This Month in *The Journal*

Kathryn D. Bungartz[1] and Robin E. Williamson[2]

Multilocus Gene-Disease Meta-Analysis

Newcombe et al., page 567

Genetic-association studies are increasingly identifying variants that are associated with disease, but because of the small effects of most of these variants, large sample sizes are needed in order to have enough power to observe these effects. Meta-analyses in which data from individual studies are combined can be useful in this regard, but the combination of the data is not always straightforward. Often, the same SNPs are not genotyped in each study, and because only summary data are usually reported, imputation of unobserved SNPs is not possible. Recently, a method was developed that could incorporate summary data from multiple studies, irrespective of the SNPs tested, to determine whether a region was associated with continuous traits. Here, Newcombe et al. extend the Bayesian model to perform meta-analysis with summary SNP data used to identify associations between markers and binary outcomes. Their focus is on determining whether variants in *PDE4D* are associated with stroke. The association that was originally reported has not always been replicated, but interpretation of the results of the independent studies has been complicated by the variety of SNPs that have been analyzed. By using their method to combine the data from the reported studies, the authors conclude that there is no association between *PDE4D* and stroke.

The Diversity in 5140 Human Mitochondrial DNA Genomes

Pereira et al., page 628

The analysis of mtDNA sequence has been instrumental in studies of human evolution as well as in the identification of pathogenic mutations in certain diseases. The increasing ability to sequence the whole mtDNA genome has fortified such work, but the increase in data has generated issues with storage and analysis. What is the best way to ensure that databases are kept up to date, and what is the easiest way to incorporate all available data into current studies? Recently, an online tool, Mitomaster, was developed to use all the mtDNA sequences deposited into GenBank as a reference; this is advantageous because static lists of identified variants can quickly become out of date, but manipulating all of the data was still a cumbersome task. Pereira et al. now introduce their mtDNA-GeneSyn software, which enables the high-throughput comparisons of complete mtDNA sequences. The authors then use mtDNA-GeneSyn to comprehensively examine the diversity in the mtDNA sequences in GenBank. After examining the nucleotide substitutions in the protein-coding genes, in the control region, and in the tRNAs and rRNAs, Pereira et al. conclude that transitions are more frequent than transversions, changes that interconvert between certain amino acids are better tolerated than others, and the functional constraints of secondary structure influence diversity. This collection of data and this data-analysis tool will serve as valuable resources for evolutionary studies and for the assessment of the pathogenicity of variants.

Intracontinental Structure in Humans

Biswas et al., page 641

Principal-component analysis (PCA) is one approach used to study human population structure. Using mathematical calculations, PCA transforms correlated variables into uncorrelated variables called principal components (PCs) in order to detect structure in the relationships between variables. Most PCAs focus on the top axes of variation because the first PC is designed to account for as much variation in the data set as possible. In the application to population structure, the variation of concern is genotype, which may be determined by use of SNPs. Here, Biswas et al. apply PCA to roughly 650,000 SNPs genotyped in unrelated people from 52 different populations of seven distinct continents in order to better understand genome-wide patterns of human population structure. They find that although the first few PCs provide much useful information, substantial information is gained from the lower PCs; they find 18 significant PCs in their analysis. In addition to this characterization of intracontinental structure, the authors identify an expanded set of ancestry-informative markers that can be used to further delineate population structure. These specific SNPs correlated to PCs are then used to identify regions in the genome that drive fine-scale structure in human populations. Together, these findings indicate genetic drift as the primary force governing human population structure.

[1]Science Editor, *AJHG*; [2]Deputy Editor, *AJHG*
DOI 10.1016/j.ajhg.2009.04.019.

Linkage of Pelvic Floor Disorders

Allen-Brady et al., page 678

Disorders of the pelvic floor include pelvic organ prolapse, stress urinary incontinence, urge urinary incontinence, and hernias. These disorders are thought to be caused by damage to or weakness in pelvic floor musculature and/or connective tissue. In addition to leading to inconvenient and embarrassing situations, these disorders pose a major health risk to women of all ages, with having given birth (parous) only posing as one potential risk factor. Other known risk factors include smoking, aging, obesity, and chronic constipation. Heavy lifting poses another risk. However, these factors do not fully explain the incidence of pelvic floor disorders, and a genetic component is thought to be present. To explore possible genetic contributions to pelvic floor disorders, Allen-Brady and colleagues performed a genome-wide linkage analysis in a large set of affected sister pairs. Their findings indicate that a region on chromosome 9q21 is associated with pelvic floor disorders with another possible associated locus on 9q31. Several genes within these regions have known function within muscle, including *TLE1*, *TLE4*, *UBQLN1*, *MAK10*, and *GOLM1*. Thus, these genes are good candidates for future association studies for pelvic floor disorders. The work presented here brings us closer to understanding the genetic component of these prevalent conditions.

Massively Parallel Sequence in Gene Identification

Brkanac et al., page 692

When a limited number of family samples are used to perform linkage analysis, the loci identified are often quite large and contain a number of positional candidate genes that can be prohibitive to sequence by traditional methods. One approach that has been pursued to limit the amount of follow-up sequencing has been to prioritize candidates and start with the assessment of genes that are predicted to most likely be involved with the disease at hand. Although this has often been successful, there are many situations in which priority predictions were not able to correctly establish the best genes to sequence. Now that sequencing technology has improved to allow for high-throughput sequencing of large regions, it has become more possible to sequence en masse the genes that reside in a locus identified via linkage analysis. Here, Brkanac et al. are studying a family with autosomal-dominant sensory/motor neuropathy with ataxia (SMNA). They identify a locus with a significant LOD score, but are only able to narrow the region to 22 Mb. On the basis of gene annotation of that 22 Mb, the authors select 3.7 Mb of DNA to target with massively parallel sequencing. After sifting through the variants identified, the authors predict that a missense mutation in *IFRD1* is etiologic for SMNA in this family.