bioRxiv preprint doi: https://doi.org/10.1101/2021.03.18.435788; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

picture(0,0)(-35,0)(1,0)30 (0,35)(0,-1)30 picture picture(0,0)(35,0)(-1,0)30 (0,35)(0,-1)30 picture

Systems Biology

BATL: Bayesian annotations for targeted lipidomics

Justin G. Chitpin [1,2,3], Graeme P. Taylor [2,3], Hongbin Xu [2,3], Thao T. Nguyen [2,3,4], Roberto Ortega [5], Julianna J. Tomlinson [6,7], Michael G. Schlossmacher [6,7], Rachel Saunders-Pullman [5], Steffany A. L. Bennett [2-4,6,7]* and Theodore J. Perkins [1,3,6]*

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

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

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 (LC-ESI-MS/MS) data acquired in either selected or multiple reaction monitoring (SRM and MRM) modes.

Results: We present here Bayesian Annotations for Targeted Lipidomics (BATL), 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 modelling 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 SRM datasets, we demonstrate over 95% of all peaks are rapidly and correctly identified.

Availability: The BATL software is available on GitHub at https://github.com/lipidomic-uottawa/batl

Contact: tperkins@ohri.ca and sbennet@uottawa.ca

Supplementary information: Supplementary data are available below.

1 Introduction

Targeted lipidomics employs liquid chromatography coupled to tandem mass spectrometry via electrospray ionization (LC-ESI-MS/MS). Using selected and multiple reaction monitoring (SRM and MRM) modes, pairs of parent 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 parent 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). These approaches have been used to successfully map fluid and cell-specific lipidomes (Quehenberger *et al.*, 2010; Slatter *et al.*, 2016; Sartain *et al.*, 2011), reveal lipidomic disruptions across biological conditions (Wang *et al.*, 2018; Alecu and Bennett, 2019), and predict changes in lipid metabolism associated with disease progression (Granger *et al.*, 2019; Alshehry *et al.*, 2016; Blasco *et al.*, 2017)

While the concept of targeting individual lipid species in SRM and MRM modes appears straightforward, identifying lipids from targeted lipidomic approaches remain a significant challenge, as exemplified in Figure 1. Multiple isobars, isomers, and isotopologues, sharing the same product ion, can elute in close proximity to the lipid target. Moreover,

picture(0,0)(-35,0)(1,0)30 (0,-35)(0,1)30 picture picture(0,0)(35,0)(-1,0)30 (0,-35)(0,1)30 picture

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.18.435788; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

picture(0,0)(-35,0)(1,0)30 (0,35)(0,-1)30 picture picture(0,0)(35,0)(-1,0)30 (0,35)(0,-1)30 picture

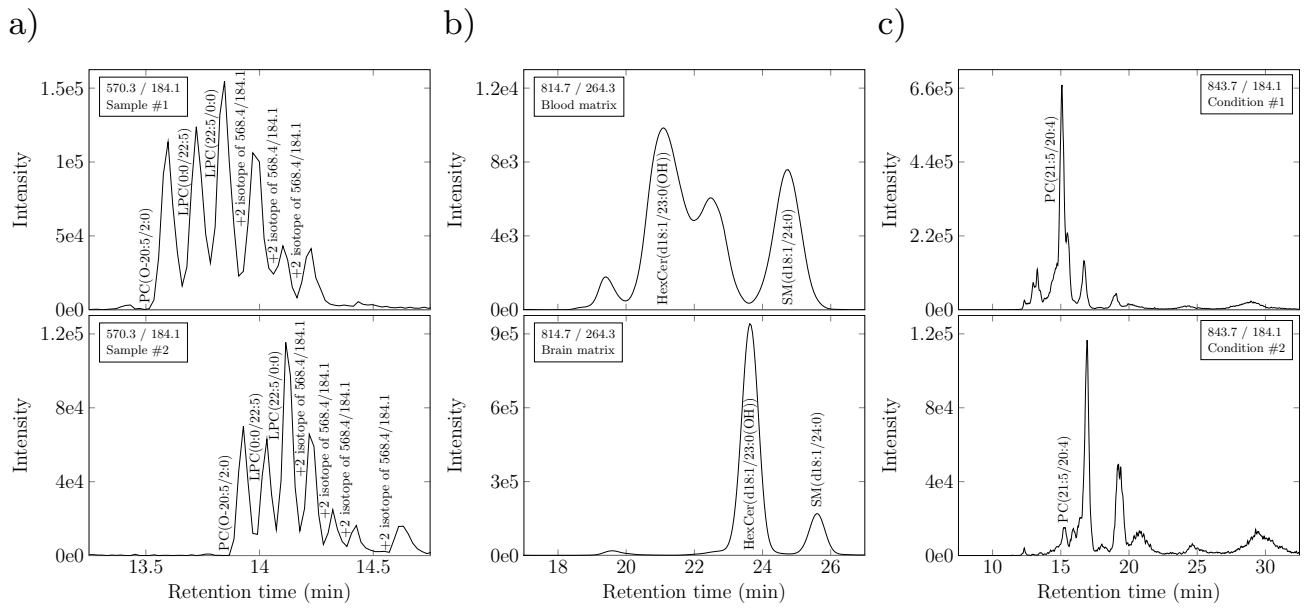

Fig. 1. Common challenges associated with SRM and MRM peak identification. a) Ambiguity occurs when multiple lipid isomers, isobars, and isotopes are detected within the same matrix at a given transition, yet technical variations in flow rate, composition of the mobile phase, temperature, pH, etc., cause their retention times to vary across samples. Data represent extracted ion chromatograms (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 one minute. Matching retention time would not align these shifted species. b) Assigning lipid identities based on peak elution order (picking the nth 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 of these differences manifest simultaneously. BATL addresses these challenges.

routine variations in chromatography can cause retention time shifts that align isobars or isotopologues with the species of interest in different MS runs (Smith *et al.*, 2015). As a result, it remains labour-intensive for different laboratories, and even within targeted lipidomic experiments from the same laboratory, to register, rapidly and reproducibly, large numbers of MS chromatograms. When multiple peaks are detected at a given transition, careful judgement is required to discriminate between lipid targets. These problems are magnified when researchers seek to match corresponding peaks and identify unique lipid species across lipidomes of different organisms or within different matrices.

Few programs have been developed to address the difficulties of SRM and MRM peak identification. Vendor-specific programs such as MultiQuant (SCIEX) 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. MultiQuant does 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.*, 2015). Additionally, Lipidyzer was designed to analyze data acquired specifically from SCIEX QTRAP 5500/5600 mass spectrometers with a SelexION DMS cell. 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). Free 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. In contrast, MRMPROBS (Tsugawa *et al.*, 2013) applies a multivariate logistic regression classifier, trained using five peak features, describing lipid retention, intensity, and shape. However, only retention time is used when classifying SRM peaks. Two further program restrictions are 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. Unless these 'new' lipids are explicitly added to the classifier dataset (i.e., raw MS data), compounds cannot be added to the MRMPROBS training set. With no means of updating the training set, MRMPROBS may misidentify lipid targets for closely eluting lipid isobars and isomers.

To address these issues, we applied a Bayesian annotation approach tailored to targeted lipidomics and present the program BATL which overcomes many of the limitations of the manual or template-based curation approaches and existing software packages. BATL is an R package applicable to any targeted lipidomics data collection mode from any LC-ESI-MS/MS platform. 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. We have developed BATL using eight specific features describing peak retention, intensity, and shape; however, BATL is trainable on any continuous feature supplied by the user. 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 is easily integrated into any lipidomics pipeline.

picture(0,0)(-35,0)(1,0)30 (0,-35)(0,1)30 picture picture(0,0)(35,0)(-1,0)30 (0,-35)(0,1)30 picture

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.18.435788; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

picture(0,0)(-35,0)(1,0)30 (0,35)(0,-1)30 picture picture(0,0)(35,0)(-1,0)30 (0,35)(0,-1)30 picture

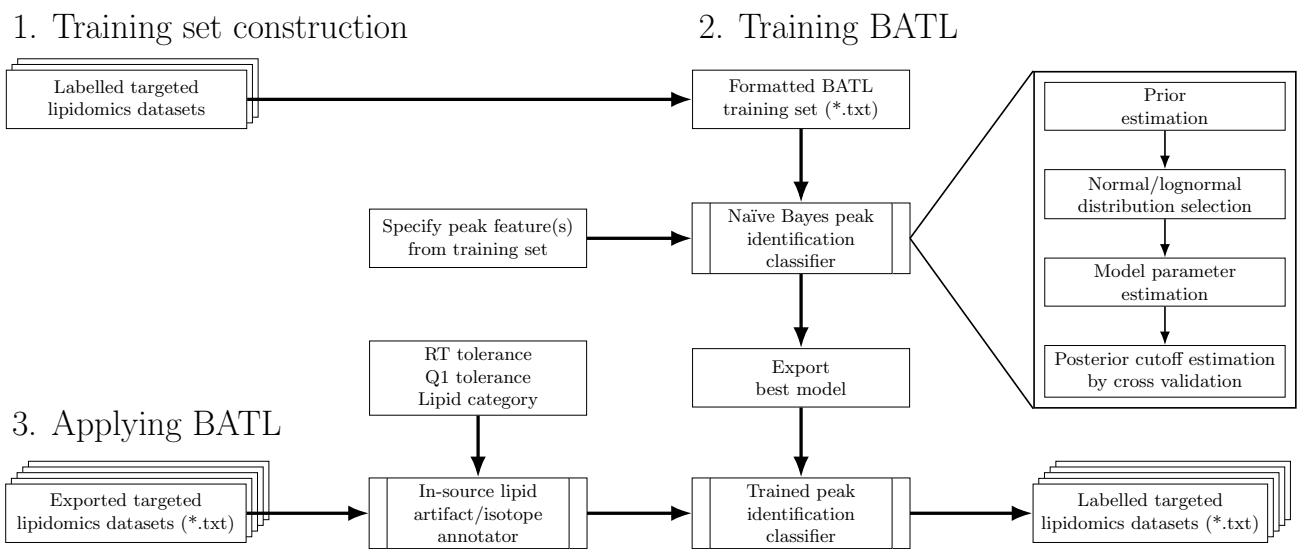

Fig. 2. Schematic of the BATL lipid identification workflow.

2 Methods

2.1 Overview of program

The BATL workflow is presented in Figure 2. First, a training set is constructed from user-labelled, 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 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 as well as sphingolipid-specific artifacts (e.g. dehydrations, deglycosylations, 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 $P_1, P_2, \ldots, P_m$ be a list of peaks within a sample described by feature vectors $F_1, F_2, \ldots, F_m$. 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 parent 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 parent 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.

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

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

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

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

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

To compute this joint probability, we make three assumptions: 1) the prior probabilities of all lipid identities are independent; 2) the peak feature vectors are statistically independent, conditional on the identities; and 3) the individual features within each vector are statistically independent, conditional on the peak identity. Thus, Equation 3 is simplified to the following probability.

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

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

$$\sum_{i=1}^{m} w_{ib}, \quad (5)$$

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

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

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 Results). 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 Results). 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.

picture(0,0)(-35,0)(1,0)30 (0,-35)(0,1)30 picture picture(0,0)(35,0)(-1,0)30 (0,-35)(0,1)30 picture

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.18.435788; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

picture(0,0)(-35,0)(1,0)30 (0,35)(0,-1)30 picture picture(0,0)(35,0)(-1,0)30 (0,35)(0,-1)30 picture

For every sample transition, a bipartite graph was constructed where the vertices represents 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 may 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 Results).

2.3 Training the model

Let $D = \{D_1, D_2, \ldots, D_p\}$ denote the labelled 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

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

where N_{B_o} 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_o^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

$$\Pr(f_{ij}|I_i) = \begin{cases} \log \mathcal{N}(f_{ij}|\mu_{I_i}, \sigma_{I_i}), & N_j^{\log} > N_j \\ \mathcal{N}(f_{ij}|\mu_{I_i}, \sigma_{I_i}), & \text{otherwise.} \end{cases} \tag{8}$$

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 labelled sphingolipid and glycerophosphocholine datasets composed of 972 MS spectra generated at the India Taylor Neurolipidomics Research Platform, University of Ottawa. To ensure all of the challenges in SRM identification outlined in Figure 1 were recapitulated in these datasets, we used: (1) 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), (2) a genotype and intervention comparison study of lipid metabolism in the temporal cortex, hippocampus, and plasma of wildtype and N5 TgCRND8 mice, a sexually dimorphic mouse model of Alzheimer's Disease (Granger *et al.*, 2016) (n=121 sphingolipid analyses; n=180 glycerophosphocholine analyses), and (3) a technical replicate study of two human plasma samples assessed in 33 sequential runs separated by blanks (n=33 sphingolipid analyses). To identify all lipids unambiguously, all molecular identities were confirmed by LC-SRM-information dependent acquisition (IDA)-enhanced product ion (EPI) experiments of samples pooled from either datasets 1-3 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. All lipids within the sphingolipid dataset were monitored at the same production 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 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 Data.

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.

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 as 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.

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

$$\text{Identification rate} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{10}$$

$$\text{Unassignment rate} = \frac{U}{\text{TP} + \text{FP} + U} \tag{11}$$

3 Results and discussion

BATL was trained on the sphingolipid and glycerophosphocholine training sets with unassigned identity weights w_{iu} learned by 10-fold cross validation and a parent/product ion tolerance of 0.5 m/z units. Models were constructed from every subset of features described in Table 1. These features described peak retention times, intensities, and shapes. Features were chosen because all, except subtracted retention time, are standard outputs of targeted lipidomic peak-picking software programs (i.e., MultiQuant, version 3.02, SCIEX). For each model, lipid identities

picture(0,0)(-35,0)(1,0)30 (0,-35)(0,1)30 picture picture(0,0)(35,0)(-1,0)30 (0,-35)(0,1)30 picture

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.18.435788; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

picture(0,0)(-35,0)(1,0)30 (0,35)(0,-1)30 picture picture(0,0)(35,0)(-1,0)30 (0,35)(0,-1)30 picture

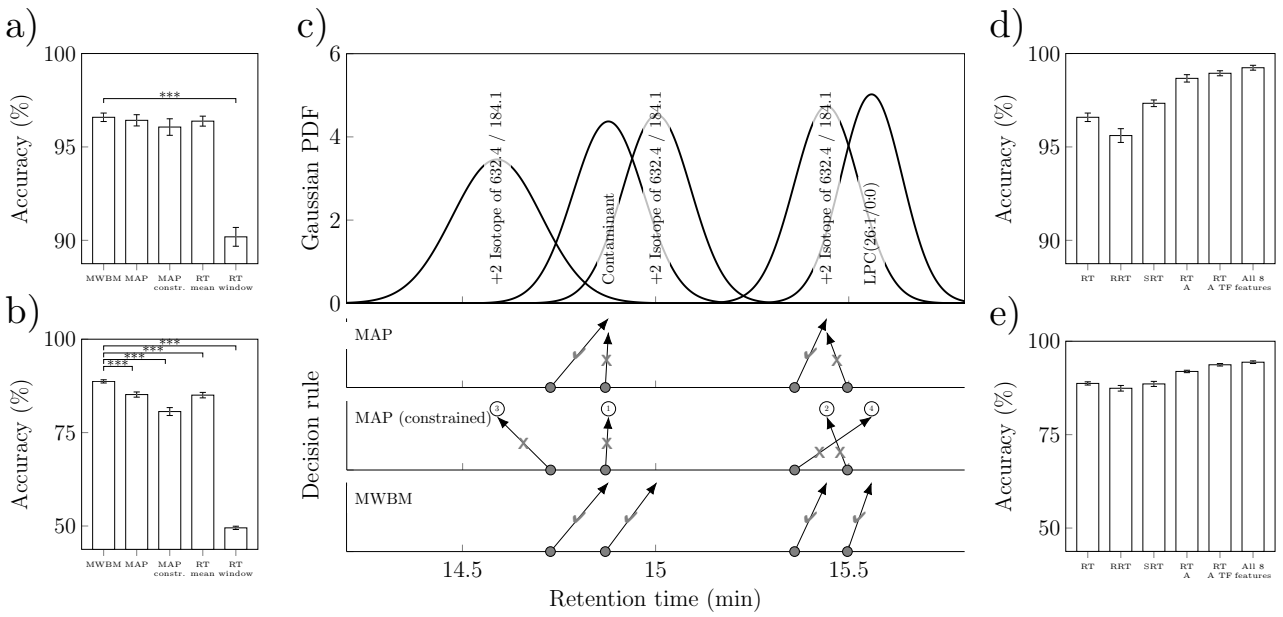

Fig. 3. Classifier performance on 10-fold cross validation sphingolipid and glycerophosphocholine datasets. 95% confidence intervals shown in panels a-b, d-e). a-b) Data represent mean accuracies of BATL models trained on retention time with each decision rule and retention time mean/window matching algorithms ($^{***}Q < 0.001$, 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 grey 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; Xs indicate incorrect assignments. The numbers for constrained MAP indicate the order of peak assignments. d-e) Data represent mean accuracies of the BATL models using MWBM decision rule trained on several features and feature combinations. The feature name codes are described in Table 1.

were assigned to peaks in the cross validation or holdout sets using either 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 identity 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 Figure 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 to four lipid isobars at the same transition in the training set (Supplementary Table S1-2). 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 Table S3-4). The BATL model, using the MWBM decision rule, achieved an 88.7% accuracy and significantly outperformed every other method (Supplementary Figure S1d-f). Performances were recapitulated when analyzing the holdout sets (Supplementary Information S1), and similar increases in accuracy were also observed when comparing decision rules of models trained using other feature subsets (See Supplementary Figures S2 and S3).

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

picture(0,0)(-35,0)(1,0)30 (0,-35)(0,1)30 picture picture(0,0)(35,0)(-1,0)30 (0,-35)(0,1)30 picture

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.18.435788; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

picture(0,0)(-35,0)(1,0)30 (0,35)(0,-1)30 picture picture(0,0)(35,0)(-1,0)30 (0,35)(0,-1)30 picture

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-e) show 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 Figure 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 Figure 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 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 revealed equal to or significantly worse identification rates between relative-versus-regular retention time (Supplementary Figure 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. 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 25 glycerophosphocholine samples. Whether trained on the less complex sphingolipid dataset or the more complex glycerophosphocholine dataset, identification rates decreased by less than 1% and unassignment rates remained under 5% when training on 10% of samples. These data demonstrate that only a small training set (i.e., 22-25 samples) is required to train the naïve Bayes model for accurate peak identification.

How does BATL compare against state-of-the-art methods for peak classification? Directly benchmarking BATL against other programs 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 identities assuming the correct peak was picked in a transition with multiple isobars. As a result, for all programs except MRMPROBS, it is not possible to separate peak detection accuracy from peak identification accuracy. 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. It was thus impossible to apply MRMPROBS to the sphingolipid or glycerophosphocholine holdout sets as they contained different number of peaks 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. This solution was cumbersome to apply when lipids were specific to the biological condition or matrix of individual samples. To address these challenges, BATL and MRMPROBS were benchmarked using 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 cutoff 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 fifteen second 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 Conclusions

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 unassignment rates remained stable when BATL was trained on as few as 22-25 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 program. BATL correctly identified more peaks than either approach with lower unassignment rates and no limitations regarding the number of lipids in the training set.

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. Learning the posterior probability cutoffs is simple to compute based on the naïve Bayes assumption, taking less than ten minutes to train each model on the sphingolipid and glycerophosphocholine training sets analyzed in this study using an Intel i5-8350U mobile processor. Further reductions in classifier training times on larger training sets are now possible with support for parallel processing. 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 parent and product ion pairs including MRM, neutral loss, precursor ion scan, and product ion scan acquisition modes. 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 SALB and RGPIN-2019-06604 to TJP from the Natural Sciences and Engineering Research Council of Canada (NSERC), as well as an NSERC

picture(0,0)(-35,0)(1,0)30 (0,-35)(0,1)30 picture picture(0,0)(35,0)(-1,0)30 (0,-35)(0,1)30 picture

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.18.435788; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

picture(0,0)(-35,0)(1,0)30 (0,35)(0,-1)30 picture picture(0,0)(35,0)(-1,0)30 (0,35)(0,-1)30 picture

CREATE Matrix Metabolomics Training grant to SALB and TJP. TTN and JGC received an NSERC CREATE Matrix Metabolomics Scholarship. JGC was further supported by an NSERC Alexander Graham Bell Canada Graduate Scholarship and an Ontario Graduate Scholarship. MGS 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 α-synucleinopathy in Parkinson's Disease. *Front. Neurosci.*, **13**, 328.

Alshehry, H., Z. *et al.* (2016). Plasma lipidomic profiles improve on traditional risk factors for the prediction of cardiovascular events in Type 2 Diabetes Mellitus. *Circulation*, **134**(21), 1637–1650.

Blasco, H. *et al.* (2017). Lipidomics reveals cerebrospinal-fluid signatures of ALS. *Sci. Rep.*, **7**(1), 1–10.

Bowden, J. *et al.* (2018). NIST lipidomics workflow questionnaire: An assessment of community-wide methodologies and perspectives. *Metabolomics*, **14**(5), 1–11.

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**(12), 2275–2288.

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**(9), 681–684.

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

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**(4), 499–517.

Murphy, R. C. and Axelsen, P. H. (2011). Mass spectrometric analysis of long-chain lipids. *Mass Spectrom. Rev.*, **30**(4), 579–599.

Quehenberger, O. *et al.* (2010). Lipidomics reveals a remarkable diversity of lipids in human plasma. *J. Lipid Res.*, **51**(11), 3299–3305.

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**(5), 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**(5), 930–944.

Smith, R. *et al.* (2015). LC-MS alignment in theory and practice: A comprehensive algorithmic review. *Brief. Bioinform.*, **16**(1), 104–117.

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**(10), 5191–5199.

Ubhi, Baljit K. amd Connor, A. *et al.* (2015). A novel lipid screening platform that provides a complete solution for lipidomics research. *SCIEX Technical Application Note*.

Wang, W. *et al.* (2018). Lipidomic profiling reveals soluble epoxide hydrolase as a therapeutic target of obesity-induced colonic inflammation. *Proc. Natl. Acad. Sci. U. S. A.*, **115**(20), 5283–5288.

picture(0,0)(-35,0)(1,0)30 (0,-35)(0,1)30 picture picture(0,0)(35,0)(-1,0)30 (0,-35)(0,1)30 picture

BATL: Bayesian annotations for targeted lipidomics Supplementary information

Justin G. Chitpin[1,2,3], Graeme P. Taylor[2,3], Hongbin Xu[2,3], Thao T. Nguyen[2,3,4], Roberto Ortego[5], Julianna J. Tomlinson[6,7], Michael G. Schlossmacher[6,7], Rachel Saunders-Pullman[5], Steffany A. L. Bennett[2-4,6,7*], and Theodore J. Perkins[1,3,6*]

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

Contact: tperkins@ohri.ca and sbennet@uottawa.ca

Contents

1. Experimental protocols 2
2. Dataset complexity 4
3. Classifier holdout performance between decision rules 6
4. Classifier cross validation performance between decision rules for other feature combinations 7
5. Classifier holdout performance between selected features 9
6. Classifier performance by best *N* feature combinations 10
7. Classifier performance by single feature 11
8. Classifier performance by training set size 12
References 13

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=319). 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. 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. 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

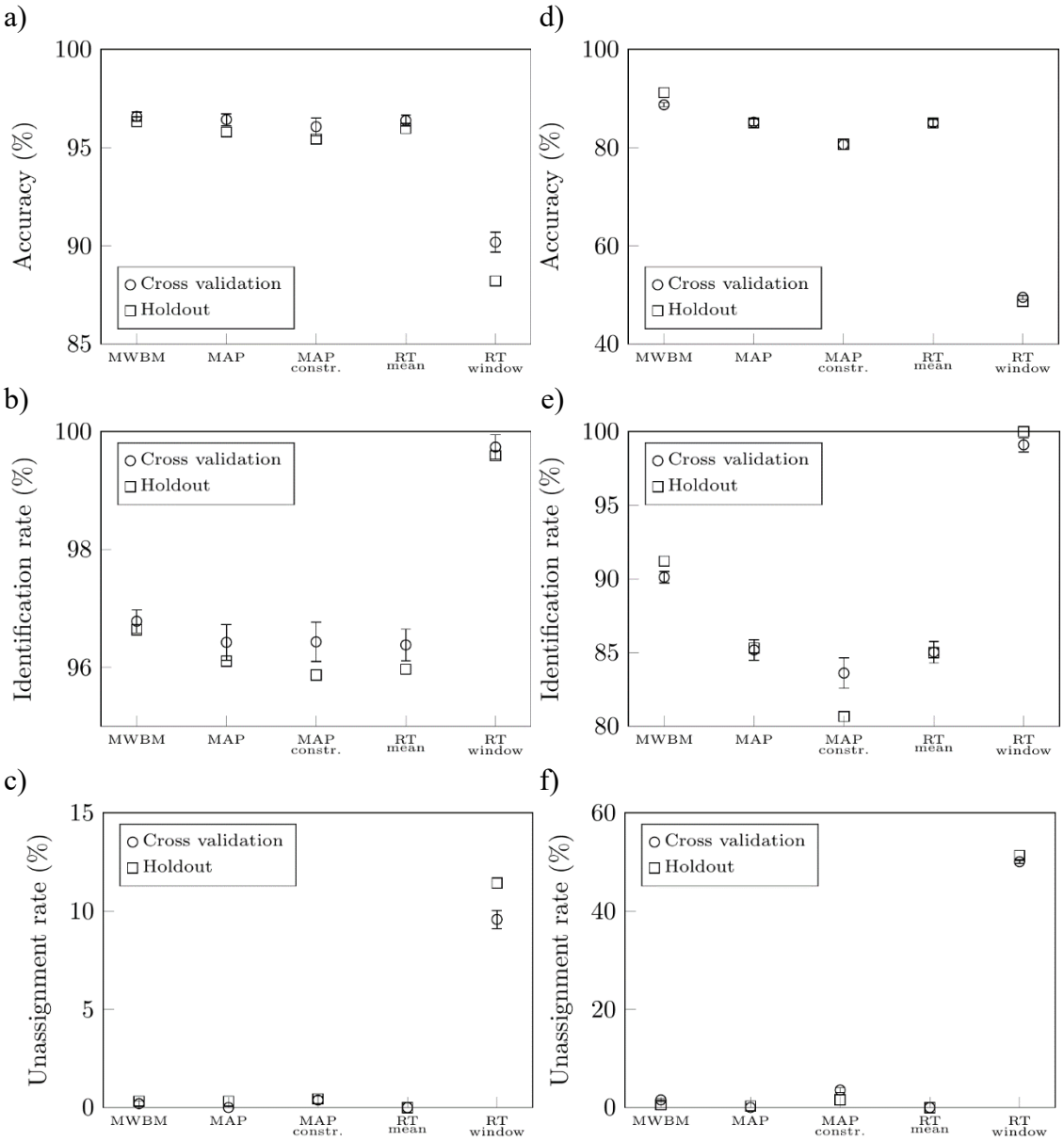

Figure S1: Detailed classifier performance on holdout and cross validation sets. For cross-validation results, point markers indicate mean accuracies/identification/unassignment rates and whiskers represent 95% confidence intervals across 10 folds. a-c) Accuracies, identification rates, and unassignment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

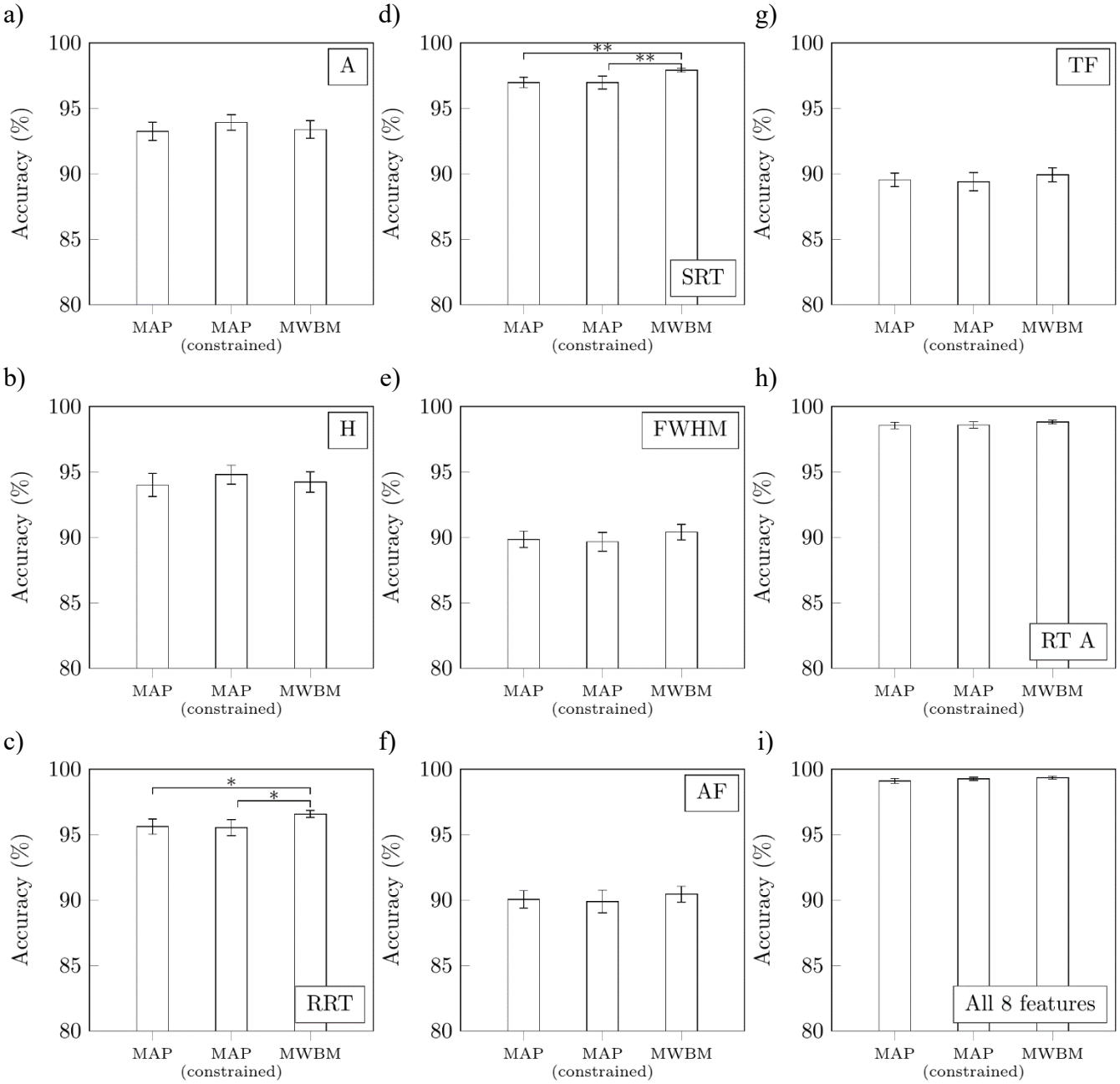

Figure S2: Classifier performance on 10-fold cross validation sphingolipid datasets. Data represent mean accuracies of BATL models trained on select feature combinations. 95% confidence intervals shown (*Q < 0.05, **Q < 0.01, ***Q < 0.001, t-test adjusted with the Benjamini-Hochberg method).

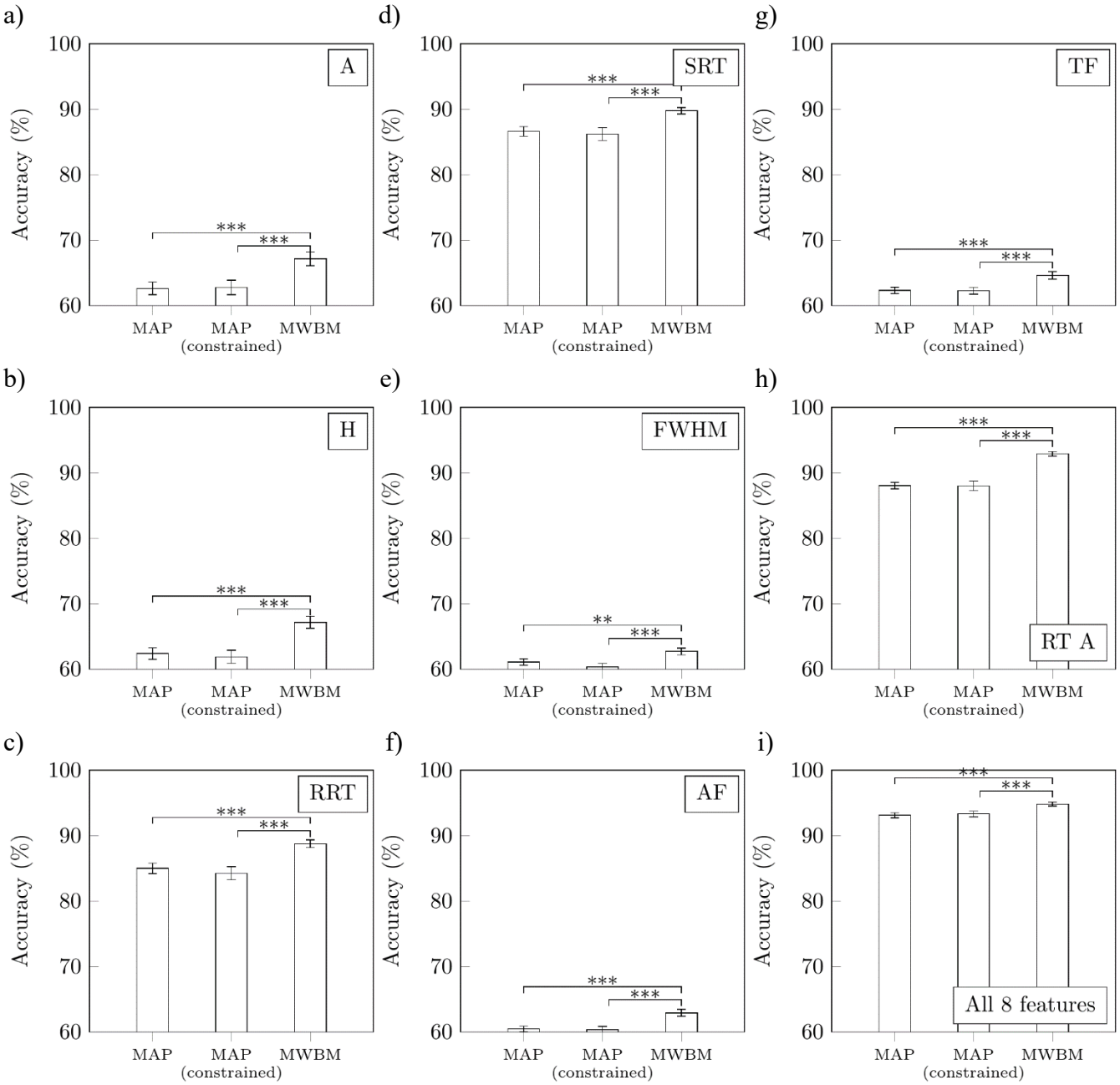

Figure S3: Classifier performance on 10-fold cross validation glycerophosphocholine datasets trained with select feature combinations. Data represent mean accuracies of BATL models trained on select feature combinations. 95% confidence intervals shown (*$Q < 0.05$, **$Q < 0.01$, ***$Q < 0.001$, *t*-test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

5. Classifier holdout performance between selected features

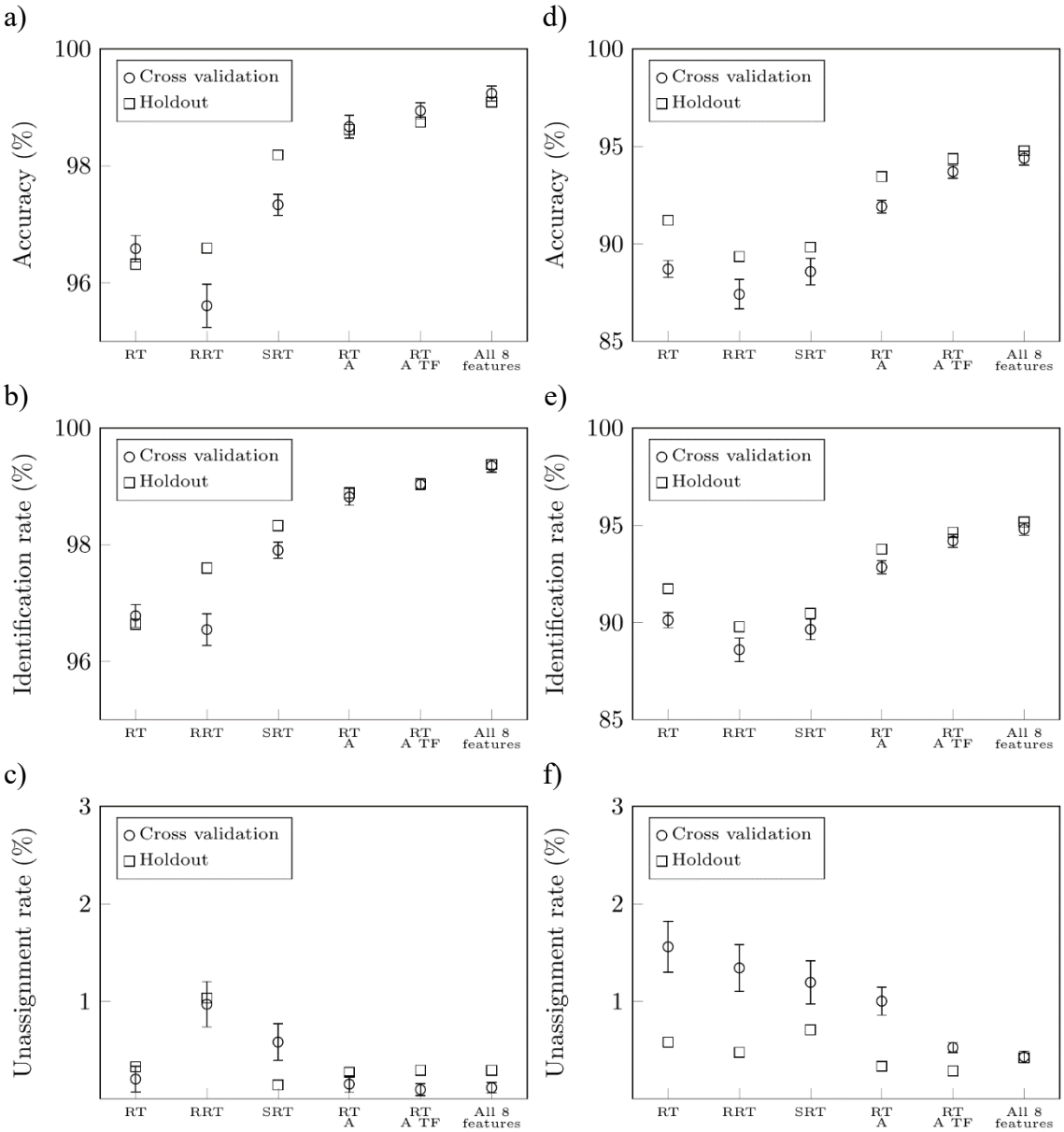

Figure S4: Classifier performance on holdout and cross validation sets trained with select feature combinations. For cross-validation results, point markers indicate mean accuracies/identification /unassignment rates and whiskers represent 95% confidence intervals across 10 folds. a-c) Accuracies, identification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

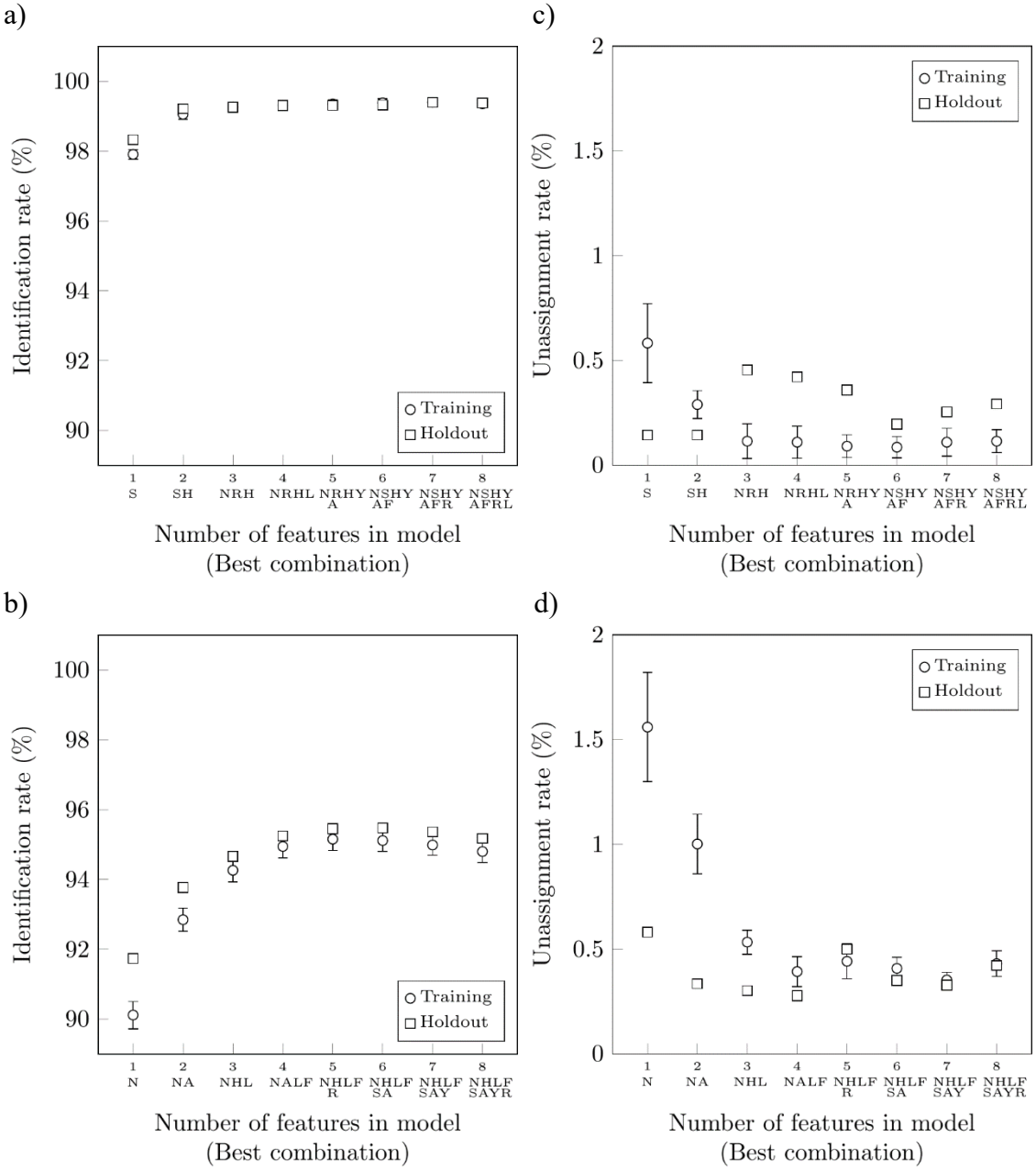

Figure S5: Models using the MWBM decision rule trained from the best combination of *N* features. X-label feature codes described in Figure S6. 95% confidence intervals for the 10-fold cross validation datasets shown. a-b) Sphingolipid and glycerophosphocholine dataset identification accuracies. c-d)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

7. Classifier performance by single feature

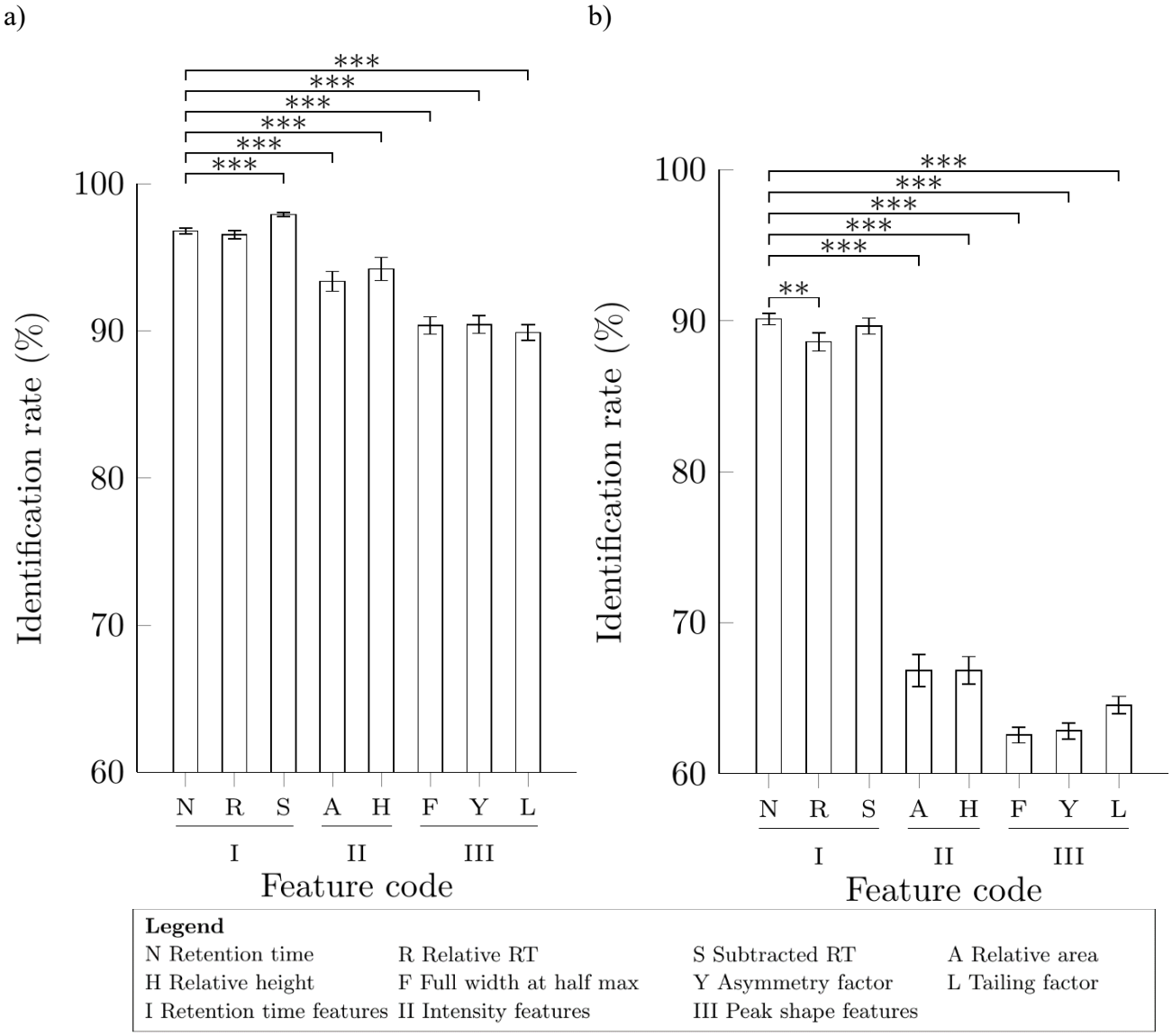

Figure S6: Comparison of models using the MWBM decision rule trained with retention time (N) versus all other single features. Ten-fold cross validation identification rates with 95% confidence intervals shown (pairwise *t*-test against the retention time feature adjusted with the Benjamini-Hochberg 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

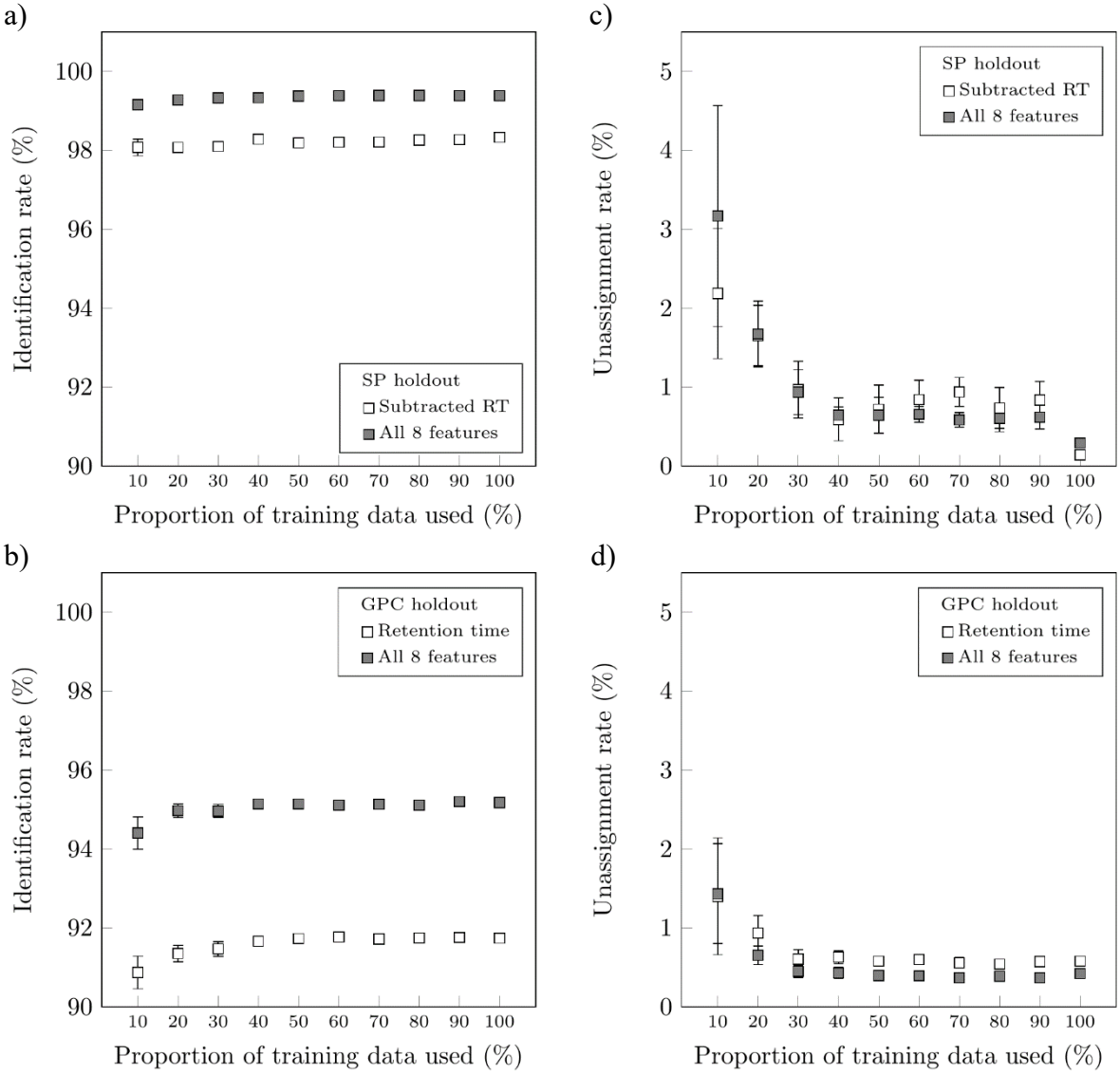

Figure S7: Classifier performance on holdout datasets with the best single feature or all features resampled ten times for each proportion of training data. Every 10% increment corresponds to 22 sphingolipid or 25 glycerophosphocholine samples. a-b) Point markers indicate mean identification rates and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