This Month in *The Journal*

Sarah Ratzel[1] and Sara B. Cullinan[2]

Exome Sequencing for Identifying and Imputing Rare Variants in a Large Population

Auer et al., page 794

Exome sequencing in small cohorts has been successfully employed for the discovery of rare, highly penetrant mutations in familial diseases. However, the use of exome sequencing for identifying low-frequency variants in large populations has been confounded by technical limitations and prohibitive costs. To apply low-frequency variants captured by exome sequencing to a large population in African Americans, Auer et al. exome sequenced a reference panel of 761 individuals and then imputed these variants into a cohort of 13,000 individuals. After imputation, a series of blood traits, including hemoglobin, hematocrit, white blood cell, and platelet counts, were tested for associations with the identified variants. Some of the interesting associations included those between variants in *MPL* (the thrombopoietin receptor) and high platelet count, between *LCT* and high white blood cell count, between *CD36* and low platelet count, and between several variants in alpha-globin and low hemoglobin count. Leveraging rare variants identified in exome sequencing to a larger population uncovered associations that have been missed by traditional genome-wide association studies, thus highlighting some of the inherent limitations of genotyping arrays. This work not only explains some of the heterogeneity in blood cell traits observed in African Americans but also demonstrates the feasibility of using this approach in other studies aimed at better understanding the role of low-frequency variants in populations.

Revealing Population Demographics

Palamara et al., page 809

A population's history can be derived, to a certain extent, from the genetic footprint left by migrations, admixture, and population expansions and contractions. Many times, high-resolution genomic data from large cohorts is used for studying genomic areas that contain long-range segments that are identical by descent and thus share a genetic history. However, lower-resolution haplotype data could represent a generally untapped wealth of information because it avoids the pitfalls of allele-frequency-spectrum analyses, which ignore subtle correlations across pairs of individuals, and of linkage-disequilibrium analysis, which relies on very strong correlations. In this issue, Palamara et al. develop a method to model the changes in a population on the basis of shared long-range haplotypes in unrelated individuals in relatively recently established populations. Their analysis of a cohort of 500 Ashkenazi Jews reveals evidence of two population expansions, separated by a founder event, and is consistent with historical and genetic evidence. Additionally, patterns of an outbred population were uncovered in 56 Maasai, in whom a high level of cryptic relatedness indicated the genetic mixing of small, migratory groups. The proposed model can be used in conjunction with several existing SNP data sets, suggesting that it could have widespread utility. Furthermore, as the sequencing of populations becomes more common and as additional population-specific variants are identified, the accuracy of identity-by-descent detection could be improved such that this analysis could be extended further into the past.

Whose DNA Is It, Anyway?

Jun et al., page 839

Discoveries based upon genotyping and next-generation sequencing have become everyday occurrences, so it can be easy to forget all of the quality-control procedures that are needed for obtaining high-quality genetic data. One potential source of error is contamination from foreign DNA. Although it is relatively easy to filter out cross-species contamination, the identification of within-species contamination, especially in low-pass sequencing studies, relies upon the detection of excess heterozygous genotypes and has proven to be much more difficult. Countless hours and dollars have been wasted in the chase for spurious results that stem from contaminated samples. In light of these concerns, Jun et al. now describe new methods for detecting DNA contamination, even at levels below 1%. By enabling the detection of low-level contamination, the authors' methods make it possible to filter out the "junk" prior to downstream analyses, i.e., variant calling in the case of sequencing studies and follow-up sequencing in the case of array-based genotyping studies. Both clinicians and basic researchers should be able to easily integrate these methods into their current workflows, potentially streamlining and reducing the cost of their analyses.

[1]Scientific Editor, *AJHG*; [2]Deputy Editor, *AJHG*
http://dx.doi.org/10.1016/j.ajhg.2012.10.004.

Y Spermatogenesis Fails

Rozen et al., page 890

Severe spermatogenic failure (SSF) affects nearly 2% of all men. The most common causes for this type of infertility are deletions involving the AZFc region, a structurally complex portion of the Y chromosome. To better understand the relative contribution of the different deletions to SFF, Rozen et al. assessed the prevalence of six recurrent AZFc deletions in over 20,000 men across five populations. They found that a rare variant (b2/b4 deletion) that increases risk by a factor of 145 and a common variant (gr/gr deletion) that doubles one's risk account for the largest proportion of SSF across populations. Although this information will no doubt be of great help to clinicians, the scope of the study should also attract attention from population geneticists. Indeed, many studies have shown that structural mutations underlie several human diseases, but much more work is needed. Most notably, we are lacking a clear understanding of which regions are prone to structural change. These findings should provide a launching pad for researchers interested in assessing mutability of structurally complex regions in a genome-wide fashion. Given that many complex regions are not well annotated, the methodology developed for studying the AZFc region might be adapted for better characterizing structural complexity on a genome-wide scale.

Mutations in *COX7B* Cause MLS

Indrieri et al., page 942

Mitochondria are essential for many reasons, including the generation of ATP by oxidative phosphorylation. Mutations in either nuclear- or mitochondria-encoded genes can disrupt this process and lead to a wide range of defects. One example of an X-linked mitochondrial disease is microphthalmia with linear skin lesions (MLS), in which affected females display linear skin lesions, neurological defects, and eye deficiencies and in which defects in males are lethal. Although this disease has been associated with mutations in *HCCS*, which encodes a component of the mitochondrial respiratory chain, not all individuals with MLS harbor *HCCS* mutations. Indrieri et al. suspected that screening for mutations in other X chromosome genes encoding mitochondrial proteins might reveal additional candidates. Indeed, mutations in *COX7B*, encoding a subunit of cytochrome C oxidase (COX), were associated with MLS in two unrelated families. These mutations reduced mitochondrial respiration by disrupting COX assembly. Furthermore, MLS phenotypes, including microcephaly and microphthalmia, were recapitulated in medaka when *COX7B* levels were decreased, suggesting that *COX7B* is important for neurological development across vertebrates.