ACKNOWLEDGEMENTS

The authors gratefully acknowledge Shinobu Abe and Nami Abe of the Institute for Advanced Biosciences, Keio University, Japan, for their help with measurement and data analysis.

FUNDING

Funding for open access charge: Yamagata Prefectural Government and Tsuruoka City, Japan.

Conflict of interest statement. None declared.

REFERENCES

1. Fiehn,O. (2002) Metabolomics–the link between genotypes and phenotypes. *Plant Mol. Biol.*, **48**, 155–171.
2. Reo,N.V. (2002) NMR-based metabolomics. *Drug Chem. Toxicol.*, **25**, 375–382.
3. Aharoni,A., Ric de Vos,C.H., Verhoeven,H.A., Maliepaard,C.A., Kruppa,G., Bino,R. and Goodenowe,D.B. (2002) Nontargeted metabolome analysis by use of Fourier Transform Ion Cyclotron Mass Spectrometry. *Omics*, **6**, 217–234.
4. Castrillo,J.I., Hayes,A., Mohammed,S., Gaskell,S.J. and Oliver,S.G. (2003) An optimized protocol for metabolome analysis in yeast using direct infusion electrospray mass spectrometry. *Phytochemistry*, **62**, 929–937.
5. Fiehn,O., Kopka,J., Trethewey,R.N. and Willmitzer,L. (2000) Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. *Anal. Chem.*, **72**, 3573–3580.
6. Plumb,R., Granger,J., Stumpf,C., Wilson,I.D., Evans,J.A. and Lenz,E.M. (2003) Metabonomic analysis of mouse urine by liquid-chromatography-time of flight mass spectrometry (LC-TOFMS): detection of strain, diurnal and gender differences. *Analyst*, **128**, 819–823.
7. Soga,T., Ohashi,Y., Ueno,Y., Naraoka,H., Tomita,M. and Nishioka,T. (2003) Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. *J. Proteome Res.*, **2**, 488–494.
8. Monton,M.R. and Soga,T. (2007) Metabolome analysis by capillary electrophoresis-mass spectrometry. *J. Chromatogr. A*, **1168**, 237–246; discussion 236.
9. Kanehisa,M., Goto,S., Furumichi,M., Tanabe,M. and Hirakawa,M. (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.*, **38**, D355–D360.
10. Caspi,R., Altman,T., Dale,J.M., Dreher,K., Fulcher,C.A., Gilham,F., Kaipa,P., Karthikeyan,A.S., Kothari,A., Krummenacker,M. *et al.* (2010) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.*, **38**, D473–D479.
11. Grafahrend-Belau,E., Weise,S., Koschutzki,D., Scholz,U., Junker,B.H. and Schreiber,F. (2008) MetaCrop: a detailed database of crop plant metabolism. *Nucleic Acids Res.*, **36**, D954–D958.
12. Croft,D., O'Kelly,G., Wu,G., Haw,R., Gillespie,M., Matthews,L., Caudy,M., Garapati,P., Gopinath,G., Jassal,B. *et al.* (2011) Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.*, **39**, D691–D697.
13. Wishart,D.S., Knox,C., Guo,A.C., Eisner,R., Young,N., Gautam,B., Hau,D.D., Psychogios,N., Dong,E., Bouatra,S. *et al.* (2009) HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.*, **37**, D603–D610.
14. Frolkis,A., Knox,C., Lim,E., Jewison,T., Law,V., Hau,D.D., Liu,P., Gautam,B., Ly,S., Guo,A.C. *et al.* (2010) SMPDB: the Small Molecule Pathway Database. *Nucleic Acids Res.*, **38**, D480–D487.
15. Kopka,J., Schauer,N., Krueger,S., Birkemeyer,C., Usadel,B., Bergmuller,E., Dörmann,P., Weckwerth,W., Gibon,Y., Stitt,M. *et al.* (2005) GMD@CSB.DB: the Golm Metabolome Database. *Bioinformatics*, **21**, 1635–1638.
16. Smith,C.A., O'Maille,G., Want,E.J., Qin,C., Trauger,S.A., Brandon,T.R., Custodio,D.E., Abagyan,R. and Siuzdak,G. (2005) METLIN: a metabolite mass spectral database. *Ther. Drug Monit.*, **27**, 747–751.
17. Horai,H., Arita,M., Kanaya,S., Nihei,Y., Ikeda,T., Suwa,K., Ojima,Y., Tanaka,K., Tanaka,S., Aoshima,K. *et al.* (2010) MassBank: a public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom*, **45**, 703–714.
18. Soga,T., Baran,R., Suematsu,M., Ueno,Y., Ikeda,S., Sakurakawa,T., Kakazu,Y., Ishikawa,T., Robert,M., Nishioka,T. *et al.* (2006) Differential metabolomics reveals ophthalmic acid as an oxidative stress biomarker indicating hepatic glutathione consumption. *J. Biol. Chem.*, **281**, 16768–16776.
19. Soga,T., Igarashi,K., Ito,C., Mizobuchi,K., Zimmermann,H.P. and Tomita,M. (2009) Metabolomic Profiling of Anionic Metabolites by Capillary Electrophoresis Mass Spectrometry. *Anal. Chem.*, **81**, 6165–6174.
20. Sato,H., Shiiya,A., Kimata,M., Maebara,K., Tamba,M., Sakakura,Y., Makino,N., Sugiyama,F., Yagami,K., Moriguchi,T. *et al.* (2005) Redox imbalance in cystine/glutamate transporter-deficient mice. *J. Biol. Chem.*, **280**, 37423–37429.
21. Sugimoto,M., Wong,D.T., Hirayama,A., Soga,T. and Tomita,M. (2010) Capillary electrophoresis mass spectrometry-based saliva metabolomics identified oral, breast and pancreatic cancer-specific profiles. *Metabolomics*, **6**, 78–95.
22. Brown,M., Dunn,W.B., Dobson,P., Patel,Y., Winder,C.L., Francis-McIntyre,S., Begley,P., Carroll,K., Broadhurst,D., Tseng,A. *et al.* (2009) Mass spectrometry tools and metabolite-specific databases for molecular identification in metabolomics. *Analyst*, **134**, 1322–1332.
23. Baran,R., Kochi,H., Saito,N., Suematsu,M., Soga,T., Nishioka,T., Robert,M. and Tomita,M. (2006) MathDAMP: a package for differential analysis of metabolite profiles. *BMC Bioinformatics*, **7**, 530.
24. Saeed,A.I., Sharov,V., White,J., Li,J., Liang,W., Bhagabati,N., Braisted,J., Klapa,M., Currier,T., Thiagarajan,M. *et al.* (2003) TM4: a free, open-source system for microarray data management and analysis. *Biotechniques*, **34**, 374–378.
25. Degtyarenko,K., Hastings,J., de Matos,P. and Ennis,M. (2009) ChEBI: an open bioinformatics and cheminformatics resource. *Curr. Protoc. Bioinformatics*, **Chapter 14**, Unit 14.19.
26. Reijeng,J.C., Martens,J.H., Giuliani,A. and Chiari,M. (2002) Pherogram normalization in capillary electrophoresis and micellar electrokinetic chromatography analyses in cases of sample matrix-induced migration time shifts. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.*, **770**, 45–51.
27. Kono,N., Arakawa,K., Ogawa,R., Kido,N., Oshita,K., Ikegami,K., Tamaki,S. and Tomita,M. (2009) Pathway projector: web-based zoomable pathway browser using KEGG atlas and Google Maps API. *PLoS One*, **4**, e7710.
28. Junker,B.H., Klukas,C. and Schreiber,F. (2006) VANTED: a system for advanced data analysis and visualization in the context of biological networks. *BMC Bioinformatics*, **7**, 109.
29. Romero,P., Wagg,J., Green,M.L., Kaiser,D., Krummenacker,M. and Karp,P.D. (2004) Computational prediction of human metabolic pathways from the complete human genome. *Genome Biol.*, **6**, R2.
30. Evsikov,A.V., Dolan,M.E., Genrich,M.P., Patek,E. and Bult,C.J. (2009) MouseCyc: a curated biochemical pathways database for the laboratory mouse. *Genome Biol.*, **10**, R84.
31. Okuda,S., Yamada,T., Hamajima,M., Itoh,M., Katayama,T., Bork,P., Goto,S. and Kanehisa,M. (2008) KEGG Atlas mapping for global analysis of metabolic pathways. *Nucleic Acids Res.*, **36**, W423–W426.
32. Baena,B., Cifuentes,A. and Barbas,C. (2005) Analysis of carboxylic acids in biological fluids by capillary electrophoresis. *Electrophoresis*, **26**, 2622–2636.
33. Ramautar,R., Mayboroda,O.A., Somsen,G.W. and de Jong,G.J. (2011) CE-MS for metabolomics: developments and applications in the period 2008–2010. *Electrophoresis*, **32**, 52–65.

34. Garcia-Perez,I., Whitfield,P., Bartlett,A., Angulo,S., Legido-Quigley,C., Hanna-Brown,M. and Barbas,C. (2008) Metabolic fingerprinting of Schistosoma mansoni infection in mice urine with capillary electrophoresis. *Electrophoresis*, **29**, 3201–3206.
35. Fan,T.W., Lane,A.N., Higashi,R.M. and Yan,J. (2011) Stable isotope resolved metabolomics of lung cancer in a SCID mouse model. *Metabolomics*, **7**, 257–269.
36. Sugimoto,M., Hirayama,A., Robert,M., Abe,S., Soga,T. and Tomita,M. (2010) Prediction of metabolite identity from accurate mass, migration time prediction and isotopic pattern information in CE-TOFMS data. *Electrophoresis*, **31**, 2311–2318.
37. Chalcraft,K.R., Lee,R., Mills,C. and Britz-McKibbin,P. (2009) Virtual quantification of metabolites by capillary electrophoresis-electrospray ionization-mass spectrometry: predicting ionization efficiency without chemical standards. *Anal. Chem.*, **81**, 2506–2515.

Metabolomics (2012) 8:99–108
DOI 10.1007/s11306-011-0290-7

ORIGINAL ARTICLE

Profiling of the charged metabolites of traditional herbal medicines using capillary electrophoresis time-of-flight mass spectrometry

Keiko Iino · Masahiro Sugimoto · Tomoyoshi Soga · Masaru Tomita

Received: 25 November 2010 / Accepted: 14 February 2011 / Published online: 27 February 2011

Abstract The quantification of a small number of bioactive components in herbal medicines is often inadequate when attempting to elucidate a medicine's biological effects. Despite rapid advances in analytical technologies, obtaining comprehensive metabolomic profiles of herbal medicines remains difficult, due to the complexity of natural product mixtures. *Toki-Shakuyaku-San* is a Chinese medicine used widely to treat gynecological and obstetric disorders, such as infertility, dysmenorrhea, toxemia during pregnancy and neural dysfunction. It consists of *Angelica acutiloba Radix* (Toki), *Cnidium officinale Rhizoma* (Senkyu), *Paeonia lactiflora Radix* (Shakuyaku), *Atractylodes lancea Rhizoma* (Sojutsu), *Alisma orientale Rhizoma* (Takusha) and *Poria cocos Hoelen* (Bukuryo). To elucidate the composition of these herbal medicines individually, we conducted non-targeted profiling analyses of extracts of these herbs using capillary electrophoresis time-of-flight mass spectrometry (CE-TOFMS), which allows the simultaneous quantification of hundreds of charged metabolites. In total, 737 ± 183.1 (average ± SD) metabolite-derived features were observed, and of these, 119 metabolites were identified. Score plots of principal component analysis (PCA) showed a clear cluster including Shakuyaku, Bukuryo, and Sojutsu, while the other three herbs were distributed over PCA spaces. Loading plots revealed that amino acids and shikimate-derived alkaloids were the predominant metabolite constituents. Hierarchical clustering analysis revealed that few clusters overlapped in the herbal medicines tested. This report is the first demonstration of the characterization of a herbal medicine using large-scale metabolomic analysis, which is complementary to traditional quality control methods.

Keywords Capillary electrophoresis time-of-flight mass spectrometry · Herbal medicine · Charged metabolite · Metabolomic profiling

1 Introduction

Herbal medicines have been used as therapeutic agents for several thousands of years in Asian countries (Wang and Ren 2002; Wang et al. 2009). The evaluation of a herbal medicine's safety and efficacy is required before its clinical use as a complementary and alternative medicine (CAM) in western countries (Miller et al. 2004; Murray and Rubel 1992; Calixto 2000; Zhang et al. 2010). The conventional pharmacological approach to characterize herbal medicines focuses on the identification and quantification of a single or several bioactive components (Murray and Rubel 1992; Liu et al. 2008), however, the complex properties of herbal medicines often render such an approach inadequate (Wang and Ren 2002; Lao et al. 2009; Liang et al. 2009; Chan 2003). The quantitative analysis of the complete metabolite profile of herbal medicines would therefore be of significant value (Liu et al. 2008).

Electronic supplementary material The online version of this article (doi:10.1007/s11306-011-0290-7) contains supplementary material, which is available to authorized users.

K. Iino · M. Sugimoto (✉) · T. Soga · M. Tomita
Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata 997-0052, Japan
e-mail: msugi@sfc.keio.ac.jp

K. Iino · T. Soga · M. Tomita
Department of Environment and Information Studies, Keio University, Fujisawa, Kanagawa 252-8520, Japan

M. Sugimoto · T. Soga · M. Tomita
Systems Biology Program, Graduate School of Media and Governance, Keio University, Fujisawa, Kanagawa 252-8520, Japan

Recent advances in the development of mass spectrometry (MS)-based profiling techniques have made a significant contribution to the study and quantification of the metabolome, the complete range of low molecular weight compounds of a natural sample, which is a field known as metabolomics. MS is typically used in combination with a separation system, such as liquid chromatography (LC), gas chromatography (GC) or capillary electrophoresis (CE). The selection of the separation system to be used is based on the chemical properties of the target compounds, since there is no single analytical methodology currently available capable of profiling complete metabolome. GC–MS and LC–MS are used routinely for targeted analysis (Liang et al. 2009) and are considered to be mature techniques (Weng and Jin 2002). GC–MS is limited to profiling volatile molecules, such as essential oils (Li et al. 2003; Gong et al. 2001), while LC–MS is suitable for separating and detecting more diverse molecules, such as flavonoids, glycosides, organic acids, saponins and lipids (Lao et al. 2009; Ma et al. 2007; Ren et al. 2008). However, the separation conditions used for LC should be optimized for individual molecular classes, which limits the number of metabolites simultaneously detected. CE coupled MS, in particular CE time-of-flight MS (CE-TOFMS), has been demonstrated to have high separation capability and sensitivity for the profiling of charged metabolites, including primary metabolites such as amino acids, amines, organic acids and nucleic acids (Soga et al. 2003, 2006; Monton and Soga 2007). CE alone has been used for the targeted analysis of alkaloids and flavonoids (Ganzera 2008; Lao et al. 2009; Hurtado-Fernandez et al. 2010) and used for quality control (QC) of herbal medicines (Ganzera 2008). LC-MS-based profiling showed variance attributed to harvesting region and processing protocol (Xie et al. 2008; Chan et al. 2007). Our CE–MS-based profiling method is also important for QC of herbal medicines, since charged metabolites are the dominant species dissolved in water extracts (Lao et al. 2009).

Toki-Shakuyaku-San, or *Danggui-Shaoyao-San*, is a Chinese herbal medicine used widely for the therapy of hematopoiesis and menstrual disorders, and has several functions, including smooth muscle relaxation and biofluid control (Akase et al. 2004). *Toki-Shakuyaku-San* consists of six individual herbal medicines, with each providing different pharmacological effects. These herbal medicines are Toki (*Angelica acutiloba Radix*), Senkyu (*Cnidium officinale Rhizoma*), Shakuyaku (*Paeonia lactiflora Radix*), Sojutsu (*Atractylodes lancea Rhizoma*), Takusha (*Alisma orientale Rhizoma*), and Bukuryo (*Poria cocos Hoelen*). Toki and Senkyu are used for the treatment of gynecological diseases (Lu et al. 2005; Kim et al. 2006; Yi et al. 2007; Bohrmann et al. 1967), while Takusha, Bukuryo and Sojutsu are used to improve water metabolism (Wang et al. 2004, 2008; Kitajima et al. 2003; Zhao et al. 2008), and Shakuyaku provides an anti-inflammatory effect (He et al. 2010; Ohta et al. 1993). While some of the active components, such as L-ornithine, gallic acid, paeoniflorin, ferulic acid and benzoic acid in this mixture, and their effects, have been well studied, the relationships between the complete compound profile and the pharmacological effects is not yet well understood (Hatip-Al-Khatib et al. 2004; Chen et al. 2009).

This study aimed to reveal the variance of the charged metabolites found in the six herbal medicines that constitute *Toki-Shakuyaku-San*, using CE-TOFMS. We conducted multivariate analysis to examine variations in the primary metabolites of the herbal medicine and to understand the relationships between the charged metabolites and the pharmacological effects. Most studies focus on the importance of secondary metabolites to the pharmacological effects of a herbal medicine, however we show that the profile of charged metabolites also varies among herbal medicines.

2 Materials and methods

2.1 Sample preparation

Six herbal medicines, Toki, Bukuryo, Sojutsu, Takusha, Shakuyaku and Senkyu, were obtained from Yatsume Pharmaceutical Co., Ltd. (Tokyo, Japan). Each herb was homogenized without solvent using a multi-bead shocker (Yasui Kikai Co., Osaka, Japan) at 2,500 rpm for 300 s. The crushed material (1.00 g) was dissolved in 3 ml of 50% MeOH containing 2 internal standards (200 μM each of methionine sulfone and 3-aminopyrrolidine). This solution was centrifuged at 3,000 rpm for 10 min at 4°C, and 0.2 ml of the supernatant fluid was transferred to a Millipore 5 kDa cutoff filter tube for centrifugal centrifugation (10,000 rpm × 2 h at 4°C). The filtrate was lyophilized and dissolved in 100 μl of Milli-Q (Millipore, Bedford, MA, USA) water prior to CE-TOFMS analysis.

2.2 Standard chemical compounds

Cysteine-glutathione disulphide was purchased from Apollo Scientific Ltd (Tokyo, Japan). 5-Methyl-2′-deoxycytidine was purchased from MP Biomedicals LLC. (Tokyo, Japan). γ-Glu-2AB and cystathionine were purchased from Toray Research Center (Tokyo, Japan). *N*-α-Dimethylhistidine was purchased from Bachem AG (Bubendorf, Switzerland). Isobutylamine and azetidine-2-carboxylate were purchased from Chem Service Inc. (West Chester, PA, USA). N_8-Acetylspermidine, 5-aminovalerate, *O*-acetylcarnitine,

betaine, glucosaminate, 7-methylguanine, 5-methoxy-3-indoleacetate and pyridoxamine were purchased from Fluka (Buchs, Switzerland). Isonicotinamide was purchased from Tokyo Chemical Industry (Tokyo, Japan). All other compounds were purchased from Sigma–Aldrich (St. Louis, MO, USA) or Wako (Osaka, Japan). All chemical standards were dissolved in Milli-Q water, 0.1 M HCl or 0.1 M NaOH to give 10 mM or 100 mM stock solutions. A standard mixture was prepared by diluting the stock solutions with Milli-Q water just prior to injection into the CE-TOFMS. The chemicals used were of analytical or reagent grade.

2.3 Instrument parameters

The instrumentation and measurement conditions used for CE-TOFMS were as described elsewhere (Sugimoto et al. 2010a). All CE-TOFMS experiments were performed using an Agilent CE capillary electrophoresis system (Agilent Technologies, Waldbronn, Germany), an Agilent G3250 AA LS/MSD TOF system (Agilent Technologies, Palo Alto, CA), an Agilent 1100 series binary HPLC pump which delivers sheath liquid, a G1603A Agilent CE–MS adapter kit, and a G1607A Agilent CE–ESI–MS sprayer kit (Agilent Technologies, Waldbronn, Germany). The CE–MS adapter kit used includes a capillary cassette that facilitates thermostatting of the capillary. The CE–ESI–MS sprayer kit simplifies the coupling of the CE system with the MS system, which was equipped with an electrospray ionization source. For system control and data acquisition, G2201AA Agilent ChemStation software for CE and Analyst QS software for TOFMS were used.

2.4 Measurement conditions for CE-TOFMS

Samples were separated in fused silica capillaries (50 μm i.d. × 100 cm total length) filled with 1 M formic acid (pH 1.8) as the background electrolyte. The sample solutions were injected at 50 mbar for 3 s and a voltage of 30 kV was applied. The capillary temperature was maintained at 20°C and the temperature of the sample tray was maintained below 5°C using the external coolant system. The sheath liquid, comprising of methanol/water (50% v/v) and 0.1 μM hexakis-(2,2-difluorothoxy)-phosphazene (Hexakis), was delivered at 10 μl/min. ESI–TOFMS was conducted in positive ion mode. The capillary voltage was set at 4 kV and the nitrogen gas flow rate (heater temperature 300°C) was 10 psig. In TOFMS, the fragmentor, skimmer and OCT RFV voltage were set at 75, 50 and 125 V, respectively. Automatic recalibration of each acquired spectrum was performed using the reference masses of reference standards (^{13}C isotopic ion of protonated methanol dimer $[2MeOH + H]^+$, *m/z* 66.06371, and $[Hexakis + H]^+$, *m/z* 622.02896). Mass spectra were acquired at a rate of 1.5 cycles/s over a 50–1,000 *m/z* range.

2.5 Data processing for CE-TOFMS

Raw data were analyzed with our proprietary software, MasterHands (Sugimoto et al. 2010b). Initially noise-filtering, baseline correction, peak detection and peak area integration from sliced electropherograms (*m/z* 0.02 width) were carried out. Subsequently, the accurate *m/z* value for each peak detected was calculated with Gaussian curve-fitting. A migration time normalization function was established using dynamic programming and the simplex optimization method (Soga et al. 2006). Peaks with small differences in *m/z* value (<20 ppm) and normalized migration time (<1.0 min) were treated as features. Subsequently, neutral compounds, salt ions of Na^+ and K^+, and redundant features such as fragments, adducts, isotopes, dimers and trimers were eliminated on the basis of established *m/z* differences (Brown et al. 2009). For the remaining features, metabolite identities were assigned by matching their *m/z* values and migration times with those of standard compounds. To quantify the detected metabolites, the injected volume for CE and sensitivity of MS were corrected using internal standards, then all annotated metabolites were further corrected with the same chemicals in a standard mixture to overcome different ionization patterns.

2.6 Statistical analysis

The concentration of each metabolite was divided by its average to scale the data, clustered on the basis of Euclidean distance, and visualized as a heat map representation using Mev TM4 software, version 4.6 (Dana-Farber Cancer Institute, Boston, MA) (Saeed et al. 2006). We conducted principal component analysis (PCA) using JMP software (version 8.0.2, SAS, Cary, NC). Metabolite data were mapped on the pathway map using Pathway Projector (Kono et al. 2009), a web-based pathway map application using KEGG Atlas (Kanehisa et al. 2010).

3 Results

3.1 PCA analysis of charged metabolites in herbal medicines and pharmacological effects

After removing peaks determined as non-metabolite, 737 ± 183.1 (average ± SD) peaks remained. Of these, 119 were assigned based on matched *m/z* and normalized migration times with compounds in our standard library. We conducted PCA using these annotated metabolites.

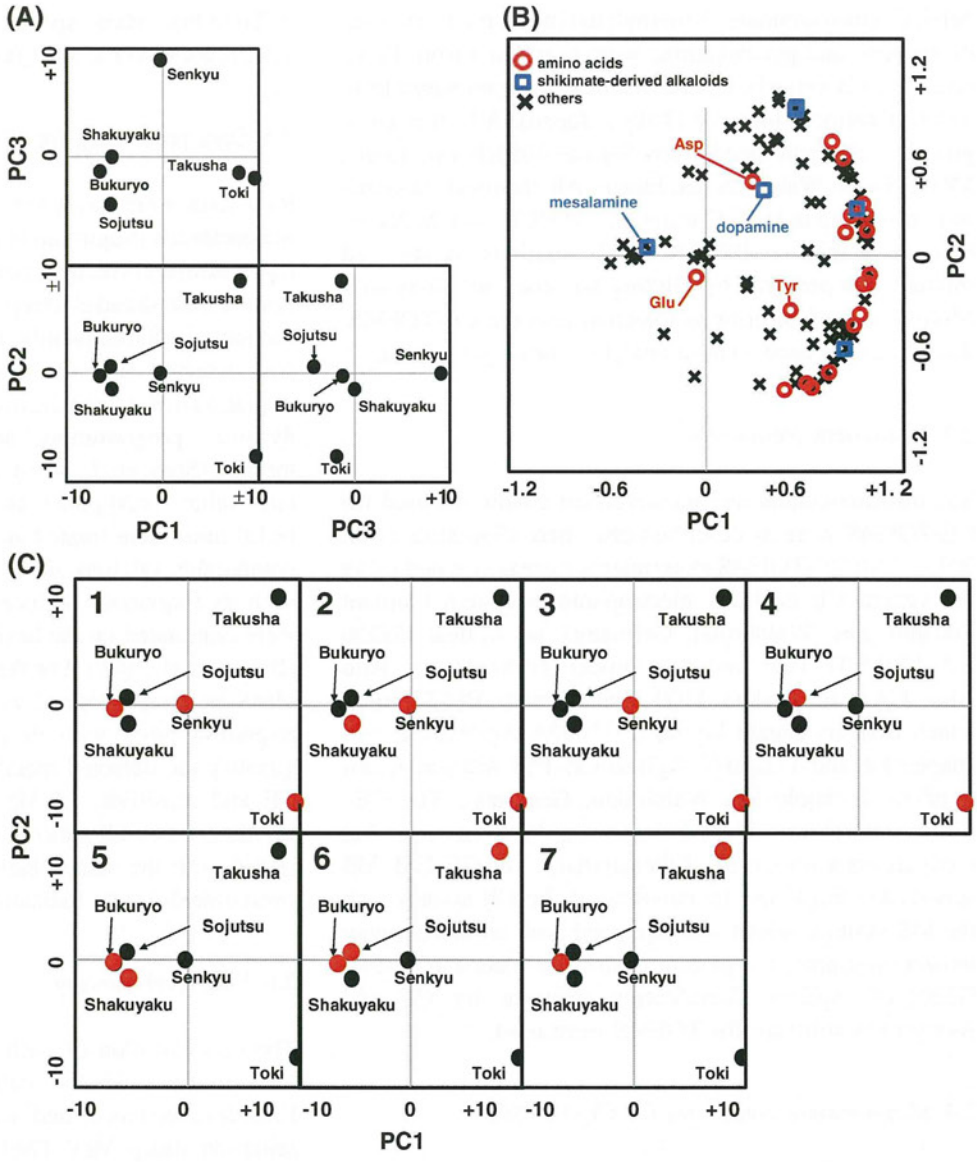

Fig. 1 PCA results of 119 annotated metabolites measured in six herbal medicines. **a** Score plots generated using all annotated metabolites. The cumulative proportions of the first, second and third PC were 43.0%, 65.9% and 83.8%, respectively. **b** Loading plots for the first and second PCs. *Red circles*, *blue squares*, and *black crosses*, represent standard amino acids, shikimate derived metabolites and other annotated metabolites, respectively. **c** Score plots for the first and second PCs. The plots colored *red* have the pharmaceutical effects of (1) tranquilization, (2) painkiller, menstruation problems (dysmenorrhea, oligomenorrhea), (3) removing blood stasis, tonic, and anemia, (4) antidiarrheal effect, (5) anti-convulsion effect (antispasmodic effect), (6) diuretic effect, and (7) oliguria, dizziness, retention of water in the stomach

Score plots (Fig. 1a) showed that, at the first principal component (PC), only Toki and Takusha were separated, showing highly positive scores (+9.5 and +7.9), while the others showed negative scores (<−0.2). However, at the second PC, Toki and Takusha were separated, showing strongly negative (−7.7) and positive scores (+8.6) respectively, while the others congregated around zero. At the third PC, only Senkyu was separated, showing a highly positive score (+8.9). Overall, loading plots (Fig. 1b; Supplementary Information Fig. 1) showed that most of the metabolites had similar loadings and no prominent features were observed. Standard amino acids, with the exception of aspartate, glutamate, and tyrosine (indicated with red circles), and shikimate-derived alkaloids, with the exception of mesalamine and dopamine (blue squares), were located furthest from the origin.

In the obtained profiles, 39 metabolites showed higher loading score values at the first PC (>0.8). Of these, 20 metabolites were amino acids and amino acid derived metabolites; threonine, valine, serine, tyrosine, glycine, phenylalanine, isoleucine, alanine, lysine, asparagine, arginine, citrulline, homoserine, *N*-ε-acetyllysine, *N*-acetylvaline, octopine, ornithine, *N*-acetylornithine, *N*-γ-ethylglutamine and *N*-methylalanine. Along the second PC axis, seven metabolites, 5-aminovalerate, acetylcholine, isoamylamine, *O*-acetyl carnitine, tryptamine, tyramine and anthranilate, showed high loading score values. Along the third PC axis, 2′-deoxyguanosine, 5′-deoxyadenosine, 5-methyl-2′-deoxycytidine, 5-methylthioadenosine, adenosine, ethanolamine phosphate, N_6-methyl-2′-deoxyadenosine and phosphorylcholine showed high loading score values. Of these, four compounds, 5′-deoxyadenosine,

5-methylthioadenosine, adenosine and N_6-methyl-2′-deoxyadenosine, were adenosine derived metabolites. In short, these three PCs characterized amino acids, amines and adenosine-derived metabolites, respectively. The pharmacological effects of the herbal medicines used in this study are summarized in Supplementary Information Table 1. To evaluate the relationships between these efficacies and the observed metabolite profiles of the herbal medicines, the efficacies shared by multiple herbal medicines were mapped on the score plot of PCA results (Fig. 1c). One distinct result found was the shared diuretic effect of Bukuryo and Sojutsu, which were closely located on the score plot, indicating a correlation between the observed metabolite profiles and this effect (Fig. 1c). Both Bukuryo and Sojutsu showed minus values in the first PC on score plots (Fig. 1a). The metabolites located at large minus values at the first PC on loading plots, such as β-alanine, isobutylamine, cytosine, cystathionine, nicotine and nicotinamide, may be characteristic of the PC scores of Bukuryo and Sojutsu (Supplementary Information Fig. 1). Senkyu and Toki are used for purposes including tranquilization, alleviation of pain, menstruation problems, removing blood stasis, as a tonic and for anemia, while they are located at a significant distance in PC score plots (Fig. 1c). Although, the PC values of Senkyu and Toki were not similar, the concentrations of several individual metabolites were clearly higher than in the other herbs; 3-methylhistidine, spermidine, adenosine, pipecolate, tryptophan, thiamine and uridine were high in Senkyu and Toki, while agmatine was high in Takusha, Bukuryo and Sojutsu (Supplementary Information Fig. 2).

3.2 Profiling of charged metabolites in herbal medicines

We conducted cluster analysis to assess the similarity of the charged metabolite profiles among the six herbs (Fig. 2). Those metabolites highlighted in yellow were present in one of the herbs at a concentration twofold higher than the average concentration of all the herbs tested. Metabolite concentrations in most clusters (labeled 1–3, and 6–8) were high only in one herb, with few clusters (4 and 5) showing high levels in multiple herb samples. For example, metabolites in clusters 1, 2, 3 and 6 were present at high concentrations in Bukuryo, Senkyu, Toki and Sojutsu, respectively, and the metabolites in clusters 7 and 8 were present at high concentrations in Takusha. The latter clusters contain three xanthine analogs, hypoxanthine, guanine and xanthosine. The amounts of these three metabolites are shown in Supplementary Information Fig. 3a. Metabolites in cluster 4 were present in significantly higher levels in Toki and Sojutsu than the other herbs tested. Metabolites in cluster 5 were present in higher levels in Toki, Takusha and Senkyu than in Shakuyaku,

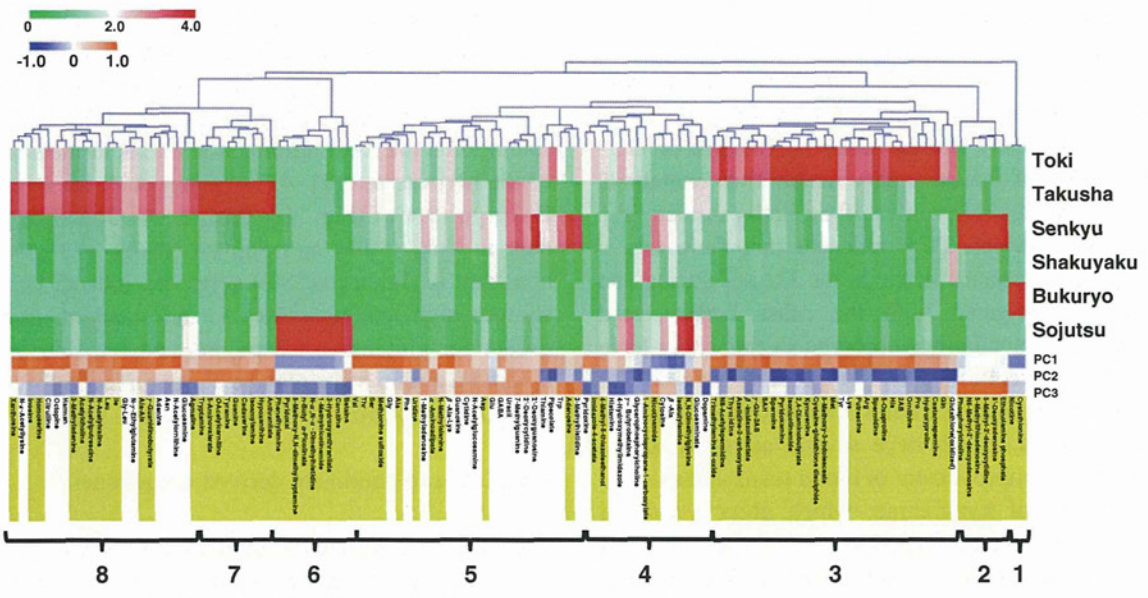

Fig. 2 Heatmap showing metabolite profiles and loading scores. *Green–red* heat map shows the quantified metabolite profiles of the six extracted herbal medicines. Each metabolite concentration shown in the heat map was divided by its average. *Blue–orange* heat map shows the loading score of the first three PCs. Metabolites highlighted in *yellow* were present in concentrations twofold higher in one herbal medicine compared with the average. See the text for numerical labels

Bukuryo and Sojutsu. A blue-orange heat map shows that the metabolites in cluster 3 were present at high loadings only in the first PC, while metabolites in clusters 7 and 8 were present at high loadings in both the first and second PC. Interestingly, metabolites in cluster 2, including five nucleic acids, ethanolaminephosphate and phosphorylcholine, showed high loading values in the third PC.

3.3 Standard amino acid concentration

Figure 3 depicts the quantification of all detected metabolites, standard amino acids, and the sum of the standard amino acids derived from the shikimate pathway. The total metabolite concentration was especially high ($>1.5 \times 10^3$ μM) in Toki, Takusha and Senkyu, while the lowest concentration (1.9×10^2 μM) was observed in Bukuryo (Fig. 3a). The sum of the standard amino acids was dominant (65–85%) except for Bukuryo (35%). The sum of tryptophan, tyrosine and phenylalanine, the shikimate derived amino acids, was notably high ($>5.9 \times 10^2$ μM) in Takusha, Toki and Senkyu (Fig. 3a). Arginine was the most dominant amino acid in Toki, Shakuyaku and Takusha (42%, 39% and 41%, respectively), and proline was present in high levels in Toki (30%) and Sojutsu (38%) (Fig. 3b).

4 Discussion

4.1 Overview of charged metabolite profiles and pharmacological effects

CE-TOFMS-based metabolomics has been used to successfully annotate mainly primary metabolites in herbal medicines, while many studies based on LC–MS or GC–MS give profiles of the secondary metabolites contained. The observed metabolite profiles of six herbs studied here showed large variances. PCA revealed that there was a particularly large variance in the herbs Toki, Takusha and Senkyu when compared with the other herbs tested (Fig. 1a). Several metabolites observed in this study have known pharmaceutical efficacies. Tryptophan is present at higher concentrations in Toki and Senkyu (Supplementary Information Fig. 2a, which have several common activities, including the improvement of blood stasis and anemia. It is a precursor of kynurenine, which plays a role in arterial vessel relaxation (Wang et al. 2010). Agmatine was present at high concentrations in Takusha, Bukuryo and Sojutsu, and is known to have a diuretic effect (Smyth and Penner 1995). Bukuryo, Sojutsu and Shakuyaku showed negative scores at the first PC and were closely plotted on the score plot (Fig. 1a) and β-alanine, isobutylamine, cytosine, cystathionine, nicotine and nicotinamide showed higher negative loading score values at the first PC (Supplementary Information Fig. 1). β-Alanine is the rate-limiting precursor of carnosine, which has an anti-inflammatory effect (Zhu et al. 2007), and single oral doses of isobutylamine have a sedative effect (Cheever et al. 1982). These metabolites therefore correlate with the clinical application of these herbs.

Clustering results, shown using a heat map, indicated that there was little similarity among the herbs tested (Fig. 2). Thus, the concentrations of many metabolites vary widely between the herbs. Izzettin Hatip-Al-Khatib et al. have shown previously that only ornithine in Toki extract is responsible for improving memory impairment (Hatip-Al-Khatib et al. 2004). Our profile revealed high concentrations of ornithine in Toki, which was consistent with their study, and in addition, other amines were also found to be present at high concentrations in Takusha. Such findings were obtained using the non-targeted profiling analysis demonstrated. Here, we discuss prominent features observed for amino acids, shikimate derived metabolites, and xanthine analogs.

4.2 Profile of amino acids

Amino acids have been reported to be the main components of Takusha extracts (Namba and Tsuda 1998), and our results are consistent with these findings; 74% of the total annotated metabolite composition was made up of standard amino acids in Takusha (Fig. 3a). However, particularly high amino acid content was not restricted to Takusha, with amino acids making up 64–85% of the composition of the other herbs tested. The only exception to this trend was Bukuryo, whose extract was composed of only 34% standard amino acids (Fig. 3a). Arginine was found to be the most dominant amino acid measured in all of the herbs tested (21–42%), again, with the exception of Bukuryo (9.4%) (Fig. 3b). Proline was found to be abundant in Toki (30%) and Sojutsu (38%), and alanine was abundant in Bukuryo (38%) (Fig. 3b). Indeed, PCA results showed high loading scores for all standard amino acids except for glutamine, asparagine and tyrosine, indicating that these amino acids may be a prominent factor useful for the characterization of these herbs (Fig. 1b).

4.3 Profile of shikimate derived metabolites

The shikimate-derived metabolites observed in our profiles are known to be bioactive, for example dopamine and kynurenine act as signal molecules in the central and peripheral nervous systems. Metabolites contained in the Toki extract are known to bind to a receptor of the central nervous system (Liao et al. 1995) and a liquid extract of Toki containing alkaloids improved retention memory in

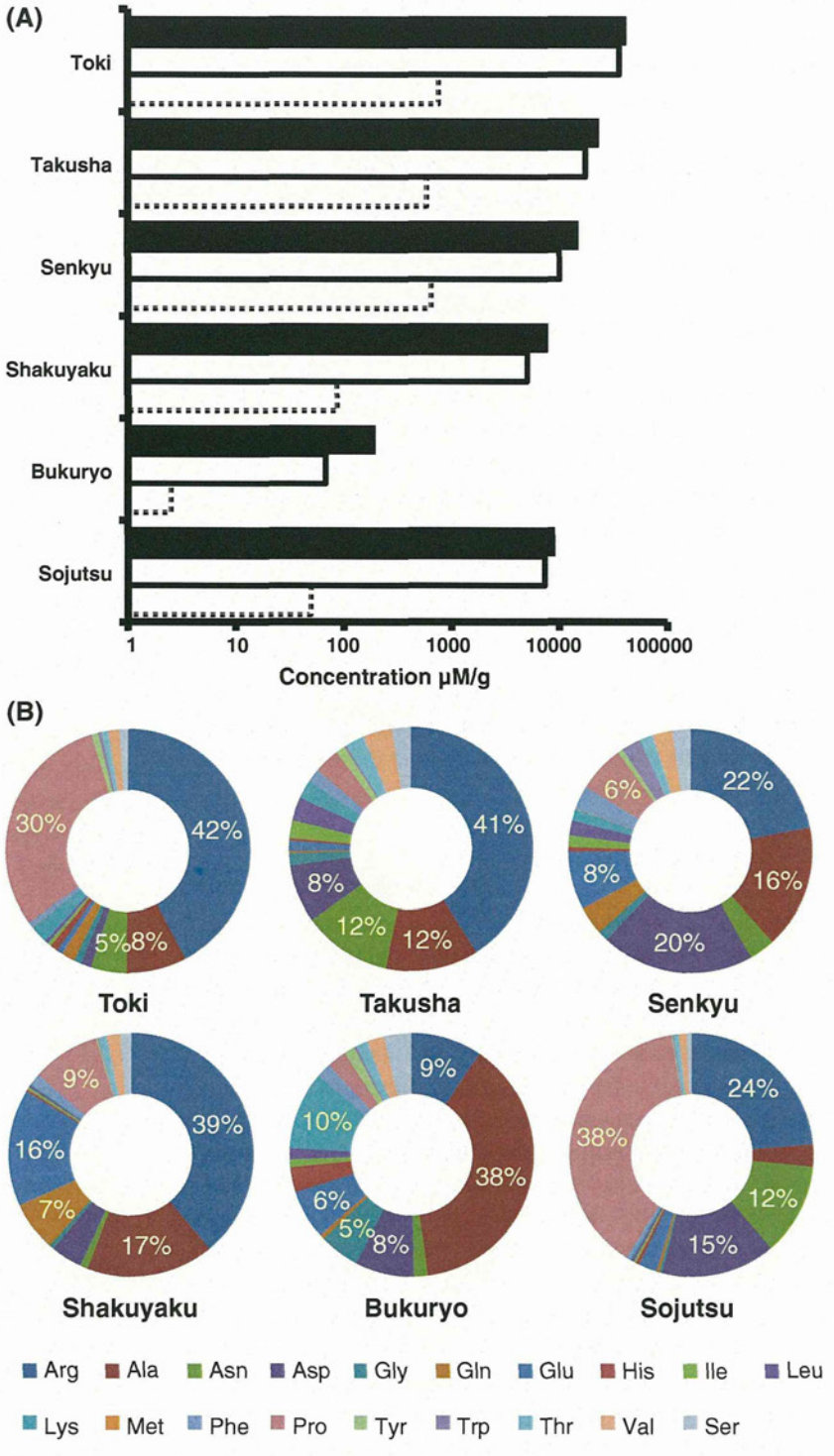

Fig. 3 Comparison of the compositions of the herbal medicines tested. **a** Total concentration of the 119 annotated metabolites (*black bar*), total concentration of standard amino acids (*white-solid bar*), and summed concentration of tryptophan, tyrosine and phenylalanine (*white-dashed bar*) are shown. **b** Standard amino acid composition of each of the six herbal medicines

rats (Hatip-Al-Khatib et al. 2004). Our profiles showed the presence of seven shikimate-derived metabolites and three precursor amino acids (tryptophan, tyrosine and phenylalanine). The total concentration of these three amino acids was higher in Toki, Senkyu and Takusha (Fig. 3a). The concentrations of the shikimate-derived metabolites detected, including kynurenine, harman, tryptamine, tyramine and anthranilate, were also high in Toki, Senkyu and Takusha (Supplementary Information Fig. 3b). These results suggest a correlation between the concentrations of shikimate-derived metabolites and those of the precursor amino acids.

Toki and Senkyu have been reported to contain ferulic acid, which is derived from tyrosine or phenylalanine

(Lu et al. 2005; Yi et al. 2007). Although, ferulic acid was not observed in our profiles, the high concentrations of the precursor amino acids of ferulic acid suggest that ferulic acid may be abundant in Toki and Senkyu. In conclusion, we have found some degree of quantitative interaction between bioactive compounds and their precursor amino acids, though there were some shikimate-derived metabolites, such as dopamine, 3-hydroxyanthranilate, which did not follow this trend.

4.4 Profiles of xanthine analogs

Xanthine analogs are known to have a wide range of biological targets. Examples include the adenosine receptor, where xanthines act as antagonists, and calcium release ryanodine-sensitive channels, where they act to sensitize the channels to calcium activation (Daly 2007). Xanthines are known to have several therapeutic benefits, including analgesic and diuretic effects (Daly 2007). Takusha was found to contain the xanthine analogs guanine (EC 3.5.4.3), hypoxanthine (EC 3.2.2.1) and xanthosine (EC 1.17.3.2 or EC 1.17.1.4) at levels more than ten, four and two times higher than the other herbs tested, respectively. Xanthine analogs may be expected to participate in the analgesic and diuretic effects of Takusha.

4.5 Limitation of this study

In the profiles we obtained, the total concentration of all annotated metabolites was higher in Toki, Takusha, and Senkyu than in the other herbs tested (Fig. 3a). The complex metabolome of these three herbal medicines, including shikimate-derived metabolites and xanthine analogs, may be responsible for the therapeutic functions associated with the herbal medicines, such as analgesia and sedation, since structurally similar compounds are known to have similar functions by receptor binding (Wambach and Casals-Stenzel 1983; Monteith et al. 1996). Further analyses are necessary to compare metabolite extraction methods, with methanol useful to achieve high recovery rates of metabolites (Ren et al. 2008; Sato et al. 2004) and water useful when considering dosage (Liang et al. 2009). More intensive studies on the individual metabolites may give insight into the effect of chemical structure on these therapeutic effects. Further study is needed to elucidate the relationship between the many polar low molecular weight compounds found in these herbal medicines and the herb's therapeutic functions, particularly their interactions with the nervous system. The quantified profiles obtained should also be validated using different samples of these herbal medicines, for example samples grown at different production sites, harvested at different times, and different batches, in order to establish the natural variability of the metabolites. Profiles of negatively charged metabolites and those obtained using other "omics" platforms should also be integrated to allow for more comprehensive analyses.

5 Concluding remarks

In this study, we have conducted a non-targeted analysis of the polar metabolite profiles of the six herbal medicines which compose *Toki-Shakuyaku-San*. Overall, there were few common characteristics observed among the herbs tested, and while the total amino acids content was consistently high among the herbs, individual amino acid content showed large variation. Notably, high concentrations of shikimate-derived metabolites were observed in Toki, Takusha and Senkyu, which helps to explain the herb's biological effects, such as sedation, analgesia and diuretic effects. Multivariable analysis revealed a high correlation between the concentrations of shikimate-derived metabolites and the concentrations of their precursor amino acids, tryptophan, tyrosine and phenylalanine, in the herbs tested. The xanthine analogs guanine, hypoxanthine and xanthosine also showed a high correlation. Analysis of the multiple pharmacological effects of these herbal medicines and comparison with the quantitative compound datasets described here would be valuable in the evaluation of the quality and efficacy of herbal medicines.

Acknowledgments This work was supported by research funds from the Yamagata Prefectural Government and the city of Tsuruoka. We thank Dr. Kazuko Otomo for technical assistance, and Wanjun Kong and Guo Jing for fruitful discussions.

References

Akase, T., Onodera, S., Matsushita, R., & Tashiro, S. (2004). A comparative study of laboratory parameters and symptoms effected by Toki-Shakuyaku-San and an iron preparation in rats with iron-deficiency anemia. *Biological and Pharmaceutical Bulletin, 27*, 871–878.

Bohrmann, H., Stahl, E., & Mitsuhashi, H. (1967). Studies of the constituents of umbelliferae plants. 8. Chromatographic studies on the constituents of *Cnidium officinale* Makino. *Chemical and Pharmaceutical Bulletin, 15*, 1606–1608.

Brown, M., Dunn, W. B., Dobson, P., et al. (2009). Mass spectrometry tools and metabolite-specific databases for molecular identification in metabolomics. *Analyst, 134*, 1322–1332.

Calixto, J. B. (2000). Efficacy, safety, quality control, marketing and regulatory guidelines for herbal medicines (phytotherapeutic agents). *Brazilian Journal of Medical and Biological Research, 33*, 179–189.

Chan, K. (2003). Some aspects of toxic contaminants in herbal medicines. *Chemosphere, 52*, 1361–1371.

Chan, E. C., Yap, S. L., Lau, A. J., Leow, P. C., Toh, D. F., & Koh, H. L. (2007). Ultra-performance liquid chromatography/time-of-flight mass spectrometry based metabolomics of raw and steamed

Panax notoginseng. *Rapid Communications in Mass Spectrometry, 21*, 519–528.

Cheever, K. L., Richards, D. E., & Plotnick, H. B. (1982). The acute oral toxicity of isomeric monobutylamines in the adult male and female rat. *Toxicology and Applied Pharmacology, 63*, 150–152.

Chen, L., Qi, J., Chang, Y. X., Zhu, D., & Yu, B. (2009). Identification and determination of the major constituents in Traditional Chinese Medicinal formula Danggui-Shaoyao-San by HPLC-DAD-ESI-MS/MS. *Journal of Pharmaceutical and Biomedical Analysis, 50*, 127–137.

Daly, J. W. (2007). Caffeine analogs: Biomedical impact. *Cellular and Molecular Life Sciences, 64*, 2153–2169.

Ganzera, M. (2008). Quality control of herbal medicines by capillary electrophoresis: Potential, requirements and applications. *Electrophoresis, 29*, 3489–3503.

Gong, F., Liang, Y. Z., Cui, H., Chau, F. T., & Chan, B. T. (2001). Determination of volatile components in peptic powder by gas chromatography-mass spectrometry and chemometric resolution. *Journal of Chromatography A, 909*, 237–247.

Hatip-Al-Khatib, I., Egashira, N., Mishima, K., et al. (2004). Determination of the effectiveness of components of the herbal medicine Toki-Shakuyaku-San and fractions of *Angelica acutiloba* in improving the scopolamine-induced impairment of rat's spatial cognition in eight-armed radial maze test. *Journal of Pharmacological Sciences, 96*, 33–41.

He, C. N., Peng, Y., Xu, L. J., et al. (2010). Three new oligostilbenes from the seeds of Paeonia suffruticosa. *Chemical and Pharmaceutical Bulletin, 58*, 843–847.

Hurtado-Fernandez, E., Gomez-Romero, M., Carrasco-Pancorbo, A., & Fernandez-Gutierrez, A. (2010). Application and potential of capillary electroseparation methods to determine antioxidant phenolic compounds from plant food material. *Journal of Pharmaceutical and Biomedical Analysis, 53*, 1130–1160.

Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., & Hirakawa, M. (2010). KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Research, 38*, D355–D360.

Kim, M. R., Abd El-Aty, A. M., Choi, J. H., Lee, K. B., & Shim, J. H. (2006). Identification of volatile components in Angelica species using supercritical-CO_2 fluid extraction and solid phase microextraction coupled to gas chromatography-mass spectrometry. *Biomedical Chromatography, 20*, 1267–1273.

Kitajima, J., Kamoshita, A., Ishikawa, T., et al. (2003). Glycosides of Atractylodes lancea. *Chemical and Pharmaceutical Bulletin, 51*, 673–678.

Kono, N., Arakawa, K., Ogawa, R., et al. (2009). Pathway projector: Web-based zoomable pathway browser using KEGG atlas and Google Maps API. *PLoS One, 4*, e7710.

Lao, Y. M., Jiang, J. G., & Yan, L. (2009). Application of metabonomic analytical techniques in the modernization and toxicology research of traditional Chinese medicine. *British Journal of Pharmacology, 157*, 1128–1141.

Li, X. N., Cui, H., Song, Y. Q., Liang, Y. Z., & Chau, F. T. (2003). Analysis of volatile fractions of *Schisandra chinensis* (Turcz.) Baill. using GC–MS and chemometric resolution. *Phytochemical Analysis, 14*, 23–33.

Liang, X. M., Jin, Y., Wang, Y. P., Jin, G. W., Fu, Q., & Xiao, Y. S. (2009). Qualitative and quantitative analysis in quality control of traditional Chinese medicines. *Journal of Chromatography A, 1216*, 2033–2044.

Liao, J. F., Jan, Y. M., Huang, S. Y., Wang, H. H., Yu, L. L., & Chen, C. F. (1995). Evaluation with receptor binding assay on the water extracts of ten CNS-active Chinese herbal drugs. *Proceedings of the National Science Council, Republic of China. Part B, Life Sciences, 19*, 151–158.

Liu, S., Yi, L. Z., & Liang, Y. Z. (2008). Traditional Chinese medicine and separation science. *Journal of Separation Science, 31*, 2113–2137.

Lu, G. H., Chan, K., Liang, Y. Z., et al. (2005). Development of high-performance liquid chromatographic fingerprints for distinguishing Chinese Angelica from related umbelliferae herbs. *Journal of Chromatography A, 1073*, 383–392.

Ma, C. M., Winsor, L., & Daneshtalab, M. (2007). Quantification of spiroether isomers and herniarin of different parts of *Matricaria matricarioides* and flowers of *Chamaemelum nobile*. *Phytochemical Analysis, 18*, 42–49.

Miller, F. G., Emanuel, E. J., Rosenstein, D. L., & Straus, S. E. (2004). Ethical issues concerning research in complementary and alternative medicine. *JAMA, 291*, 599–604.

Monteith, D. K., Emmerling, M. R., Garvin, J., & Theiss, J. C. (1996). Cytotoxicity study of tacrine, structurally and pharmacologically related compounds using rat hepatocytes. *Drug and Chemical Toxicology, 19*, 71–84.

Monton, M. R., & Soga, T. (2007). Metabolome analysis by capillary electrophoresis–mass spectrometry. *Journal of Chromatography A, 1168*, 237–246. discussion 236.

Murray, R. H., & Rubel, A. J. (1992). Physicians and healers—Unwitting partners in health care. *New England Journal of Medicine, 326*, 61–64.

Namba, T., & Tsuda, Y. (1998). *Shoyakugakugairon* (3rd ed.). Tokyo: Nankodo.

Ohta, H., Ni, J. W., Matsumoto, K., Watanabe, H., & Shimizu, M. (1993). Peony and its major constituent, paeoniflorin, improve radial maze performance impaired by scopolamine in rats. *Pharmacology, Biochemistry and Behavior, 45*, 719–723.

Ren, M. T., Chen, J., Song, Y., Sheng, L. S., Li, P., & Qi, L. W. (2008). Identification and quantification of 32 bioactive compounds in *Lonicera* species by high performance liquid chromatography coupled with time-of-flight mass spectrometry. *Journal of Pharmaceutical and Biomedical Analysis, 48*, 1351–1360.

Saeed, A. I., Bhagabati, N. K., Braisted, J. C., et al. (2006). TM4 microarray software suite. *Methods in Enzymology, 411*, 134–193.

Sato, S., Soga, T., Nishioka, T., & Tomita, M. (2004). Simultaneous determination of the main metabolites in rice leaves using capillary electrophoresis mass spectrometry and capillary electrophoresis diode array detection. *The Plant Journal, 40*, 151–163.

Smyth, D. D., & Penner, S. B. (1995). Renal I1-imidazoline receptor-selective compounds mediate natriuresis in the rat. *Journal of Cardiovascular Pharmacology, 26*(Suppl. 2), S63–S67.

Soga, T., Baran, R., Suematsu, M., et al. (2006). Differential metabolomics reveals ophthalmic acid as an oxidative stress biomarker indicating hepatic glutathione consumption. *The Journal of Biological Chemistry, 281*, 16768–16776.

Soga, T., Ohashi, Y., Ueno, Y., Naraoka, H., Tomita, M., & Nishioka, T. (2003). Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. *Journal of Proteome Research, 2*, 488–494.

Sugimoto, M., Koseki, T., Hirayama, A., et al. (2010a). Correlation between sensory evaluation scores of Japanese sake and metabolome profiles. *Journal of Agriculture and Food Chemistry, 58*, 374–383.

Sugimoto, M., Wong, D. T., Hirayama, A., Soga, T., & Tomita, M. (2010b). Capillary electrophoresis mass spectrometry-based saliva metabolomics identified oral, breast and pancreatic cancer-specific profiles. *Metabolomics, 6*, 78–95.

Wambach, G., & Casals-Stenzel, J. (1983). Structure-activity relationship of new steroidal aldosterone antagonists. Comparison of the affinity for mineralocorticoid receptors in vitro and the

antialdosterone activity in vivo. *Biochemical Pharmacology, 32*, 1479–1485.

Wang, H. X., Liu, C. M., Liu, Q., & Gao, K. (2008). Three types of sesquiterpenes from rhizomes of Atractylodes lancea. *Phytochemistry, 69*, 2088–2094.

Wang, Y., Liu, H., Mckenzie, G., et al. (2010). Kynurenine is an endothelium-derived relaxing factor produced during inflammation. *Nature Medicine, 16*, 279–285.

Wang, Z. G., & Ren, J. (2002). Current status and future direction of Chinese herbal medicine. *Trends in Pharmacological Sciences, 23*, 347–348.

Wang, J., Van Der Heijden, R., Spruit, S., et al. (2009). Quality and safety of Chinese herbal medicines guided by a systems biology perspective. *Journal of Ethnopharmacology, 126*, 31–41.

Wang, Y., Zhang, M., Ruan, D., et al. (2004). Chemical components and molecular mass of six polysaccharides isolated from the sclerotium of Poria cocos. *Carbohydrate Research, 339*, 327–334.

Weng, Q., & Jin, W. (2002). Carbon fiber bundle–Au–Hg dual-electrode detection for capillary electrophoresis. *Journal of Chromatography A, 971*, 217–223.

Xie, G., Plumb, R., Su, M., et al. (2008). Ultra-performance LC/TOF MS analysis of medicinal Panax herbs for metabolomic research. *Journal of Separation Science, 31*, 1015–1026.

Yi, T., Leung, K. S., Lu, G. H., & Zhang, H. (2007). Comparative analysis of *Ligusticum chuanxiong* and related umbelliferous medicinal plants by high performance liquid chromatography–electrospray ionization mass spectrometry. *Planta Medica, 73*, 392–398.

Zhang, A., Sun, H., Wang, Z., Sun, W., Wang, P., & Wang, X. (2010). Metabolomics: Towards understanding traditional Chinese medicine. *Planta Medica, 76*, 2026–2035.

Zhao, M., Xu, L. J., & Che, C. T. (2008). Alisolide, alisols O and P from the rhizome of *Alisma orientale*. *Phytochemistry, 69*, 527–532.

Zhu, Y. Y., Zhu-Ge, Z. B., Wu, D. C., et al. (2007). Carnosine inhibits pentylenetetrazol-induced seizures by histaminergic mechanisms in histidine decarboxylase knock-out mice. *Neuroscience Letters, 416*, 211–216.

Research Article

EASL EUROPEAN ASSOCIATION FOR THE STUDY OF THE LIVER | JOURNAL OF HEPATOLOGY

Serum metabolomics reveals γ-glutamyl dipeptides as biomarkers for discrimination among different forms of liver disease

Tomoyoshi Soga[1,*], Masahiro Sugimoto[1], Masashi Honma[2], Masayo Mori[1], Kaori Igarashi[1], Kasumi Kashikura[1], Satsuki Ikeda[1], Akiyoshi Hirayama[1], Takehito Yamamoto[2], Haruhiko Yoshida[3], Motoyuki Otsuka[3], Shoji Tsuji[4], Yutaka Yatomi[4], Tadayuki Sakuragawa[5], Hisayoshi Watanabe[6], Kouei Nihei[7], Takafumi Saito[6], Sumio Kawata[6], Hiroshi Suzuki[2], Masaru Tomita[1], Makoto Suematsu[5]

[1]*Institute for Advanced Biosciences, Keio University, Tsuruoka 997-0052, Japan;* [2]*Department of Pharmacy, The University of Tokyo Hospital, Hongo, Bunkyo-ku, Tokyo 113-8655, Japan;* [3]*Department of Gastroenterology, Faculty of Medicine, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-8655, Japan;* [4]*Department of Neurology, Division of Neuroscience, Graduate School of Medicine, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-8655, Japan;* [5]*Department of Biochemistry, JST ERATO Suematsu Gas Biology Project, School of Medicine, Keio University, Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan;* [6]*Department of Gastroenterology, Yamagata University School of Medicine, Yamagata 990-9585, Japan;* [7]*Department of Surgery, Shonai Hospital, 4-20 Izumi-cho, Tsuruoka 997-8515, Japan*

Background & Aims: We applied a metabolome profiling approach to serum samples obtained from patients with different liver diseases, to discover noninvasive and reliable biomarkers for rapid-screening diagnosis of liver diseases.

Methods: Using capillary electrophoresis and liquid chromatography mass spectrometry, we analyzed low molecular weight metabolites in a total of 248 serum samples obtained from patients with nine types of liver disease and healthy controls.

Results: We found that γ-glutamyl dipeptides, which were biosynthesized through a reaction with γ-glutamylcysteine synthetase, were indicative of the production of reduced glutathione, and that measurement of their levels could distinguish among different liver diseases. Multiple logistic regression models facilitated the discrimination between specific and other liver diseases and yielded high areas under receiver-operating characteristic curves. The area under the curve values in training and independent validation data were 0.952 and 0.967 in healthy controls, 0.817 and 0.849 in drug-induced liver injury, 0.754 and 0.763 in asymptomatic hepatitis B virus infection, 0.820 and 0.762 in chronic hepatitis B, 0.972 and 0.895 in hepatitis C with persistently normal alanine transaminase, 0.917 and 0.707 in chronic hepatitis C, 0.803 and 0.993 in cirrhosis type C, and 0.762 and 0.803 in hepatocellular carcinoma, respectively. Several γ-glutamyl dipeptides also manifested potential for differentiating between nonalcoholic steatohepatitis and simple steatosis.

Conclusions: γ-Glutamyl dipeptides are novel biomarkers for liver diseases, and varying levels of individual or groups of these peptides have the power to discriminate among different forms of hepatic disease.

Keywords: γ-Glutamyl dipeptides; Metabolomics; Biomarker; Capillary electrophoresis mass spectrometry; Oxidative stress; Glutathione; Hepatocellular carcinoma; Nonalcoholic steatohepatitis; Hepatitis C virus.

Received 15 July 2010; received in revised form 17 January 2011; accepted 24 January 2011; available online 18 February 2011

* Corresponding author. Address: Institute for Advanced Biosciences, Keio University, 246-2 Mizukami, Kakuganji, Tsuruoka, Yamagata 997-0052, Japan. Tel.: +81 235 29 0528; fax: +81 0235 29 0574.

E-mail address: soga@sfc.keio.ac.jp (T. Soga).

Abbreviations: HCC, hepatocellular carcinoma; AST, aspartate transaminase; ALT, alanine transaminase; γ-GTP, γ-glutamyl transpeptidase; CT, computed tomography; NAFLD, nonalcoholic fatty liver disease; SS, simple steatosis; NASH, nonalcoholic steatohepatitis; CE-TOFMS, capillary electrophoresis time-of-flight mass spectrometry; GSH, reduced glutathione; GC, gastric cancer; GCS, γ-glutamylcysteine synthetase; C, healthy control; DI, drug-induced liver injury; AHB, asymptomatic hepatitis B virus infection; CHB, chronic hepatitis B; CNALT, hepatitis C with persistently normal alanine transaminase; CHC, chronic hepatitis C; CIR, cirrhosis type C; HBs, hepatitis B surface; HBV, hepatitis B virus; HCV, hepatitis C virus; AFP, α-fetoprotein; PIVKA, protein induced by vitamin K antagonist; LC–MS/MS, liquid chromatography–electrospray tandem mass spectrometry; MLR, multiple logistic regression; APAP, acetaminophen; GS, glutathione synthetase; BSO, buthionine sulfoximine; DEM, diethylmaleate; ROS, reactive oxygen species.

Introduction

Acute or chronic viral hepatitis affects populations around the world, and the disease often progresses from chronic hepatitis and cirrhosis to hepatocellular carcinoma (HCC) [1]. Accurate diagnosis at earlier stages is necessary for improved therapeutic outcome. However, the diagnostic procedures are laborious and not risk-free. Patients with suspected liver damage are initially subjected to liver function tests that include the assessment of aspartate transaminase (AST), alanine transaminase (ALT), and γ-glutamyl transpeptidase (γ-GTP) serum levels. If these levels are abnormal, patients are then subjected to diagnostic imaging, such as ultrasound and computed tomography (CT), and assays to determine the presence of antibodies against hepatitis virus. Finally, a liver biopsy may be recommended to evaluate the severity of inflammation or fibrosis and to confirm the indications for antiviral therapy.

Recently, nonalcoholic fatty liver disease (NAFLD) has become the most common liver disease in western countries. It

encompasses a wide spectrum of conditions associated with over-accumulation of fat in the liver, ranging from simple steatosis (SS) to nonalcoholic steatohepatitis (NASH), and cirrhosis [2]. Although SS typically follows a benign non-progressive clinical course, NASH may eventually develop into cirrhosis and HCC. To date, a liver biopsy remains the gold standard for the diagnosis of NASH [3]. However, since the biopsy procedures carry the risk of mortality [4,5], the noninvasive identification of biomarkers, that can provide reliable differential diagnoses for the characterization of liver diseases, is desirable.

Metabolomics, which can be defined as measurement of the levels of all cellular metabolites, has emerged as a powerful new tool for discovering new low molecular weight biomarkers. Its utility has been demonstrated by the identification of new biomarkers for prostate cancer [6], Parkinson's disease [7], type 2 diabetes mellitus [8], acute myocardial ischemia [9], and pre-eclampsia [10].

Recently, we developed new metabolomic profiling approaches based on capillary electrophoresis mass spectrometry [11] and capillary electrophoresis–time-of-flight mass spectrometry (CE–TOFMS) [12–14]. The efficacy of CE–TOFMS was demonstrated by the discovery of ophthalmate (γ-glutamyl-2-aminobutyrylglycine) as a biomarker; in mice, reduced glutathione (GSH) depletion produced acetaminophen-induced hepatotoxicity [12,14]. In this study, to discover new noninvasive biomarkers for human liver diseases, we comprehensively analyzed the serum metabolites in a total of 248 samples from patients with nine types of liver disease or gastric cancer (GC) and from normal individuals using our metabolomic approaches, and found increased levels of γ-glutamyl dipeptides in the majority of the liver diseases. Moreover, we found that γ-glutamyl dipeptides were synthesized via the ligation of glutamate with various amino acids and amines by the γ-glutamylcysteine synthetase (GCS), an enzyme that is feedback-inhibited by GSH, and that the levels of γ-glutamyl dipeptides were indicative of the amount of GSH production. The concentrations of serum γ-glutamyl dipeptides varied with the stage and type of liver disease and can, therefore, act as new biomarkers for liver diseases. Here, we report that a highly specific set of γ-glutamyl dipeptides, alone or in combination with transaminases and methionine sulfoxide, can effectively distinguish specific liver diseases from other hepatic injuries and healthy control samples.

Materials and methods

Serum samples

A total of 248 serum samples were obtained from three institutes, Yamagata University Hospital (YUH; Yamagata, Japan), University of Tokyo Hospital (UTH; Tokyo, Japan) and Shonai Hospital (SH; Tsuruoka, Japan). The 162 YUH cases comprised 53 healthy controls (C) and patients with drug-induced liver injury (DI; $n = 10$), asymptomatic hepatitis B virus infection (AHB; $n = 9$), chronic hepatitis B (CHB; $n = 7$), hepatitis C with persistently normal alanine transaminase (CNALT; $n = 10$), chronic hepatitis C (CHC; $n = 24$), cirrhosis type C (CIR; $n = 10$), HCC ($n = 19$), SS ($n = 9$) and NASH ($n = 11$). The 75 UTH cases comprised four controls and patients with DI ($n = 17$), AHB ($n = 7$), CHB ($n = 7$), CNALT ($n = 8$), CHC ($n = 11$), CIR ($n = 8$) and HCC ($n = 13$). The 11 SH cases were all GC patients. Written informed consent was obtained from all the participants and the study protocol conformed to the ethical guidelines of the 1975 Declaration of Helsinki as reflected in a priori approval by the appropriate institutional review boards of YUH, UTH, and SH. The study subjects were patients with viral liver diseases, drug-induced hepatotoxicity or NAFLD who were referred to the Department of Gastroenterology and Hepatology at YUH, UTH, or SH.

Clinical diagnosis

All the healthy controls had normal liver function and no viral hepatitis infection, and none were alcoholics. The AHB and CNALT patients were confirmed to have normal liver function and to be positive for hepatitis B surface (HBs) antigen and hepatitis B virus (HBV) DNA, or for anti-hepatitis C virus (HCV) antibodies and HCV RNA, respectively. DI was diagnosed based on abnormal values on biochemical tests, absence of other hepatic diseases, and a history of treatment with drugs suspected of being probable causes of DI. The suspected medications were different, and the biochemical test results in each patient normalized after their withdrawal.

CHC and CIR were diagnosed on the basis of physical examination, biochemical tests, ultrasonography, and CT findings. Some patients with chronic hepatitis provided informed consent for a liver biopsy, and the procedure was performed to confirm the accuracy of the diagnosis. The diagnosis of CHB and CHC was based on increased ALT levels (above the upper limit of the normal range) in at least two blood samples assayed over a 6-month period, and the presence of detectable HBs antigen and HBV DNA or detectable anti-HCV antibodies and HCV RNA, respectively. HCV infection was causative in all cirrhosis patients, and they manifested symptoms of portal hypertension, such as splenomegaly, esophageal varices, encephalopathy, or ascites.

The diagnosis of HCC was based on ultrasonography, CT, and MRI findings that revealed features typical of HCC. HCV was causative in all cases, and the α-fetoprotein (AFP) and protein induced by vitamin K antagonist (PIVKA)-II levels were assayed in all HCC patients.

All of the SS and NASH patients underwent liver biopsy. The tissue samples were stained with hematoxylin–eosin, reticulin, and Masson trichrome; and examined by the same experienced pathologist who was blinded to the clinical data. The histological criterion for the diagnosis of NAFLD was the presence of fatty changes in hepatocytes. When hepatocytes exhibited macrovesicular steatosis, the differential diagnosis was SS or NASH. The criteria for a diagnosis of steatohepatitis were the presence of lobular inflammation and either ballooning cells or perisinusoidal/pericellular fibrosis, in addition to steatosis in the liver specimen. No patient with autoimmune hepatitis, primary biliary cirrhosis, sclerosing cholangitis, hemochromatosis, α1-antitrypsin deficiency, Wilson's disease, or alcoholic liver injury was included. All patients with GC were diagnosed by pathologic studies of biopsy tissues.

Analytical and statistical technologies for biomarker discovery

Using a total of 237 samples from YUH (training cohort, $n = 162$) and UTH (validation cohort, $n = 75$) (Table 1), we performed CE–TOFMS for a comprehensive analysis of the metabolite changes to discover new biomarkers in the diagnosis of human liver diseases. To facilitate peak identification and quantification, we analyzed 162 metabolic standards listed in the KEGG LIGAND database [15] before analyzing the samples. Global mass scanning over a 50–1000 m/z range was applied in the CE–TOFMS mode [12]. To focus on γ-glutamyl peptides, we employed a highly sensitive method using liquid chromatography electrospray tandem mass spectrometry (LC–MS/MS) with multiple reactions monitoring for analyses of the patient serum samples. The Kruskal–Wallis test and Dunn's post-test were used to assess the statistical significance of differences among C, DI, AHB, CHB, CNALT, CHC, CIR, and HCC. The Mann–Whitney test was used to evaluate the statistical significance of differences between SS and NASH. The algorithm of the feature selection for the multiple logistic regression (MLR) models is described in the Supplementary data.

Results

Discovery of γ-glutamyl dipeptides in serum by metabolomic profiling

The CE–TOFMS analysis quantified the levels of 49 metabolites in the serum samples (Supplementary Tables 1 and 2) and revealed increases in many compounds in most liver diseases. We identified these compounds as γ-glutamyl dipeptides (e.g., γ-Glu-Gly, γ-Glu-Ala, γ-Glu-Ser, γ-Glu-Val, γ-Glu-Thr, γ-Glu-Taurine, γ-Glu-Leu, γ-Glu-Gln, γ-Glu-Lys, γ-Glu-Glu, γ-Glu-Met, γ-Glu-His, γ-Glu-Phe, γ-Glu-Arg, γ-Glu-Citrulline, γ-Glu-Tyr, and γ-Glu-Trp) by comparing their migration times and exact molecular

Table 1. Summary of patient information.

Clinical information		Training cohort (n = 162)	Defect no.	Validation cohort (n = 75)	Defect no.	*p* value
Age (years)						
	Median	61	0	66	0	0.47
	Interquartile range	51-73	0	55-70	0	
Sex (n)	Male	73	0	52	0	0.0007*
	Female	89	0	23	0	
AST (UL^{-1})						
	C	21.5 ± 5.40	0	25.3 ± 3.60	0	0.074
	DI	274 ± 567	0	81.2 ± 84.9	0	0.15
	AHB	25.0 ± 6.81	2	23.9 ± 6.90	0	0.71
	CHB	109 ± 164	0	150 ± 146	0	0.0059
	CNALT	24.1 ± 3.80	0	23.8 ± 6.00	0	0.72
	CHC	62.8 ± 65.3	0	110 ± 51.0	0	0.0010
	CIR	54.6 ± 27.1	0	58.0 ± 26.1	0	0.69
	HCC	71.3 ± 52.8	0	35.0 ± 24.5	0	0.0010
	SS	41.2 ± 11.5	0			
	NASH	78.6 ± 48.0	0			
ALT (UL^{-1})						
	C	17.7 ± 4.70	0	25.0 ± 8.30	0	0.062
	DI	253 ± 343	0	115 ± 132	0	0.15
	AHB	26.6 ± 18.6	2	23.1 ± 5.60	0	0.40
	CHB	117 ± 162	0	173 ± 131	0	0.0060
	CNALT	17.9 ± 4.10	0	21.5 ± 3.60	0	0.074
	CHC	79.4 ± 81.0	0	160 ± 116	0	0.0036
	CIR	40.7 ± 21.9	0	57.3 ± 42.4	0	0.69
	HCC	57.9 ± 58.8	0	25.0 ± 21.6	0	0.0026
	SS	72.2 ± 24.5	0			
	NASH	121 ± 140	0			
γ-GTP						
	C	20.7 ± 8.60	0	—	4	—
	DI	190 ± 236	0	46.2 ± 29.5	5	0.010
	AHB	31.1 ± 24.1	2	—	7	—
	CHB	52.8 ± 38.1	1	—	7	—
	CNALT	150 ± 5.70	0	—	8	—
	CHC	48.5 ± 36.4	0	—	11	—
	CIR	28.8 ± 17.9	0	49.6 ± 53.1	0	0.17
	HCC	51.2 ± 31.1	0	—	13	—
	SS	61.8 ± 43.7	0			
	NASH	98.7 ± 99.1	0			
AFP						
	CHC	6.40 ± 7.40	3	—	11	—
	CIR	35.1 ± 71.8	0	14 ± 15.6	0	0.63
	HCC	9.79×10^2 ± 1.73×10^3	0	7.04×10^3 ± 2.52×10^4	0	0.024
PIVKA-II						
	HCC	1.57×10^2 ± 1.87×10^2	0	7.78×10^3 ± 2.77×10^4	0	0.022

*Chi-square test. The others *p* values were obtained by the Mann–Whitney *U*-test.

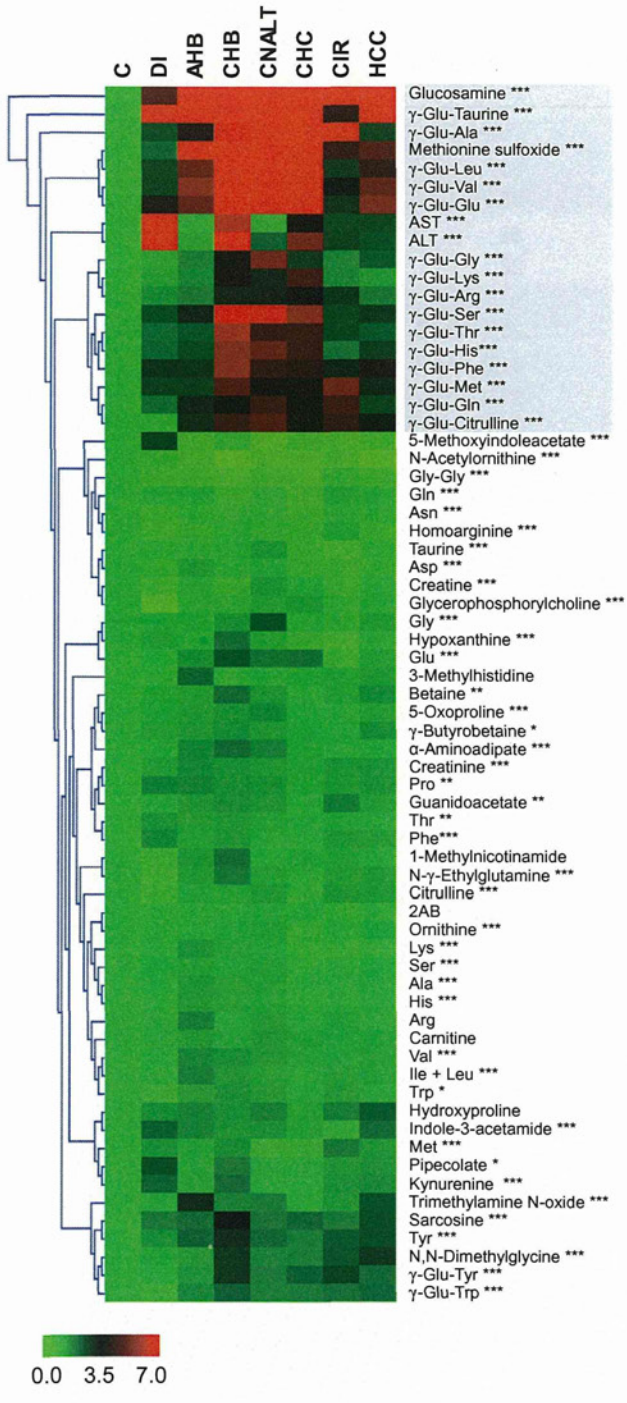

Fig. 1. Heat map representing the hierarchical clustering of 67 compounds in serum samples from controls and patients with various types of liver disease in both cohorts. Each row shows data for a specific metabolite or transaminase, and each column shows data for the healthy controls and patients with liver diseases. The compound concentration in each individual was divided by the average concentration in the healthy controls and the obtained values were then averaged again for each disease. The metabolites highlighted in blue showed large fold changes (disease/control ratios of >2.5) in an average of seven liver diseases. *p <0.05, **p <0.01, ***p <0.0001, significance difference by the Kruskal–Wallis test. The compounds were clustered based on elucidation distances. Red and green denote relatively high and low concentrations, respectively, compared with the average concentration.

weights with those of the standards. Significant differences were observed among controls and liver diseases (p <0.0001; Kruskal–Wallis test) except for γ-Glu-Met in the validation data (Supplementary Tables 1 and 2). Correlational cluster analyses of 67 compounds showed that all the γ-glutamyl dipeptides except for γ-Glu-Tyr and γ-Glu-Trp were clustered with AST, ALT, and metabolites involved in oxidative stress responses, namely glucosamine [16] and methionine sulfoxide [17–19] (Fig. 1).

Statistical analysis and validation for biomarker discovery

From the serum samples obtained at YUH, we selected 89 liver disease patients including DI, AHB, CHB, CNALT, CHC, CIR, and HCC patients, and 53 healthy controls with no significant differences in the age distribution between the training and validation cohorts (Table 1). As shown in the whisker box plots for the training cohort (Fig. 2), the levels of γ-glutamyl dipeptides and of AST and ALT, as commonly used hepatocyte biomarkers, were increased in different patterns in comparison with C. For example, the AST and ALT levels were significantly increased in patients with DI, CHB, CHC, CIR, and HCC (p <0.05; Dunn's post-test), but not in those with AHB and CNALT (Fig. 2). On the other hand, significant increases were observed in the levels of γ-Glu-Ser, γ-Glu-Val, γ-Glu-Thr, γ-Glu-Leu, and γ-Glu-Phe (p <0.05; Dunn's post-test) in AHB and in the levels of all the γ-glutamyl derivatives of amino acids (p <0.05; Dunn's post-test) except for ophthalmate, γ-Glu-Thr, and γ-Glu-Trp in CNALT (Fig. 2 and Supplementary Table 1). Oxidative metabolites, methionine sulfoxide, and glucosamine were significantly increased in all diseases (p <0.05; Dunn's post-test) and in CHB, CNALT, and CHC (p <0.0001; Dunn's post-test), respectively (Fig. 2).

To assess their abilities to discriminate specific liver diseases from other liver diseases, we developed MLR models using combinations of several components of the γ-glutamyl dipeptides, transaminases, and oxidative metabolites using the training dataset. For example, an MLR model incorporating four selected biomarkers (γ-Glu-Ala, γ-Glu-Citrulline, γ-Glu-Thr, and γ-Glu-Phe) was able to differentiate HCC from the other groups (C, DI, AHB, CHB, CNALT, CHC, and CIR) with an area under the receiver-operating characteristic (ROC) curve (AUC) value of 0.762 (95% CI 0.647–0.877, $p = 0.00025$). The probability (p) of HCC is calculated by: $\log(p/(1-p)) = -1.87 - 1.13 \times$ γ-Glu-Ala $+ 3.51 \times$ γ-Glu-Citrulline $- 1.65 \times$ γ-Glu-Thr $+ 6.99 \times$ γ-Glu-Phe (Table 2). When the concentrations of γ-Glu-Ala, γ-Glu-Citrulline, γ-Glu-Thr, and γ-Glu-Phe are 1.7, 0.84, 0.54, and 0.34 μM, respectively, the probability of HCC is 65.5%. All the MLR models achieved high AUC values at statistically significant levels (between 0.754 and 0.972, p <0.011) (Fig. 3, Table 2 and Supplementary Table 3).

The developed MLR models were evaluated in a blinded manner using an independent cohort (YUH) consisting of 75 individuals who were not members of the training cohort (Supplementary Table 2). We found that all of the MLR models also produced high AUC values at statistically significant levels (between 0.707 and 0.993, p <0.023) (Fig. 3, Table 2 and Supplementary Table 3). Although C, CHB, and CHC were each differentiated from the other groups by a single γ-glutamyl dipeptide (γ-Glu-Phe, γ-Glu-Thr, and γ-Glu-Lys, respectively), the MLR models for the other diseases required multiple biomarkers to achieve accurate discrimination (Table 2). The odds ratios of ALT, AST, and methionine sulfoxide were close to 1.0 compared with the odds ratios of the γ-glutamyl dipeptides, indicating their

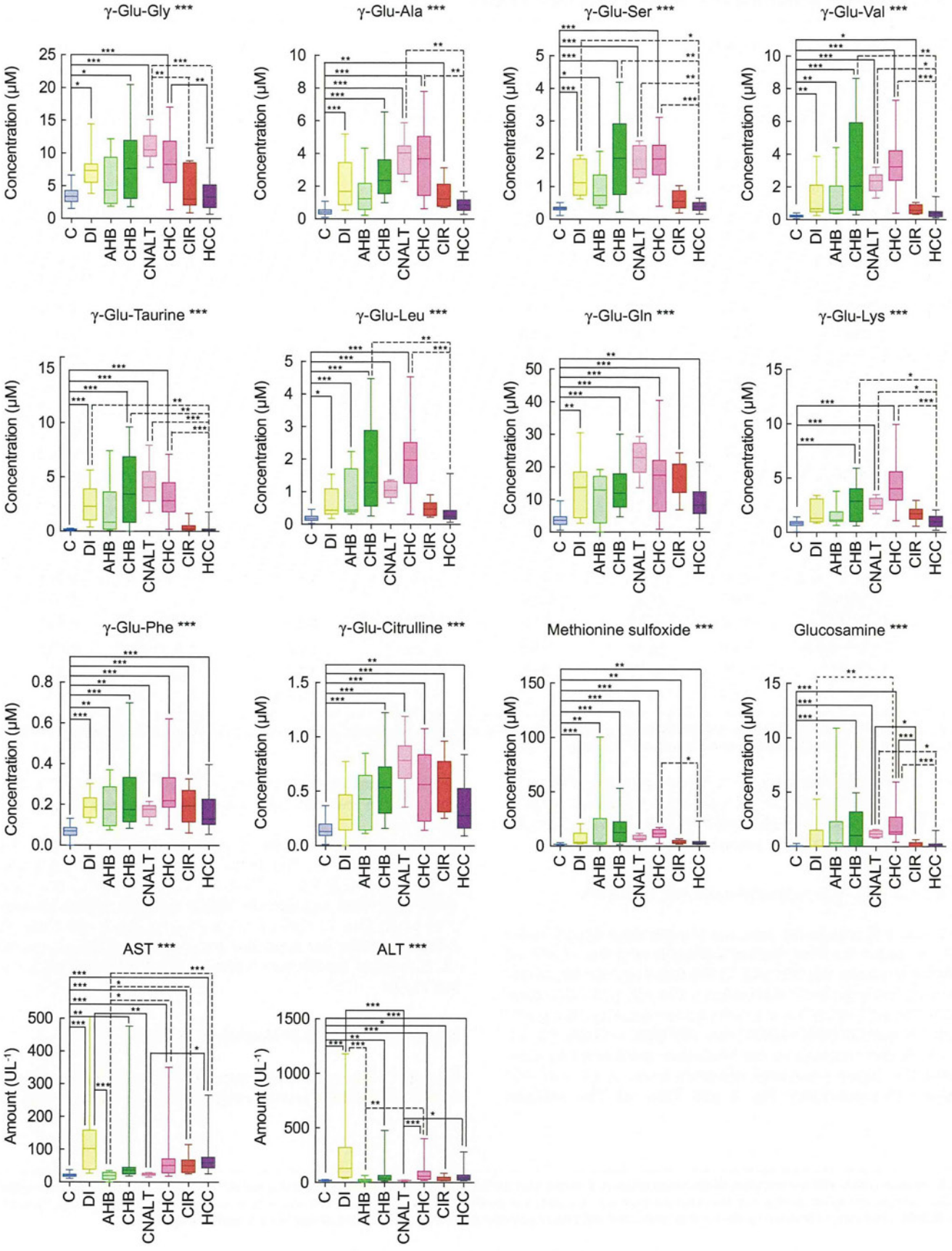
γ-Glu-Gly ***
γ-Glu-Ala ***
γ-Glu-Ser ***
γ-Glu-Val ***
γ-Glu-Taurine ***
γ-Glu-Leu ***
γ-Glu-Gln ***
γ-Glu-Lys ***
γ-Glu-Phe ***
γ-Glu-Citrulline ***
Methionine sulfoxide ***
Glucosamine ***
AST ***
ALT ***
Concentration (μM)
Amount (UL-1)
C
DI
AHB
CHB
CNALT
CHC
CIR
HCC

Table 2. Biomarkers for discriminating each liver disease selected by MLR models.

Group	Biomarker	Coefficient	95% CI		Odds ratio	95% CI		*p* value
C	(Intercept)	5.77	3.84	8.32	—	—	—	<0.0001
	γ-Glu-Phe	-58.2	-84.3	-39.0	5.16×10^{-26}	2.47×10^{-37}	1.15×10^{-17}	<0.0001
DI	(Intercept)	-3.08	-4.49	-1.94	—	—	—	<0.0001
	ALT	0.020	7.89×10^{-3}	0.034	1.02	1.01	1.03	2.00×10^{-3}
	γ-Glu-Citrulline	-1.55	-5.01	1.13	0.21	6.68×10^{-3}	3.11	0.31
AHB	(Intercept)	-1.52	-3.35	0.63	—	—	—	0.12
	AST	-0.057	-0.15	-4.96×10^{-3}	0.94	0.86	1.00	0.12
	Methionine sulfoxide	0.072	0.018	0.15	1.08	1.02	1.17	0.047
CHB	(Intercept)	-4.52	-6.33	-3.24	—	—	—	<0.0001
	γ-Glu-Thr	1.52	0.65	2.63	4.58	1.91	13.9	2.30×10^{-3}
CNALT	(Intercept)	-0.76	-3.15	1.94	—	—	—	0.55
	ALT	-0.16	-0.34	-0.049	0.85	0.71	0.95	0.032
	γ-Glu-Taurine	0.80	0.43	1.31	2.23	1.54	3.72	3.00×10^{-4}
CHC	(Intercept)	-4.73	-6.39	-3.47	—	—	—	<0.0001
	γ-Glu-Lys	1.27	0.85	1.82	3.57	2.34	6.14	<0.0001
CIR	(Intercept)	-2.79	-4.05	-1.55	—	—	—	<0.0001
	γ-Glu-Ala	1.80	0.42	3.52	6.05	1.52	33.7	0.020
	γ-Glu-Leu	-0.066	-3.06	2.24	0.94	0.047	9.42	0.96
	γ-Glu-Ser	-1.35	-5.35	1.86	0.26	4.77×10^{-3}	6.44	0.41
	γ-Glu-Taurine	-2.28	-5.07	-0.33	0.10	6.27×10^{-3}	0.72	0.064
HCC	(Intercept)	-1.87	-2.90	-0.90	—	—	—	2.00×10^{-4}
	γ-Glu-Ala	-1.13	-2.44	-0.14	0.32	0.087	0.87	0.050
	γ-Glu-Citrulline	3.51	0.45	7.00	33.4	1.57	1.10×10^{3}	0.033
	γ-Glu-Thr	-1.65	-5.12	0.49	0.19	5.95×10^{-3}	1.63	0.27
	γ-Glu-Phe	6.99	-0.52	14.7	1.09×10^{3}	5.92×10^{-1}	2.50×10^{6}	0.063

Note: The en-dashes in the 95% CI columns indicate that these values could not be calculated. Biomarker and coefficients are used in MLR model to calculate the probability of each disease. Intercept indicates the constant term in MLR models.

relatively lower contributions to the separation ability of the MLR models (Table 2). Overall, for all types of liver diseases, the MLR models mostly based on γ-glutamyl dipeptides provided complementary results, even in the second (validation) cohort.

γ-Glutamyl dipeptides as biomarkers for HCC and NAFLD

To evaluate the diagnostic potential of γ-glutamyl dipeptides for HCC, we compared their diagnostic abilities with that of AFP, an established marker for HCC (Fig. 4). We found that the MLR models using four γ-glutamyl dipeptides (γ-Glu-Ala, γ-Glu-Citrulline, γ-Glu-Thr, γ-Glu-Phe) (Table 2) were better at distinguishing HCC from CHC and CIR (AUC = 0.881) than AFP (AUC = 0.760) (Fig. 4).

We further investigated the biomarker specificities by comparing the serum γ-glutamyl dipeptide levels in GC and HCC patients (Supplementary Fig. 2 and Table 4). The analyses revealed significant differences, with the exception of γ-Glu-Phe, and the levels of γ-glutamyl dipeptides were notably low in GC.

Differences in the levels of γ-glutamyl dipeptides were also observed in NAFLD. The levels of six γ-glutamyl dipeptides (γ-Glu-Val, γ-Glu-Thr, γ-Glu-Leu, γ-Glu-His, γ-Glu-Phe, and γ-Glu-Arg) were significantly higher ($p < 0.05$; Mann–Whitney test) in SS than in NASH (Supplementary Fig. 3 and Table 5). Although further investigations are necessary, these dipeptides can be used as noninvasive biomarkers in rapid screening for SS and NASH.

Mechanism of γ-glutamyl dipeptide biosynthesis

To confirm the γ-glutamyl dipeptide biosynthesis pathway, the hepatic metabolism was investigated using a mouse model. In

Fig. 2. Representative whisker box plots of the serum levels of detected transaminases and metabolites in the training cohort. The horizontal lines indicate the upper median, median, and lower median, and the whiskers show the maximum and minimum levels. One plot for AST was outside the range (>500 U/L). *$p < 0.05$, **$p < 0.01$, ***$p < 0.0001$, significance difference by the Kruskal–Wallis test and Dunn's post-test for each marker and two groups in each marker, respectively.

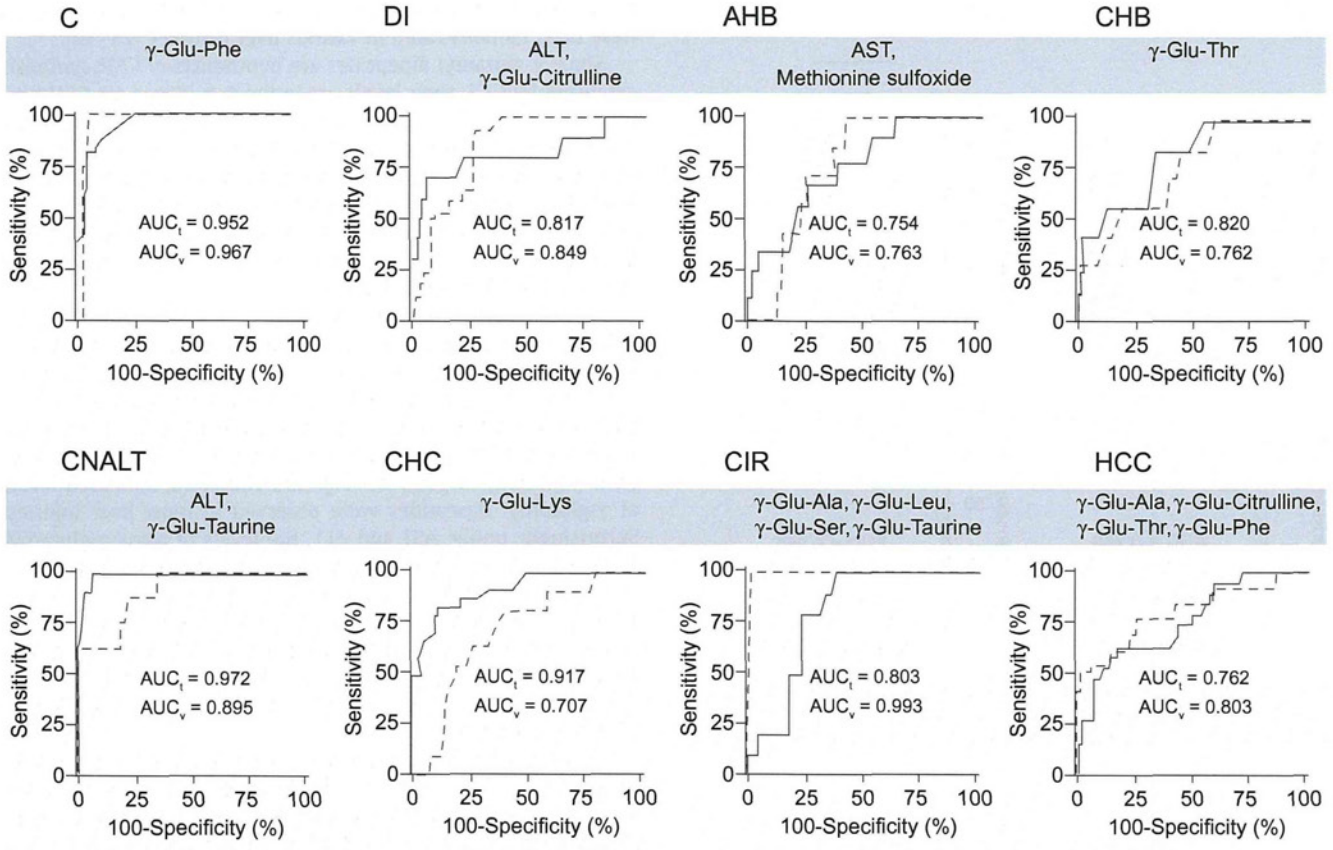

Fig. 3. ROC curve analyses of the ability of γ-glutamyl peptides alone or in combination with AST, ALT and methionine sulfoxide to discriminate each group from all other liver diseases and healthy controls. The solid and dashed curves represent the ROC curves for the training and validation cohorts, respectively. AUCt and AUCv in each panel indicate the AUC values in the training and validation cohorts, respectively. The group label indicates the discriminated group from all the other groups by an MLR model. The biomarkers in each panel were used in the MLR model for discriminating the group, e.g., ALT and γ-Glu-Taurine were the biomarkers for discriminating CNALT from the other groups. The coefficients and constant term of the MLR model of these biomarkers were summarized in Table 2.

acetaminophen (APAP)-treated mice [12], ophthalmate, a γ-glutamyl tripeptide, was synthesized through consecutive reactions with GCS and glutathione synthetase (GS), the same enzymes that play a role in GSH synthesis [12] (Fig. 5). Therefore, we investigated the alterations in the levels of hepatic amino acids, amines, γ-glutamyl dipeptides, and tripeptides after administration of buthionine sulfoximine (BSO), diethylmaleate (DEM) or APAP (Supplementary Fig. 4). BSO treatment resulted in GCS inhibition [20] and marked reductions in most of the hepatic γ-glutamyl dipeptide and tripeptide levels (Fig. 5 and Supplementary Fig. 4A). In contrast, DEM treatment led to GSH depletion by oxidation of the thiol group in GSH [21], resulting in GCS activation and considerable increases in the hepatic γ-glutamyl dipeptide and tripeptide levels compared with the controls (Fig. 5 and Supplementary Fig. 4A). The hepatic levels of several γ-glutamyl dipeptides and tripeptides were increased with concurrent GSH depletion in APAP-treated mice (Supplementary Fig. 4B and C). These results indicated that in mice, γ-glutamyl dipeptides and tripeptides were certainly synthesized via the ligation of glutamate by various amino acids through consecutive reactions with GCS and GS when GSH was depleted (Fig. 5). The identification details for the γ-glutamyl dipeptide biosynthetic pathway are described in the Supplementary data.

Discussion

Our analyses of 237 serum samples from patients with liver diseases and healthy controls revealed that γ-glutamyl dipeptides were increased in liver injuries and could provide specific information for different liver diseases. In APAP-induced liver injury in mice, ophthalmate, a γ-glutamyl tripeptide, was markedly increased as a byproduct of GSH synthesis [21] (Fig. 5 and Supplementary Fig. 4B). However, in liver diseases in humans, many γ-glutamyl dipeptides were primarily synthesized and secreted from hepatocytes into the blood (Figs. 1 and 5). Although the reason for the difference is unclear, it may be attributable to species differences in the levels and activities of enzymes and transporters [22,23].

In all types of liver disease, oxidative stress resulting from an imbalance between the production of reactive oxygen species (ROS) and the ability of a biological system to detoxify reactive intermediates plays a crucial role in the induction and progression of liver damage independently of its etiology [1]. In patients with hepatitis, oxidative stress is produced by inflammation induced by immunological mechanisms. Upon viral infection, NADPH oxidase produces ROS in neutrophils and macrophages, and ROS are also generated from free iron through the Fenton reaction [24–26]. ROS are further produced in hepatocytes upon

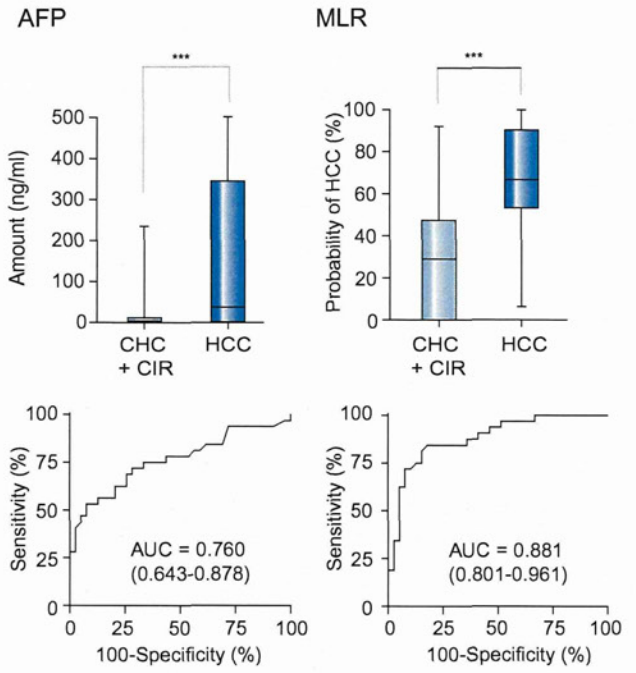

Fig. 4. Whisker box plots and ROC curves of AFP and MLR analyses based on γ-Glu-Ala, γ-Glu-Citrulline, γ-Glu-Thr and γ-Glu-Phe for discriminating patients with HCC (n = 32) from patients with CHC (n = 35) and CIR (n = 18).

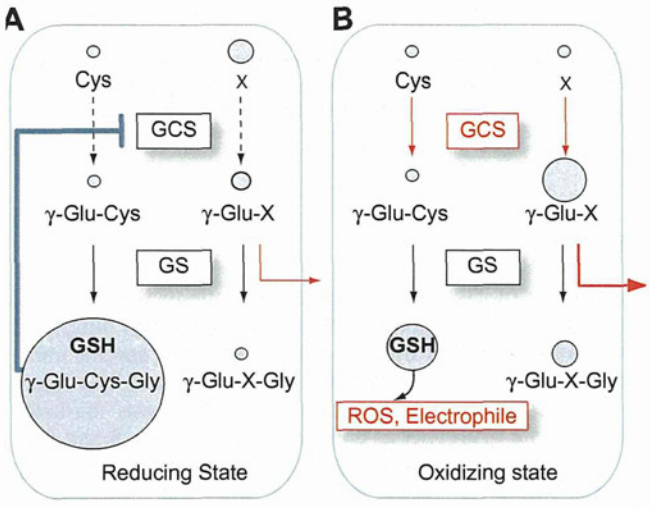

Fig. 5. Biosynthetic mechanism of γ-glutamyl peptides in hepatocytes under (A) reducing conditions and (B) oxidative stress. GCS is feedback-inhibited by GSH under reducing conditions and small amounts of γ-glutamyl dipeptides are synthesized. During oxidative stress, GSH is consumed, leading to GCS activation. This could result in biosynthesis of γ-glutamyl dipeptides, which are then effluxed across the hepatocellular membrane. γ-Glutamyl dipeptides and tripeptides are indicated by γ-Glu-X and Glu-X-Gly, respectively (X = amino acid or amine).

the release of inflammatory cytokines, such as tumor necrosis factor-α and interleukin-1β from inflammatory cells [27]. GSH is the most abundant antioxidant in hepatocytes, and helps to protect cells against ROS. Upon depletion of GSH, ROS induce oxidative stress resulting in liver damage, and reduced GSH levels have been demonstrated in various liver diseases [28–30].

Since γ-glutamyl dipeptides are byproducts of GSH synthesis catalyzed by GCS, their levels are indirect evidence for GSH production (Fig. 5). Different levels of γ-glutamyl dipeptides were observed in different types of liver disease and each γ-glutamyl dipeptide showed a somewhat different variation pattern among liver diseases (Fig. 2). This might be attributed to differences in hepatic levels of amino acids (the substrate of GCS) among liver diseases, though further studies are necessary to understand the details of this observation.

In healthy controls, the γ-glutamyl dipeptide levels were low. This occurred because under reducing conditions, the level of hepatic GSH was high and a small amount of GSH was biosynthesized (Fig. 5A). However, in the patients with liver diseases, GSH was consumed to neutralize the generated ROS, which in turn led to GCS activation, resulting in the biosynthesis of GSH together with γ-glutamyl dipeptides (Fig. 5B). Therefore, increased levels of γ-glutamyl dipeptides were observed in most liver injuries. Surprisingly, unlike AST and ALT, the levels of most γ-glutamyl dipeptides were markedly increased in asymptomatic individuals with AHB and CNALT (Fig. 2 and Supplementary Fig. 1), possibly because viral infection induced ROS generation followed by GSH depletion, which led to the biosynthesis of GSH and γ-glutamyl dipeptides (Fig. 5). We hypothesize that sufficiently high levels of GSH production neutralized ROS, resulting in lower incidences of AHB and CNALT.

There are relationships between liver diseases attributable to HCV infection and oxidative stress parameters, such as ROS, antioxidants, and inflammation. Oxidative stress increased with hepatic disease progression in HCV-infected patients [31]. Consistent with that report, among all the patients with HCV-related liver diseases, the serum levels of γ-glutamyl dipeptides, as indicators of hepatic GSH production, were markedly increased in CNALT and tended to decrease with disease progression (CNALT ⩾ CHC > CIR > HCC) (Fig. 2). These observations led us to conclude that at the time of viral infection (CNALT), a sufficient amount of GSH production can neutralize ROS and thus weaken the pathogenesis of liver damage. However, when GSH production falls below ROS generation, oxidative stress followed by inflammation is induced, resulting in the development and progression of liver diseases. Similarly, the levels of several γ-glutamyl dipeptides were significantly lower in NASH patients than in SS patients (Supplementary Fig. 3), indicating low levels of GSH production in NASH patients. Based on the present observations, we suggest that NASH is susceptible to oxidative stress and progression to liver fibrosis and cirrhosis.

HCC is one of the most common cancers in humans, and primarily develops in patients with chronic liver disease. Its early detection is important because effective treatments are available for the management of non-advanced cancers [32]. Until now, the diagnosis of HCC has relied on combinations of imaging techniques and measurements of the serum levels of AFP [33] and PIVKA-II [34]. Although they are reliable tumor markers for the diagnosis and monitoring of primary HCC, high levels of serum AFP and plasma PIVKA-II have also been observed in some gastric carcinomas [34,35]. However, the serum γ-glutamyl dipeptide levels in GC and HCC patients revealed significant differences, and the levels of several γ-glutamyl dipeptides was notably low in GC (Supplementary Fig. 2). We suspect that this occurred through differences in the tissue activities of the glutathione