Am. J. Hum. Genet. 61:1000–1005, 1997

SOMATIC GENETICS '97
"Mistakes Happen": Somatic Mutation and Disease

Feng Qian and Gregory G. Germino

Department of Medicine, Division of Nephrology, The Johns Hopkins University School of Medicine, Baltimore

An intriguing quality of many genetic diseases in humans is the significant intrafamilial variability observed in their clinical presentation. It seems to be likely that definition of the factors responsible for such differences will provide important clues to the normal function of the genes and to the pathobiology of their associated diseases. For some diseases, such as fragile X syndrome and Huntington disease, it has been shown that genetically related individuals do not necessarily have identical germ-line mutations, and instability at the mutant loci may account for much of the observed clinical variation. However, these diseases are the exception to the rule, and other explanations must be sought.

The results of breeding experiments on mice suggest that variants at other genetic loci may play an important role in modification of the phenotypic expression of many mutant genotypes. Transfer of a mutant gene from one strain to another may greatly alter the severity of disease. For example, it is well recognized for *Drosophila* that, in the heterozygous state, certain recessive alleles can result in a phenotype if introduced into a genetic background that has so-called enhancer alleles at other loci. This strategy has been exploited to dissect out signaling pathways and to identify interacting molecular partners (Simon et al. 1991). One can easily imagine how family members who share the same mutant genotype might have different alleles at modifying loci and thus might present with a milder or a more severe disease.

However, genetic background is unlikely to explain all aspects of phenotypic variability in human disease. This, perhaps, is illustrated best by autosomal dominant diseases that present with focal manifestations, such as autosomal dominant polycystic kidney disease (ADPKD) (Gabow 1993). ADPKD describes a group of at least three genetically distinct disorders with essentially identical phenotypes. The clinical hallmark of this group of diseases is the focal formation of cysts within the kidney and liver, which increase in both size and number over the lifetime of an individual. It is estimated that <5% of all nephrons (~106 nephrons per kidney) acquire cysts. Hepatic cysts, presumed to be derived from the epithelia of intrahepatic biliary ductules, develop in only ~40% of affected individuals and vary greatly in number. Cerebral aneurysms, one of the most feared complications associated with ADPKD, are focal and develop in a relatively small portion of all affected individuals.

At one point, an unstable mutation had been postulated to explain these findings (Reeders 1992), but, with the subsequent discovery of the genes for the two most common forms of ADPKD, *PKD1* (chromosome 16; ~85%) and *PKD2* (chromosome 4; ~15%), this explanation has been ruled out (Burn et al. 1995; Mochizuki et al. 1996). For either gene, all germ-line mutations that have been identified thus far have been stable within families. In one family, for instance, a child with a severe infantile presentation was shown to carry the same nonsense mutation of *PKD1* as that found in another, clinically unaffected child (Peral et al. 1996).

If every tubular cell within an affected individual's kidney or biliary tract inherits the same germ-line mutation and the same alleles at modifying loci, why do only a small fraction become cystic? Several models are worth brief consideration. One obvious explanation is that the temporal and spatial patterns of expression of the normal gene may determine which tissues are affected. It also is possible that mutations that promote abnormal patterns of gene expression may confer a phenotype specifically on cells in which the mutant gene product accumulates. The phenotype may reflect the abnormal expression pattern of the mutant gene product. An alternative explanation is that some subsets of cells have redundant pathways that compensate for the effect of the mutation. Similarly, disease expression may be defined by the differential expression of other pathway components. Local factors such as metabolic load, environmental milieu, or stochastic effects also may play an important role. Finally, somatic mutation has been

Received September 12, 1996; accepted for publication September 22, 1997; electronically published October 29, 1997.

Address for correspondence and reprints: Dr. Gregory G. Germino, Department of Medicine, Division of Nephrology, The Johns Hopkins University School of Medicine, Baltimore, MD 21205-2196. E-mail: ggermino@welchlink.welch.jhu.edu

This article represents the opinion of the authors and has not been peer reviewed.

0002-9297/97/6105-0003$02.00

shown to be an important second step in the pathogenesis of many familial cancer syndromes.

The Two-Hit Model and Cyst Formation in ADPKD

Which of these mechanisms might explain the variability of focal cyst formation in ADPKD? Recent studies have provided partial answers to this question, for PKD1. Although there are conflicting reports regarding the protein's pattern of expression, it appears to be likely that the *PKD1* gene product, polycystin, is widely expressed in both fetal and adult tissues and appears to be more abundant in cystic tissue (Bacallao and Carone 1997). These data have been interpreted by some to suggest that *PKD1* mutant products must act by means of a gain-of-function or a dominant-negative mechanism. However, if either mechanism is to blame, the normal pattern of expression does not easily explain why cysts are focal. These observations suggest that cyst formation requires at least two steps, an inherited susceptibility and a subsequent so-called activating event, which is rate limiting.

Recent data from our laboratory (Qian et al. 1996), which were corroborated by Brasier and Henske (1997), suggest that the second step in the pathogenesis of PKD1 is likely to be an acquired mutation of the normal allele of the *PKD1* gene. We developed a new method for isolating DNA from the epithelial cells lining single renal cysts and used it to show that renal cysts in ADPKD are monoclonal. We demonstrated loss of heterozygosity (LOH) within individual cysts, for two closely linked polymorphic markers located within the *PKD1* gene. Genetic analysis confirmed that the normal haplotype was lost or mutated, in somatic tissues. Brasier and Henske (1997), using two different techniques (primary cell culture and cyst scraping), reported similar findings.

On the basis of a relatively low rate of detection of LOH (~20% of cysts), some have suggested that the phenomenon may be a consequence of the cystic state rather than an important step in its pathogenesis (Ong and Harris 1997). Somatic involvement of the neighboring *TSC2* gene also had been suggested as a possible explanation. Although more data will be required in order to resolve the matter, several pieces of evidence support the so-called two-hit genetic model, for PKD1. First, it is very likely that the low rate of detection reflects the limits of our ability to identify mutations rather than a lack of somatic mutation. Indeed, one of the cyst samples had acquired a 2-bp deletion in exon 46 in its normal allele, a lesion that would have been overlooked by focusing on large-scale deletions that affect markers linked to *PKD1* (Qian et al. 1996). We subsequently have discovered other subtle intragenic mutations that appear to represent so-called second hits in cyst formation (T. Watnick, personal communication). It is important to note that only a small number of germ-line mutations have been reported since the initial report of the gene's sequence, in 1994, despite intense screening by many groups (The European Polycystic Kidney Disease Consortium 1994; Peral et al. 1997; Watnick et al. 1997). Second, the discovery that the somatic mutations can be intragenic argues against a role for *TSC2* in the pathogenesis. Finally, somatic mutation was found to have altered the normal haplotype in all cysts for which a haplotype could be determined (10/10 cysts). The probability that this could occur by chance is <1/1,000.

An obvious question is whether a similar two-step process is responsible for cyst formation in the other forms of ADPKD. The essentially indistinguishable clinical presentation of the three forms suggests that this is likely to be the case. Although germ-line mutations affecting *PKD2* and *PKD3* undoubtedly are responsible for the so-called first hit in the respective forms of ADPKD (Mochizuki et al. 1996), the nature of the second events is not yet known. An important consideration is the extraordinary number of somatic mutations that apparently must occur to account for the number of cysts observed (10^2 to 10^4). For PKD1, we speculated that an unusual polypyrimidine tract within intron 21 of the gene might be responsible for both the high frequency of the disease within the population and the large number of second hits observed in cystic epithelia (Qian et al. 1996). For PKD2, however, no similar elements have been discovered.

The apparently high rate of cyst formation in PKD2 and PKD3 implies one of two possibilities: (1) the second-hit rates of *PKD1, PKD2,* and *PKD3* reflect the average rate of somatic mutation in the genome rather than an unusually high rate of somatic mutation resulting from a locally mutable structure within each gene, or (2) somatic inactivation of *PKD1* may be the second step that leads to clonal expansion in the other forms of ADPKD. Reports showing that the *PKD1* and *PKD2* gene products are likely to interact is consistent with the latter hypothesis (Qian et al. 1997). This model predicts that the relative frequencies of the forms of the disease in the population are determined by the rate of germline mutation at each genetic locus, whereas the frequency with which cysts form in each disorder is determined by the rate of somatic mutation of *PKD1*. Further study is required, to distinguish between the two models. The results may have important implications for investigators interested in the problem of somatic mutability.

The Two-Hit Model in Other Nonmalignant Diseases

These observations suggest that ADPKD is one of a growing list of disorders to which the two-hit model, first proposed by Knudson (1971) to explain the origin of retinoblastoma, seems to apply. Although most ex-

amples involve tumor-suppressor genes implicated in the origin of various other types of cancer, it increasingly has been recognized that benign tumors may develop by a similar mechanism. The tumors that develop in individuals suffering from either tuberous sclerosis (TSC) (type I or type II) or neurofibromatosis type I (NF1) are examples.

TSC is an autosomal dominant trait that, like ADPKD, is characterized by significant intrafamilial phenotypic variability and focal manifestations. One hallmark of the disease is the development of benign tumors, known as "hamartomas," in multiple organs. LOH has been found for markers flanking the loci for *TSC1* (chromosome 9q34) and *TSC2* (chromosome 16p13.3). A study by Sampson et al. (1997), reported in the October issue of the *Journal,* suggests that the two-hit mechanism may be involved in causing other clinical manifestations in TSC2. The authors studied 27 patients with TSC and multiple bilateral renal cysts. They found that 22 patients had contiguous gene deletions that removed sizable portions of *PKD1* and *TSC2,* 3 had gross rearrangements involving only *TSC2,* and 2 had no detectable mutations of *TSC2*. Since no significant cystic disease had been documented for the affected members of the TSC1 families, the authors concluded that significant renal cystic disease in TSC usually reflects mutational involvement of *PKD1*. We speculate that somatic mutation of *PKD1* is also likely to be discovered in the renal cysts of patients with TSC. In addition, one might speculate that some of the other uncommon extrarenal manifestations of TSC that are also observed in PKD1—such as hepatic cysts, cerebral and aortic aneurysms, and cardiac valvular disease—might arise by a similar mechanism.

For NF1, the malignancies that complicate this disease were thought to follow the two-hit hypothesis, but the etiology of the benign tumors was less clear, since a number of studies initially failed to detect LOH in neurofibromas. Demonstration of LOH for chromosome 17q11 markers in the benign neurofibromas was more difficult, because the lesions were comprised of an admixture of cell types. Colman et al. (1995) resolved the matter by using a panel of intragenic markers to demonstrate a decrease in signal intensity for alleles of a haplotype in the tumor DNA. Sawada et al. (1996) subsequently identified a 4-bp somatic deletion in *NF1* exon 4b of the previously normal allele from a microdissected benign neurofibroma with a known constitutional germline mutation.

One disease that recently has been proposed as an excellent candidate for the two-hit model is hereditary hemorrhagic telangiectasia (HHT), an autosomal dominant disease of blood vessels, characterized by mucocutaneous telangiectasias and vascular malformations in the lung and brain. The lesions are discrete, appear in a random distribution, and increase in size and number with the age of the individual. In most families, the disease arises from mutations of either (1) *ENG,* the gene for endoglin, a TGF-β (transforming growth-factor β) binding protein of endothelial cells (HHT1) (Shovlin et al. 1997), or (2) *ALK-1,* which encodes a type I receptor of the TGF-β superfamily (HHT2) (Berg et al. 1997). The results of the studies by Shovlin et al. (1997) and Berg et al. (1997) suggest that mutations of both *ENG* and *ALK-1* result in functionally null alleles. In two individuals with HHT2, the mutant mRNA was absent or hardly detectable, by reverse-transcriptase PCR, in the cDNA of peripheral blood leukocytes. Similarly, a stable mutant *ENG* transcript was not produced in lymphocytes of two individuals with HHT1. Five other HHT1 mutations were discovered to produce severely altered mRNAs that were predicted to encode greatly truncated proteins.

Whereas Shovlin et al. (1997) suggested that haploinsufficiency is likely to be the mechanism of disease for HHT1, Berg et al. (1997) proposed a two-hit mechanism for HHT2, on the basis of the focal nature of the disease. They noted, however, that this hypothesis may be very difficult to prove. The telangiectasias appear to result from dilatation of blood vessels rather than from proliferative growth, and loss of *ALK-1* function may induce remodeling of the vascular bed rather than alteration of the rate of endothelial cell proliferation. LOH or somatic mutation may occur only in the single layer of endothelium lining the structure. Berg et al. (1997) acknowledged that formal proof of their hypothesis will require the use of methods that are capable of isolating relatively pure populations of the endothelial cells that line the vascular malformation. Given the overlap in the clinical presentations of HHT1 and HHT2, one must consider a two-hit model for HHT1 as well.

The PKD1 and NF1 examples illustrate one of the difficulties often faced by investigators seeking to identify somatic mutations. Focal lesions often are small and lack clear boundaries with the surrounding normal tissue. Moreover, the cells in these structures frequently are heterogeneous in origin. Contamination by normal tissue may complicate the analyses and may obscure LOH. In the case of PKD1, our initial attempts at demonstrating clonality for cysts failed when we used microdissection techniques to isolate the epithelia, because the cyst walls contained nonepithelial cells. Thus, in the pathogenesis of the focal manifestations of diseases other than cancers, somatic mutation may play a role that is much more common than has been shown heretofore.

The Two-Hit Model and the Haploinsufficient State

These observations prompt one to consider whether somatic mutations also might play a role in the patho-

genesis of other diseases for which haploinsufficiency has been proposed as the molecular mechanism. Fisher and Scambler (1994) previously suggested this possibility, to account for the phenotypic variability observed in at least some of the haploinsufficiency syndromes. Fisher and Scambler cited the CATCH22 (cardiac defect, abnormal facies, thymic hypoplasia, cleft palate, hypocalcemia, and 22q11 deletion) syndrome as an example for which the clinical features could be explained nicely by a two-hit process. Fisher and Scambler suggested that the pattern of disease might be determined by the specific cell lineage affected by the second hit. Somatic mutation in one neural-crest line might result in a conotruncal defect, whereas mutation in another might result in parathyroid aplasia.

Holoprosencephaly type 3 (HPE3) is another autosomal dominant disorder for which the clinical features could be explained by a similar process. This disease is characterized by considerable intrafamilial phenotypic variability and a broad spectrum of clinical severity. Roessler et al. (1996) recently reported that loss of one allele of the sonic hedgehog gene (*SHH*) is sufficient to cause the disease in humans, whereas loss of both alleles is necessary to cause the characteristic phenotype in mice. Although a species-specific difference in the *SHH* pathway cannot be excluded, a two-hit model also could explain the phenotypic variability. Mice have a smaller target size and lower general mutation frequency than those of humans and, thus, may acquire somatic mutations of *SHH* much less frequently. This difference could explain why mice with one mutant allele do not develop abnormalities.

It is reasonable to speculate that somatic mutation may be required for some forms of piebaldism. This autosomal dominant disorder of pigmentation is characterized by patches of white skin and hair and is genetically heterogeneous. The *KIT* proto-oncogene and its ligand, steel factor, have been implicated in the pathogenesis of this disorder. A wide spectrum of mutations of the *KIT* proto-oncogene have been described and include both dominant-negative and null alleles (Spritz et al. 1992). One explanation for the apparent paradox is that the pathway is exquisitely sensitive to the level of functional protein. The two-hit model provides an attractive alternative solution. The likely net effect of a dominant-negative mutation is a reduction of functional activity much below that seen in the haploinsufficient state. A similar decrease in net functional activity may result in cells with null alleles only after they have acquired a second hit (from either *KIT* or another gene in its pathway).

Crosby et al. (1993) proposed the same phenomenon as an explanation for the origin of sporadic congenital anomalies in mice with the so-called disorganization (*Ds*) mutation. The *Ds* mutation results in a dominant condition with variable expression and incomplete penetrance, presenting with a wide range of congenital abnormalities. Crosby et al. showed that the occurrence of the abnormalities in *Ds* mice followed a Poisson distribution, suggesting that multiple anomalies in a given mouse result from independent second events. The truly dominant nature of the *Ds* mutation suggests that the second events are likely to be somatic mutations affecting other genes.

Somatic Mutation and Clinical Variability

These data lend further support to the hypothesis that, in phenotypic variation and perhaps even in the aging process, somatic mutation may play a role much larger than previously suspected. Whereas most acquired mutations are expected to be silent in diploid organisms, a number of situations might allow these effects to be manifested:

1. Germ-line mutations, such as those in *Ds* (or *PKD2?*), might combine with a somatic mutation that disrupts the function of an interacting partner or a pathway member in a manner analogous to that observed for enhancer alleles in *Drosophila* (Simon et al. 1991).

2. The somatic mutation might inactivate the second allele in a heterozygous individual. Virtually all individuals are likely to be heterozygous carriers for multiple recessive alleles. Cell lineages that acquire somatic mutations at any of these loci might be predicted to express a recessive phenotype. In some genes, this might result in cell death and gradual loss of organ function (particularly if the cells are not proliferative, such as with mature neurons). In developmental pathways, focal loss could reduce the number of progenitor cells, with variable consequences.

3. The mutation might occur at a locus that is subject to X inactivation. This group of genes is functionally hemizygous. Cells that acquire inactivating mutations affecting any gene that undergoes X inactivation will lose that gene's function. Paroxysmal nocturnal hemoglobinuria (PNH) is a well recognized example of this phenomenon. The PNH abnormality has been demonstrated to be due to a somatic mutation of the X-linked *PIG-A* gene (Bessler et al. 1994). The disease does not show sex preference, because the acquired mutation occurs after X-chromosome inactivation.

4. The somatic mutation might result in products with a true dominant or dominant-negative activity. The role of acquired activating mutations of various proto-oncogenes has been well established in the origin of cancer. McCune-Albright syndrome is another example of this phenomenon. This sporadic disease is characterized by polyostotic fibrous dysplasia, cafe-au-lait pigmentation of the skin, and multiple endocrinopathies. It results from an acquired mutation of the stimulatory G protein,

$G_s\alpha$, which occurs early in development and causes a mosaic population of cells (Weinstein et al. 1991). The pattern of expression of the disease is determined by the number and location of cells bearing the mutation.

This discussion naturally leads to the final question: how frequently do somatic mutations actually occur? The rate would greatly determine the relative role this process has in contributing to normal phenotypic variation, the pathogenesis of disease, or the biology of aging. ADPKD may offer a unique opportunity to study this problem in vivo (but, see Tischfield 1997 [in this issue] for a complementary approach using primary cell culture). As noted earlier, the three forms of the disease have identical patterns of disease expression, with relatively similar rates of cyst formation (PKD2 may be slightly milder). If analysis of PKD2 and PKD3 cysts reveals LOH for their respective genes, the data would suggest that the rate of somatic mutation at three independent loci is approximately similar. These results could provide a minimal estimate of the average rate of somatic mutation in the human genome, in at least one solid organ.

Acknowledgments

The authors thank Drs. Onuchic, Piontek, and Watnick for their helpful discussions and important contributions to the study of this problem. We are grateful to the ADPKD family members who have participated in our studies and to the Polycystic Kidney Research Foundation (PKRF) and the many physicians and staff members who have kindly helped in procuring samples. This work was supported by grants from the PKRF, the McKutcheon Foundation, and the NIH (DK48006). G.G.G. is the Irving Blum Scholar of the Johns Hopkins University School of Medicine.

References

Bacallao RL, Carone FA (1997) Recent advances in the understanding of polycystic kidney disease. Curr Opin Nephrol Hypertens 6:377–383

Berg JN, Gallione CJ, Stenzel TT, Johnson DW, Allen WP, Schwartz CE, Jackson CE, et al (1997) The activin receptor–like kinase 1 gene: genomic structure and mutations in hereditary hemorrhagic telangiectasia type 2. Am J Hum Genet 61:60–67

Bessler M, Mason P, Hillmen P, Luzzatto L (1994) Somatic mutations and cellular selection in paroxysmal nocturnal haemoglobinuria. Lancet 343:951–953

Brasier JL, Henske EP (1997) Loss of the polycystic kidney disease (PKD1) region of chromosome 16p13 in renal cyst cells supports a loss-of-function model for cyst pathogenesis. J Clin Invest 99:194–199

Burn TC, Connors TD, Dackowski WR, Petry LR, Van Raay TJ, Millholland JM, Venet M, et al (1995) Analysis of the genomic sequence for the autosomal dominant polycystic kidney disease (PKD1) gene predicts the presence of a leucine-rich repeat. Hum Mol Genet 4:575–582

Colman SD, Williams CA, Wallace MR (1995) Benign neurofibromas in type 1 neurofibromatosis (NF1) show somatic deletions of the NF1 gene. Nat Genet 11:90–92

Crosby JL, Varnum DS, Nadeau JH (1993) Two-hit model for sporadic congenital anomalies in mice with the disorganization mutation. Am J Hum Genet 52:866–874

European Polycystic Kidney Disease Consortium, The (1994) The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within a duplicated region on chromosome 16. Cell 77:881–894

Fisher E, Scambler P (1994) Human haploinsufficiency—one for sorrow, two for joy. Nat Genet 7:5–7

Gabow PA (1993) Autosomal dominant polycystic kidney disease. N Engl J Med 329:332–342

Knudson AG Jr (1971) Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci USA 68:820–823

Mochizuki T, Wu G, Hayashi T, Xenophontos SL, Veldhuisen B, Saris JJ, Reynolds DM, et al (1996) PKD2, a gene for polycystic kidney disease that encodes an integral membrane protein. Science 272:1339–1342

Ong AC, Harris PC (1997) Molecular basis of renal cyst formation—one hit or two? Lancet 349:1039–1040

Peral B, Gamble V, Strong C, Ong AC, Sloane-Stanley J, Zerres K, Winearls CG, et al (1997) Identification of mutations in the duplicated region of the polycystic kidney disease 1 gene (*PKD1*) by a novel approach. Am J Hum Genet 60: 1399–1410

Peral B, Ong AC, San Millan JL, Gamble V, Rees L, Harris PC (1996) A stable, nonsense mutation associated with a case of infantile onset polycystic kidney disease 1 (PKD1). Hum Mol Genet 5:539–542

Qian F, Germino FJ, Cai Y, Zhang X, Somlo S, Germino GG (1997) PKD1 interacts with PKD2 through a probable coiled-coil domain. Nat Genet 16:179–183

Qian F, Watnick TJ, Onuchic LF, Germino GG (1996) The molecular basis of focal cyst formation in human autosomal dominant polycystic kidney disease type I. Cell 87:979–987

Reeders ST (1992) Multilocus polycystic disease. Nat Genet 1:235–237

Roessler E, Belloni E, Gaudenz K, Jay P, Berta P, Scherer SW, Tsui LC, et al (1996) Mutations in the human sonic hedgehog gene cause holoprosencephaly. Nat Genet 14:357–360

Sampson JR, Maheshwar MM, Aspinwall R, Thompson P, Cheadle JP, Rivine D, Roy S, et al (1997) Renal cystic disease in tuberous sclerosis: role of the polycystic kidney disease 1 gene. Am J Hum Genet 61:843–851

Sawada S, Florell S, Purandare SM, Ota M, Stephens K, Viskochil D (1996) Identification of NF1 mutations in both alleles of a dermal neurofibroma. Nat Genet 14:110–112

Shovlin CL, Hughes JMB, Scott J, Seidman CE, Seidman JG (1997) Characterization of endoglin and identification of novel mutations in hereditary hemorrhagic telangiectasia. Am J Hum Genet 61:68–79

Simon MA, Bowtell DD, Dodson GS, Laverty TR, Rubin GM (1991) Ras1 and a putative guanine nucleotide exchange factor perform crucial steps in signaling by the sevenless protein tyrosine kinase. Cell 67:701–716

Spritz RA, Holmes SA, Ramesar R, Greenberg J, Curtis D,

Beighton P (1992) Mutations of the *KIT* (mast/stem cell growth factor receptor) proto-oncogene account for a continuous range of phenotypes in human piebaldism. Am J Hum Genet 51:1058–1065

Tischfield JA (1997) Loss of heterozygosity or: how I learned to stop worrying and love mitotic recombination. Am J Hum Genet 61:995–999 (in this issue)

Watnick TJ, Piontek KB, Cordal TM, Weber H, Gandolph MA, Qian F, Lens XM, et al (1997) An unusual pattern of mutation in the duplicated portion of PKD1 is revealed by use of a novel strategy for mutation detection. Hum Mol Genet 6:1473–1481

Weinstein LS, Shenker A, Gejman PV, Merino MJ, Friedman E, Spiegel AM (1991) Activating mutations of the stimulatory G protein in the McCune-Albright syndrome. N Engl J Med 325:1688–1695