The Jackson Laboratory

Leading the search for tomorrow's cures

Prof. Raymond Dalgleish
Department of Genetics and Genome Biology
University of Leicester

Wednesday, March 13, 2019

Dear Raymond,

As you know, I am a physician and computer scientist by training, and my laboratory has been responsible for the development of the Human Phenotype Ontology (HPO), which provides comprehensive bioinformatic resources for the analysis of human diseases and phenotypes, offering a computational bridge between genome biology and clinical medicine. The HPO has been adopted by major international projects such as the NIH Undiagnosed Diseases Network, several European Reference Networks for Rare Disease, and Genomics England's 100,000 Genomes project, and has become the *de facto* standard for computational phenotype analysis in rare disease (http://www.human-phenotype-ontology.org).

In addition to developing the HPO, we develop algorithms for phenotype-driven genomic diagnostics. For this, we have performed detailed biocuration of published disease-association variants to generate training sets for machine learning algorithms (*e.g.*, our Genomiser software: Am J Hum Genet. 2016;99:595-606). The curation of an individual case requires us to extract the variant from the publication, which is usually given as a transcript using the HGVS nomenclature. For our machine learning, we require the genomic coordinates, and we apply several Q/C approaches to ensure that we have the correct position (otherwise, our training set would be unusable). In the initial phases of our project, we would do this with the UCSC Genome Browser, which took usually about 30–60 minutes per case. When we discovered VariantValidator, we took advantage of the fact that the tool easily converted between both coordinate sets. More recently, we have been using the VariantValidator API directly in our curation tool, and have brought down the time to curate a variant to as little as a few minutes, roughly a ten-fold improvement in curation efficiency. We have also used what is probably VariantValidator's main competitor, Mutalyzer, but have chosen to use VariantValidator for our work because of the substantially greater amount of information that is presented to the user, the ease of use and the flexibility of the API and of parameterized URLs, and the much better graphical user interface, all of which improves the efficiency of our work and saves us time.

Sincerely,

Peter Robinson
Professor and Donald A. Roux Chair, Genomics and Computational Biology
The Jackson Laboratory for Genomic Medicine
10 Discovery Drive
Farmington, CT 06032