Bioinformatics, 2021, 1–7
https://doi.org/10.1093/bioinformatics/btab854
Advance Access Publication Date: 24 December 2021
Original Paper

OXFORD

Systems biology

BATL: Bayesian annotations for targeted lipidomics

Justin G. Chitpin[1,2,3,4,5], Anuradha Surendra[6], Thao T. Nguyen [3,4,5,7], Graeme P. Taylor[3,4,5], Hongbin Xu[3,4,5], Irina Alecu[3,4,5], Roberto Ortega[8], Julianna J. Tomlinson[9,10], Angela M. Crawley[2,5], Michaeline McGuinty[2], Michael G. Schlossmacher[9,10], Rachel Saunders-Pullman[8], Miroslava Cuperlovic-Culf[5,6,*], Steffany A. L. Bennett [2,3,4,5,7,9,10,*] and Theodore J. Perkins [1,2,4,5,*]

[1]Regenerative Medicine Program, Ottawa, ON K1H 8L6, Canada, [2]Ottawa Hospital Research Institute, Ottawa, ON K1H 8L6, Canada, [3]Neural Regeneration Laboratory and India Taylor Lipidomics Research Platform, University of Ottawa Brain and Mind Research Institute, Ottawa, ON K1H 8M5, Canada, [4]Ottawa Institute of Systems Biology, University of Ottawa, Ottawa, ON K1H 8M5, Canada, [5]Department of Biochemistry, Microbiology and Immunology, University of Ottawa, Ottawa, ON K1H 8M5, Canada, [6]Digital Technologies Research Center, National Research Council, Ottawa, ON K1A 0R6, Canada, [7]Department of Chemistry and Biomolecular Sciences, Centre for Catalysis Research and Innovation, University of Ottawa, Ottawa, ON K1N 6N5, Canada, [8]Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, NY 10003, USA, [9]Department of Cellular and Molecular Medicine, University of Ottawa, Ottawa, ON K1H 8M5, Canada and [10]Neuroscience Program, Ottawa Hospital Research Institute, Ottawa, ON K1H 8L6, Canada

*To whom correspondence should be addressed.

Associate Editor: Olga Vitek

Received on March 18, 2021; revised on November 25, 2021; editorial decision on December 6, 2021; accepted on December 20, 2021

Abstract

Motivation: Bioinformatic tools capable of annotating, rapidly and reproducibly, large, targeted lipidomic datasets are limited. Specifically, few programs enable high-throughput peak assessment of liquid chromatography–electrospray ionization tandem mass spectrometry data acquired in either selected or multiple reaction monitoring modes.

Results: We present here Bayesian Annotations for Targeted Lipidomics, a Gaussian naïve Bayes classifier for targeted lipidomics that annotates peak identities according to eight features related to retention time, intensity, and peak shape. Lipid identification is achieved by modeling distributions of these eight input features across biological conditions and maximizing the joint posterior probabilities of all peak identities at a given transition. When applied to sphingolipid and glycerophosphocholine selected reaction monitoring datasets, we demonstrate over 95% of all peaks are rapidly and correctly identified.

Availability and implementation: BATL software is freely accessible online at https://complimet.ca/batl/ and is compatible with Safari, Firefox, Chrome and Edge.

Contact: miroslava.cuperlovic-culf@nrc-cnrc.gc.ca; tperkins@ohri.ca; sbennet@uottawa.ca

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Targeted lipidomics employs high performance liquid chromatography - electrospray ionization tandem mass spectrometry (LC-ESI-MS/MS)]. Using selected and multiple reaction monitoring (SRM and MRM) modes, pairs of precursor and product ions (transitions), are monitored to quantify lipids of interest. Targeted transition lists are constructed based on prior knowledge of lipid fragmentation pathways as reported in literature (e.g., Murphy and Axelsen, 2011), obtained through exploration of MS/MS spectra for untargeted lipidomic analyses, and/or by performing other semi-targeted, unbiased lipid approaches, such as prior assessment of a given matrix in precursor ion scan mode (Sartain *et al.*, 2011). Once precursor and product ion pairs are identified, parking on a single product ion effectively reduces interfering signals generated by isobaric lipids from other classes, enabling SRM and MRM modes to excel at high-throughput quantitation of both high- and low-abundance species (Bowden *et al.*, 2017). Conversely, high-resolution mass spectrometry-based targeted lipidomics is done by parallel reaction monitoring (PRM). This method exploits instrument setups that combine a quadrupole, a high-energy collisional dissociation (HCD) cell, and a high-resolution mass analyzer, such as an Orbitrap or time-of-flight (ToF). The targeted precursor ion is isolated by the quadrupole, fragmented by the HCD, and product ions are simultaneously monitored and quantified by the high-resolution mass analyzer (Gallien *et al.*, 2014; Peterson *et al.*, 2012). Parallel monitoring of all product ions eliminates the needs for *a priori* targeted transition lists. Because the precursor ion is selected by the low-resolution quadrupole, this targeted approach remains subject to isobaric contamination and thus requires additional bioinformatic tools to confirm peak identities. Together, these approaches have been used to successfully map fluid and cell-specific lipidomes (Quehenberger *et al.*, 2010; Sartain *et al.*, 2011; Slatter *et al.*, 2016), reveal lipidomic disruptions across biological conditions (Alecu and Bennett, 2019; Wang *et al.*, 2018) and predict changes in lipid metabolism associated with disease progression (Alshehry *et al.*, 2016; Blasco *et al.*, 2017; Granger *et al.*, 2019).

Despite the power of SRM, MRM and PRM approaches to quantify lipid analytes, it remains challenging to annotate lipid identities rapidly and reproducibly across large numbers of MS chromatograms, notably when collected from different organisms or matrices using different mass spectrometry methodologies. While the concept of targeting individual lipid species in SRM, MRM and PRM modes appears straightforward, ensuring peaks are correctly assigned is labor-intensive and not trivial, as exemplified in Figure 1. Multiple isobars, isomers and isotopologues, sharing the same product ion, can elute in close proximity to the targeted lipid. Moreover, routine variations in chromatography can cause retention time shifts that align isobars or isotopologues to the species of interest in different MS runs (Smith, 2015). When multiple peaks are detected at a given transition, careful judgment is required to discriminate between lipid targets. These problems are magnified when researchers seek to match corresponding peaks and identify unique lipid species (i) across lipidomes of different organisms or (ii) within different matrices where peak features may change drastically.

Few programs have been developed to address the difficulties of SRM, MRM and PRM peak identification. MRMPROBS is the most well-recognized SRM/MRM peak identification program, using a multivariate logistic regression classifier to assign annotations from a library of lipid species (Tsugawa *et al.*, 2013). The program computes the posterior probability of a peak belonging to a lipid in the training set conditioned on five peak features describing lipid retention time, intensity, and shape. However, these features are reduced to only retention time when classifying SRM peaks. Two further program restrictions of MRMPROS lie in the fact that the number of lipid identities in the training set cannot exceed the number of transitions acquired in the raw MS data and that the compound names in the training set must match the lipid target names in the SRM or MRM method. These restrictions become problematic when new lipid species are discovered in different biological matrices or conditions and users seek to match corresponding lipids across these datasets. mProphet uses a conceptually similar linear discriminant analysis method to identify peptides from SRM and MRM data but further includes addition of decoy transitions that act as negative controls to parameterize the null model and derive false discovery rates. These additions improve identification confidence (Reiter *et al.*, 2011). However, identifying a sufficient number of decoy transitions universally applicable to all lipidomes has proven difficult. Vendor-specific programs, such as MultiQuant (SCIEX), MassHunter (Agilent), MassLynx (Waters) and LipidSearch (Thermo Fisher Scientific) are peak-picking algorithms where users can specify retention time windows and compute retention time ratios based on predetermined internal standards to assist in peak identification. MultiQuant, MassHunter and MassLynx do not, however, assign peak identities. LipidSearch (Thermo Fisher Scientific) assigns peak identity to the closest matching retention time within a user-defined retention time window to a proprietary internal library. Similarly, Lipidyzer, using the Lipidomics Workflow Manager program (SCIEX), assigns lipid identities from differential mobility spectrometry (DMS) data acquired by direct infusion SRM mode (Ubhi *et al.*, 2016). Additionally, Lipidyzer was designed to analyze data acquired specifically from SCIEX QTRAP 5500/5600 mass spectrometers with a SelexION DMS cell. However, even when using these software packages, manual curation remains the most common peak identification method when extracted ion chromatograms (XICs) do not match exactly to reference samples (Bowden *et al.*, 2018). Finally, academic programs, such as METLIN-MRM (Domingo-Almenara *et al.*, 2018), use a similar approach to LipidSearch, first aligning XIC peaks by retention time before assigning lipid identities to the closest peak within the retention time window. While these approaches excel in identifying compounds within the same condition in simple matrices, any of the common scenarios described in Figure 1 can lead to peak misidentification.

To address this problem, we applied a Bayesian annotation approach tailored to annotate targeted lipidomic datasets and present the program Bayesian Annotations for Targeted Lipidomics (BATL), which overcomes many of the limitations of the manual or template-based curation approaches. BATL is an R package, implemented through an online GUI at CompLiMet: Computational Lipidomics and Metabolomics https://complimet.ca/batl/ . The input format is based on results tables generated using MultiQuant (SCIEX) but is applicable to any targeted lipidomics data collection mode from any LC-ESI-MS/MS platform once the user formats their results tables to match the format provided. The program models lipid-specific peak features obtained from a user-curated training set using Gaussian distributions and computes the joint posterior probability of all peak identities in a given sample. BATL was developed using eight specific features, describing peak retention, intensity, and shape. he online version allows users to train on any combination of features. We show here that our approach accurately identifies over 95% of all sphingolipid and glycerophosphocholine peaks in SRM datasets analyzed across matrices and disease conditions. Thus, BATL is a useful tool for accurate, targeted lipid identification and, with online access, is easily integrated into any lipidomic pipeline.

2 Materials and methods

2.1 Overview of program

The BATL workflow is presented in Figure 2. First, a training set is constructed from user-labeled, targeted lipidomic datasets. Second, BATL uses both the training set and the specified input features to construct a naïve Bayes statistical model. Third, the model and associated metadata are exported and used by BATL to annotate peaks in query SRM, MRM, or PRM datasets. If a peak cannot be assigned to an identity present in the training set, an annotation of 'unassigned' is returned, enabling the user to assess and validate a potentially novel peak at that transition. An optional BATL function is further included, which annotates isotopes in all lipid categories as well as sphingolipid-specific artifacts (e.g., dehydrations, deglycosylations and dimers).

2.2 Naïve Bayes model

Our approach to peak identification is based on maximizing the joint posterior probability of all peak identities within each sample. Let P_1, P_2,. . ., P_m be a list of peaks within a sample described by feature vectors F_1, F_2,. . ., F_m. Each feature vector contains k features, where $F_i = \{f_{i1}, f_{i2}, \ldots, f_{ik}\}$ describes the ith peak. Each peak is detected at precursor ion m_i and product ion p_i under the same Q1 and Q3 mass analyzer tolerance δ. Let $I = \{B_1, B_2, \ldots, B_n\}$ be the set of all lipid identities, where the bth identity is detected at precursor ion n_b and product ion q_b under the Q1 and Q3 mass analyzer tolerance δ. Thus, the possible lipid identities for each peak are those detected within the machine tolerance of the lipid identity and peak transition.

$$f(P_i) = \{|B_b|\ b \leq n, |m_i - n_b| \leq 2\delta, |p_i - q_b| \leq 2\delta\}. \qquad (1)$$

To denote the assigned identity for P_i, let I_1, I_2, . . ., I_m take lipid identities drawn from $f(P_1)$, $f(P_2)$, . . ., $f(P_m)$. The posterior probability of some joint assignment of peak identities is

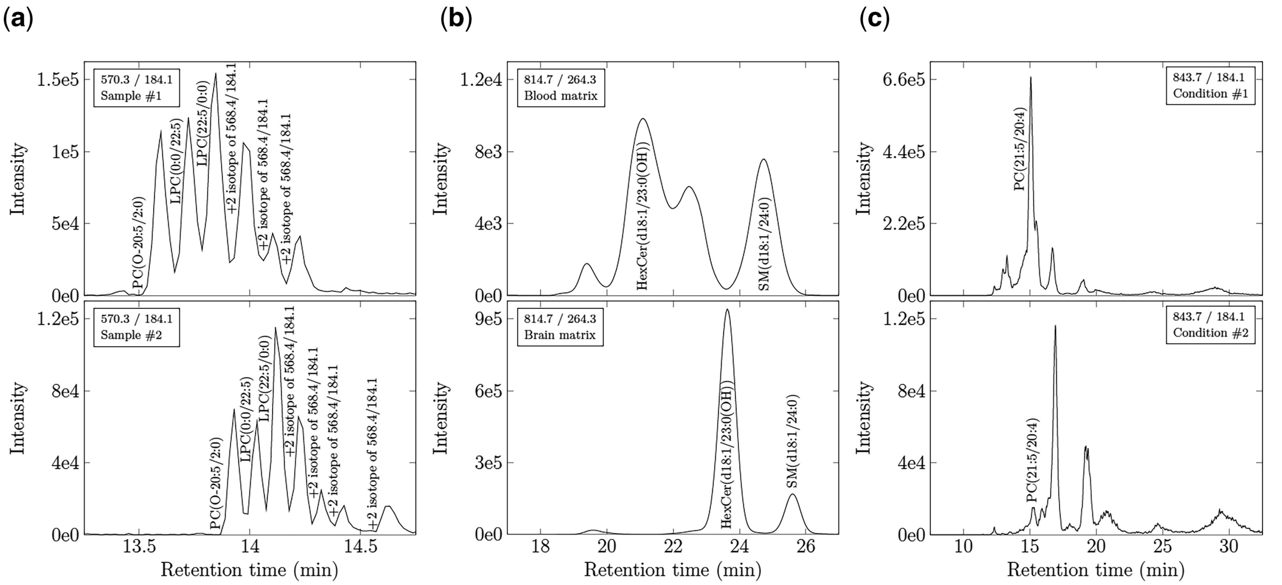

Fig. 1. Common challenges associated with SRM, MRM and PRM peak identification. (**a**) Ambiguity occurs when multiple lipid isomers, isobars, and isotopes are detected within the same matrix at a given transition, yet technical variations in flow rate, composition of the mobile phase, temperature, pH, etc., cause their retention times to vary across samples. Data represent XICs of the same matrix (murine plasma) in animals fed different diets. Note six peaks are observed in one sample at a given transition. Seven peaks are observed in a different sample shifted by 1 min. Matching retention time would not align these shifted species. (**b**) Assigning lipid identities based on peak elution order (picking the *n*th eluting peak) will also lead to misidentifications when comparing lipid species across matrices. Data represent XICs of plasma and brain (temporal cortex) lipidomes from the same animal. Note both the retention time shift and the fundamentally different number of species within each matrix. Matching by either retention time or peak elution order would confound identification. (**c**) Matching lipids based on peak intensity features is complicated by pathological changes detected in lipid metabolism. Data represent XICs of the human plasma lipidome of patients with different neurodegenerative diseases. Note the marked change in abundances between conditions that impacts on lipid identification. While algorithms exist to address each of these challenges, few are applicable to datasets wherein all differences manifest simultaneously. BATL addresses these challenges

$$\Pr(I_1, I_2, \ldots, I_m | F_1, F_2, \ldots, F_m). \tag{2}$$

This joint probability is expanded using Bayes's Theorem as in Equation (3).

$$\frac{\Pr(I_1, I_2, \ldots, I_m | F_1, F_2, \ldots, F_m)\Pr(I_1, I_2, \ldots, I_m)}{\Pr(F_1, F_2, \ldots, F_m)}. \tag{3}$$

To compute this joint probability, we make three assumptions: (i) the prior probabilities of all lipid identities are independent; (ii) the peak feature vectors are statistically independent, conditional on the identities; and (iii) the individual features within each vector are statistically independent, conditional on the peak identity. Thus, Equation (3) is simplified to the following probability.

$$\frac{\prod_{i=1}^{m}\prod_{j=1}^{k}\Pr(f_{ij}|I_i)\Pr(I_i)}{\Pr(F_1, F_2, \ldots, F_m\ )}. \tag{4}$$

The denominator is a data-dependent constant and can be ignored when comparing the probabilities of different joint assignments. The log posterior probability of a joint assignment is thus proportional to

$$\sum_{i=1}^{m} w_{ib}, \tag{5}$$

where weight w_{ib} is the unnormalized, log posterior probability of assigning peak i to lipid identity B_b.

$$w_{ib} = \log\prod_{j=1}^{k}\Pr(f_{ij}|I_i)\Pr(I_i). \tag{6}$$

The joint assignment of lipid identities is determined by the classifier decision rule. To optimize BATL, we tested three classifier rules. First, we assessed choosing lipid identities that maximize w_{ib} following the maximum *a posteriori* (MAP) decision rule typical of naïve Bayes classifiers. We found that a disadvantage of this decision rule was that lipid identities were assigned independently. Although peaks detected in the same sample clearly corresponded to unique lipid identities, the MAP decision rule could assign an identity more than once per sample (see Section 3). To address this problem, we evaluated a constrained MAP decision rule wherein lipid identities were assigned by the ranked order of their log posteriors, such that no lipid identity was assigned more than once per sample. We found that this method was not guaranteed to maximize Equation (5) and thus did not yield the optimal assignment of lipid identities (see Section 3). Third, we resolved the shortcomings of MAP and constrained MAP with the maximum weighted bipartite matching (MWBM) decision rule, which considers the simultaneous identification of all peak identities within a sample under the naïve Bayes model.

For every sample transition, a bipartite graph was constructed where the vertices represent peaks P_i and their possible lipid identities $f(P_i)$ with corresponding edges weighted by w_{ib}. The optimal set of matching peaks and lipid identities was then solved by MWBM, thereby maximizing Equation (5) while ensuring a unique lipid identity was assigned to each peak detected per sample. Finally, under certain conditions, the true identity of a peak would be absent from set I, representing a novel lipid species detected in the sample of interest. To account for this possibility, every peak was matched to an 'unassigned' identity U in addition to $f(P_i)$. The weights w_{iu} were found to be specific to each transition and estimated by cross validation (see Section 3).

2.3 Training the model

Let $D = \{D_1, D_2, \ldots, D_p\}$ denote the labeled training set containing the instances $D_o = (F_o, B_o)$ for samples $n = 1, \ldots, N$. F_o is the feature vector of length k, where $F_o = (f_{o1}, f_{o2}, \ldots, f_{ok})$, and B_o is the true lipid identity. Each lipid identity in the training set contains a unique sample index because the same lipid can only be detected once per sample. The prior probability of each lipid identity is computed by maximum likelihood estimation

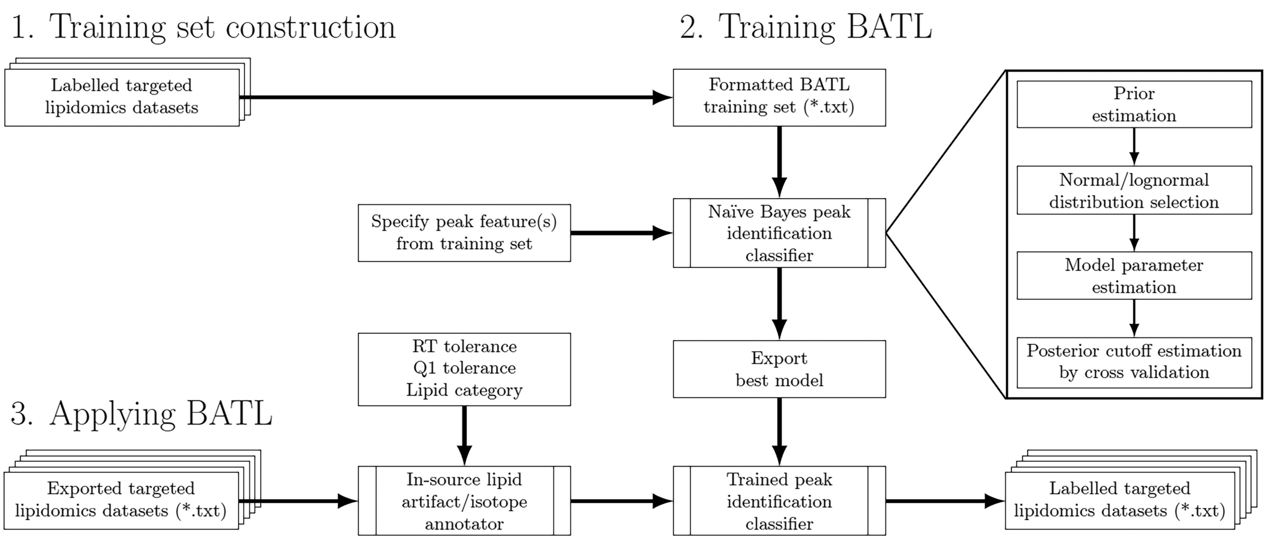
1. Training set construction
Labelled targeted lipidomics datasets
2. Training BATL
Formatted BATL training set (*.txt)
Prior estimation
Normal/lognormal distribution selection
Model parameter estimation
Posterior cutoff estimation by cross validation
Specify peak feature(s) from training set
Naïve Bayes peak identification classifier
RT tolerance
Q1 tolerance
Lipid category
Export best model
3. Applying BATL
Exported targeted lipidomics datasets (*.txt)
In-source lipid artifact/isotope annotator
Trained peak identification classifier
Labelled targeted lipidomics datasets (*.txt)

Fig. 2. Schematic of the BATL lipid identification workflow. BATL follows three steps: (i) users are asked to identify training datasets for which they have unambiguous knowledge of peak identities. (ii) These datasets are used to train BATL, constructing a naïve Bayes statistical model based on the peak features users select. (iii) The model and associated metadata are used by the BATL algorithm to annotate peaks in subsequent query SRM, MRM or PRM datasets

$$\Pr(B_o) = \frac{N_{B_o}}{N}, \tag{7}$$

where N_{Bo} is the number of lipid identities B_o in the training set. The feature likelihoods are computed using either a normal or lognormal distribution with parameters μ_o and σ^2 estimated using the sample mean and variance from the training set. The choice of distribution is assessed using a KS-test for normality and lognormality of feature j for lipid identity B_o.

$$\Pr(f_{ij}|I_i) = \begin{cases} logN\left(f_{ij}|\mu_{I_{ij}}, \sigma_{I_{ij}}\right), & if\ N_j^{\log} < N_j \\ N\left(f_{ij}|\mu_{I_{ij}}, \sigma_{I_{ij}}\right), & \text{otherwise} \end{cases}. \tag{8}$$

N_j and $N_j^{\log}$ are the number of lipid identities failing the KS-test for normality and lognormality, respectively, for feature j at a P-value threshold of 0.05.

Lastly, the unassigned identity weights w_{iu} are estimated per transition by k-fold cross validation. Looping over the $k-1$ folds of the training set, the naïve Bayes model is trained and unnormalized log posteriors are computed from the testing fold. Across all k iterations, the weights w_{iu} for each transition are set to the minimum unnormalized posterior of a correct peak assignment.

2.4 Datasets

To train and test BATL, we curated and labeled sphingolipid and glycerophosphocholine datasets composed of 1008 MS spectra generated at the India Taylor Neurolipidomics Research Platform, University of Ottawa. To ensure all of the challenges in MRM, SRM, and PRM identification outlined in Figure 1 were recapitulated in these datasets, we used: (i) a population-based study of circulating lipids in human plasma of cognitively normal controls, and patients suffering from Alzheimer's disease, mild cognitive impairment, dementia with Lewy bodies, or Parkinson's disease ($n=319$ sphingolipid analyses; $n=319$ glycerophosphocholine analyses), (ii) a genotype and intervention comparison study of lipid metabolism in the temporal cortex, hippocampus, and plasma of wild-type and N5 TgCRND8 mice, a sexually dimorphic mouse model of Alzheimer's disease (Granger, 2016) ($n=121$ sphingolipid analyses; $n=180$ glycerophosphocholine analyses), (iii) a technical replicate study of two human plasma samples assessed in 33 sequential runs separated by blanks ($n=33$ sphingolipid analyses), and as sample data provided online (iv) two datasets of human plasma of persons positive or negative for SARS-CoV-2 (Galipeau, 2021) ($n=24$ glycerophosphocholine longitudinal analyses provided as two datasets) and (v) a test dataset of human plasma of persons positive or negative for SARS-CoV-2(Galipeau, 2021) ($n=12$ glycerophosphocholine analyses).

To identify all lipids unambiguously in Datasets 1–3, all molecular identities were confirmed by LC-SRM-information dependent acquisition (IDA)-enhanced product ion (EPI) experiments of samples pooled across all datasets in which the SRM was used as a survey scan to identify target analytes and an IDA of an EPI spectra was acquired in the linear ion trap and examined to confirm molecular identities. For Datasets 4 and 5, each lipid identity was confirmed by LC-IDA-EPI-ESI-MS/MS using SRM as the survey scan. These structural analyses of EPI spectra were further validated by analyzing each lipid (for which commercial standards existed) individually as a standard. All lipids within the sphingolipid dataset were monitored at the same product ion m/z of 264.3 detecting sphingolipids with a d18:1 sphingoid base backbone (sphingosine). All lipids within the glycerophosphocholine dataset were monitored at the same product ion m/z of 184.1 detecting glycerophospholipids and sphingomyelins with a phosphocholine headgroup. Samples from both the sphingolipid and glycerophosphocholine datasets were equally stratified by acquisition date into training sets for cross validation and holdout sets for model validation. Complete LC-ESI-MS/MS details are provided in Supplementary Material.

2.5 Performance metrics

Classifier performance was assessed using metrics of accuracy, identification rate, and unassignment rate. These metrics evaluated how well BATL assigned lipid identities and the calibration of the unassigned identity weights. A correct peak assignment (true positive or TP) was defined as occurring when the classifier assigned the same identity established by IDA-EPI analysis. An incorrect peak assignment (false positive or FP) was defined when the classifier assigned a different identity than the one determined by IDA-EPI structural validation. An unassigned peak (U) refers to when the classifier assigned no identity to the peak (unassigned). Any unassigned peaks were considered incorrectly unassigned when the true identity of all peaks was present in the annotated training set.

$$\text{Accuracy} = \frac{\text{TP}}{\text{TP}+\text{FP}+U} \tag{9}$$

$$\text{Identification rate} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (10)$$

$$\text{Unassignment rate} = \frac{U}{\text{TP} + \text{FP} + U} \quad (11)$$

2.6 Availability and implementation

To facilitate use of BATL, we have developed a user friendly R/Shiny (Chang *et al.*, 2021) Web application that enables labeling of MultiQuant SCIEX data utilizing user and BATL-labeled training datasets. The application with user instruction pages is available at https://complimet.ca/batl/. Users with result tables generated through other acquisition packages can simply use the program by downloading the sample data and formatting their training and test datasets accordingly.

3 Results

BATL was trained on the sphingolipid and glycerophosphocholine training sets with unassigned identity weights w_{iu} learned by 10-fold cross validation and a precursor/product ion tolerance of 0.5 m/z units. Models were constructed from every subset of features presented in Table 1. These features described peak retention times, intensities, and shapes calculated from the standard outputs of all targeted lipidomic peak-picking software programs (e.g., MultiQuant, version 3.02, SCIEX). To train BATL, labeled validated datasets were used as training sets by adding an additional column 'Lipid_identifier'. This identifier can be any standardized character string used by a laboratory to annotate lipid identity. To calculate BATL-specific peak features (Relative RT, Subtracted RT, Relative Area, and Relative Height), an internal standard must be specified by the user and can be identified in the GUI. This internal standard must be present in all samples and all datasets (training and test). For each model, lipid identities were assigned to peaks in the cross validation or holdout sets using the MAP, constrained MAP or MWBM decision rules. Two peak identification algorithms, retention time mean and retention time window, were also devised as benchmarks recapitulating manual curation performed on-the-fly by users using MultiQuant to target desired peaks. The retention time mean approach assigned peaks to the single lipid identity in the training set with the closest mean retention time. The retention time window approach computed a retention time range for each lipid identity based on their minimum and maximum observed retention times in the training set. Lipid identities were only assigned to peaks whose retention times unambiguously fell within the window of a single lipid species.

To identify the best decision rule, cross validation accuracies were compared between BATL models trained using retention time only but differing in decision rule. For comparison, the accuracies of the two retention time window/mean matching algorithms were included to benchmark the BATL models where Figure 3a shows over 95% accuracies on the sphingolipid dataset using any method except the retention time window approach. As peaks were only assigned if they fell within the retention time window of a single lipid identity, this method incurred a 10% unassignment rate on the sphingolipid dataset, which was two orders of magnitude greater than any of the BATL models (see Supplementary Fig. S1a–c).

Similar accuracies were observed across the naïve mean approach and three BATL models, given the relatively low isobaric complexity of the sphingolipid dataset. Only 56.3% of the peaks in the validation sets matched between two and four lipid isobars at the same transition in the training set (Supplementary Tables S1 and S2). Thus, a large proportion of peaks were guaranteed to match to their corresponding lipid identity. When single lipid targets were detected at a transition, the MWBM decision rule assigned the same peak identities as the MAP or constrained MAP decision rule.

The strengths of different BATL models emerged when classifying the more complex glycerophosphocholine dataset in Figure 3b, where 95.3% of all peaks in the validation datasets were present in transitions that contained at least two and up to eight unique lipid isomers (Supplementary Tables S3 and S4). The BATL model, using the MWBM decision rule, achieved 88.7% accuracy and significantly outperformed every other method (Supplementary Fig. S1d–f). Performances were recapitulated when analyzing the holdout sets (Supplementary Material S1), and similar increases in accuracy were also observed when comparing decision rules of models trained using other feature subsets (see Supplementary Figs S2 and S3).

Table 1. Specified SRM peak features for naïve Bayes model

Feature	Description
Retention time (RT)	Peak retention time
Relative RT (RRT)	Peak divided by internal standard retention time
Subtracted RT (SRT)	Peak subtracted by internal standard retention time
Relative area (A)	Peak divided by internal standard area
Relative height (H)	Peak divided by internal standard height
Full width at half max (FWHM)	Peak width at half maximum height
Asymmetry factor (AF)	Quotient between centerline to back slope and centerline to front slope at 10% max peak height
Tailing factor (TF)	Distance between the front and back slope of a peak divided by twice the distance between the centerline and front slope at 5% max peak height

To understand why the MWBM decision rule outperformed the other methods, retention time likelihoods were assessed for the glycerophosphocholine cross validation analyses. Figure 3c shows the Gaussian likelihoods of five glycerophosphocholine isomers based on the retention time feature. When peak retention times were close together, both the naïve mean approach and MAP decision rule assigned multiple peaks to the same lipid identities. While the constrained MAP decision rule conceptually improved on the MAP decision rule, accuracies were significantly worse on the glycerophosphocholine dataset. Constrained MAP assigned lipid identities by ranked order of posterior probability. These rankings are denoted in Figure 3c by the ordinal numbers above the assignment arrows. However, interestingly, the most correct peak assignment was not necessarily the one with the greatest posterior probability. As discussed in Figure 1, retention time shifts can cause peak retention times in one sample to misalign to different peaks present in another sample. A similar problem arises when computing the likelihoods of peaks in samples experiencing retention time shifts. Variations in retention time altered the posterior rankings, increasing the likelihood of an incorrect-versus-correct peak-lipid assignment. Thus, once one lipid identity was incorrectly assigned to a given peak, subsequent peaks with similar retention times were misclassified (or not assigned an identity). In contrast, the best performing MWBM decision rule resolved these two types of misidentifications.

While retention time is the most common feature for peak identification, it is not the only lipid-specific peak feature or necessarily the most discriminative one. Figure 3d and e shows cross validation accuracies of selected BATL models using the MWBM decision rule trained using different retention time, intensity and shape features. Across both the less complex sphingolipid and more complex glycerophosphocholine datasets, additional features describing peak intensity and shape increased classification accuracies and identification rates, while decreasing unassignment rates (Supplementary Fig. S4). When comparing models trained on the best subset of N features, the use of all eight features consistently resulted in the best identification and unassignment rates on the holdout sets (Supplementary Fig. S5). When adding statistically dependent features to the model, diminishing performance returns were observed, although identification and unassignment rates remained equal to or greater than less complex models on the holdout sets. Of the three retention time

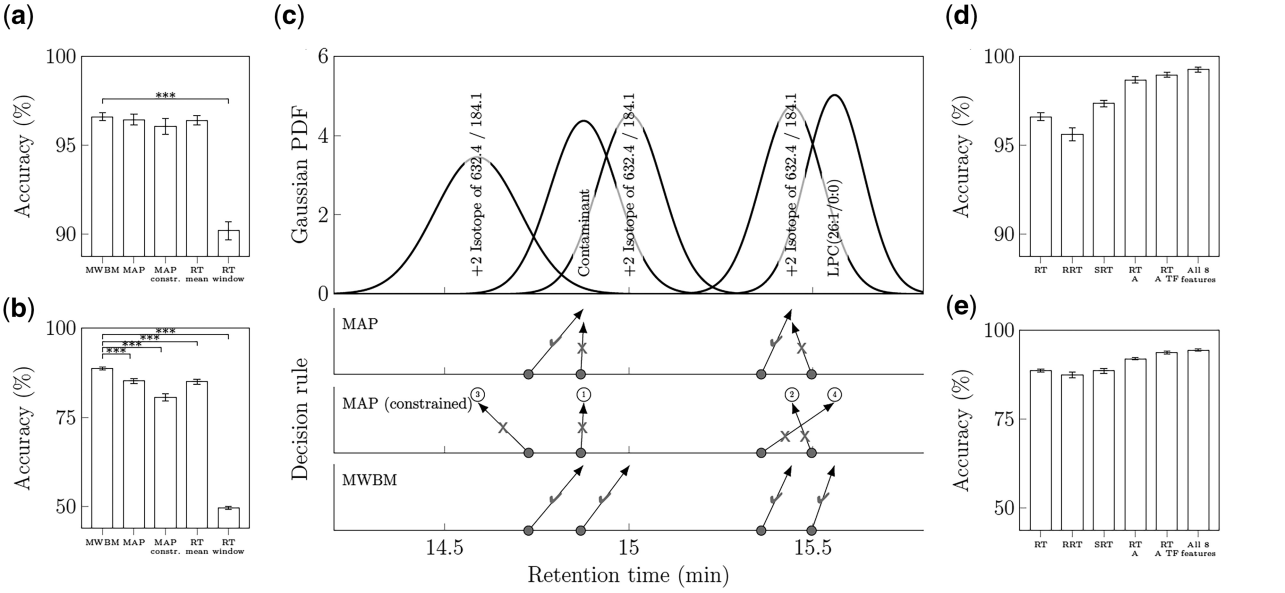

Fig. 3. Classifier performance on 10-fold cross validation sphingolipid and glycerophosphocholine datasets. The 95% confidence intervals are shown in panels (**a, b, d** and **e**). In **a** and **b**),data represent mean accuracies of BATL models trained on retention time with each decision rule and retention time mean/window matching algorithms for (**a**) sphingolipids or (**b**) glycerophosphocholines ($***Q < 0.001$, *t*-test adjusted with the Benjamini–Hochberg method of all models against the MWBM decision rule). (**c**) Lipid assignment differences between MAP, constrained MAP, and MWBM decision rules during cross validation and trained using retention time. In the top panel, data represent the Gaussian likelihoods of five glycerophosphocholine isomers based on the retention time feature. The rows of gray dots indicate the retention times of four peaks from the same sample in the validation set. Each row indicates the outcome of the three decision rules. Arrows indicate the lipid assignments; checkmarks indicate correct assignments; and Xs indicate incorrect assignments. The numbers for constrained MAP indicate the order of peak assignments. In **d** and **e**,data represent mean accuracies of the BATL models using MWBM decision rule trained on several features and feature combinations for (**d**) sphingolipids or (**e**) glycerophosphocholines. The feature name codes are described in Table 1

features explored, models trained using subtracted retention time performed equal to or significantly greater than those trained using regular retention time. Notably, this method of accounting for variations in LC retention time has not been reported in literature. Software programs, such as MultiQuant, can report both regular and relative retention times if an internal standard is specified. Although relative retention time is designed to correct against retention time shifts, this method of normalization was sometimes found to induce retention time shifts when no systematic retention time differences were observed across samples (i.e., only transient component level variation was detected). A comparison of models trained on each feature using the MWBM decision rule revealed equal to or significantly worse identification rates between relative-versus-regular retention time (Supplementary Fig. S6). Overall, models trained using retention time features significantly outperformed peak intensity and shape features, which were the least discriminative, while combinations of multiple features outperformed models focusing on single feature characteristics.

To ensure BATL can be used across platforms, researchers are required to develop their own curated training sets specific to their LC methodologies. Limitations of BATL are that the annotations returned by BATL depend on the accuracy of the identifications assigned in the training set and on the size of the training dataset. Supplementary Figure S7 shows the performance of BATL on the holdout sets when trained on 10% increments of the sphingolipid or glycerophosphocholine training sets. Models were trained using the best single feature or all eight features and every 10% increment corresponded to 22 sphingolipid or 24 glycerophosphocholine samples. Whether trained on the less complex sphingolipid dataset or the more complex glycerophosphocholine dataset, identification rates decreased by <1% and unassignment rates remained under 5% when training on 10% of samples. These data demonstrate that only a small rigorously validated training set (i.e., 22–24 samples) is required to train the naïve Bayes model for accurate peak identification.

Benchmarking BATL against other state-of-the-art methods for peak classification is challenging because BATL assigns lipid identities to a list of curated SRM peaks provided by the user, as is the nature of a targeted lipidomic approach, while vendor-specific (e.g., LipidSearch, MultiQuant) and free programs (e.g., METLIN-MRM) pick peaks automatedly and output the pre-assigned targeted identities assuming peak-picking accuracy. As a result, for all programs except MRMPROBS, it is not possible to separate peak detection accuracy from peak identification accuracy. Indeed, this is one of the problems BATL seeks to address. BATL was thus benchmarked against MRMPROBS (Tsugawa *et al.*, 2013). A notable shortcoming of MRMPROBS, however, is that the number of lipids in the training set cannot exceed the number of lipid targets in the SRM method, meaning that MRMPROBS can only compare identical acquisition methods and cannot annotate a peak as ‘unassigned’ or indicate a new isobar has been selected not already present in the training set. It was thus impossible to apply MRMPROBS to the sphingolipid or glycerophosphocholine holdout sets as they contained different numbers of isobars at a given transition in the training set. This problem was overcome by applying MRMPROBS to multiple training sets containing all combinations of lipid isobars, not exceeding the number of sample peaks. In practice, however, MRMPROBS cannot be used to compare matrices wherein different numbers of isobars are present and a user seeks to annotate which lipids are corresponding between two tissues. To compare, BATL and MRMPROBS, we used a technical replicate dataset, which applied the exact same SRM method to monitor sphingolipid species present in 33 replicate runs of two human plasma samples. Thus, both training and testing sets contained the same number of lipids *de facto*. For this analysis, 75% of the samples in the dataset were used to train MRMPROBS and BATL. To construct the MRMPROBS training set, the mean retention times of each lipid were computed from the training set, the logistic regression probability threshold was set to 70% and the retention time deviation parameter was empirically computed following the MRMPROBS guidelines (Tsugawa *et al.*, 2013). On the remaining 25% of the technical replicate holdout set, 8.9% of all peaks were not detected by MRMPROBS using a 15-s retention time window to account for peak detection differences between MRMPROBS and MultiQuant, which was used to pick peaks in the longitudinal dataset. Excluding the 8.9% undetected peaks, MRMPROBS achieved a 94.5% identification rate and 4.7% unassignment rate, while BATL, trained using all eight features, achieved 100% identification rate and 0.02% unassignment rate.

4 Discussion

We present here a targeted lipidomics classifier BATL, which uses a naïve Bayes model and MWBM decision rule to simultaneously assign lipid annotations to all SRM or MRM peaks in a sample. Using sphingolipid and glycerophosphocholine SRM datasets, BATL was validated on holdout sets with accuracies of 95% or greater when trained using all eight features. As a simple probabilistic classifier, identification and assignment rates remained stable when BATL was trained on as few as 22–24 samples. Lastly, BATL was benchmarked against a retention time window and mean matching approach, comparable to many peak identification programs as well as to the MRMPROBS software. BATL correctly identified more peaks than either approach with lower unassignment rates and no limitations regarding the number of lipids labeled in the training set nor number of transitions present in the test sets.

In summary, we emphasize that BATL is trainable on any continuous feature and applicable to targeted lipidomics data from any vendor or LC-ESI-MS/MS platform with proper data input formatting. Learning the posterior probability cutoffs is simple to compute based on the naïve Bayes assumption, taking <10 min to train each model on the sphingolipid and glycerophosphocholine training sets analyzed in this study using an Intel i5-8350U mobile processor. To facilitate user experience, we provide BATL at https://complimet.ca/batl/ wherein instructions and user prompts ensure a fluid training and test pipeline. The sample datasets take <10 min to process online with respect to building the BATL model. Larger training datasets of >100 samples can take up to 20 min to generate the BATL statistical model and the model can be emailed to the user. Multiple training datasets can be compressed as .zip files and uploaded as single .zip file. Users are invited to use the sample biological data wherein the minimum number of training datasets ($n = 24$) are provided as two files that can be .zip for upload and used to annotate an $n = 12$ test dataset. While BATL was validated using SRM data, the program is flexible to operate on other targeted lipidomics data acquisition modes that output lists of peaks detected at precursor and product ion pairs including MRM, neutral loss, precursor ion scan, and product ion scan acquisition modes with the caveat that file format must be identical to the sample datasets provided. As BATL does not operate on raw MS data, users can continue using their preferred software program to select their lipid targets and conveniently output peak text files into BATL for identification.

Funding

This work was supported in part by operating grants RGPIN-2019-06796 to S.A.L.B. and RGPIN-2019-06604 to T.J.P. from the Natural Sciences and Engineering Research Council of Canada (NSERC), AI-4D-102-3 to S.A.L.B. and M.C.-C. from the National Research Council AI for Design Challenge Program, as well as an NSERC CREATE Matrix Metabolomics Training grant to S.A.L.B. and T.J.P. T.T.N. and J.G.C. received an NSERC CREATE Matrix Metabolomics Scholarship. J.G.C. was further supported by an NSERC Alexander Graham Bell Canada Graduate Scholarship and an Ontario Graduate Scholarship. I.A. received a Parkinson Research Consortium Crabtree Family Fellowship. M.G.S. was supported by the Department of Medicine and the Sam and Uttra Bhargava Family.

Conflict of Interest: none declared.

References

Alecu,I. and Bennett,S.A.L. (2019) Dysregulated lipid metabolism and Its role in alpha-synucleinopathy in Parkinson's disease. *Front. Neurosci.*, **13**, 328.

Alshehry,Z.H. *et al.* (2016) Plasma lipidomic profiles improve on traditional risk factors for the prediction of cardiovascular events in type 2 diabetes mellitus. *Circulation*, **134**, 1637–1650.

Blasco,H. *et al.* (2017) Lipidomics reveals cerebrospinal-fluid signatures of ALS. *Sci. Rep.*, **7**, 17652.

Bowden,J.A. *et al.* (2017) Harmonizing lipidomics: NIST interlaboratory comparison exercise for lipidomics using SRM 1950-metabolites in frozen human plasma. *J. Lipid Res.*, **58**, 2275–2288.

Bowden,J.A. *et al.* (2018) NIST lipidomics workflow questionnaire: an assessment of community-wide methodologies and perspectives. *Metabolomics*, **14**, 53.

Chang,W.C. *et al.* (2021) *shiny: Web Application Framework for R*. https://cran.r-project.org/web/packages/shiny/index.html (28 December 2021).

Domingo-Almenara,X. *et al.* (2018) XCMS-MRM and METLIN-MRM: a cloud library and public resource for targeted analysis of small molecules. *Nat. Methods*, **15**, 681–684.

Galipeau,Y. *et al.*. (2021) Relative Ratios of Human Seasonal Coronavirus Antibodies Predict the Efficiency of Cross-Neutralization of SARS-CoV-2 Spike Binding to ACE2. EBioMedicine, **74**, https://doi.org/10.1016/j.ebiom.2021.103700.

Gallien,S. *et al.* (2014) Technical considerations for large-scale parallel reaction monitoring analysis. *J. Proteomics*, **100**, 147–159.

Granger,M.W. *et al.*. (2016) A TgCRND8 Mouse Model of Alzheimer's Disease Exhibits Sexual Dimorphisms in Behavioral Indices of Cognitive Reserve. Journal of Alzheimer's Disease : JAD, **51**, 757–773. https://doi.org/10.3233/JAD-150587.

Granger,M.W. *et al.* (2019) Distinct disruptions in Land's cycle remodeling of glycerophosphocholines in murine cortex mark symptomatic onset and progression in two Alzheimer's disease mouse models. *J. Neurochem.*, **149**, 499–517.

Murphy,R.C. and Axelsen,P.H. (2011) Mass spectrometric analysis of long-chain lipids. *Mass Spectrom. Rev.*, **30**, 579–599.

Peterson,A.C. *et al.* (2012) Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol. Cell. Proteomics*, **11**, 1475–1488.

Quehenberger,O. *et al.* (2010) Lipidomics reveals a remarkable diversity of lipids in human plasma. *J Lipid Res.*, **51**, 3299–3305.

Reiter,L. *et al.* (2011) mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nat. Methods*, **8**, 430–435.

Sartain,M.J. *et al.* (2011) Lipidomic analyses of mycobacterium tuberculosis based on accurate mass measurements and the novel "Mtb LipidDB". *J. Lipid Res.*, **52**, 861–872.

Slatter,D.A. *et al.* (2016) Mapping the human platelet lipidome reveals cytosolic phospholipase A2 as a regulator of mitochondrial bioenergetics during activation. *Cell Metab.*, **23**, 930–944.

Smith,R. *et al.*. (2015) LC-MS alignment in theory and practice: a comprehensive algorithmic review. Briefings in Bioinformatics, **16**, 104–117. https://doi.org/10.1093/bib/bbt080.

Tsugawa,H. *et al.* (2013) MRMPROBS: a data assessment and metabolite identification tool for large-scale multiple reaction monitoring based widely targeted metabolomics. *Anal. Chem.*, **85**, 5191–5199.

Ubhi,B.K. *et al.* (2016) *A Novel Lipid Screening Platform that Provides a Complete Solution for Lipidomics Research. SCIEX Technical Application Notes*. https://www.technologynetworks.com/analysis/application-notes/a-novel-lipid-screening-platform-that-provides-a-complete-solution-for-lipidomics-research-228552 (20 September 2021, date last accessed).

Wang,W. *et al.* (2018) Lipidomic profiling reveals soluble epoxide hydrolase as a therapeutic target of obesity-induced colonic inflammation. *Proc. Natl. Acad. Sci. USA*, **115**, 5283–5288.

BATL: Bayesian annotations for targeted lipidomics Supplementary information

Justin G. Chitpin[1-5], Anuradha Surendra[6], Thao T. Nguyen[3,4,5,7], Graeme P. Taylor[3-5], Hongbin Xu[3-5], Irina Alecu[3-5], Roberto Ortega[8], Julianna J. Tomlinson[9,10], Angela M. Crawley[2,5], Michaeline McGuinty[2], Michael G. Schlossmacher[9,10], Rachel Saunders-Pullman[8], Miroslava Cuperlovic-Culf[5,6*], Steffany A.L. Bennett[2-5,7,9,10*], Theodore J. Perkins[1,2,4*]

[1]Regenerative Medicine Program,
[2]Ottawa Hospital Research Institute, Ottawa, ON, Canada,
[3]Neural Regeneration Laboratory and India Taylor Lipidomics Research Platform, University of Ottawa Brain and Mind Research Institute,
[4]Ottawa Institute of Systems Biology,
[5]Department of Biochemistry, Microbiology and Immunology, University of Ottawa, Ottawa, ON, Canada,
[6]Digital Technologies Research Center, National Research Council, Ottawa, ON, Canada,
[7]Department of Chemistry and Biomolecular Sciences, Centre for Catalysis Research and Innovation, University of Ottawa, Ottawa, ON, Canada,
[8]Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, New York, USA
[9]Department of Cellular and Molecular Medicine, University of Ottawa, Ottawa, ON, Canada,
[10]Neuroscience Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada

Contact: miroslava.cuperlovic-culf@nrc-cnrc.gc.ca, tperkins@ohri.ca, and sbennet@uottawa.ca

Contents

1. Experimental protocols

Human plasma, collected and prepared in K_2EDTA Lavender BD Hemogard tubes (#367863), was extracted from cognitively normal persons, and patients suffering from Alzheimer's Disease, Mild Cognitive Impairment, Dementia with Lewy Bodies, or Parkinson's Disease (n=321 participants) or persons positive or negative for SARS-CoV-2 (n=36 participants). Consent was obtained in accordance with the Ottawa Hospital Research Institute Research Ethics Committee and in agreement with the National Institutes of Health and NINDS and the Icahn School of Medicine at Mount Sinai Research Ethics Committee. Use of human samples for this study was approved ty the University of Ottawa Ethics Review Board, certificate H-06-20-5864 and H-02-19-3400. Murine plasma and brain, specifically hippocampus and temporal cortex, extracts were generated as part of an ongoing genotype and intervention study of lipid metabolism in n=180 wildtype and N5 TgCRND8 mice, a sexually dimorphic mouse model of Alzheimer's Disease (Granger *et al.,* 2016). All procedures were approved by the Animal Care Committee of the University of Ottawa and performed in accordance with the ethical guidelines for experimentation of the Canada Council for Animal Care. Lipid extraction methodology was as described in detail in Xu *et al.* (2013).

Briefly, brain tissue was homogenized in 4 mL acidified methanol (A412P-4; Fisher, Nepean, ON, Canada) containing 2% acetic acid (351271-212; Fisher, Nepean, ON, Canada) using a tissue tearer (985370; BioSpec, Bartlesville, OK, USA). For plasma samples, 100 µl of human plasma or 50 µl of murine plasma were added directly to 4 mL methanol containing 2% acetic acid. Internal standards, 90.7 ng PC(13:0/0:0) [LM1600], 99.54 ng PC(12:0/13:0) [LM1000], 249 ng PE(12:0/13:0) [LM1100], 249 ng PS(12:0/13:0) [LM1300], 133.7 ng Cer(d18:1/16:0-D31) [868516], 133.7 ng GlcCer(d18:1/8:0) [860540], 133.7 ng GalCer(d18:1/8:0) [860538], and 75 ng SM(d18:1/18:1-d9) [860740] were added at time of extraction. All standards were from Avanti Polar Lipids, Alabaster, AL, USA. Sodium acetate (0.1 M; S-2889; Sigma-Aldrich, Oakville, ON, Canada) and chloroform (C298-500; Fisher) were added sequentially to each sample to a final ratio of 2:1.9:1.6 (acidified methanol/chloroform/sodium acetate). Samples were vortexed, incubated on ice for 15 minutes, and centrifuged for 5 minutes at 600 x g at 4°C. The organic phase was collected, and the aqueous phase back-extracted 3 times using chloroform. Each organic phase was combined and dried under a steady stream of nitrogen. Lipids were re-solubilized in 300 µl of anhydrous ethanol (P016EAAN; Commercial Alcohols, Toronto, ON, Canada) and stored under nitrogen gas at -80°C in amber vials (C779100AW; BioLynx, Brockville, ON, Canada).

HPLC was performed on an Agilent 1290 Infinity LC system, equipped by a binary pump, with an autosampler maintained at 4°C. Aqueous mobile phase (Solvent A) contained 0.1% formic acid and 10 mM ammonium acetate. Organic mobile phase (Solvent B) contained acetonitrile/isopropanol (5:2 v/v) with 0.1% formic acid and 10 mM ammonium acetate. Reversed-phase liquid chromatography for sphingolipid assessment was performed on a 100 mm x 250 µm (inner diameter) capillary column packed with either ReproSil-Pur 120 C8 (particle size of 3 µm and pore size of 120 Å or ReproSil-Pur 200 C18 (particle size of 3 µm and pore size of 200 Å, Dr. A. Maisch, Ammerbruch, Germany) for glycerophosphocholine assessment. For the GPC analysis, five µl of sample were mixed with 5 µl of an internal standard mixture consisting of PC(O-16:0-d4/0:0) [2.5 ng, 360906; Cayman Ann Arbor, MI, USA], PC(O-18:0-d4/0:0) [2.5 ng, 10010228; Cayman], PC(O-16:0-d4/2:0) [1.25 ng, 360900; Cayman], and PC(O-18:0-d4/2:0) [1.25 ng, 10010229; Cayman] in EtOH, PC(15:0/18:1-d7) [1.25 ng, 791637; Avanti Polar Lipids], PE(15:0/18:1-d7) [1.25 ng, 791638; Avanti Polar Lipids], LPC(18:1-d7/0:0) [1.25 ng,791643; Avanti Polar Lipids], LPE(18:1-d7/0:0) [1.25 ng, 791644; Avanti Polar Lipids], and PS(15:0/18:1-d7) [1.25 ng, 791639; Avanti Polar Lipids] and 13.5 µl of Solvent A. For sphingolipid analysis, five µl of sample were mixed with 2.5 µl of EtOH and 16 µl of Solvent A. The LC method operated at a flow rate of 10 µl/min with 3 µl of sample injection for GPC analysis and 5 µl for sphingolipid analysis. The LC gradient used for GPC analysis started at 30% Solvent B, reached 100% Solvent B at 8 minutes, and remained at 100% B for 45 min. At 45 min, composition was returned to 30% Solvent B and the column was regenerated for 15 min. For sphingolipid analysis, the gradient began at 30% Solvent B, was ramped to 100% Solvent B over 5 min, and was maintained for 30 min. The composition was returned to 30% Solvent B at 30 min and maintained for 20 min to regenerate the column. A blank run, wherein 3 µl or 5 µl of Solvent A was injected, followed each sample run.

ESI-MS/MS acquisition and instrument control were performed using Analyst software (version 1.6.2, SCIEX) on a QTRAP 5500 triple quadrupole mass spectrometer equipped with a Turbo V ion source (SCIEX). The ion source operated at 5500 V and 0°C for GPC analysis, 250°C for sphingolipid analysis. Nebulizer/heated gas (GS1/GS2), curtain gas (CUR), and collision gas (CAD) were set to 20/0 psi, 20 psi, and medium, respectively, for GPC analysis. For sphingolipid analysis, the values for these source parameters are 20/20 psi, 20 psi, and medium, respectively. Nitrogen was used as GS1/GS2, curtain and collision gas. Compound parameters (declustering potential, entrance potential, collision energy, and collision cell exit potential) were individually optimized for each transition. All data acquisitions were performed in the positive ion mode. The GPC lipidome monitored the diagnostic product ion at m/z 184.1, indicative of the glycerophosphocholine headgroup, while the sphingolipidome was profiled using the diagnostic product ion 264.3 indicative of the sphingosine backbone. MultiQuant software (version 3.0.2 AB SCIEX) was used for peak picking and processing quantitative SRM data.

The identification of lipid species was performed by employing Information-Dependent-Acquisition Enhanced-Product-Ion-Scan (IDA-EPI) following the quantitative MRM acquisition which served as a survey scan. The IDA method triggered EPI scans following analysis of MRM signals with dynamic background subtraction from the survey scan. The IDA criteria were set to select one to three most intense peaks, and intensity threshold was set to exceed 1000 cps. The EPI experiment operated in the positive mode, scanning mass range from m/z 200-1000 at a scan rate of 10,000 Da/s with dynamic fill in the trap. Two different collision energies were applied, 35 and 50 eV with collision energy spread (CES) of 15 eV to ensure a broad coverage of fragmentation.

2. Author contributions

JGC developed the BATL method, conducted the analyses, and drafted the manuscript. AS developed the GUI interface and online implementation of BATL at http://complimet.ca/batl/. RO, JJT, AMC, MM, MGS, and RSP provided the human samples. TTN, GPT, HX, and IA generated all animal samples, prepared all human and animal samples and generated all of the mass spectrometry data. TJP and SALB guided the analysis of the core BATL program and revised the manuscript. MCC and SALB guided the analysis of the online implementation and revised the manuscript. All authors read, edited, and approved the manuscript.

2. Dataset complexity

Table 1: Sphingolipid training set isomeric/isobaric complexity.

Number of candidate assignments	Detected peaks	Exact mass assignments[1]
1	13222	13222
2	11403	704
3	4590	94
4	1023	0

[1] Peak matches the correct lipid target based on Q1 and Q3 m/z alone, regardless of the feature value(s).

Table 2: Sphingolipid holdout set isomeric/isobaric complexity.

Number of candidate assignments	Detected peaks	Exact mass assignments[1]
1	13069	13069
2	11307	800
3	4545	159
4	1008	0

[1] Peak matches the correct lipid target based on Q1 and Q3 m/z alone, regardless of the feature value(s).

Table 3: Glycerophosphocholine training set isomeric/isobaric complexity.

Number of candidate assignments	Detected peaks	Exact mass assignments[1]
1	9492	9492
2	16880	1034
3	39002	0
4	52940	173
5	45202	0
6	25033	0
7	7490	0
8	5831	0

[1] Peak matches the correct lipid target based on Q1 and Q3 m/z alone, regardless of the feature value(s).

Table 4: Glycerophosphocholine holdout set isomeric/isobaric complexity.

Number of candidate assignments	Detected peaks	Exact mass assignments[1]
1	9474	9474
2	16963	1050
3	38838	0
4	53361	181
5	44937	0
6	25760	0
7	7483	0
8	5810	0

[1] Peak matches the correct lipid target based on Q1 and Q3 m/z alone, regardless of the feature value(s).

Table 5: Feature codes unless otherwise stated in the following supplementary figure captions.

Feature	Abbreviation
Retention time	RT
Relative retention time	RRT
Subtracted retention time	SRT
Relative area	A
Relative height	H
Full width at half maximum	FWHM
Asymmetry factor	AF
Tailing factor	TF

3. Classifier holdout performance between decision rules

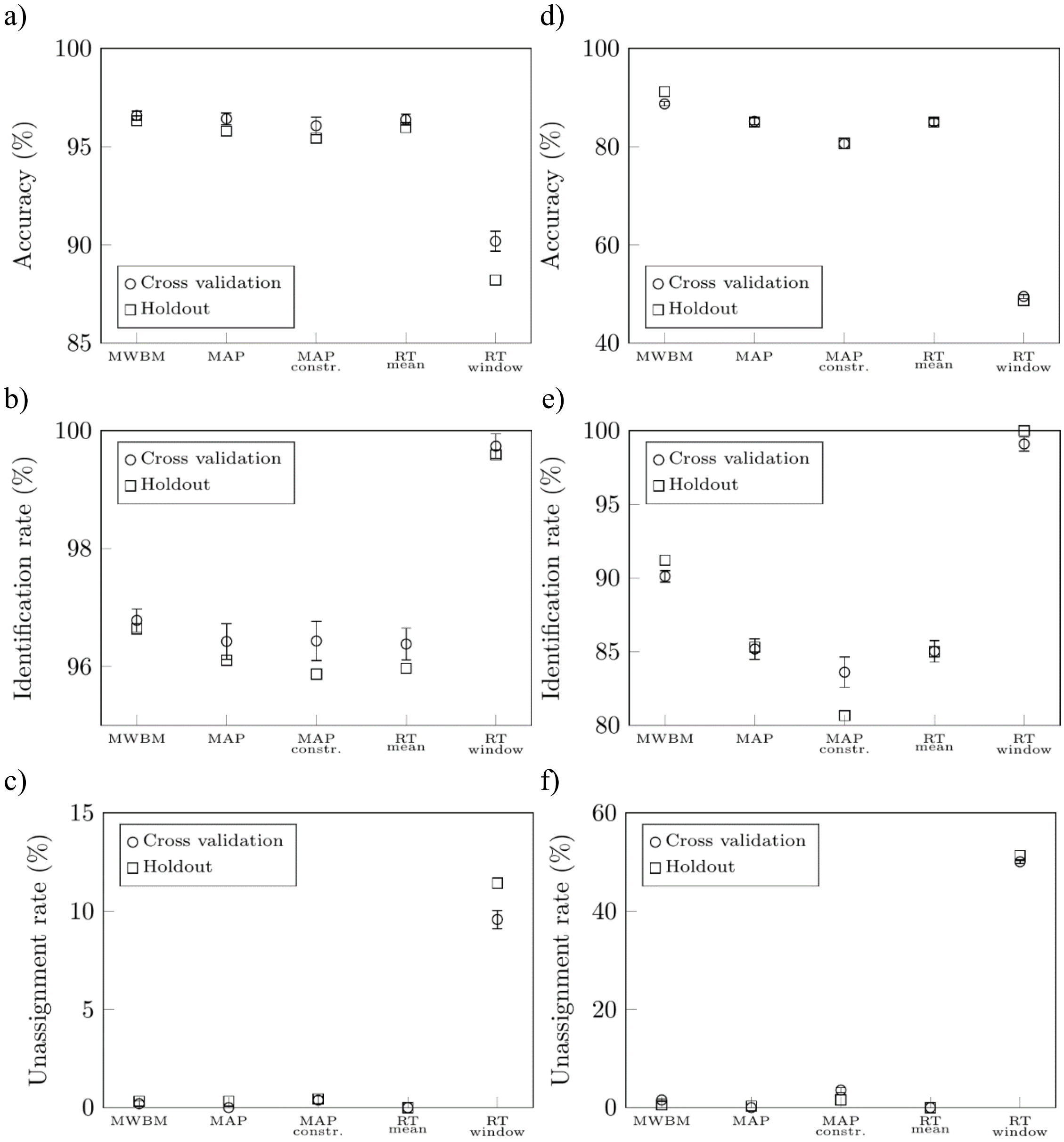

Figure S1: Detailed classifier performance on holdout and cross validation sets. For cross-validation results, point markers indicate mean accuracies/identification/unassignment rates and whiskers represent 95% confidence intervals across 10 folds. a-c) Accuracies, identification rates, and unassignment rates for the sphingolipid datasets. d-f) Accuracies, identification rates, and unassignment rates for the glycerophosphocholine datasets.

4. Classifier cross validation performance between decision rules for other feature combinations

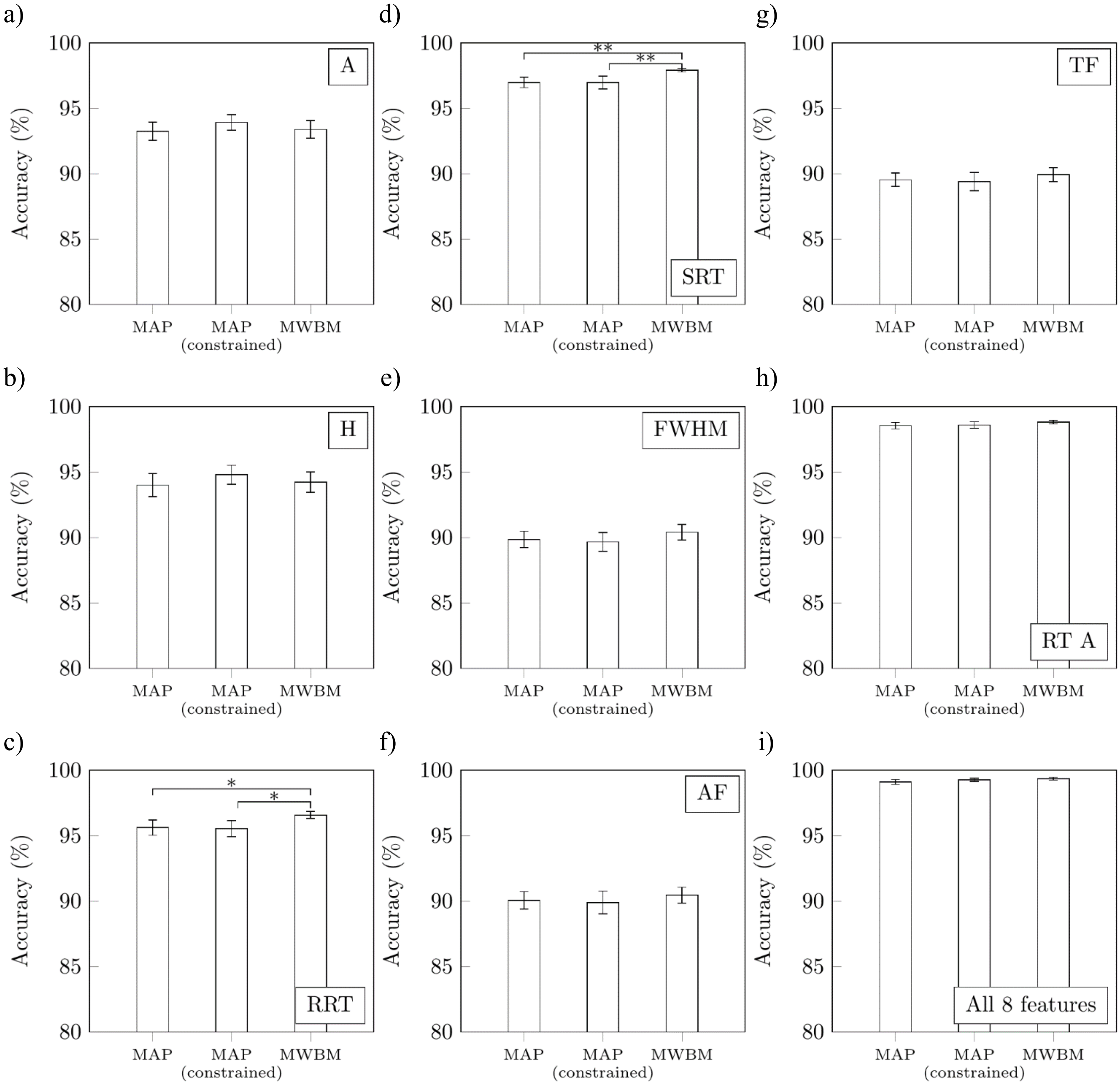

Figure S2: Classifier performance on 10-fold cross validation sphingolipid datasets. Data represent mean accuracies of BATL models trained on select feature combinations. 95% confidence intervals shown (*Q < 0.05, **Q < 0.01, ***Q < 0.001, *t*-test adjusted with the Benjamini-Hochberg method).

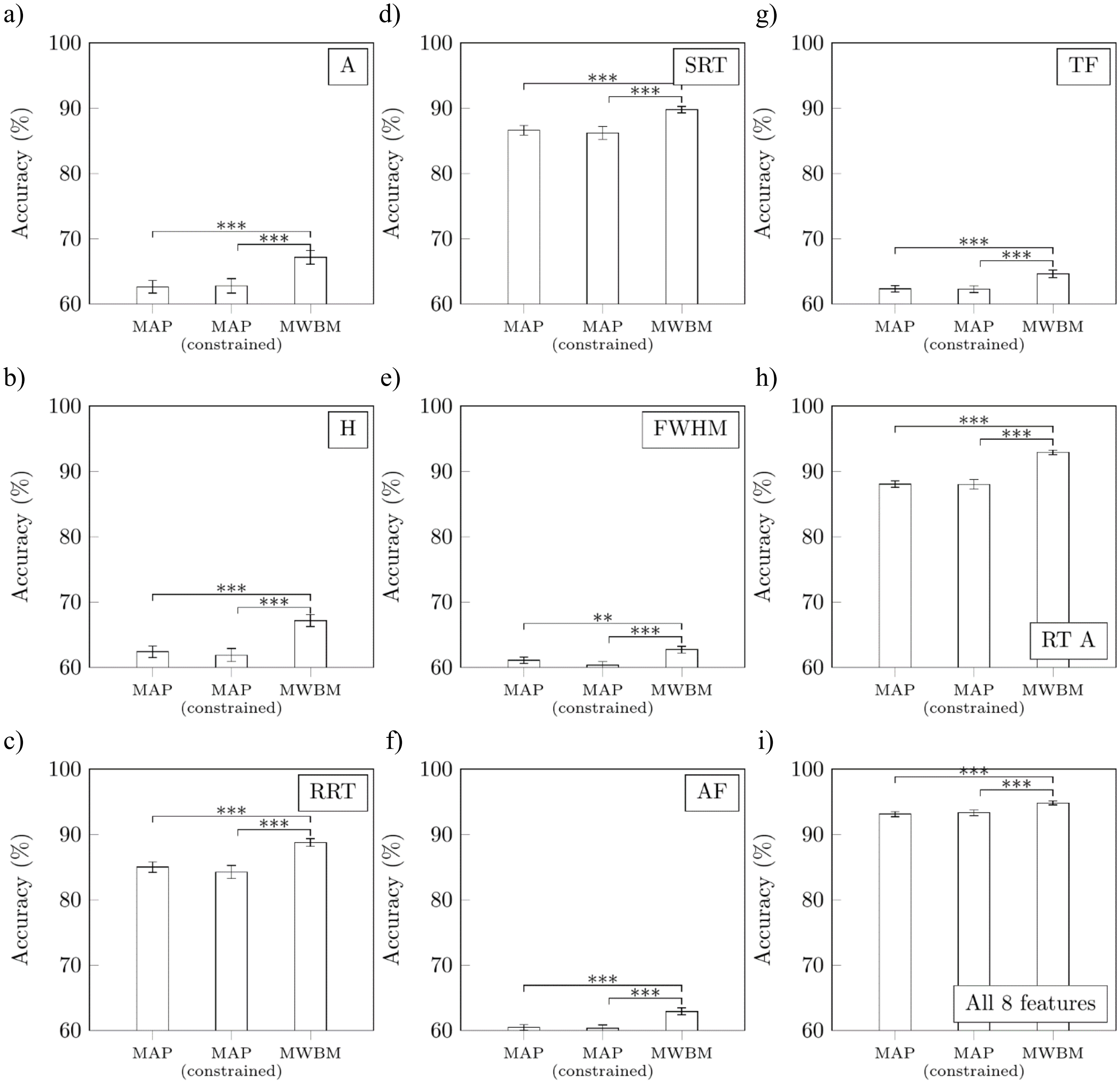

Figure S3: Classifier performance on 10-fold cross validation glycerophosphocholine datasets trained with select feature combinations. Data represent mean accuracies of BATL models trained on select feature combinations. 95% confidence intervals shown (**Q* < 0.05, ***Q* < 0.01, ****Q* < 0.001, *t*-test adjusted with the Benjamini-Hochberg method).

5. Classifier holdout performance between selected features

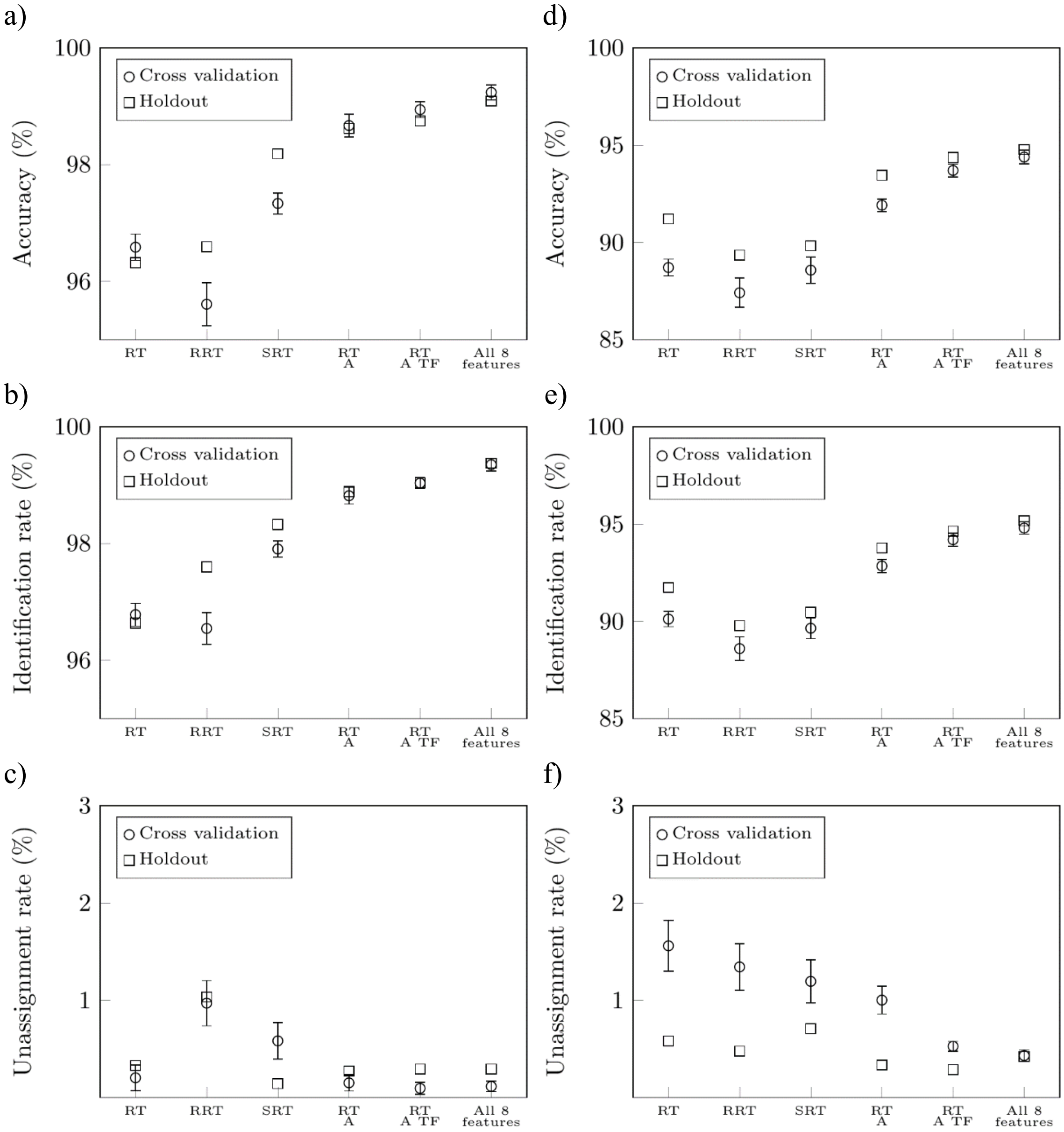

Figure S4: Classifier performance on holdout and cross validation sets trained with select feature combinations. For cross-validation results, point markers indicate mean accuracies/identification /unassignment rates and whiskers represent 95% confidence intervals across 10 folds. a-c) Accuracies, identification rates, and unassignment rates for the sphingolipid datasets. d-f) Accuracies, identification rates, and unassignment rates for the glycerophosphocholine datasets.

6. Classifier performance by best *N* feature combinations

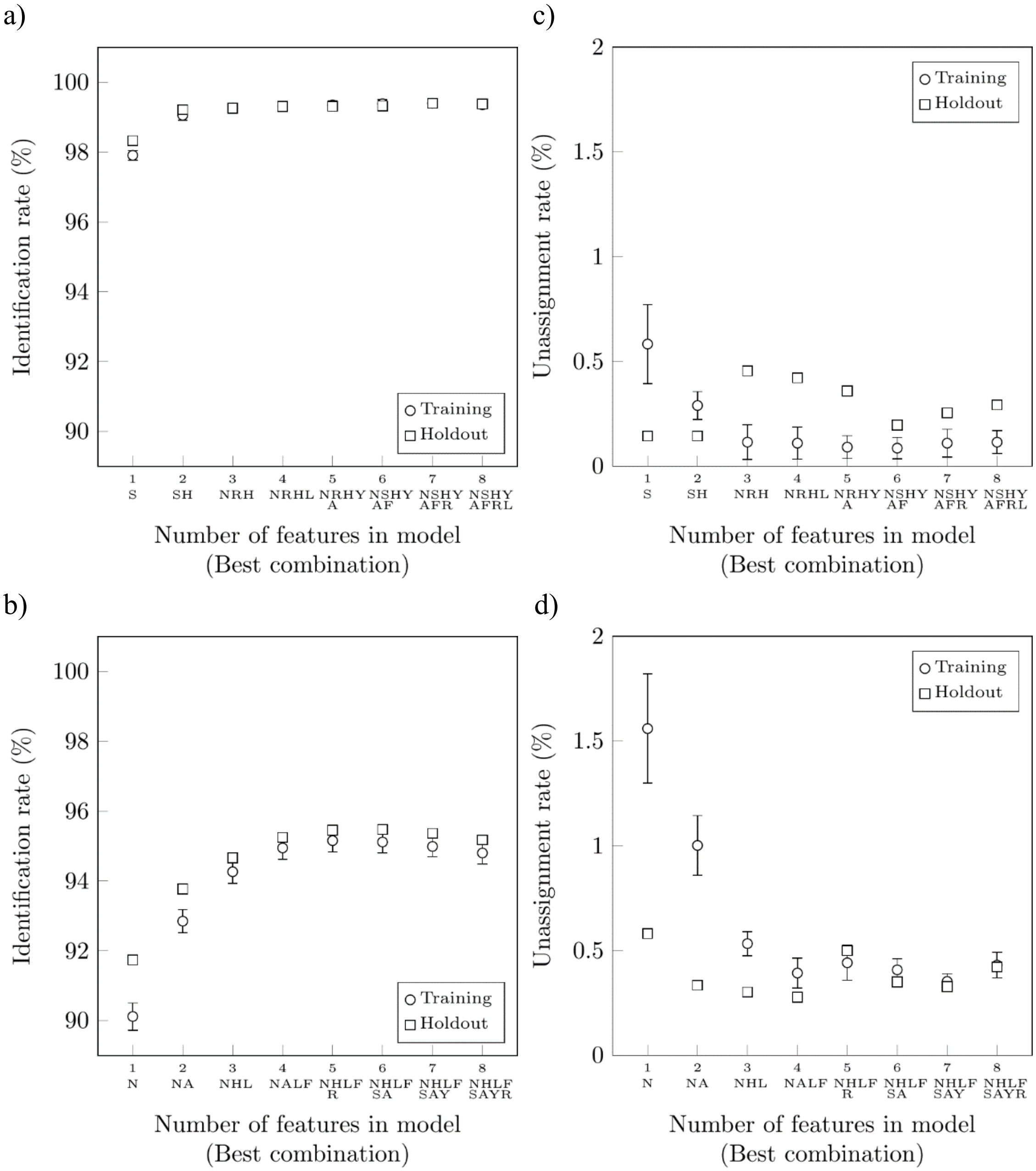

Figure S5: Models using the MWBM decision rule trained from the best combination of *N* features. X-label feature codes described in Figure S6. 95% confidence intervals for the 10-fold cross validation datasets shown. a-b) Sphingolipid and glycerophosphocholine dataset identification accuracies. c-d) Sphingolipid and glycerophosphocholine dataset unassignment rates.

7. Classifier performance by single feature

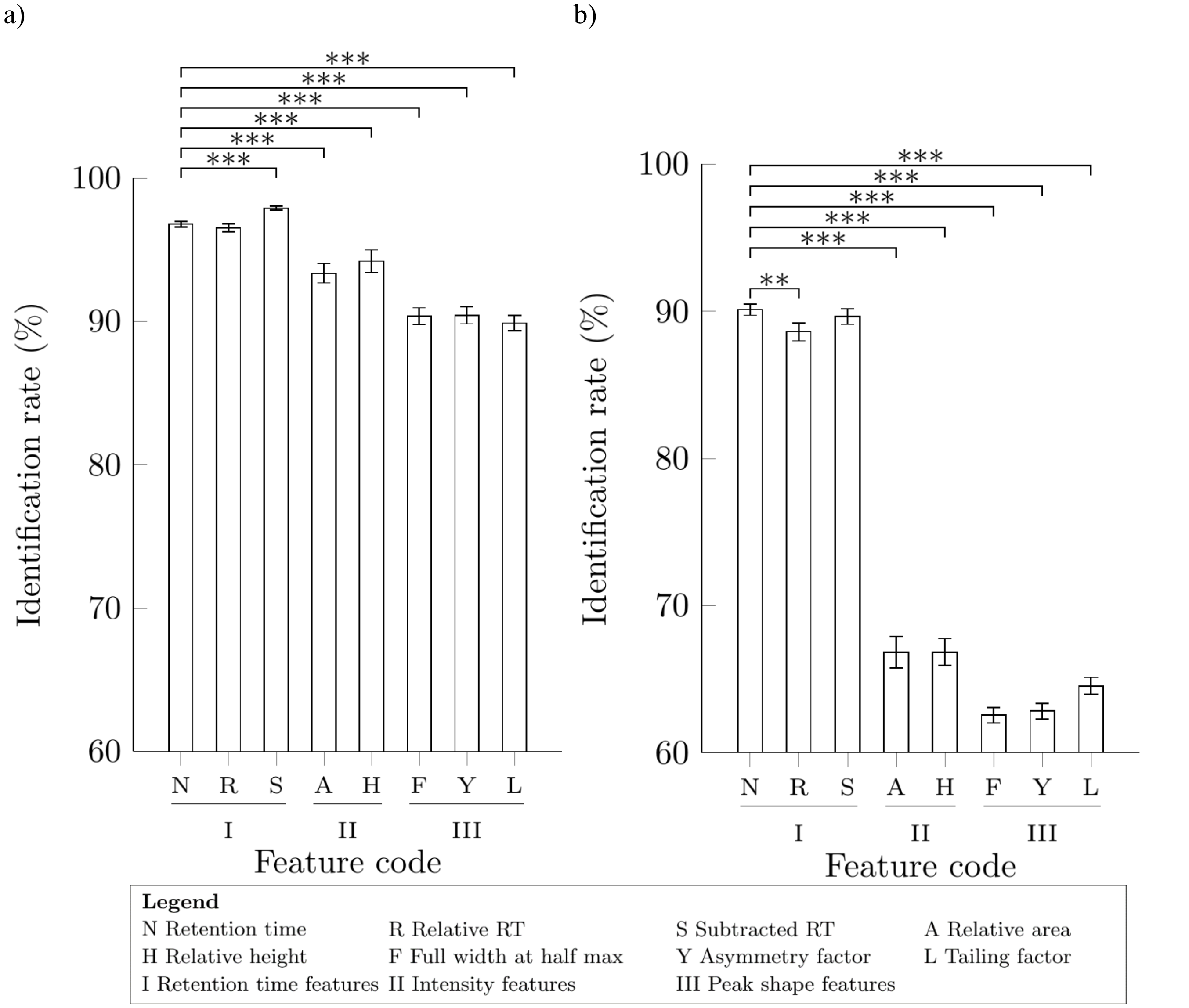

Figure S6: Comparison of models using the MWBM decision rule trained with retention time (N) versus all other single features. Ten-fold cross validation identification rates with 95% confidence intervals shown (pairwise *t*-test against the retention time feature adjusted with the Benjamini-Hochberg method *Q < 0.05, **Q < 0.01, ***Q < 0.001). a) Sphingolipid cross validation dataset. b) Glycerophosphocholine cross validation dataset.

8. Classifier performance by training set size

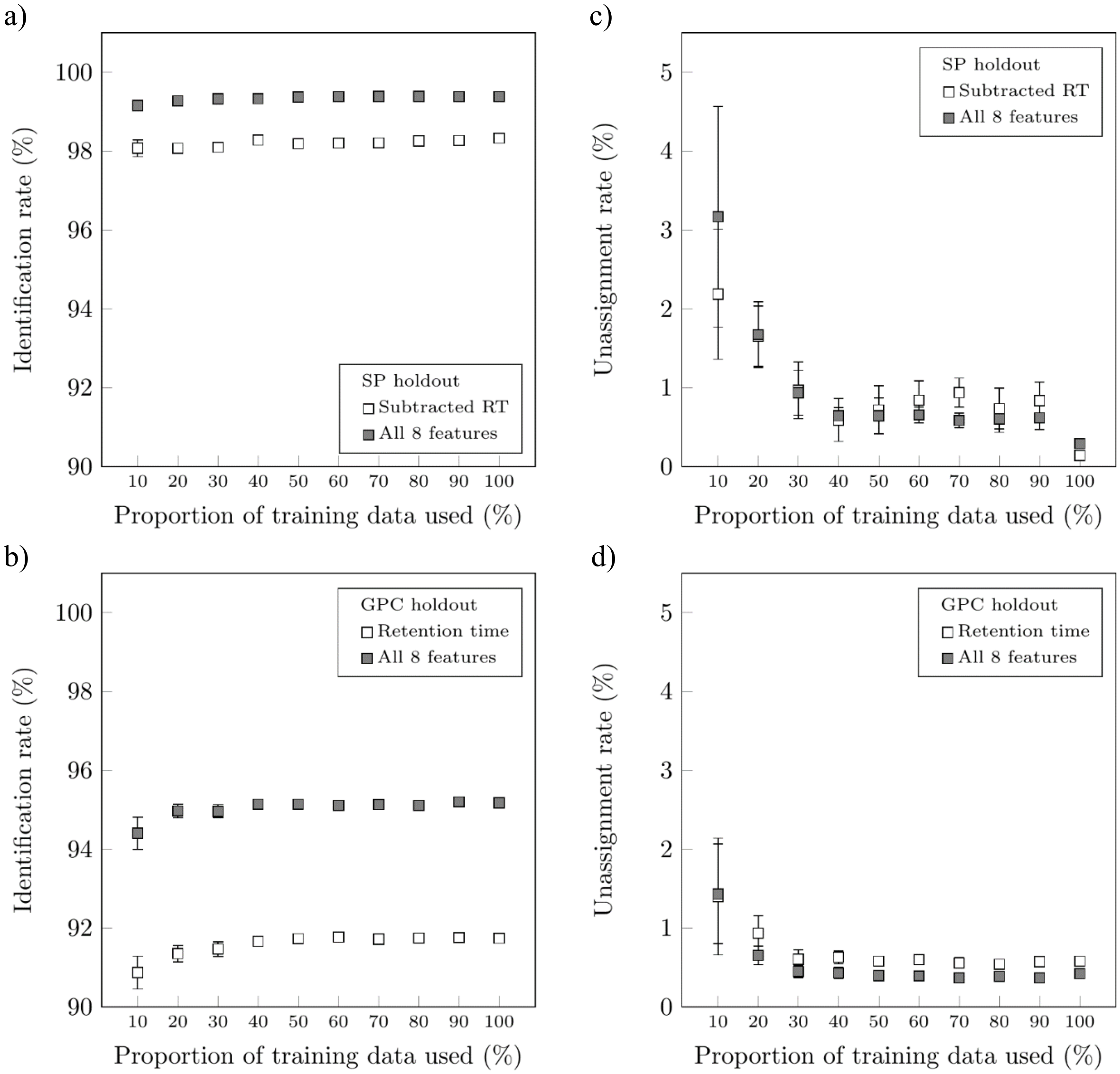

Figure S7: Classifier performance on holdout datasets with the best single feature or all features resampled ten times for each proportion of training data. Every 10% increment corresponds to 22 sphingolipid or 25 glycerophosphocholine samples. a-b) Point markers indicate mean identification rates and whiskers represent 95% confidence intervals for the sphingolipid and glycerophoshocholine holdout sets, respectively. c-d) Point markers indicate mean unassignment rates and whiskers represent 95% confidence intervals for the sphingolipid and glycerophosphocholine holdout sets, respectively.

References

Granger, M. W. *et al*. (2016). A TgCRND8 mouse model of Alzheimer's Disease exhibits sexual dimorphisms in behavioral indices of cognitive reserve. *J. Alzheimers Dis*., **51**(3), 757-773.

Xu, H. *et al*. (2013). Targeted lipidomics – advances in profiling lysophosphocholine and platelet-activating factor second messengers. *FEBS J*., **280**(22), 5652-5667.