analytical chemistry

Article

Cite This: *Anal. Chem.* 2018, 90, 12198−12206

pubs.acs.org/ac

Combining Desorption Electrospray Ionization Mass Spectrometry Imaging and Machine Learning for Molecular Recognition of Myocardial Infarction

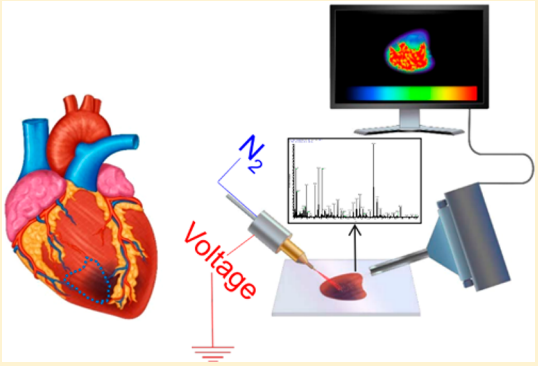
Katherine Margulis,[†] Zhenpeng Zhou,[†] Qizhi Fang,[‡] Richard E. Sievers,[‡] Randall J. Lee,[‡] and Richard N. Zare*[,†]

[†]Department of Chemistry, Stanford University, Stanford, California 94305, United States

[‡]Cardiovascular Research Institute and Department of Medicine, University of California San Francisco, San Francisco, California 94131, United States

Supporting Information

ABSTRACT: Lipid profile changes in heart muscle have been previously linked to cardiac ischemia and myocardial infarction, but the spatial distribution of lipids and metabolites in ischemic heart remains to be fully investigated. We performed desorption electrospray ionization mass spectrometry imaging of hearts from *in vivo* myocardial infarction mouse models. In these mice, myocardial ischemia was induced by blood supply restriction via a permanent ligation of left anterior descending coronary artery. We showed that applying the machine learning algorithm of gradient boosting tree ensemble to the ambient mass spectrometry imaging data allows us to distinguish segments of infarcted myocardium from normally perfused hearts on a pixel by pixel basis. The machine learning algorithm selected 62 molecular ion peaks important for classification of each 200 μm-diameter pixel of the cardiac tissue map as normally perfused or ischemic. This approach achieved very high average accuracy (97.4%), recall (95.8%), and precision (96.8%) at a spatial resolution of ∼200 μm. In addition, we determined the chemical identity of 27 species, mostly small metabolites and lipids, selected by the algorithm as the most significant for cardiac pathology classification. This molecular signature of myocardial infarction may provide new mechanistic insights into cardiac ischemia, assist with infarct size assessment, and point toward novel therapeutic interventions.



Ischemic heart disease is a leading cause of death worldwide,[1] whereas acute myocardial infarction (MI, or heart attack) is the first manifestation of ischemic heart disease in approximately 50−70% of patients.[2] Silent myocardial infarctions are heart attacks with no typical symptom manifestation but with similar clinical significance,[3] and they account for almost 50% of incident MIs.[1] Both symptomatic and silent MIs can go undetected by an electrocardiogram;[4,5] hence, ideally, a simple chemical screening should be established to detect the presence of infarcted myocardial tissue (myocyte necrosis). Clinical tests that are currently being employed for MI diagnostics detect elevated levels of a few protein biomarkers in blood, mostly cardiac troponin and creatine kinase.[5] The levels of these proteins rise rapidly following infarction onset but also return to baseline within several days, making later diagnostics impossible.[5] Uncovering additional biomolecules whose levels are perturbed as a result of MI may potentially improve our long-term diagnostic capabilities and may allow prediction of residual cardiac tissue viability. Moreover, identifying the molecular signature of infarction in affected myocardium may provide new mechanistic insights into cardiac ischemia and point toward therapeutic intervention to limit viable tissue loss and heart remodeling following MI.[6,7]

In previous works, lipid profile changes were shown to have a strong correlation to cardiac ischemia and ischemic reperfusion injury.[8−14] The mammalian heart acquires lipids both from circulating free fatty acids (FAs) and esterified FAs bound to lipoproteins.[15] Being the most energy-requiring organ of the body, the heart heavily relies on fatty acid oxidation for energy metabolism.[15] During ischemia, the heart switches to ATP generation from glucose through excessive glycolysis, which drastically modifies the lipid and small metabolite profile in the myocardium.[15] However, owing to the complexity of molecular events accompanying cardiac ischemia, variability in individual lipid marker levels, and inability to directly monitor myocardial metabolic modifica-

Received: July 30, 2018
Accepted: September 6, 2018
Published: September 6, 2018

DOI: 10.1021/acs.analchem.8b03410
Anal. Chem. 2018, 90, 12198−12206

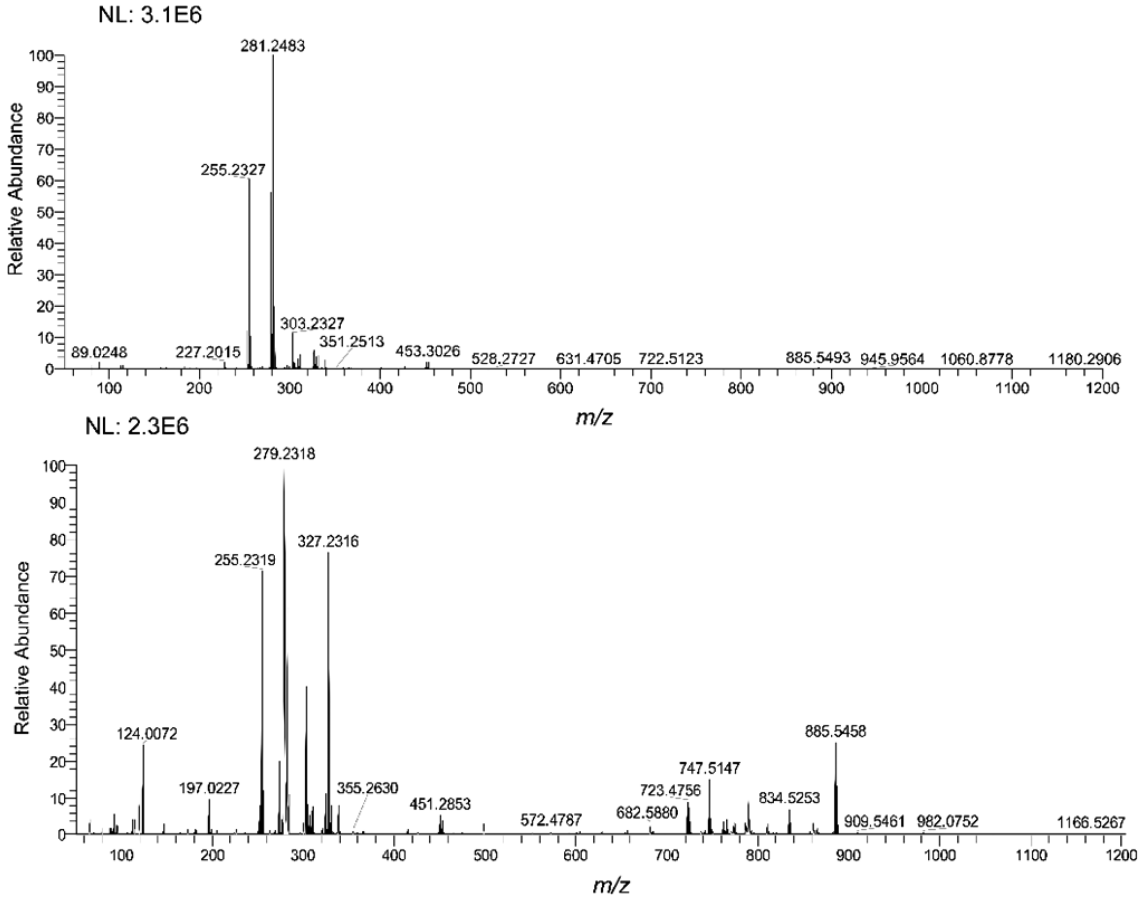


Figure 1. Representative mass spectra of full infarct region (upper spectrum) and perfused myocardium (lower spectrum).

tions, no specific lipid biomarkers of cardiac ischemia are currently in clinical use.[16]

Mass spectrometry (MS) holds much promise as a tool for MI biomarker interrogation, including lipids and metabolites. Recently, GC/MS and LC/MS-based metabolite profiling was used to elucidate possible markers in post-infarction serum, cardiac tissues, and cardiac tissue homogenates,[17−19] alongside with the more established MS proteomics profiling of MI.[20,21] A pioneering study of spatially resolved MS imaging analysis of MI-bearing cardiac tissue sections was performed a few years ago using matrix-assisted laser desorption/ionization (MALDI) MS imaging.[7] This study allowed delineation of infarcted tissue from normal myocardium in a MI rat model. Specifically, imaging in positive ion mode uncovered an increase in ion signals from lysophospholipids and a decrease in the ion signals from intact phospholipids in the infarcted area, consistent with known accelerated activity of the phospholipase A2 enzyme in MI. Creatinine abundance decrease was also detected in infarcted areas owing to the release of creatine kinase enzyme following MI.[7]

In the present study, we employed an ambient condition imaging method, desorption electrospray ionization mass spectrometry imaging (DESI-MSI),[22−29] operating in negative ion mode to study the spatial distribution of various lipids and metabolites in murine ischemic heart muscle vs adjacent perfused myocardium. DESI-MSI enables a high-throughput monitoring of lipids and metabolites with very high sensitivity, because the mass-to-charge (m/z) ratios of the analytes lie within the optimal detection range of this method. Briefly, a beam of charged, solvent droplets is directed onto the myocardium surface to desorb and ionize molecules, while the splash of these droplets carries the resultant ions into a mass spectrometer for analysis. A 2D imaging stage moves the myocardium section at a controlled speed to scan the specimen, while the mass spectra are recorded as a function of x,y-position on the tissue, allowing this information to be subsequently converted into 2D images of molecular ion distributions. Thousands of molecular species, whose absolute and relative intensities change according to myocardium pathology, are detectable from the specimen within each run at spatial resolution of ∼200 μm, and a 2D distribution image with relative signal intensity can be generated for any of those species.

The immense amount of chemical information obtained in the process of imaging invites employing machine-learning techniques to develop algorithms capable of predicting myocardial pathology based on mass spectral fingerprints. In this study, we explored the ability of the machine learning algorithm, gradient boosting tree ensemble,[30] to process imaging data and select features (mass spectral peak patterns) discriminative between normally perfused and infarcted tissue. We recently reported developing a machine learning technique to categorize mass spectra obtained from the sweat of human volunteers.[30] In this present study, we extended this machine learning technique to imaging data by including the spatial localization of each pixel, which enabled us to use the information obtained from adjacent pixels to make a better prediction. We further compared the performance of the algorithm we established with other classification algorithms frequently used in mass spectrometry. The developed

DOI: 10.1021/acs.analchem.8b03410
Anal. Chem. 2018, 90, 12198−12206

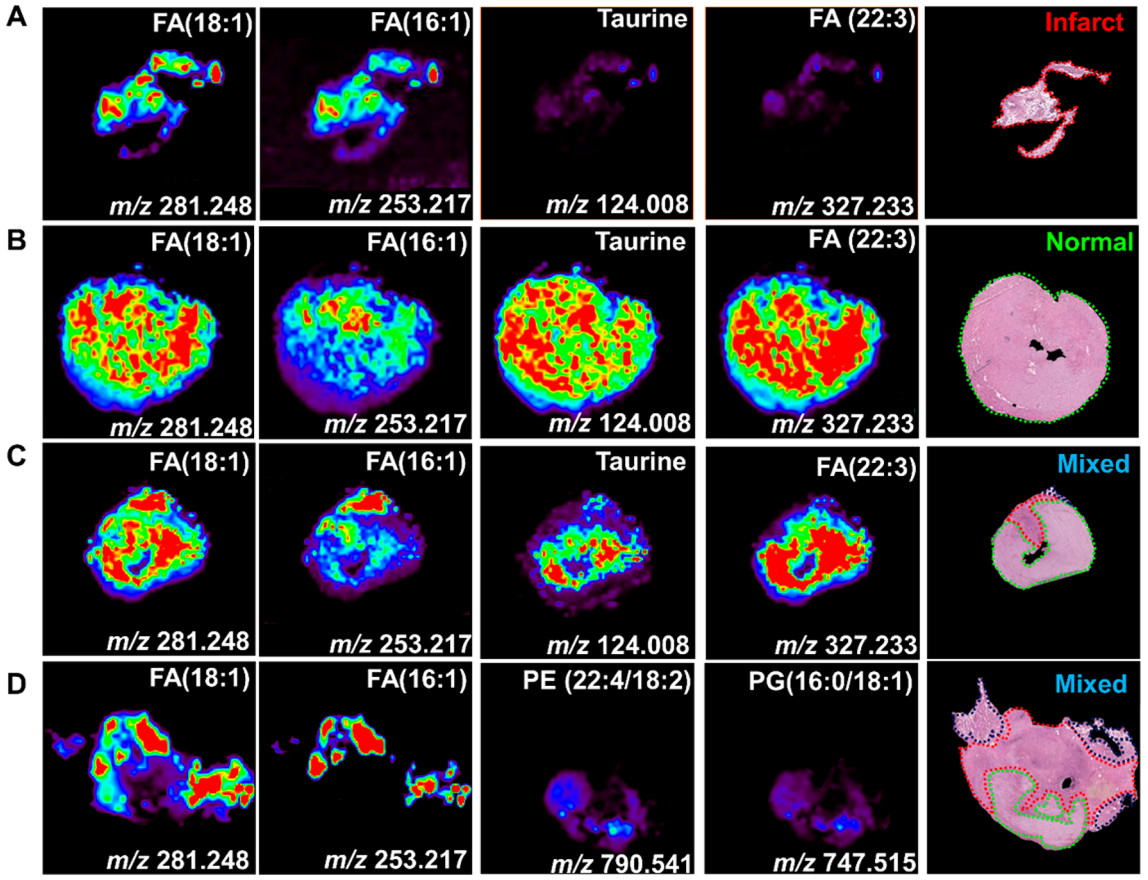


Figure 2. 2D distribution images of various molecular ions in myocardial sections throughout regions with various pathologies: (A) infarct section, (B) normal myocardium section, and (C and D) sections with mixed pathology. Histological staining is shown in the far-right column, whereas infarcted regions are delineated in red; normal myocardium is marked in green, and epicardium tissue is marked in dark blue.

algorithm was applied to mass spectra extracted from 2D ion distribution images in accordance with pathological evaluation of each 200-μm diameter picture element (pixel). This feature selection combined with tandem mass spectrometry allowed us to uncover ion species important for molecular recognition of cardiac ischemia vs normally perfused myocardium.

RESULTS

Animal Models of MI. Controlled apical MIs were successfully induced in mice by a permanent ligation of the left anterior descending coronary artery (LAD). The LAD was ligated with one single stitch, forming immediate ischemia in a confined region at heart apex while the surrounding myocardial tissues were nearly unaffected.[31] This surgical procedure resulted in controlled heart attacks imitating well the pathophysiological processes occurring in infarction-related myocardial ischemia.[31] Post-sacrifice, the infarct locations were macroscopically confirmed, and harvested hearts were cryosectioned to slices that included apical sections (from the infarct region), sections from the infarct border, and sections from the normal myocardium distant from the ischemic area (three sections from each region for each mouse).

Molecular Imaging. For the construction of 2D imaging maps, the sections of myocardium were imaged in a negative ion mode, in a m/z range of 50−1200. In this range a wide variety of molecular ions can be detected, including amino acids and small metabolites (m/z 50−200), free fatty acids (FAs) (m/z 200−400), and complex glycerophospholipids (m/z 700−1000). The sections were histologically stained after imaging using hematoxylin and eosin (H&E) staining protocol for unfixed tissue.[32] Striking differences between MS spectra of infarcted and normally perfused myocardium were observed (Figure 1). Specifically, infarcted myocardium had diminished relative and absolute abundances of glycerophospholipids (m/z 700−1000), and elevated relative abundances of monounsaturated and some saturated fatty acids (FAs) (Figures 1 and 2). On the other hand, the abundances of polyunsaturated FAs and some small metabolites were markedly decreased in infarct. Thus, the most abundant ion in the infarct spectra was identified as the deprotonated oleic acid, (FA(18:1), at m/z 281.248). Prominent peaks of deprotonated palmitoleic (FA(16:1), at m/z 253. 217), stearic (FA(18:0), at m/z 283.264), and palmitic (FA(16:0), at m/z 255.232) acids were also detected in the infarct region. Normally perfused myocardium presented prominent peaks of many glycerophospholipids, e.g., deprotonated phosphatidylinositol (PI), (PI(22:4/16:0), at m/z 885.546), phosphatidylglycerol (PG), (PG(16:0/18:1), at m/z 747.516), and phosphatidylethanolamine (PE), (PE(22:6/18:0), at m/z 790.538), as well as of polyunsaturated FAs, such as docosatrienoic acid (FA(22:6), at m/z 327.232), and of small metabolites, such as the amino acid taurine at m/z 124.007. Though some changes in the spectra and corresponding 2D images could be detected by the naked eye (Figures 1 and 2), the immense amount of chemical information obtained

DOI: 10.1021/acs.analchem.8b03410
Anal. Chem. 2018, 90, 12198−12206

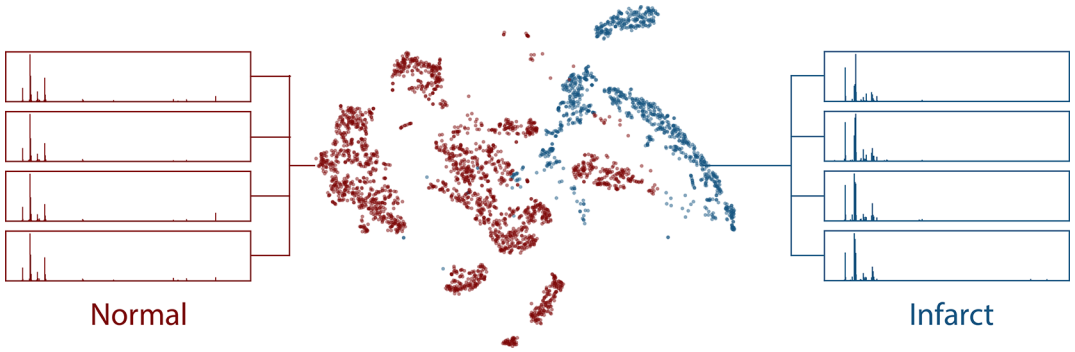


Figure 3. Visualization of the selected features by t-SNE embedding. The red-colored point clouds show perfused myocardium, while the blue-colored ones show infarct samples. The boxes show the mass spectra corresponding to sample points.

in the process of imaging required development of a machine learning algorithm to be adequately processed.

Machine Learning Model. A machine learning algorithm of gradient boosting tree ensemble (GBDT)[30] was developed to be applied to the mass spectra extracted from 2D images of the specimens using MSIReader software in accordance with each segment's pathology. This algorithm enabled classification of mass spectral features (peak intensities) originating from each pixel as characteristic of normally perfused heart or infarct segments. Instead of looking at all peaks available, we only focused on the peaks that appear in more than 0.1% of the samples in the data set, which allowed us to generalize our prediction model better. GBDT was also trained to localize each pixel and extract the information from adjacent pixels to improve its prediction abilities. This is based on the assumption that if all neighboring pixels are infarct, the center pixel is more likely to be infarct. Boosting is a technique that generates a prediction model in the form of an ensemble of weak predictors, in our case, decision trees. It combines these weak predictors into a single, strong prediction model in an iterative way. It uses an additive strategy: fixes what was learned and adds one new predictor at a time. The whole data set was separated into a training set, a cross-validation set, and a test set with ratio of 7:1:2. A discriminative prediction model was trained on the training set of randomly selected infarcted or normal regions and was optimized on the cross-validation set. The algorithm selected 62 peaks important for classification of each 200 μm-diameter pixel of heart image as normally perfused or infarcted myocardium. These peaks were assigned with a relative weight of their importance for discriminative prediction model and arranged in the descending order of their weight (Table S1). The difference between the mean relative abundance of each peak in the infarct and in the normal heart was also calculated (Table S1, mean difference column). This difference was negative if the molecular ion was more abundant in the normal myocardium and positive if its abundance was increased in infarct (Table S1). We visualized the features selected by our model using t-SNE (t-distributed stochastic neighbor embedding)[33] as shown in Figure 3. Each point represents a 2D projection from 62-dimensional vectors of selected features. We can see the clusters of normal and infarct points with an obvious decision boundary between them, demonstrating the effectiveness of our algorithm.

The performance of the trained model was evaluated on the test set. Overall using the developed machine learning algorithm resulted in a very high average accuracy of 97.43 ± 0.45% of correct diagnosis for each 200 μm-diameter pixel. High values were also obtained for the average precision and average recall (or sensitivity), which reflect the fraction of correctly diagnosed pixels out of all pixels with a certain diagnosis, and the fraction of correctly diagnosed pixels out of all pixels with a certain pathology, respectively (Table 1).

Table 1. Statistical Analysis on Pixel-to-Pixel Basis Using Developed GBDT Algorithm

prediction infarct vs normal for each pixel	%
average accuracy	97.43 ± 0.45
average recall	95.82 ± 1.46
average precision	94.84 ± 0.96

We compared the GBDT model described above with several other classification techniques or algorithms that have been previously used for data analysis in mass spectrometry, namely, principal component analysis (PCA)[34,35] combined with linear discriminant analysis (LDA),[36] support vector machines (SVM),[37,38] L1 regularized logistic regression,[39] recursive maximum margin criterion (RMMC),[40−44] and random forest (RF).[45,46] GBDT outperformed all other tested methods in average accuracy, average recall, and average precision when applied to our data (Table S2). Specifically, compared to the algorithm that shows the second highest accuracy (RF), the error rate using GBDT was decreased by 25%, which is significant for pathological diagnosis.

Peak Identification. We used collision-induced dissociation (CID) experiments and isotopic peak distribution for identification of molecular ions selected as important for myocardial pathology by GBDT algorithm. Twenty-six peaks with a relative importance greater than 0.01 were identified, whereas the most prevalent tandem MS fragmentation pattern was used to assign each peak's molecular identity. The identified molecular ions for each peak are listed in Table 2 and selected CID patterns with proposed fragments are shown in Figure 4. For saturated fatty acids and fatty acids with up to three unsaturated bonds, fragmentation was dominated by elimination of H_2O from the carboxyl moiety, whereas longer fatty acids bearing three or more double bonds, fragmented both by elimination of water and loss of CO_2 from the carboxyl moiety. Complex phospholipids showed characteristic losses of acyl chains as ketenes from the precursor anion, neutral losses of carboxylic acid chains from the anion, and neutral losses of both carboxylic acid chain and phosphate group substituent. We used isotopic distribution patterns (listed in the far-right

DOI: 10.1021/acs.analchem.8b03410
Anal. Chem. 2018, 90, 12198−12206

Table 2. High Mass Resolution, Isotopic Distribution, and Tandem Mass Spectrometry Data Used for Identification of Molecular Ions

observed m/z[a]	main fragment ions	ion attribution[b]	exact m/z	mass error (ppm)	proposed ion formula[c]	selected isotopic peaks[d]
124.0071	106.98, 79.96	taurine	124.0068	2.4	$C_2H_6NO_3S$	126.0028 (^{34}S)
253.2168	235.21	palmitoleic FA(16:1)	253.2168	0	$C_{16}H_{29}O_2$	
327.2323	283.24, 309.22	docosahexaenoic FA(22:6)	327.2324	−0.3	$C_{22}H_{31}O_2$	
281.2481	263.24	oleic FA(18:1)	281.2481	0	$C_{18}H_{33}O_2$	
282.2514		oleic FA(18:1) (^{13}C)	282.2514	0	$C_{18}H_{33}O_2$ (^{13}C)	
309.2793	265.29, 291.27	eicosenoic acid FA(20:1)	309.2794	−0.3	$C_{20}H_{37}O_2$	
303.2323	259.24, 285.22	arachidonic FA(20:4)	303.2324	−0.3	$C_{20}H_{31}O_2$	
790.5377	480.31, 327.23, 283.26	PE(22:6/18:0)	790.5386	−1.1	$C_{45}H_{77}NO_8P$	
329.2479	285.26, 311.24	docosapentaenoic FA(22:5)	329.2481	−0.6	$C_{22}H_{33}O_2$	
284.2671		stearic FA(18:0) (^{13}C)	284.2671	−0.6	$C_{18}H_{35}O_2$ (^{13}C)	
283.2637	265.25	stearic FA(18:0)	283.2637	0	$C_{18}H_{35}O_2$	
299.2010	255.21	retinoic acid	299.2011	−0.3	$C_{20}H_{27}O_2$	
304.2357		arachidonic acid FA(20:4) (^{13}C)	304.2357	0	$C_{20}H_{31}O_2$ (^{13}C)	
89.0241		lactic acid	89.0239	2.2	$C_3H_5O_3$	90.0277 (^{13}C)
328.2356		docosahexaenoic FA(22:6) (^{13}C)	328.2357	−0.3	$C_{22}H_{31}O_2$ (^{13}C)	
255.2325	237.22	palmitic FA(16:0)	255.2324	0.4	$C_{16}H_{31}O_2$	
331.2636	287.27, 313.25	docosatetraenoic acid FA (22:4)	331.2637	−0.3	$C_{22}H_{35}O_2$	
307.2637	289.25	eicosadienoic acid FA(20:2)	307.2637	0	$C_{20}H_{35}O_2$	
227.2013	209.19	myristic FA(14:0)	227.2011	0.9	$C_{14}H_{27}O_2$	
297.2429	279.23	hydroxyoleic acid	297.2430	−0.3	$C_{18}H_{33}O_3$	
277.2167	259.21	linolenic acid	277.2168	−0.4	$C_{18}H_{29}O_2$	
280.2358		linoleic FA(18:2) (^{13}C)	280.2357	0.4	$C_{18}H_{31}O_2$ (^{13}C)	
256.2358		palmitic FA(16:0) (^{13}C)	256.2357	0.4	$C_{16}H_{31}O_2$ (^{13}C)	
885.5476	599.32, 581.31, 439.22, 419.26, 303.23, 283.26	PI(18:0/20:4)	885.5493	−1.9	$C_{47}H_{81}O_{13}P$	
279.2324	261.22	linoleic FA(18:2)	279.2324	0	$C_{18}H_{31}O_2$	
305.2480	261.26, 287.24	dihomo-γ-linolenic FA(20:3)	305.2481	−0.3	$C_{20}H_{33}O_2$	
		importance < 0.01				
747.5158	483.28, 391.22, 281.25, 255.23	PG(16:0/18:1)	747.5176	−2.4	$C_{40}H_{76}O_{10}P$	

[a]High mass resolution analysis was performed with the Orbitrap mass analyzer. [b]FA = fatty acid; PE = glycerophosphoethanolamine; PI = glycerophosphoinositol; PG = glycerophosphoglycerol. (*X*:*Y*) denotes the total number of carbons and double bonds in the fatty acid chains, respectively. The most abundant isomer based on the fragments is listed. [c]Proposed formula for the ion detected. [d]Selected isotopic peaks listed in this column are important for molecular ion characterization. All ions have abundant isotopic peaks such as ^{13}C peaks.

column of Table 2) to prove identification of species at m/z 124.0071 as taurine, based on the presence of ^{34}S-$[M]^-$ isotopic peak, and identification of species at m/z 89.0241 as lactate, based on ^{13}C-$[M]^-$ distribution characteristic of singly charged lactic acid anion.[28] However, ^{13}C-$[M]^-$ isotopic peaks that were chosen by GBDT as important for myocardial pathology classification were listed in Table 2 as separate features.

DISCUSSION

Despite documented dysregulation of lipid levels following myocardial infarction, there is no routine use of lipid and small metabolites as cardiac pathology predictors. We show that applying a machine learning algorithm of GBDT on imaging data collected by DESI-MSI allows selection of 62 ion signals important for cardiac pathology detection and predicts with high accuracy (97.4%) the location of infarcts in murine heart tissues. GBDT showed a superior performance when applied to our imaging data compared with several other classification algorithms. Molecular ions selected by GBDT mostly originate from lipids and small metabolites, and some of these molecules can also shed light on pathophysiological processes occurring within ischemic heart. DESI-MSI/GBDT revealed, for example, that the small amino acid taurine is more abundant in normally perfused myocardium than in infarct. Taurine is a physiologically ubiquitous sulfur-containing amino acid that does not participate in peptide bond formation.[47] Its concentration in mammalian hearts is higher than that in most other tissues,[48] and it is easily detected in whole blood.[49] Taurine plays an important role as mitochondrial matrix buffer for stabilization of mitochondrial oxidation.[50] A growing body of evidence points out that taurine depletion in diet leads to the development of a cardiomyopathy, whereas its supplementation attenuates cardiac remodeling after myocardial infarction and reduces generation of reactive oxygen species within cardiomyocytes,[51,52] suggesting taurine's protective function.[47,51] Other animal studies have shown, however, that taurine is released from the heart during episodes of cardiac ischemia to prevent Ca^{2+} overload and minimize cardiac damage[52] and is elevated in blood following ischemia.[49] Moreover, detection of elevated levels of taurine in blood of patients with an acute cardiac pain suggests the diagnosis of either myocardial infarction or unstable angina pectoris.[49,53,54] The concentration of taurine may help to differentiate between the two conditions.[49] The taurine in blood is elevated probably owing to its release from the infarcted heart muscle. In our

DOI: 10.1021/acs.analchem.8b03410
Anal. Chem. 2018, 90, 12198−12206

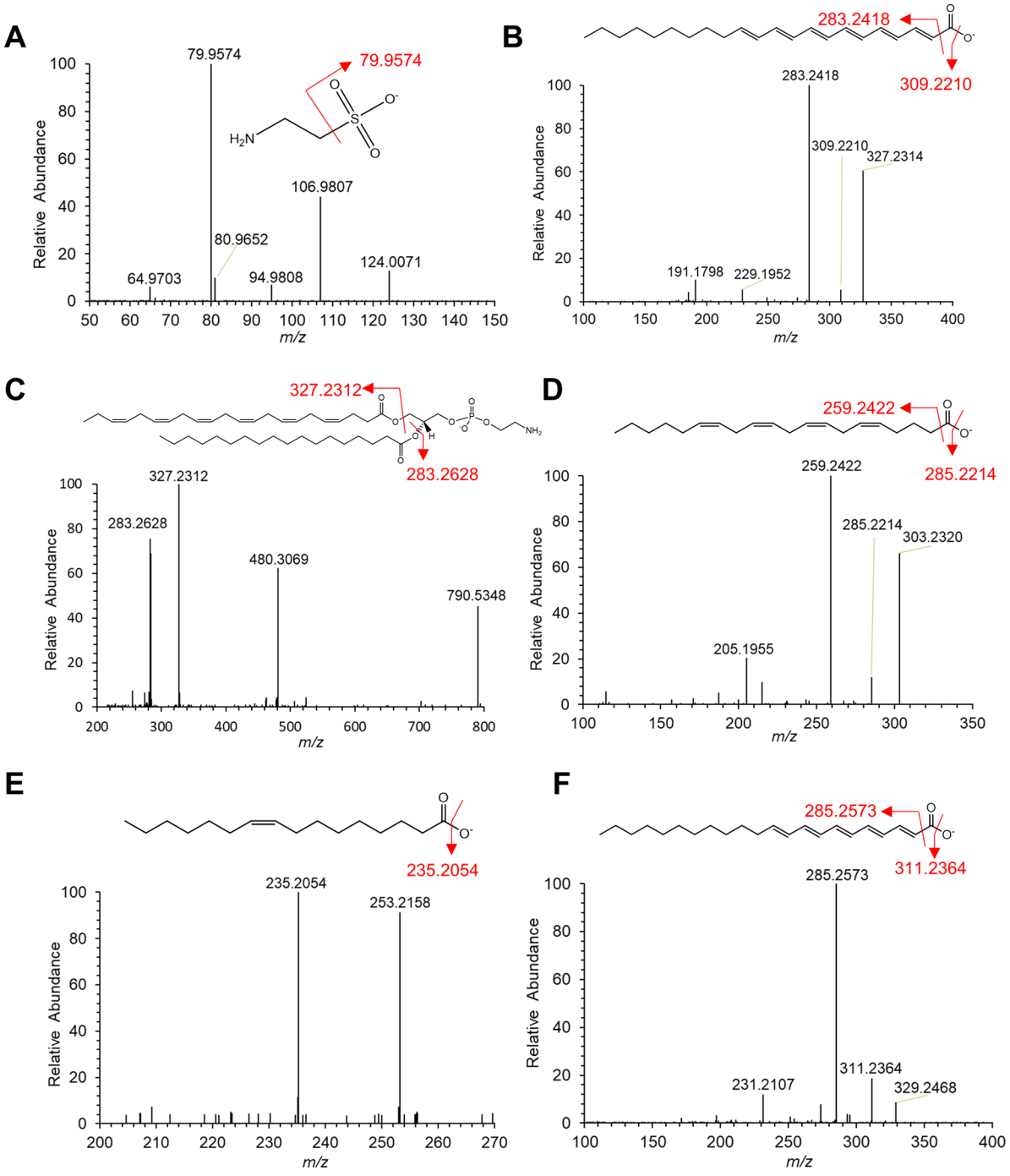


Figure 4. CID identification of selected molecular ions identified by GBDT as significant for cardiac pathology identification: (A) taurine; (B) docosahexaenoic acid FA(22:6), (C) phosphatidylethanolamine PE(22:6/18:0), (D) arachidonic acid FA(20:4), (E) palmitoleic acid FA(16:1), and (F) docosapentaenoic acid FA(22:5).

study, we identified taurine as one of the 62 molecular ion markers, whose changes in the collective abundance pattern may predict MI with very high accuracy, thus potentially improving predictive potential of taurine alone.

Another molecular species prominently depleted in infarcted areas are the polyunsaturated long-chain fatty acids, such as docosahexaenoic acid FA(22:6), docosapentaenoic acid FA(22:5), and docosatetraenoic acid FA(22:4). Recent evidence shows that infusion of docosahexaenoic acid (DHA) before or after the onset of cardiac ischemia in rats protects the heart from damage.[55] Moreover, adequate dietary intake of DHA is associated with reduced cardiovascular mortality in humans.[56] High blood levels of docosapentaenoic acid (DPA) were also demonstrated to prevent episodes of MI.[57] On the other hand, monounsaturated and saturated fatty acids have increased abundancy in infarct. Prominent rise in plasma levels of free fatty acids during MI as well as FA deposition in ischemic areas are well-documented and can be partially caused by the release of norepinephrine during MI episodes.[58,59] Combined with other molecular ions identified by DESI-MSI/GBDT, the aforementioned species generate a molecular signature characteristic of MI and may promote

DOI: 10.1021/acs.analchem.8b03410
Anal. Chem. 2018, 90, 12198−12206

diagnostic capabilities in silent MI and cardiac ischemia. Whereas MS imaging process directly on a patient heart muscle is extremely challenging, the signature we uncovered may reflect some changes in the blood-born metabolites during cardiac ischemia and improve diagnostic capabilities based on molecules detected in other tissues, e.g., pericardial fat tissue, which are readily accessible through routine cardiac procedures such as valve replacement, angioplasty, and artery bypass surgeries. The combination of DESI-MSI/GBDT offers a high prediction accuracy of MI based on the cardiac tissue-specific markers, but future studies are required to assess the applicability of this signature to early detection of infarct through analysis of blood, plasma, or pericardial fat in humans.

EXPERIMENTAL SECTION

MI Induction and Specimen Collection. Controlled apical MIs were induced in 11 female mice (8−10 weeks) by a permanent ligation of the LAD. Mice were sacrificed 5 ± 2 days postligation; the infarct locations were macroscopically confirmed, and harvested hearts were snap-frozen and later cryosectioned to 12 μm-thick sections. When not analyzed immediately, these fresh specimens were stored at −80 °C and briefly dried in vacuum desiccator before DESI-MS analysis.

DESI-MS Analysis. A custom build DESI-MS imaging stage coupled to an LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific) was used for DESI-MSI. The mass spectra were acquired in the negative ion mode using the Orbitrap as the mass analyzer at 60 000 resolving power, with the spray voltage set to −5 kV, the capillary voltage set to −65 V, and the tube lens voltage set to −120 V. Ion injection time was 100 ms, and one microscan was performed. At least three cardiac muscle sections for each sample were imaged. We used tissue nondestructive solvent system dimethylformamide:acetonitrile [1:1 (vol:vol)] at a flow rate of 0.8 μL/min assisted by a nebulizing gas (N2) at a pressure of 175 psi to desorb and ionize molecules of interest from the tissue. This allowed a subsequent histological evaluation of the same specimen to delineate infarcted regions. DESI-MS spray tip-to-surface distance was 2 mm; spray incident angle of 56°, and spray-to-inlet distance was 6.5 mm. Step size in the moving stage was set to 0.2 mm, and automatic gain control of mass spectrometer was switched off. These parameters were empirically found to yield the optimal MS signal from cardiac tissues. All experiments were carried out under identical experimental conditions to allow comparison between the measurements.

The software ImageCreator (ver. 3.0) was used to convert the Xcalibur 2.2 mass spectra files (Thermo Fisher Scientific) into a format compatible with BioMap (freeware, ver. 3.8.0.4, http://www.ms-imaging.org) to construct spatially accurate 2D ion images. Rainbow color palette was used in the BioMap for signal intensity visualization.

After DESI-MS imaging, the same cardiac section was subjected to H&E staining for unfixed tissue[32] and histopathologic evaluation. The location of delineated infarcted regions was compared with DESI-MS images. Another software, MSIReader (ver. 0.09),[60] was employed to extract regions of interest for GBDT analysis. To use the MSIReader, an additional freeware tool, MSConvert (tool of ProteoWizard software, ver. 2.1x.),[61] was utilized to convert the Xcalibur 2.2 mass spectra files (.raw files) into .mzML format files, and then imzMLConverter (ver. 1.3.0)[62] was used to combine .mzML files into .imzML format file, readable by the MSIReader.

For further structural identification of lipids and metabolites, tandem MS analyses were performed. Heart sections obtained from both ischemic and normally perfused regions were used; they were carefully removed from the glass slides, and the analytes were extracted in 500 μL of methanol:water 70:30 solution. The undissolved tissue was separated from the extract by centrifugation at 4000 g for 5 min. Supernatant was collected and introduced into the mass spectrometer for tandem MS analysis via electrospray at the solvent flow rate of 5 μL/min. Nebulizing gas pressure was set to 120 psi, and other parameters remained the same. A normalized collision energy of 20−40% was applied; the isolation window width was set to 1.0 m/z, and an activation Q value was 0.2−0.25. Ion injection time was 1000 ms, and one microscan was performed. Lipid and metabolite compositions were assigned based on high mass resolution analysis, isotope distribution, and the most prevalent tandem MS fragmentation pattern. The LipidMaps (http://www.lipidmaps.org/), MassBank (http://www.massbank.jp), and Metlin (https://metlin.scripps.edu/) databases were employed to assist with species identification. It should be noted that, while tandem MS experiments elucidated the exact length of acyl groups in glycerophospholipids, they could not accurately assign the position and stereochemistry of the double bonds in FAs and glycerophospholipids.

Machine Learning Algorithm. The 2D mass spectra files obtained by DESI-MS imaging were converted to Microsoft excel files and exported for GBDT analysis according to regions of interest of varying pathology using the MSIReader (see above).

Regions of interest were delineated within the specimens to categorize each 200-μm-diameter pixel of investigated heart muscle map as infarct or normally perfused myocardium. For the purpose of GBDT analysis, we defined each individual pixel as a "sample". The number of normal samples was 2427, while the number of infarct samples was 1075. Each sample was vectorized by a hand-written peak finding algorithm with a resolution of 0.05 m/z. A total of 938 peaks were found in all samples. Notably, the peak vectors around a pixel are concatenated to the feature vector of that pixel as well. Mathematically, let $P \in \mathbb{R}^{n\times m\times d}$ be a mass spectrometry image, where n is the width of the image, m is the length of the image, and $d = 938$ is the number of peaks included in each pixel P_{ij}. Then, the feature vector of P_{ij} is constructed as $[P_{ij}, P_{(i+1)j}, P_{(i-1)j}, P_{i(j+1)}, P_{i(j-1)}]$, where the brackets denote vector concatenation. Here, concatenation stands for the operation of joining vectors end-to-end. For example, concatenating the vectors (0, 1) and (2, 3) will yield the vector (0, 1, 2, 3).

The samples were then shuffled to compose the data set, and the whole data set was randomly separated into a training set, a cross-validation set, and a test set with ratio of 7:1:2. Samples were normalized by l_∞ norms of each samples vector, where l_∞ norm of a vector is the maximum absolute value of elements in the vector. Classification algorithms of logistic regression, support vector machines, random forests, gradient tree boosting, nearest neighbors, and Bayesian regression were tested. Model selection was based on the performance of the cross-validation set. The algorithm adapted from XGboost[63] yielded the best result and was chosen as the final classifier.

DOI: 10.1021/acs.analchem.8b03410
Anal. Chem. 2018, 90, 12198−12206

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.8b03410.

Mass spectral peaks and performance of classification algorithms (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: zare@stanford.edu.

ORCID

Zhenpeng Zhou: 0000-0002-3282-9468

Richard N. Zare: 0000-0001-5266-4253

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors are thankful to Ramin E. Beygui for many helpful disscussions. This research reported here was supported by the American Heart Association under award 16IRG27330012, by the Air Force Office of Scientific Research under AFOSR grant FA9550-16-1-0113, and by the National Science Foundation under the Data-Driven Discovery Science in Chemistry (D3SC) for Early Concept Grants for Exploratory Research grant CHE-1734082.

■ REFERENCES

(1) Benjamin, E. J.; Blaha, M. J.; Chiuve, S. E.; Cushman, M.; Das, S. R.; Deo, R.; de Ferranti, S. D.; Floyd, J.; Fornage, M.; Gillespie, C.; Isasi, C. R.; Jiménez, M. C.; Jordan, L. C.; Judd, S. E.; Lackland, D.; Lichtman, J. H.; Lisabeth, L.; Liu, S.; Longenecker, C. T.; Mackey, R. H.; Matsushita, K.; Mozaffarian, D.; Mussolino, M. E.; Nasir, K.; Neumar, R. W.; Palaniappan, L.; Pandey, D. K.; Thiagarajan, R. R.; Reeves, M. J.; Ritchey, M.; Rodriguez, C. J.; Roth, G. A.; Rosamond, W. D.; Sasson, C.; Towfighi, A.; Tsao, C. W.; Turner, M. B.; Virani, S. S.; Voeks, J. H.; Willey, J. Z.; Wilkins, J. T.; Wu, J. H.; Alger, H. M.; Wong, S. S.; Muntner, P. Heart Disease and Stroke Statistics—2017 Update: A Report From the American Heart Association. *Circulation*; American Heart Association, 2017.

(2) Manfroi, W. C.; Peukert, C.; Berti, C. B.; Noer, C.; Gutierres, D. d. A.; Silva, F. T. B. G. C. d. *Arquivos Brasileiros de Cardiologia* **2002**, *78*, 392−395.

(3) Dörr, M. *Heart* **2010**, *96* (18), 1434−1435.

(4) Caceres, L.; Cooke, D.; Zalenski, R.; Rydman, R.; Lakier, J. B. *Clin. Cardiol.* **1995**, *18* (10), 563−568.

(5) Morrow, D. A.; Cannon, C. P.; Jesse, R. L.; Newby, L. K.; Ravkilde, J.; Storrow, A. B.; Wu, A. H. B.; Christenson, R. H. *Circulation* **2007**, *115* (13), e356−e375.

(6) van Rooij, E. *N. Engl. J. Med.* **2016**, *374* (1), 85−87.

(7) Menger, R. F.; Stutts, W. L.; Anbukumar, D. S.; Bowden, J. A.; Ford, D. A.; Yost, R. A. *Anal. Chem.* **2012**, *84* (2), 1117−25.

(8) FREDRICKSON, D. S. *Circulation* **1969**, *40* (5S4), IV-99−IV-110.

(9) Evanochko, W. T.; Pohost, G. M. *NMR Biomed.* **1994**, *7* (6), 269−77.

(10) Shug, A. L.; Paulson, D. J. Fatty Acid and Carnitine-Linked Abnormalities During Ischemia and Cardiomyopathy. In *Myocardial Ischemia and Lipid Metabolism*; Ferrari, R., Katz, A. M., Shug, A., Visioli, O., Eds.; Springer US: Boston, MA, 1984; pp 203−224.

(11) Ferrari, R.; Di Lisa, F.; Raddino, R.; Bigoli, C.; Curello, S.; Ceconi, C.; Albertini, A.; Visioli, O. Factors Influencing the Metabolic and Functional Alterations Induced by Ischemia and Reperfusion. In *Myocardial Ischemia and Lipid Metabolism*; Ferrari, R., Katz, A. M., Shug, A.; Visioli, O., Eds.; Springer US: Boston, MA, 1984; pp 135−157.

(12) Perman, J.; Borén, J. *Atheroscler. Suppl.* **2011**, *12* (1), 4.

(13) Pena, Y.; Fernandez-Britto, J. E.; Bacallao, J.; Batista, J. F.; de Leon, M. L. *MEDICC review* **2012**, *14* (1), 18−24.

(14) Batchu, S. N.; Chaudhary, K.; Zlobine, I.; Pawa, J.; Seubert, J. M. Chapter 3: Fatty Acids and Cardiac Ischemia Reperfusion Injury A2. In *Handbook of Lipids in Human Function*; Meester, F. D., Ed.; AOCS Press: 2016; pp 39−83.

(15) Goldberg, I. J.; Trent, C. M.; Schulze, P. C. *Cell Metab.* **2012**, *15* (6), 805−812.

(16) Li, S.; Guo, Y.-L.; Zhao, X.; Zhang, Y.; Zhu, C.-G.; Wu, N.-Q.; Xu, R.-X.; Qing, P.; Gao, Y.; Li, X.-L.; Sun, J.; Liu, G.; Dong, Q.; Li, J.-J. *Sci. Rep.* **2017**, *7* (1), 360.

(17) Ali, S. E.; Farag, M. A.; Holvoet, P.; Hanafi, R. S.; Gad, M. Z. *Sci. Rep.* **2016**, *6*, 36359.

(18) Sun, M.; Sun, L.; Miao, L.; Lin, L.; Huang, S.; Yang, B.; Fu, J.; Ge, Z.; Jin, L.; Liu, J. *Chromatographia* **2016**, *79* (19), 1309−1316.

(19) Wang, X.; Wang, D.; Wu, J.; Yu, X.; Lv, J.; Kong, J.; Zhu, G.; Su, R. *Int. Heart J.* **2017**, *58* (3), 441−446.

(20) Crutchfield, C. A.; Thomas, S. N.; Sokoll, L. J.; Chan, D. W. *Clin. Proteomics* **2016**, *13*, 1.

(21) Marshall, J.; Kupchak, P.; Zhu, W.; Yantha, J.; Vrees, T.; Furesz, S.; Jacks, K.; Smith, C.; Kireeva, I.; Zhang, R.; Takahashi, M.; Stanton, E.; Jackowski, G. *J. Proteome Res.* **2003**, *2* (4), 361−72.

(22) Takats, Z.; Wiseman, J. M.; Gologan, B.; Cooks, R. G. *Science* **2004**, *306* (5695), 471−3.

(23) Wiseman, J. M.; Ifa, D. R.; Song, Q.; Cooks, R. G. *Angew. Chem., Int. Ed.* **2006**, *45* (43), 7188−7192.

(24) Bilkey, J.; Tata, A.; McKee, T. D.; Porcari, A. M.; Bluemke, E.; Woolman, M.; Ventura, M.; Eberlin, M. N.; Zarrine-Afsar, A. *Anal. Chem.* **2016**, *88* (24), 12099−12107.

(25) Tata, A.; Woolman, M.; Ventura, M.; Bernards, N.; Ganguly, M.; Gribble, A.; Shrestha, B.; Bluemke, E.; Ginsberg, H. J.; Vitkin, A.; Zheng, J.; Zarrine-Afsar, A. *Sci. Rep.* **2016**, *6*, 35374.

(26) Eberlin, L. S.; Tibshirani, R. J.; Zhang, J.; Longacre, T. A.; Berry, G. J.; Bingham, D. B.; Norton, J. A.; Zare, R. N.; Poultsides, G. A. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (7), 2436−2441.

(27) Zhang, J.; Feider, C. L.; Nagi, C.; Yu, W.; Carter, S. A.; Suliburk, J.; Cao, H. S. T.; Eberlin, L. S. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (6), 1166−1174.

(28) Banerjee, S.; Zare, R. N.; Tibshirani, R. J.; Kunder, C. A.; Nolley, R.; Fan, R.; Brooks, J. D.; Sonn, G. A. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (13), 3334−3339.

(29) Eberlin, L. S.; Margulis, K.; Planell-Mendez, I.; Zare, R. N.; Tibshirani, R.; Longacre, T. A.; Jalali, M.; Norton, J. A.; Poultsides, G. A. *PLoS Med.* **2016**, *13* (8), e1002108.

(30) Zhou, Z.; Zare, R. N. *Anal. Chem.* **2017**, *89* (2), 1369−1372.

(31) Kolk, M. V.; Meyberg, D.; Deuse, T.; Tang-Quan, K. R.; Robbins, R. C.; Reichenspurner, H.; Schrepfer, S. *J. Visualized Exp.* **2009**, No. 32, 1.

(32) Calligaris, D.; Norton, I.; Feldman, D. R.; Ide, J. L.; Dunn, I. F.; Eberlin, L. S.; Cooks, R. G.; Jolesz, F. A.; Golby, A. J.; Santagata, S.; Agar, N. Y. *J. Mass Spectrom.* **2013**, *48* (11), 1178−1187.

(33) van der Maaten, L.; Hinton, G. *J. Mach Learn Res.* **2008**, *9*, 2579−2605.

(34) Plumb, R. S.; Stumpf, C. L.; Gorenstein, M. V.; Castro-Perez, J. M.; Dear, G. J.; Anthony, M.; Sweatman, B. C.; Connor, S. C.; Haselden, J. N. *Rapid Commun. Mass Spectrom.* **2002**, *16* (20), 1991−6.

(35) Huanwen, C.; Arno, W.; Renato, Z. *J. Mass Spectrom.* **2007**, *42* (9), 1123−1135.

(36) Skrobot, V. L.; Castro, E. V. R.; Pereira, R. C. C.; Pasa, V. M. D.; Fortes, I. C. P. *Energy Fuels* **2007**, *21* (6), 3394−3400.

(37) Zhang, X.; Lu, X.; Shi, Q.; Xu, X.-q.; Leung, H.-c. E.; Harris, L. N.; Iglehart, J. D.; Miron, A.; Liu, J. S.; Wong, W. H. *BMC Bioinf.* **2006**, *7* (1), 197.

(38) Groseclose, M. R.; Massion, P. P.; Chaurand, P.; Caprioli, R. M. *Proteomics* **2008**, *8* (18), 3715−24.

DOI: 10.1021/acs.analchem.8b03410
Anal. Chem. 2018, 90, 12198−12206

(39) Hendriks, M. M.; Smit, S.; Akkermans, W. L.; Reijmers, T. H.; Eilers, P. H.; Hoefsloot, H. C.; Rubingh, C. M.; de Koster, C. G.; Aerts, J. M.; Smilde, A. K. *Proteomics* **2007**, *7* (20), 3672−80.
(40) Veselkov, K. A.; Mirnezami, R.; Strittmatter, N.; Goldin, R. D.; Kinross, J.; Speller, A. V. M.; Abramov, T.; Jones, E. A.; Darzi, A.; Holmes, E.; Nicholson, J. K.; Takats, Z. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (3), 1216−1221.
(41) Guenther, S.; Muirhead, L. J.; Speller, A. V. M.; Golf, O.; Strittmatter, N.; Ramakrishnan, R.; Goldin, R. D.; Jones, E.; Veselkov, K.; Nicholson, J.; Darzi, A.; Takats, Z. *Cancer Res.* **2015**, *75* (9), 1828−1837.
(42) Dória, M. L.; McKenzie, J. S.; Mroz, A.; Phelps, D. L.; Speller, A.; Rosini, F.; Strittmatter, N.; Golf, O.; Veselkov, K.; Brown, R.; Ghaem-Maghami, S.; Takats, Z. *Sci. Rep.* **2016**, *6*, 39219.
(43) Mirnezami, R.; Veselkov, K.; Strittmatter, N.; Kinross, J. M.; Goldin, R. D.; Speller, A.; Jones, E. A.; Holmes, E.; Abramov, T.; Nicholson, J. K.; Darzi, A. W.; Takats, Z. *Journal of Clinical Oncology* **2013**, *31* (15), 14620.
(44) Mirnezami, R.; Veselkov, K.; Strittmatter, N.; Goldin, R. D.; Kinross, J. M.; Stebbing, J.; Holmes, E.; Darzi, A. W.; Nicholson, J. K.; Takats, Z. *J. Clin. Oncol.* **2016**, *34* (15), 15104.
(45) Fusaro, V. A.; Mani, D. R.; Mesirov, J. P.; Carr, S. A. *Nat. Biotechnol.* **2009**, *27* (2), 190−8.
(46) Datta, S. *Stat. Appl. Genet. Mol. Biol.* **2008**, *7* (2), 1.
(47) Schaffer, S. W.; Ju Jong, C.; Kc, R.; Azuma, J. *J. Biomed. Sci.* **2010**, *17* (1), S2.
(48) Jacobsen, J. G.; Smith, L. H. *Physiol. Rev.* **1968**, *48* (2), 424−511.
(49) Bhatnagar, S. K.; Welty, J. D.; Yusuf, A. R. A. *Int. J. Cardiol.* **1990**, *27* (3), 361−366.
(50) Hansen, S. H.; Andersen, M. L.; Birkedal, H.; Cornett, C.; Wibrand, F. *Advances in experimental medicine and biology* **2006**, *583*, 129−35.
(51) Ardisson, L. P.; Rafacho, B. P. M.; Santos, P. P.; Assalin, H.; Gonçalves, A. F.; Azevedo, P. S.; Minicucci, M. F.; Polegato, B. F.; Okoshi, K.; Marchini, J. S.; Barbisan, L. F.; Fernandes, A. A. H.; Seiva, F. R. F.; Paiva, S. A. R.; Zornoff, L. A. M. *Int. J. Cardiol.* **2013**, *168* (5), 4925−4926.
(52) Schaffer, S. W.; Jong, C. J.; Ito, T.; Azuma, J. *Amino Acids* **2014**, *46* (1), 21−30.
(53) Cooper, M. W.; Lombardini, J. B. *Advances in experimental medicine and biology* **1982**, *139*, 191−205.
(54) Lombardini, J. B.; Cooper, M. W. *Journal of laboratory and clinical medicine* **1981**, *98* (6), 849−859.
(55) Richard, D.; Oszust, F.; Guillaume, C.; Millart, H.; Laurent-Maquin, D.; Brou, C.; Bausero, P.; Visioli, F. *Prostaglandins, Leukotrienes Essent. Fatty Acids* **2014**, *90* (4), 139−143.
(56) Holub, B. J. *Prostaglandins, Leukotrienes Essent. Fatty Acids* **2009**, *81* (2−3), 199−204.
(57) Byelashov, O. A.; Sinclair, A. J.; Kaur, G. *Lipid Technol.* **2015**, *27* (4), 79−82.
(58) Jurand, J.; Oliver, M. F. *Atherosclerosis* **1970**, *11* (1), 157−170.
(59) Kurien, V. A.; Oliver, M. F. *Prog. Cardiovasc. Dis.* **1971**, *13* (4), 361−373.
(60) Robichaud, G.; Garrard, K. P.; Barry, J. A.; Muddiman, D. C. *J. Am. Soc. Mass Spectrom.* **2013**, *24* (5), 718−21.
(61) Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P. *Bioinformatics* **2008**, *24* (21), 2534−6.
(62) Race, A. M.; Styles, I. B.; Bunch, J. *J. Proteomics* **2012**, *75* (16), 5111−2.
(63) Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; ACM: San Francisco, California, 2016; pp 785−794.

DOI: 10.1021/acs.analchem.8b03410
Anal. Chem. 2018, 90, 12198−12206