

## Conventional wisdom

Recent agreement on stable reference sequences for reporting human genetic variants now allows us to mandate the use of the allele naming conventions developed by the Human Genome Variation Society.

By agreement between stakeholders and two principal databases, it has been proposed (R. Dalglish *et al.*, *Genome Med.* 2, 24, 2010, doi:10.1186/gm145) that human genetic variants be reported relative to a new set of stable reference sequences, “Locus Reference, Genomic” (LRG, pronounced “large” <http://www.lrg-sequence.org/page.php>). These sequences have been developed from the initial NCBI RefSeqGene concept and are provided by NCBI and EBI according to agreed rules and in consultation with community users of locus-specific genetic information and locus-specific databases. It is anticipated that the LRG will be stable and supported for many years, long enough to serve as a bridge between existing and future clinical gene tests.

The core nucleotide sequence, principal transcripts, exons and open reading frames will be presented as a fixed layer and will not be subject to revision (‘versioning’) of the reference sequence because this practice was previously a source of confusion in the literature. There will also be an updatable layer to the LRG construct that will provide the coordinates needed to map variants to future updates of the human reference genome sequence as well as to legacy coordinate systems.

The LRG conventions are not intended to replace the—universal and, in our view, superior—coordinates of reference genomes, rather to provide a convenient input and output tool compatible with genome coordinates, so we are not asking authors to convert variants already identified with genomic sequence. Genetic variants unambiguously identified in a central database by a unique genomic reference sequence may continue to be cited by their accession number and allele (for example, SNPs: rs123 A). The mnemonic practice of citing a SNP together with a nearby gene will be tolerated provided the variant is within the transcription unit or is in perfect linkage disequilibrium with one or more transcribed variants. However, authors should consider that other researchers may be trying to integrate information about variants reported from various link-

age, resequencing and marker association studies and so keep allele descriptions commensurate with the method by which their data were generated.

The LRG reference sequences should be used in conjunction with standard HGNC gene abbreviations (<http://www.genenames.org/>) that we already require as a condition of publication. All human genetic variants must now be described—in abstracts and at first use—in accordance with the Human Genome Variation Society (HGVS) conventions (<http://www.hgvs.org/mutnomen/>) also as a condition of publication. We continue to encourage authors to use HGVS nomenclature for unambiguous reference in all tables and figures and throughout the paper. Reporting variants relative to a reference sequence is a good research practice that can improve the usability and citation of your publications. The LRG construct contains taxonomic information and may be adopted by other model organism communities.

Obviously the LRG concept is more important for communities working with many variants in few genes than for those working on few variants in many loci, but the practice of reporting variants relative to a reference sequence is essential as an increasing number of researchers rely on global genomic data integration. Applying for, designing and using the new reference sequences may take some time, so we will not delay publication where variants are described relative to an existing RefSeqGene or reported in genomic coordinates relative to a named genome build.

We are also pleased to note that the ‘sandwich’ of XML code that makes up a LRG construct contains fields for attribution to the original databases that housed the variant data as well as to publications that led to the locus annotation, model construction and specification of the reference sequence itself. In time, it should be possible to use this metadata to display quantitative credit to the data generators and curators who wisely enabled research and clinical testing with these new tools. ■