CLINICAL IMPLICATIONS OF BASIC RESEARCH

Structural Variation in the Human Genome

James R. Lupski, M.D., Ph.D.

The completion of the Human Genome Project was a remarkable feat and provided 3 billion bases of reference nucleotides for comparative studies. An analogy often used to conceptualize human genetic information is that of an encyclopedia, in which each volume of the set would represent 1 of the 23 pairs of human chromosomes. Sections within each volume would represent the (approximately) 25,000 genes of the human genome, and letters of the alphabet would represent the individual bases of DNA encoding specific amino acids that are the building blocks of proteins. To date, the molecular medicine model that is promulgated in every medical school is based on sickle cell disease, in which the predominant type of mutation is a base-pair change (according to the analogy to the encyclopedia, the substitution of a single letter), which alters the coding sequence and results in the synthesis of a mutant protein. In this model, genetic variation between individuals and between populations arises from variant bases, also known as single-nucleotide polymorphisms.

However, the completion of the reference sequence of the human genome and the development of new technologies to detect the extent and position of genomic alterations within a single human genome have made it apparent that large fragments of our genome have been deleted or duplicated.[1-3] These genomic rearrangements can change the copy number of the genes that lie within the affected regions and alter gene regulation.[4]

A recent study by Redon and colleagues[5] provides the first map of genomic variation in copy number in the complete human genome. The map shows that this variation encompasses two to three times as many nucleotides (by analogy, letters) as are present in single base-pair changes between individual genomes and thus represents a substantial source of genomic variation. In fact, there is a tremendous degree of large structural variation in the human genome among closely related people and between world populations that is caused by large deletions and duplications of genomic segments (Fig. 1). Redon and colleagues identified copy-number variations in 1400 regions that overlap with 14.5% of the genes implicated in human disease, which are listed in Online Mendelian Inheritance in Man (www.ncbi.nlm.nih.gov/omim).

Many variations in copy number are probably benign, but specific variations are associated with common mendelian (single-locus) conditions such as color blindness, Charcot–Marie–Tooth disease type 1A, and other disorders of the nervous system. Variation in copy number can also influence susceptibility to complex diseases such as lupus glomerulonephritis and to infection with the human immunodeficiency virus, as well as to Parkinson's disease, Alzheimer's disease, and Crohn's disease.[4]

The clinical application of genomic arrays — DNA chips that detect changes in the copy number of specific parts of the genome — has made it possible to detect many submicroscopical genomic deletions and duplications that cause complex mental-retardation syndromes. The demonstration of aberrations in gene dosage (the number of copies of a given gene present in a cell or nucleus) as a disease mechanism opens the way to new, more easily developed approaches to treatment in which the goal is not to correct abnormal or mutant proteins but instead to modify their abnormal dosage.

The finding of a high degree of variation in copy number in the human genome will also change our investigations into the causes of human genetic diseases. From now on, all genetic linkage and association studies should incorporate an evaluation of the variations in copy number in the study population to determine wheth-

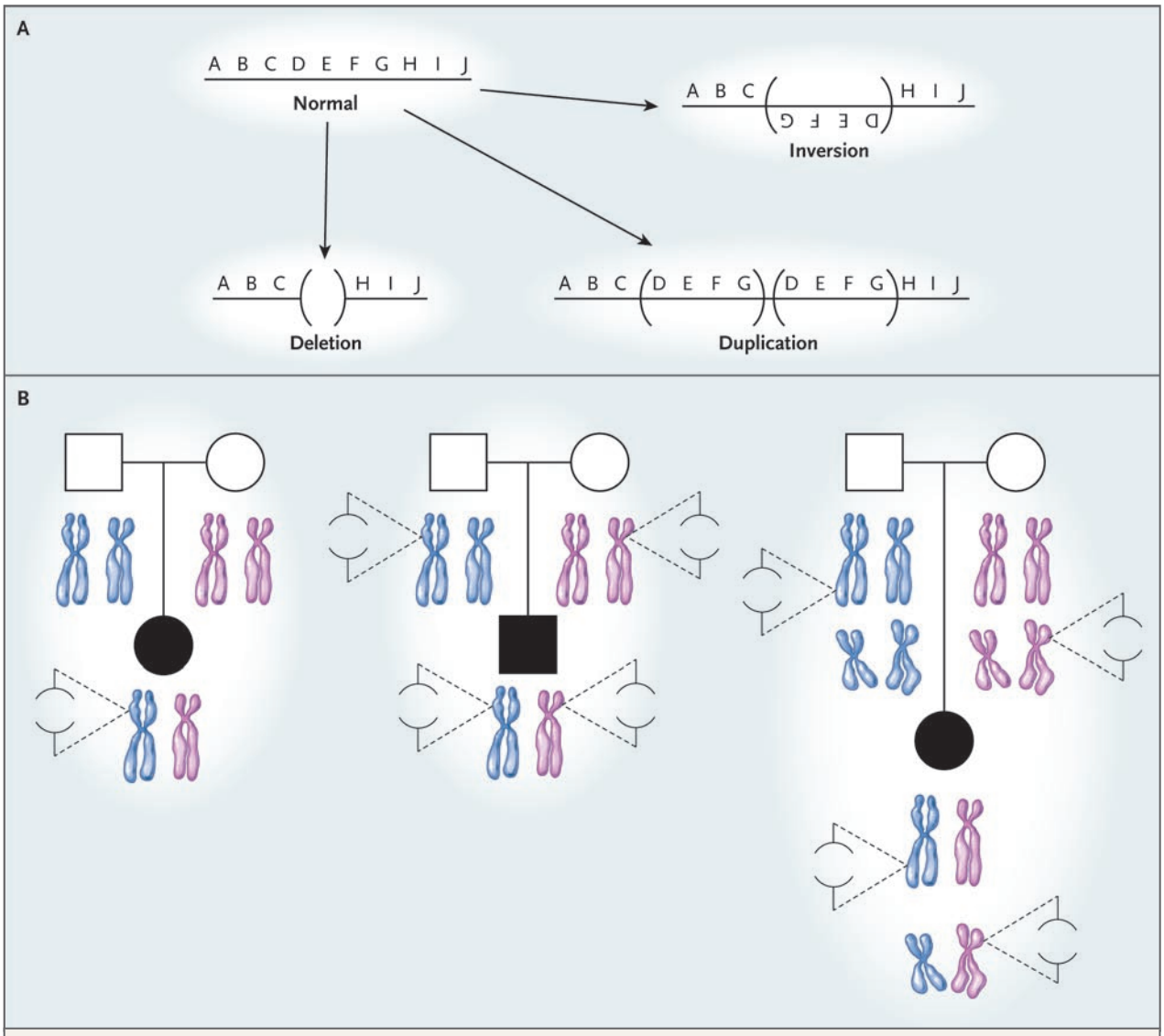

Figure 1. Variation in Copy Number and Sporadic Disease.

Panel A depicts structural variation of the human genome. Horizontal lines with letters above them represent the genome. Only one chromosome is shown. Deletions and duplications result in variations in the copy number in the sequences contained within the rearrangements. Panel B shows three potential mechanisms by which variations in copy number or combinations of such variations result in a sporadic phenotype. Squares represent males, circles females, and solid symbols persons with the disease phenotype. Below the pedigrees, the paternal (blue) and maternal (pink) chromosomes are shown. Dashed lines represent the enlargement of submicroscopical regions, vertical lines the genome, and parentheses flanking a deleted genome an interval causing a variation in copy number. In the first case in this panel (left), a new dominant deletion has occurred on the paternally inherited chromosome; new rearrangement mutations might cause sporadic traits more frequently than do new point mutations. In the second case (center), a deletion at the same locus is inherited recessively from each carrier parent. In the third case (right), the inheritance of the deletion of two different loci on two separate chromosomes is shown.

er an individual variation in copy number, rather than a single-nucleotide polymorphism, might be responsible for the trait being investigated. Furthermore, sporadic disease might be found to result from a new genomic alteration that causes variation in the copy number or from a combination of two or more variations in copy number inherited from two parents, in each of whom the uncombined variation did not provide a genetic burden that was great enough to cause disease

(Fig. 1). One might even speculate that variation in copy number underlies common normal traits such as those involved in behavior.

Dr. Lupski reports receiving consulting fees from Athena Diagnostics. No other potential conflict of interest relevant to this article was reported.

From the Departments of Molecular and Human Genetics and Pediatrics, Baylor College of Medicine and Texas Children's Hospital, Houston.

1. Shaw-Smith C, Redon R, Rickman L, et al. Microarray based comparative genomic hybridisation (array-CGH) detects submicroscopic chromosomal deletions and duplications in patients with learning disability/mental retardation and dysmorphic features. J Med Genet 2004;41:241-8.

2. Iafrate AJ, Feuk L, Rivera MN, et al. Detection of large-scale variation in the human genome. Nat Genet 2004;36:949-51.

3. Sebat J, Lakshmi B, Troge J, et al. Large-scale copy number polymorphism in the human genome. Science 2004;305:525-8.

4. Lee JA, Lupski JR. Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. Neuron 2006;52:103-21.

5. Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. Nature 2006;444:444-54.