This Month in *The Journal*

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

Population Structure of Qatar

Hunter-Zinck et al., page 17

The Qatari Peninsula protrudes into the Persian Gulf and is part of the larger Arabian Peninsula located in Southwest Asia. Although expatriates make up a large proportion of the present-day Qatari population, Qatar natives have their origins in the Middle East. Qatari citizens can be divided into three primary populations: Arabs constitute the majority of the Qatari population and originate from the Arabian Peninsula, the Ajam are primarily a Farsi-speaking people that migrated to Qatar from Persia, and the Abd are the third main Qatari population and are descendants of African slaves from Zanzibar that migrated to Qatar via Oman. Each of these populations has distinct geographic, linguistic, and presumably genetic histories. Islam is the predominant religion of Qatar, with the majority of the population being Sunni Muslim. Although the practice of Islam has strict laws of marriage and prohibits marriage between very close blood relations, first-cousin marriages are permissible and are thought to have been common in the Qatari Peninsula. Here, Hunter-Zinck and colleagues perform an in-depth genetic analysis of the Qatari population and examine the imprint of consanguinity in the native populations. Through principal-components anlaysis, these authors are able to clearly distinguish the three Qatari populations and define the extent of linkage disequlibrium (LD) among the three. Although the LD structure reflects substantial consanguinity, the authors report exceptional variance in the runs of homozygosity within and between the Qatari populations. Their findings suggest that Qatar genetics may be useful in mapping genes associated with complex disorders.

Rearrangements at Common Fragile Sites

Mitsui et al., page 75

Throughout the genome, there are sites that are hotspots for chromosomal breaks and deletions. These common fragile sites (CFSs) are often sites of rearrangements in cancer cells, and disruption of genes at the sites can also cause Mendelian disease. There is evidence that delayed replication at CFSs plays a role in the increased frequency of breaks at the sites, but questions remain about the sequence motifs and mechanisms that contribute to the sensitive state. Here, Mitsui and others closely examine the sequence of the breakpoints of rearrangements involving CFSs in an effort to establish why these sites are so sensitive. The sites studied, those containing *PARK2* and *DMD*, are selected because disruptions are found not only in cancer cells, but also in the cells of patients with juvenile Parkinsonism and Duchenne and Becker muscular dystrophy, respectively. This allows the authors to look for similarities and differences between the somatic rearrangements of cancer cells and the germline disruptions of Mendelian disease. They find that the breakpoints in the regions are clustered together in the same region in both cell types, but that the breakpoints in the cancer cell lines are generally more varied than those in the germ cell lines. The authors are also able to evaluate how the presence of homology and repetitive sequence at the breakpoint might contribute to the fragility.

Exome Sequencing in Hearing Loss

Walsh et al., page 90

Linkage analysis is a historically common first step in mutation identification for rare diseases with Mendelian inheritance. The defined interval of homozygosity may subsequently be refined by including more affected families or by utilizing more markers. Even after the associated locus has been narrowed down, several to hundreds of genes are often candidates for harboring a pathologic mutation. Often, researchers use a method of gene prioritization and begin sequencing exons gene by gene until the causative mutation is found. If a group gets lucky, they may find a pathogenic mutation in the first gene sequenced; however, this seems to be the exception. Compound heterozygosity and digenic cases can further complicate such mutation analyses. In addition to being time consuming and frustrating, such mutation analyses are also costly. Whole-exome sequencing, in which the protein-coding portion of the genome is sequenced, is a recently developed technology that provides a quick and potentially cost-effective alternative to the traditional methods. Here, Walsh and colleagues use exome sequencing to analyze the DFNB82 locus associated with autosomal-recessive nonsyndromic hearing loss in a Palestinian family. After ruling out variants present in the dbSNP database and using additional processes of elimination, Walsh and colleagues successfully identify a single homozygous nonsense mutation in *GPSM2* that is found to segregate with hearing loss in the family and

[1]Science Editor, *AJHG*; [2]Deputy Editor, *AJHG*
DOI 10.1016/j.ajhg.2010.06.012.

is not found in control populations. This report indicates that exome sequencing is an effective means of identifying causative genetic mutations in large disease loci and will likely be utilized in future studies.

Spoiling the Whole Bunch

Pluzhnikov et al., page 123

In this age of genome-wide analysis, cases and controls are often shared between studies and research institutes across the world. Although this sharing is a positive and useful outcome of large case-control studies, it does present some complications in data interpretation. High-throughput genotyping consists of microarray platforms containing hundreds of variables to which individual DNA samples are added. Thus, each microarray plate contains a defined set of sample DNA consisting of cases and/or controls. A series of quality control (QC) steps are taken when the samples are prepared and the original genotyping performed. Although the QC used for the original study is often rigorous and appropriate, the original QC measures may not always be sufficient for subsequent analyses by different investigators. Here, Pluzhnikov and colleagues examine this issue. Using plates generated for the Genetics of Kidneys in Diabetes (GoKinD) study, these authors discover a set of individuals on one plate having miscalled relatedness. This leads to an allele-frequency bias on that plate that affects the results of the entire study. They then demonstrate and suggest methods for filtering such data. They recommend using the Affymetrix method of Contrast QC (CQC) and the Skewness QC (SQC) tests, but they note that these methods cannot be used to eliminate all plate effects. This report emphasizes that QC transparency is critical for data to be successfully utilized by the public, and the authors suggest randomizing cases and controls in initial studies to minimize plate bias. As the title indicates, one bias plate can be sufficient to spoil the whole association study.

Genome-wide Detection of Mosaic Variants

Rodriguez-Santiago et al., page 129

When a mutational event or chromosomal abnormality occurs during development, the alteration will be present only in a percentage of cells and the individual will be a mosaic; i.e., the person will have a population of cells with the original genotype and a second population of cells with the new genotype. If the disruption occurs early in development, a large number of tissues and organs can end up with the abnormality, whereas if the event occurs later in development, the new genotype can be restricted to isolated tissues. Detection of mosaicism is usually performed by techniques that examine full-chromosome imbalances to identify mosaic aneuploidies, but recent evidence has suggested that copy-number variants (CNVs) can also occur postzygotically. In addition, individuals can be mosaic for uniparental disomy (UPD), a state in which both copies of a chromosome or a segment of a chromosome have been inherited from a single parent. Containing a proportion of cells with such structural events has most often been linked to disease phenotypes, but Rodriguez-Santiago and colleagues find that mosaicism for a number of chromosomal changes can be present in individuals with no overt clinical phenotype. In this issue, the authors use SNP arrays to perform genome-wide searches for mosaicism in their samples. They find individuals to be mosaic for UPD regions, large CNVs, trisomies, and even more complex rearrangements. In addition to causing disease in certain cases, mosaicism for structural events is most likely also a source of genetic variation in normal individuals.