GENETICS

Allelic Variation in Human Gene Expression

Hai Yan,[1] Weishi Yuan,[2] Victor E. Velculescu,[1] Bert Vogelstein,[1] Kenneth W. Kinzler[1]*

Understanding the genetic basis of human variation is a vital goal of biomedical research. Studies in other organisms suggest that differences in gene expression levels account for a major part of the variation within and among species (*1*, *2*). To address this in humans, we developed methods to quantitatively evaluate allelic variation in gene expression.

The analysis of variation in gene expression is complicated by the potentially small differences associated with alterations in a single allele as well as by potential variations between individuals that arise from environmental or physiological rather than genetic factors. To circumvent these analytic problems, we compared the relative expression levels of two alleles of the same gene within the same cellular sample. To make these comparisons, we used a fluorescent dideoxy terminator–based method (*3*) to distinguish the mRNA products of alleles from normal individuals who were heterozygous for a single nucleotide polymorphism (SNP) in the transcript of interest (fig. S1). We estimated that this approach could confidently identify variations when the differences between expression of the two alleles differed by more than 20% (*3*).

We examined SNPs for 13 genes in 96 individuals from the CEPH families and identified 17 to 37 individuals who were heterozygous for any given gene (Table 1). Their mRNA was then used to assess relative expression of the two alleles. Significant differences in allelic variation were observed in 6 of the 13 genes studied. The fraction of individuals exhibiting such variation ranged from 3% (*Catalase*) to 30% (*p73*) (Table 1). In individuals whose alleles were differentially expressed, the ratio of transcripts varied from 1.3:1.0 (*FBN1*) to 4.3:1.0 (*p73*). Given that these variations were each observed in only a minority of individuals, it is unlikely that they were due to parental imprinting.

We next examined the families of nine individuals exhibiting allelic variation. Three families were informative and displayed expression fully consistent with Mendelian inheritance. Two families had allelic variation of *Calpain-10* expression, and one family had allelic variation of *PKD2* expression (Fig. 1). Altered expression was consistently inherited together with a single haplotype defined by at least two adjacent microsatellite markers. The altered allelic expression of *PKD2* was due to increased expression of the affected allele, whereas in both *Calpain-10* families, it was due to decreased expression of the affected allele.

Thus, cis-acting inherited variations in gene expression are relatively common among normal individuals. Moreover, these measurements likely represent an underestimate, as additional variations in allelic expression may manifest in a cell type or state-specific manner. These results suggest approaches for connecting genotype to disease susceptibility based on changes in gene expression as opposed to changes in the structure of the encoded protein. Such analyses do not require any assumptions about the genetic structure of the population studied and may yield direct information about the causative gene and its mechanism of action.

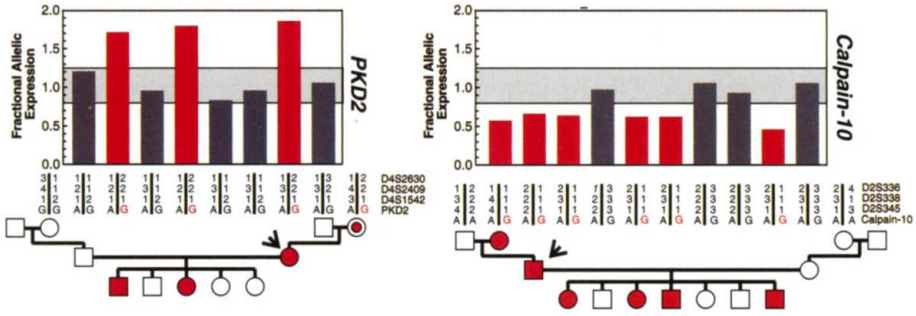

Fig. 1. Mendelian inheritance of altered *PKD2* or *Calpain-10* expression. Only individuals who were heterozygous for the SNP or were used to deduce haplotypes are shown. Red indicates altered fractional allelic expression; arrows indicate probands. An obligate carrier in the *PKD2* pedigree who could not be scored is indicated with a red dot within a circle. The results of genotype and expression analyses are shown above each member of the pedigrees. A single *Calpain-10* genotype suggesting a recombination is italicized.

Table 1. Allelic variation in gene expression.

Gene	SNP	Heterozygous individuals tested	Individuals displaying variations	Magnitude of variation (fold)
APC	486C/T	17	0	–
BRCA1	4449T/C	19	0	–
Calpain-10	2037A/G	27	3 (11%)	1.7–1.9
Catalase	1235T/C	37	1 (3%)	1.4
COMT	388C/T	21	0	–
DNT	195A/G	20	0	–
FBN1	2008T/C	19	2 (11%)	1.3, 1.6
LDLR	2325G/A	24	0	–
NOD2	1866T/G	25	1 (4%)	1.6
p53	466G/C	18	0	–
p73	629T/C	20	6 (30%)	1.5–4.3
PKD2	4208G/A	26	1 (4%)	1.7
UCP2	544C/T	26	0	–

References and Notes

1. N. A. Johnson, A. H. Porter, *J. Theor. Biol.* **205**, 527 (2000).
2. M. Levine, *Nature* **415**, 848 (2002).
3. Materials and methods are available as supporting material on *Science* Online.
4. Supported by NIH grants CA57345, CA 62924, and CA43460.

Supporting Online Materials

www.sciencemag.org/cgi/content/full/297/5584/1143/DC1

Materials and Methods

Fig. S1

[1]The Sidney Kimmel Comprehensive Cancer Center and the Howard Hughes Medical Institute, Johns Hopkins Medical Institutions, Baltimore, MD 21231, USA. [2]Department of Mathematical Sciences, Johns Hopkins University, Baltimore, MD 21218, USA.

*To whom correspondence should be addressed. E-mail: kinzlke@jhmi.edu