

Whole genome expression as a quantitative trait

John Hardy¹, Danyah Trabzuni and Mina Ryten

Reta Lila Weston Laboratories, UCL Institute of Neurology, Queen Square, London WC1H 3BG, U.K.

Abstract

Surprisingly, whole genome analyses of complex human neurological and psychiatric disorders have revealed that many genetic risk factors are likely to influence gene expression rather than alter protein sequences. Previous analyses of neurological diseases have shown that genetic variability in gene expression levels of deposited proteins influence disease risk. With this background, we have embarked on a comprehensive project to determine the effects of common genetic variability on whole genome gene expression.

Introduction

Over the last 20 years, it has become increasingly easy to identify the genes underlying Mendelian disease. Almost without exception, the identified pathogenic mutations result in changes in protein sequence or gene copies (duplications or deletions). This led to the expectation that the risk loci identified by GWASs (genome-wide association studies) would also be largely restricted to coding regions of the genome. However, while GWASs have allowed the systematic identification of loci capable of altering risk for common complex disease by factors from 1.1- to 5.0-fold, unexpectedly most of these loci have not mapped to coding changes: indeed many have not even mapped to protein open reading frames [1]. This has led to the realization that much biologically and clinically important genetic variability is likely to result in quantitative differences in gene expression and splicing [2] as opposed to protein coding.

Use of mRNA as a quantitative trait

Since much of the genetic variability that underpins disease susceptibility is likely to act in the RNA world, we have an imperative to develop a better understanding of the genetic regulation of RNA and how it might translate into clinically important differences in protein production [2]. A simple consideration of the structure of RNA reveals how it is almost built for regulation, with the possibility of unstable hairpins which can be reversibly opened or closed and flexible antisense control. This remarkable area of biology has barely been addressed, and perhaps its importance has been largely overlooked because Mendelian diseases are nearly all explained by protein sequence variability or large changes in protein concentrations caused by structural changes to the genome.

There are many possible mechanisms by which genetic variability could quantitatively and qualitatively influence

gene expression. Genetic variability in promoter and enhancer elements could influence mRNA production, while genetic variability within transcripts could change mRNA stability or the efficiency of translation. Common genetic variants within introns could influence splicing to produce qualitative differences in gene expression. Furthermore, the importance of genetic variability in the regulation of a given gene's expression may change. In some circumstances, cellular responses may be 'hard wired' within the genome, while at other times gene expression is primarily influenced by environmental factors. Almost certainly the importance of genetic variability in gene expression is likely to be cell- and tissue-type-selective.

Mapping genotypic gene expression

Until recently the study of genetic variability in gene expression has been restricted to specific candidate genes. We and others have shown that genetic variability in apolipoprotein E and APP (amyloid precursor protein) expression influence the risk of developing Alzheimer's disease [3]. Similarly, genetic variability in the expression of α -synuclein, MAPT (microtubule-associated protein tau) and the prion gene have all been shown to increase the risk of protein deposition (α -synuclein, tau and prion protein) and the relevant sporadic neurodegenerative diseases [4]. In all these cases, missense mutations led to the pathogenic protein being deposited and autosomal dominant disease, whereas those individuals at the high end of normal expression (probably ~20–40% above the mean expression) had an increase in the risk of sporadic disease.

More recently, we and others have started to assess the effect of genetic variability on gene expression in a systematic genome wide hypothesis-free manner. The first such analysis was performed by Cheung et al. [6] who studied lymphoblastoid cells derived from the genotyped individuals from the CEPH (Centre Etude Polymorphisme Humaine). This study demonstrated that many genes showed evidence of both *cis* genetic variability (variability within the locus affecting expression) and *trans* genetic variability (variability at other places in the genome) in expression [5, 6].

Key words: brain, gene expression, genome-wide association study (GWAS), Mendelian disease, protein sequence.

Abbreviation used: GWAS, genome-wide association study.

¹To whom correspondence should be addressed (email j.hardy@ion.ucl.ac.uk).

Interestingly, *trans* variability was more pronounced when lymphoblasts were stimulated [6].

Of direct relevance to disease risk, Moffat et al. [7] also used lymphoblast expression profiling to dissect a disease association with immune diseases. Similarly, Schadt et al. [8] looked at liver expression to assess variability in cholesterol metabolism and Emilsson et al. [9] investigated genetic variability in adipose tissue to assess variants which were implicated in obesity. In all these studies, as well as in our own study of human cerebral cortex expression [10], disease associations found in GWASs were found to be at least partly mediated through variability in the levels of expression of candidate genes at a disease-associated locus.

Tissue specificity in genotypic gene expression

Clearly these studies and the databases generated from them could potentially allow investigators to look at any disease-associated polymorphism and then determine whether it is associated with variation in gene expression or splicing. What is not clear is the extent to which this variability will be tissue-specific and therefore whether we will require expression analysis in multiple tissues. Emilsson et al. [9] investigated this issue by studying genotypic gene expression in adipose tissue and lymphoblasts obtained from the same individuals. They showed that genetic variability in gene expression within adipose tissue was a more useful predictor of obesity than lymphoblast expression. This finding suggests that tissue-specific (and maybe eventually, cell type-specific) databases of the genetic analysis of gene expression will be needed.

Our primary interest is in neurological and psychiatric disease. Of all tissues, the brain has the most complex and diverse cellular architecture. Furthermore, many neurological disorders are clinically and pathologically characterized by selective vulnerability, with only specific brain regions affected by the disease process. Therefore, it will be necessary to develop databases of genotypic gene expression for multiple regions of the brain. How many regions will be necessary to capture the variability in expression, which underpins disease is still unclear. Roth et al. [11] carried out unsupervised cluster analysis of gene expression in 20 brain regions and showed that the resulting expression profiles clustered largely in accord with known developmental relationships. This suggests that much of the variability in expression can be assayed using a less than exhaustive complement of brain regions.

Future directions

Our strategy, informed by the results of the Roth et al. [11] analysis, involves the collection and sampling of post-mortem brains, largely collected through the MRC Sudden Death Brain and Tissue Bank (Edinburgh, U.K.). This brain bank

was established in 2005 to access tissues from cases of sudden death as this is recognized to be the circumstance in which brain tissue is most likely to be normal. Our immediate goal is to collect ~10 brain regions from ~200 individuals. After extracting and assessing RNA quality from each sample we will analyse gene expression using Affymetrix GeneChip® Human Exon 1.0 ST arrays. By combining these results with genotypic data generated from each brain donor, we will be able to determine both *cis* and *trans* effects on gene and exon expression. In the future, we hope to use whole transcriptome sequencing to more completely characterize these samples and better understand the role of genetic variability in gene expression in the human brain.

Acknowledgements

We thank Dr Colin Smith and Dr Robert Walker at the MRC Sudden Death Brain and Tissue Bank in Edinburgh for their assistance and support.

References

- Hindorf, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S. and Manolio, T.A. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 9362–9367
- Mattick, J.S. (2009) The genetic signatures of noncoding RNAs. *PLoS Genet.* **5**, e1000459
- Lambert, J.C., Pérez-Tur, J., Dupire, M.J., Galasko, D., Mann, D., Amouyel, P., Hardy, J., Delacourte, A. and Chartier-Harlin, M.C. (1997) Distortion of allelic expression of apolipoprotein E in Alzheimer's disease. *Hum. Mol. Genet.* **6**, 2151–2154
- Singleton, A., Myers, A. and Hardy, J. (2004) The law of mass action applied to neurodegenerative disease: a hypothesis concerning the etiology and pathogenesis of complex diseases. *Hum. Mol. Genet.* **13**, R123–R126
- Morley, M., Molony, C.M., Weber, T.M., Devlin, J.L., Ewens, K.G., Spielman, R.S. and Cheung, V.G. (2004) Genetic analysis of genome-wide variation in human gene expression. *Nature* **430**, 743–747
- Cheung, V.G., Spielman, R.S., Ewens, K.G., Weber, T.M., Morley, M. and Burdick, J.T. (2005) Mapping determinants of human gene expression by regional and genome-wide association. *Nature* **437**, 1365–1369
- Moffatt, M.F., Kabisch, M., Liang, L., Dixon, A.L., Strachan, D., Heath, S., Depner, M., von Berg, A., Bufe, A., Rietschel, E. et al. (2007) Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* **448**, 470–473
- Schadt, E.E., Molony, C., Chudin, E., Hao, K., Yang, X., Lum, P.Y., Kasarskis, A., Zhang, B., Wang, S., Suver, C. et al. (2008) Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* **6**, e107
- Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A.S., Zink, F., Zhu, J., Carlson, S., Helgason, A., Walters, G.B., Gunnarsdottir, S. et al. (2008) Genetics of gene expression and its effect on disease. *Nature* **452**, 423–428
- Myers, A.J., Gibbs, R., Webster, J., Rohrer, K., Zhao, A., Marlowe, L., Kaleem, M., Leung, D., Bryden, L., Nath, P. et al. (2007) A survey of genetic cortical gene expression. *Nat. Genet.* **39**, 1494–1499
- Roth, R.B., Hevezi, P., Lee, J., Willhite, D., Lechner, S.M., Foster, A.C. and Zlotnik, A. (2006) Gene expression analyses reveal molecular relationships among 20 regions of the human CNS. *Neurogenetics* **7**, 67–80

Received 19 August 2009
doi:10.1042/BST0371276