This Month in *The Journal*

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

MtDNA Mutation-Level Variance

Wonnapinij et al., page 540

Heteroplasmy most often refers to a single cell that contains more than one type of mtDNA. In the case of inherited mtDNA mutations, different cells and tissues throughout the body vary in their degree of heteroplasmy. For example, the load of mutant mtDNA may be great in hematopoietic cells but nearly absent in skeletal muscle. Likewise, some hematopoietic cells may have an abundance of mutant mtDNA, whereas others may have less or even lack mutant mtDNA. This variability is also true for cells in the germ line, leading to a great deal of variance in the degree of inherited heteroplasmy, even in siblings with the same affected mother. To date, this variance has made it difficult for physicians and genetics counselors to predict the likelihood of mitochondrial disease transmission. Here, Wonnapinij and colleagues introduce a method for determining the standard error of variance and determine the necessary number of cells in which heteroplasmy is measured to accurately estimate mtDNA mutation-level variance. They find that most studies are underpowered to draw reliable conclusions regarding the mtDNA mutation level because not enough measurements of heteroplasmy are made. They suggest that more than 30 measurements are required for detecting a 2-fold (or greater) difference in variance for mutation levels near 50%, and they recommend 50 or more measurements for detecting mutation levels on both the low (10%) and high (90%) ends of the spectrum. Furthermore, these authors find the variance of heteroplasmy in humans to be greater than that in mice. These results can be used to shed further light on the size and timing of the mtDNA bottleneck.

Mutations in *FKBP10*, which Encodes FKBP65, Cause Osteogenesis Imperfecta

Alanay et al., page 551

The brittle-bone disease osteogenesis imperfecta (OI) is caused by the disruption of the process in which type I procollagen is modified and converted to type I collagen, which is then deposited into the extracellular matrix. The disease is very heterogeneous, both in terms of phenotypic severity and in the types of genes in which mutations are etiologic. Although the majority of OI cases are due to mutations in the genes encoding type I procollagen, dysfunctional chaperon proteins and proteins involved in modification of the collagen chains can also lead to disease. In this issue, Alanay and colleagues report the results of their gene-mapping study in families with a severe and progressively deforming type of OI. The authors identify mutations in *FKBP10*, the gene encoding FKBP65, in their families. FKBP65 is a member of a family of proteins with chaperon function, and the protein was previously shown to interact with type I procollagen. Functional work with patient samples demonstrates that the mutations do not affect posttranslational processing of type I procollagens and collagens but result in a defect in protein secretion.

Genetic-Association Studies Using Electronic Medical Record Data

Ritchie et al., page 560

It is well accepted that using larger sample sizes in association studies will yield more power, but collecting additional cases and controls is not a trivial task. One possible source of such data may be DNA databanks that are linked to electronic medical record (EMR) systems. EMRs were created to improve patient care through improving the accessibility and organization of vast amounts of patient data; therefore, these EMRs potentially contain the information necessary to designate whether someone could be a case or control in a study. If the EMRs were then connected to a biorepository containing DNA that could be genotyped, these samples could serve as members of association-study data sets. A first step in the development of such a resource is establishing whether the phenotype-data contained in the EMRs are comprehensive and specific enough for genetic-association analyses. In this issue, Ritchie and colleagues evaluate whether validated genetic associations can be identified in the samples from a DNA databank linked to EMR data. They genotype their samples for variants known to be associated with atrial fibrillation, Crohn disease, multiple sclerosis, rheumatoid arthritis, and type 2 diabetes. Cases and controls for each disease are then selected on the basis of classifications used in the EMRs. The authors report that, of the associations for which they had enough power to observe

[1]Science Editor, *AJHG*; [2]Deputy Editor, *AJHG*
DOI 10.1016/j.ajhg.2010.03.007.

a significant signal, the majority of known associations are replicated. These results suggest that the data from EMRs and biorepositories might serve as a valuable resource for genetic-association studies.

Integrating GWAS Pathway Analyses with Genetics of Gene Expression Data

Zhong et al., page 581

Genetics of gene expression (GGE) studies are studies in which SNPs associated with gene expression are identified. These expression SNPs, or eSNPs, might be found within or near the gene whose expression level they are modifying, or they could affect gene expression from afar via more indirect mechanisms. Either way, these eSNPs are likely to be biologically relevant and are potentially strong candidates for association with disease. In this issue, Zhong and colleagues integrate GGE data into GWAS pathway analyses to identify variants that are associated with type 2 diabetes (T2D). They start by creating a list of genes whose expression levels are affected by eSNPs. These genes are assigned to pathways, and the eSNPs that are associated with the expression of the genes are grouped together. These groups of eSNPs are then evaluated for determining whether they are associated with T2D. The idea is that, if they are, the gene pathway that they represent is likely relevant to disease etiology. By analyzing data from the Wellcome Trust Case Control Cohort and the Diabetes Genetics Replication and Meta-analysis Consortium, the authors provide support for pathways that have been previously implicated in T2D and identify additional pathways whose involvement in the disease has yet to be reported.

Gender-Biased Admixture in the SAC Population

Quintana-Murci et al., page 611

South Africa comprises separate groups of people having distinct geographic and genetic histories. Nearly 80% of the South African population is black and represents a multitude of ethnicities speaking a number of different languages. Around 10% of the South African population consists of white individuals. This group has a predominantly European background derived from settlers of the 1600s, and coloured individuals make up around 9% of the South African population. The South African Coloured (SAC) population is thought to be derived from these early European settlers, their slaves, and native people of South Africa, which make up the remaining 1%–2% of the South African population. The SAC population has among the highest degree of recent admixture. Here, using markers from mtDNA and the nonrecombining portion of the Y chromosome (NRY), Quintana-Murci and colleagues examine the maternal and paternal components of the SAC. Their findings reveal a strong gender bias in SAC admixture, with very little mtDNA derived from Europeans but most of the NRY markers contributed by Europeans. These data indicate the ancestry of the SAC population is largely derived from English men and African and Southeast Asian women. This genetic data fits well with historical accounts of the origins of contemporary South Africa. History tells us that in 1652, a post was established at Table Bay to provide refreshments for sailors of the Dutch East India Company and that Cape Town soon became an outstation of this company. This led to European settlers, predominantly men, who brought along slaves, often from Southeast Asia. Combining such historical and genetic data helps to promote an understanding of the diversity and dynamics of admixed populations.