

Dear LSDB Curators

On behalf of the HGVS and the GEN2PHEN project we are writing about an international collaborative project to integrate LSDB data with central repository and genome browser information. The project builds on previous efforts including projects such as PhenCode and Sue Povey's providing an extract of some of her TSC2 data to NCBI. Indeed, Donna Maglott or Ray Tully at the NCBI may have already contacted you regarding reference gene sequences.

An international collaboration, which includes the EBI, the NCBI, and the GEN2PHEN project, aims to create integrated data resources and visual interfaces based on consolidating core information from LSDBs with genome resources such as dbSNP and Ensembl. The overall goal is to create a useful public resource, while giving full acknowledgement and links to the LSDBs that will continue to maintain their own detailed records.

To participate, we ask that you submit merely the core content from your LSDB to either the NCBI or the EBI. Suggested data elements for exchange are listed below, but additional or fewer or alternative items may be submitted based on the needs of your LSDB. The question of 'what is core LSDB content' has been addressed by Johan den Dunnen in a document recently sent to HGVS members for comment (see accompanying attachment). It is not necessary for the data from your LSDB to be made public as part of participation in this project. However, if you are happy for us to release the core data from your LSDB to the public, that content will be integrated and released through public genome browsers and other resources.

If you are interested in participating, please send a message to either Paul Flicek (flicek@ebi.ac.uk) or Donna Maglott (maglott@ncbi.nlm.nih.gov) to discuss data transfer. Data previously provided to either the EBI or the NCBI does not need to be resubmitted.

With best regards,

Anthony Brookes (ajb97@le.ac.uk)
on behalf of the GEN2PHEN Consortium
(<http://www.gen2phen.org>)

Richard Cotton (cotton@unimelb.edu.au)
on behalf of the HGVS
(<http://www.hgvs.org/>)

DATA EXCHANGE, LSDB TO BROWSER: Suggested core content

- 1) Contact details for PI(s) running the LSDB
- 2) Gene Name(s)/Symbol(s) [from HGNC] and utilised Reference Sequence(s)
- 3) Name and URL of the LSDB
- 4) URL syntax for linking to individual LSDB records
- 5) Core Content to exchange per variant
 - a) HGNC Gene Symbol (e.g. 'CAPN3')
 - b) Variant in HGVS nomenclature (e.g., c.-766A>G)
 - c) Variant ID (for direct URL linking to this record in your LSDB)
 - d) dbSNP ID (e.g., rs1801496) [if exists / optional]
 - e) OMIM ID (e.g., OMIM114240:0004) [if exists / optional]
 - f) PubMed ID for paper describing variant(s) [if exists / optional]

Notes:

- If a particular variant exists as multiple independent records in your LSDB, then submit those variants multiple times with their own specific linking information for point 5)c).
- Phenotype and pathogenicity data are not considered core information at this stage of the project as standards for this must first be agreed to avoid any risk of misinterpretation.