Minireview

Variations in abundance: genome-wide responses to genetic variation… and vice versa
Bruce A Hamilton

Address: Department of Medicine (Division of Genetics), Department of Cellular and Molecular Medicine, and Rebecca and John Moores UCSD Cancer Center, University of California San Diego School of Medicine, La Jolla, CA 92093-0644, USA. E-mail: bah@ucsd.edu

Abstract

How do naturally occurring polymorphisms in DNA sequence relate to variation in gene expression? Recent work to map genetic sources of expression variation has shown a surprising balance between *cis* and *trans* effects. Other work suggests some chromosomal clustering of genes by expression pattern. A synthesis of approaches may provide new insight in to adaptive mechanisms in evolution and the population basis of complex traits.

Interpreting the functional significance of genetic polymorphisms in natural populations poses a major challenge. Here I review recent work in yeast, flies, mice and primates that examines the influences of naturally occurring sequence variation, chromosomal order and speciation on genome-wide expression profiles of both RNA and protein. A synthetic view from these experiments would suggest that gene expression is not randomly distributed along chromosomes, that variations in mRNA and protein expression within a single species result from a surprising balance between polymorphisms acting *in cis* and polymorphisms acting *in trans* to the regulated gene, and consequently that relatively few adaptive changes could have major impacts in remodeling gene expression patterns over the course of evolution.

Sequence variation across genomes and across populations
Sequence polymorphism and genome-wide haplotype maps have begun to catalog the extent and structure of genetic variation present in human populations [1,2] and to a lesser extent inbred strains of model organisms [3-6]. An emerging theme from such maps is that linked polymorphisms tend to travel through a population together, creating haplotype blocks [2,7], by virtue of either recombination hotspots or population expansions of chromosomes carrying ancestral recombination events. Recent reports have also documented variation in gene expression profiles between individuals and, more importantly, reproducible variations between inbred strains [8]. However, assigning functional significance to individual polymorphisms at the level of either sequence or expression is a different problem. A series of recent papers has taken a complementary approach by examining the sources of variation in expression profiles between strains of model organisms.

Polymorphism and clustering in expression profiles
The availability of nearly complete genome sequences and large sets of gene expression data has led several groups to consider whether gene expression profiles are structured by chromosomal order [9-13]. Individual gene clusters have been observed and analyzed since almost the beginnings of molecular genetics, and genome sequences have demonstrated that genomes are not randomly organized. But how prevalent is chromosomal organization by gene expression? New comparisons of transcript maps and expression profiles with genome sequence have indicated nonrandom clustering of genes into chromosomal expression domains in flies [10], nematodes [11], and humans [12-14]. Lercher *et al.* [13] have shown that housekeeping genes cluster more readily than do other gene classes (as measured by breadth of representation in serial analysis of gene expression (SAGE) experiments), but clustering patterns that are more specific have also been

reported. Strikingly, Qiu *et al.* [14] report in *BMC Genomics* that SAGE tags found in normal brain or brain tumor samples cluster along chromosomes more than would be expected by chance, and that the normal and tumor clusters are distinct from each other. Although more work to understand the underlying reasons for these statistical clusters is warranted, such clustering would make a ready substrate for evolutionary remodeling of gene expression patterns.

Mapping sources of variation in gene-product expression: yeast strains and mouse brains

One intriguing question is to what extent variation in gene expression within a population is controlled by sequence variation within the gene itself (*in cis*) compared with variations in unlinked regulatory genes (*in trans*). Multiple layers of control of gene expression have been selected in evolution (transcription, processing, export and stability of RNA; and translation, folding, modification, trafficking and degradation of protein), and variation in genes that act in any layer could produce systematic variation in gene expression among individuals in a population. This raises an interesting question about the genetic architecture of gene expression differences among individuals in natural populations, between inbred strains of laboratory organisms, and between evolving species.

Starting with a relatively simple case, Kruglyak and colleagues [15] have mapped sources of mRNA expression differences between a laboratory strain and a wild isolate of *Saccharomyces cerevisiae*. Among 6,215 RNAs monitored, 1,528 showed highly significant expression differences between the two parental strains ($p < 0.005$ and $< 2\%$ expected false-positive rate by permutation testing). By examining both expression profiles and genetic markers in the same cross progeny, Brem *et al.* [15] were able to estimate the heritability of expression variations and map the loci that control the variation. The heritability estimates indicated that 84% of the expression difference was genetic; surprisingly, however, only 308 of the 1,528 expression differences showed significant genetic linkage among 40 haploid segregants. A power calculation puts this in perspective: a cross of this size and marker density should detect nearly all loci that control all of an expression difference and nearly 30% of loci that control as much as a third of the variation for an expression difference. As Brem *et al.* [15] found linkage for only about 20% of the expression differences - and only a third of those map back to the structural gene - this suggests that interstrain expression differences (and therefore inter-individual differences in a broader population) in even a 'simple' eukaryote can be genetically complex and that the majority can not be accounted for simply by changes in *cis*-acting sequences.

Similarly, Klose and colleagues [16] have taken a genetic approach for an initial look at a more complex question:

protein expression in the mouse brain. Using inbred strains of *Mus musculus* and *Mus spretus*, Klose *et al.* [16] identified 8,767 distinct two-dimensional gel spots, representing isoforms of an estimated 2,770 proteins (based on mass spectrometry of identified spots) from soluble extracts. Of these, 1,324 spots (an estimated 936 proteins) were polymorphic between species. Protein polymorphisms can be either quantitative or qualitative (or both). Clearly, a large number of mechanisms can operate on both abundance and electrophoretic mobility of proteins compared to RNA, but which of them is most prevalent between recently diverged species? About half of the polymorphic spots (40% of the proteins) varied primarily in a qualitative manner, having altered migration patterns, with the remainder being primarily quantitative. Are these differences intrinsic to the allelic proteins or reflective of altered regulation?

Not surprisingly, qualitative differences proved easier to follow than purely quantitative differences in backcross progeny (one limitation of this experimental design is that only one sex of the F1 hybrid is fertile and only one of the possible backcrosses was examined, and thus about 343 protein spots for which the backcross parent has a dominant allele could not be scored). Klose *et al.* [16] followed linkage of 409 polymorphic spots (273 qualitative, 176 quantitative) in 200 backcross (F1 x *M. spretus*) progeny from the previously genotyped panel. Consistent with the yeast RNA experiments, several of the protein spots appeared to be affected by more than one gene, and some by as many as three. Among 150 polymorphic proteins identified by mass spectrometry, 42 mapped to the known location of the structural gene, suggesting either amino-acid substitutions or *cis*-regulatory changes, and 41 polymorphisms mapped to sites other than the structural gene. As with yeast RNA expression [15], this again points to a high degree of unlinked regulatory polymorphism in the control of proteome expression. Nor does this analysis exhaust the approach; Klose *et al.* [16] focused on the soluble fraction of their extracts, leaving membrane and chromatin fractions for later consideration. It will be of interest to see whether proteins with additional constraints on their trafficking and localization have any significant difference in the kind or number of linkages for protein polymorphisms.

One wonders about other patterns of change that may be observed at the edge of speciation. Several reproductively isolated but interfertile subspecies of mice have been inbred, including *Mus musculus castaneus* and *Mus musculus molossinus*. Polymorphism rates between these strains and canonical lab mice are not quite as high as for *M. spretus* but still average about one amino-acid change per protein, and both sexes of the hybrid progeny are fertile, allowing an intercross design. Comparisons among multiple sibling species (or subspecies) may provide an additional layer of functional annotation by highlighting proteins that are either highly constrained or highly plastic

in their two-dimensional gel profiles. How the balance of *cis* versus *trans* effects might change over different evolutionary distances would also be of interest.

An evolving framework

The nature and distribution of polymorphisms in humans and experimental animals is important for disease and modifier gene hunts, but also for understanding mechanisms of selection and adaptation in evolution. The distributions of amino-acid substitutions and gene expression changes are particularly interesting in view of the changes in global patterns of gene expression recently observed in primate speciation [17]. To the human eye, the two mouse species examined above [16] seem relatively similar. By contrast, humans and chimpanzees, which differ by roughly the same level of nucleotide changes, seem quite different. Is this just a consequence of our own anthropocentrism? The protein expression analysis of Enard *et al.* [17] would suggest otherwise. While brain protein differences between mouse species are about evenly split between quantitative and qualitative changes, a similar analysis of humans and chimps, even after accounting for inter-individual differences, shows a several-fold increase in the rate of quantitative differences [17]. It will be of great interest to see how these changes distribute between *cis*-acting regulatory changes and changes in *trans*-acting regulatory factors; and whether this index differs with morphological divergence.

Life is messy - and that is to its credit. Robust performance in a noisy environment is a hallmark of well-engineered systems, from telecommunications networks to the human brain. Biological populations require variation among individuals for adaptation to changing environments and to take advantage of new niche opportunities. A multipronged approach that includes global analysis of sequence variants, expression variants and genetic mapping promises to provide a new understanding of how the population structure of the genome-wide response to genetic variants influence biological traits.

References

1. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, *et al.*: **A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms.** *Nature* 2001, **409:**928-933.
2. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, *et al.*: **The structure of haplotype blocks in the human genome.** *Science* 2002, **296:**2225-2229.
3. Lindblad-Toh K, Winchester E, Daly MJ, Wang DG, Hirschhorn JN, Laviolette JP, Ardlie K, Reich DE, Robinson E, Sklar P, *et al.*: **Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse.** *Nat Genet* 2000, **24:**381-386.
4. Berger J, Suzuki T, Senti KA, Stubbs J, Schaffner G, Dickson BJ: **Genetic mapping with SNP markers in *Drosophila*.** *Nat Genet* 2001, **29:**475-481.
5. Wicks SR, Yeh RT, Gish WR, Waterston RH, Plasterk RH: **Rapid gene mapping in *Caenorhabditis elegans* using a high density polymorphism map.** *Nat Genet* 2001, **28:**160-164.
6. Jordan B, Charest A, Dowd JF, Blumenstiel JP, Yeh RF, Osman A, Housman DE, Landers JE: **Genome complexity reduction for SNP genotyping analysis.** *Proc Natl Acad Sci USA* 2002 **99:**2942-2947.
7. Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, Kautzer CR, Lee DH, Marjoribanks C, McDonough DP, *et al.*: **Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21.** *Science* 2001, **294:**1719-1723.
8. Sandberg R, Yasuda R, Pankratz DG, Carter TA, Del Rio JA, Wodicka L, Mayford M, Lockhart DJ, Barlow C: **Regional and strain-specific gene expression mapping in the adult mouse brain.** *Proc Natl Acad Sci USA* 2000, **97:**11038-11043.
9. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, *et al.*: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409:**860-921.
10. Spellman PT, Rubin GM: **Evidence for large domains of similarly expressed genes in the *Drosophila* genome.** *J Biol* 2002, **1:**5.
11. Roy PJ, Stuart JM, Lund J, Kim SK: **Chromosomal clustering of muscle-expressed genes in *Caenorhabditis elegans*.** *Nature* 2002, **418:**975-979.
12. Caron H, van Schaik B, van der Mee M, Baas F, Riggins G, van Sluis P, Hermus MC, van Asperen R, Boon K, Voute PA, *et al.*: **The human transcriptome map: clustering of highly expressed genes in chromosomal domains.** *Science* 2001, **291:**1289-1292.
13. Lercher MJ, Urrutia AO, Hurst LD: **Clustering of housekeeping genes provides a unified model of gene order in the human genome.** *Nat Genet* 2002, **31:**180-183.
14. Qiu P, Benbow L, Liu S, Greene JR, Wang L: **Analysis of a human brain transcriptome map.** *BMC Genomics* 2002, **3:**10.
15. Brem RB, Yvert G, Clinton R, Kruglyak L: **Genetic dissection of transcriptional regulation in budding yeast.** *Science* 2002, **296:**752-755.
16. Klose J, Nock C, Herrmann M, Stuhler K, Marcus K, Bluggel M, Krause E, Schalkwyk LC, Rastan S, Brown SD *et al.*: **Genetic analysis of the mouse brain proteome.** *Nat Genet* 2002, **30:**385-393.
17. Enard W, Khaitovich P, Klose J, Zollner S, Heissig F, Giavalisco P, Nieselt-Struwe K, Muchmore E, Varki A, Ravid R, *et al.*: **Intra- and interspecific variation in primate gene expression patterns.** *Science* 2002, **296:**340-343.