations in renal renin-angiotensin system activity, and increased fetal glucocorticoid exposure [15]. Nevertheless, even though the impact of fetal metabolic programming on adult health is well documented, the underlying mechanisms are poorly understood.

In the present study, the naturally occurring form of IUGR in the pig was used as an experimental model for fetal metabolic programming. The objective was to identify possible mechanisms during fetal development that can couple metabolism during fetal life with later development of the metabolic disorders.

2. Materials and Methods

2.1. Animals. Offspring from 6 Danish Landrace sows mated with one of 6 Danish Landrace boars were used in this study. After mating, the sows were reared under normal production conditions at the Faculty of Agricultural Sciences, Aarhus University, Denmark, until day 110 of gestation. Gestation length in pigs is 113–115 days. At day 110 of gestation sows were stunned using a captive bolt pistol, and immediately after bleeding the uterus was taken out. The umbilical cord of each fetus/piglet was cut and as much blood as possible was collected through the umbilical cord. Blood was used for production of plasma, which was kept at -80°C until analysis was performed. Piglets were weighed and their position within the uterus horns recorded. Organs were weighed and some anatomical measures recorded. All procedures were carried out after permission from the Danish Animal Experiments Inspectorate.

2.2. Study Design. The data reported in this study is for a total of 24 piglets. Within each of the 6 litters, plasma from the 2 piglets with the lowest birth weight (LW) and the 2 piglets with the highest birth weight (HW) were analyzed by NMR and GC-MS methods as described in what follows.

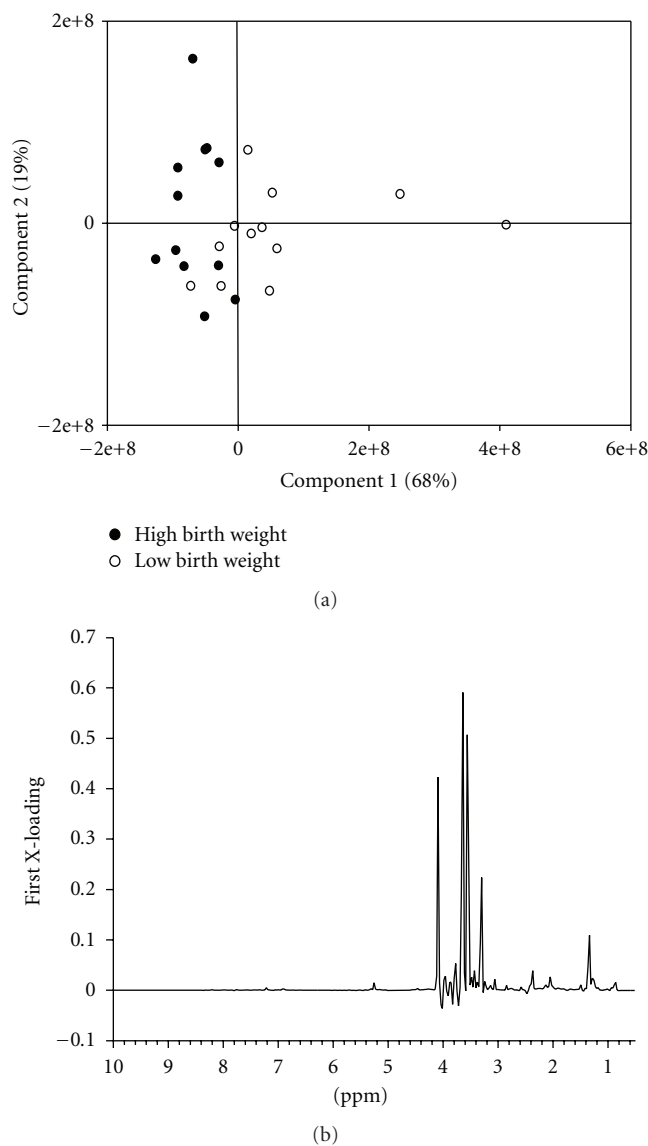


FIGURE 1: Differences in the ^1H NMR metabolite profile of plasma samples from low birth weight (LW) and high birth weight (HW) piglets. (a) PCA score plot showing the two first principal components for HW piglet samples (closed circles) and LW piglet samples (open circles). (b) Loading plot of first principal component.

2.3. NMR Measurements. The NMR measurements were performed at 310 K on a Bruker Avance III 600 spectrometer, operating at a ^1H frequency of 600.13 MHz, and equipped with a 5-mm ^1H TXI probe (Bruker BioSpin, Rheinstetten, Germany). Prior to the measurements, plasma samples were thawed and 400 μL aliquots were mixed with 200 μL D_2O . Sodium trimethylsilyl-[2,2,3,3- $^2\text{H}_4$]-1-propionate (TSP) was added as an internal chemical shift reference (0.17 mg/ml). ^1H NMR spectra of plasma samples were obtained using a Carr-Purcell-Meiboom-Gill (CPMG) delay added in order to attenuate broad signals from high-molecular-weight components. The total CPMG delay was 50 ms. Water suppression

TABLE 1: Anatomical characteristics of low birth weight (LW) and high birth weight (HW) piglets. LSMean values and standard errors of the LSMeans (SEM) are given.

| Traits | LW | HW | SEM | P |
|-------------------------------|------|-------|-------|--------|
| No. of piglets | 12 | 12 | | |
| No. of female piglets | 6 | 6 | | |
| No. of male piglets | 6 | 6 | | |
| Body weight, g | 687 | 1,179 | 0.05 | < .001 |
| CRL, cm* | 21.7 | 26.3 | 0.058 | < .001 |
| CRL/100 g body weight | 33.0 | 22.1 | 1.9 | < .001 |
| Pancreas, g/100 g body weight | 1.03 | 1.04 | 0.06 | NS |
| Liver, g/100 g body weight | 29.1 | 32.2 | 1.3 | .052 |

* Crown-rump-length.

was achieved by irradiating the water peak during the relaxation delay of 3 s. A total of 64 transients of 32 K data points spanning a spectral width of 17.36 ppm were collected. An exponential line-broadening function of 0.3 Hz was applied to the free induction decay (FID) prior to Fourier transform (FT). All spectra were referenced to the TSP signal at 0.0 ppm. The spectra were subdivided into 0.026 ppm integral regions and integrated, reducing each spectrum into 345 independent variables in the region 0.52–4.60 and 5.02–10.0 ppm.

2.4. GC-MS Measurements. Prior to GC-MS measurements, 50 μ l plasma from each sample was extracted with 960 μ l of MeOH (8:1v/v) and 100 μ l of 0.1 mg/ml D6 myo-inositol (Isotech) was added. The sample was then centrifuged at 14,000 g for 10 min at room temperature. Thereafter 100 μ l supernatant was transferred to GC-vials, evaporated and derivatized with 30 μ l of metoxyamin in pyridine (15 mg/ml) for 90 min at 30°C. The sample was then added to 30 μ l of MSTFA (1% TMCS; PIERCE), vortexed and left for 30 min at 37°C. Thereafter, 40 μ l of heptan was added and the sample was vortexed before GC-MS.

GC-MS was performed using an Agilent 7890 GCsystem with a CTC COMBI PAL autosampler coupled to a single quadrupole mass spectrometer (Agilent 5975). Gas chromatographic separation was performed using a 30 m * 250 μ m (i.d.) * 0.25 μ m film HP-5MS column (J & W Scientific). A volume of 1 μ l derivatized extract was injected into the GC-MS using split mode 1:10 with a split flow of 12 ml/min. The inlet temperature was 270°C. Oven temperature was increased from an initial 70°C to 230°C (15°C per min) and thereafter from 230°C to 300°C (10°C per min). The run time was 21 min. Helium 6.0 was used as a carrier gas at a flow rate of 1.2 ml/min. The ion source and quadrupole temperatures were 230 and 150°C, respectively. Chromatograms and mass spectra were evaluated using the chemstation E.02.00.493 software. Myo-inositol-C-d-d6 (Isotech) was used as an internal standard. D-(+)-chiro-inositol (cas 643-12-9; Aldrich) and Myo-inositol (Fluka)

was used to perform standard curves for each compound over the expected range relative to the internal standard. Semiquantitative concentrations of plasma myo-inositol and D-chiro-inositol were obtained against the standard curves.

2.5. Data Analysis and Statistics. Multivariate analysis of NMR data was performed using the Unscrambler software version 9.2 (Camo, Oslo, Norway). Principal component analysis (PCA) was applied to the centered data to explore any clustering behavior of the samples, and partial least square regression (PLS) was carried out using NMR spectra as x -variables and birth weight as y -variable. Martens' uncertainty test [16] was used to eliminate noisy variables, and all models were validated using full cross-validation [17].

Statistical analysis of GC-MS data and anatomical properties were performed using the SAS version 9.2 (SAS Institute Inc., Cary, NC, USA) using the MIXED procedure.

The MIXED model generalizes the standard linear model as follows:

$$y = X\beta + Z\gamma + \varepsilon \quad (1)$$

In this expression, y represents a vector of observed data, β is an unknown vector of fixed-effects parameters with known design matrix X , γ is a vector of random-effects parameters with known design matrix Z , and ε is an unknown random error vector whose elements are not required to be independent or homogeneous.

The model included the fixed effects of gender and birth weight group (LW or HW) and their interaction and sow as a random factor.

3. Results

3.1. Anatomical Properties of Piglets. The average birth weight of LW and HW piglets were 687 and 1,179 g, respectively ($P < .001$; Table 1). Also the crown-rump-length (CRL) was measured and LW piglets had an average CRL of 21.7 cm whereas HW piglets had an average of 26.3 cm. The proportional length of LW piglets was significantly higher

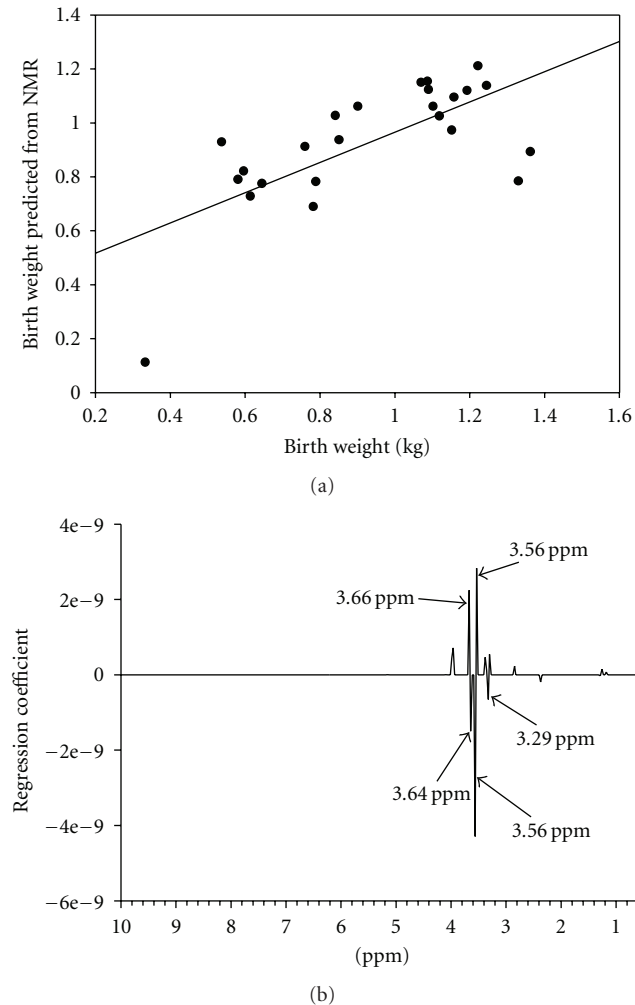


FIGURE 2: Prediction of birth weight by the ^1H NMR metabolite profile of plasma samples from low birth weight (LW) and high birth weight (HW) piglets. (a) Predicted versus measured birth weight from PLS regression with NMR-variables as x -variables and measured birth weight as y -variable. Values are given in Kg, $R^2 = 0.47$, Root mean squared error of prediction = 0.21 Kg. (b) The corresponding regression coefficients. The regression coefficients show the relationship between the NMR variables used as predictors for the birth weight. A positive coefficient shows a positive link with birth weight, and a negative coefficient shows a negative link. NMR variables with a small coefficient are negligible.

than of HW piglets ($P < .001$; Table 1), indicating that the LW piglets were very thin. This is a trait which is often recognized in IUGR subjects [2]. The extreme differences in size at birth were evident in the present study, where the variation in birth weight differed from 334 g to 1,453 g in one of the litters. This LW piglet had an extremely low birth weight, and in the animal science literature this extreme LW piglet is referred to as a runt [18–20].

In this study, pancreas and liver were weighed (Table 1). The relative weight (g/100 g body weight) of pancreas did not differ between LW and HW piglets, but the relative weight of

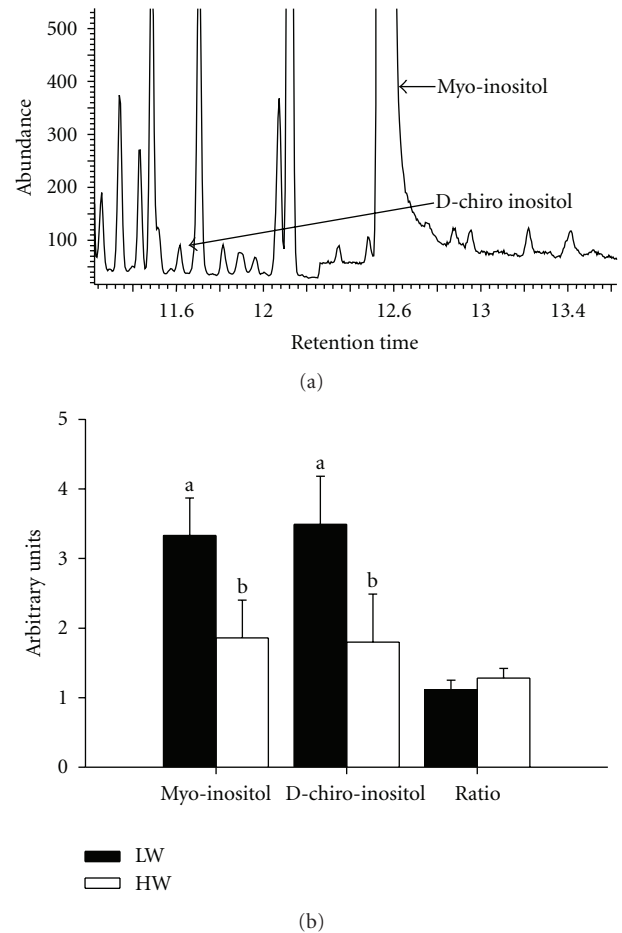


FIGURE 3: Myo-inositol and D-chiro-inositol levels in plasma from low birth weight (LW) and high birth weight (HW) piglets. (a) GC-MS spectra of a representative plasma sample. Peak at retention time 11.61 is D-chiro-inositol and peak at retention time 12.60 is myo-inositol. The spectra reveal the large difference in concentration of D-chiro-inositol and myo-inositol in the plasma samples. (b) Arbitrary plasma inositol levels in LW and HW piglets, showing that LW piglets have a significantly higher plasma concentration of both myo- and D-chiro-inositol than HW piglets at birth. Each bar represents the LSMeans of 12 piglets \pm SEM. ab, $P < .05$. Bars named ratio is representing the ratio between D-chiro-inositol and myo-inositol.

the liver tended ($P = .052$) to be higher in HW than in LW piglets (32.2 v 29.1 g/100 g body weight, resp.).

3.2. NMR Metabolomics. To investigate the early consequences of IUGR and possible relation with later development of adult health, a metabolomic approach was applied to plasma samples from LW and HW piglets by ^1H NMR spectroscopy. Principal component analysis (PCA), which is an unsupervised method, was performed on the pre-processed ^1H NMR spectra. The resulting plot of score 1 versus score 2 for mean-centered data shows a clear separation of plasma samples from LW and HW piglets, respectively, (Figure 1(a)). The corresponding X-loadings for the first component reveal that signals at 3.29, 3.56,

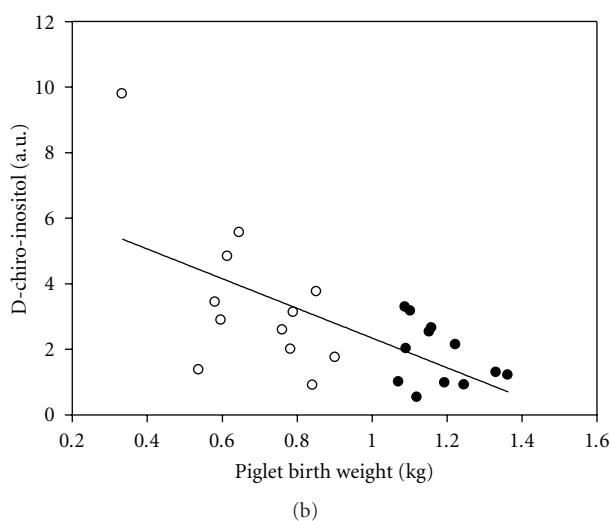
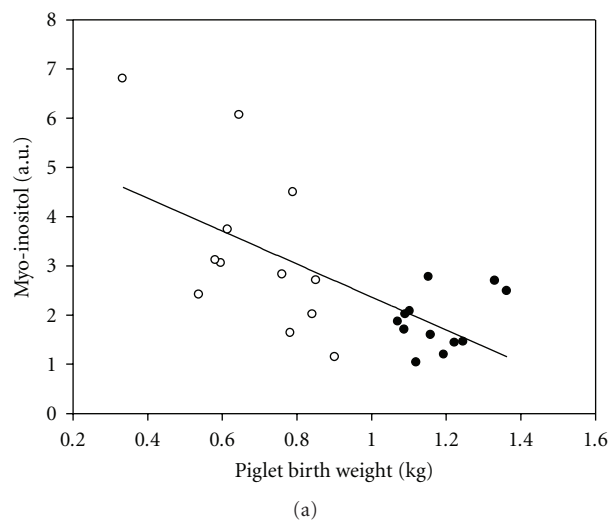


FIGURE 4: Relationship between piglet birth weight and (a) plasma myo-inositol and (b) plasma D-chiro-inositol. For both inositols a negative relationship (linear regression) exists between the plasma inositol concentration and piglet birth weight. $R^2 = 0.43$ for myo-inositol; $R^2 = 0.42$ for D-chiro-inositol. Each point represents a single piglet.

and 3.64 ppm are determining the position of samples along the first score, and thereby clearly dominate in the differentiation between LW and HW piglets (Figure 1(b)). These chemical shift values are equivalent with reference values for myo-inositol [21]. For a further investigation of the relationship between the ^1H NMR metabolite profiles and birth weight of the piglets, partial least squares (PLS) regression was carried out with NMR spectra as X -variables and birth weight as y -variable, which established a clear correlation ($R^2 = 0.47$) (Figure 2(a)). Analysis of the PLS model's regression coefficients reveals that intensities of signals at 3.53 and 3.66 ppm are positively correlated to birth weight (Figure 2(b)), and these signals are tentatively assigned to glucose. In addition, the regression coefficients disclose that NMR variables at 3.30, 3.56, and 3.64 ppm, which are equivalent with reference values for myo-inositol

[21], contribute to the correlation between the ^1H NMR metabolite profiles and birth weight of the piglets, as the intensity of these are negatively correlated with birth weight.

3.3. Plasma Inositol Concentrations. In order to verify that the found effect could be ascribed to a higher myo-inositol in LW piglets, semi-quantitative plasma myo-inositol concentrations were determined together with semi-quantitative plasma concentrations of D-chiro-inositol by GC-MS. The LW piglets in the present study had a significant higher plasma concentration of both myo-inositol ($P < .01$) and D-chiro-inositol ($P < .05$) than HW piglets, whereas the myo-/D-chiro-inositol ratio did not differ between LW and HW piglets (Figure 3(b)). Also it can be seen from Figure 3(a) that the plasma level of myo-inositol is much higher than the level of D-chiro-inositol.

3.4. Correlations between Plasma Inositol Concentrations and Piglet Birth Weight. The above shown results verify that LW piglets have a higher plasma concentration of both myo-inositol and D-chiro-inositol than HW piglets. In this study, piglets from 6 litters have been examined but the variation in birth weight within each litter differs between litters. Thus, in some litters the LW piglets are not extremely small and may be more in the category of average weigh piglets in other litters. We therefore tested the correlation between piglet birth weight and both myo-inositol and D-chiro-inositol (Figure 4). These results verify that there is a negative correlation between piglet birth weight and myo-inositol ($R^2 = 0.43$) and D-chiro-inositol ($R^2 = 0.42$) plasma concentrations. It can also be seen in Figure 4 that the variation in the plasma concentration of each compound seems to be higher in the LW piglets compared to the HW piglets. The extreme LW piglet also has a much higher concentration of myo-inositol and especially of D-chiro-inositol compared to the other LW and HW piglets.

4. Discussion

In the present study we used the naturally occurring form of IUGR that exists in pigs to study the influences of IUGR on the plasma metabolomic. Even though pigs are a litter-bearing species (9–14 piglets/litter) they seem to be a good model for human IUGR as there is increasing evidence that low birth weight pigs develop the same metabolic changes in relation to hypertension, obesity and type 2 diabetes as seen in human IUGR [11, 12, 22]. Thus, pigs are considered an excellent model for studying fetal metabolic programming and the adult consequences of IUGR.

In the present study, the 2 piglets with the lowest (LW) and the 2 piglets with the highest (HW) birth weight within each of 6 litters were studied in order to select pigs that had been subjected to IUGR (LW) and compared with pigs expected to have a normal nutrient supply during fetal development (HW). Piglets were taken from the uterus at day 110 of gestation (gestation length in sows is 113–115 days) in order to study piglets close to birth, but at the same time

ensuring that the piglets did not suckle and thereby ingest food.

A fuel-mediated teratogenesis has been proposed [23]. Thus, fetal islets, fetal fat stores, fetal muscles, and a large range of other cells may be altered mediated by changes in fetal fuels [23]. Accordingly, the organs of IUGR subjects may have a disproportionate growth compared to normal subject, where for example, the weight of pancreas and liver relative to whole body weight is lower in protein-restricted subjects [24]. The disproportional growth of body length (CRL) and liver in relation to body weight found in this study indicates that some tissues and organs are more affected by the IUGR than others, which may have long-term effects on postnatal growth and metabolism in adulthood.

The NMR-based metabolomic data in the present study shows that the plasma concentration of glucose is positively correlated with birth weight. Fetal plasma glucose originates mainly from transport across the placenta and the fetus is therefore highly dependent on delivery of glucose from the dam [10]. Possibly some gluconeogenesis takes place in fetal liver at the late part of gestation, and the low-glucose concentration in LW piglets in the present study thus indicates either a low maternal-fetal glucose gradient and/or decreased fetal gluconeogenesis. It is well recognized that the placental transport of nutrients to the fetus is of utmost importance for fetal growth, and that placental size and efficiency of nutrient transport is reduced in relation to IUGR [10, 25, 26].

Intriguingly, the explorative NMR-based metabolomic investigation demonstrated differences in plasma myo-inositol between piglets with high and low birth weight, respectively. These findings were confirmed by GC-MS measurements, which also showed a difference in D-chiro-inositol between high and low birth weight, respectively. Since the relationship between birth weight and health seems to be U-shaped [2], it could be questioned if the group of high birth weight piglets is an appropriate reference group. However, comparison of NMR data from an intermediate birth weight group revealed that the LW pigs also had a higher plasma myo-inositol level than the intermediate birth weight group (data not shown). Myo-inositol and D-chiro-inositol, which in the present study are negatively correlated to birth weight, have been related to glucose intolerance and type 2 diabetes in several studies [27–29]. Consequently, these metabolites are interesting candidates as markers for fetal programming of metabolic disease in adult life. Myo-inositol is synthesized from glucose and is the most abundant inositol in the body. D-chiro-inositol is either synthesized from myo-inositol by epimerization or obtained from the diet. Previous studies have shown that in subjects with noninsulin-dependent and insulin-dependent diabetes, urinary excretion of myo-inositol and D-chiro-inositol was higher than in normal subjects both in humans and rats [27, 30], whereas others have found a decreased D-chiro-inositol concentration in human urine [29, 31]. Reports on plasma concentrations of myo- and D-chiro-inositol in relation to diabetes are sparse, but a higher plasma concentration of inositols in acute diabetic compared to normal or mild diabetic rats was found in one study

[32], whereas no differences were found in another study comparing normal patients with non-insulin-dependent and insulin-dependent diabetic patients [30].

Thus, while it remains unclear if myo-inositol and D-chiro-inositol are useful biomarkers in adults, the present study suggests that they act as biomarkers in newborns. D-chiro-inositol seems to inhibit glucose-stimulated insulin release [33], suggesting a negative feedback mechanism after insulin-stimulated inositol phosphoglycan synthesis. Consequences of the high concentration of D-chiro-inositol found in LW piglets in the present study could be deficiency of insulin during fetal development. This may leave the insulin-responsive tissues like muscle highly sensitive to insulin [34]. In fact, highly insulin-sensitive tissues have been found in a low-maternal protein model for IUGR [35]. It could therefore be speculated that in the long-term, this could cause development of glucose intolerance and insulin resistance, and low-insulin concentrations would also have an impact on growth [34]. Also, the high D-chiro-inositol found in LW piglets may be a mechanism to decrease the glucose uptake in insulin-sensitive tissues, through the negative feedback on insulin release, leaving the small amount of available glucose for the more important tissues/organs in relation to survival, like the brain and heart. The biological significance of myo-inositol is less well understood. Thus, the present study suggests that myo-inositol and/or D-chiro inositol is useful markers in newborns, however, the consequences of the high plasma myo-inositol and D-chiro-inositol in LW piglets is not clear.

5. Conclusions

The present study demonstrated a clear positive relationship between plasma glucose concentration and birth weight and a negative relationship between myo-inositol and D-chiro-inositol plasma concentrations on the one hand and birth weight on the other hand in the naturally occurring pig model of IUGR. Our results show that low birth weight (LW) piglets have a significant higher concentration of both inositols in plasma compared to larger littermates. As Myo-inositol and D-chiro-inositol have been coupled with glucose intolerance and insulin resistance in adults, the present study indicates that these metabolites could be novel biomarkers for fetal programming.

Acknowledgments

This paper was supported by the Danish Research Council through the project “Foetal metabolic programming: Effects of intrauterine nutrition on metabolism and health later in life” and the Danish Technology and Production Research Council through the project “Advances in food quality and nutrition research through implementation of metabolomic technologies”. The authors also appreciate the technical assistance of Marianne Rasmussen during NMR measurements. The Danish Agriculture & Food Council is acknowledged for delivering the gilts for the present study.

References

- [1] A. Forsdahl, "Living conditions in childhood and subsequent development of risk factors for arteriosclerotic heart. The cardiovascular survey in Finnmark 1974-75," *Journal of Epidemiology and Community Health*, vol. 32, no. 1, pp. 34–37, 1978.
- [2] T. Harder, E. Rodekamp, K. Schellong, J. W. Dudenhausen, and A. Plagemann, "Birth weight and subsequent risk of type 2 diabetes: a meta-analysis," *American Journal of Epidemiology*, vol. 165, no. 8, pp. 849–857, 2007.
- [3] C. N. Hales and D. J. P. Barker, "Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis," *Diabetologia*, vol. 35, no. 7, pp. 595–601, 1992.
- [4] K. M. Godfrey and D. J. P. Barker, "Fetal nutrition and adult disease," *American Journal of Clinical Nutrition*, vol. 71, no. 5, pp. 1344S–1352S, 2000.
- [5] L. Aerts and F. A. Van Assche, "Animal evidence for the trans-generational development of diabetes mellitus," *International Journal of Biochemistry and Cell Biology*, vol. 38, no. 5-6, pp. 894–903, 2006.
- [6] K. L. Kind, P. M. Clifton, P. A. Grant et al., "Effect of maternal feed restriction during pregnancy on glucose tolerance in the adult guinea pig," *American Journal of Physiology*, vol. 284, no. 1, pp. R140–R152, 2003.
- [7] S. P. Burns, M. Desai, R. D. Cohen et al., "Gluconeogenesis, glucose handling, and structural changes in livers of the adult offspring of rats partially deprived of protein during pregnancy and lactation," *The Journal of Clinical Investigation*, vol. 100, no. 7, pp. 1768–1774, 1997.
- [8] M. J. De Blasio, K. L. Gatford, I. C. McMillen, J. S. Robinson, and J. A. Owens, "Placental restriction of fetal growth increases insulin action, growth, and adiposity in the young lamb," *Endocrinology*, vol. 148, no. 3, pp. 1350–1358, 2007.
- [9] J. B. Susa, C. Neave, P. Sehgal, D. B. Singer, W. P. Zeller, and R. Schwartz, "Chronic hyperinsulinemia in the fetal rhesus monkey. Effects of physiologic hyperinsulinemia on fetal growth and composition," *Diabetes*, vol. 33, no. 7, pp. 656–660, 1984.
- [10] I. Cetin and G. Alvino, "Intrauterine growth restriction: implications for placental metabolism and transport. A review," *Placenta*, vol. 30, pp. 77–82, 2009.
- [11] A. L. Fowden and A. J. Forhead, "Endocrine mechanisms of intrauterine programming," *Reproduction*, vol. 127, no. 5, pp. 515–526, 2004.
- [12] K. R. Poore and A. L. Fowden, "The effect of birth weight on glucose tolerance in pigs at 3 and 12 months of age," *Diabetologia*, vol. 45, no. 9, pp. 1247–1254, 2002.
- [13] J. J. Meier, "Linking the genetics of type 2 diabetes with low birth weight: a role for prenatal islet maldevelopment?" *Diabetes*, vol. 58, no. 6, pp. 1255–1256, 2009.
- [14] R. M. Freathy, A. J. Bennett, S. M. Ring et al., "Type 2 diabetes risk alleles are associated with reduced size at birth," *Diabetes*, vol. 58, no. 6, pp. 1428–1433, 2009.
- [15] K. Zandi-Nejad, V. A. Luyckx, and B. M. Brenner, "Adult hypertension and kidney disease: the role of fetal programming," *Hypertension*, vol. 47, no. 3, pp. 502–508, 2006.
- [16] H. A. Martens and P. Dardenne, "Validation and verification of regression in small data sets," *Chemometrics and Intelligent Laboratory Systems*, vol. 44, no. 1-2, pp. 99–121, 1998.
- [17] H. Martens and M. Martens, "Modified Jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR)," *Food Quality and Preference*, vol. 11, no. 1-2, pp. 5–16, 2000.
- [18] S. E. Handel and N. C. Stickland, "Muscle cellularity and birth weight," *Animal Production*, vol. 44, pp. 311–317, 1987.
- [19] S. E. Powell and E. D. Aberle, "Skeletal muscle and adipose tissue cellularity in runt and normal birth weight swine," *Journal of Animal Science*, vol. 52, no. 4, pp. 748–756, 1981.
- [20] P. V. J. Hegarty and C. E. Allen, "Effect of pre-natal runting on the post-natal development of skeletal muscles in swine and rats," *Journal of Animal Science*, vol. 46, no. 6, pp. 1634–1640, 1978.
- [21] J. C. Lindon, J. K. Nicholson, and J. R. Everett, "NMR spectroscopy of biofluids," *Annual Reports NMR Spectroscopy*, vol. 38, pp. 1–88, 1999.
- [22] F. Gondret, L. Lefaucheur, H. Juin, I. Louveau, and B. Lebret, "Low birth weight is associated with enlarged muscle fiber area and impaired meat tenderness of the longissimus muscle in pigs," *Journal of Animal Science*, vol. 84, no. 1, pp. 93–103, 2006.
- [23] N. Freinkel, "Banting Lecture 1980. Of pregnancy and progeny," *Diabetes*, vol. 29, no. 12, pp. 1023–1035, 1980.
- [24] M. Desai, N. J. Crowther, A. Lucas, and C. N. Hales, "Organ-selective growth in the offspring of protein-restricted mothers," *British Journal of Nutrition*, vol. 76, no. 4, pp. 591–603, 1996.
- [25] C. B. Doherty, R. M. Lewis, A. Sharkey, and G. J. Burton, "Placental composition and surface area but not vascularization are altered by maternal protein restriction in the rat," *Placenta*, vol. 24, no. 1, pp. 34–38, 2003.
- [26] K. L. Kind, J. A. Owens, J. S. Robinson et al., "Effect of restriction of placental growth on expression of IGFs in fetal sheep: relationship to fetal growth, circulating IGFs and binding proteins," *Journal of Endocrinology*, vol. 146, no. 1, pp. 23–34, 1995.
- [27] J. M. Kawa, R. Przybylski, and C. G. Taylor, "Urinary chiro-inositol and myo-inositol excretion is elevated in the diabetic db/db mouse and streptozotocin diabetic rat," *Experimental Biology and Medicine*, vol. 228, no. 8, pp. 907–914, 2003.
- [28] G. Sarashina, M. Yamakoshi, M. Noritake et al., "A study of urinary myo-inositol as a sensitive marker of glucose intolerance," *Clinica Chimica Acta*, vol. 344, no. 1-2, pp. 181–188, 2004.
- [29] T.-S. Jung, J.-R. Hahm, J.-J. Kim et al., "Determination of urinary myo-/chiro-inositol ratios from Korean diabetes patients," *Yonsei Medical Journal*, vol. 46, no. 4, pp. 532–538, 2005.
- [30] R. E. Ostlund Jr., J. B. McGill, I. Herskowitz, D. M. Kipnis, J. V. Santiago, and W. R. Sherman, "D-chiro-inositol metabolism in diabetes mellitus," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 90, no. 21, pp. 9988–9992, 1993.
- [31] I. Asplin, G. Galasko, and J. Larner, "Chiro-inositol deficiency and insulin resistance: a comparison of the chiro-inositol- and the myo-inositol-containing insulin mediators isolated from urine, hemodialysate, and muscle of control and type II diabetic subjects," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 90, no. 13, pp. 5924–5928, 1993.
- [32] K. P. Palmano, P. H. Whiting, and J. N. Hawthorne, "Free and lipid myo-inositol in tissues from rats with acute and less severe streptozotocin-induced diabetes," *Biochemical Journal*, vol. 167, no. 1, pp. 229–235, 1977.
- [33] J. Larner, "D-chiro-inositol-its functional role in insulin action and its deficit in insulin resistance," *International Journal of Experimental Diabetes Research*, vol. 3, no. 1, pp. 47–60, 2002.

- [34] R. Bhandari, K. R. Juluri, A. C. Resnick, and S. H. Snyder, "Gene deletion of inositol hexakisphosphate kinase 1 reveals inositol pyrophosphate regulation of insulin secretion, growth, and spermiogenesis," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 7, pp. 2349–2353, 2008.
- [35] M. Thamocharan, B.-C. Shin, D. T. Suddirikku, S. Thamocharan, M. Garg, and S. U. Devaskar, "GLUT4 expression and subcellular localization in the intrauterine growth-restricted adult rat female offspring," *American Journal of Physiology*, vol. 288, no. 5, pp. E935–E947, 2005.

Research Article

Etiological Analysis of Neurodevelopmental Disabilities: Single-Center Eight-Year Clinical Experience in South China

Li Guo,¹ Bing-Xiao Li,¹ Mei Deng,¹ Fang Wen,¹ Jian-Hui Jiang,²
Yue-Qiu Tan,³ Yuan-Zong Song,¹ Zhen-Huan Liu,⁴ Chun-Hua Zhang,⁵
Keiko Kobayashi,⁶ and Zi-Neng Wang⁷

¹Department of Pediatrics, The First Affiliated Hospital, Jinan University, No.613, Huangpu Dadao Xi, Guangzhou 510630, China

²Guangzhou Neonatal Screening Center, Guangzhou Women and Children's Medical Center, Guangzhou 510180, China

³Institute of Reproduction and Stem Cell Engineering, Xiangya School of Medicine, Central South University, Changsha 410078, China

⁴Department of Pediatric Neurorehabilitation, Nanhai Maternity and Child Care Hospital, Guangzhou University of Chinese Medicine, Foshan 528200, China

⁵Department of Research and Development, Matsumoto Institute of Life Science International, Kanazawa 921-8154, Japan

⁶Department of Molecular Metabolism and Biochemical Genetics, Kagoshima University Graduate School of Medical and Dental Sciences, Kagoshima 890-8544, Japan

⁷Department of Gynecology and Obstetrics, The First Affiliated Hospital, Jinan University, Guangzhou 510630, China

Correspondence should be addressed to Yuan-Zong Song, songyuanzong@tom.com

Received 13 May 2010; Accepted 30 July 2010

Academic Editor: Veikko Salomaa

Copyright © 2011 Li Guo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Etiology determination of neurodevelopmental disabilities (NDDs) currently remains a worldwide common challenge on child health. We herein reported the etiology distribution feature in a cohort of 285 Chinese patients with NDDs. Although concrete NDD etiologies in 48.4% of the total patients could not be identified, genetic diseases (with the proportion of 35.8% in the total cases) including inborn errors of metabolism (IEM) and congenital dysmorphic diseases, constituted the commonest etiology category for NDDs in this study. The two key experimental technologies in pediatric metabolomics, gas chromatography-mass spectrometry (GC-MS), and tandem mass spectrometry (MS-MS), proved to be substantially helpful for the exploration of the NDD etiologies in this clinical investigation. The findings in this paper provided latest epidemiologic information on the etiology distribution of NDDs in Chinese, and the syndromic NDDs caused by citrin deficiency and the novel chromosomal karyotype, respectively, further expanded the etiology spectrum of NDDs.

1. Introduction

With global developmental delay (GDD) and mental retardation (MR) as two main clinical subtypes, neurodevelopmental disabilities (NDDs), which are defined as a group of chronic clinically distinct disorders that all share a documented disturbance, quantitative, qualitative, or both, in developmental progress in one or more developmental domains compared with established norms [1], are conventionally categorized into syndromic type which is characterized by associated clinical, radiological, metabolic or biological features, and nonsyndromic type in which NDD

represents the only manifestation. The precise prevalence of NDDs remains unclear, but this entity has been estimated to affect 5% to 10% of children [2]. In developed countries, MR has become the most frequent cause of severe handicap in children and one of the main reasons for referral in clinical genetic practice [3]. Actually, 1% to 3% of children younger than 5 years have been reasonably given the prevalence of MR in a specific population [4]. As the largest developing country in the world with a population over 1.3 billion, China also faces the difficult challenge of NDDs on its child health. An investigation in the year 2000 has revealed that the MR incidence in children below 6 years of age was 0.931%,

and 136,000 children with MR were increased annually in mainland of our country [5]. Etiology determination of NDDs was essential not only for the option of therapeutic imperatives and evaluation of clinical outcomes and recurrence risks but also for other benefits including avoidance of unnecessary tests and access to appropriate patients for accumulating management experiences, however, this issue also remains far from resolved at the current stage in pediatric practice. In this paper, we reported our eight-year findings on NDD etiologies in a medical center in south China.

2. Subjects and Methods

2.1. Patients. The research subjects recruited in this study were all patients referred to, from April 2002 to March 2010, Department of Pediatrics, First Affiliated Hospital, Jinan University, Guangdong, China. The GDD/MR diagnosis in most cases was made by at least 2 pediatric physicians in different hospitals in accordance with the updated concepts in the review [1]. For some cases who suffered from other clinical problems such as liver diseases and malformations, NDD was noticed and then confirmed in our department. The patients in this study came from 22 provinces, municipalities, and autonomous regions in China, respectively, with most of them from Guangdong Province.

2.2. Clinical Data. History inquiry and physical examination were performed on all the NDDs patients in our pediatric clinic or ward and the positive findings were recorded and preserved by the authors. Most of the laboratory and imaging results were collected from the corresponding databases in our hospital, besides some provided by parents of the patients at their referral to the authors for clinical counseling. In this study we, by means of a cross-sectional study, retrospectively analyzed and summarised the clinical information collected in the past 8 years.

2.3. Gas Chromatography-Mass Spectrometry (GC-MS). Selective screening of inborn errors of metabolism (IEMs) in this study was conducted by analysis of the urinary components, using an urease pretreatment GC-MS procedure, mainly with a Finnigan GC-MS instrument (TRACE DSQ), with detailed information described previously by our group [6].

2.4. Tandem Mass Spectrometry (MS-MS). Amino and acyl carnitine in dried blood stains was analyzed by means of a MS-MS procedure, and sample preparation, apparatus settings, and data analysis were based on the detailed information described in [7]. The analysis was conducted with an API 3200 tandem mass spectrometer purchased from Applied Biosystems. Neutral loss scan and precursor scan were used for the analysis of most amino acids and acyl carnitines, respectively, while multiple reaction monitoring (MRM) was utilized for the detection of glycine, ornithine, arginine, and citrulline as well.

TABLE 1: Main clinical manifestations besides NDDs and the positive laboratory and imaging findings in syndromic NDDs.

| No. | Positive findings | Cases |
|-----|---------------------------------|-------|
| 01 | Failure to thrive | 81 |
| 02 | Seizure/convulsion | 37 |
| 03 | Hearing disability | 28 |
| 04 | Dysmorphic facial features | 25 |
| 05 | Abnormal urine odor | 20 |
| 06 | Eye movement obstacles | 19 |
| 07 | Vomiting | 18 |
| 08 | Hair depigmentation | 15 |
| 09 | Microcephaly | 11 |
| 10 | Skin abnormalities | 10 |
| 11 | Hepato/splenomegaly | 10 |
| 12 | Impaired swallowing and chewing | 5 |
| 13 | Fondus oculi abnormalities | 4 |
| 14 | Vision problem | 4 |
| 15 | Abnormal lens | 3 |
| 16 | Genitalia malformation | 3 |
| 17 | Metabolic acidosis | 46 |
| 18 | Hyperammonemia | 12 |
| 19 | Abnormal EEG | 22 |
| 20 | Skeleton abnormality on X ray | 5 |
| 21 | CT/MRI abnormal findings | 85 |

2.5. Chromosome Karyotype Analysis. Traditional chromosomal banding was performed in NDD patients suspected to have chromosomal aberrations. Fluorescence in situ hybridization (FISH) was further used, when necessary, to determine the complex karyotypes as previously described [8]. Briefly, the peripheral blood lymphocytes were cultured under phytohemagglutinin (PHA) stimulation and treated with colcemid and harvested by standard methods. Metaphases were spread on clean slides, and standard G-banding with trypsin-Giemsa was performed. The slides for FISH were stored at -20°C before use. The denatured FISH probe (Abbott-Vysis, Downers Grove, IL, USA) was added to the denatured slides with metaphase spreads in a moist chamber for hybridizing over night. After washing, the slide was counterstained with DAPI in an antifade solution. The hybridized metaphase chromosomes were captured and analyzed using a digital image analysis system containing an Olympus BX51 microscope equipped with LUCIA Cytogenetics system (Prague, Czech Republic).

2.6. SLC25A13 Gene Mutation Analysis. The diagnosis of citrin deficiency was confirmed by mutation analysis of the causative gene SLC25A13. Four hotspot mutations, 851-854del(851del4), IVS6 + 5G > A, IVS16ins3kb, and 1638-1660dup were screened by means of the routine approaches described in reference [9]. In this study, the sequences of the forward and backward primers for PCR amplification of the mutation 851del4 were 5'-ggatatattgttctgtgttg-3' and 5'-tcttccagaggagcaatccg-3', respectively.

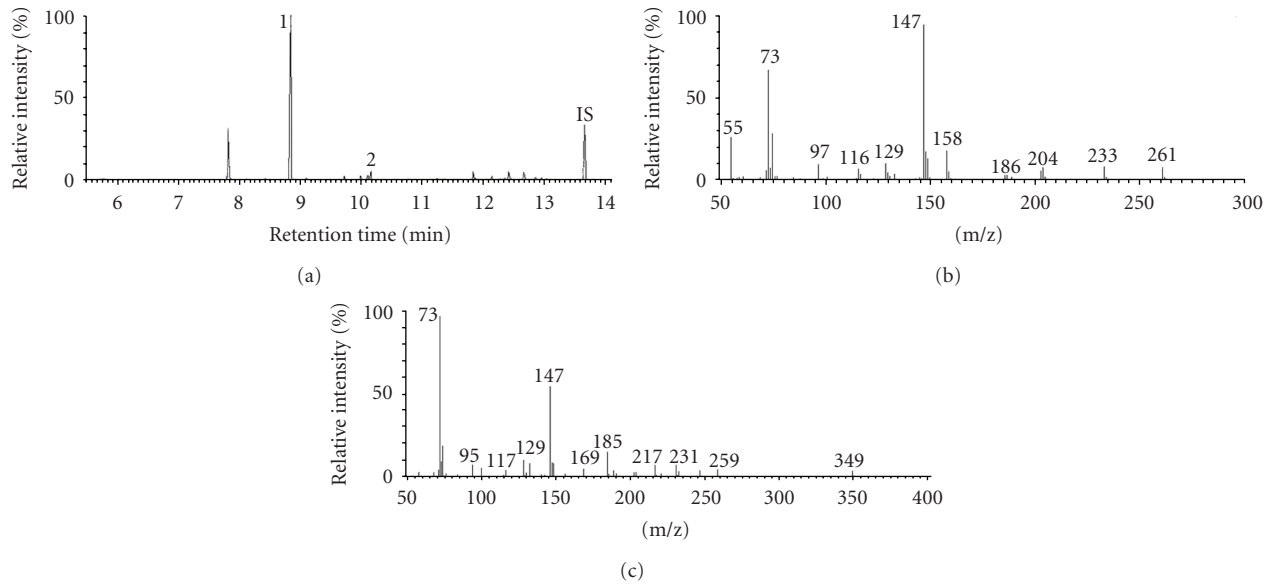


FIGURE 1: Chemical diagnosis of glutaric acidemia type I by GC-MS analysis of urinary metabolites for a 9-month-old female with motor and language retardation. Figure 1(a) is a representative GC-MS total ion current (TIC) profile, in which intensities of peak 1 and 2 were both dramatically increased. IS is the abbreviation of internal standard. Figures 1(b) and 1(c), the mass spectra for peak 1 and 2 in Figure 1(a), revealed their identifications as trimethylsilyl derivatives of glutarate and 3-hydroxyglutarate, respectively.

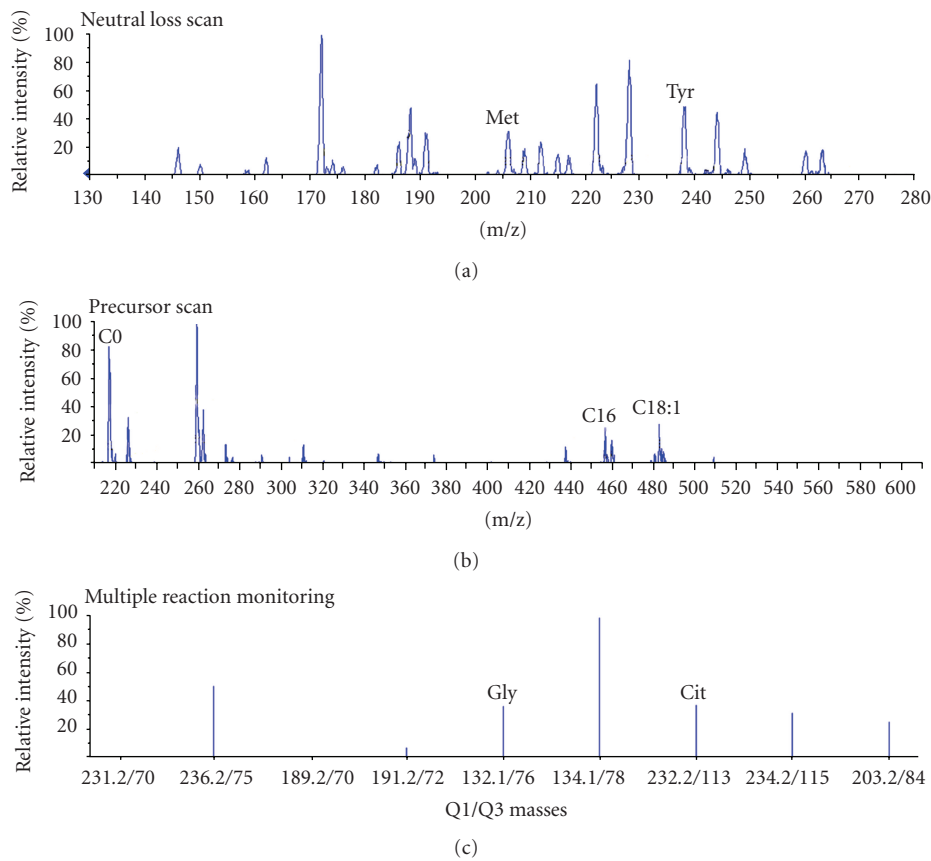


FIGURE 2: Chemical diagnosis of citrin deficiency by MS-MS analysis of amino and acyl carnitines in dried blood stain from a 16-month-old male toddler (C0013) with syndromic GDD. Figures 2(a), 2(b), and 2(c) are profiles of neutral loss scan, precursor scan, and multiple reaction monitoring, respectively. The amino and acyl carnitines with increased levels were labeled as abbreviations, with Met, Tyr, Gly, Cit, C0, C16, and C18:1 representing methionine, tyrosine, glycine, citrulline, free carnitine, palmitate, and hydroxypalmitate, respectively.

TABLE 2: Etiology distribution in the whole cohort of NDDs.

| Etiology categories | Disease types | Case number | Proportion | Concrete disease (case number) |
|-------------------------|---------------|-------------|------------|---|
| Genetic diseases | 50 | 102 | 35.8% | Detailed information in Table 3 |
| Psychobehaviour | 3 | 23 | 8.1% | Autism (21, including 5 cases of Rett syndrome); ADHD (2) |
| Acquired brain injuries | 2 | 7 | 2.4% | Kernicterus (4); HIE (3) |
| Other etiology | 2 | 15 | 5.3% | Cerebral palsy (7), Epilepsy (8) |
| Unknown | 1 (NDDs) | 138 | 48.4% | No concrete etiologies were identified at the current stage |
| In total | 58 | 285 | 100% | — |

TABLE 3: Etiology distribution in the patients with genetic diseases in Table 2.

| Etiology categories | Disease types | Case number | Proportion | Concrete diseases (case number) |
|--------------------------------|---------------|-------------|------------|---|
| IEMs | 27 | 66 | 64.7% | Detailed information in Table 4 |
| Congenital dysmorphic diseases | 14 | 22 | 21.6% | Detailed information in Table 5 |
| Chromosomal aberrations | 4 | 6 | 5.9% | Detailed information in Section 3.5 |
| Endocrine disorders | 3 | 4 | 3.9% | Hypoparathyroidism (1), Pseudohypoparathyroidism (1), Congenital hypothyroidism (2) |
| Others | 2 | 4 | 3.9% | Congenital muscular dystrophy (3), Progressive muscular dystrophy (1) |
| In total | 50 | 102 | 100% | — |

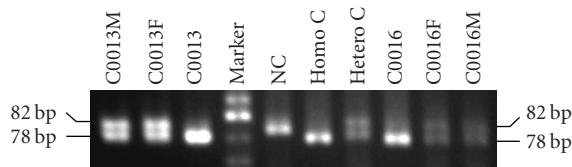


FIGURE 3: PCR-gel electrophoresis analysis of mutation 851del4 in the gene SLC25A13 of the two families with citrin-deficient patients C0013 and C0016. NC, Homo C and Hetero C in this figure are abbreviations of Normal Control, Homozygous Control and Heterozygous Control, respectively. F and M in the two families represent Father and Mother, respectively. The 78 bp PCR products in both patients are 4 bp shorter than the normal size 82 bp, suggesting that the 2 patients are both 851del4 homozygotes, and their parents all carriers of the same mutation.

2.7. Electronic Microscopy. The muscular ultrastructure changes were observed by transmission electronic microscopy in patients suspected to have Leigh syndrome and muscular dystrophy, with biopsy samples from musculi gastrocnemius. Muscular tissue was fixed in 2.5% glutaraldehyde and post-fixed in 1% osmium tetroxide solution, and then embedded in epoxy resin before semithin sectioning, as described in reference [10]. Two electronic microscope instruments (JOEL-CX100 and Philips-Tecai-10) were utilized for ultrastructure observation in our investigation.

This study was performed with the informed consents from the parents of the patients, adhering to the principles

of the Declaration of Helsinki. In particular, SLC25A13 mutation analysis has been approved by the Committee for Medical Ethics in the First Affiliated Hospital, Jinan University, while chromosome karyotype analysis by that in Reproductive and Genetic Hospital of Citic-Xiangya, Central South University.

3. Results

3.1. General Information and Semiology. The NDDs cohort in this study was composed of 285 cases in total, including 191 males and 94 females. The median age at referral was 1 year and 7 months, with minimum 1 week, and maximum 16 years. The whole cohort included 240 syndromic and 45 nonsyndromic NDD cases, with the relative proportion of 84.2% and 15.8%, respectively. The main clinical manifestations besides NDD and the laboratory and imaging findings in the syndromic NDDs were summarized in Table 1.

3.2. Etiology Distribution. As shown in Table 2, no concrete etiologies, unfortunately, were found for 48.4% of the total patients, but NDDs in the remaining 51.6% could be attributed to complex and diverse etiologies, among which genetic diseases were on top of the list of the identified causes. Further analysis in Table 3 revealed that IEMs took the first place in genetic etiologies, and then came congenital dysmorphic disorders. These 2 entities constituted

TABLE 4: Feature of etiology distribution in the patients with IEMs in Table 3.

| Categories | Diseases* | Cases | Major diagnostic evidences | Clinical outcomes |
|--------------------------------------|---------------------------------|-------|--|---|
| Disorders of Carbohydrate metabolism | Galactosemia | 1 | Clinical features including congenital cataract and leukodystrophy, and GC-MS analysis | Lost contact |
| | Fructosuria | 2 | GC-MS analysis | Both lost contact |
| Disorders of Amino acid metabolism | Phenylketonuria | 7 | GC-MS analysis in 6 cases, and PAH gene analysis in 1 case | Referred to local network of management and 2 died after treatment withdrawal |
| | Histidinemia | 1 | Repeated MS-MS analysis | Lost contact |
| | Hyperhomocysteinemia | 2 | Total plasma homocysteine levels and MS-MS analysis | 1 died, and 1 stable without obvious clinical or biochemical improvement |
| | Pyroglutamic acidemia | 1 | GC-MS analysis | Lost contact |
| | Tyrosinemia type I | 1 | GC-MS and MS-MS findings | Died due to acute liver failure |
| | Hyperglycinemia | 1 | GC-MS and MS-MS analysis | Intractable seizures and behavioral problem |
| | Canavan's disease | 1 | GC-MS analysis | Lost contact |
| Organic acidemia | Methylmalonic acidemia | 11 | GC-MS, MS-MS and MMACHC gene analysis, with 5 combined with hyperhomocysteinemia | 5 died after withdrawal of treatment, 3 improved and 3 lost contact |
| | Maple syrup urine disease | 2 | GC-MS and MS-MS analysis | Both died |
| | Ethylmalonic acidemia | 1 | GC-MS analysis | Lost contact |
| | Propionic acidemia | 3 | GC-MS analysis | 2 stable with episodic hyperammonemia, and 1 lost contact |
| | Glutaric acidemia type I | 2 | GC-MS and MS-MS analysis | 1 lost contact and 1 stable |
| | Glutaric acidemia type II | 1 | GC-MS analysis | Stable |
| | 2-hydroxyglutaric acidemia | 1 | GC-MS analysis | Lost contact |
| | 4-hydroxybutyric aciduria | 1 | GC-MS and ALDH5A1 gene analysis | Stable but with seizure episodes |
| | Multiple carboxylase deficiency | 4 | GC-MS, biotinidase activity, and HLCS gene analysis | 1 died, 3 recovered/improved clinically |
| Urea cycle disorders | OTCD | 2 | GC-MS and MS-MS analysis | Recovered clinically |
| | Hyperammonemia | 4 | Markedly increased serum ammonia levels, but with etiologies undetermined yet | All lost contact |
| | Citrin deficiency | 2 | SLC25A13 mutation analysis | 1 died due to liver cirrhosis, 1 improved |
| Mitochondrial disease | Leigh syndrome | 5 | Clinical and imaging features, serum/CSF lactate levels, and electronic microscopy findings on muscle biopsy samples | 3 died already, and the remaining 2 stable at follow-up |
| Lysosome storage diseases | Mucopolysaccharidosis type I | 1 | Typical clinical manifestations | Improved after bone marrow transplantation |
| | Mucopolysaccharidosis type II | 2 | Activity analysis of iduronate-2-sulphatase | Lost contact |
| Peroxisomal disorders | X-linked adrenoleukodystrophy | 2 | Clinical manifestations, CT/MRI findings, and MS-MS analysis of VLCFA | Both Died |
| Others | Glyceroluria | 4 | GC-MS analysis | 1 died after severe infection, 3 lost contact |
| | 3-aminoisobutyric aciduria | 1 | GC-MS analysis | Lost contact |

* Some diseases have been reported in [6] as GC-MS screening results, and this list herein is the latest update of our findings, just focusing on the IEMs associated with NDDs.

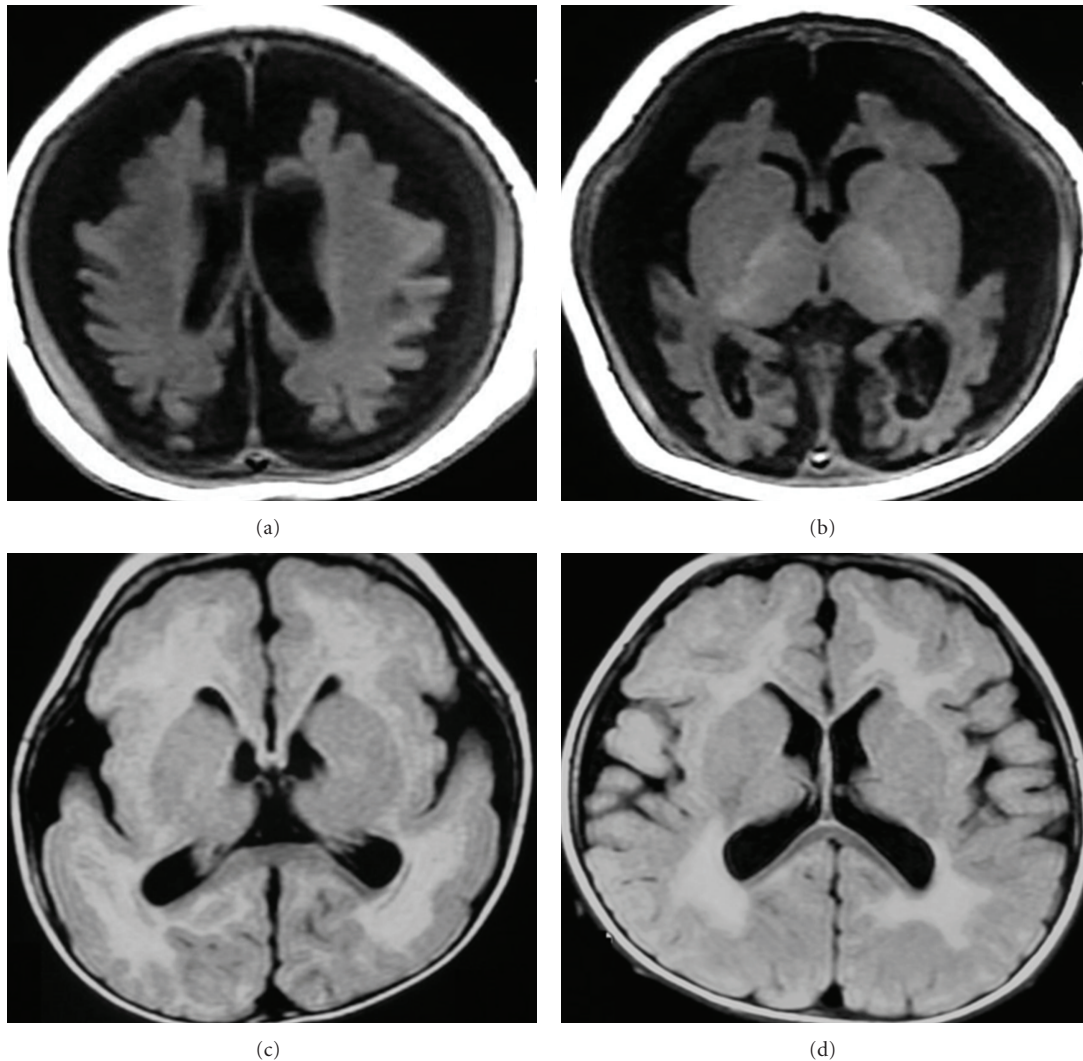


FIGURE 4: Representative MRI findings in different malformations of cortical development (MCD). Figures 4(a) and 4(b) showed severe cortex dysplasia in bilateral parietal lobes and frontal, temporal and occipital lobes, respectively, in the telencephalon of a 5-month-old male with NDD. Figure 4(c) demonstrated typical lissencephaly in a 9-month-old female with Miller-Dieker syndrome. The white matter volume was decreased while the cortex was thick and smooth due to lack of enough sulcation, forming the so-called pachygyria malformation, and the thickened and irregular cortex in Figure 4(d) revealed the cobblestone cortical malformation in a patient with muscle-eye-brain disease.

the overwhelming majority in the genetic diseases. Other causes such as chromosome and endocrine abnormalities were also identified, just accounting for a minority less than 14% in the total genetic etiologies.

3.3. IEMs. As listed in Table 4, 66 patients with IEMs of 27 types and 8 categories were diagnosed in this NDD cohort. The traditional clinical, biochemical, and imaging findings were indispensable during the diagnostic processes, however, the applications of two metabolome tools, GC-MS and MS-MS, were substantially helpful in the exploration of the NDD etiologies in this study. Figure 1 demonstrated the diagnostic evidences of glutaric acidemia type I by GC-MS analysis of urine sample. In particular, we diagnosed 2 GDD patients secondary to citrin deficiency. Figure 2 illustrated the

MS-MS findings suggestive of citrin deficiency in a male toddler (C0013) at his age of 1 year and 4 months, who presented with persistent GDD due to prolonged hepatosplenomegaly and recurrent ascites that progressed into lethal hepatic encephalopathy at his age of 1 year and 10 months. GDD was transient in another 7-month-old infant (C0016) with neonatal intrahepatic cholestasis caused by citrin deficiency (NICCD, OMIM #605814), who demonstrated catch-up development after recovery of dyslipidemia and abnormal liver function indices. As shown in Figure 3, mutation analysis of the causative gene *SLC25A13* clearly confirmed their diagnosis of 851del4 homozygotes.

3.4. Congenital Dysmorphic Disorders. In this NDDs cohort, 14 kinds and 22 cases of congenital dysmorphic disorders

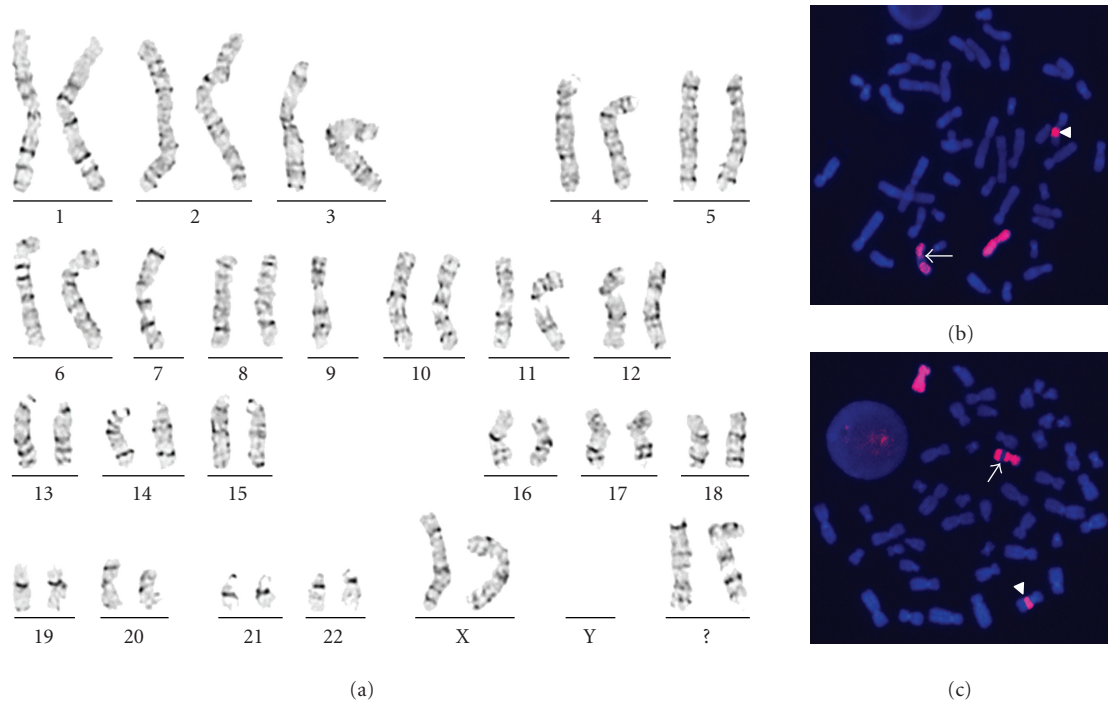


FIGURE 5: Chromosome aberration in a 6-year-old female with mental retardation (MR). High-resolution GTG-banding in Figure 5(a) revealed the derivative chromosomes 7 and 9 (question mark), and their detailed identities were further illustrated by the results of FISH analysis. Figure 5(b) showed four red signals in a metaphase, by means of utilization of whole chromosome 7 painting probe (WCP 7, red). The normal chromosome 7 had one intact red signal, but the derivative chromosome 7 (arrow) had two dispersed red signals, and a fragment of chromosome 7 (arrowhead) was inserted into a chromosome 9, forming a derivative chromosome 9. Similarly, FISH analysis with WCP 9 (red) in Figure 5(c) showed the normal chromosome 9 with intact red signal, the derivative chromosome 9 (arrow) with two dispersed red signals, and the derivative chromosome 7 with a inserted fragment of chromosome 9 (arrowhead). Finally, the chromosome karyotype in this patient was identified as 46, XX, ins (7;9) (p13; q32q22) inv(7) (p11.2 q11.23), ins (9;7) (q22; q22q32), ish ins(7;9) (WCP7+, WCP9+), ins(9;7) (WCP7+, WCP9+).

were diagnosed. Most of them were, as shown in Table 5, malformations of cortical development (MCD) such as Miller-Dieker syndrome, Muscle-Eye-Brain disease, and isolated lissencephaly sequence. Figure 4 demonstrated representative MRI findings in the different MCD types. Other CNS malformations like Dandy-Walker syndrome and spinocerebellar ataxia, and some rare syndromes including Silver-Russell syndrome, Noonan syndrome, Poland-Moebius syndrome, and Crisponi syndrome, were also found in our clinical practice. To our knowledge, the patient with Crisponi syndrome reported here is the first case in China.

3.5. Chromosome Karyotypes. By traditional chromosome analysis, karyotype abnormalities were found in 6 patients. Among them, 3 patients were diagnosed as having Down's syndrome (trisomy 21) and 1 Patau's syndrome (trisomy 13), respectively. The remaining 2 were both complex karyotypes, with mos. 47, XX, +der (15) (pter → q14::q14 → pter) [11]/48, XX, +der (15) (pter → q14::q14 → pter) × 2 [12]. ish der(15) (WCP15+, UBE3A++, PML-) in 1 case, and 46, XX, inv ins (7;9) (p13; q32q22) inv(7) (p11.2 q11.23), ins (9;7) (q22; q22q32), ish(7;9) (WCP7+, WCP9+), ins(9;7) (WCP7+, WCP9+) in another, as illustrated in Figure 5. So far as we know, the last karyotype is a novel one that has

never been reported in any other references. The patient with this novel abnormal karyotype was a 6-year-old female with mental retardation. History inquiry revealed motor retardation, and examination uncovered dysmorphic facial features including hypertelorism, downward eyeslant, and low-set ears, constituting a clinical phenotype of syndromic NDD.

3.6. Endocrine and Other Genetic Disorders. As shown in Table 3, endocrine and other genetic disorders also played important roles in NDD development in this cohort. Figure 6 clearly demonstrated not only the muscular lesions on electronic microscopy observation, but also the existence of leukodystrophy in MRI which indicated the involvement of central nervous system and thus explained in part the neurodevelopmental retardation in the patient with congenital muscular dystrophy. It is also noteworthy that a hypothyroidism patient in this study was combined with hypophosphatasia with remarkable delay in bone ossification as the radiological feature due to the reduced activity of alkaline phosphatase. To our knowledge, the combined clinical spectra of hypothyroidism and hypophosphatasia in the same patient, once again, have never been reported in other references.

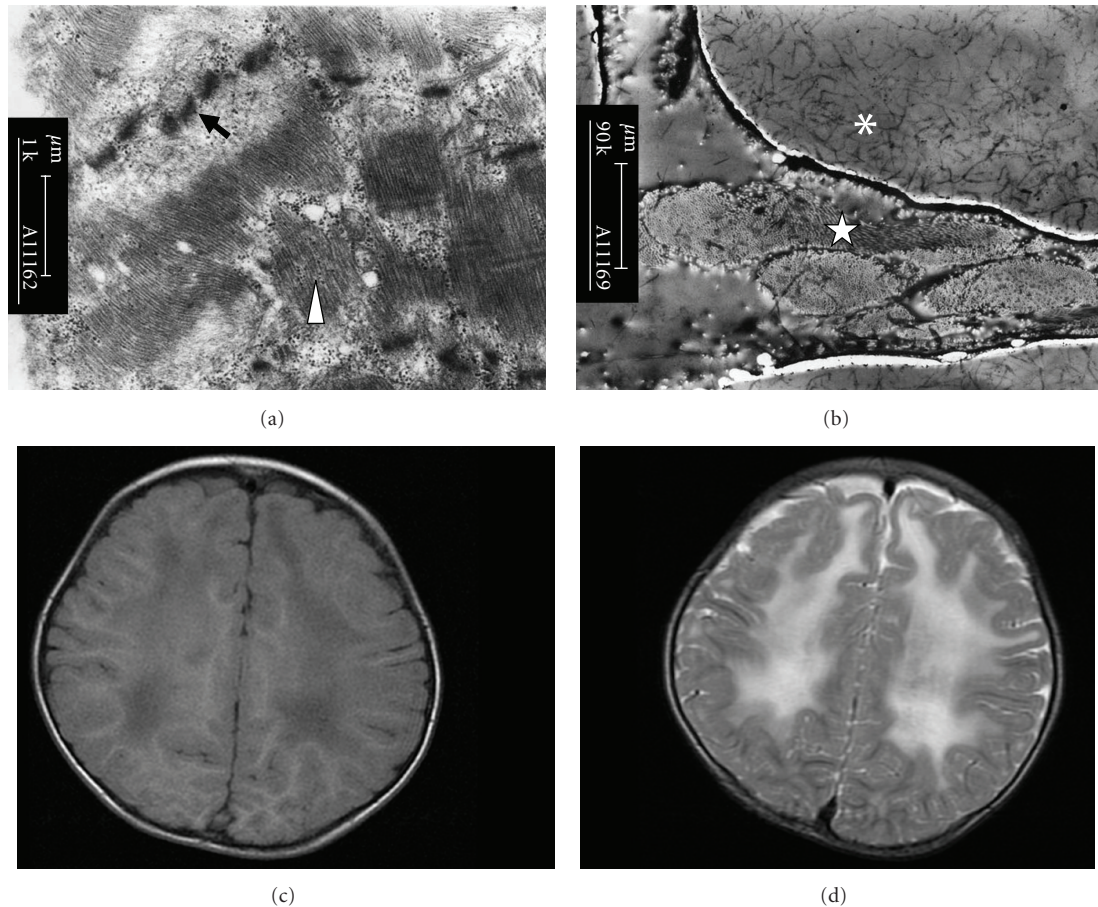


FIGURE 6: Muscular ultrastructure and brain MRI findings in a patient with congenital muscular dystrophy. In Figure 6(a), the myofibril Z lines in a muscle cell were ruptured (†), and the filaments between Z lines were arranged disorderly (Δ). Figure 6(b) showed the fatty degeneration of myofibrils (*) in a muscle cell with completely vanished sarcolemma, and large quantity of collagenous fibers (☆) were observed in the endomysium (Bar = 1 μm). Figures 6(c) and 6(d) were cranial MRI findings revealing the reduced signal intensity in T1WI while increased one in T2WI, respectively, in bilateral frontal and parietal lobes, both indicating leukodystrophy.

4. Discussion

The findings in this study provided the latest clinical epidemiology information on the etiology distribution of NDDs in Chinese children. As shown in Table 2, nearly half of the NDD cases could not be attributed to concrete etiologies. This finding indicated the current issue or challenge that we are facing and was consistent with the well-known fact that the etiologies in a substantial percentage of NDDs patients are undiagnosed even after a comprehensive evaluation [13]. On the other hand, concrete etiologies were identified for 51.6% of the total NDD cases, and the etiology distribution in this cohort demonstrated a rather heterogeneous feature. Autism spectrum disorders (ASDs) have been well recognized as pervasive NDDs entities, and there have been evidences suggesting that some genes or chromosomal aberrations are associated with ASDs [14–16]. However, genetic studies have not provided substantial insight into the 90% of cases of autism whose cause is idiopathic, and the relative genetic contribution to a susceptibility to autism from de novo mutations, rare

mutations, and common polymorphisms has been debated extensively [17]. Therefore, autism was categorized into psychobehavioral disorders other than genetic diseases in this paper. This investigation also found 7 NDD cases caused by kernicterus and hypoxic and ischemic encephalopathy (HIE) which were secondary to 2 curable and even preventable diseases, hyperbilirubinemia and birth asphyxia, respectively. The existence of these 7 cases further suggested the brain vulnerability to exogenous injuries in children, especially in fetuses and neonates. We also found cerebral palsy and epilepsy as NDD etiologies in this NDD cohort, with several cases of refractory epilepsy. However, it is noteworthy that, considering the brain vulnerability once again and the possible effects of antiepileptic drugs (AEDs) on neurodevelopment [18, 19], every child with epilepsy must be evaluated on an individual basis as to the risk and benefit of any particular AED used and the role of ongoing treatment [20].

In this eight-year clinical study, various laboratory technologies were conducted in the fields of clinical biochemistry, neuroimaging, biochemical genetics, cytogenetics, molecular

TABLE 5: Congenital dysmorphic disorders in the patients with genetic diseases in Table 3.

| Disorders | Cases | Main clinical/imaging features |
|---|-------|--|
| Cortical dysplasia | 1 | Motor developmental retardation and microcephaly. Severe cortex dysplasia in parietal lobe and frontal, temporal and occipital lobes, respectively, on cranial MRI scanning. |
| Tuberous sclerosis | 1 | Intelligence and motor retardation, seizures, cutaneous hypomelanotic macules, fundus oculi depigmentation, and subependymal nodules and calcified lesions in the cortex of parietal and temporal lobes on cranial CT scanning |
| Miller-Dieker syndrome | 2 | Intelligence and motor retardation, microcephaly, prominent occiput, narrow forehead, small nose and chin. Seizure in 1 case and hypertonia in another one. Agyria/pachygyria cortical malformations on MRI. |
| Muscle-Eye-Brain disease | 2 | Sibling sisters with global developmental delay. Abnormal pupils and vitreous bodies in both cases on ophthalmologic examination. Convulsions in 1 case and small right eyeball in another. Both have increased creatine kinase levels and cobblestone cortical malformations on MRI. |
| Isolated lissencephaly sequence | 1 | Intelligence and motor retardation, and bilateral thickened and irregular cortex on MRI |
| Lissencephaly with cerebellar hypoplasia | 1 | Mental/language retardation, drooling, and bilateral pachygyria malformation and hypoplasia of cerebellum revealed by MRI. |
| Other malformations of cortical development | 6 | All have intelligence and motor retardation. Including 2 cases of cobblestone cortical malformations and 1 classic lissencephaly revealed by CT/MRI. |
| Dandy-Walker syndrome | 1 | Intelligence and motor retardation with low-set and everted ears, and hypoplasia and upward rotation of the cerebellar vermis and cystic dilation of the fourth ventricle on MRI. |
| Spinocerebellar ataxia | 1 | Intelligence and motor retrogression, and severe cerebellar and pons atrophy together with tiger-eye-like sign at the basal ganglia level and cross-sign at pons level, respectively, on MRI. |
| Neurofibromatosis type I | 1 | Intelligence and motor retardation, and 7 cutaneous cafe-au-lait patches with diameter over 10 mm |
| Silver-Russell syndrome | 2 | Both have dysmorphic facial features including triangular face, low-set ears, flat nasal bridge with extroversion of nostrils and down-curving mouth corners. Normal head circumference. Asymmetry of the lower extremities. Postnatal failure to thrive. Intrauterine retardation in 1 case and linea alba hernia in another. |
| Noonan syndrome | 1 | Short stature, short neck with redundancy of skin, hypertelorism, downward eyeslant, low-set ears, cryptorchidism, and poor sucking. Atrial septal defect and right pulmonic stenosis on ultrasonography. |
| Poland-Moebius syndrome | 1 | Signs of facial palsy, disappeared corneal reflex, and poor sucking and swallowing. Micrognathia and high-arched palate. Small left hand, ipsilateral brachydactyly and hypoplasia of the nails and pectoralis major muscle. |
| Crisponi syndrome | 1 | Convulsions in response to stimuli like crying and bathing. Camptodactyly in hands. Round face, broad nose with anteverted nostrils, and micrognathia. Major sucking difficulty and frequent apnea. Hyperthermia that led to death. |

genetics, and electronic microscopy as well. Our experiences strongly supported the viewpoint that detailed history inquiry and physical examination are paramount in the evaluation of NDDs, while judicious investigations can be useful adjunct in determining etiologies [21]. As the commonest etiology category in this NDD cohort, genetic diseases, as shown in Table 3, also covered a wide profile of different entities, with IEMs on top of the list. The application of two metabolomic technologies, GC-MS and MS-MS, played irreplaceable roles in the identification of various IEM entities in this study (Table 4). Although other skills such as enzymatic activity and gene mutation

analysis help us to confirm the diagnosis of IEM in some cases, selective screening for IEMs by means of GC-MS and MS-MS provides the basis or prerequisite for further intensive investigation. The clinical application of these metabolomic tools in mainland of China was initiated from the beginning of this century, however, their indispensable function in chemical diagnosis of IEMs has been affirmed by more and more clinical evidences in our pediatric practice [6, 22, 23]. Actually, mass spectrometry has occupied an increasingly prominent place in clinical chemistry and laboratory diagnostics during the past few decades [24], and nowadays has been recognized as one of the most frequently

employed methods of detection in the field of metabolome [25]. However, the metabolomic technologies including GC-MS and MS-MS analysis in this study sometimes yields nonspecific and nondiagnostic abnormalities. An example is the metabolome feature of NICCD. This entity was prone to be misdiagnosed as tyrosinemia or galactosemia in case just based on the GC-MS and/or MS-MS findings [26, 27].

Citrin deficiency and a rare chromosome karyotype were found in this study as novel causes for GDD and MR, respectively, further expanding the etiology profile of NDDs. Many “new” IEMs have been discovered in the recent years, and some of them have been found to have features of NDDs [28]. Citrin deficiency, first described in Japan and East Asia, is a newly-established IEM now considered as a panethnic disease distributed worldwide [27, 29]. We have no direct evidence to explain why the 2 citrin-deficient patients in this study presented with syndromic GDDs. However, their GDD manifestations could not be explained with brain injury directly due to deficient citrin, since this aspartate/glutamate carrier (AGC) is predominantly found in liver, kidney, heart, and intestine, but not in brain [29]. Interplay of multiple factors may represent the possible underlying mechanism. For example, secondary galactosemia and tyrosinemia in citrin deficiency, including the 2 cases in this paper, are quite common [26, 27], and brain injuries caused by disturbances in the metabolism of galactose and tyrosine have been reported [12, 30]. The detailed mechanism(s) for the child with novel karyotype to develop MR remains unclear at the current stage. However, the chromosome aberration in this patient produced 4 chromosome breakpoints, changing the genome architecture inevitably, and thus the syndromic MR may result from one or more mechanisms including gene interruption, gene fusion, position effect, unmasking of a recessive allele, presence of a functional polymorphism, and gene transvection effect [11, 31], as proved in other disorders of genome architecture such as acute lymphoblastic leukemia [32] and retinitis pigmentosa [33].

5. Conclusions

Although the etiologies in nearly half of the patients still remain an unresolved issue in this NDD cohort, genetic diseases including IEMs and congenital dysmorphic diseases constituted the commonest identified etiology category. GC-MS and MS-MS, the two key experimental technologies in the field of pediatric metabolomics, proved to be substantially helpful for the exploration of the NDD etiologies in our clinical practice. The findings in this paper provided latest epidemiologic information on the etiology distribution of NDDs in Chinese, and syndromic NDDs caused by citrin deficiency and the novel chromosomal karyotype, respectively, further expanded the etiology spectrum of NDDs.

Acknowledgments

The authors are grateful to all the patients and their parents for their cooperation, and to those who referred NDD

patients to our department. This paper was supported financially in part by Medical Research Fund of Guangdong Province, China (Nos. A2008358 & A2009366) and Project 81070279 supported by National Natural Science Foundation of China (NSFC), and by Grants-in-Aid for Scientific Research (B: Nos. 16390100 & 19390096) and for Asia-Africa Scientific Platform Program (AASPP) from the Japan Society for the Promotion of Science.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

References

- [1] M. Shevell, “Global developmental delay and mental retardation or intellectual disability: conceptualization, evaluation, and etiology,” *Pediatric Clinics of North America*, vol. 55, no. 5, pp. 1071–1084, 2008.
- [2] M. Shevell, S. Ashwal, D. Donley et al., “Practice parameter: evaluation of the child with global developmental delay: report of the quality standards subcommittee of the American Academy of Neurology and The Practice Committee of the Child Neurology Society,” *Neurology*, vol. 60, no. 3, pp. 367–380, 2003.
- [3] J. Chelly, M. Khelifaoui, F. Francis, B. Chérif, and T. Bienvenu, “Genetics and pathophysiology of mental retardation,” *European Journal of Human Genetics*, vol. 14, no. 6, pp. 701–713, 2006.
- [4] M. Yeargin-Allsopp, C. C. Murphy, J. F. Cordero, P. Decouflé, and J. G. Hollowell, “Reported biomedical causes and associated medical conditions for mental retardation among 10 year old children, metropolitan Atlanta, 1985 to 1987,” *Developmental Medicine and Child Neurology*, vol. 39, no. 3, pp. 142–149, 1997.
- [5] Y. W. Jiang, “Enhance neurogenetic etiologic studies actively and standardize diagnosis of neurodevelopmental disorders,” *Zhonghua Er Ke Za Zhi*, vol. 47, pp. 561–564, 2009 (Chinese).
- [6] Y. Z. Song, B. X. Li, H. Hao et al., “Selective screening for inborn errors of metabolism and secondary methylmalonic aciduria in pregnancy at high risk district of neural tube defects: a human metabolome study by GC-MS in China,” *Clinical Biochemistry*, vol. 41, no. 7-8, pp. 616–620, 2008.
- [7] L. S. Han, J. Ye, W. J. Qiu, X. L. Gao, Y. Wang, and X. F. Gu, “Selective screening for inborn errors of metabolism on clinical patients using tandem mass spectrometry in China: a four-year report,” *Journal of Inherited Metabolic Disease*, vol. 30, no. 4, pp. 507–514, 2007.
- [8] D. H. Cheng, Y. Q. Tan, Y. F. Di, L. Y. Li, and G. X. Lu, “Crypt Y chromosome fragment resulting from an X;Y translocation in a patient with premature ovarian failure,” *Fertility and Sterility*, vol. 92, no. 2, pp. 828–e3, 2009.
- [9] Y. Z. Song, M. Ushikai, K. Kobayashi, and T. Saheki, “Citrin deficiency is an important etiology for cholestatic liver disease in children,” *Zhonghua Er Ke Za Zhi*, vol. 47, no. 8, pp. 624–627, 2009 (Chinese).
- [10] M. G. Mohaupt, R. H. Karas, E. B. Babiychuk et al., “Association between statin-associated myopathy and skeletal muscle damage,” *Canadian Medical Association Journal*, vol. 181, no. 1-2, pp. E11–E18, 2009.
- [11] D. Kumar, “Disorders of the genome architecture,” *Genomic Medicine*, vol. 2, no. 3-4, pp. 69–76, 2008.

- [12] A. M. Bosch, H. D. Bakker, A. H. van Gennip, J. V. van Kempen, R. J. A. Wanders, and F. A. Wijburg, "Clinical features of galactokinase deficiency: a review of the literature," *Journal of Inherited Metabolic Disease*, vol. 25, no. 8, pp. 629–634, 2002.
- [13] J. B. Moeschler, "Genetic evaluation of intellectual disabilities," *Seminars in Pediatric Neurology*, vol. 15, no. 1, pp. 2–9, 2008.
- [14] J. T. Glessner, K. Wang, G. Cai et al., "Autism genome-wide copy number variation reveals ubiquitin and neuronal genes," *Nature*, vol. 459, no. 7246, pp. 569–572, 2009.
- [15] P. Szatmari, A. D. Paterson, L. Zwaigenbaum et al., "Mapping autism risk loci using genetic linkage and chromosomal rearrangements," *Nature Genetics*, vol. 39, no. 3, pp. 319–328, 2007.
- [16] L. A. Weiss, Y. Shen, J. M. Korn et al., "Association between microdeletion and microduplication at 16p11.2 and autism," *The New England Journal of Medicine*, vol. 358, no. 7, pp. 667–675, 2008.
- [17] X. Zhao, A. Leotta, V. Kustanovich et al., "A unified genetic theory for sporadic and inherited autism," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 31, pp. 12831–12836, 2007.
- [18] A. P. Aldenkamp, G. Baker, O. G. Mulder et al., "A multicenter, randomized clinical study to evaluate the effect on cognitive function of topiramate compared with valproate as add-on therapy to carbamazepine in patients with partial-onset seizures," *Epilepsia*, vol. 41, no. 9, pp. 1167–1178, 2000.
- [19] D. W. Loring and K. J. Meador, "Cognitive side effects of antiepileptic drugs in children," *Neurology*, vol. 62, no. 6, pp. 872–877, 2004.
- [20] J. H. Cross, "Neurodevelopmental effects of anti-epileptic drugs," *Epilepsy Research*, vol. 88, no. 1, pp. 1–10, 2010.
- [21] L. McDonald, A. Rennie, J. Tolmie, P. Galloway, and R. McWilliam, "Investigation of global developmental delay," *Archives of Disease in Childhood*, vol. 91, no. 8, pp. 701–705, 2006.
- [22] L. S. Han, J. Ye, W. J. Qiu et al., "Diagnosis of inborn errors of metabolism using tandem mass spectrometry and gas chromatography mass spectrometry," *Zhonghua Yi Xue Za Zhi*, vol. 88, no. 30, pp. 2122–2126, 2008 (Chinese).
- [23] X. P. Luo, M. T. Wang, H. Wei et al., "Application of gas chromatography-mass spectrometry analysis on urine filter paper in the high-risk screening and diagnosis of inherited metabolic diseases," *Zhonghua Er Ke Za Zhi*, vol. 41, no. 4, pp. 245–248, 2003 (Chinese).
- [24] A. R. Spitzer and D. Chace, "Proteomics- and metabolomics-based neonatal diagnostics in assessing and managing the critically ill neonate," *Clinics in Perinatology*, vol. 35, no. 4, pp. 695–716, 2008.
- [25] S. G. Villas-Bôas, S. Mas, M. Åkesson, J. Smedsgaard, and J. Nielsen, "Mass spectrometry in metabolome analysis," *Mass Spectrometry Reviews*, vol. 24, no. 5, pp. 613–646, 2005.
- [26] T. Ohura, K. Kobayashi, Y. Tazawa et al., "Clinical pictures of 75 patients with neonatal intrahepatic cholestasis caused by citrin deficiency (NICCD)," *Journal of Inherited Metabolic Disease*, vol. 30, no. 2, pp. 139–144, 2007.
- [27] Y. Z. Song, B. X. Li, F. P. Chen et al., "Neonatal intrahepatic cholestasis caused by citrin deficiency: clinical and laboratory investigation of 13 subjects in mainland of China," *Digestive and Liver Disease*, vol. 41, no. 9, pp. 683–689, 2009.
- [28] M. A. Kayser, "Inherited metabolic diseases in neurodevelopmental and neurobehavioral disorders," *Seminars in Pediatric Neurology*, vol. 15, no. 3, pp. 127–131, 2008.
- [29] T. Saheki, K. Inoue, A. Tushima, K. Mutoh, and K. Kobayashi, "Citrin deficiency and current treatment concepts," *Molecular Genetics and Metabolism*, vol. 100, pp. S59–S64, 2010.
- [30] A. M. Sgaravatti, B. A. Vargas, B. R. Zandoná et al., "Tyrosine promotes oxidative stress in cerebral cortex of young rats," *International Journal of Developmental Neuroscience*, vol. 26, no. 6, pp. 551–559, 2008.
- [31] J. R. Lupski and P. Stankiewicz, "Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes," *PLoS Genetics*, vol. 1, no. 6, article e49, 2005.
- [32] C. M. Rubin, J. J. Carrino, M. N. Dickler, D. Leibowitz, S. D. Smith, and C. A. Westbrook, "Heterogeneity of genomic fusion of BCR and ABL in Philadelphia chromosome-positive acute lymphoblastic leukemia," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 85, no. 8, pp. 2795–2799, 1988.
- [33] L. Köhn, S. J. Bowne, L. S. Sullivan et al., "Breakpoint characterization of a novel ~59 kb genomic deletion on 19q13.42 in autosomal-dominant retinitis pigmentosa with incomplete penetrance," *European Journal of Human Genetics*, vol. 17, no. 5, pp. 651–655, 2009.