Towards Comprehensive 3D Proteome Imaging of the Mouse Brain by Voxelation Integrated with High-Throughput Proteomics

Vladislav A. Petyuk[1], Wei-Jun Qian[1], Mark H. Chin[2], Haixing Wang[1], Eric A. Livesay[1], Matthew E. Monroe[1], Joshua N. Adkins[1], Navdeep Jaitly[1], Xiuxia Du[1], David J. Anderson[1], David G. Camp II[1], Desmond J. Smith[2], and Richard D. Smith[1]

[1]Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, P.O. Box 999, Richland, WA 99352
[2]Department of Molecular and Medical Pharmacology, UCLA School of Medicine, Los Angeles, CA 90095

Overview

We describe a methodology that combines a voxelation technique [1] with microscale automated sample processing and an LC-FTICR MS-based quantitative proteomic approach [2] for spatial imaging of protein abundances in mouse brain with 1 mm resolution.

As a proof-of-concept trial we applied this methodology for analysis of protein distribution within a coronal slice of a C57BL/6J mouse brain. For accurate quantitation of peptides across the slice we spiked an isotopically labeled reference sample into each voxel sample and used isotopic intensity ratios as accurate measurements of peptide abundances. As a reference sample we used the ^{18}O-labeled tryptic peptides derived from a whole control coronal slice, thus providing a reference for virtually all peptides.

In total, 1028 proteins were identified and quantified using an accurate mass and time (AMT) tag strategy and 2D protein abundance images for these proteins are generated. The 1 mm spatial resolution was sufficient to distinguish major structural components of the brain in this coronal section, i.e., cerebral cortex, striatum, diencephalon, white matter and olfactory tubercles.

We classified 283 proteins into 5 clusters having apparent biases in protein abundance toward specific brain structures. 29 proteins had an increased abundance in the ventromedial area collocating with diecephalon, 59 in the cortex, 17 in the striatum, and 28 in the central part of the section collocating with white matter structures like corpus callosum, fornix and anterior commissure. In addition, 150 patterns were classified as uniform, i.e. without any significant bias in abundance across the slice. Of those protein distributions 68%, 82%, 100%, and 52%, correspondingly, agreed with mRNA distributions from GENSAT and Allen Brain Atlas project providing an evidence for validity of methodology.

Introduction

The immense molecular complexity of the mammalian brain is a significant challenge in applying high-throughput technologies in a discovery-driven fashion. Besides that, an additional challenge is dictated by the importance of taking into account the brain's complicated anatomy and 3D structure for understanding of its functioning. A few currently ongoing efforts (e.g., Allen Brain Atlas Project and GENSAT) independently aim at comprehensive spatial gene expression characterization of mouse brain using *in situ* hybridization technology.

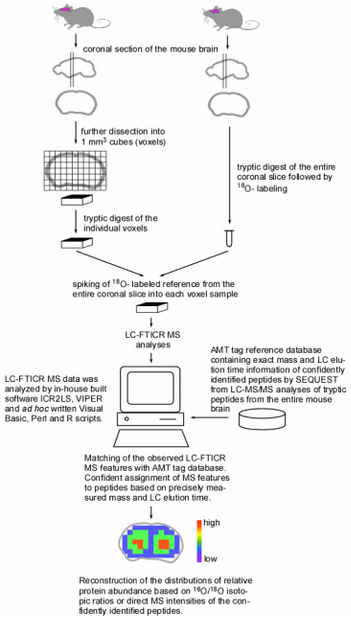
Despite high automation comprehensive discovery-based analyses of multiple types of brains that represent different disease models and their progression states is a very challenging task.

To contribute to 3D-proteome mapping method development we designed an approach that utilizes tissue voxelation and high-throughput quantitative proteomics and tested it on a mouse brain coronal section.

Methods

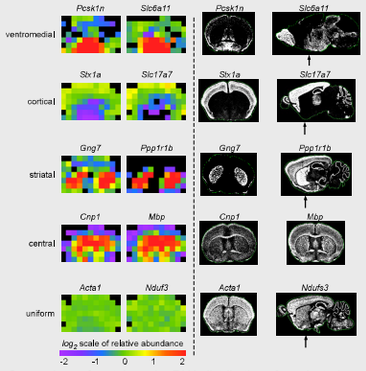
- **Mouse brain tissue samples and voxel samples** were prepared from C57BL6 mice at UCLA.
- **Sample preparation:** Brain tissue samples were lysed using hypotonic buffer and trifluoroethanol (TFE) and digested with trypsin. Sample processing for voxel samples was automated and performed on a Beckman Biomek FX liquid handling robot.
- **LC-FTICR MS analyses** were performed using a PNNL-developed 11T FTICR mass spectrometer interfaced with 150 mm i.d. x 65 cm long LC column packed with 3 mm Jupiter C18 particles (Phenomenex)
- **Data analysis:** FTICR spectra deisotoping was performed by Horn mass transform using ICR2LS (an in house developed software package available at http://ncrr.pnl.gov). Detection of MS features and matching to the peptides from AMT tag database was done using VIPER software. *Ad hoc* written Visual Basic, Perl and R scripts were used for statistical filtering of MS feature/peptide matches and robust quantitation of protein abundances.

Overview of the Approach



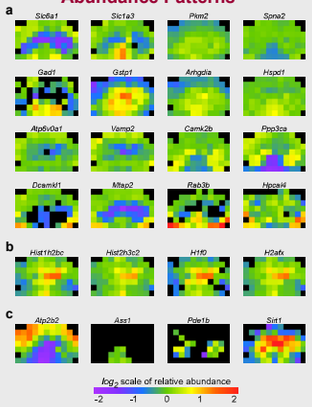
Results

Selected Representatives of the Major Abundance Patterns



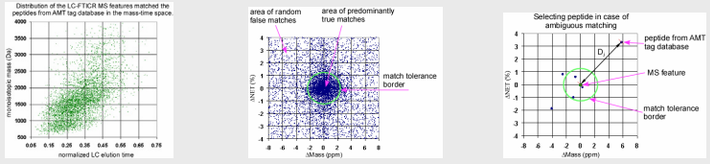
Comparison of protein and mRNA distributions for selected genes with characteristic abundance patterns: cortical, ventromedial, striatal, central and uniform. The determined protein relative abundance data are on the left side. On the right side are the mRNA abundance data provided under license by The Allen Institute for Brain Science (www.brainatlas.org). Sagital sections were used when the coronal section data were not available. The arrows on the sagital sections indicate the corresponding position of bregma 0 coronal section.

Selected Representative Proteins with Complex and Newly Characterized Abundance Patterns



(a) Examples of proteins with complex or indeterminate abundance patterns. (b) Histone proteins classified as with predominant abundance in the center have no correlation with mRNA abundance data. (c) Selected examples of protein abundance patterns of genes with previously uncharacterized expression patterns in the mouse brain.

Basic Principles and Statistics of Peptide Identification in the Brain Voxel Samples Using AMT tag Approach



Probabilistic distance D_t. Where $\sigma\Delta$Mass and $\sigma\Delta$NET are standard deviations of the distributions of mass and NET measurement errors, respectively.

$$D_t = \sqrt{\left(\frac{\Delta Mass_t}{\sigma_{\Delta Mass}}\right)^2 + \left(\frac{\Delta NET_t}{\sigma_{\Delta NET}}\right)^2}$$

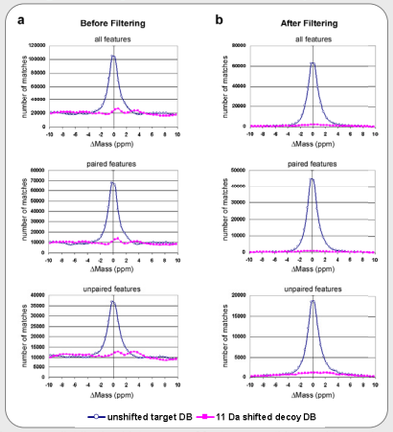
Probability density function of a true match to be at the distance D_t assuming bivariate normal distribution of mass and NET measurement error residuals with no mutual correlation.

$$p(D_t \mid +) = \frac{1}{2\pi\sigma_{\Delta Mass}\sigma_{\Delta NET}} e^{(-D_t^2/2)}$$

Assuming equal *a priori* probability π of all peptide matches, the probability of the peptide match at distance D_j to be true $P(+|D_j)$ (SLiC score) according to the Bayes law is:

$$P(+ \mid D_t) = \frac{\pi_t p(D_t \mid +)}{\sum_{i=1}^{N} \pi_i p(D_i \mid +)} = \frac{e^{(-D_t^2/2)}}{\sum_{i=1}^{N} e^{(-D_i^2/2)}}$$

Estimation of False Discovery Rates for Peptide and Protein Identifications using Composite Target/Decoy Database for AMT approach



Mass error distribution histograms of feature - peptide matching to the merged target/decoy AMT tag database, before (**a**) and after (**b**) applying step 1 filtering* except mass error tolerance cutoff. (**a**) The density of mass error distribution of matches to the normal unshifted database is a combination of unimodal normal-like distribution of true matches and uniform distribution of random false matches. The uniform component of the distribution representing random matches is well approximated by the distribution of matches to the 11 Da shifted database. (**b**) The FDR rate after step 1 filtering* can be estimated as a ratio of area beneath the magenta curve representing matches to the shifted database to the area beneath the blue curve representing matches to the normal database within respective mass error tolerances: ±2.3 ppm for paired features and ±2.0 ppm for unpaired features.

	paired	unpaired	overall
step1*, peptides	2.7	8.6	4.5
step2**, peptides	1.9	6.4	3.2
step2**, proteins			2.6

* Step1 data filtering criteria is based on mass and elution time differences between observed features and matched peptide from reference database. We retained only those matches that had D < 1.645 for paired features and D < 1.44 for unpaired features, thereby retaining 90% and 85%, respectively, of true matches (assuming normality of mass and NET measurement error distributions). Probabilistic distances D = 1.645 and D=1.44 define ellipses with *a* and *b* parameters corresponding to tolerances of ±2.3 ppm mass, ±1.2 % NET, and ±2.0 ppm mass, ±1.1 % NET, respectively. Further, we used only those peptide matches that had a SLiC (spatially localized confidence score) score >0.3 for paired features and only the top matches, i.e., peptides with the highest SLiC score matching the same feature, for unpaired features.

** A Step2 filtering criterion is based on number of peptide hits per protein. Peptides from proteins having only one hit were discarded.

Conclusions

- In a proof-of-ability experiment on a C57BL/6J mouse bregma 0 coronal section using the AMT tag approach we identified and quantified 1028 proteins having two or more unique peptides.
- The "universal" reference sample, made by ^{18}O labeling of peptides derived from a tryptic digest of the whole coronal section of control mouse, made robust quantitation possible for the majority of the cases: 838 proteins out of 1028.
- The majority of the protein abundance patterns correlates well with corresponding mRNA abundance patterns from GENSAT and Allen Brain Atlas, which highlights the validity of the methodology.
- To evaluate the confidence of the protein identification and quantitation we applied a shifted database approach. The behavior of matches to the 11 Da shifted database emulates the behavior of false positive matches to the normal database. The resulting estimated percentages of false positive peptide and protein identifications are 3.2% and 2.6%, respectively.

Acknowledgements

Portions of the research were supported by the NIH National Center for Research Resources (RR18522 to Richard D. Smith) and the NIH grants R01 DA015802 and R01 NS050148 to Desmond J. Smith.

Proteomic analyses were performed in the Environmental Molecular Sciences Laboratory, a U.S. Department of Energy (DOE) national scientific user facility located at the Pacific Northwest National Laboratory (PNNL) in Richland, Washington. PNNL is a multiprogram national laboratory operated by Battelle for the DOE under Contract No. DE-AC05-76RLO 1830.

The authors also would like to thank The Allen Institute for Brain Science for permission to present their data.

References

1. Liu, D. & Smith, D.J. Voxelation and gene expression tomography for the acquisition of 3-D gene expression maps in the brain. *Methods* **31**, 317-325 (2003).
2. Qian, W.J., Camp, D.G., 2nd & Smith, R.D. High-throughput proteomics using Fourier transform ion cyclotron resonance mass spectrometry. *Expert Rev Proteomics* **1**, 87-95 (2004).
3. Petyuk, V.A., Qian, W.J., Chin, M.H., Wang, H., Livesay, E., Monroe, M., Adkins, J., Jaitly, N., Anderson., D.J., Camp, D.G. II, Smith, D.J., Smith, R.D. Towards Comprehensive 3D Proteome Imaging of the Mouse Brain by Voxelation Integrated with High-Throughput Proteomics. *Submitted to Nature Methods*.

Contact Information

Vlad Petyuk, Ph. D.
Biological Sciences Division, K8-98
Pacific Northwest National Laboratory
P.O. Box 999, Richland, WA 99352
e-mail: name vladislav.petyuk@pnl.gov
Phone: (509) 376-0056