profiles include nuclear magnetic resonance (NMR) (Reo 2002), as well as gas chromatography (GC) (Fiehn et al. 2000), liquid chromatography (LC) (Plumb et al. 2003), and capillary electrophoresis (CE) combined with mass spectrometry (MS) (Soga et al. 2003). Typically, the data analysis workflow, starting with raw data, includes filtering or baseline correction, peak detection, alignment of peaks across multiple datasets, generation of a processed data matrix, and statistical analysis such as principal component analysis and partial least squares discriminant analysis to identify significant differences between datasets (Katajamaa and Oresic 2007). Although software packages for automatic processing are available, most of the existing tools were developed or optimized for NMR (Wang et al. 2009; Zhao et al. 2006), LC–MS, and GC–MS (Bellew et al. 2006; Bunk et al. 2006; Fischer et al. 2006; Katajamaa et al. 2006; Katajamaa and Oresic 2005; Smith et al. 2006; Styczynski et al. 2007; Tautenhahn et al. 2008), or for MS alone (Broeckling et al. 2006; Haimi et al. 2006; Karpievitch et al. 2007; Wong et al. 2005). There are currently relatively few tools optimized for CE–MS data analysis (Wittke et al. 2003).

CE–MS is a versatile system, which is well suited for metabolome studies that require high-resolution separation of metabolites and high-detection sensitivity for the analysis of numerous charged and low molecular weight molecules. CE allows for temporal separation of components based on their charge and size and, using MS, most compounds that co-migrate in CE can be resolved (Monton and Soga 2007). However, a major challenge in CE–MS is the variability in migration time. This run-to-run variability in electro-osmotic flow (EOF) is mainly due to changes in the capillary wall or electrolyte solution induced by the sample matrix that results in greater migration time variation compared with other separation methods such as GC or LC. On the other hand, even in a single run, fluctuations of capillary electric condition and run-to-run variability also cause migration time shifts. Although good reproducibility in electrophoretic mobilities was reported for amino acids in CE–MS (Lee et al. 2007), accurate and versatile migration time correction applicable to a large variety of metabolites is necessary. With regard to migration time, once it has been corrected, the actual electrophoretic mobility of molecules in CE can be highly reproducible. In addition, the peak shapes in CE–MS show more diversity and differences compared with those derived from chromatographic techniques such as LC–MS and GC–MS, making the peak detection problem particularly challenging. Thus, software that implements robust migration time alignments and efficient feature analyses is needed for CE–MS data processing. To address these issues, we previously developed MathDAMP, a collection of tools running as a Mathematica package (Baran et al. 2006; Baran et al. 2007). MathDAMP was instrumental in the discovery of metabolite biomarkers (Soga et al. 2006) and for elucidating enzyme and gene functions (Saito et al. 2006; Yoshida et al. 2007). However, the use of complex scripts with large datasets in a generic mathematical environment involves large computation overhead costs and, consequently, a relatively limited throughput. Specifically, the alignment procedures in the electrophoretic dimension are sensitive to measurement quality and require iterative quality control steps and manual optimization of multiple parameters to avoid incomplete alignment due to outlier peaks or large migration time-shifts between datasets. In addition, the datapoint-by-datapoint method for difference detection, as implemented in MathDAMP, which detects significant differences among groups without peak-selection, can yield a number of false-positive results.

The objective of this project was to develop a user-friendly and high-performance platform suitable for differential analysis of CE–MS metabolite profiles that is complementary to existing tools. Therefore, we developed *JDAMP* (Java application for Differential Analysis of Metabolome Profiles), which offers a graphical user interface (GUI) and is designed to facilitate iterative analyses with graphical confirmation of findings. It also uses a specific file converter that allows direct conversion of standard data formats such as NetCDF and CSV (text) file, or Agilent-specific CE-TOFMS raw data to the *JDAMP* original file format. The possibility of directly using Agilent-specific CE-TOFMS raw data has the added benefit of avoiding the large size of intermediate standard file formats based on text or XML. In addition, the newly developed difference detection algorithm allowed for a reduced number of false-positive peaks, which can accelerate discovery-oriented applications of CE–MS-based metabolomics.

2 Materials and methods

2.1 File conversion

The first step in the data processing workflow is file conversion from either standard or vendor-specific raw data files to the *JDAMP* input file. Because a large number of samples are usually analyzed simultaneously to identify statistically reliable differences, the huge file size of conventional standard file formats such as netCDF or mzXML (Hardy and Taylor 2007; Pedrioli et al. 2004) can constitute a significant barrier to large-scale and high-throughput analyses in terms of performance and data storage. Therefore, we implemented a separate program, *dotMZ*, to convert *wiff* data files generated by Analyst QS for Agilent TOF software (Applied Biosystems, CA, USA; MDS SCIEX, ON, Canada) and binary data files (called dot D

dataset) generated by the MassHunter software (Agilent, Santa Clara, CA), which controls the latest versions of Agilent TOF mass spectrometers. To support other vendor platforms as well as non-CE–MS data formats, ASCII-based comma-separated values (CSV) files formatted as generated by Analyst QS and MassHunter, Tab delimited files formatted as generated by MassLynx software (Waters Corporation, Milford, MA), and NetCDF format files that are generated from most types of instruments can also be used as the input. NetCDF, CSV and Tab delimited data files can be converted by *dotMZ* to a specifically designed binary file format (named *ciff* files). Using the application programming interfaces (APIs) of Analyst QS or MassHunter, the Agilent-supplied binary files (*wiff* file or D dataset) can also be directly converted to *ciff* files. In this case, because some Analyst QS or MassHunter libraries are required during conversion, the converter must be installed on a system hosting the Analyst QS or MassHunter software, which is usually provided to owners of Agilent TOF systems.

2.2 Data processing and analysis

The analytical workflow includes data preprocessing, normalization of time-shift (alignment) and signal intensities, and difference-detection, all of which are commonly used feature-detection steps in metabolomics processing of LC–MS and GC–MS data (reviewed in Katajamaa and Oresic 2007). The strategy for data analysis in *JDAMP* is shown in Fig. 1. Overall, it corresponds with the workflow of MathDAMP and its basic algorithm (Baran et al. 2006). Briefly, in the preprocessing step, raw datasets undergo primary binning along the *m/z* dimension to fine resolution (default 0.02 *m/z*) while subtracting the baseline from each electropherogram by polynomial curve-fitting using a nonlinear regression method (Ruckstuhl et al. 2001) and by fixing signals under a specified threshold to 0. Noise values are calculated from signals between 2 and 3 min, where metabolite signals are not usually found. Values obtained in the first minute are not usually used because of unstable signals. The resulting datasets are then further binned to 1 *m/z* unit resolution along the *m/z* axis (secondary binning). Directly binning electropherograms into 1 *m/z* units without primary narrow binning and background-subtraction and noise reduction will result in a low signal/noise ratio for small (narrow peaks (in *m/z* axis)) peaks. Therefore, a primary narrow binning step is preferred to facilitate and maintain the detection of these peaks for subsequent procedures. For the secondary binning process, MathDAMP used n ± 0.5 *m/z* (n; integer) as edges of binning electropherograms. By contrast, *JDAMP* uses n − 0.3 to n + 0.7 *m/z* to limit the possibility of separating isotopic peaks derived from a divalent peak into two different bins. Subsequently, migration time correction (optimized for CE–MS-specific variation) is performed by a dynamic time-warping method (Bylund et al. 2002). This step (1) executes peak selection using the Douglas-Peucker algorithm (Wallace et al. 2004) for each electropherogram (peaks detected at this step are called representative peaks), (2) matches the peaks across datasets by dynamic programming (DP), (3) changes the parameters of the time-normalization function with the optimization method, and (4) returns to (2) until the improvement of the objective function is reduced to a specific limit value. The score produced by DP is used to evaluate the two numerical parameters, α and γ, of the normalization Eq. 1 derived for CE migration (Reijenga et al. 2002), as previously described (Baran et al. 2006).

$$t_R = \frac{1}{(1/\alpha t) - (\gamma/2)}, \tag{1}$$

where t_R and t are the normalized and original migration times, respectively. Briefly, to enhance the robustness of the alignment, the optimization loop steps from (2) to (4) are performed twice using different gap penalties; a larger gap penalty is used to generate a primary normalization function for rough alignment and a smaller gap penalty is used for secondary fine-tuning of the function. The resulting function is then used to rescale the migration times of each dataset, thus eliminating the time shifts for each run. Signal intensities are adjusted to compensate for the compression or expansion of peaks during the normalization and thus conserve the same peak areas, as previously implemented in MathDAMP (Baran et al. 2006). Finally, differences are detected from complete, aligned datasets on a datapoint-by-datapoint basis using a novel difference-detection function that was not implemented in MathDAMP. The results are visualized as numerical values or statistical scores on overlaid electropherograms and 2D maps.

Except for the difference-detection phase, all of the steps include parameters that can be tuned by the user based on the input datasets. This is an important step that can involve considerable user time and input. Therefore, the GUI was designed to facilitate quality control and optimization of iterative parameters by the user. The GUI is implemented in Java language. The GUI is easy to use and allows interactive data processing with visualization. On the other hand, the calculation engines are written in C++ for rapid performance. Each process was implemented as a separate program to benefit programmers who want to write scripts to create directly executable files for routine analyses.

A datapoint-by-datapoint approach was originally implemented in MathDAMP to highlight differences between multiple datasets. This approach enables the

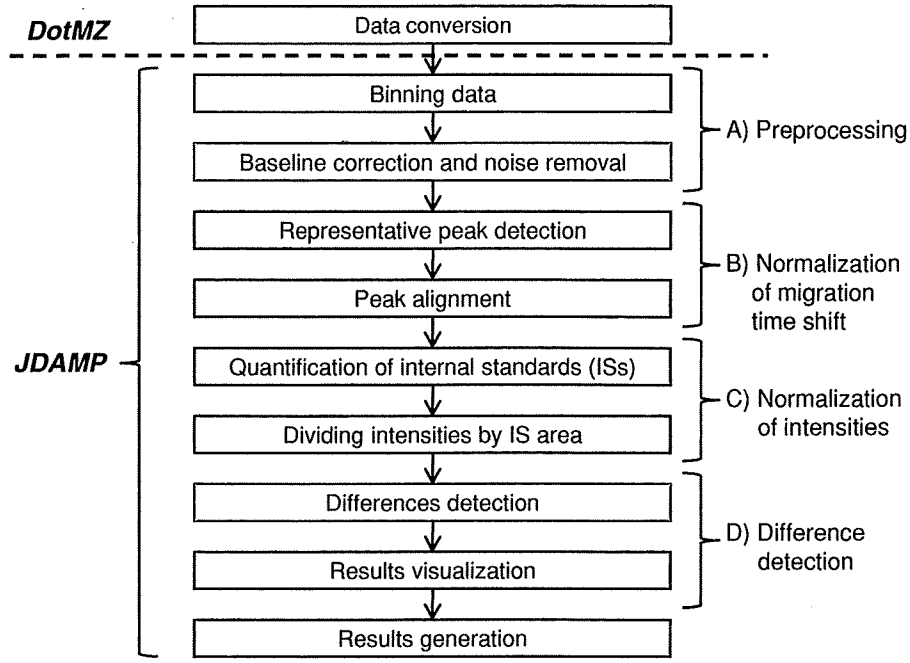

Fig. 1 Schematic representation of the analytical workflow performed by the *dotMZ* converter and *JDAMP*. In preprocessing (A), binning datasets, baseline correction for eliminating background drift and noise removal to delete small-intensity signals below a user specified S/N are performed. In the migration-time normalization procedure (B), representative peak detection, peak matching across datasets with dynamic programming, and correction of migration times are conducted. In the normalization of intensities step (C), internal standards selected by the user are used to normalize the intensities in the entire datasets. For users who do not use internal standards, this process can be omitted. In the difference detection step (D), significant differences are detected depending on multiple criteria and are visualized as overlaid electropherograms with 2D plots

identification of differences while avoiding the limitations of peak-selection for CE–MS electropherograms and the common resulting problem of missing values. However, this method yields a number of false-positives, e.g., a data point at the edge of a peak exhibiting a significant difference is recursively selected as a different result. In addition, the noise-related regions of the electropherograms are sometimes highlighted for reasons such as incomplete background-correction and noise removal. To eliminate such false-positive results, we defined an additional intuitively interpretable, simple evaluation function E;

$$E = \frac{\sum_{t\in\Phi} I_t}{\mathrm{E}_{t\in\varepsilon}|I_t - G_t|} \times \frac{AR}{AR_{\max}} \times \frac{T}{T_{\max}}, \qquad (2)$$

where AR and T represent the intensity differences that are significant in both absolute and relative terms (absolute × relative difference, named ABSRel) (Baran et al. 2006) and the t-score of intensities at the selected time point, respectively. $\mathrm{AR_{max}}$ and $\mathrm{T_{max}}$ are the maximum ABSRel and t-score values in the dataset, respectively. $\mathrm{I_t}$ is the signal intensity for the actual datapoint and $\mathrm{G_t}$ is the height of the Gaussian curve at time-point t. Because Φ and ε are the peak area and the Gaussian area along the time axis, respectively, the numerator and denominator of the first term become the peak area and the degree of distortion from the Gaussian curve. First, to determine the peak area, the electropherograms in a group in which the average intensities of the points of interest are larger than that of the others, are averaged. Second, both the leading and trailing peak edges are identified by moving away, in both directions, from the local maximal intensities. The peak edges are assigned to the first datapoints that are below the threshold (5% of the local maximal intensity). Third, a Gaussian curve is fitted to the peak shape using the simplex method and the differences between the curve and the peak are summed. Then, the datapoint-by-datapoint detection score, using function E, is used to increase the weight of the contribution of datapoints located in regions with larger and more statistically significant differences, and with better Gaussian peak shapes.

The performance of peak-selection based on the difference-detection function using the Douglas-Peucker algorithm (Wallace et al. 2004) was compared with the datapoint-by-datapoint method with and without evaluation using function E.

2.3 Test data

To test the utility of the software to detect differential features in complex datasets, we processed data collected by CE–MS analysis of mixtures of standards in which a few metabolites were spiked at different levels. The

preparation of individual standard solutions and the CE-TOFMS condition and instruments were as described elsewhere (Hirayama et al. 2009). We prepared 304 standard metabolites for cation datasets. The concentration of all standard metabolites was 50 μM and 200 μM of methionine sulfone was added as an internal standard. Each mixture was separated into four containers, and then three selected metabolites were additionally spiked into the three bottles at different levels to increase their concentration by 15, 30 and 50%. The selected cationic metabolites were N-α-benzenolarginine ethylester, 2,4-dimethylaniline, and S-(5′-Adenosyl)-L-homocysteine (SAH) and were selected based on their different detection sensitivity. For SAH, the divalent ion peaks were used for the following benchmark experiments. Three replicates of all samples were measured on the same instrument on the same day.

The biological test datasets used for other validations originate from previous studies (Soga et al. 2006). We used serum samples from control mice and mice treated with acetaminophen for 2 h prior to analysis. All numerical experiments were conducted on Windows XP x64 with a Xeon 3 GHz CPU and 8 GB memory.

3 Results and discussion

3.1 File converter

To reduce the file size to be generated, the lowest *m/z* values common to all time-points are memorized and only the difference in the adjacent *m/z* values is stored. The actual *m/z* values for all datapoints are then reconstituted using the sum of the lowest *m/z* and their respective differences. In addition, all data stored in *ciff* files are wrapped in a zlib library (http://www.zlib.net/) to further compress the file size. Details on the file format are available from the *JDAMP* website (http://software.iab.keio.ac.jp/jdamp).

Under our routine measurement conditions (Soga et al. 2006), for each CE–MS run, Analyst QS stores raw data in approximately 100 MB for the cation mode and in approximately 150 MB for the anion mode. Analyst QS can export the raw data to CSV or NetCDF. However, this conversion, without any masking of low abundance intensities, results in an approximately 10-fold increase in file size to approximately 1.0–1.5 GB in the CSV, NetCDF and mzXML formats. On the other hand, the *JDAMP* converter produces *ciff* files that are approximately only 120 and 180 MB for cation and anion data, respectively, which can be easily imported into *JDAMP*. Compared with the use of CSV, NetCDF or mzXML files, the file conversion time is also reduced from 20–40 to 3–4 min, on average. These significant improvements contribute to reduce the processing time for subsequent analysis because file-access time, an important variable in processing numerous large data files, is shortened. Common file conversion tools, such as mzStar (http://tools.proteomecenter.org/mzStar.php), Analyst QS and MassHunter, include an option to eliminate signals below a user-defined intensity threshold to prevent this enlargement of outputs. However, such data reduction should not be implemented during file conversion because this might reduce the possibility of finding significance related to the small peaks; therefore, such functions should be implemented in subsequent analytical processes to enable the users to use small data files without additional file conversion.

The *ciff* file contains data indexes to separate data along the mass spectral and electrophoretic axes, and to reduce the access time to a specific data block in the *ciff* file. Compared with common data formats that represent a series of mass spectra, as in CSV or TXT format, which only allow fast data access to mass spectral data, the *ciff* format enables rapid access to both mass and time dimensions, which significantly reduces calculation times for handling electropherograms.

3.2 Software features

Screenshots of the graphical user interface are shown in Figs. 2, 3, and Supplementary Information Fig. S1. First, the data files (converted with *dotMZ*) for two or more groups to be compared are imported (Fig. S1A). Then, the user specifies the options for preprocessing such as a threshold for the signal/noise ratio. The baseline correction with primary and secondary binning is then executed. Spike noise, defined as signals that are continuous in time for less than the user-defined threshold, is also eliminated at this step (Fig. S1B). In the next step, the user can specify criteria for peak selection and select the DP parameters to be used for migration-time normalization (Fig. S1C); these include the distribution of representative peaks over time or the *m/z* axis, and the gap penalties (Baran et al. 2006). After the migration-time alignment is completed (Figs. S1D and S1E), the internal standard(s), commonly used in CE–MS systems to compensate for changes such as ionization efficiency, injection volume and sensitivity of MS (Ohnesorge et al. 2005), must be chosen to normalize the signal intensities to account for systematic bias between separate measurements and to limit variation to biologically significant variation. However, this step can be omitted if not necessary. The detected differences are visualized directly on 2D density plots (time and *m/z* dimension in Fig. 2A). As recently demonstrated (Erny and Cifuentes 2007), such 2D maps of CE–MS data facilitate intuitive visual inspection of large datasets, which enables the identification of relevant redundant ions such as

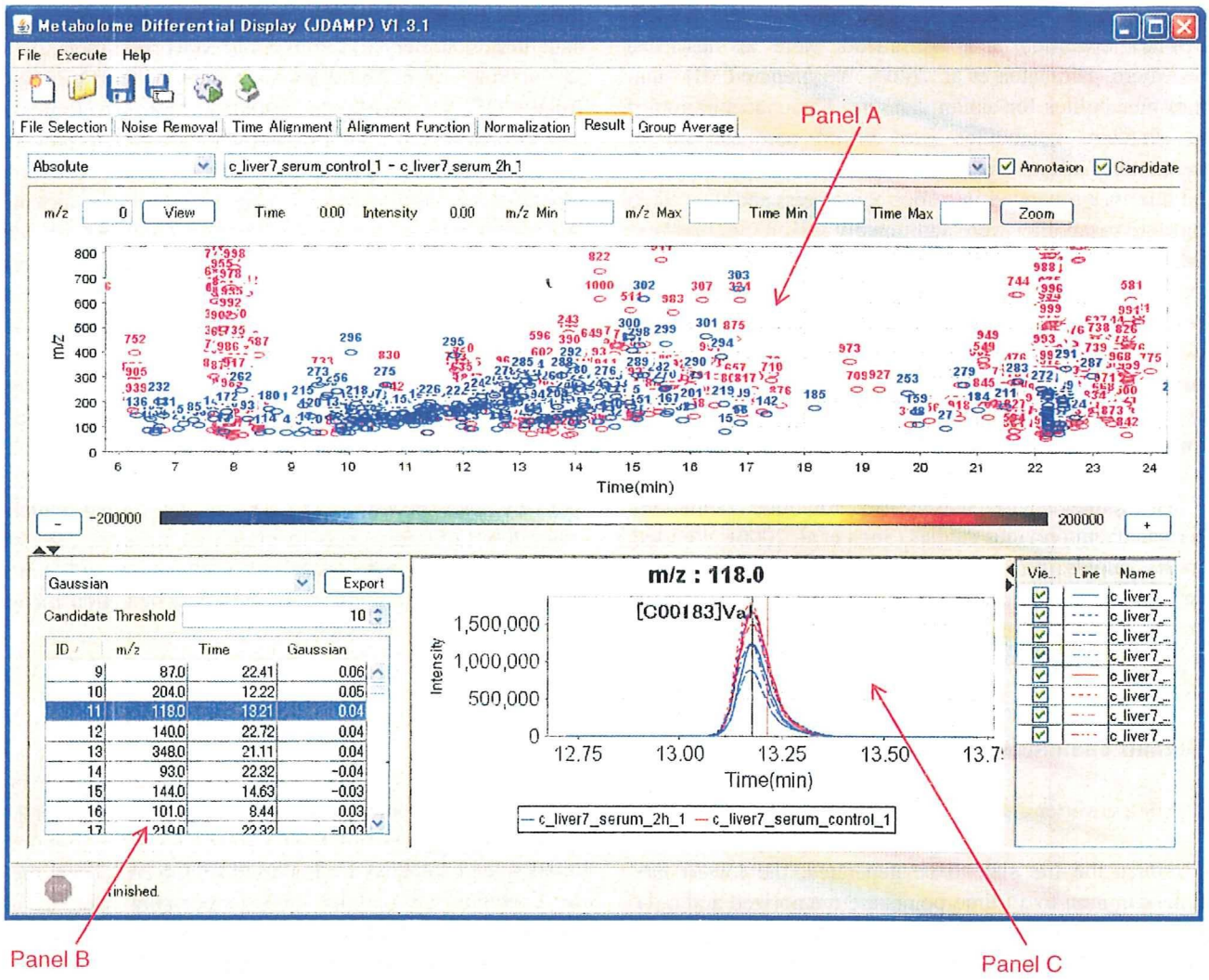

Fig. 2 Screenshots of *JDAMP* results windows. Panel A displays the location of detected significant differences (*red* labels) and of known compounds (*blue* labels). Details of the differences identified are shown in Panel B. An electropherogram overlay is shown for the selected features in Panel C. Other windows, e.g., 2D plots to visualize the averaged intensities within a group and electropherograms of normalized internal standards, are accessible when the respective tab is clicked and the setup window for each process is spawned from the menu or gear icons

fragment ions and adducts, and to differentiate between multiple samples. The map also allows quick overall evaluation of run quality, which is more comprehensive than the total ion electropherogram alone, and yields more readily interpretable information. For example, we empirically know that our CE–MS data always include peaks derived from salts and neutral molecules that appear as vertical smear lines during the first few and last minutes of measurements, respectively. Because of their peak-like appearance, they are not completely removed by baseline correction and the noise-filtering process; however, they are clearly visualized on 2D maps. Such peaks should be eliminated when performing differential analysis using CE–MS data by selecting the corresponding migration time windows for data removal.

To aid visual confirmation of automatically detected differences, a list of significant differences and the corresponding overlaid electropherograms are displayed and linked to each other for easy access to the datapoints of interest (Fig. 2B, C). A user-supplied list of known compounds (chemical standards) can be used to annotate the data and can be visualized on the same figure to facilitate the identification of metabolites in the dataset (Fig. 2A, C), even though further confirmation, such as spiking experiments may be required for reliable identification. *JDAMP* generates structured summary reports, including the detected difference matrix, and corrected electropherograms for whole datasets and a list of detected individual differences for further external analysis with other tools.

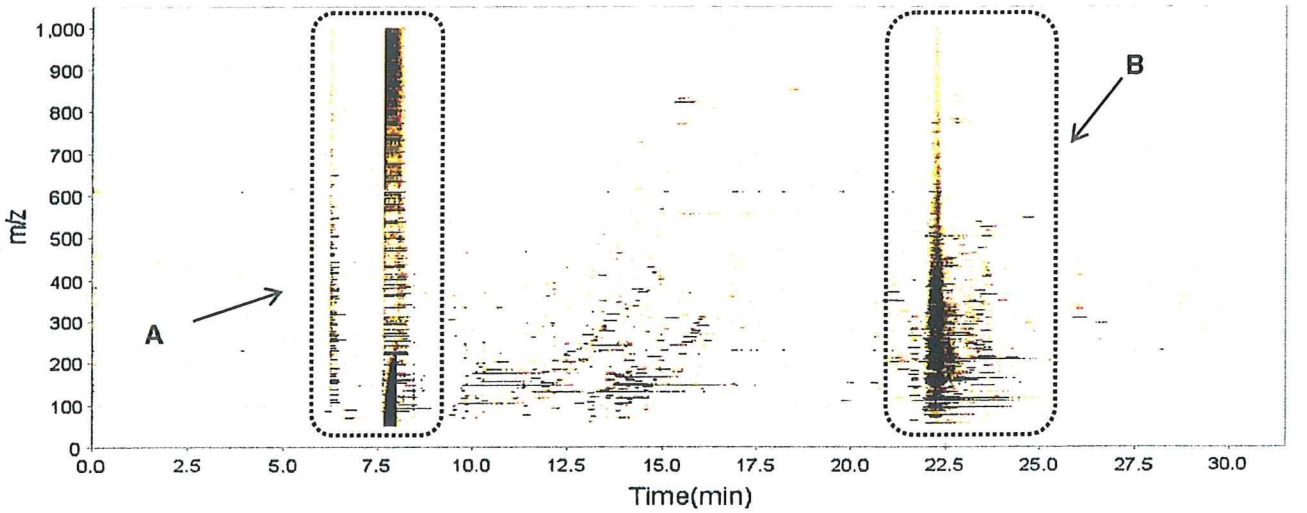

Fig. 3 A typical 2D plot of CE–MS data (time and *m/z* axis) generated after background subtraction and noise filtering. (A) Double vertical smears originating from the early-eluting salt ions or from a sharp baseline drift often occur just after the elution of salt ions. (B) A wider vertical smear derived from a cohort of late-eluting neutral peaks

As described by others (Robinson et al. 2007), the MathDAMP alignment procedure for migration times has some limitations when the datasets are highly dissimilar and users must tune the alignment options to accommodate datasets. As an alternative, we devised the GUI to facilitate prompt quality confirmation by including parameters for alignment algorithms and the range for eliminating unnecessary/undesirable data, and to execute the process iteratively. The optimization options or parameters of the alignment procedure are described in Supplementary Information Text S1 with an example of processing results (Fig. 3).

3.3 Preprocessing for noise reduction

In the preprocessing step, we used a single region of the electropherogram to calculate the noise value, which was used as a threshold to remove noise of low intensity. Supplementary Information Fig. S2 shows the total ion electropherograms and extracted electropherograms of mouse serum datasets. Except for the region around the peaks derived from the analytes and neutral molecules, the deviations are almost constant, and noise was clearly removed. When *JDAMP* is applied to non-CE–MS systems, the current denoising method may not completely eliminate all of the noise across the chromatogram because in LC–MS, for example, such noise generally changes due to a variable mobile phase composition (gradient) resulting in more variable background drift and noise levels.

3.4 Alignment of multiple datasets

Figure 4 and Supplementary Information Fig. S3 depict the differences in migration times between matched peaks in two samples before and after the alignment procedure. The average standard deviations of the migration time difference between five comparisons were reduced from 0.260 min (0.64%) to 0.0190 min (0.047%). In the alignment procedure, although the datasets included a few mismatched representative peaks in the DP phase, most of the correctly matched peaks allowed us to optimize the parameters for Eq. 1 and to produce accurate alignments.

Overall, migration alignment is very useful to correctly match the corresponding signals for differential analysis. However, the electric current condition in the capillary during measurement shows different profiles and is a possible factor that affects the migration time shift and therefore the quality of the alignment results, (Supplementary Information Fig. S4). Variation in the pH of the formic acid solution is also a possible factor responsible for migration time variation. Although the peaks with faster electrophoretic mobility were correctly aligned, the peaks derived from neutral molecules migrating after 22.5 min (Fig. 4A) showed greater variance and were not accurately aligned (Fig. 4B). These peaks represent the main source of poorly aligned signals. However, this part of the data should be discarded or should not be used in subsequent processing because the separation is non-electrophoretic and this part of the data represents neutral molecules.

3.5 Differential detection performance

Methods for peak detection and deconvolution for LC–MS and GC–MS have been developed (Halket et al. 1999; Vivo-Truyols et al. 2005a, b). Although a similar method for CE peaks has been proposed (Garcia-Alvarez-Coque et al. 2005; Wee et al. 2008), its application to actual data

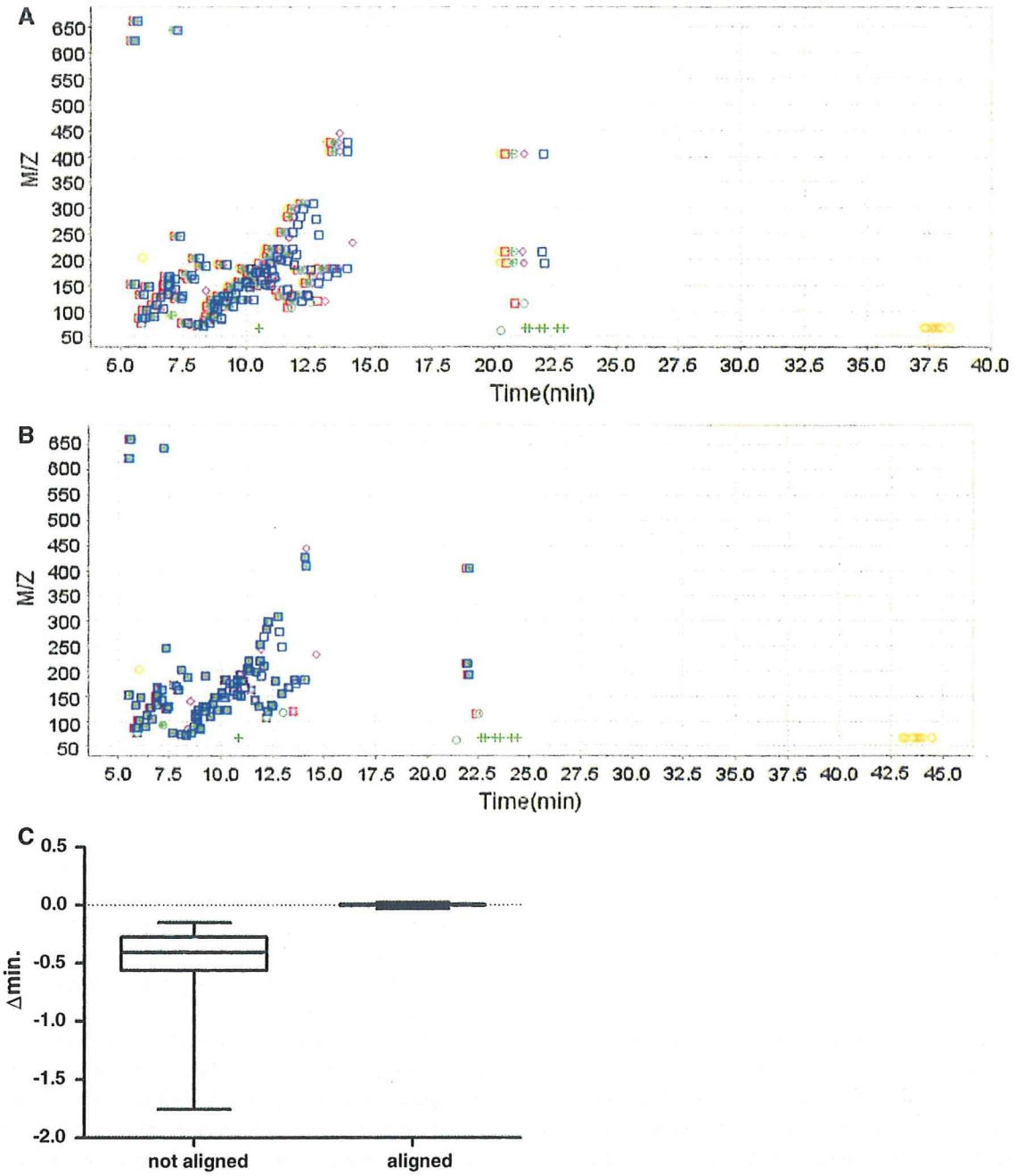

Fig. 4 Migration time alignment results using a standard metabolite mixture. 2D plots (migration time and *m/z* axis) of (**A**) and (**B**) shows actual and normalized representative peak locations, respectively. A total of six samples were aligned simultaneously in this case, and the representative peaks derived from a sample are colored with a single color. Box-and-whisker plots show the difference in migration times of the same representative peaks between two samples (Y-axis, Δmin) in (**C**). The horizontal lines in the box indicate the first quartile, median, third quartile, and the whiskers indicate the maximum and minimum values. Orphan peaks that did not have a matching peak in the corresponding sample and the misaligned peaks in DP phases were eliminated. Plots for other sample combinations and plots showing all differences without elimination of the unmatched peak data are shown in Supplementary Information Figure S3

requires smoothing for noise reduction (Liu et al. 2003; Vivo-Truyols et al. 2005a, b), a process that remains controversial because smoothing distorts the peak area (Wallace et al. 2004). MathDAMP uses the Douglas–Peucker algorithm (Wallace et al. 2004) to select peaks, but only for migration-time alignment, and avoids peak area-based differential feature identification to bypass CE–MS peak detection difficulties. To evaluate the two approaches, we implemented a peak area-based method for difference detection (named area-based detection) and compared its

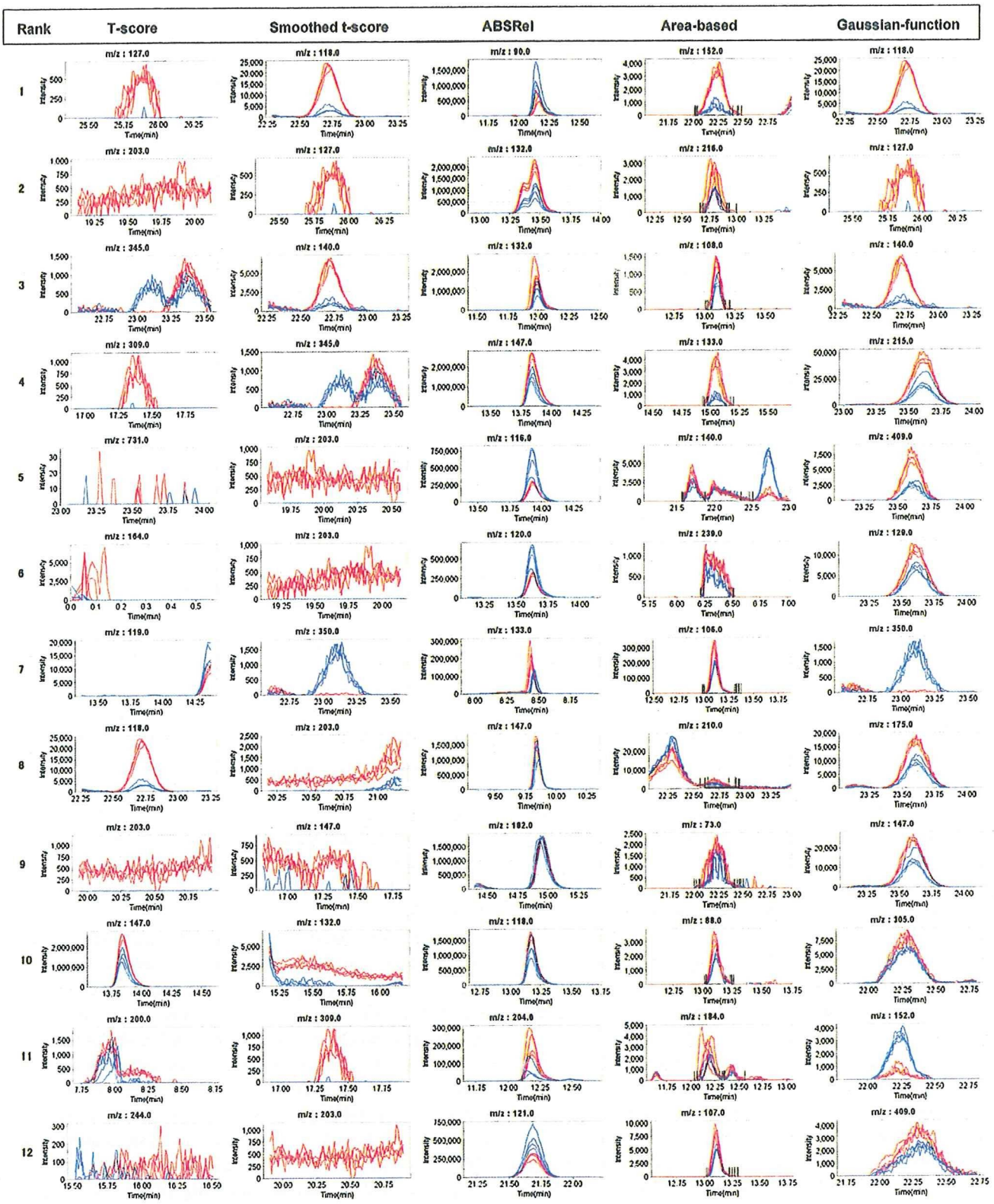

Fig. 5 Overlaid electropherograms of the results ranked in the top 12 from calculations performed using datapoint-by-datapoint *t*-score, smoothed *t*-score, ABSRel, area or Gaussian area functions. The *red* and *blue* curves represent the peaks for the samples and control datasets, respectively

performance with the datapoint-by-datapoint methods. The criteria for the latter included ABSRel, moving average *t*-score using the selected datapoints and the four preceding and subsequent datapoints in the time dimension (named smoothed *t*-score), and Eq. 1 (named Gaussian function).

The ability of *JDAMP* to detect differences was tested using a standard mixture and the results are summarized in Supplementary Information Table S1. Overall the Gaussian function best ranked the N-a-benzenol arginine ethylester metabolites whose concentration was increased compared with the other detection criteria, while the *t*-score showed the worst performance. For example, N-a-benzenol arginine ethylester, which showed high detection sensitivity, was ranked first in the 30 and 50% differentiated solutions and third in the 15% differentiated solution. By contrast, the divalent ion of SAH, which showed low detection sensitivity, was differentially selected only when spiked at an additional level of 30 and 50% and was not found in the 15% spiked samples among 1000 signal rankings based on the *t*-score and Gaussian criteria. For the Gaussian-based results with 2,4-dimethylaniline, even though the accuracy was greater than with the smoothed *t*-score, the rapid deterioration of the results with decreasing spiked amounts suggest that the Gaussian method did not improve the accuracy of peak detection for small peaks or smaller differences. In the datasets used for these validation experiments, a relatively high baseline (background noise), possibly due to lock mass errors or related phenomena, was observed and incomplete elimination of the background yielded a large number of false-positives, which contributed to the deterioration of the differential ranking of 2,4-dimethylaniline and divalent SAH.

For *JDAMP* analysis results using biological samples, Fig. 5 depicts the overlaid electropherograms that were ranked in the top 12 based on these criteria. The overlaid electropherograms for all features ranked within the top 13–50 are listed in Supplementary Information Fig. S5. To reduce false-positive results in the area-based method, peaks that were only found in a few samples across the datasets were eliminated. Here, we used three samples (i.e., 75% of samples in a group of 4 contain the peak) as the threshold and the missing values were set to 0.

Overall, the *t*-score- or smoothed *t*-score datapoint-by-datapoint-based algorithms can detect discriminating peaks when most of the peaks in a group are clearly higher or lower than the peaks in the other groups. However, the electropherogram at 203 *m/z*, ranked 2nd and 9th by the *t*-score method and 5th, 6th, 8th and 12th by the smoothed *t*-score method, showed no clear peaks and the features were scored as significant because of baseline levels that were reproducible between replicates but very different between the two groups. Such false-positives can be rejected by visual inspection of the confirmation plots, demonstrating the importance of this feature. On the other hand, these tests can detect small but clear differences, such as the results at 127 *m/z* and 309 *m/z*, which were ranked 1st and 4th by the *t*-score method and 2nd and 11th by the smoothed *t*-score methods, but were not apparent in the area-based method. This is an important feature of *t*-score-based methods that can be missed (false-negative) by other procedures. Compared with the *t*-score-based method, the results ranked as most significant by the ABSRel index include mainly clear, smooth and high-intensity peaks, even though the algorithm evaluates the datapoints without actual peak detection. Although some peaks manifest significant differences, such as the electropherograms ranked 2nd and 3rd (at 132 *m/z* of $P = 0.017$ and $P = 7.29 \times 10^{-4}$, respectively), the peaks ranked 8th and 9th (147 and 182 *m/z*) exhibited no significant differences ($P = 0.097$ and 0.13, respectively). This is derived from a bias in the ABSRel index, which sometimes highlights signals that are statistically less significant but which show large differences in absolute intensities. The ABSRel index was previously implemented to reduce such bias, which is common when only the absolute difference index is used. However, it cannot be completely eliminated for overwhelmingly large peaks (Baran et al. 2006). By contrast, imperfect alignment or jagged or distorted peaks appear to be responsible for the differences observed in the large internal standard peak at 182 *m/z*, which would be expected to show no difference. Finally, the area-based method could detect peak-like shapes which could be ranked as small, but clearly different peaks (e.g., ranks 1 to 4). However, for effective performance in areas where multiple peaks exist in close proximity, e.g., those ranked 5th and 11th, a more sophisticated peak edge detection algorithm may be needed because some of the peak edges were incorrectly assigned to the neighboring peak and may compromise statistical comparisons.

For the differential analysis of hyphenated MS profiles, both *MZmine* and XCMS perform peak detection, produce lists of statistically significant differences by comparing detected peaks and allow imputation of missing data (Katajamaa and Oresic 2005; Nordstrom et al. 2006; Smith et al. 2006). In addition, a rerun of the integration procedure after dataset alignment to facilitate statistical comparisons is possible because not all of the peaks are detected and aligned in all samples (Katajamaa and Oresic 2005; Nordstrom et al. 2006; Smith et al. 2006). However, the power of their deconvolution algorithms for complex peak shapes and overlapping peaks is unclear. Although datapoint-by-datapoint-based difference detection can bypass such additional procedures, this method alone cannot directly cope with peak deconvolution. However, it can highlight clear differences in irregular and overlapping

peaks, such as the result at 345 *m/z*, which was ranked 3rd by the *t*-score method and 4th by the smoothed *t*-score method. When the objective is to find only statistically significant differences, a low threshold for the peak detection process should be set to allow for the detection of small but significantly different peaks. However, such a procedure involves trade-offs that can compromise either the sensitivity or specificity of the area-based method. Using the Gaussian-based method, the results ranked within the first 12 include signals from both small and large intensities that display, by definition, Gaussian peak-like shapes and also yield small *P*-values. The problem of whether the differences, which are small in absolute terms but statistically significant, represent biologically significant differences needs to be evaluated by further experiments and analyses. Although all methods generate false-positives, Gaussian-based difference-detection appears to minimize their occurrence by combining the high sensitivity of the datapoint-by-datapoint approach and the enhanced specificity of Gaussian fit to normal electrophoretic peaks, thus avoiding noise-related signals. This improvement in accuracy is important to reliably identify discriminating features from large-scale CE–MS datasets. Therefore, the multiple different calculations performed by *JDAMP* represent a major advantage over existing tools and are useful to maximize the detectability of significantly different features.

3.6 Comparison with MathDAMP and MZmine

Using the two criteria, smoothed *t*-score and ABSRel, which are implemented in both MathDAMP and *JDAMP*, the similarities in ranking of differences for the top 50 features in the mouse liver samples are depicted in Supplementary Information Figs. S6A and S6B. Of the detected differences, 64% by ABSRel and 44% by smoothed *t*-score were detected by both tools. ABSRel showed similar profiles to the smoothed *t*-score in MathDAMP and *JDAMP*. Although these differences might predominantly arise from differences in bin borders, the profiles determined using the smoothed *t*-score method were markedly different and were sensitive to the quality of the processing steps prior to the difference detection process. Because the *t*-score method tends to find smaller peaks compared with ABSRel, this discrepancy between MathDAMP and *JDAMP* might explain the differences observed. In the results based on ABSRel, although several peak-shaped results (e.g., a peak at 122 *m/z* (Fig. S6C)) were included only by MathDAMP, a high ranking was assigned to these peaks was due to an overestimation of the significance resulting from incomplete migration time normalization. In the results obtained using the smoothed *t*-score method, those derived from incomplete baseline adjustment, such as in Figs. S6D and S6E, were observed using MathDAMP. Although the former results might be common to both *JDAMP* and MathDAMP and should be eliminated by tuning the options to improve the alignment, the Gaussian-based method implemented in *JDAMP* reduces the detection of the latter cases, as shown in Fig. 5.

With respect to the computation times, the preprocessing takes about 40 to 50 min in both *JDAMP* and MathDAMP because they are based on the same external C++ code module. The migration time alignment process of *JDAMP* requires only a few seconds per dataset while MathDAMP takes about 1 to 2 min under the same conditions. The subsequent steps require 1–2 min for JDAMP and 4–5 min for MathDAMP. Using mouse serum samples (eight datasets), the subsequent procedures including alignment and peak detection took 12 and 38 min for *JDAMP* and MathDAMP, respectively.

We also analyzed the data for the standard metabolite mixture using *MZmine*, a tool that provides peak detection-basis analysis for LC–MS data (Katajamaa et al. 2006). The processing procedure for comparative experiments using *MZmine* is described in Supplementary Information Text S2. Supplementary Information S7 shows typical results obtained using *MZmine*. When the data are converted to mzMXL after eliminating low-intensity signals (<100 cps) to decrease the converted file size, *MZmine* did not detect the expected metabolites and mostly produced false-positives (Supplementary Information Figs. S7A and S7B). In fact, these noise peaks were much larger than other peaks derived from actual metabolite (Figs S7-C and S7-D); therefore, the small deviations in these peaks were, although unexpectedly, detected as differences by *JDAMP* or as peaks by *MZmine*. Only the mzXML data converted without filtering, although each file becomes larger than 1 GB, was successfully used in the subsequent analyses, which might limit the throughput in larger analyses. Using the successfully detected results, alignments with migration time tolerance of 1 and 5% failed to match the peaks even though the average standard deviation of the migration times was 0.64% (Figs. S7E, S7F, and S7G). This result was presumably due to the existence of nearby peaks and, therefore, peak detection with larger peak detection threshold might reduce such instances of misalignment. However, such options will limit the chance of discovery. While the power and utility of *MZmine* for LC–MS data analysis is not questioned here, our results suggest that, at least in its current form, its applicability to the specificities of CE–MS data processing may be limited.

3.7 Advantages and disadvantages of JDAMP

The development of a fully automatic procedure is the ultimate goal to increase throughput for large-scale

metabolomic analysis based on CE–MS data. However, current algorithms optimized for CE–MS data processing such as denoising, peak detection, and migration-time alignment include arbitrary parameters that need to be optimized by the data analysts. To facilitate these tasks, we have developed software tools that feature a simple user interface, improved performance and easier optimization of processing parameters using simple operations with intuitive visual confirmation of the results.

Binning datapoints in the *m/z* domain, as performed by *JDAMP*, results in the loss of high mass resolution obtained by TOF–MS or Fourier Transform Ion Cyclotron Resonance (FT-ICR)–MS, and can limit the identification of adducts, isotopic or fragment-derived peaks. However, while it can considerably facilitate compound identification, the differential detection of features using high-resolution data often requires undesirable or unrealistic computational power and time, and introduces additional steps and hurdles. These include, for example, the need for *m/z* correction across datasets that arise from incomplete *m/z* correction by the MS instrument mass lock feature (Hack and Benner 2002; Soga et al. 2006; Wu and McAllister 2003), an intensity-dependent *m/z* shift due to the signal processing capacity of MS detector (Mihaleva et al. 2008), or peak distortion in the *m/z* dimension (Kempka et al. 2004). For these reasons, we elected to use *m/z* binning as a reasonable trade-off. Once the candidate features are found, the users can easily return to the original high-resolution data using vendor-specific software to extract accurate *m/z* values to facilitate compound identification. In addition, external software should be used to confirm that the observed differences do not originate from different but closely spaced peaks in the *m/z* and migration-time direction, or from corresponding peaks that were assigned to different *m/z* bins due to values near the bin limits.

JDAMP implements metabolite difference detection methods based on both area-based criteria with peak selection and on datapoint-by-datapoint criteria without peak selection. The latter method has significant advantages over peak selection methods for handling irregularly shaped or erroneously missing peaks and can thus enhance the sensitivity of difference detection. Although, empirical mathematical functions to describe electrophoretic peaks have been developed, (Garcia-Alvarez-Coque et al. 2005), the actual peak shapes are, as shown in Figs. 5 or S5, more complicated in biological samples. Multiple factors can influence peak broadening in CE–MS including diffusion, Joule heating, interactions of analytes with the capillary wall, pressure-induced parabolic flow, and negative pressure at the capillary outlet originating from the nebulizing gas (Axen et al. 2007); these can make the peak detection problem more difficult. Although the datapoint-by-datapoint approach is hardly affected by this increased complexity, good results require more accurate migration time normalization than the general approach with peak detection and matching. While most generally used alignment methods to generate matched peak matrix result in other difficulties related to peak splitting or merging (reviewed in Robinson et al. 2007), they require only good peak matching. By contrast, the datapoint-by-datapoint approach requires that the peak maximum is properly matched on the normalized electropherograms, otherwise false-positive signals are often generated. However, easy visualization of the original overlaid electropherogram as implemented in *JDAMP* allows to rapidly exclude these signals.

Because of uncertainty in the number of total features or peaks in the dataset, we did not implement *P*-value corrections such as Bonferroni's correction (Shaffer 1995), which can conservatively correct for multiple hypothesis testing in the *t*-test. Users should be aware that false-positive results will be generated from any such multivariate analyses (more likely for larger *P*-values) and could perform simple correction by estimating the total number of peaks or preferably perform additional experiments to confirm the reproducibility of the original findings. For the same reason that peaks are not used for many calculations, the annotation or elimination of redundant data—arising from isotopic peaks, alternatively charged ions, adducts or fragment ions—is not part of the current *JDAMP* features. However, inspection of the 2D maps can reveal such occurrences as characteristically spaced signals that are vertically well aligned, and allow the user to eliminate these apparently significant but potentially misleading features. In addition, the 2D maps can assist the users to identify and eliminate regions where salt and neutral molecules migrate (visualized as obvious vertical streaks across the datasets). However, further developments are necessary for automatic elimination of those undesirable results using objective criteria. Instrument-specific artifacts previously reported for Orbitrap MS (Brown et al. 2009), such as instrument-dependent and run-to-run difference, were also observed in CE-TOFMS data. For example, unclear but weak vertical lines sometimes appear migrating just prior (left) to the neutral molecule-derived band. These occasionally observed horizontal bands along electropherograms at 92 *m/z*, which are distinct from background ions used for lock mass, may be derived from contamination of the nitrogen gas. Further studies are needed to store these empirical rules and to implement general or ad hoc noise filters.

The *JDAMP* file converter and specific file format provide important benefits, even when handling a relatively small number of datasets and are essential when hundreds of datasets are analyzed on a routine basis to optimize data storage and improve performance. *JDAMP* is a powerful

and rapid tool that identifies significant differences, and is thus useful for initial high-throughput screening of metabolomics datasets. High accuracy *m/z* values to generate compositional formulae and the manual interpretation of mass spectra may be necessary for reliable identification. A number of vendor-supplied software packages, such as Analyst QS, Mass Hunter and Mass Lynx, are user-friendly and are useful for such tasks. However, they lack specific features for automated and reliable differential feature selection between numerous datasets and are thus complementary to *JDAMP*. On the other hand, many other useful tools based on statistical/mathematical software, such as XCMS (Smith et al. 2006), which is based on the R statistical language (University of Auckland; http://www.r-project.org/), MathDAMP (Baran et al. 2006), which is based on Mathematica (Wolfram Research, Inc.; http://www.wolfram.com/), or other recently described software (Allard et al. 2008) based on Matlab (Mathworks, Inc; http://www.mathworks.com/), remain relatively difficult to use, but can offer extra flexibility that is useful for routine analyses or to combine tools with external packages for further analyses. *MZmine* is another powerful tool with the benefit of a sophisticated user interface, but it was developed primarily for LC–MS data analysis (Katajamaa et al. 2006) and, as shown, may be less useful for CE–MS data analysis. The various difference detection methods implemented in *JDAMP* are currently limited to the comparison of two groups, and to evaluate candidate features individually (univariate testing). Pattern recognition technologies, such as support vector machine or partial least square-discriminant analysis, and artificial neural networks, as well as multivariate analyses such as principal components analysis or partial least squares discriminant analysis, have been widely used to simultaneously evaluate multiple peaks and enhance the potential to discriminate between given samples (Acevedo et al. 2007; Mahadevan et al. 2008). To facilitate such multivariable analyses and to enable multiple comparisons between a greater number of groups (>2), *JDAMP* can export intermediate or final results in several formats for downstream use in other software tools. Further development of visual methods for simultaneous comparison of multiple groups is needed (Baran et al. 2007).

JDAMP might be used for instruments other than the ESI-TOFMS used in this study but, for the differential detection approach of metabolic profiles, accurate quantification of signals is a prerequisite to correctly evaluate the significance of the difference. The wider linearity range for quantification in ESI-TOFMS compared with MALDI–MS provides advantages to quantify the difference in biological sources (Ohnesorge et al. 2005). With the use of a supported data converter, *JDAMP* might also be used with data obtained from other types of mass spectrometers, e.g., ion-trap or quadrupole instruments. However, the higher sensitivity of ESI-TOF–MS compared with these techniques (Simo et al. 2008) enhances the limit of detection of small but significant differences with *JDAMP*.

Finally, with the exception of MathDAMP, most of the other currently available software solutions are not optimized for some of the specificities of CE–MS-derived data (peak shape and migration time shifts) and are also based almost exclusively on standard peak detection-based analysis, which offers advantages but also has limitations, as described above. Therefore, rather than replace these tools, *JDAMP* was designed to fill a gap in metabolic data processing and provide an easy-to-use, complementary tool that offers versatile methods to compare metabolite profiles obtained with CE–MS.

4 Concluding remarks

We developed *JDAMP* to offer simplified and faster quantitative differential analysis of high-throughput CE–MS-based metabolomics data. Our software rapidly processes large datasets, detects differences among multiple datasets using different operations, allows visualization of the results using an intuitive and easy-to-use GUI, and can export analysis reports. *JDAMP* enables complementary peak area-based and datapoint-by-datapoint differential feature identification. We expect the software to considerably simplify the analysis of large CE–MS datasets and the identification of discriminatory features such as potential biomarkers. For academic research purposes, the software, manual and animated tutorials are freely available at http://software.iab.keio.ac.jp/jdamp and the source code is available upon request.

Acknowledgments We thank Dr. Yusuke Tanigawara and Dr. Akito Nishimuta of the School of Medicine, Keio University, Dr. Satoshi Yoshida and Dr. Hideki Koizumi of Kirin Holdings, Dr. Akira Oikawa of Riken, and Dr. Eri Shimizu and Dr. Tadahiro Ozawa of Kao Corporation, for valuable discussions. We also thank Maki Sugawara, Hiroko Ueda, Shinobu Abe, and Kazuki Sugisaki of IAB for measurement, data analyses, and programming, and Dr. Ursula Petralia for editing the manuscript. This work was supported by research grants from the Yamagata Prefectural Government and the City of Tsuruoka.

References

Acevedo, F. J., Jimenez, J., Maldonado, S., Dominguez, E., & Narvaez, A. (2007). Classification of wines produced in specific regions by UV-visible spectroscopy combined with support vector machines. *Journal of agricultural and food, 55*, 6842–6849.

Allard, E., Backstrom, D., Danielsson, R., Sjoberg, P. J., & Bergquist, J. (2008). Comparing capillary electrophoresis-mass spectrometry fingerprints of urine samples obtained after intake of coffee, tea, or water. *Analytical chemistry, 80*, 8946–8955.

Axen, J., Axelsson, B. O., Jornten-Karlsson, M., Petersson, P., & Sjoberg, P. J. (2007). An investigation of peak-broadening effects arising when combining CE with MS. *Electrophoresis, 28*, 3207–3213.

Baran, R., Kochi, H., Saito, N., et al. (2006). MathDAMP: A package for differential analysis of metabolite profiles. *BMC Bioinformatics, 7*, 530.

Baran, R., Robert, M., Suematsu, M., Soga, T., & Tomita, M. (2007). Visualization of three-way comparisons of omics data. *BMC Bioinformatics, 8*, 72.

Bellew, M., Coram, M., Fitzgibbon, M., et al. (2006). A suite of algorithms for the comprehensive analysis of complex protein mixtures using high-resolution LC–MS. *Bioinformatics, 22*, 1902–1909.

Broeckling, C. D., Reddy, I. R., Duran, A. L., Zhao, X., & Sumner, L. W. (2006). MET-IDEA: Data extraction tool for mass spectrometry-based metabolomics. *Analytical chemistry, 78*, 4334–4341.

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.

Bunk, B., Kucklick, M., Jonas, R., et al. (2006). MetaQuant: A tool for the automatic quantification of GC/MS-based metabolome data. *Bioinformatics, 22*, 2962–2965.

Bylund, D., Danielsson, R., Malmquist, G., & Markides, K. E. (2002). Chromatographic alignment by warping and dynamic programming as a pre-processing tool for PARAFAC modelling of liquid chromatography-mass spectrometry data. *Journal of Chromatography. A, 961*, 237–244.

Erny, G. L., & Cifuentes, A. (2007). Simplified 2-D CE–MS mapping: Analysis of proteolytic digests. *Electrophoresis, 28*, 1335–1344.

Fiehn, O., Kopka, J., Dormann, P., et al. (2000). Metabolite profiling for plant functional genomics. *Nature biotechnology, 18*, 1157–1161.

Fischer, B., Grossmann, J., Roth, V., et al. (2006). Semi-supervised LC/MS alignment for differential proteomics. *Bioinformatics, 22*, e132–e140.

Garcia-Alvarez-Coque, M. C., Simo-Alfonso, E. F., Sanchis-Mallols, J. M., & Baeza-Baeza, J. J. (2005). A new mathematical function for describing electrophoretic peaks. *Electrophoresis, 26*, 2076–2085.

Hack, C. A., & Benner, W. H. (2002). A simple algorithm improves mass accuracy to 50–100 ppm for delayed extraction linear matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry, 16*, 1304–1312.

Haimi, P., Uphoff, A., Hermansson, M., & Somerharju, P. (2006). Software tools for analysis of mass spectrometric lipidome data. *Analytical Chemistry, 78*, 8324–8331.

Halket, J. M., Przyborowska, A., Stein, S. E., et al. (1999). Deconvolution gas chromatography/mass spectrometry of urinary organic acids–potential for pattern recognition and automated identification of metabolic disorders. *Rapid Communications in Mass Spectrometry, 13*, 279–284.

Hardy, N. W., & Taylor, C. F. (2007). A roadmap for the establishment of standard data exchange structures for metabolomics. *Metabolomics, 3*, 1573–3890.

Hirayama, A., Kami, K., Sugimoto, M., et al. (2009). Quantitative metabolome profiling of colon and stomach cancer microenvironment by capillary electrophoresis time-of-flight mass spectrometry. *Cancer Research, 69*, 4918–4925.

Karpievitch, Y. V., Hill, E. G., Smolka, A. J., et al. (2007). PrepMS: TOF MS data graphical preprocessing tool. *Bioinformatics, 23*, 264–265.

Katajamaa, M., Miettinen, J., & Oresic, M. (2006). MZmine: Toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics, 22*, 634–636.

Katajamaa, M., & Oresic, M. (2005). Processing methods for differential analysis of LC/MS profile data. *BMC Bioinformatics, 6*, 179.

Katajamaa, M., & Oresic, M. (2007). Data processing for mass spectrometry-based metabolomics. *Journal of Chromatography. A, 1158*, 318–328.

Kempka, M., Sjodahl, J., Bjork, A., & Roeraade, J. (2004). Improved method for peak picking in matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry, 18*, 1208–1212.

Lee, R., Ptolemy, A. S., Niewczas, L., & Britz-McKibbin, P. (2007). Integrative metabolomics for characterizing unknown low-abundance metabolites by capillary electrophoresis-mass spectrometry with computer simulations. *Analytical Chemistry, 79*, 403–415.

Liu, B. F., Sera, Y., Matsubara, N., Otsuka, K., & Terabe, S. (2003). Signal denoising and baseline correction by discrete wavelet transform for microchip capillary electrophoresis. *Electrophoresis, 24*, 3260–3265.

Mahadevan, S., Shah, S. L., Marrie, T. J., & Slupsky, C. M. (2008). Analysis of metabolomic data using support vector machines. *Analytical Chemistry, 80*, 7562–7570.

Mihaleva, V., Vorst, O., Maliepaard, C., et al. (2008). Accurate mass error correction in liquid chromatography time-of-flight mass spectrometry based metabolomics. *Metabolomics, 4*, 171–182.

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

Nicholson, J. K., & Wilson, I. D. (2003). Opinion: Understanding 'global' systems biology: Metabonomics and the continuum of metabolism. *Nature Reviews. Drug Discovery, 2*, 668–676.

Nordstrom, A., O'Maille, G., Qin, C., & Siuzdak, G. (2006). Nonlinear data alignment for UPLC–MS and HPLC–MS based metabolomics: Quantitative analysis of endogenous and exogenous metabolites in human serum. *Analytical Chemistry, 78*, 3289–3295.

Ohnesorge, J., Neususs, C., & Watzig, H. (2005). Quantitation in capillary electrophoresis-mass spectrometry. *Electrophoresis, 26*, 3973–3987.

Pedrioli, P. G., Eng, J. K., Hubley, R., et al. (2004). A common open representation of mass spectrometry data and its application to proteomics research. *Nature Biotechnology, 22*, 1459–1466.

Plumb, R., Granger, J., Stumpf, C., et al. (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.

Reijenga, J. C., Martens, J. H., Giuliani, A., & Chiari, M. (2002). Pherogram normalization in capillary electrophoresis and micellar electrokinetic chromatography analyses in cases of sample matrix-induced migration time shifts. *Journal of Chromatography B, Analytical Technologies in the Biomedical and Life Sciences, 770*, 45–51.

Reo, N. V. (2002). NMR-based metabolomics. *Drug and Chemical Toxicology, 25*, 375–382.

Robinson, M. D., De Souza, D. P., Keen, W. W., et al. (2007). A dynamic programming approach for the alignment of signal peaks in multiple gas chromatography-mass spectrometry experiments. *BMC Bioinformatics, 8*, 419.

Ruckstuhl, A. F., Jacobson, M. P., Field, R. W., & Dodd, J. A. (2001). Baseline subtraction using robust local regression estimation. *Journal of Quantitative Spectroscopy and Radiative Transfer, 68*, 179–193.

Saito, N., Robert, M., Kitamura, S., et al. (2006). Metabolomics approach for enzyme discovery. *Journal of Proteome Research, 5*, 1979–1987.

Shaffer, J. P. (1995). Multiple hypothesis testing. *Annual Review of Psychology, 46*, 561–584.

Simo, C., Moreno-Arribas, M. V., & Cifuentes, A. (2008). Ion-trap versus time-of-flight mass spectrometry coupled to capillary electrophoresis to analyze biogenic amines in wine. *Journal of Chromatography. A, 1195*, 150–156.

Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R., & Siuzdak, G. (2006). XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical Chemistry, 78*, 779–787.

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

Soga, T., Ohashi, Y., Ueno, Y., et al. (2003). Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. *Journal of Proteome Research, 2*, 488–494.

Styczynski, M. P., Moxley, J. F., Tong, L. V., et al. (2007). Systematic identification of conserved metabolites in GC/MS data for metabolomics and biomarker discovery. *Analytical Chemistry, 79*, 966–973.

Tautenhahn, R., Bottcher, C., & Neumann, S. (2008). Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics, 9*, 504.

Vivo-Truyols, G., Torres-Lapasio, J. R., van Nederkassel, A. M., Vander Heyden, Y., & Massart, D. L. (2005a). Automatic program for peak detection and deconvolution of multi-overlapped chromatographic signals part I: Peak detection. *Journal of Chromatography. A, 1096*, 133–145.

Vivo-Truyols, G., Torres-Lapasio, J. R., van Nederkassel, A. M., Vander Heyden, Y., & Massart, D. L. (2005b). Automatic program for peak detection and deconvolution of multi-overlapped chromatographic signals part II: Peak model and deconvolution algorithms. *Journal of Chromatography. A, 1096*, 146–155.

Wallace, W. E., Kearsley, A. J., & Guttman, C. M. (2004). An operator-independent approach to mass spectral peak identification and integration. *Analytical Chemistry, 76*, 2446–2452.

Wang, T., Shao, K., Chu, Q., et al. (2009). Automics: An integrated platform for NMR-based metabonomics spectral processing and data analysis. *BMC Bioinformatics, 10*, 83.

Wee, A., Grayden, D. B., Zhu, Y., Petkovic-Duran, K., & Smith, D. (2008). A continuous wavelet transform algorithm for peak detection. *Electrophoresis, 29*, 4215–4225.

Wittke, S., Fliser, D., Haubitz, M., et al. (2003). Determination of peptides and proteins in human urine with capillary electrophoresis-mass spectrometry, a suitable tool for the establishment of new diagnostic markers. *Journal of Chromatography. A, 1013*, 173–181.

Wong, J. W., Cagney, G., & Cartwright, H. M. (2005). SpecAlign–processing and alignment of mass spectra datasets. *Bioinformatics, 21*, 2088–2090.

Wu, J., & McAllister, H. (2003). Exact mass measurement on an electrospray ionization time-of-flight mass spectrometer: Error distribution and selective averaging. *Journal of Mass Spectrometry, 38*, 1043–1053.

Yoshida, S., Hashimoto, K., Tanaka-Kanai, K., Yoshimoto, H., & Kobayashi, O. (2007). Identification and characterization of amidase-homologous AMI1 genes of bottom-fermenting yeast. *Yeast, 24*, 1075–1084.

Zhao, Q., Stoyanova, R., Du, S., Sajda, P., & Brown, T. R. (2006). HiRes–a tool for comprehensive assessment and interpretation of metabolomic data. *Bioinformatics, 22*, 2562–2564.

Metabolomics (2010) 6:78–95
DOI 10.1007/s11306-009-0178-y

ORIGINAL ARTICLE

Capillary electrophoresis mass spectrometry-based saliva metabolomics identified oral, breast and pancreatic cancer-specific profiles

Masahiro Sugimoto · David T. Wong · Akiyoshi Hirayama · Tomoyoshi Soga · Masaru Tomita

Received: 8 April 2009 / Accepted: 18 August 2009 / Published online: 10 September 2009

Abstract Saliva is a readily accessible and informative biofluid, making it ideal for the early detection of a wide range of diseases including cardiovascular, renal, and autoimmune diseases, viral and bacterial infections and, importantly, cancers. Saliva-based diagnostics, particularly those based on metabolomics technology, are emerging and offer a promising clinical strategy, characterizing the association between salivary analytes and a particular disease. Here, we conducted a comprehensive metabolite analysis of saliva samples obtained from 215 individuals (69 oral, 18 pancreatic and 30 breast cancer patients, 11 periodontal disease patients and 87 healthy controls) using capillary electrophoresis time-of-flight mass spectrometry (CE-TOF-MS). We identified 57 principal metabolites that can be used to accurately predict the probability of being affected by each individual disease. Although small but significant correlations were found between the known patient characteristics and the quantified metabolites, the profiles manifested relatively higher concentrations of most of the metabolites detected in all three cancers in comparison with those in people with periodontal disease and control subjects. This suggests that cancer-specific signatures are embedded in saliva metabolites. Multiple logistic regression models yielded high area under the receiver-operating characteristic curves (AUCs) to discriminate healthy controls from each disease. The AUCs were 0.865 for oral cancer, 0.973 for breast cancer, 0.993 for pancreatic cancer, and 0.969 for periodontal diseases. The accuracy of the models was also high, with cross-validation AUCs of 0.810, 0.881, 0.994, and 0.954, respectively. Quantitative information for these 57 metabolites and their combinations enable us to predict disease susceptibility. These metabolites are promising biomarkers for medical screening.

Keywords Salivary metabolome · Capillary electrophoresis-mass spectrometry · Oral cancer · Breast cancer · Pancreatic cancer

M. Sugimoto and D. T. Wong contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s11306-009-0178-y) contains supplementary material, which is available to authorized users.

M. Sugimoto · D. T. Wong · A. Hirayama · T. Soga · M. Tomita (✉)
Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata 997-0052, Japan
e-mail: mt@sfc.keio.ac.jp

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

D. T. Wong
School of Dentistry and Dental Research Institute, University of California, Los Angeles, CA 90095-1668, USA

1 Introduction

Saliva is an important biological fluid that provides various functions, including lubrication for speech, digestion of food, and protection from microorganisms. It is produced by multiple salivary glands; particularly the three major salivary glands parotid, submandibular and sublingual, and several minor glands. Saliva is comprised of 99% water with minerals, mucus, electrolytes, nucleic acids and proteins such as enzymes, enzyme inhibitors, growth factors, cytokines, immunoglobulins, and other glycoproteins (de Almeida Pdel et al. 2008). Saliva is a filtration of blood,

reflecting the physiological conditions of the body; thus it could be used to monitor clinical status and predict systemic diseases. Compared with blood, saliva offers distinct advantages for diagnostic or research purposes; its collection is cost-effective, safe, easy and non-invasive. Indeed, many of the characteristics of bodily fluids, such as blood and urine, are applicable to saliva including diurnal variation and the presence of diverse diagnostic analytes.

Cancer is a leading cause of death and oral cancer annually affects more than 400,000 individuals worldwide. Despite advances in treatment, the overall 5-year survival rate of patients with oral cancer is approximately 50% and has not improved over the past 30 years (Epstein et al. 2002; Mao et al. 2004). The mortality rate associated with oral cancer is particularly high because it is routinely discovered late, commonly after metastasis to the lymph nodes or neck has already occurred. Worldwide, more than 200,000 patients with pancreatic cancer are registered annually, and 98% of the patients die of the disease (Parkin et al. 2005). The high mortality rate from this cancer is thought to be due to a lack of adequate systemic therapies and the high rate of metastasis at the time of diagnosis. Therefore, novel diagnostic tests are urgently needed to detect these cancers at the premalignant stage.

Studies using molecular-based biomarkers in blood or urine to detect the progress of malignant tumors have mainly focused on altered DNA methylation or mutation, or on changes in the RNA or protein levels (Sidransky 2002). In addition, several molecular biomarker candidates have been identified by analyzing the transcriptome or proteome of saliva (Hu et al. 2007, 2008; Zimmermann and Wong 2008). However, sufficiently sensitive and reproducible saliva-based diagnostic methods are not yet available. In addition, conventional tumor markers, such as serum cancer antigen 19-9, which is widely used in the diagnosis of pancreatic cancer, are known to have less specificity for particular lesions. With the exception of breast cancer (Streckfus et al. 2008), few studies have used saliva to detect tumors remote from the oral cavity. Tumor markers that can discriminate individual cancer-specific differences and which are sensitive are required for clinical applications.

Metabolomics, the measurement of all intracellular metabolites, has become a powerful new tool to gain insight into cellular function. So far, several metabolomic approaches have been reported (Aharoni et al. 2002; Fiehn et al. 2000; Plumb et al. 2003; Reo 2002). In this marriage of methodologies, CE offers rapid analysis and efficient resolution, and MS provides excellent selectivity and sensitivity (Soga et al. 2006). A number of clinical applications of CE-MS exploring urinary or serum proteomics biomarkers, were developed to detect and identify the charged peptide content, which demonstrates their potential to assess the profiles of small molecules in biofluids (Fliser et al. 2005; Kolch et al. 2005; Metzger et al. 2009; Schiffer et al. 2006, 2008; Zurbig and Mischak 2008). Although diverse saliva analyses with CE have been proposed (Lloyd 2008), salivary metabolomic analysis to determine cancer-specific profiles for early cancer detection has not yet been conducted. In this study, we, for the first time, obtained and compared comprehensive salivary metabolic profiles of patients with oral, breast or pancreatic cancer, or periodontal disease, and healthy controls. We then identified individual cancer-specific markers with high discriminative ability, demonstrating the potential use of salivary metabolomics in cancer diagnosis.

2 Materials and methods

2.1 Patient selection

This study was approved by the UCLA Institutional Review Board. Patients with oral, breast or pancreatic cancer or periodontal disease and the healthy controls were recruited at the UCLA Medical Center. All patients had recently been diagnosed with primary disease and were without metastasis; none had received any prior treatment in the form of chemotherapy, radiotherapy, surgery or alternative therapy. No subjects had a history of prior malignancy, immunodeficiency, autoimmune disorders, hepatitis or HIV infection. Written, informed consent was obtained from all patients and from volunteers who agreed to serve as saliva donors.

2.2 Sample collection and sample preparation

The subjects were asked to refrain from eating, drinking, smoking or using oral hygiene products for at least 1 h prior to saliva collection. The subjects rinsed their mouth with water and, 5 min later, they were instructed to spit into 50-cc Falcon tubes, which were placed in a Styrofoam cup filled with crushed ice. The subjects were reminded not to cough up mucus. Five milliliters of unstimulated saliva could usually be collected in 5–10 min. Saliva collection was performed in a private room. The saliva samples were centrifuged at $2600\times g$ for 15 min at 4°C and spun for another 20 min for incomplete separation. Equal amounts of supernatant were transferred to two fresh tubes and the samples were processed and frozen within 30 min. The protocols used for sample collection are described in more detail elsewhere (Li et al. 2004).

Saliva fluid samples were obtained from patients with oral ($n = 69$), breast ($n = 30$) and pancreatic cancer ($n = 18$), patients with periodontal diseases ($n = 11$) and healthy controls ($n = 87$). The race, ethnicity, sex and age

Table 1 Subject characteristics

Groups	Control	Oral cancer	Breast cancer	Pancreatic cancer	Periodontal diseases
Age					
Min–Max (median)	20–75 (43)	34–87 (59.5)	29–77 (57)	11–87 (67)	23–76 (60)
Missing	2	5	10	2	2
Sex					
Male	42	41	N/A		
Female	27	23			
Missing	18	5	30	18	11
Race or ethnic group					
Total	87	69	30	18	11
Caucasian	37	41	N/A		
Asia	15	5			
African-American	12	4			
Hispanic	5	5			
Missing	18	14	30	18	11

N/A not available

of the subjects are summarized in Table 1. Except for age, clinical parameters were not collected for the non-oral cancer groups.

Frozen saliva was thawed and dissolved at room temperature, and 27 μl of each sample (69 patients with oral cancer and 70 healthy control samples) were added to a 1.5-ml Eppendorf tube, to which 3 μl of water containing 2 mM methionine sulfone and 2 mM 3-aminopyrrolidine as internal standards was added and mixed well. Similarly, individual thawed saliva samples (24 μl) from patients with breast or pancreatic cancer, and patients with periodontal disease and 17 healthy controls were admixed with 6 μl water containing internal standards (1 mM each of methionine sulfone and 3-aminopyrrolidine). These internal standards were selected because they were not included in the human endogenetic metabolites. Furthermore, they migrated to the center of the metabolite distribution, which was used to confirm the quality of the alignment results. Even though a unified dilution was preferred for the preparation of all samples, a greater dilution ratio was required for the control, breast, pancreatic cancer, and periodontal disease samples because of their high electrolyte content, which decreases the electrical current during the measurement.

2.3 Metabolite standards, instrumentation, and CE-TOF-MS conditions

The metabolite standards, instrumentation and CE-TOF-MS condition were used in this study as previously described (Soga et al. 2006), with slight modifications in the lock mass system setting. All chemical standards were of analytical or reagent grade and were obtained from commercial sources. They were dissolved in Milli-Q water (Millipore, Bedford, MA, USA), 0.1 mol/l HCl or 0.1 mol/l NaOH to obtain 1, 10 or 100 mmol/l stock solutions. The working solution was prepared prior to use by diluting with Milli-Q water to the appropriate concentration.

All CE-MS experiments were performed using an Agilent CE capillary electrophoresis system (Agilent Technologies, Waldbronn, Germany), an Agilent G3250AA LC/MSD TOF system (Agilent Technologies, Palo Alto, CA, USA), an Agilent 1100 series binary HPLC pump, and the G1603A Agilent CE-MS adapter and G1607A Agilent CE-ESI-MS sprayer kit. System control and data acquisition were done with G2201AA Agilent Chemstation software for CE and Analyst QS software for TOF-MS (ver. 1.1).

All samples were measured in single mode (see below); separation was done in fused-silica capillaries (50 μm i.d. × 100 cm total length) filled with 1 M formic acid as the background electrolyte. 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 kept below 5°C using an external thermostatic cooler. The sheath liquid, comprising methanol/water (50% v/v) and 0.5 μM reserpine, was delivered at 10 μl/min. ESI-TOF-MS was conducted in the positive ion mode. The capillary voltage was set at 4 kV; the flow rate of nitrogen gas (heater temperature 300°C) was set at 10 psig. In TOF-MS, the fragmentor, skimmer and OCT RFV voltage were set at 75, 50 and 125 V, respectively. In the present study, we used a methanol dimer adduct ion ($[2MeOH + H]^+$, *m/z* 65.059706) and hexakis phosphazene ($[M + H]^+$, *m/z*

622.028963) to provide the lock mass for exact mass measurements. Exact mass data were acquired at the rate of 1.5 cycles/s over a 50–1000 *m/z* range.

2.4 Processing of CE-TOF-MS data

Raw data were analyzed with our proprietary software called MasterHands, which has already been used in several CE-TOF-MS-based profiling studies (Hirayama et al. 2009; Minami et al. 2009; Saito et al. 2009). The data analysis workflow starting with the raw data included noise-filtering, baseline correction, peak detection and integration of the peak area from sliced electropherograms (the width of each electropherogram was 0.02 *m/z*). Such functions are commonly used by data processing software such as MassHunter from Agilent Technologies, or XCMS (Smith et al. 2006) for liquid chromatography-MS or gas chromatography-MS data. The accurate *m/z* value for each peak detected within the time domain was calculated with Gaussian curve-fitting to the mass spectrum on the *m/z* domain peak. The alignment of peaks in multiple measurements was done by dynamic programming (DP)-based techniques (Baran et al. 2006; Soga et al. 2006) with slight modifications. The method picked up a few representative peaks using the Douglas-Peucker algorithm (Wallace et al. 2004) from unit *m/z* electropherograms, found corresponding peaks across multiple samples by DP, and optimized the numerical parameters of the normalization function for CE-migration (Reijenga et al. 2002). Instead of representative peaks, we used the detected peaks with accurate *m/z* values and regarded the peaks whose *m/z* difference was less than 20 ppm as ones that were derived from the same electropherograms.

All peak areas were divided by the area of the internal standard (relative area) to normalize the signal intensities, and to avoid injection-volume bias and mass-spectrometry detector sensitivity bias among multiple measurements. Undetected peaks with a threshold signal-to-noise ratio of 2 were given a peak area of 0. The relative areas of the 17 healthy control samples and of the pancreatic and breast cancer, and the periodontal disease samples were multiplied by 1.25/1.1 to standardize the sample concentration.

The peaks derived from salt and neutral molecules were found in the first and the last few minutes, respectively. Then, isotopic compounds, ringing, spikes and fragment and adduct ions were eliminated and the peak data sets were compared across the sample profiles and aligned according to *m/z* and migration time. Although all of the metabolites were quantified separately, the sum of the quantified values of leucine and isoleucine were counted as a single marker owing to the low separation of these peaks. Peaks showing $P < 0.05$ in the non-parametric, multiple comparison Steel–Dwass test, between the controls and at least one disease cohort were selected as candidate markers.

2.5 Metabolite identification

The peaks were identified based on the matched *m/z* values and normalized migration times of the corresponding standard compounds if available. Of the peaks that did not match with any standard compounds, the concomitant peaks, such as isotopic peaks and fragment peaks, were removed based on the difference in *m/z* values and the normalized migration time of the two peaks with an error tolerance of 20 ppm and 0.01 min to yield only the peaks, or referred to as components, which might be derived from metabolites (Brown et al. 2009). Although CE-TOF-MS provides accurate molecular mass at the milli *m/z* level, the *m/z* alone is seldom successful to identify the metabolite (Kind and Fiehn 2006, 2007). Therefore, we used their *m/z* values and the migration times predicted by the Artificial Neural Networks (ANNs) (Sugimoto et al. 2005) to identify the metabolite. Briefly, the ANN model was first trained using the measured migration times of standard compounds and molecular descriptors with the net charge calculated from the pKa values. The trained ANN model then predicted the migration times of the candidate metabolites. Here, we used compounds available from the Kyoto Encyclopedia of Gene and Genomics (KEGG) database (Goto et al. 2002) and the Human Metabolome Database (HMDB) (Wishart et al. 2007) as candidates. The composition formulae obtained using the measured mass spectrometry and the matched candidates were confirmed by their isotope distribution patterns.

2.6 Statistical analysis

To evaluate the ability of the detected peaks to discriminate diseases, we conducted an unsupervised method, principal component analysis (PCA). The same analyses were also conducted to discriminate only between controls and oral samples between males and females, and between race and ethnic groups. The analyses were not performed for the other patient groups due to the unavailability of clinical parameters. Supervised classification techniques, such as partial least squares-discriminant analysis (Jonsson et al. 2005; Michell et al. 2008; Woo et al. 2009), support vector machine (SVM) (Mahadevan et al. 2008) and multiple logistic regression (MLR), are commonly used to separate subjects and to identify important features for the separation. Here, we developed independent MLR models to discriminate healthy individuals and each disease cohort using a stepwise variable selection method (backward procedure to eliminate non-predictive peaks with a threshold of $P > 0.10$) to construct the predictive models.

The models were trained with the complete dataset and we evaluated their versatilities by tenfold cross-validation (CV). The data were randomly separated into training sets and remaining data and this process was repeated ten times for all of the values selected in the training set. The non-parametric Mann–Whitney test was used to compare two groups, e.g. comparison of metabolites in males and females.

Statistical analyses using the Steel–Dwass test were performed using the R package with the Design, Hmisc, and Lexis libraries (available at http://lib.stat.cmu.edu/R/CRAN/). Statistical analyses using the Mann–Whitney test and the heat maps were generated with TM4 software (Saeed et al. 2003). The CV data were generated using WEKA (Witten and Frank 2005). The PCA and MLR models were developed using JMP Version 7 (SAS Institute Inc., Cary, NC, USA, 1989–2007; http://www.jmp.com/software/jmp.shtml).

3 Results and discussion

3.1 Statistical results of discriminative metabolites

On average, CE-TOF-MS detected 3041 peaks (minimum 1585, maximum 8400, standard deviation (SD) 1137) in each saliva sample. After removing the concomitantly observed peaks such as the isotopic and fragment peaks, and noise peaks including spike and ringing peaks, an average of 90 peaks were derived from the metabolites (minimum 48, maximum 128, SD 15). The standard deviation of the relative peak areas of the metabolite-derived peaks was 1.14 (no unit), and the SDs of the migration times before and after the time normalization procedure were 1.75 min and 3.02×10^{-3} min, respectively. Of the remaining peaks, we identified 57 metabolites that were significantly different between the patients and healthy controls ($P < 0.05$; Steel–Dwass test).

The marker pool used to discriminate between individuals with oral cancer and healthy controls revealed 28 metabolites; namely pyrroline hydroxycarboxylic acid, leucine plus isoleucine, choline, tryptophan, valine, threonine, histidine, pipecolic acid, glutamic acid, carnitine, alanine, piperidine, taurine, and two other metabolites with a significance of $P < 0.001$ (Steel–Dwass test); piperidine, alpha-aminobutyric acid, phenylalanine and an additional metabolite with a significance of $P < 0.01$ (Steel–Dwass test); and betaine, serine, tyrosine, glutamine, beta-alanine, cadaverine, and two other metabolite with a significance of $P < 0.05$ (Steel–Dwass test). The overlaid electropherograms of these CE-TOF-MS peaks with a 2-dimensional map (migration time and *m/z*) visualizing the difference in intensity between the averaged control and oral cancer samples are shown in Fig. 1. The vertical smear lines in the first few minutes (5–7 min) and those at a later time (at 19 min) were derived from salt ions and neutral molecules, respectively, and most of the peaks derived from charged metabolites were distributed between these times. Using a similar strategy, we identified 28 metabolites for breast cancer, 48 for pancreatic cancer and 27 for periodontal disease ($P < 0.05$; Steel–Dwass test) as biomarker candidates. The detected markers and the statistical results are listed in Table 2; dot plots of the quantified peak areas are shown in Fig. 2 and Supplementary Fig. S1. Although, several metabolites in the dot plots achieved a statistically significant difference, individual metabolites could not separate any two groups with high sensitivity and specificity. The score plots of the PCA analyses for all individuals are shown in Fig. 3 and in Supplementary Fig. S2. Although the PCA developed using the metabolite profiles of all subjects showed no unequivocal group-specific clusters, PCAs developed individually for the control and each disease group showed partial discriminative separation of the subjects, which might be attributed to the reduced complexity of the given datasets, or the extinction in the overlap between the distribution of the score plots for all disease groups.

The MLR model developed for oral cancer yielded a high AUC (0.865), and the trained models also showed high separation ability in the CV (AUC = 0.810). The receiver operating characteristic (ROC) curves and selected parameters of the MLR models for each disease are shown in Fig. 4 and Supplementary Table S1, respectively. The MLR models for pancreatic cancer and periodontal disease yielded high AUCs in the CV test (0.944 and 0.954, respectively), using only five and two metabolic markers, respectively; while oral and breast cancers (0.810 and 0.881, respectively) used 9 and 14 metabolites, respectively, with lower AUCs. On the metabolite heat map (Fig. 5), the control group and the periodontal disease group were relatively lower and the pancreatic cancer group tended to be homologically higher, while the oral and breast cancers exhibited more diverse profiles compared with the other groups. This suggests that our MLR models for oral and breast cancer require additional parameters for accurate classification. The heterogeneous nature of oral cancers, including oral squamous cell carcinoma (OSCC), oropharyngeal, tongue and neck cancer, may produce different profiles; this diminishes the discriminative capability of a single classification model. The diverse profiles associated with breast cancer may result in a similar situation because breast cancer comprises structurally differing types according to the expression of hormone receptors such as estrogen and progesterone, and is affected by clinical parameters, such as the patient's age or menopause status. Three metabolites, taurine, piperidine, and a peak at

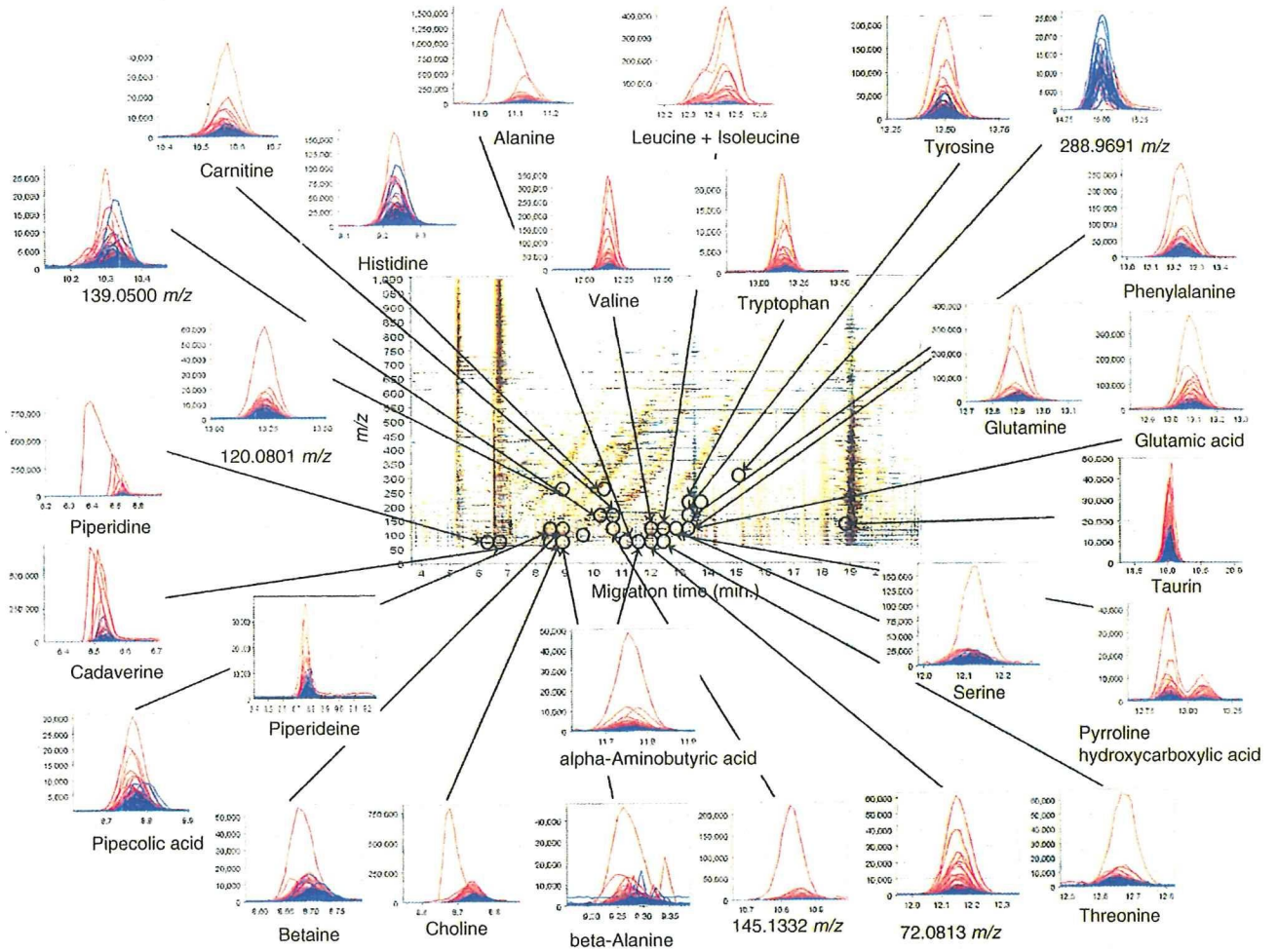

Fig. 1 A summary of the different metabolome profiles of cations obtained from CE-TOF-MS analyses of salivary metabolites from control (n = 87) and oral cancer samples (n = 69). The X and Y axes represent the migration time and the *m/z* value, respectively. The color density reflects the difference in intensity between the averaged control and oral cancer samples. *Black circles* indicate peaks that are significantly different between healthy control and oral cancer samples ($P < 0.05$; Steel–Dwass test). The small linked figures include overlaid electropherograms of control (*blue*) and oral cancer samples (*red*)

120.0801 *m/z*, were oral cancer-specific markers (different from all of the other groups at $P < 0.05$; Steel–Dwass test) and eight metabolites (leucine with isoleucine, tryptophan, valine, glutamic acid, phenylalanine, glutamine, and aspartic acid) were pancreatic cancer-specific markers. Although several metabolites in breast cancer patients yielded a statistically significant difference between breast cancer and healthy controls, including taurine and lysine ($P < 0.001$ for both; Steel–Dwass test), there were no differences in metabolites between breast cancer and other cancer, and they were not unique for breast cancer.

3.2 Comparison of the obtained metabolites with previous studies

Of the metabolite profiles obtained, the annotated metabolites included carnitines (betaine, choline, carnitine, glycerophosphocholine), polyamines (cadaverine and putrescine), a purine (hypoxanthine), amino alcohols (ethanolamine), aliphatic and aromatic amine (trimethylamine), and amino acids (the others), in accordance with the defined chemical class category in HMDB. Because each MLR model developed to discriminate between control and patient groups reached high accuracy by incorporating quantified multiple metabolites, the quantitative associations between the multiple metabolites and the individual markers are important. Changes in the individual metabolites were generally consistent with those of earlier studies. For example, polyamines are correlated with cell growth and proliferation (Casero and Marton 2007; Gerner and Meyskens 2004; Tabor and Tabor 1984), and with tumor growth in oral cancer (Dimery et al. 1987), while putrescine is used to monitor the effect of chemotherapy on oral cancer cells (Okamura et al. 2007). The serum