This Month in *The Journal*

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

Meta-analysis Methods for RVASs

Tang and Lin, page 35

Efforts aimed at gaining insight into the genetic architecture of complex-trait disorders turn, with increasing frequency, to rare-variant association studies (RVASs). Indeed, in recent years, many methods have been developed to detect rare variants that influence human traits. The increase in large-scale sequencing studies suggests that like genome-wide association studies (GWASs), which study the association between common variants and complex phenotypes, successful implementation of RVASs might rely on the integration of data obtained from multiple groups. However, whereas GWASs rely on the genotyping of common variants, a consensus has yet to emerge regarding the best way to perform meta-analysis in the context of RVASs. When embarking on an RVAS meta-analysis, one must consider several key points—points that differ from those routinely implemented in GWAS meta-analyses. In this issue, Tang and Lin provide an overview of these considerations and also describe PreMeta, a unified framework that integrates the most popular RVAS meta-analysis packages. Notably, PreMeta makes it possible for investigators to combine summary statistics generated across different platforms. Such a tool should be of great help for large, multinational consortia, whose initial statistical analyses are often performed on different platforms.

Deletion Might Explain High-Altitude Adaptation in Tibetans

Lou et al., page 54

Several studies of high-altitude adaptation have implicated *EPAS1* in conferring adaptation to low oxygen conditions; however, the mechanism by which variation in this region contributes to fitness is unclear. In this issue, Lou et al. used WinXPCNVer to detect a previously identified, but not well-characterized, deletion in a Tibetan population living at high altitude. Follow-up deep sequencing in a second Tibetan population ruled out additional structural variation at this locus. Because the deletion was identified in 90% of Tibetans and only 3% of other worldwide populations, it is referred to as a Tibetan-enriched deletion (TED). Testing based on extended haplotype homozygosity supports the idea that the TED might underlie the signal of positive selection previously observed in the region. Indeed, the TED lies near *EPAS1* and is in linkage disequilibrium (LD) with variants influencing hemoglobin concentrations. Interestingly, the TED is in complete LD with an introgressed Denisovan SNP motif. The TED itself is absent from the Denisovan genome, however, suggesting that the deletion was acquired at some point after introgression. The TED is estimated to have originated 13,000 years ago—the same time at which the Tibetan population underwent an expansion. Future studies to identify the potential functional consequences of the TED will help to determine whether the deletion itself confers a selective advantage or whether it is in LD with a functional variant.

When Did that Mutation Occur?

Acuna-Hidalgo et al., page 67

Many researchers undertake a search for de novo mutations when they wish to discover the genetic cause of a disease that appears not to run in families. This approach, most often a trio analysis focusing on the unaffected parents and the affected child, has proven successful in both small- and large-scale sequencing projects. Although some rare diseases, for example, overgrowth conditions such as Proteus syndrome, have been attributed to somatic mutations affecting only a subset of an individual's cells, the underlying assumption is that most deleterious mutations arise during gametogenesis and are present in all cells. New findings from Acuna-Hidalgo et al. challenge this paradigm and instead suggest that many apparently de novo mutations actually represent somatic mosaicism; that is, the mutations arise during early development or are passed along from a parent who harbors the mutation in just a subset of their cells (and who might therefore be unaffected or experience only mild symptoms). Similar to recent findings related to somatic mosaicism for copy-number variation, these new results serve as an important reminder that one's genome is not identical across all cells in the body. Indeed, the authors estimate that over 6% of an individual's genetic variation stems from somatic mosaicism. If confirmed, these findings could have far-reaching implications for the ways in which disease risk is estimated and prospective parents are counseled.

[1]Scientific Editor, AJHG; [2]Deputy Editor, AJHG
http://dx.doi.org/10.1016/j.ajhg.2015.06.006.

A *TWIST2* the Story

Marchegiani et al., page 99

In this issue, Marchegiani et al. provide evidence that two congenital ectodermal dysplasias, ablepharon-macrostomia syndrome (AMS) and Barber-Say syndrome (BSS), share more than just overlapping clinical features. Through careful phenotyping and sequencing, the authors demonstrated that recurrent de novo mutations in *TWIST2* cause both disorders. Interestingly, all identified mutations affect the same amino acid in TWIST2; however, whereas a p.Glu75Lys substitution was identified in all of the study's AMS-affected individuals, a p.Glu75Gln change was found in all individuals with BSS. *TWIST2* encodes a transcription factor whose activity has been implicated in diverse developmental processes, including stem cell differentiation. The authors' work in zebrafish and cell-culture models supports the idea that the altered TWIST2 proteins could function in either a dominant-negative or a neomorphic fashion. It is possible that the mutants differentially titrate the wild-type TWIST2 away from its heterodimeric binding partners, perhaps enacting distinct downstream transcriptional programs. Further experimentation will be necessary for testing these ideas and reaching an understanding of exactly how the two substitutions are able to lead to such discrete outcomes; the answers should provide insight into the genetic underpinnings of AMS and BBS, as well as normal TWIST2 function.

Human Phenotype Ontology: Now Taking On Common Disease

Groza et al., page 111

The Human Phenotype Ontology (HPO) has been widely used in the rare disease community to aid in diagnosis and identification of disease-associated candidate genes by drawing upon more than 11,000 terms to describe phenotypes that might be encountered in a clinical evaluation. However, the HPO was not optimized for common disease, and the utility for such an application has been unclear. Now, Groza et al. have expanded the HPO by including information mined from over five million abstracts, generating over 130,000 annotations for 3,145 common diseases. After they incorporated these new annotations, a highly significant phenotype network showed dense clusters and patterns that reflect the phenotypic continuum, supporting the use of the data-mining approach for producing the database. In a demonstration of how the HPO can be used, the authors investigated the similarities between the networks of phenotypes to look for shared underlying etiologies, and they identified overlaps between common and rare or Mendelian conditions. Ongoing curation of the HPO to improve annotations and phenotypes will enhance its value as a bioinformatics tool for common disease, in addition to its established role in the rare-disease community.