Part I

REVIEWS

1

GENOTYPE TO PHENOTYPE

Sue Malcolm
Clinical and Molecular Genetics Unit, Institute of Child Health, London, UK

■ INTRODUCTION

One might expect that close study of an undergraduate textbook would be all that is needed to understand this topic. If a codon is mutated to a stop codon, then full-length protein cannot be produced. If the start codon is mutated, then translation cannot take place, and if the third nucleotide of a codon is mutated so that the genetic code predicts no change in amino acid, then there will be no mutation. In reality, the situation is far from this simple, and exceptions to all the above exist and will be presented. The incredible complexity involved in splicing and the central importance of splicing mutations will also be discussed and provide a uniting theme.

It is also important to appreciate that specific mutations will have very different effects on the function of the protein that one expects to be reflected in a range of disorders corresponding to mutations in a particular gene. This is well illustrated in the craniosynostoses, which are autosomal dominant disorders in which there is premature fusion of the cranial sutures. These are associated with mutations in the fibroblast growth factor receptor (FGFR) gene family. The mutations found in Crouzon and Pfeiffer syndromes predominantly create or remove a cysteine to leave an unpaired cysteine in one of the three extracellular immunoglobulin-like domains. This results in cross-linking and covalent dimerization of subunits leading to constitutive activation of the receptor molecules, which become independent of the need for a ligand (fibroblast growth factor, FGF). On the other hand, when a specific mutation occurs in the linker region between the second and third loop a distinct clinical variant, Apert syndrome, occurs. In contrast, the specific amino acid

Guide to Mutation Detection, Edited by Graham R. Taylor and Ian N. M. Day.
ISBN 0-471-23444-3 Copyright © 2005 by John Wiley & Sons, Inc.

substitutions cause enhanced or changed binding affinities to FGFs. The majority of cases are caused by either a Ser252Trp or a Pro253Arg mutation in FGFR2. The only other missense change leading to Apert syndrome is Ser252Phe. This will occur much less often than Ser252Trp because, as a result of the constraints of the genetic, code a double substitution, CG_TT, is required, but it is significant that both result in replacement of serine by a large hydrophobic residue. In one family where Ser252 mutated to leucine, which has completely different properties as an amino acid and is achieved by a single nucleotide substitution, the phenotype was so mild that some members were considered to be unaffected (reviewed in 1).

■ NONSENSE MUTATIONS

Contrary to intuition, the predominant consequence of nonsense mutations is not the synthesis of truncated proteins. The majority of nonsense transcripts are recognized and efficiently degraded by the cell via a pathway known as nonsense-mediated mRNA decay (NMD). This mechanism is thought to protect the organism from potentially harmful dominant-negative or gain-of-function effects of truncated proteins that could result if nonsense transcripts were stable (2).

Exon skipping often results when either the intronic g or t at a splice donor site is mutated. However, Dietz et al. (3) were the first to describe the skipping of constitutive exons in vivo induced by nonsense mutations. Maintenance of an open reading frame provides an additional level of scrutiny during splice site selection. A deletion found at the cDNA level should always be confirmed at the genomic level. It remains to be seen what proportion of these skipped exons actually arise because of disruption of an exonic splice enhancer (see below).

■ CHANGES INVOLVING SPLICING

Nucleotide changes within and around splice donor and acceptor sites have long been recognized as a major source of mutation in humans and are routinely screened for. However, they will seriously underrepresent the proportion of mutations that are in fact operating via the splicing system. A systematic study of mutations in ataxia-telangiectasia showed that 30/62 (48%) resulted from defective splicing (4), but fewer than half involved the canonical AG splice acceptor site or GT splice donor site. Others occurred at less stringently conserved sites, including silent mutations of the last nucleotide of exons, mutations elsewhere in the consensus splice sites, and creation of splice-acceptor or splice-donor sites in either introns or exons. These splicing mutations led to a variety of consequences, including exon skipping and, to a lesser degree, intron retention, activation of cryptic splice sites, or creation of new splice sites. Several cases of exon skipping in both normal controls and patients for whom no underlying defect could be found in genomic DNA were also observed, suggesting other so far undetected sites of mutation. Another cDNA-based study of neurofibromatosis identified splice mutations in 50% of patients for whom a mutation was found (5). Again, most of the mutations did not involve the conserved AG/GT motif. Several mutations within both exons and introns resulted in the activation of cryptic splice sites. For example, a mutation 332 bases into IVS30 resulted in the change of AGACataagt to AGACgtaagt, which gives a Shapiro and Senapathy (6) score of 86.49.

A recent interesting case involves two missense mutations in the Krit1 gene in families with cerebral cavernous malformations. The point mutations were predicted to lead to the changes D137G and Q210E, but further RNA analysis revealed that both point mutations actually activate cryptic splice-donor sites, causing aberrant splicing and leading to a frameshift (7).

More recently it has been demonstrated that some mutations within coding regions may in fact be exercising their effect through splicing by disrupting exonic splice enhancers (ESEs). Accurate splicing is a complicated business requiring an array of small nuclear ribonucleoproteins and other factors that are components of the spliceosome. ESEs are present in constitutive and alternatively spliced exons and are required for efficient splicing of those exons. The ESEs in pre-mRNAs are recognized by serine/arginine-rich (SR) proteins, a family of essential splicing factors that also regulate alternative splicing. ESEs contain a wide spectrum of sequences of approximately 6–8 nucleotides, but they are hard to detect because of their degeneracy. When missense mutations are identified in genomic DNA, particularly in a diagnostic laboratory, the usual interpretation is that the affected amino acid is crucial for the function of the protein, particularly if the change is nonconservative or the residue changed is highly conserved in evolution. But if these mutations disrupt an ESE, the consequence may be incorrect splicing of the transcript. Experimental evidence for this comes from studies in the breast cancer-causing gene BRCA1 (8). A Glu1694Ter nonsense mutation in exon 18, which causes skipping of the entire exon, has been found in a number of families with a breast cancer predisposition. Although it is difficult to search directly for ESE disruption because of the sequence degeneracy, the G_T change was predicted to disrupt a high scoring motif in a matrix analysis. Minigene constructs were tested for splicing in a HeLa cell nuclear extract. Exon jumping correlated with the disruption of high-scoring motifs rather than the presence of nonsense codons. It follows that the high-scoring motifs may also be disrupted by changes leading to missense mutations or silent third base changes. and because of the degeneracy of the ESEs these events may be frequent.

The effects on phenotype may be varied and subtle. The molecular basis for the distinction between Duchenne and the clinically less severe Becker muscular dystrophy is generally explained by the frameshift hypothesis. Both disorders can result from deletions in the dystrophin gene. but the outcome depends on the exact combination of exons removed. Deletions resulting in Duchenne have been shown to shift the translational open reading frame of triplet codons (because not all exons contain an exact multiple of three nucleotides), leading to in-frame stop codons and a truncated protein. Deletions identified in Becker result from a combination of exons that maintain the open reading frame and predict a shorter but partially functional protein. In a minority of Duchenne cases nonsense mutations are observed, again predicting a truncated protein. A Japanese patient was reported with a nonsense mutation in exon 27 of the dystrophin gene but a phenotype of Becker. It was shown by functional studies using an in vitro splicing system that the mutation led to exon skipping of exon 27 by interruption of an exon-specific enhancer.

Exon splice enhancers also provided the explanation for initially puzzling correlations of genotype and phenotype in spinal muscular atrophy (SMA). Two closely related survival motor neuron genes, predicted to encode identical proteins, map to the SMA locus. SMA is caused by mutations within the telomeric copy of the gene (SMN1), with 96% of patients showing a homozygous deletion of SMN1. Not only is SMN2 not able to compensate for loss of SMN1, but homozygous absence of

SMN2 is found in 5% of control individuals with no phenotypic effect. SMN1 produces full-length mRNA, but SMN2 expresses dramatically reduced full-length mRNA and abundant levels of incorrectly spliced transcript lacking exon 7. The splicing of exon 7 is determined by a single nucleotide difference at position +6 in exon 7 (C in SMN1 and T in SMN2), which disrupts an exon splice enhancer (9). A most exciting development for possible therapy has been the identification in *Drosophila* of a splicing factor, Htra2-β1, which promotes the inclusion of exon 7 into full-length SMN2 transcripts and works across species in human and mouse cells carrying an SMN2 minigene (10).

■ TROUBLE IN INTRONS

As mutation detection becomes more exhaustive, an increasing number of mutations are being reported deep within introns. This must pose enormous problems for diagnostic laboratories as the introns are in general many times longer than the exons and because functional tests for the significance of the change will be difficult.

A recent study of patients with the cardiac version of Fabry disease found a novel midintronic base substitution that causes an insertion of a "pseudo exon" of 57 base pairs between the normal exons 4 and 5 (11). The base change is not within the canonical GT/AG but instead increases the activity of a rare alternative splice site that normally only contributes <5% of the α-galactosidase mRNA levels but in this case produces >70% of the total α-galactosidase mRNA. Consequently, residual activity of the normal product is reduced to about 10% of normal, leading to a cardiac phenotype. The base change occurs within the additional exon, 4 base pairs before the 3′ end of it, and has been suggested to act by exon splice enhancement.

Exon splice silencers are even less well understood in humans, but a fascinating example was recently reported (12). A 4-base pair deletion within intron 20 of the ATM gene results in the incorporation of an additional pseudo-exon into the mRNA. The deletion was shown to occur in a new splice processing element that is essential for accurate intron removal and is part of a new U1 small nuclear ribonucleoprotein (snRNP) binding site.

■ HOW MANY HITS ARE NEEDED?

Increasingly, our simple models of autosomal dominant, recessive or, X-linked modes of inheritance are inadequate. Despite heroic efforts to unravel the molecular genetics of common disorders, progress has been slow. Two recent success stories give us some clue to why it has been so difficult. An association has been found between an insertion mutation in NOD2 gene and Crohn disease (inflammatory bowel disorder) in German and British populations (13). The variant is a C-insertion mutation in exon 11 in a leucine-rich region leading to a premature stop codon at amino acid 1007 of the protein and was initially submitted to a public database as a polymorphism. The nature of the change would imply that it is almost certainly pathogenic. This is indeed true, as a strong association was observed with the allele occurring at 4.6% in the general population and at 19% in the cases. However, even in such an apparently clear-cut case the extra risk conferred by carrying

the heterozygous genotype is less than 3-fold, although the homozygous extra risk is 42-fold.

A second example has arisen during the search for minor genes contributing to breast cancer (14). CHEK2 is a cell cycle checkpoint kinase directly connected to the DNA damage response pathway. A previously described protein-truncating mutation in *CHEK2*, 1100delC, was found to segregate with breast cancer in a branch of a large family EUR60 that mapped to neither of the known high-penetrance breast cancer genes, *BRCA1* or *2*. An association between breast cancer and 1100delC was found in *BRCA1/BRCA2*-negative families compared with controls. The variant was found in 1.2% of the healthy population and 5.1% of the breast cancer patients. They estimated that the *CHEK2**1100delC variant results in an approximately 2-fold increase of breast cancer risk in women and a 10-fold increase of risk in men. Even in EUR60, it was carried by only 8 of the 17 affected women in the family.

Both these examples are proven examples of pathogenic mutations, and both are functionally nonsense mutations. However, the remarkably low penetrance shows how little we understand of the modifying factors, including environmental factors and modifying genes, that control the genotype-phenotype correlations.

A tantalizing glimpse of such interactions comes from studies showing the need for three mutant alleles that are emerging in the case of Bardet–Biedl syndrome (15).

■ ACKNOWLEDGMENTS

I thank Adrian Krainer for stimulating my interest in RNA splicing.

■ REFERENCES

1. Wilkie AOM, Morris-Kay GM (2001) Genetics of craniofacial development and malformation. Nat Rev Genet 2: 458–468.
2. Mendell JT, Dietz HC (2001) When the message goes awry: disease-producing mutations that influence mRNA content and performance. Cell 107: 411–414.
3. Dietz HC, Valle D, Francomano CA, Kendzior RJ, Pyeritz RE, Cutting GR (1993) The skipping of constitutive exons in vivo induced by nonsense mutations. Science 259: 680–683.
4. Teraoka SN, Telatar M, Becker-Catnia S, Liang T, Onengut S, Tolun A, Chessa L, Sanal O, Bernatowska E, Gatti RA, Concannon P (1999) Splicing defects in the ataxia-telangiectasia gene, ATM: Underlying mutations and consequences. Am J Hum Genet 64: 1617–1631.
5. Ars E, Serra E, Garcia J, Kruyer H, Gaona A, Lazaro C, Estevill X (2000) Mutations affecting mRNA splicing are the most common molecular defect in patients with neurofibromatosis type I. Hum Mol Genet 22: 237–247.
6. Shapiro MB, Senapathy P (1987) RNA splice junctions of different classes of eukaryotes: Sequence statistics and functional implications in gene expression. Nucleic Acids Res. 15: 7155–7174.
7. Verlaan DJ, Siegel AM, Rouleau GA (2002) Krit1 missense mutation lead to splicing errors in cerebral cavernous malformation. Am J Hum Genet 70: 1564–1567.
8. Liu H-X, Cartegni L, Zhang MQ, Krainer AR (2001) A mechanism for exon skipping caused by nonsense or missense mutations in BRCA1 and other genes. Nat Genet 27: 55–58.

9. Lorson CL, Androphy EJ (2000) An exonic enhancer is required for inclusion of an essential exon in the SMA-determining gene SMN. Hum Mol Genet 22: 259–265.
10. Hofmann Y, Lorson CL, Stamm S, Androphy EJ, Wirth B (2000) Htra2-β1 stimulates an exonic splicing enhancer and can restore full-length SMN expression to survival motor neuron 2 (SMN2). Proc Natl Acad Sci USA 97: 9618–9623.
11. Ishii S, Nakao S, Minamikawa-Tachino R, Desnick RJ, Fan J-Q (2002) Alternative splicing in the α-Galactosidase A gene: Increased exon inclusion results in the Fabry cardiac phenotype. Am J Hum Genet 70: 994–1002.
12. Pagani F, Buratti E, Stuani C, Bendix R, Dork T, Baralle FE (2002) A new type of mutation causes a splicing defect in ATM. Nat Genet 30: 426–429.
13. Hampe J, Cuthbert A, Croucher PJ, Mirza MM, Mascheretti S, Fisher S, Frenzel H, King K, Hasselmeyer A, MacPerson AJ, Bridger S, van Deventer S, Forbes A, Nikolaus S, Lennard-Jones JE, Foelsch U, Krawczak M, Lewis C, Mathew CG (2001) Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. Lancet 357: 1925–1928.
14. Meijers-Heijboer et al. (2002) Low penetrance susceptibility to breast cancer due to CHEK2*1100del C in non-carriers of BRCA1 or BRCA2 mutations. Nat Genet 31: 55–59.
15. Badano JL, Kim JC, Hoskins BE, Lewis RA, Ansley SJ, Cutler DJ, Castellan C, Beales PL, Leroux MR, Katsanis N (2003) Heterozygous mutations in BBS1, BBS2 and BBS6 have a potential epistatic effect on Bardet–Biedl patients with two mutations at a second BBS locus. Hum Mol Genet 12: 1651–1659.