Two Paths from LSDBs: Core versus Deeper Data

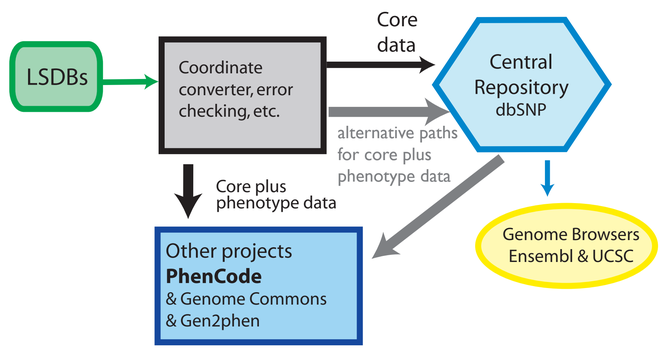
Belinda M. Giardine, Riemer C., Miller W., Hardison R.C.
Center for Comparative Genomics and Bioinformatics, The Pennsylvania State University, University Park, PA USA

phencode.bx.psu.edu

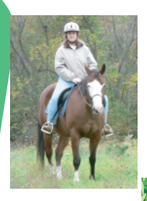
Data Sources

SRD5A2 · LMDp · RPGR · BGMUT · OSTM1base · KinMutBase · CLCN7base · IARC TP53 · CA2base · HbVar · ALPL · ARdb · IDbases · AD&FTDMD · PAHdb · MMR · CBS · CASRdb · database x-ald · F12base · dbRIP a database of retrotransposon insertion polymorphisms in humans

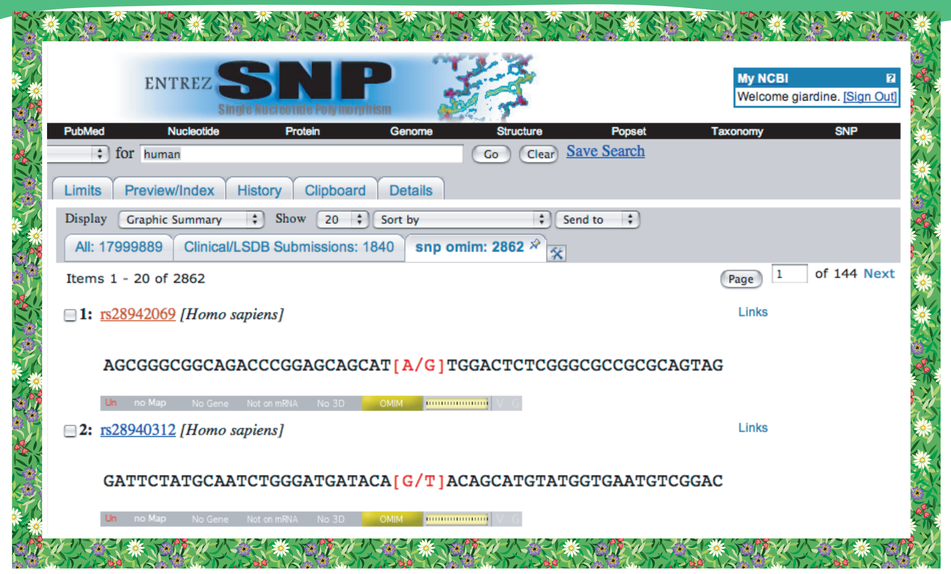
Data flow

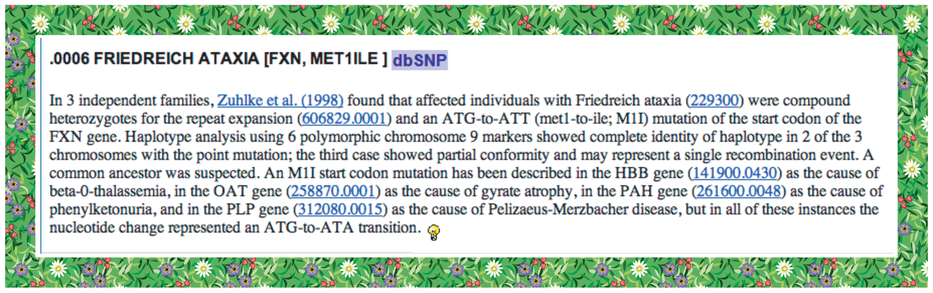


1. For the first example I start at dbSNP and query to see what data is available. Of over 17 million SNPs 1,840 of them are Clinical/LSDB Submissions, and 2,862 are in OMIM.



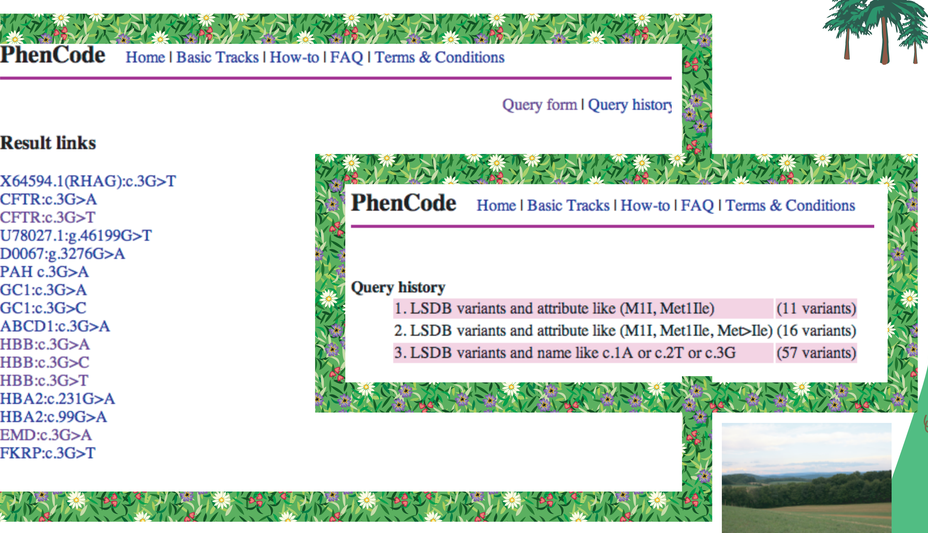
Core data





2. Following hyperlinks to OMIM, we learn that this variant alters the translation start codon. Mutations of the start codon are associated with pathology in other genes.

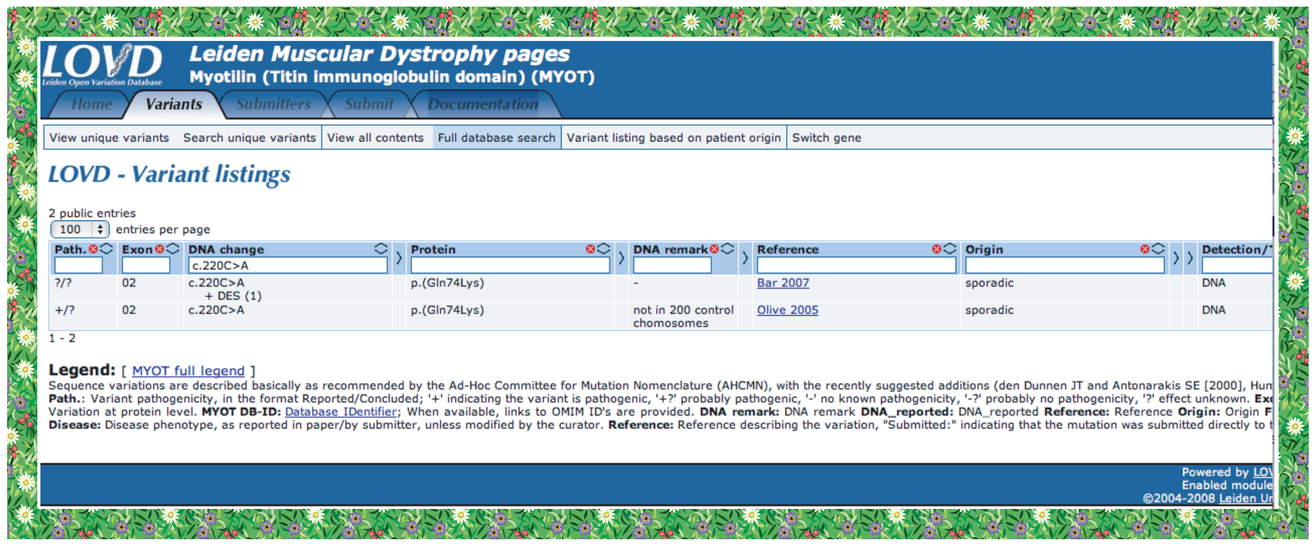
3. How often are translation start codons mutated in disease-associated alleles? By querying at PhenCode I found several additional examples.



4. For the second example I browse personal genomes in the UCSC Genome Test Browser. The mutant form of this variant is found in the YRI HapMap population, Dr. James Watson, and the reference chromosome sequence. The LSDB uses a different reference where the wildtype matches the other species shown in the multiple alignment.

Deeper data

5. To find out whether this variant is associated with a pathology, I followed the hyperlinks to the source LSDB, LMDp. Here we learn that the pathogenicity of this variant is unknown.



- While the core data is useful the deeper data is also needed in a central location to make full use of the data.
- The LSDBs are also useful for providing information that is specific to the locus and more specialized query access.