

MutRes and parsing LSDB data into HGVbase

Pablo Marín-García¹, Stefan Wecks², Tony Brookes³, Heikki Lehtväslaiho¹

¹ European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambs. CB10 1SD, United Kingdom

² Bioinformatics Unit, department of Molecular Genetics, Flanders Interuniversity Institute for Biotechnology – VIB, University of Antwerp Universiteitsplein 1 B-2610 Wilrijk – Antwerpen

³ Karolinska Institute, Center for Genomics Research, Berzelius väg 35, 171 77 Stockholm, Sweden

The MutRes database of mutation databases and related resources (Lehtväslaiho 2000) has been updated and rewritten. It has been moved from flat file into a relational database with perl object layer providing command line and web interfaces to it. It now contains nearly 300 Locus Specific Data Bases (LSDB) and metadata entries with detailed information about URLs, genes, reference sequences and maintainers. The web interface allows the researchers perform full and boolean keyword searches, as well as create new and updated entries. The database can be easily be output to various text formats.

The new MutRes database provides the crucial metadata for collecting and parsing LSDB mutation details into HGVbase. We are now creating a highly automated system to accomplish this in clearly defined steps:

1. Mirror LSDB data files
2. Parse mutation entries from distribution files
3. Find changed entries since previous update
4. Parse entry details into memory
5. Create standardized mutation representation
6. Validate mutations against reference data
7. Write out new data for inclusion into HGVbase

With this new pipeline, we are finally ready to start distributing mutations in to wider audience together with SNPs.

References

Lehtväslaiho H, Stupka E, Ashburner M: **Sequence variation database project at the European Bioinformatics Institute**. *Hum.Mutat.* 2000, **15**:52-56.

MutRes Database: see <http://www.ebi.ac.uk/mutations/>