Genetics of Childhood Disorders: VIII. Making Sense Out of Nonsense

JOHN M. GREALLY, M.B., B.CH., B.A.O., AND PAUL J. LOMBROSO, M.D.

DNA is a biological polymer that encodes information essential for all physiological processes. The cell devotes considerable resources to maintain the integrity of this information. High priority is given to repairing DNA damage in an ongoing manner, and stringent proofreading mechanisms are in place in dividing cells. Despite these safeguards, however, alterations do occur to the original sequence. These changes may have no effect or can actually benefit the organism as occurs during evolutionary changes. When the alterations are deleterious we recognize these changes as mutations.

Mutations may cause inherited disease when they are present in parental gametes and are passed on to the next generation. They may cause acquired disease when they develop de novo in somatic cells. An example of the former would include achondroplasia, a disorder in which affected individuals have short-limbed dwarfism. This illness results from a mutation in the fibroblast growth factor receptor-3 gene (*FGFR3*). Examples of acquired disorders are characteristically found in neoplasms, such as those affecting the p53 tumor suppressor gene. A key distinction between these 2 forms of mutations is that inherited mutations are present in all cells of the offspring, whereas acquired mutations affect only a subgroup of cells in an individual.

Ten percent of the DNA in the human genome codes for genes. This indicates that a random mutation is more likely to affect intergenic DNA than DNA coding for proteins. While the principal disease-causing effect of mutations seems to be on coding sequences within genes, it is now appreciated that some mutations of the DNA lying between genes have functional consequences on gene expression. This realization is based on the recognition that the intergenic DNA contains sequences essential for the normal regulation of nearby genes. This regulation is usually referred to as epigenetic (*epi:* upon, higher than), so mutations are also classified as genetic (affecting coding sequences) or epigenetic (affecting gene regulation).

Most genes may be divided into functional regions. The promoter area is where the transcriptional machinery gains access to the DNA strand to initiate transcription. Nearby regulatory sequences determine the correct timing of transcription, the level of gene expression, and the specific cell type in which transcription will occur. In addition to these regulatory regions, a gene also contains the exons which actually encode for protein and which must be spliced together to form the processed or mature messenger RNA (mRNA) molecule. The mRNA is then read by the translational machinery as a sequence of triplets. Each triplet of nucleotides encodes individual amino acids that are strung together to form a protein molecule. The sequence of triplets is termed the coding sequence or open reading frame of the mRNA. Much of this material may be reviewed in 2 earlier columns on transcriptional control (April and May, 1998).

When the DNA of a gene is damaged, the cell attempts to repair it. For the most part, these repairs are successful and the cell resumes its normal activity. Occasionally, the repairs are not successful and mutations arise. If the mutation involves the coding region DNA, it may or may not have an effect on the function of the protein. For example, one of the DNA triplets encoding the amino acid leucine is GAA. DNA damage may change this to GAC. The latter triplet also encodes leucine, so no alteration in amino acid sequence of the protein results. If, however, the change were from GAA to GCA, the amino acid encoded is now arginine. This change might still not affect the activity of the protein. If the leucine serves a critical role for the protein, then its replacement with an arginine would in all likelihood alter its function. This type of mutation is recognized as a missense mutation.

The integrity of the process allowing translation of mRNA to protein depends on the maintenance of the series of triplets in an unbroken manner. When a single nucleotide has been changed, the mutation is termed a point mutation. When several nucleotides are either removed or added, the mutation is termed a deletion or an insertion. If the deletion or insertion occurs to a number of nucleotides other than 3 (or a multiple thereof) in the coding sequence, then the sense of the coding sequence message is lost downstream of the mutation. For example, if a series of triplet nucleotides is represented by the series of numbers 123 123 123 123 123 . . . , the insertion of an additional 2 nucleotides would cause the following pattern: 123 1NN 231 231 231 231 . . . . This usually gives rise to the premature termination of the protein as one of the codons that now lies beyond the mutation will have been changed into a termination signal. Moreover, the protein that is produced will bear no resemblance to the original protein beyond the point of the mutation. The end result of such mutations are often nonfunctional proteins. Deletions and insertions that cause alterations in the frame of ref-

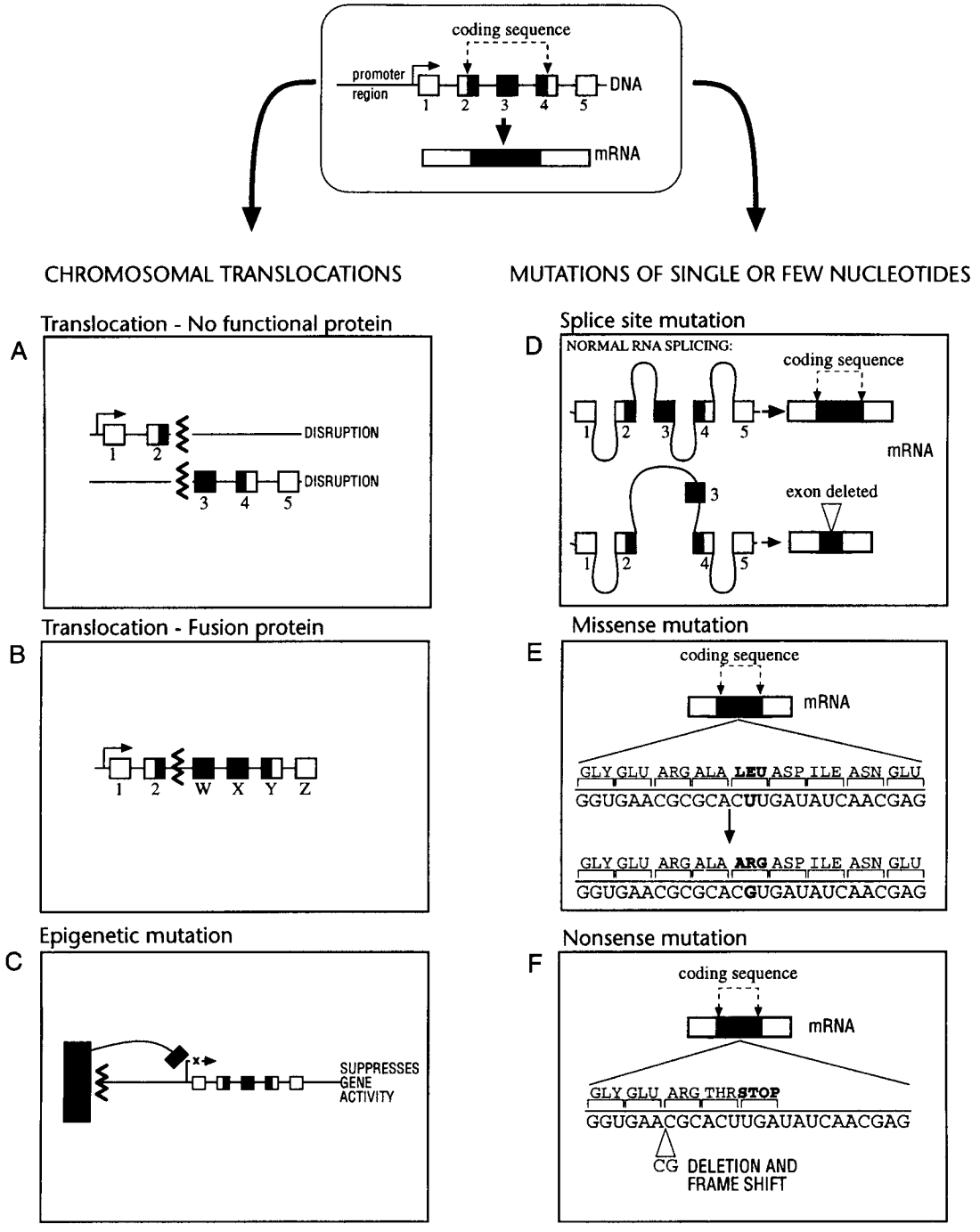

Fig. 1 Two broad mechanisms of mutation are illustrated: structural alterations of chromosomes and nucleotide alterations within chromosomes. If a translocation breaks a gene as shown in panel A, no functional protein is produced. If the translocation places one open reading frame into continuity with that of another gene, a fusion protein consisting of parts of 2 proteins may be formed as depicted in panel B. Even translocations not in the immediate vicinity of a gene can alter gene expression. Panel C illustrates a breakpoint upstream from a gene having a suppressive effect on the gene from a distance, a phenomenon known as position effect. More subtle mutations involving nucleotides within the gene itself can lead to aberrant mRNA splicing (panel D), the substitution of one amino acid for another in the gene's protein product (panel E), or the complete disruption of the downstream message by introduction of a nonsense mutation as shown in panel F.

erence for the amino acid–encoding triplets are referred to as nonsense mutations.

On occasion, the mutation that is present affects the processing of the RNA molecule. This occurs when the mutation lies within regulatory sequences. For example, the formation of exons depends on specific nucleotide sequences flanking the exons and these sequences must be present for normal RNA processing to occur. Mutations affecting these splice donor and acceptor sites can cause exons to be lost or introns to be included in the processed mRNA. These insertions and deletions can also affect the function of the proteins produced.

RNA is not merely a passive molecular messenger. The stability of RNA is variable, and this stability is affected by the nucleotide sequence that is present. Evolutionary selection dictates that functional mRNA molecules encoded by the genomes of complex organisms are stable enough to allow translation within the cell. Thus, a further mechanism that leads to a failure of protein formation appears to be the generation of unstable mRNAs by mutation of stabilizing nucleotide sequences.

Mutations do not always confer a loss of function. Some mutations alter the protein to render it continuously active by removing the normal physiological mechanisms that normally modulate its activity. Such a mechanism accounts for the dominant inheritance pattern of certain genetic diseases. In these disorders, only 1 of the 2 alleles at a particular locus is mutated while the other allele produces a normal protein. However, the normal copy is unable to compensate for the overactive mutated form.

An example of such a process is multiple endocrine neoplasia, type IIA (MEN2A). The mutations causing this disorder affect the *RET* proto-oncogene on 10q11.2. This protein normally resides on the plasma membrane and serves as a receptor for a specific growth factor. Upon binding of the proper ligand, a signal is transferred to the interior of the cell and a specific pathway is activated. The mutation causes an alteration of cysteine residues in the molecule's extracellular domain. This induces the dimerization of the RET receptor on the cell surface and leads to the unregulated and continuous activation of the signaling pathway inside the cell. Even the presence of normal copies of the receptor is unable to modify the effects of the overactive form. This type of mutation is referred to as a gain of function mutation.

RET is a gene that also exhibits the interesting phenomenon of pleiotropy. Pleiotropy is the ability of different mutations within the same gene to cause distinct phenotypes. As mentioned, a gain of function mutations of *RET* leads to MEN2A. Loss of function mutations, on the other hand, leads to a phenotype in which the normal nervous supply fails to form in the colon, leading to a disorder termed Hirschsprung disease. No endocrine tumors occur in Hirschsprung disease, and colonic aganglionosis is very rare in MEN2A. The mutations that occur in Hirschsprung disease are once again located in the *RET* gene, but they do not cause overactivity of the receptor. Instead, they arise in different parts of the receptor molecule and result in its inability to function as a receptor at all. These mutations either disturb its ability to bind to the ligand or interfere with the intracellular transmission of the signal.

The methods by which mutations are detected are rapidly increasing. Direct sequencing of coding regions is the most direct way of finding a mutation, but the presence of multiple exons and large coding sequences can make this process costly and time-consuming. Indirect means of determining the presence of a mutation include the comparison of electrophoretic mobility of single stranded DNA isolated from patients and from normal individuals. Alterations in the nucleotide sequence will cause mobility changes and will target the DNA for subsequent sequencing experiments. Such techniques are particularly useful for assessing large genes with numerous exons where mutations causing disease can occur at multiple loci.

So far, the focus has been on subtle mutations affecting at most several base pairs at a time. Larger-scale events also occur. A germline mutation such as trisomy 21, where an entire extra chromosome 21 is inherited by the offspring, causes the Down syndrome phenotype. Monosomy for the X chromosome (only one X chromosome, without a second sex chromosome, X or Y) can result in spontaneous abortion or, less frequently, in girls with the Turner syndrome phenotype.

Deletions involving multiple genes located in a cluster can be associated with disease. An example of such a disease is Prader-Willi syndrome (PWS). Deletions of chromosome 15q11-q13 affecting multiple genes have the recognizable consequence of PWS, whereas mutations involving individual genes within that cluster do not. Duplications of certain chromosomal regions increase the "dosage" of the genes that are present, sometimes with phenotypic consequences. One process in which this commonly occurs is malignancy.

Rearrangement of chromosomes can have a number of pathogenic consequences. Translocations occur when 2 chromosomes inappropriately come together and exchange genetic material. The breakpoints of such translocations can occur in or near genes and affect their expression. A number of disease-causing genes have been identified by studying translocations in people with specific phenotypes, including neurofibromatosis type I at 17q11.2. Translocations can also place the upstream part of one gene into continuity with the downstream part of another gene, resulting in the production of a "fusion protein." The classic example of this process is termed the Philadelphia translocation that fuses the *BCR* gene with the *ABL* oncogene, resulting in chronic myeloid leukemia due to the unregulated expression of the activated oncogene.

Translocations may have effects on gene expression without physically disrupting a gene. The genome is very heterogeneous

in terms of its ability to support gene expression. Chromosomal banding patterns represent this heterogeneity. Gene-poor regions of the genome are generally found in areas that are heterochromatic and are reflected in the G banding pattern of the chromosome. Gene-rich regions are generally located in areas that are euchromatic and are reflected in the R banding pattern. Placement of a normally euchromatic gene into or adjacent to a heterochromatic region can alter the chromatin environment of the gene, causing suppression of its expression. This phenomenon is well characterized in model organisms such as *Drosophila melanogaster* and is referred to as position effect variegation. A translocation that places heterochromatin in the vicinity of a gene located in a euchromatic region is now recognized as a mechanism that accounts for a number of human genetic diseases. This type of indirect influence on gene regulation is another example of epigenetic dysregulation of gene expression.

Other epigenetic processes can cause human disease. Genomic imprinting is a phenomenon characterized by silencing of a locus on a chromosome from a specific parental origin. For example, the tumor suppressor *CDKN1C* is active on the maternal chromosome 11p15.5 but silenced on the paternal chromosome. Mutation of the paternally inherited gene has no obvious effect, whereas mutation of the maternal copy causes the expression of the gene to drop, not to the 50% level associated with nonimprinted loci, but to drastically low levels, leading to a partial Beckwith-Wiedeman syndrome phenotype. Other reported examples of epigenetic dysregulation, such as alteration of regulatory methylation patterns in triplet repeat disorders, are increasing the interest in epigenetic processes in human disease. The mechanism by which triplet repeat mutations cause illnesses will be reviewed in the next column.

WEB SITES OF INTEREST

http://anatomy.med.unsw.edu.au/cbl/teach/genetics/omimgene.htm
http://www.ncbi.nlm.nih.gov/disease/Signals.html
http://www.ncbi.nlm.nih.gov/disease/Werner.html
http://www.biochemj.org/bj/314/bj3140397.htm
http://www.ncbi.nlm.nih.gov/Omim/

ADDITIONAL READINGS

Hatano M, Aoki T, Dezawa M et al. (1997), A novel pathogenesis of megacolon in Ncx/Hox11L1 deficient mice. *J Clin Invest* 100:795–801

Heim R, Kam-Morgan L, Binnie C et al. (1995), Distribution of 13 truncating mutations in the neurofibromatosis 1 gene. *Hum Mol Genet* 4:975–981

Ledbetter D, Rich D, O'Connell P, Leppert M, Carey J (1989), Precise localization of NF1 to 17q11.2 by balanced translocation. *Am J Hum Genet* 44:20–24

Mulligan L, Kwok J, Healey C et al. (1993), Germ-line mutations of the RET proto-oncogene in multiple endocrine neoplasia type 2A. *Nature* 363:458–460

Rousseau F, Bonaventure J, Legeai-Mallet L et al. (1994), Mutations in the gene encoding fibroblast growth factor receptor-3 in achondroplasia. *Nature* 371:252–254

Shiang R, Thompson L, Zhu Y-Z et al. (1994), Mutations in the transmembrane domain of FGFR3 cause the most common genetic form of dwarfism, achondroplasia. *Cell* 78:335–342

Accepted May 19, 1999.

Dr. Greally is Associate Research Scientist, Department of Genetics, and Dr. Lombroso is Associate Professor, Child Study Center, Yale University School of Medicine, New Haven, CT.

Correspondence to Dr. Lombroso, Child Study Center, Yale University School of Medicine, 230 South Frontage Road, New Haven, CT 06520; e-mail: paul.lombroso@yale.edu.

To read all the columns in this series, visit the Web site at http://info.med.yale.edu/chldstdy/plomdevelop/