analytical chemistry

Article

Cite This: Anal. Chem. 2018, 90, 3802−3810

pubs.acs.org/ac

Top-Down Proteomics Enables Comparative Analysis of Brain Proteoforms Between Mouse Strains

Roderick G. Davis,[†,‡] Hae-Min Park,[†,‡] Kyunggon Kim,[‡,§] Joseph B. Greer,[‡] Ryan T. Fellers,[‡] Richard D. LeDuc,[‡] Elena V. Romanova,[∥] Stanislav S. Rubakhin,[∥] Jonathan A. Zombeck,[§,⊥] Cong Wu,[§,∥] Peter M. Yau,[#] Peng Gao,[‡] Alexandra J. van Nispen,[‡] Steven M. Patrie,[‡] Paul M. Thomas,[‡] Jonathan V. Sweedler,[∥] Justin S. Rhodes,[⊥] and Neil L. Kelleher*[,‡]

[‡]Departments of Chemistry, Molecular Biosciences and the Proteomics Center of Excellence, Northwestern University, 2145 North Sheridan Road, Evanston, Illinois 60208, United States

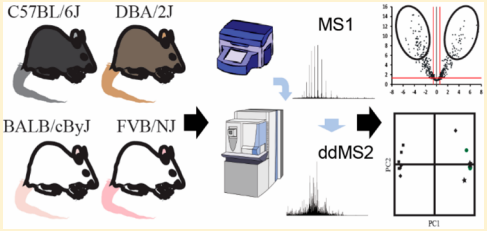
[∥]Department of Chemistry, University of Illinois, Urbana−Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801, United States

[⊥]Department of Psychology, University of Illinois, Urbana−Champaign, 405 North Mathews Avenue, Urbana, Illinois 61801, United States

[#]Roy J. Carver Biotechnology Center, Protein Sciences Facility, University of Illinois, Urbana−Champaign, 505 South Mathews Avenue, Urbana, Illinois 61801, United States

Supporting Information

ABSTRACT: Over the past decade, advances in mass spectrometry-based proteomics have accelerated brain proteome research aimed at studying the expression, dynamic modification, interaction and function of proteins in the nervous system that are associated with physiological and behavioral processes. With the latest hardware and software improvements in top-down mass spectrometry, the technology has expanded from mere protein profiling to high-throughput identification and quantification of intact proteoforms. Murine systems are broadly used as models to study human diseases. Neuroscientists specifically study the mouse brain from inbred strains to help understand how strain-specific genotype and phenotype affect development, functioning, and disease progression. This work describes the first application of label-free quantitative top-down proteomics to the analysis of the mouse brain proteome. Operating in discovery mode, we determined physiochemical differences in brain tissue from four healthy inbred strains, C57BL/6J, DBA/2J, FVB/NJ, and BALB/cByJ, after probing their intact proteome in the 3.5−30 kDa mass range. We also disseminate these findings using a new tool for top-down proteomics, TDViewer and cataloged them in a newly established Mouse Brain Proteoform Atlas. The analysis of brain tissues from the four strains identified 131 gene products leading to the full characterization of 343 of the 593 proteoforms identified. Within the results, singly and doubly phosphorylated ARPP-21 proteoforms, known to inhibit calmodulin, were differentially expressed across the four strains. Gene ontology (GO) analysis for detected differentially expressed proteoforms also helps to illuminate the similarities and dissimilarities in phenotypes among these inbred strains.



The laboratory mouse is the primary model organism for the study of mammalian biology and disease because of their genetic suitability and accessibility.[1] Inbred strains of mice represent unique fixed genotypes with predictable phenotypic traits defined by fixed allelic composition.[2] Genome-wide mapping of inbred mouse strain gene expression can resolve genetic variations between the mouse strains using various genome analysis technologies such as next generation sequencing.[3−6] A wide variety of behavioral phenotypes of the inbred strains have been reported for learning and memory tasks as well as behavior responses to drugs such as nicotine and cocaine, among others.[7] Many investigators in neurobiology have focused on gene expression to characterize and compare different phenotypes of these strains.[8,9] However, given that proteins are one of the most important classes of molecules within a cell with important structural features, localizations and turnover rates which cannot be fully predicted by genomic or transcriptomic evaluations, assessing how genetic differences between strains affect the brain proteome could provide an important missing link between genetics and complex phenotypes.

Traditionally, bottom-up proteomics (BUP) approaches have been used to identify and to quantify the mouse brain proteome with protein separation techniques such as large-gel two-dimensional electrophoresis, providing new insights into molecular mechanisms of brain function and brain-associated

Received: October 6, 2017
Accepted: February 14, 2018
Published: February 26, 2018

DOI: 10.1021/acs.analchem.7b04108
Anal. Chem. 2018, 90, 3802−3810

diseases.[10−13] Recently, Sharma et al. used BUP together with existing transcriptome databases for an in-depth investigation of the mouse brain and its major subregions and cell types.[14] A mouse brain atlas that covers 17 anatomically distinct regions was reported by Jung et al.[15] Their bottom-up data revealed comprehensive and regiospecific expression of over 12000 gene products across the entire brain. However, because BUP is rarely able to fully characterize full-length endogenous proteoforms,[16] there are limitations in the amount of qualitative and quantitative information that can be provided.

Compared to BUP, an advantage of the top-down proteomics (TDP) is its ability to deduce proteoform-resolved information on combinatorial PTMs, coding polymorphisms and alternative splicing events on the same molecule. Additionally confidence in proteoform identification and complete characterization is often greater with TDP. Our group has demonstrated that quantitative top-down proteomics is possible in a multitarget discovery space using a model yeast system.[17] Recently, we used TDP to differentiate abundant proteoforms derived from patient-derived breast tumor xenografts[18] and peripheral blood mononuclear cells associated with kidney and liver transplantation.[19] In rodent brains, Li et al. used a detergent-free sample preparation protocol to identify 736 proteoforms representing 471 proteins in an extract from C57BL/6 mice using a TDP method.[20] In the same study 151 known or potential neuropeptides were identified.

Here we report the first application of a label-free TDP approach to characterize and to quantify proteoforms extracted from the brains of mice from four inbred mouse strains often used in drug and alcohol addiction research: C57BL/6J, DBA/2J, FVB/NJ, and BALB/cByJ. C57BL/6J, most widely used as a genetic background reference, readily ingest ethanol solutions and shows a behavior pattern indicative of the rewarding effects of cocaine, morphine, and other opiates.[21] DBA/2J mostly avoids drinking alcohol and shows differential sensitivity to rewarding effects of cocaine, and morphine relative to C57BL/6J.[21] BALB/cByJ does not appear to self-administer cocaine[22] but readily self-administers opioids, which produces profound behavioral effects in the strain.[23] FVB/NJ has been shown to have a greater locomotor stimulant response to ethanol relative to C57BL/6J.[24] We have previously shown that these strains differ in their behavior and neurochemical responses to drug rewards.[25,26]

Additionally, this work allows us to test and introduce a high performance computing environment for top-down protein database searching and reporting. These advances in informatics infrastructure enabled highly confident qualitative analysis of proteins and proteoforms using a global FDR calculations as well as integrated label-free quantitation, analyzed on a local institutional high-performance computing environment. Finally, we establish the Mouse Brain Proteoform Atlas (http://mousebrain.northwestern.edu) as a web-based repository for proteoforms characterized with high confidence available for use (or proteoform deposit) by other investigators in the murine top-down proteomics field.

MATERIALS AND METHODS

Materials and Reagents. Acetonitrile (Optima HPLC grade), water (Optima HPLC grade), acetone (Optima HPLC grade), methanol (Optima HPLC grade), and chloroform were from Fisher. Water for non-HPLC use in the experiments was purified by a Milli-Q system from Millipore Co. (Bedford, MA, U.S.A.). Formic acid (FA, 28905) and Halt Phosphatase and Protease Single-Use Inhibitor Cocktail (78443) were purchased from Thermo Scientific (Rockford, IL, U.S.A.). Tris base (T1503), hydrochloric acid (HCl, #320331), benzonase nuclease (E1014), and magnesium chloride ($MgCl_2$, M8266) were purchased from Sigma-Aldrich (St. Louis, MO, U.S.A.). Sodium dodecyl sulfate (SDS) was purchased from Bio-Rad (Hercules, CA, U.S.A.). Sodium butyrate (A11079) was purchased from Alfa Aesar (Tewksbury, MA, U.S.A.). 1,4-Dithiothreitol (DTT, CAS# 3483−12−3) was purchased from EMD Millipore (Darmstadt, Germany).

Mouse Brain Tissue Preparation and Homogenization. Five nine-week old (± 3 days) female mice from each of four strains, C57BL/6J, DBA/2J, FVB/NJ, and BALB/cByJ, were acquired from the Jackson Laboratory (Bar Harbor, ME). Animals were euthanized by asphyxia using CO_2 according to the approved UIUC IACUC animal use protocol in accordance with local and federal regulations. The circulatory system of euthanized animals was quickly flushed with 20 mL of ice cold Modified Gey's balanced salt solution (mGBSS) using transcardial perfusion. mGBSS consists of 1.5 mM $CaCl_2$, 4.9 mM KCl, 0.2 mM KH_2PO_4, 11 mM $MgCl_2$, 0.3 mM $MgSO_4$, 138 mM NaCl, 27.7 mM $NaHCO_3$, 0.8 mM Na_2HPO_4, and 25 mM HEPES dissolved in Milli-Q water with the pH adjusted to 7.2 using NaOH in Milli-Q water. Mouse brains were quickly, surgically removed and frozen in liquid nitrogen for storage. These brains were then pulverized to a fine powder with a tissue pulverizer (BioSpec, Bartlesville, OK) and cooled to below −78.5 °C using liquid nitrogen. Tissue powders placed in individual, marked, and chilled on dry ice tubes and stored at −78.5 °C until sample processing.

Mouse brain tissues are soft. Moreover, pulverization of frozen samples produces mostly fine powder. Therefore, additional homogenization with a Teflon homogenizer in lysis buffer was sufficient to effectively lyse cells present in this sample. Approximately one-third of the pulverized brain tissue volume was transferred to a Dounce homogenizer containing ice cold lysis buffer (4% sodium dodecyl sulfate (SDS), 15 mM Tris-HCl (pH 7.4), 10 mM sodium butyrate, 10 mM DTT, and 1× Halt protease and phosphatase inhibitor cocktail). A total of 100 strokes of a Teflon homogenizer was used to visually dematerialize the brain tissue. Next, benzonase nuclease (750 units) was added to degrade sample DNA and RNA after adding $MnCl_2$ to a final concentration of 1 mM. Samples were incubated at 37 °C for 30 min. After incubating at 95 °C for 5 min, the soluble fraction located in the supernatant was obtained after centrifugation (21000*g*) at 4 °C for 10 min.

Protein Separation Using GELFrEE and SDS Removal. Proteins in 300 μL of homogenized and lysed samples were precipitated by adding cold acetone at a ratio of 4:1 (acetone/sample) and the resulting protein pellet was resuspended in 1% SDS. The protein concentration was determined at this point using a BCA assay (Pierce, Rockford, IL). Next, samples containing three hundred micrograms of total protein were prepared for GELFrEE separation per manufacturer's instructions (GELFrEE 8100 Fractionation System, Expedeon, Cambridgeshire, UK). Samples were randomized for fractionation across three 8% GELFrEE cartridges (eight lanes per cartridge). Fraction 1 containing proteins of ∼3.5−30 kDa was collected for each sample and stored at −80 °C until analysis by LC-MS/MS. Prior to LC-MS/MS analysis, fractions were thawed and SDS was removed by methanol/chloroform/water extraction.[27] Proteins were resuspended in 50 μL Solvent A (95% H_2O, 5% ACN, 0.2% FA) by vigorous pipetting and

DOI: 10.1021/acs.analchem.7b04108
Anal. Chem. 2018, 90, 3802−3810

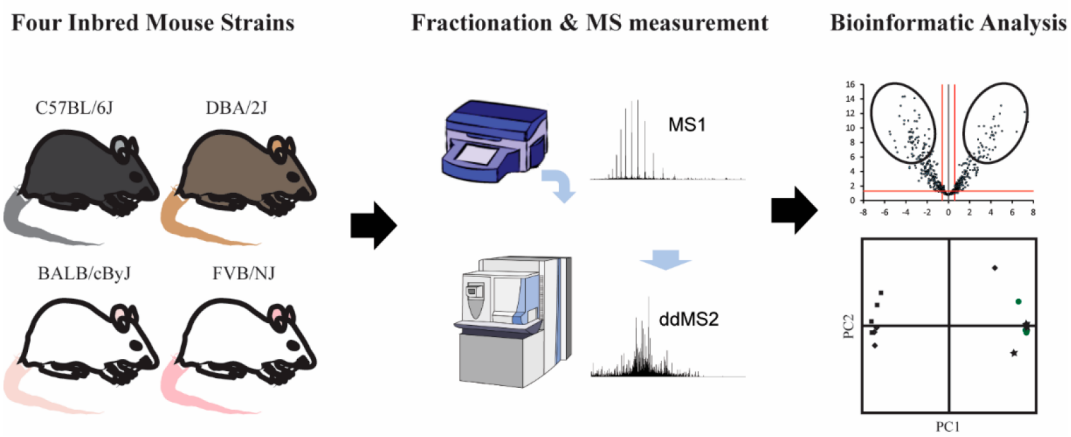


Figure 1. Overview of study for comparative label-free quantitation of mouse brain tissue proteoforms using top-down proteomics. The proteins were obtained from mouse brains after several steps including tissue homogenization and liquid extraction. Proteins below ~30 kDa were isolated using a GELFrEE fractionation system. High-resolution LC-MS/MS runs were carried out with a randomized order, and then data analysis was performed using TDPortal 1.3. For proteoform characterization, deconvoluted MS^1 and MS^2 data were processed using a three-tiered search tree and visualized with TDViewer. For proteoform quantitation, a hierarchical linear model was employed using normalized MS signal intensities. Strain differentiation based upon TDP data was gauged using principal component analysis and Pearson correlation analysis of TDP data.

sonication in an ice bath for 10 min. Samples were then centrifuged for 10 min at 21000*g* at 4 °C prior to transferring the supernatant to HPLC autosampler vials for top-down LC-MS/MS analysis.

LC-MS Methods. Resuspended protein fractions (6 μL) were injected onto a PepSwift RP-4H trap column (150 μm ID × 2 cm). After loading the sample at 10 uL/min, the trap was switched in-line with a Thermo ProSwift RP-4H column (100 μm ID × 50 cm) for protein separation. Mobile phases were delivered using a Dionex Ultimate 3000 RSLCnano system. Proteins were eluted from the ProSwift RP-4H column with a mobile phase flow rate of 1 μL/min. Solvent A was described above. Solvent B was 5% water, 95% acetonitrile, 0.2% formic acid. A linear gradient was used with slope change points: 5% B at 0 min.; 15% B at 5 min.; 55% B at 80 min.; 95% B hold from 83 to 102 min. Eluent was directed to a 15 μm nano-electrospray tip (New Objective, Waltham, MA) held at 1.9–2.1 kV. Mass spectrometry measurements were performed on an Orbitrap Elite (Thermo Scientific, Bremen, Germany) mass spectrometer operating in "protein mode" and fitted with a custom nanoelectrospray ionization source. A top-2 data-dependent acquisition strategy was employed as described previously using higher-energy collisional dissociation (HCD).[17] Biological and technical replicates were randomized across the study to minimize bias.

Data Analysis. Raw data files were processed in TDPortal 1.3, a Galaxy[28] platform containing a bioinformatics pipeline that detects features in top-down data, performs a qualitative database search, and quantifies intensity changes for any proteoform passing a FDR cutoff. The shotgun annotated mouse database was generated using biological variability reported in UniProt in the manner described by Pesavento et al.[29] (For further information, see Method S-1.) Search results were visualized using TDViewer software (available for download at http://topdownviewer.northwestern.edu). Venn diagrams were generated using Venny (http://bioinfogp.cnb.csic.es/tools/venny/index.html). Raw and processed data can be downloaded from ftp://massive.ucsd.edu/MSV000081435.

Statistical Analysis. Relative differential abundances in proteoform signal intensities (proteoform-level extracted ion chromatographic peak intensities) was determined using methods described in detail elsewhere.[17] In brief, a hierarchical linear models were used with biological replicates (random effect) nested within mouse strains (fixed effect), and mass spectrometric technical replicates (repeat injections, random effect) nested within the biological replicates. Signal intensity measures were log base 2 transformed and standardized prior to analysis. Each proteoform with measured intensity data from at least 50% of the LC runs for each strain was analyzed independently. For this study, each pairwise comparison of proteoform intensity between mouse strains was performed using a Student's *t* test on the least-squares means estimates from the linear models. This test used the least significant difference (LSD) for the finding of pairwise differences, which is appropriate given the randomized-complete-block design of the study.[30] Lastly, all p-scores were corrected for multiple testing at a false discovery rate of $\alpha = 0.05$.[31] Hereafter, FDR for quantitation is referred to as "quantitative FDR" to distinguish it from the FDR associated with protein identification, termed a "qualitative FDR". Statistical analyses were performed within SAS 9.4, (SAS Institute; Cary, NC).

Mouse Brain Proteoform Atlas. The proteoforms detected and fully characterized (C-Score > 40) in this project were assigned proteoform records (with PFR accession numbers) and deposited into the Mouse Brain Proteoform Atlas (http://mousebrain.northwestern.edu). The repository was established to aid investigators who study proteins present in murine brain tissues and can be useful for wider range of investigations of Metazoa. The proteoforms cataloged on the site have been segregated by strain and top-down mass spectrometric data set. Currently, 896 proteoforms from 327 gene products from four mouse strains have been cataloged on the site (including data from this paper and previously unpublished data sets). As regiospecific data become available, the atlas will be updated to allow filtering by region and treatment.

RESULTS AND DISCUSSION

Inbred mice are widely used to model various pathological conditions in humans. The murine brain has been widely studied to gain insights into the mammalian neuro-proteome

DOI: 10.1021/acs.analchem.7b04108
Anal. Chem. 2018, 90, 3802–3810

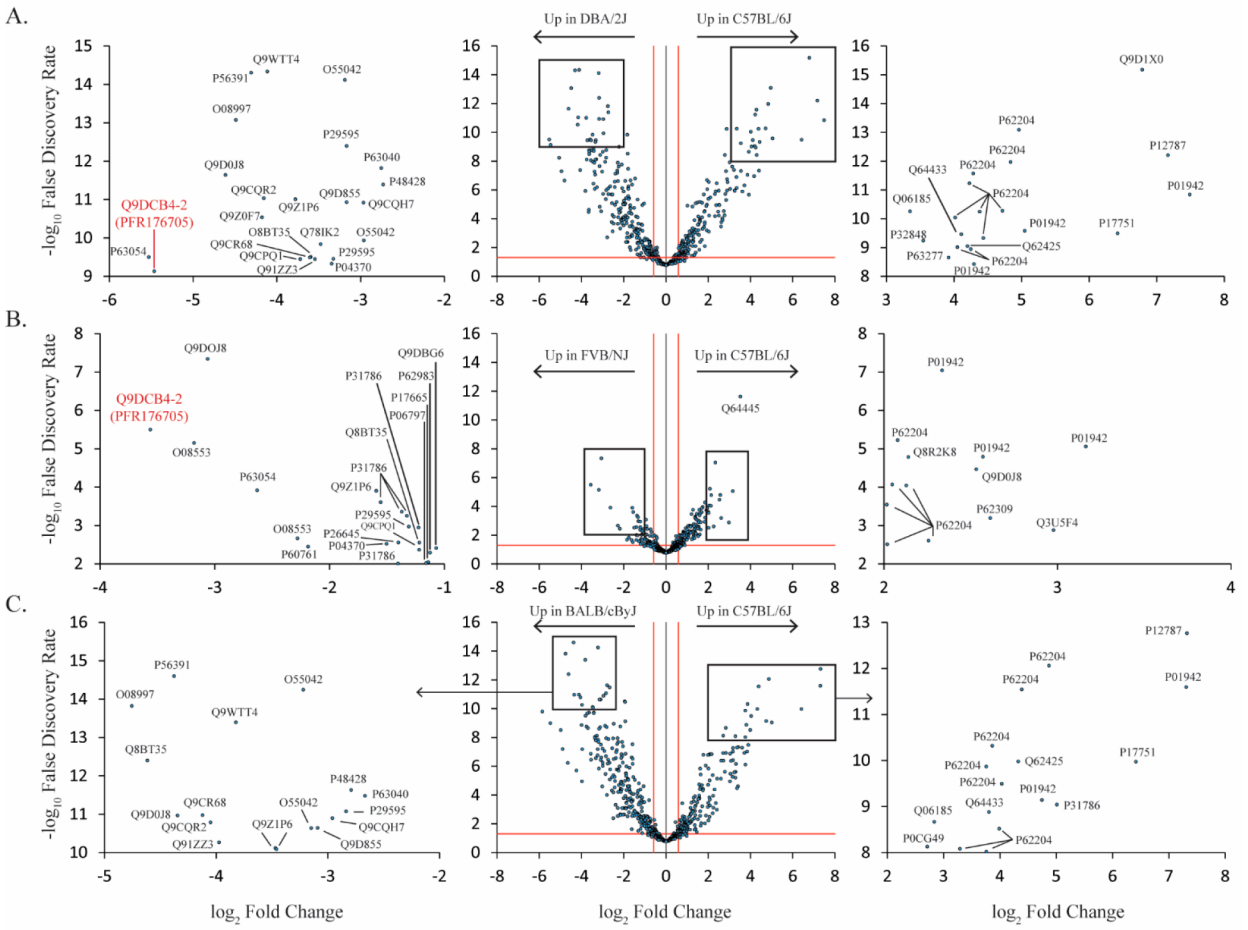


Figure 2. Quantitative comparison of 593 proteoforms from C57BL/6J, DBA/2J, FVB/NJ, and BALB/cByJ. The volcano plots (full versions in the center column with magnified areas in left and right columns) for the pairwise comparisons were generated, showing each proteoform (blue dots) as a function of relative signal abundance difference (x-axis, $\log_2$ fold-change) between the strains, and the statistical confidence (y-axis, quantitative FDR) that the difference is significant in normalized MS signal intensity. For C57BL/6J v/s DBA/2J (A), C57BL/6J v/s FVB/NJ (B), C57BL/6J v/s BALB/cByJ (C), 438, 211, and 422 proteoforms were identified, respectively, below a 5% FDR cutoff (horizontal red line) for quantitation and above a 1.5-fold change (vertical red lines) in proteoform signal abundance between the strains.

and to better understand the molecular mechanisms underlying brain functions. While several investigators have used bottom-up proteomics in these types of studies, this quantitative top-down study of the mouse brain proteome provides unique information on specific proteoforms and their variability in different strains, C57BL/6J, DBA/2J, FVB/NJ, and BALB/cByJ. The study was performed following the workflow shown in Figure 1.

The Venn diagrams in Figure S-1 show the overlap in proteins identified (1% qualitative FDR) and proteoforms characterized (1% qualitative FDR and C-Score >40) across each strain. The complete listing for these two groups is given in Supporting Information, Table S-1 (proteins) and Table S-2 (proteoforms). Across the experiment, a total of 131 proteins were identified and 343 proteoforms were fully characterized as per the C-score model.[31] On average, 102 proteins and 195 proteoforms were identified for each strain. Surprisingly, several proteins were repeatedly detected in only one strain, indicating relatively higher levels of such proteoforms among genetically similar animals and higher specific analyte recovery associated with different physicochemical properties of the examined tissues. Myristoylated alanine-rich C-kinase substrate (MARCKS, P26645) and secretogranin-2 (Q9CQV4) were detected only in C57BL/6J. MARCKS proteins are known binders of synapsins, a family of proteins long thought to regulate neurotransmitter release at synapses. Tubulin beta-5 chain (P99024) and Src substrate cortactin (Q60598) were detected only in DBA/2J. Synapsin-2 (Q64332), a neuronal phosphoprotein that coats synaptic vesicles, was found only in FVB/NJ. Protein S100-A5 (P63084), mitochondrial import receptor subunit TOM22 (Q9CPQ3), MICOS complex subunit Mic10 (Q7TNS2), and cholecystokinin (P09240) were found only in BALB/cByJ.

The study design presented here nests multiple levels of biological and experimental complexity, four mouse strains with four biological replicates per strain and with five technical injections per biological replicate. The number of conditions tested, biological replicates, and technical replicates require conducting reproducible fractionation across multiple GEL-FrEE cartridges and maintenance of LC-MS performance over a period of a week at a level of repeatability that would still allow a determination of the interstrain effect, in this case, the detection of differences in the brain proteome from the four mouse strains. Typical chromatographic repeatability is shown in Figure S-2. Applying a random effects model allowed attribution of the variance in proteoform signal intensity across

DOI: 10.1021/acs.analchem.7b04108
Anal. Chem. 2018, 90, 3802−3810

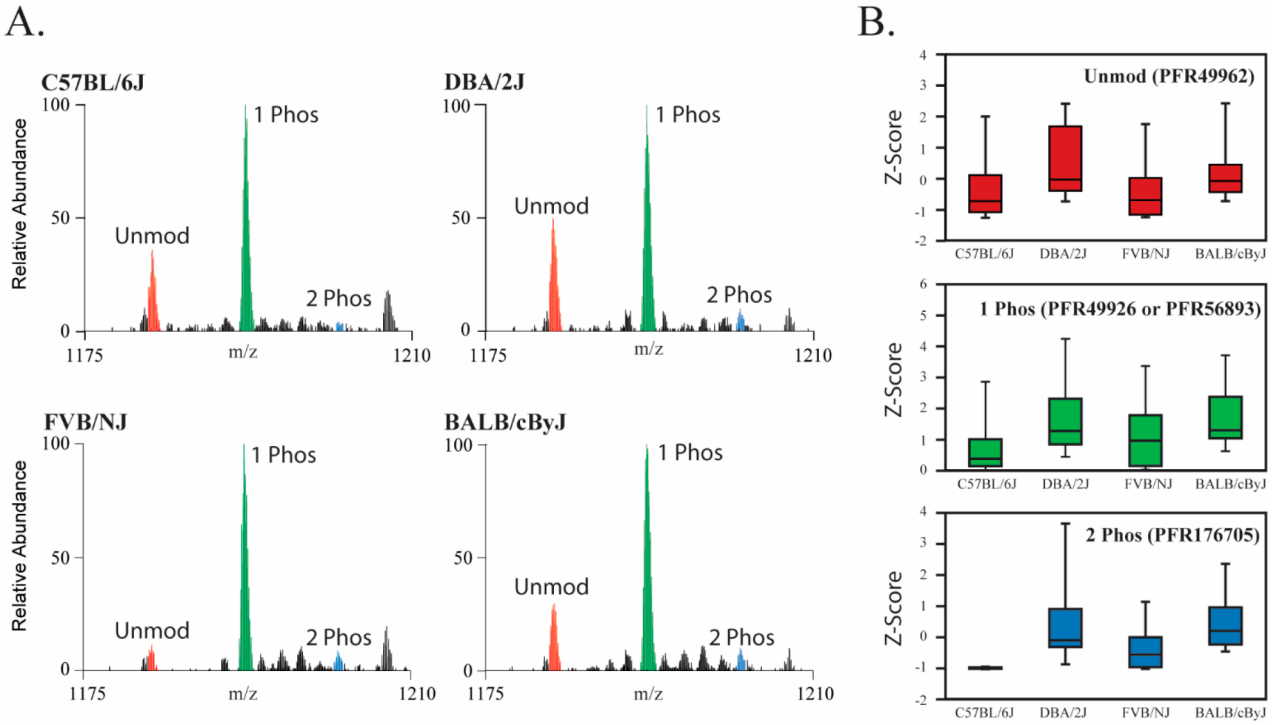


Figure 3. Differential analysis of ARPP-21 across mouse strains. (A) Representative spectra showing the relative signal abundance of the 8+ charge state for three phosphorylation states for ARPP-21 from C57BL/6J, DBA/2J, FVB/NJ, and BALB/cByJ mouse strains. (B) Box and whisker plots showing the Z-scores for the three ARPP-21 proteoforms from the four strains. For the three pairwise comparisons involving C57BL/6J and one of the other strains, only in the case of unmodified C57BL/6J against FVB/NJ does differential analysis not pass our threshold of significance (5% quantitative FDR and fold change >1.5) across all the proteoforms.

the levels of the experiment. Figure S-3 shows the relative contributions of the four major sources of variation, interstrain variation, biological variability between replicates, and mass spectrometric replicates, with remaining unassignable variation being "Residual". The variance components were calculated separately for each of the quantified proteoforms, and box plots were made showing the overall behavior across the experiment. The percent variation attributable to technical variation is indicative of consistent bench work. In general, most proteoforms show a strain effect that is larger than the biological replicate effect, suggesting that the difference between strains is larger than the difference within strains for the proteoforms examined. This suggests that label-free TDP is an appropriate method for detecting differentially expressed proteoforms in this system.

The candidate pool of proteoforms considered for quantitative analysis was expanded using a qualitative FDR cutoff of 10% and removing the C-score cutoff. In total, 593 proteoforms were found consistently across all four strains (hence, the difference from the fully characterized proteoform count reported in Figure S-1, where a 1% qualitative FDR threshold was used). For each pairwise comparison, we mined data from the SAS statistical report and generated a volcano plot. For each proteoform, the estimated effect size (in $\log_2$ fold-change) and the statistical confidence (quantitative false discovery rate) that there was a difference in normalized intensity are plotted for the two strains being compared. Figure 2 shows the three pairwise plots for C57BL/6J and DBA/2J, C57BL/6J and FVB/NJ, and C57BL/6J and BALB/cByJ. The plots for the other pairwise comparisons can be found in Figure S-3. The horizontal red line in each plot marks the 5% FDR statistical confidence level. The vertical red lines mark intensity differences that are 1.5-fold above and below a no-change value for convenience when viewing the plots.

Our quantitative analysis showed that of the 593 proteoforms, abundance changes for 211 proteoforms were found to be statistically significant between C57BL/6J and FVB/NJ at a 5% global quantitative FDR threshold while 438 and 422 proteoforms from this set were found to be statistically different between C57BL/6J and DBA/2J, and between C57BL/6J and BALB/cByJ, respectively (Figure 2). There are also 120 proteoforms with significant abundance differences between BALB/cByJ and DBA/2J at the same FDR cutoff whereas there are 401 and 419 proteoforms with those between DBA/2J and FVB/NJ, and BALB/cByJ and FVB/NJ, respectively (Figure S-4).

Among the results, we differentiated three phosphorylation states of isoform-2 of cAMP-regulated phosphoprotein 21 (UniProt Accession Q9DCB4−2). The three states were assigned the following PFR identifiers: unmodified (PFR49962), monophosphorylated (PFR49926 or PFR56893), and diphosphorylated (PFR176705; labeled in red text in Figure 2). Within the genome, full length cAMP-regulated phosphoprotein 21 (isoform 1) is an 88 kDa protein composed of 807 amino acids. Isoform 2 is a splice variant of cAMP-regulated phosphoprotein 21, known in the literature as ARPP-21, consists of the first 88 amino acids of the full length species with an ES-to-TL substitution at positions 87−88 due to an alternative splice between exons 6 and 7 that introduces a stop codon within exon 7 of isoform 2. Figure 3A shows mass spectra for the three ARPP-21 proteoforms across each strain. Figure 3B shows the differential expression levels given by Z-score of the proteoforms that was derived by our statistical platform. For each proteoform, all pairwise comparisons

DOI: 10.1021/acs.analchem.7b04108
Anal. Chem. 2018, 90, 3802−3810

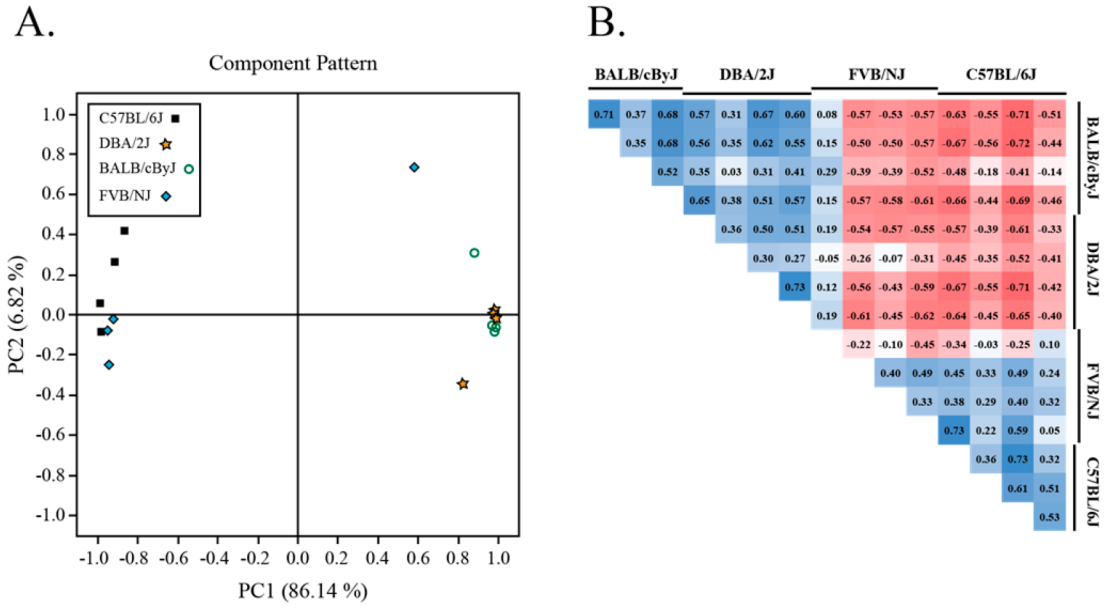


Figure 4. Discriminating Mouse Strains with Top-Down Proteomics. (A) PCA scores scatterplot of principal components 1 and 2. Component plot from the unsupervised principal component analysis showing the discrimination of data points for strains C57B/6J and FVB/NJ from data points for strains DBA/2J and BALB/cByJ by the first component. On the far right, three DBA/2J data points are clustered above three data points for BALB/cByJ. (B) Correlation matrix revealed positive correlations in label-free quantitation intensity between biological replicates of the same strains. The color code is based on their Pearson R correlation coefficient values from −1 (red) to 1 (blue).

involving C57BL/6J and one of the other strains pass our threshold of significance (5% quantitative FDR and fold change >1.5) except that of unmodified ARPP-21 in C57BL/6J versus FVB/NJ. Graphical fragment maps obtained for these three proteoforms can be found in Figure S-5. Interestingly, Caporaso et al., showed that drugs of abuse modulate phosphorylation of ARPP-21.[33] They used a phosphorylation state-specific antibody selective for the detection of ARPP-21 phosphorylated on Ser^{55} to measure increased phosphorylation at the site, in C57BL/6 mice treated with either methamphetamine or cocaine. Whereas in our work, two phosphorylation states of this protein were quantified in the same experiment. Given that multisite phosphorylation mediates many protein functions,[34] ARPP-21 phosphorylation on the both Ser^{32} and Ser^{55} may regulate some of the physiological effects of drugs of abuse in mouse brain. The confident identification, characterization, and quantification of these three proteoforms of isoform 2 could not have been achieved during a standard bottom-up proteomics experiment. This example shows how the top-down approach enables direct determination of the relative abundance of intact proteoforms and demonstrates a new screening modality for this important analyte.

Within the analysis, we also detected other known neuronal proteins including complexin-2, myelin basic protein, neurogranin, synapsin-2, and Src substrate cortactin. Few neuropeptides were observed and only 5% (30 of 593) of quantified proteoforms had masses less than 5 kDa, this is likely due to the 3 kDa molecular weight cutoff filter within the GELFrEE cartridges.

To determine whether quantitative TDP data could be used to discriminate between multiple mouse strains, we used both a multivariate principal component analysis (PCA) and a Pearson correlation analysis. In Figure 4A, we show a plot of PC1 against PC2 from the PCA. Eighty-six percent of the variance can be attributed to the first principal component, and nearly 93% of the total variance in the data can be explained by the first two components. Strains C57BL/6J and FVB/NJ are clearly discriminated from strains BALB/cByJ and DBA/2J by principal component one. In Figure 4B, the autocorrelation matrix, comparing each biological replicate with each other reveals that, in general, there are positive correlations in normalized MS intensity between biological replicates of the same strains, indicating good experimental repeatability. In addition, comparing biological replicates across the 4 strains shows negative Pearson's correlation coefficients for strains that are dissimilar. Combined the PCA and the correlation plots infer that more similarity is shared between the pairs C57BL/6J and FVB/NJ, and likewise between the pairs DBA/2J and BALB/cByJ, at least at the brain proteoform expression level. A previous comprehensive genetic study has been reported to reconstruct the phylogenetic relationships among the 102 inbred strains including those four strains using an Amplifluor genotyping system.[35] In that work, C57BL/6J, DBA/2J, FVB/NJ, and BALB/cByJ strains were organized into different groups based on single-nucleotide polymorphism (SNP) markers. We further revealed here that, at the proteoform level, more similarity is shared between the pairs C57BL/6J and FVB/NJ, and likewise between the pairs DBA/2J and BALB/cByJ showing the power of TDP to complement traditional genetic markers for strain differentiation.

Multiple phenotypic and behavioral differences among various inbred laboratory strains of mice are well-known and often obvious (body coloration and size).[36−38] Multiple efforts have been made in the past to determine the relationship between these parameters and the genomes[5,39,40] or transcriptomes[41,42] of these particular strains. These measurements have revealed significant differences in the molecular blueprints of the studied strains. However, despite the importance of genomic and transcriptomic information, there is noticeable lack of comprehensive proteomics data. Such data are critical for understanding the structural and functional organization of biological systems. A number of reports demonstrate strain-

DOI: 10.1021/acs.analchem.7b04108
Anal. Chem. 2018, 90, 3802−3810

dependent proteome characteristics including plasma,[43] hair,[44,45] and liver.[46] Not surprisingly, low levels of correlation have been observed between transcripts and proteins.[46] Our data illustrate important biological differences that can be inferred from proteomic measurements as demonstrated by gene ontology analysis.

For this assessment, we evaluated our quantitative results using GeneGo MetaCore in an effort to discover an underlying biological perspective in the lists (Figures S-6 to S-9). We used as input the proteins underlying the subset of proteoforms in Figure 3 that passed the significance threshold described therein. We relate the results of the GeneGo analysis to known phenotypes of strains C57BL/6J and DBA/2J. In studies that examine the rewarding effects of cocaine, C57BL/6 mice show behavior indicative of high reward, unlike DBA/2 mice.[21] Following opioid administration these two strains show an opposite response in locomotor stimulant testing.[21] Drug addiction is widely regarded as a brain disease because neuronal dysfunction and neurotoxicity accompany the abuse of drugs. Drugs of abuse such as amphetamines, cocaine, and opioid drugs, have been shown to induce mitochondrial dysfunction.[47] Interestingly, the top six scoring overestimated terms in GO localization for DBA/2J relative to C57BL/6J were tracked to the mitochondria (Figure S-8). Also, the top six scoring overestimated terms in GO Processes and the top six in GO Molecular Function for DBA/2J relative to C57BL/6J were mitochondria focused (Figures S-7 and S-8). Hence, the strain resolved brain proteoform differences identified here by TDP could be involved in how the two strains respond differently to cocaine and opioids. Furthermore, the response profile for BALB/cByJ was similar to that of DBA/2J, whereas the GeneGO results show that the FVB/NJ and C57BL/6J strains are more similar. These results mirror the data shown in Figure 4. We emphasize here that the strain similarities and dissimilarities displayed in Figure 4 are based on relative ion intensities. The strain similarities and dissimilarities displayed in Figure S-6 are rooted in the biology of the proteoforms that yielded those ion intensities.

CONCLUSIONS

We present the first application of a label-free top-down proteomics method for the analysis of mouse brain tissue and demonstrate that physiochemical differences among four inbred strains can be determined by sampling proteoforms in the 3.5−30 kDa mass range. We achieved this differentiation despite the upper mass limitation which is set by the ability of the mass spectrometer to resolve isotopologues in a chromatographic time scale as required by downstream data analysis algorithms for accurate mass assignment. Assessing differences in protein complement and its dynamics in the murine brain could contribute to a better understanding of phenotypical differences among strains. Neuroscientists study the murine brain as a model organism to better understand the human brain and human biology because of the genetic and physiological similarities between the species. Most often researchers attempt to gain phenotype knowledge through genetic screening or with bottom-up proteomics. The new screening approach of rodent model tissue described here is important because top-down proteomics analyzes the entire protein molecule without conversion to peptides, unlike BUP. Thus, deep phenotypic information is retained enabling the detection of degradation products, sequence variants, and combinations of post-translational modifications. Moreover, we report the use of two new software tools, TDPortal and TDViewer, which provide significant improvements over processes used in our previous top-down proteomic studies. They enable faster processing of raw data and highly confident qualitative and label-free quantitative analysis of proteins and proteoforms. Finally, a fully characterized and mouse strain-resolved proteoform database has been established and offered as a web-based resource to the neuroscience community.

With the validation afforded by this work, in future studies we look to enhance throughput by reducing the number of technical replicates while simultaneously increasing the number of samples several fold. Furthermore, the technologies utilized in TDP are still improving. Faster scan speeds, more efficient protein fragmentation techniques available on mass spectrometers, as well as improved low resolving power data deconvolution algorithms for accurate determination of protein mass and signal intensity will extend the mass range that is accessible by top-down proteomics within a chromatographic time scale. Future experiments are planned to extend this work to study the effect of drugs of abuse specifically on these different inbred mouse strains as a means to better understand the mechanisms of drug dependency in humans.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.7b04108.

Method S1 and Figures S1−S9 (PDF).
Table S1 (XLSX).
Table S2 (XLSX).

AUTHOR INFORMATION

Corresponding Author

*Tel.: 1 847-467-4362. Fax: 1 847 467-3276. E-mail: n-kelleher@northwestern.edu.

ORCID

Roderick G. Davis: 0000-0002-9314-6644
Elena V. Romanova: 0000-0002-2040-3312
Stanislav S. Rubakhin: 0000-0003-0437-1493
Steven M. Patrie: 0000-0002-0308-5498
Paul M. Thomas: 0000-0003-2887-4765
Jonathan V. Sweedler: 0000-0003-3107-9922
Neil L. Kelleher: 0000-0002-8815-3372

Present Address

§Department of Convergence Medicine, Asan Medical Center, University of Ulsan College of Medicine, 88 Olympic-ro 43-gil, Songpa-gu, Seoul, 05505, Korea (K.K.); Process Development, Amgen Inc., One Amgen Center Drive, Thousand Oaks, CA 91320, United States (C.W.); Massachusetts Institute of Technology, NE18−501, 255 Main Street, Cambridge, MA 02142, United States (J.Z.).

Author Contributions

The manuscript was written by R.G.D., H.-M.P., and P.M.T. with contributions from R.T.F., R.D.L., S.M.P., E.V.R., and S.S.R. under the direction of J.V.S. and N.L.K. R.G.D., K.K., and H.-M.P. designed the work flow of the project and performed experiments. E.V.R., S.S.R., C.W., P.M.Y., J.A.Z., and J.S.R. prepared brain tissues from all animals and performed preliminary experiments. R.D.L., R.T.F., J.B.G., and A.J.N.

DOI: 10.1021/acs.analchem.7b04108
Anal. Chem. 2018, 90, 3802−3810

performed data and statistical analysis. All authors have given approval to the final version of the manuscript.

Author Contributions

†These authors contributed equally to this work.

Funding

The following grants are acknowledged: NIH P30 DA018310, NIH P41 GM108569.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The project described was supported by Award No. P30 DA018310 from the National Institute on Drug Abuse (J.V.S. and N.L.K.). The development of TDPortal 1.3 and TDViewer was conducted under the guidance of the National Resource for Translational and Developmental Proteomics under Grant P41 GM108569 from the National Institute of General Medical Sciences, National Institutes of Health.

■ ABBREVIATIONS

ACN, acetonitrile; AGC, automatic gain control; BCA, bicinchoninic acid assay; FDR, false discovery rate; GELFrEE, gel elution liquid fractionation entrapment electrophoresis; HCD, higher energy collisional dissociation; MeOH, methanol; MS, mass spectrometry; MS1, intact/precursor scan; MS2 (or MS/MS), tandem mass spectrometry scan, fragmentation; nLC, nanoliquid chromatography; SDS, sodium dodecyl sulfate; UIUC IACUC, University of Illinois at Urbana−Champaign Institutional Care and Use Committee.

■ REFERENCES

(1) Peterson, A. S. *Genome Res.* **2002**, *12*, 217−218.
(2) Lloyd, B. J.; et al. *Nat. Genet.* **2000**, *24*, 23−25.
(3) Flint, J.; Eskin, E. *Nat. Rev. Genet.* **2012**, *13*, 807−817.
(4) Frazer, K. A.; Eskin, E.; Kang, H. M.; Bogue, M. A.; Hinds, D. A.; Beilharz, E. J.; Gupta, R. V.; Montgomery, J.; Morenzoni, M. M.; Nilsen, G. B.; Pethiyagoda, C. L.; Stuve, L. L.; Johnson, F. M.; Daly, M. J.; Wade, C. M.; Cox, D. R. *Nature* **2007**, *448*, 1050−1053.
(5) Keane, T. M.; Goodstadt, L.; Danecek, P.; White, M. A.; Wong, K.; Yalcin, B.; Heger, A.; Agam, A.; Slater, G.; Goodson, M.; Furlotte, N. A.; Eskin, E.; Nellaker, C.; Whitley, H.; Cleak, J.; Janowitz, D.; Hernandez-Pliego, P.; Edwards, A.; Belgard, T. G.; Oliver, P. L.; et al. *Nature* **2011**, *477*, 289−294.
(6) Yang, H.; Bell, T. A.; Churchill, G. A.; Pardo-Manuel de Villena, F. *Nat. Genet.* **2007**, *39*, 1100−1107.
(7) Crawley, J. N. *Trends Neurosci.* **1996**, *19*, 181−182 discussion 188−189.
(8) Chesler, E. J.; Lu, L.; Shou, S.; Qu, Y.; Gu, J.; Wang, J.; Hsu, H. C.; Mountz, J. D.; Baldwin, N. E.; Langston, M. A.; Threadgill, D. W.; Manly, K. F.; Williams, R. W. *Nat. Genet.* **2005**, *37*, 233−242.
(9) Wahlsten, D.; Bachmanov, A.; Finn, D. A.; Crabbe, J. C. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 16364−16369.
(10) Ishihama, Y.; Sato, T.; Tabata, T.; Miyamoto, N.; Sagane, K.; Nagasu, T.; Oda, Y. *Nat. Biotechnol.* **2005**, *23*, 617−621.
(11) Kislinger, T.; Cox, B.; Kannan, A.; Chung, C.; Hu, P.; Ignatchenko, A.; Scott, M. S.; Gramolini, A. O.; Morris, Q.; Hallett, M. T.; Rossant, J.; Hughes, T. R.; Frey, B.; Emili, A. *Cell* **2006**, *125*, 173−186.
(12) Klose, J.; Nock, C.; Herrmann, M.; Stuhler, K.; Marcus, K.; Bluggel, M.; Krause, E.; Schalkwyk, L. C.; Rastan, S.; Brown, S. D.; Bussow, K.; Himmelbauer, H.; Lehrach, H. *Nat. Genet.* **2002**, *30*, 385−393.
(13) Price, J. C.; Guan, S.; Burlingame, A.; Prusiner, S. B.; Ghaemmaghami, S. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 14508−14513.
(14) Sharma, K.; Schmitt, S.; Bergner, C. G.; Tyanova, S.; Kannaiyan, N.; Manrique-Hoyos, N.; Kongi, K.; Cantuti, L.; Hanisch, U. K.; Philips, M. A.; Rossner, M. J.; Mann, M.; Simons, M. *Nat. Neurosci.* **2015**, *18*, 1819−1831.
(15) Jung, S. Y.; Choi, J. M.; Rousseaux, M. W.; Malovannaya, A.; Kim, J. J.; Kutzera, J.; Wang, Y.; Huang, Y.; Zhu, W.; Maity, S.; Zoghbi, H. Y.; Qin, J. *Mol. Cell. Proteomics* **2017**, *16*, 581−593.
(16) Nesvizhskii, A. I.; Aebersold, R. *Mol. Cell. Proteomics* **2005**, *4*, 1419−1440.
(17) Ntai, I.; Kim, K.; Fellers, R. T.; Skinner, O. S.; Smith, A. D. t.; Early, B. P.; Savaryn, J. P.; LeDuc, R. D.; Thomas, P. M.; Kelleher, N. L. *Anal. Chem.* **2014**, *86*, 4961−4968.
(18) Savaryn, J. P.; Toby, T. K.; Catherman, A. D.; Fellers, R. T.; LeDuc, R. D.; Thomas, P. M.; Friedewald, J. J.; Salomon, D. R.; Abecassis, M. M.; Kelleher, N. L. *Proteomics* **2016**, *16*, 2048−2058.
(19) Toby, T. K.; Abecassis, M.; Kim, K.; Thomas, P. M.; Fellers, R. T.; LeDuc, R. D.; Kelleher, N. L.; Demetris, J.; Levitsky, J. *Am. J. Transplant.* **2017**, *17*, 2458−2467.
(20) Li, W.; Petruzziello, F.; Zhao, N.; Zhao, H.; Ye, X.; Zhang, X.; Rainer, G. *Proteomics* **2017**, *17*, 1600419.
(21) Crawley, J. N.; Belknap, J. K.; Collins, A.; Crabbe, J. C.; Frankel, W.; Henderson, N.; Hitzemann, R. J.; Maxson, S. C.; Miner, L. L.; Silva, A. J.; Wehner, J. M.; Wynshaw-Boris, A.; Paylor, R. *Psychopharmacology* **1997**, *132*, 107−124.
(22) Deroche, V.; Caine, S. B.; Heyser, C. J.; Polis, I.; Koob, G. F.; Gold, L. H. *Pharmacol., Biochem. Behav.* **1997**, *57*, 429−440.
(23) Roberts, A. J.; Polis, I. Y.; Gold, L. H. *Eur. J. Pharmacol.* **1997**, *326*, 119−125.
(24) Blednov, Y. A.; Metten, P.; Finn, D. A.; Rhodes, J. S.; Bergeson, S. E.; Harris, R. A.; Crabbe, J. C. *Alcohol.: Clin. Exp. Res.* **2005**, *29*, 1949−1958.
(25) Zombeck, J. A.; Swearingen, S. P.; Rhodes, J. S. *Genes, Brain, and Behavior* **2010**, *9*, 892−898.
(26) Romanova, E. V.; Rubakhin, S. S.; Ossyra, J. R.; Zombeck, J. A.; Nosek, M. R.; Sweedler, J. V.; Rhodes, J. S. J. *J. Neurochem.* **2015**, *135*, 1038−1048.
(27) Wessel, D.; Flügge, U. I. *Anal. Biochem.* **1984**, *138*, 141−143.
(28) Afgan, E.; Baker, D.; van den Beek, M.; Blankenberg, D.; Bouvier, D.; Cech, M.; Chilton, J.; Clements, D.; Coraor, N.; Eberhard, C.; Gruning, B.; Guerler, A.; Hillman-Jackson, J.; Von Kuster, G.; Rasche, E.; Soranzo, N.; Turaga, N.; Taylor, J.; Nekrutenko, A.; Goecks, J. *Nucleic Acids Res.* **2016**, *44*, W3−w10.
(29) Pesavento, J. J.; Kim, Y.; Taylor, G. K.; Kelleher, N. L. J. *J. Am. Chem. Soc.* **2004**, *126*, 3386−3387.
(30) Littell, R. C.; Stroup, W. W.; Freund, R. J.; Littell, R. C. *Books 24 × 7 Inc., Wiley Series in Probability and Statistics*; SAS Institute: Cary, N.C., U.S.A., 2002.
(31) LeDuc, R. D.; Boyne, M. T., 2nd; Townsend, R. R.; Bose, R. *RECOMB Satellite Conference on Computational Proteomics 2010*; University of California: San Diego, 2010.
(32) LeDuc, R. D.; Fellers, R. T.; Early, B. P.; Greer, J. B.; Thomas, P. M.; Kelleher, N. L. J. *Proteome Res.* **2014**, *13*, 3231−3240.
(33) Caporaso, G. L.; Bibb, J. A.; Snyder, G. L.; Valle, C.; Rakhilin, S.; Fienberg, A. A.; Hemmings, H. C., Jr; Nairn, A. C.; Greengard, P. *Neuropharmacology* **2000**, *39*, 1637−1644.
(34) Cohen, P. *Trends Biochem. Sci.* **2000**, *25*, 596−601.
(35) Petkov, P. M.; Ding, Y.; Cassell, M. A.; Zhang, W.; Wagner, G.; Sargent, E. E.; Asquith, S.; Crew, V.; Johnson, K. A.; Robinson, P.; Scott, V. E.; Wiles, M. V. *Genome Res.* **2004**, *14*, 1806−1811.
(36) International Mouse Phenotyping Consortium; http://www.mousephenotype.org (accessed September, 2017).
(37) Mouse Genome Informatics; http://www.informatics.jax.org (accessed September, 2017).
(38) JAX Mice Clinical and Research Services Catalog (June 2017−May 2018, p. 47); Retrieved from http://jackson.jax.org/rs/444-BUH-304/images/LT0119_01_Catalog_2017-18_06082017_WEB.pdf.
(39) Fernandes, C.; Paya-Cano, J. L.; Sluyter, F.; D'Souza, U.; Plomin, R.; Schalkwyk, L. C. *European Journal of Neuroscience* **2004**, *19*, 2576−2582.

DOI: 10.1021/acs.analchem.7b04108
Anal. Chem. 2018, 90, 3802−3810

(40) Doran, A. G.; Wong, K.; Flint, J.; Adams, D. J.; Hunter, K. W.; Keane, T. M. *Genome Biol.* **2016**, *17*, 167.
(41) Morris, J. A.; Royall, J. J.; Bertagnolli, D.; Boe, A. F.; Burnell, J. J.; Byrnes, E. J.; Copeland, C.; Desta, T.; Fischer, S. R.; Goldy, J.; Glattfelder, K. J.; Kidney, J. M.; Lemon, T.; Orta, G. J.; Parry, S. E.; Pathak, S. D.; Pearson, O. C.; Reding, M.; Shapouri, S.; Smith, K. A.; Soden, C.; Solan, B. M.; Weller, J.; Takahashi, J. S.; Overly, C. C.; Lein, E. S.; Hawrylycz, M. J.; Hohmann, J. G.; Jones, A. R. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 19049−19054.
(42) Turk, R.; 't Hoen, P. A.; Sterrenburg, E.; de Menezes, R. X.; de Meijer, E. J.; Boer, J. M.; van Ommen, G.-J. B.; den Dunnen, J. T. *BMC Genomics* **2004**, *5*, 57.
(43) Pamir, N.; Hutchins, P.; Ronsein, G.; Vaisar, T.; Reardon, C. A.; Getz, G. S.; Lusis, A. J.; Heinecke, J. W. J. *J. Lipid Res.* **2016**, *57*, 246−257.
(44) Rice, R. H.; Rocke, D. M.; Tsai, H.-S.; Silva, K. A.; Lee, Y. J.; Sundberg, J. P. J. *J. Invest. Dermatol.* **2009**, *129*, 2120−2125.
(45) Rice, R. H.; Bradshaw, K. M.; Durbin-Johnson, B. P.; Rocke, D. M.; Eigenheer, R. A.; Phinney, B. S.; Sundberg, J. P. *PLoS One* **2012**, *7*, e51956.
(46) Ghazalpour, A.; Bennett, B.; Petyuk, V. A.; Orozco, L.; Hagopian, R.; Mungrue, I. N.; Farber, C. R.; Sinsheimer, J.; Kang, H. M.; Furlotte, N.; Park, C. C.; Wen, P.-Z.; Brewer, H.; Weitz, K.; Camp, D. G., II; Pan, C.; Yordanova, R.; Neuhaus, I.; Tilford, C.; Siemers, N.; et al. *PLoS Genet.* **2011**, *7*, e1001393.
(47) Cunha-Oliveira, T.; Rego, A. C.; Oliveira, C. R. *Brain Res. Rev.* **2008**, *58*, 192−208.

DOI: 10.1021/acs.analchem.7b04108
Anal. Chem. 2018, 90, 3802−3810