

Transcriptome and proteome analysis: a perspective on correlation

Opinion

Druggable proteins (such as enzymes, receptors, transporters and channel proteins, signal transduction proteins, oncogenes and transcription factors) are powerful drug targets for therapeutics of diverse diseases.¹⁻⁴ Among the approximately 22,000 human proteins, novel drug targets are likely to emerge.⁵ Numerous bioinformatics tools are used to monitor the gene expression at mRNA level, including UniGene and Expressed Sequence Tag, EST,^{6,7} Serial Analysis of Gene Expression, SAGE,⁸ Digital Differential Display, DDD,^{9,10} Cancer Genome Anatomy Project X-Profiler and Digital Gene Expression Display.^{11,12} High throughput transcriptome analysis is often performed using Microarray,¹³⁻¹⁵ and Next-Generation RNA sequencing, NGS RNA-seq.^{16,17}

In recent years proteome analysis has been greatly aided by various protein expression tools, the Allan Brain Atlas,^{18,19} Human Protein Atlas,²⁰ Multi Omics Profiling Expression Database, MOPED,²¹ the Human Protein Reference Database, HPRD^{22,23} and the recently described Human Proteome Map²⁴ and proteomics Db. The last two mentioned studies together account for approximately 84% of the total annotated protein-coding genes in humans. Together, these two studies identified >18,000 proteins encoded by both known genes and uncharacterized Open Reading Frames. Over 20,000 protein isoforms expressions were also characterized.^{25,26}

Numerous normal tissues, body fluids and cell lines were used in these two studies. These recently described protein expression databases greatly expand our ability to establish correlative evidence for mRNA and protein expression for most of the human proteome.

Based on mRNA expression using a high throughput transcriptome analysis, interpretations are often made about the functional relevance, pathways, interacting proteins and the possibility of drug therapy use. Protein expression is often inferred, but frequently a correlation of mRNA versus protein is missing in studies. The complex regulation of protein expression at the level of noncoding RNAs, DNA methylation, epigenetic changes, gene amplifications, copy number variations, mutations, post translational modifications such as acetylation, amidation, myristylation, phosphorylation, sumoylation, Ubiquitination etc., stability and degradation as well as the interacting proteins adds to the complexity of transcriptome-based interpretations.^{27,28}

For identifying candidate driver genes for therapeutic target discovery and functional elucidation, knowledge of mRNA and protein correlation is critical. The vast amount of data generated by microarray and the NGS RNA-sequencing technology can be reduced to a manageable number of putative targets using the correlation at the mRNA and the protein levels. In a recent study, Zhang et al.,²⁹ using The Cancer Genome Atlas (TCGA) tissues (n=87) demonstrated that the mRNA transcript abundance and gene amplifications did not often correlate. Among 3,764 genes analyzed, only 32% showed statistically significant correlation. Further, copy number alterations had no significant effects on the protein levels. Our own results as well as several other reports support such a lack of correlation.³⁰⁻³⁴

Volume 1 Issue 5 - 2014

Ramaswamy Narayanan,¹ Wim J M Van De Ven²¹Department of Biological Sciences, Florida Atlantic University, USA²Department of Human Genetics, University of Leuven, Belgium

Correspondence: Ramaswamy Narayanan, Department of Biological Sciences, Charles E. Schmidt College of Science, Florida Atlantic University, Boca Raton, FL 33431 USA, Tel 5612972247, Fax 5612973859, Email rnarayan@fau.edu

Received: September 07, 2014 | **Published:** September 29, 2014

During the early stages of microarray technology, vast data sets were generated with no standards. The establishment of Minimum Information About a Microarray Experiment, MIAME,³⁵ greatly reduced the noise in subsequent array-based datasets. Similar guidelines for RNA and protein expression are needed to develop meaningful interpretations. Where possible, the investigators and authors should be expected to demonstrate such a correlation using the protein expression tools, which are becoming increasingly available. This would greatly help reduce the noise in the published literature, verify functional roles for the proteins and optimize the chances of identifying candidate driven genes for druggableness.

Further, the development of single cell transcriptome and proteome analysis capability (in contrast to whole tissue-based analysis as is currently done) would allow for a single cell-based correlation between mRNA and proteins. This will allow for cellular heterogeneity in expression profiling in a tissue environment. An offshoot of the protein expression databases, such as creation of a separate human proteome database showing correlation with the transcriptome, would greatly aid in the therapeutic target discovery.

Acknowledgements

We thank Jeanine Narayanan for editorial assistance.

Conflict of interest

The author declares no conflict of interest.

References

1. Hopkins AL, Groom CR. The druggable genome. *Nature reviews Drug discovery*. 2002;1(9):727-730.
2. Landry Y, Gies JP. Drugs and their molecular targets: an updated overview. *Fundam clin pharmacol*. 2008;22(1):1-18.
3. Russ AP, Lampel S. The druggable genome: an update. *Drug Discov today*. 2005;10(23-24):1607-1610.
4. Workman P, Al-Lazikani B, Clarke PA. Genome-based cancer therapeutics: targets, kinase drug resistance and future strategies for precision oncology. *Curr Opin Pharmacol*. 2013;13(4):486-496.

5. Griffith M, Griffith OL, Coffman AC, et al. DGIdb: mining the druggable genome. *Nat Methods*. 2013;10(12):1209–1210.
6. Wheeler DL, Church DM, Federhen S, et al. Database resources of the National Center for Biotechnology. *Nucleic Acids Res*. 2003;31(1):28–33.
7. Boguski MS, Schuler GD. E S Tablishing a human transcript map. *Nat Genet*. 1995;10(4):369–371.
8. Velculescu VE, Zhang L, Vogelstein B, et al. Serial analysis of gene expression. *Science*. 1995;270(5235):484–487.
9. Strausberg RL, Buetow KH, Emmert–Buck MR, et al. The cancer genome anatomy project: building an annotated gene index. *Trends genet*. 2000;16(3):103–106.
10. Scheurle D, DeYoung MP, Binninger DM, et al. Cancer gene discovery using digital differential display. *Cancer Res*. 2000;60(15):4037–4043.
11. Strausberg RL. The Cancer Genome Anatomy Project: new resources for reading the molecular signatures of cancer. *J Pathol*. 2001;195(1):31–40.
12. Lauriola M, Ugolini G, Rosati G, et al. Identification by a Digital Gene Expression Displayer (DGED) and test by RT–PCR analysis of new mRNA candidate markers for colorectal cancer in peripheral blood. *Int J Oncol*. 2010;37(2):519–525.
13. Rhodes DR, Kalyana–Sundaram S, Mahavisno V, et al. Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia*. 2007;9(2):166–180.
14. Liu F, White JA, Antonescu C, et al. GCOD – GeneChip Oncology Database. *BMC Bioinformatics*. 2011;12:46.
15. Parkinson H, Sarkans U, Shojatalab M, et al. ArrayExpress—a public repository for microarray gene expression data at the EBI. *Nucleic acids Res*. 2005;33(Database issue):D553–D555.
16. Mortazavi A, Williams BA, McCue K, et al. Mapping and quantifying mammalian transcriptomes by RNA–Seq. *Nature Methods*. 2008;5(7):621–628.
17. Nagalakshmi U, Wang Z, Waern K, et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*. 2008;320(5881):1344–1349.
18. Miller JA, Ding SL, Sunkin SM, et al. Transcriptional landscape of the prenatal human brain. *Nature*. 2014;508(7495):199–206.
19. Shen EH, Overly CC, Jones AR. The Allen Human Brain Atlas: comprehensive gene expression mapping of the human brain. *Trends Neurosci*. 2012;35(12):711–714.
20. Uhlen M, Oksvold P, Fagerberg L, et al. Towards a knowledge–based Human Protein Atlas. *Nat Biotechnol*. 2010;28(12):1248–1250.
21. Kolker E, Higdon R, Haynes W, et al. MOPED: Model Organism Protein Expression Database. *Nucleic acids Res*. 2012;40(Database issue):D1093–D1099.
22. Mathivanan S, Ahmed M, Ahn NG, et al. Human Proteinpedia enables sharing of human protein data. *Nat Biotechnol*. 2008;26(2):164–167.
23. Keshava Prasad TS, Goel R, Kandasamy K, et al. Human Protein Reference Database—2009 update. *Nucleic acids Res*. 2009;37(Database issue):D767–D772.
24. Kim MS, Pinto SM, Getnet D, et al. A draft map of the human proteome. *Nature*. 2014;509(7502):575–581.
25. Wilhelm M, Schlegl J, Hahne H, et al. Mass–spectrometry–based draft of the human proteome. *Nature*. 2014;509(7502):582–587.
26. Lawrence RT, Villen J. Drafts of the human proteome. *Nat Biotechnol*. 2014;32(8):752–753.
27. Deribe YL, Pawson T, Dikic I. Post–translational modifications in signal integration. *Nat Struct Mol Biol*. 2014;17(6):666–672.
28. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet*. 2012;13(4):227–232.
29. Zhang B, Wang J, Wang X, et al. Proteogenomic characterization of human colon and rectal cancer. *Nature*. 2014;513(7518):382–387.
30. Delgado AP, Brandao P, Chapado MJ, et al. Open Reading Frames Associated with Cancer in the Dark Matter of the Human Genome. *Cancer Genomics & Proteomics*. 2014;11(4):201–213.
31. Delgado AP, Brandao P, Narayanan R. Diabetes Associated Genes from the Dark Matter of the Human Proteome. *MOJ Proteomics Bioinform*. 2014;1(4):00020.
32. Greenbaum D, Colangelo C, Williams K, et al. Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol*. 2003;4(9):117.
33. Gedeon T, Bokes P. Delayed protein synthesis reduces the correlation between mRNA and protein fluctuations. *Biophys J*. 2012;103(3):377–385.
34. Gry M, Rimini R, Stromberg S, et al. Correlations between RNA and protein expression profiles in 23 human cell lines. *BMC Genomics*. 2009;10:365.
35. Brazma A, Hingamp P, Quackenbush J, et al. Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nat Genet*. 2001;29(4):365–371.