This Month in the *Journal*

Y-Chromosomal Variation in India, *by Sengupta et al.* (*p. 202*)

Examination of mitochondrial and Y-chromosomal DNA has yielded conflicting results regarding the origins of tribes and castes in South Asia. Whereas some studies have shown tribes and castes to have independent origins, other studies have suggested that the genetic variation is due to recent input by nomads from Central Asia. There has also been disagreement about the development of Dravidian languages in Asia, with some believing that the languages spread along with the spread of agriculture into South Asia from Central Asia and with others supporting a preagricultural origin in South Asia. Sengupta et al. address these issues by using increased molecular resolution to more carefully define Y-chromosome haplogroups in populations from well-defined ethnic groups in India, Pakistan, and East Asia. Their analysis of the distribution of Indian haplogroups supports hypotheses that nomads from Central Asia had little effect on the genetic differences seen in India. Haplotypic variation reveals that regional differentiation predates possible input from Central Asia. Additionally, examination of the predominant Dravidian haplogroups reveals that the language originated in South Asia and was not introduced along with agriculture from Central Asia.

SNPs in NAGNAG Splice Acceptors, *by Hiller et al.* (*p. 291*)

SNPs that affect splicing are increasingly being recognized as important factors in disease. Hiller et al. closely examine SNPs that specifically affect a certain type of splice acceptor, the NAGNAG tandem acceptor. NAGNAG splice-acceptor sites have been found in ~30% of human genes; in some cases, either AG can be used by splicing machinery, and alternative transcripts are produced. There is experimental evidence that these alternative transcripts can have different functions as well as different tissue-expression patterns. Here, the authors observe that, whereas cells expressing a NAGNAG allele generate two transcripts, cells in which a SNP changes one of the NAG sites express only one transcript. It is predicted that these splice-affecting SNPs can have an even larger effect if the change converts the second NAG to a new amino acid. In these cases, not only would splicing occur only at the first site, but also the expressed product would be different from either of the transcripts expressed from the original allele. Conversely, SNPs can also create a NAGNAG tandem acceptor so that alternative splicing at that site is introduced. Awareness of NAGNAG sites might aid in the recognition of SNPs that affect splicing.

Spectrum of* CHD7 *Mutations in CHARGE Syndrome, *by Lalani et al.* (*p. 303*)

CHARGE syndrome is a disorder comprising various combinations of congenital anomalies, including coloboma, choanal atresia or cleft lip and/or palate, cranial nerve dysfunction, distinct external ear malformations with temporal-bone anomalies, cardiovascular malformations, and hypogonadotropic hypogonadism with genitourinary anomalies. Recently, mutations in *CHD7,* the gene encoding chromodomain helicase DNA-binding protein, were found in a large number of individuals with CHARGE syndrome. Here, Lalani et al. examine a larger cohort of patients, to confirm the prevalence of *CHD7* mutations and also to search for any possible genotype-phenotype correlations. Whereas no relationship was found between the type of *CHD7* mutation patients carry and the phenotypes with which they present, significant differences are found between CHARGE patients with and without *CHD7* mutations. Those with mutations have a higher incidence of cardiovascular malformations, coloboma of the iris or of the retinal or optic nerve, and facial asymmetry. It is hypothesized that *CHD7* is involved in gene regulation through chromatin remodeling in these tissues.

LCT Mutated in Congenital Lactase Deficiency, *by Kuokkanen et al.* (*p. 339*)

Lactose intolerance, or adult-type hypolactasia, is a common enzyme deficiency in which lactase activity declines as an individual ages. Elsewhere, it was determined that lactose intolerance was associated with a mutation in the promoter region of the gene encoding lactase, *LCT.* There is a more severe form of this condition, congenital lactase deficiency (CLD), in which lactase activity is virtually absent in the intestine at birth, and infants suffer from watery diarrhea, dehydration, acidosis, and weight loss when fed breast milk or lactose-containing formula. Although the disease is not very common worldwide, its prevalence is increased in Finnish populations. Previous work had excluded *LCT* from the CLD locus, but Kuokkanen et al. performed higher-resolution mapping that suggests that *LCT* is, in fact, a candidate gene for CLD. By screening 32 patients from 24 families, the authors

Am. J. Hum. Genet. 2006;78:1–ii. 0002-9297/2006/7802-0001$15.00

found five different *LCT* mutations, with 90% of the disease chromosomes containing the nonsense mutation Y1390X that leads to nonsense-mediated decay of the truncated transcript. This mutation is termed "Fin_{major}," and analysis of individuals from different areas of Finland reveals significant carrier frequencies in populations from the later settlement regions.

Robust Genomic Control, *by Zheng et al.* (*p. 350*)

Case-control studies are a powerful way to test for association between a marker and a disease, but, when some statistical methods are used, the presence of cryptic population substructure can lead to false-positive results, or type I error. A variety of approaches have been developed to deal with this problem, and genomic control (GC) is one such method. Previous GC work has developed strategies that work best when the underlying genetic model is specified as dominant, recessive, or additive. But these parametric methods are not always ideal, since the mode of inheritance is often not known. Zheng et al. propose a modification of established statistics that allows for robust analyses, even when the genetic model is unknown. Simulation studies demonstrate that, although the parametric methods perform well on the models for which they were designed, they lose a great deal of power when used on data of a different mode of inheritance. The new robust GC method, on the other hand, has good power for all models. The authors then evaluate how each method handles type I error if population stratification is a factor. Although the traditional χ^2 test can be used when the underlying genetic model is unknown, type I error rates are inflated with population stratification. Here, the robust GC modifications keep the error rates within acceptable limits, even in the presence of population substructure.

This Month on the Cover

Also known as "photograph B51," the cover image is the x-ray diffraction pattern of the B form of DNA obtained by Rosalind Franklin in 1952. Franklin's x-ray diffraction data greatly helped James Watson and Francis Crick solve the three-dimensional helical structure of a DNA molecule in 1953. (Image reproduced with permission from Franklin R, Gosling R [1953] Acta Crystallog 6:673 and Franklin R, Gosling R [1953] Nature 171: 740.)

ROBIN E. WILLIAMSON
Deputy Editor