Update TRENDS in Genetics Vol.20 No.3 March 2004

Research Focus

Allele-specific gene expression uncovered

Julian C. Knight

Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK

Genetic variation in populations can result in variation in levels of gene expression but the extent to which this occurs has been unclear. In this article, recent studies of allele-specific expression among autosomal non-imprinted genes are reviewed. These new data provide evidence that differential expression is relatively common and that allelic differences are heritable and can be highly context specific.

ALLELE-specific (see Glossary) differences in levels of gene expression have classically been associated with the EPIGENETIC phenomena of X-chromosome inactivation and GENOMIC IMPRINTING. Several recent studies have emphasized the extent to which gene expression varies within and between populations [1–4], and have shown that allele-specific expression is relatively common among non-imprinted autosomal genes [5–7]. Futhermore, it appears that these allele-specific differences in expression are heritable [6]. At present it is difficult to predict what proportion of genes will show such differences in expression and what the magnitude of the effect will be. It is likely that differential expression will be highly context specific with regard to, for example, cell type and stimulus, and that for some genes small differences in levels of expression (e.g. ~20% difference in gene expression) between alleles might be physiologically important. Allele-specific differences in gene expression provide evidence for a model whereby *cis*-acting genetic variation results in differential expression between alleles. Analysis of allele-specific expression has the potential to enhance considerably our ability to identify such regulatory genetic variation in a population where the haplotypic structure of the locus has been clearly defined.

Allele-specific expression and epigenetics

Allele-specific expression arising during the process of development is well described, with the epigenetic phenomena of X-chromosome inactivation and genomic imprinting being noteworthy examples. In both cases, epigenetic marking by covalent modification of DNA and core histones creates molecular landmarks that differentiate between active and inactive chromatin [8]. The process whereby one of the two X chromosomes becomes inactivated early in embryogenesis involves specific developmental signals coordinated by a chromosomal locus, the X-inactivation center, including *Xist* [9]. The silenced X chromosome is characterized by a coating of untranslated *Xist* RNA and the acquisition of multiple epigenetic modifications. These result in stable repressed HETEROCHROMATIN, which is faithfully propagated in subsequent cell divisions.

The differential expression of the alleles of imprinted genes that occurs during development results from differential epigenetic modification of the genome in male and female gametes (e.g. DNA methylation, histone acetylation and methylation). Deletion of differentially methylated regions has been shown to result in a loss of imprinting [10,11], whereas knockout mice for *Dnmt-1* (which encodes DNA methyltransferase 1) show silencing of some alleles that are normally expressed paternally or maternally, and show bialleic expression of other imprinted genes [12]. Typically, one allele at an imprinted locus is transcriptionally silent but epigenetic heterogeneity is recognized between individuals [13].

Allele-specific expression in non-imprinted autosomal genes

Interest in the existence of allele-specific expression in non-imprinted autosomal genes has increased with awareness of the important role that variation in non-coding DNA sequences can play in determining phenotypic diversity. If such regulatory variation does indeed modulate the levels of gene expression, one would expect to find evidence of allele-specific differences in gene expression. Several recent studies suggest that allelic differences in gene expression occur among autosomal non-imprinted genes [5–7,14]. The crucial question is whether such differences are heritable, and the work of Yan and colleagues [6] has shown that these differences are inherited. The authors investigated 13 genes in a panel of Epstein-Barr virus (EBV)-immortalized lymphoblastoid cell lines (LCLs) from 96 individuals in the Centre d'Etude du Polymorphisme Humain (CEPH) family collection.

Glossary

Allele: one of the variant forms of a gene at a particular location on a chromosome.
Epigenetic: any heritable influence (in the progeny of cells or of individuals) on chromosome or gene function that is not accompanied by a change in DNA sequence.
Genomic imprinting: a phenomenon by which the two alleles of certain genes are differentially expressed according to their parental origin.
Heterochromatin: a region of the genome that remains highly condensed throughout the cell cycle and shows little or no evidence of active gene expression.
Heterozygous: an individual that possesses two different forms of a particular gene, one inherited from each parent.
Haplotype: a set of genetic markers that is present on one chromosome.
Linkage: the tendency of genes or other DNA sequences at specific loci to be inherited together as a consequence of their physical proximity on a single chromosome.
Linkage disequilibrium: particular alleles at nearby sites co-occur on the same haplotype more often than is expected by chance.

Corresponding author: Julian C. Knight (julian@well.ox.ac.uk).

They identified between 17 and 37 individuals who were informative heterozygotes for a given transcribed marker single nucleotide polymorphism (SNP) among these genes, and quantified the relative transcript abundance that originated from each allele by primer-extension analysis of RT–PCR products. They observed that allele-specific differences in expression were surprisingly common, occurring in six out of the 13 genes that were assayed, with a 1.3–4.3-fold difference between alleles. However, for a given marker SNP, the number of HETEROZYGOUS individuals who showed an allele-specific effect was low, varying from 3% to 30%. The authors demonstrated, using a pedigree analysis of families of individuals who exhibited allele-specific differences, that differential expression was indeed consistent with Mendelian inheritance. They identified three informative families and found that altered expression of the genes *CAPN10* (which encodes calpain 10) and *PKD2* (which encodes polycystic kidney disease 2) was consistently inherited with a single HAPLOTYPE defined by at least two adjacent microsatellite markers.

Interestingly, when Bray and colleagues examined allele-specific differences in native human tissue rather than in transformed cultured cell lines, a similar proportion of genes appeared to show allele-specific effects [14]. They investigated human post-mortem brain tissue from 60 adults, and observed allele-specific differences of $>20\%$ in transcript abundance in seven out of 15 genes screened; the magnitude of the difference between alleles was 20–66%. Consistent with the study by Yan *et al.* [6], only a minority of individuals who were heterozygous for a given marker SNP showed this allele-specific difference; in most cases this was restricted to a single individual. This might reflect the underlying complexities of the haplotypic structure in which the haplotype that contains the functionally important regulatory SNP(s) is only partially resolved by the marker SNP used to distinguish the relative transcript abundance. In the studies by Bray [14] and Yan [6] there was no evidence of monoallelic expression, which would indicate genomic imprinting.

Recently, Cowles and colleagues screened murine genes for allele-specific effects by the elegant approach of studying F1 intercrosses from four inbred mouse strains [5]. Their approach resolved that the differences in transcript levels were due to *cis*-acting sequence variation because in the F1 mice the alleles were exposed to common *trans*-acting factors and environmental influences. Similar to the work described in human cells, this does not define which of the specific *cis*-acting regulatory variants are important but rather that such regulatory variation is present. The authors used a transcribed marker SNP to distinguish between transcripts derived from each of the two parental alleles. Relative allelic abundance was measured by a robust quantitative primer-extension method using fluorescently labeled dideoxynucleotides, which is able to reliably detect a 1.5-fold or greater difference between alleles. The initial screen focused on genes that showed detectable expression in liver, spleen or brain and had a suitable transcribed SNP present in at least two mouse strains. This revealed that seven of 69 genes investigated showed allele-specific differences in expression of 1.5-fold or greater in duplicate animals. Analysis of the sex of the parental allele demonstrated that the differential expression between alleles was not due to parental imprinting.

Cowles and colleagues confirmed these results by identifying and testing additional marker SNPs within the transcripts of the seven genes in four mouse strains. For the four genes that possessed the greatest difference between alleles, *Il9r* (which encodes interleukin 9 receptor), *Ccnf* (which encodes cyclin F), *Uros* (which encodes uroporphyrinogen III synthase) and *Hmgcr* (which encodes HMG CoA reductase), the results were impressively consistent for the different SNPs analyzed on a given transcript. Moreover, for *Ccnf* and *Hmgcr* the allelic difference was tissue-specific, emphasizing that allele-specific effects on gene expression are likely to be highly context specific, and depend on cell or tissue type, developmental stage and environmental condition. Smaller allele-specific differences below the threshold for reliable detection of 1.5 fold might be physiologically important. The percentage of genes that showed allele-specific differences in their levels of expression ($\sim$6% of those analyzed) might therefore be an underestimate, particularly because the available marker SNPs enabled less than half of all possible pairwise combinations to be tested among the four mouse strains. It would be interesting to extend this analysis to F2 populations to investigate *trans*-acting effects. Such *trans*-acting regulatory variation, in which a polymorphism in one gene can affect the expression of other genes, has been shown to be important by recent research that regarded expression differences as quantitative traits and mapped differences using LINKAGE analysis between strains of yeast [3,15] and mice [16].

One challenge ahead is how to scale up analysis of allele-specific gene expression to interrogate differential expression at the genomic level. Lo and coworkers [7] have recently published research using a microarray platform to screen for allele-specific differences in transcript abundance in human fetal liver and kidney tissue, which might enable high-throughput analysis. The approach is an attractive one: the authors have used a genotyping technology, the Affymetrix HuSNP chip system, which enables alleles to be distinguished on the basis of hybridization specificity to probes matching the two allelic forms of the marker SNP. Such an approach enables high-throughput screening but is a method that relies on differential hybridization sufficiently quantitative to enable reproducible and accurate measurement of relative transcript abundance? The data suggest that it might be, at least for detecting large allelic differences. The authors analyzed the variation for the allelic ratios measured in genomic DNA (in which a 1:1 ratio in allelic abundance would be expected) for 39 SNPs and found that the average 95% confidence interval was 0.5–2.0, from which they determined that the approach had a threshold that was sufficient to confidently detect genes with a greater than twofold difference in expression between the alleles.

The authors analyzed transcribed SNPs selected on the basis that at least one of the seven fetuses studied was heterozygous for a given SNP, with the gene expressed in

fetal liver or kidney tissue. The results appear striking: 54% of the 602 genes analyzed showed at least a twofold difference in transcript abundance between alleles in at least one individual, whereas 28% of genes showed a greater than fourfold difference. The magnitude of difference between alleles for a given gene varied significantly between the different individuals studied and the genes showing allelic differences in expression appear scattered widely across the genome. Several known imprinted genes were present on the array and four out of five of these had the expected monoallelic expression, whereas the fifth imprinted gene, *WT1* (which encodes Wilms tumor 1), showed biallelic expression but is known to have tissue-restricted imprinting. When a limited set of seven genes from the array was analyzed by allele-specific RT–PCR the results appeared consistent with the array data; however, further work remains to be done to establish the robustness of this approach. It would be interesting to determine in a larger panel of individuals whether using multiple SNPs to quantify the relative allelic abundance for a given transcript (as was performed by Cowles *et al.* [5]) would lead to the same estimate of 20–50% of all genes showing allele-specific differences. It is intriguing that the magnitude of this estimate should be consistent with recent work that screened promoter SNPs for regulatory effects using reporter gene analysis [17]. When promoter haplotypes for 170 human genes with experimentally derived promoters were analyzed, ~1/3 of SNPs were found to modulate reporter gene expression.

Quantifying allele-specific expression

Several issues arise with regard to the quantification of allele-specific gene expression. The use of a transcribed marker polymorphism to distinguish between alleles in cells that are heterozygous for that marker (Figure 1) presents an attractive internally controlled system that avoids many potential genetic, environmental and technical confounders of any comparison between individuals or cells with different genotypes [6,14,18,19]. The major limitation is that such analysis is restricted to genes, and more specifically haplotypes of those genes, in which a transcribed marker is present. The use of phosphorylated RNA polymerase II (Pol II) as a marker of transcriptional activity abrogates the requirement for transcribed polymorphisms when analyzed by the haplotype-specific chromatin immunoprecipitation (haploChIP) method [20]. Here, the phosphorylated Pol II is crosslinked to chromatin in living cells and, following sonication, immunopreciptiation and reversal of crosslinks, the relative abundance of the resulting DNA fragments provides a surrogate measure of transcriptional activity. Given that any polymorphism in non-coding or coding DNA within 1 kb of the transcriptional start site or the 3′ untranslated region (UTR) can be used as a marker for allele-specific analysis of these immunoprecipitated DNA fragments, this considerably broadens the number of genes, and haplotypes of genes that can be analyzed for allele-specific differences in expression.

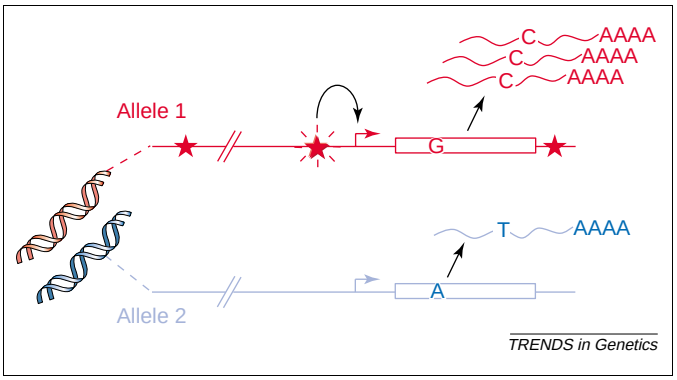

Figure 1. Allele-specific differences in gene expression might usefully be determined using variation within the coding regions to distinguish the relative abundance of transcript arising from the two alleles. When cells from an individual who is heterozygous for a coding polymorphism (e.g. an individual with genotype GA) are analyzed, the allelic origin of transcript can be defined. In this example, allele 1 is a high-producer allele (transcript shown in red) relative to allele 2 (transcript shown in blue). It is important to note however that, although the coding single nucleotide polymorphism (SNP) might be functional, it is being used here as a marker to distinguish the relative abundance of the two alleles. In this example, several other SNPs are present on the high-producer allele (denoted by red stars), one of which is a functionally important regulatory polymorphism that modulates gene expression (denoted by red star with surrounding flash lines).

The question arises, whether analyzing mRNA or immunoprecipitated DNA fragments, as to how best to accurately quantify their relative allelic abundance. Various methods have been employed to measure relative allelic abundance, notably those based on the primer extension of PCR-amplified products. The use of fluorescently labeled dideoxynucleotides [21] enables quantification on gel electrophoresis and this appears robust, being able to detect differences of as little as 20% between alleles [6]. More attractive in terms of its capacity for high-throughput detection of primer-extension products would be the use of time-of-flight mass spectrometry; this technique also appears to be highly sensitive and quantitative provided that sufficient replicates are used [20]. With all methods using a marker SNP to distinguish relative allelic abundance, the use of multiple markers and the use of primers designed for forward and reverse complementation in large panels of informative heterozygous individuals will be required.

Implications of resolving allele-specific expression for defining genetic traits

Functionally important genetic diversity is likely to underlie much of the observed human phenotypic diversity. Classically, this genetic variation has been considered in terms of coding-region polymorphisms with the capacity to alter protein structure and function but it is recognized that variation in the non-coding DNA is probably at least as important [22]. Resolving the host genetic variation that influences susceptibility to multifactorial common diseases is an active research area. The extraordinary genetic diversity and LINKAGE DISEQUILIBRIUM within our genomes, which enables such interrogation to take place, is also a stumbling block because of the difficulty in resolving the true functionally important variants. Defining allele-specific gene expression will help us to appreciate the extent of functionally important regulatory variation, and to focus on candidate haplotypes that have allelic differences in expression for detailed molecular characterization of specific polymorphisms.

References

1 Cheung, V.G. *et al.* (2003) Natural variation in human gene expression assessed in lymphoblastoid cells. *Nat. Genet.* 33, 422–425
2 Oleksiak, M.F. *et al.* (2002) Variation in gene expression within and among natural populations. *Nat. Genet.* 32, 261–266
3 Brem, R.B. *et al.* (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science* 296, 752–755
4 Enard, W. *et al.* (2002) Intra- and interspecific variation in primate gene expression patterns. *Science* 296, 340–343
5 Cowles, C.R. *et al.* (2002) Detection of regulatory variation in mouse genes. *Nat. Genet.* 32, 432–437
6 Yan, H. *et al.* (2002) Allelic variation in human gene expression. *Science* 297, 1143
7 Lo, H.S. *et al.* (2003) Allelic variation in gene expression is common in the human genome. *Genome Res.* 13, 1855–1862
8 Li, E. (2002) Chromatin modification and epigenetic reprogramming in mammalian development. *Nat. Rev. Genet.* 3, 662–673
9 Plath, K. *et al.* (2002) *Xist* RNA and the mechanism of X chromosome inactivation. *Annu. Rev. Genet.* 36, 233–278
10 Tremblay, K.D. *et al.* (1995) A paternal-specific methylation imprint marks the alleles of the mouse *H19* gene. *Nat. Genet.* 9, 407–413
11 Thorvaldsen, J.L. *et al.* (1998) Deletion of the *H19* differentially methylated domain results in loss of imprinted expression of *H19* and *Igf2*. *Genes Dev.* 12, 3693–3702
12 Li, E. *et al.* (1993) Role for DNA methylation in genomic imprinting. *Nature* 366, 362–365
13 Sakatani, T. *et al.* (2001) Epigenetic heterogeneity at imprinted loci in normal populations. *Biochem. Biophys. Res. Commun.* 283, 1124–1130
14 Bray, N.J. *et al.* (2003) *Cis*-acting variation in the expression of a high proportion of genes in the human brain. *Hum. Genet.* 113, 149–153
15 Yvert, G. *et al.* (2003) *Trans*-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat. Genet.* 35, 57–64
16 Schadt, E.E. *et al.* (2003) Genetics of gene expression surveyed in maize, mouse and man. *Nature* 422, 297–302
17 Hoogendoorn, B. *et al.* (2003) Functional analysis of human promoter polymorphisms. *Hum. Mol. Genet.* 12, 2249–2254
18 Singer-Sam, J. *et al.* (1992) Parental imprinting studied by allele-specific primer extension after PCR: paternal X chromosome-linked genes are transcribed prior to preferential paternal X chromosome inactivation. *Proc. Natl. Acad. Sci. U. S. A.* 89, 10469–10473
19 Singer-Sam, J. *et al.* (1992) A sensitive, quantitative assay for measurement of allele-specific transcripts differing by a single nucleotide. *PCR Methods Appl.* 1, 160–163
20 Knight, J.C. *et al.* (2003) *In vivo* characterization of regulatory polymorphisms by allele-specific quantification of RNA polymerase loading. *Nat. Genet.* 33, 469–475
21 Norton, N. *et al.* (2002) Universal, robust, highly quantitative SNP allele frequency measurements in DNA pools. *Hum. Genet.* 110, 471–478
22 King, M.C. and Wilson, A.C. (1975) Evolution at two levels in humans and chimpanzees. *Science* 188, 107–116

0168-9525/$ - see front matter
doi:10.1016/j.tig.2004.01.001

Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes

Dario Leister

Abteilung für Pflanzenzüchtung und Ertragsphysiologie, Max-Planck-Institut für Züchtungsforschung, Carl-von-Linné Weg 10, D-50829 Köln, Germany

NBS-LRR genes are the major class of disease resistance genes in flowering plants, and are arranged as single genes and as clustered loci. The evolution of these genes has been investigated in *Arabidopsis thaliana* by combining data on their genomic organisation and position in phylogenetic trees. Tandem and segmental duplications distribute and separate NBS-LRR genes in the genome. It is, however, unclear by which mechanism(s) NBS-LRR genes from different clades are sampled into heterogeneous clusters. Once physically removed from their closest relatives, the NBS-LRR genes might adopt and preserve new specificities because they are less prone to sequence homogenization.

Plant resistance (*R*) genes mediate phenotypic resistance against pests and pathogens expressing avirulence genes (a situation known as 'gene-for-gene' interaction). Genes that encode proteins containing a nucleotide-binding site (NBS) and C-terminal leucine-rich repeats (LRRs) represent the largest class of *R* genes in flowering plants [1]; NBS-LRR genes also exist in gymnosperms, non-vascular plants and mammals [2–4]. Based on their N-termini, two subclasses of NBS-LRR resistance proteins are known: the first is characterized by the TIR-domain homologous to the *Drosophila* Toll and mammalian Interleukin-1 receptors, and the second is characterized by a coiled-coil (CC) structure. Truncated versions of NBS-LRR genes exist encoding proteins that lack either a domain close to the N-terminal of the NBS, or the LRR region, or consist only of a TIR-domain. In grass species, TIR-NBS-LRR genes have not yet been identified, but the CC-type is very common [5].

In different plants, NBS-LRR loci are found both as isolated genes (singletons) and as tightly linked arrays of related genes (gene clusters) [6]. In some cases, gene clusters contain copies of NBS-LRR genes from different phylogenetic clades [HETEROGENEOUS CLUSTERS (see Glossary)] [7]. Before the complete sequences of plant genomes became available, analyses of NBS-LRR gene evolution in diverse species were based on relatively few loci. Recently, with the complete sequence of the genome of *Arabidopsis thaliana* available, several groups have carried out genome-wide analyses of the organisation and evolution of NBS-LRR genes [8–12]. The distribution of NBS-LRR genes in the genome has been, in general,

Corresponding author: Dario Leister (leister@mpiz-koeln.mpg.de).