

This Month in *The Journal*

Sarah Ratzel¹ and Sara B. Cullinan²

Improving Heritability Estimates with LDAK

Speed et al., page 1011

The use of SNP-based genotyping for estimating heritability in unrelated individuals has garnered significant interest recently as a way to help explain the “missing heritability” of many phenotypes. Many believe that “missing heritability” can be found in either rare variants with modest effects or common variants with weak effects. For evaluating the idea that many common variants with weak effects contribute to a complex phenotype, SNP-based mixed-model analysis is widely used. In this paper, Speed et al. use several simulations to find that, although this method is very robust to perturbations in some underlying assumptions, it is highly sensitive to linkage disequilibrium (LD). For example, estimated heritability can be too low for low-frequency causal variants in low-LD regions and too high for high-frequency variants in high-LD regions. To correct for this bias, the authors developed Linkage-Disequilibrium-Adjusted Kinship (LDAK), a software program that modifies the kinship matrix for local patterns of LD by equalizing the level of SNP tagging on the basis of how well a SNP tags its neighbor. With the use of data from the Wellcome Trust Case Control Consortium and LDAK, heritability estimates were better for several traits. The estimates improved by approximately 25% for hypertension and type 2 diabetes, suggesting that the causal variants are not well tagged for these diseases. Conversely, the estimation of heritability was reduced for diseases caused by well-tagged variants in high LD, e.g., rheumatoid arthritis and type 1 diabetes, which are both associated with variants in the major-histocompatibility-complex region. The incorporation of LDAK should improve SNP-based genomic partitioning methods currently in use and might also accelerate association analysis and phenotype prediction.

Cycles of Variation

Koren et al., page 1033

The amount of human genetic diversity is immense, and we are just now beginning to understand how and when changes arise. The need for cells—both germ and somatic—to replicate their DNA and divide provides numerous opportunities for errors to be introduced into the genome. The introduction of errors is not random;

previous work has shown that early-replicating genomic areas, where most protein-coding genes lie, accumulate fewer mutations than do early-replicating regions. In this issue, Koren et al. extend these analyses and provide an in-depth view of the intersection between human genome replication, mutation, and recombination. Notably, the authors show that transversion and transition mutations are the most prevalent mutation type at late-replicating loci. Further work will be needed for identifying the mechanisms by which this type of biochemical specificity is achieved; indeed, this study highlights the need for cross-disciplinary collaboration in tackling tough biological problems. The authors also show that recombination and replication timing are more tightly associated in males than in females. The underlying cause for this difference remains unknown and will probably require technological innovation so that it is addressed properly, but for now, many interesting thought experiments will occupy the readers' minds.

Are They Really Pathogenic?

Xue et al., page 1022

Accurate annotation of disease-causing mutations (DMs) in databases such as the Human Gene Mutation Database (HGMD) is imperative for the development of personalized medicine and diagnosis of genetic disease. When faced with a lack of additional evidence, investigators must carefully interpret variant pathogenicity when evaluating missense alleles that are predicted to be damaging. In this study, Xue et al. analyzed 1000 Genomes Pilot Project sequence to estimate the average number of predicted harmful missense mutations in healthy individuals and to compare these variants to DMs annotated in HGMD. Their work shows that healthy individuals can harbor 281–515 missense substitutions (40–85 of which are homozygous) that are predicted to be damaging. On average, each individual had 40–110 heterozygous variants and 3–24 homozygous variants that are annotated in HGMD as being DMs. Further analysis of 71 DMs suggested that 3 are functional polymorphisms and 22 are not pathogenic, although reduced penetrance might explain the lack of phenotype in some cases. With the increasing number and diversity of genomes being sequenced, the identification of variants predicted as being DMs in healthy populations could prompt the

¹Scientific Editor, *AJHG*; ²Deputy Editor, *AJHG*

<http://dx.doi.org/10.1016/j.ajhg.2012.11.005>. ©2012 by The American Society of Human Genetics. All rights reserved.

continual refining of annotations in databases such as HGMD. Although cumbersome, persistent curation of these databases will no doubt improve genetic testing and treatment.

New Insights into HSPs

Oz-Levi et al., page 1065; Tesson et al., page 1051; Schuurs-Hoeijmakers et al., page 1122

Hereditary spastic paraplegias (HSPs) are one of the most heterogeneous groups of inherited disease. Characterized by progressive spasticity of the lower limbs, HSPs can be further divided by the presence or absence of additional neurological and/or extraneurological phenotypes. The era of exome sequencing has enabled the molecular diagnoses of many previously uncharacterized forms of HSP, and this issue contains three additional reports of mutations that lead to HSP. Although the mutations lie in different genes, together, the papers highlight the interconnected nature of the genetic defects that underlie these complex disorders. By identifying a mutation in *TECPR2*, Oz-Levi et al. reveal a connection between HSPs and autophagy. Tesson et al. and Schuurs-Hoeijmakers et al. identify in two intracellular phospholipases A₁ family members, *DDHD1* and *DDHD2*, mutations causing HSP subtypes that are linked by the presence of defects in lipid metabolism. Moreover, Tesson et al. also show that mutations in *CYP2U1*, which encodes a cytochrome P450 enzyme, cause a distinct form of HSP. Taken together, these papers uncover the importance of fatty-acid and phospholipid metabolism in neurological health. The connections revealed in these papers suggest that additional HSPs, as well as other neurological disorders, are

caused by disruptions in this vast network of cellular metabolic control.

lincing RNA to Human Disease

Talkowski et al., page 1128

Recent years have seen great advances in the understanding of how gene mutations and disruptions cause human disease. Yet, as we all know, protein-coding sequences constitute just a small portion of the genome. Until whole-genome sequencing becomes commonplace, our comprehension of how the remaining portions of the genome contribute to disease will probably lag behind. In this issue, Talkowski et al. provide a glimpse of the insights that we might hope to glean in the coming years. They first identified an individual who is affected by a neurodevelopmental abnormality and who harbors a balanced chromosomal translocation only disrupting *LINC00299*, a gene whose RNA product is a large intergenic noncoding RNA (lincRNA). Follow-up work identified copy-number variants affecting *LINC00299* in several individuals with neurodevelopmental phenotypes. Interestingly, *LINC00299* expression levels increase during the course of neural-progenitor differentiation, thus providing some insight into the biology that underlies these clinical manifestations. lincRNAs, a newly identified class of RNA, are thought to regulate gene transcription, but the number of characterized family members remains low. In pointing to a role for a lincRNA in human neurodevelopment, this study highlights just how little we know about the genome and provides hints of exciting findings that will no doubt be made in the quest for understanding the molecular bases of human disease.