nature genetics

What is the Human Variome Project?

The successor to the Human Genome Project intends to establish, by international cooperation, an encyclopedic catalog of sequence variants indexed to the human genome sequence.

Genomics is not just for rich countries any more. Anyone can contribute to the Human Variome Project (HVP; see Commentary, page 433). Indeed, the project might just be ambitious enough that everyone really will need to contribute. By stating that all human genetics and genomics contributes to a single aim, the HVP essentially reduces duplication of effort while increasing credit for participation.

However, it will have to find ways to coordinate the disparate activities of clinicians, researchers, database curators and bioinformaticians by providing the means and incentives to lodge the variants they have found in public databases. Variome aims to get all to use compatible nomenclature and phenotype reporting systems and to index variant and phenotype data to gene models in the coordinate system generated by the Human Genome Project. Automation and expert curation, and open comment and expert review, will all have a place in this endeavor. How will we do this without creating more than a necessary minimum of new databases, procedures and bureaucracy?

A very important point, but a tough one to get across, is that much of the necessary work is currently happening across the globe—but is just insufficiently coordinated. The individuals already hard at work aren't getting the credit they deserve. In a sense, the rest of the world's geneticists deserve the kind of service that US researchers receive from the excellent coordinating work of the National Human Genome Research Institute and the repositories of the National Center for Biotechnology Information (NCBI), together with the kind of attention afforded by international journals. If only these kinds of coordination, recording and attention could be brought to bear, however briefly, on publication units as small as single instances of a variant gene! Thus, Variome aims to add value to databases such as OMIM, GenBank, dbSNP, dbGAP and the HapMap and organizations including NCBI and the European Bioinformatics Institute (EBI) by working with them all. It will start gene by gene, evaluating variants already found and curated for mendelian diseases, and will add rare and common variants in common diseases as they are reported. As it does so, HVP participants will develop mechanisms to expedite and automate reporting of variants and their occurrence.

In the consensus-building exercise of the first Human Variome meeting (page 433), delegates constructed a wish list of recommendations that numerically exceeded the number of participants at the meeting. We think that two points emerge as particularly important to the success of the project: publication and credit.

To be successful in persuading clinical and diagnostic laboratories to contribute variations and persuading researchers to evaluate the pathogenic potential of each variant, the HVP will need to introduce publishing innovations at both ends of the citation spectrum. It will need to track the citation of each variant's accession code in papers, database entries and across the web. This closing of the online publication loop might be termed microattribution. Perhaps existing journals could be persuaded to take responsibility for monitoring and highlighting the citation of database entries in their papers, so that the HVP can readily aggregate this information. A journal devoted to the human variome could commission peer-reviewed, gene-based synopses of mendelian mutations based on information in locus-specific databases (see pages 425 and 427), meta-analyses of association studies and resequencing data such as those reported by Jonathan Cohen and colleagues in this issue (page 513, with News and Views on page 439). Phenotypic and diagnostic information might be linked to these synopses from existing databases such as the dysmorphology databases, PharmGKB (page 426) and GeneTests (http://www.genetests.org). Genome browsers including Ensembl and UCSC might then be persuaded to display a Variome track. We envisage such synopses to be a gene-based extension of the disease-based annual synopses for association studies we proposed last year (*Nat. Genet.* **38**, 1; 2006). The first of these, on Alzheimer disease, was published by Lars Bertram and colleagues (*Nat. Genet.* **39**, 17–23; 2007) using their newly created AlzGene database.

Which genes should the HVP annotate first to demonstrate the utility and impact of its coordinating activities? Perhaps we can learn from one of the most impressive recent exercises in evidence-based medicine: namely, the American College of Medical Genetics' systematic prioritization of genes for newborn screening (http://mchb.hrsa.gov/screening/). Variome synopses would take into account the prevalence, seriousness and treatability of the clinical condition(s), the value added by combining all three types of genetic study listed above and the availability of all three kinds of evidence in existing laboratories, databases and publications.

There are, inevitably, limits to what can be achieved by a gene-based view of human variation. Gene models are revised and re-annotated, and structural genomic variation plays havoc with reference genome builds and the context within which point variants and haplotypes are found. Physicians and the general public will want a disease-based view—and the associated diagnostic genetic tests, rather than genome annotation. Delaying the appearance of such alternative views, there is often a many-to-many correspondence between genes and disease phenotypes. On the brighter side, this complexity should provide good business for database designers and review journals.

As the participants of the Variome meeting note in their Commentary, the effort to index and evaluate all of human variation will provide many new opportunities in genomics for researchers whose home countries did not participate in the initial human genome sequencing project. They are right that this is both the project and the time to achieve the globalization of genomics.