PhenCode: Linking Human Mutations and Phenotype

PhenCode is a collaborative project to bring the deep information on genotypes and phenotypes in locus specific databases (LSDBs) into a common database in a manner that can be displayed and analyzed using genome sequence coordinates. By doing so, users can interrogate the extensive information on human variation (as a Locus Variants track in a genome browser) along with associated information on genes, transcripts, chromatin modifications, evolutionary conservation, and other genomic information. This gives power to the exploration of links between genotype and phenotype.

The information in LSDBs is quite comprehensive and often deeply annotated for phenotypes. We suspected that these would have more mutations than generic SNP sources, such as dbSNP. To test this, we ran the following analyses, which illustrate the richness and depth of information derived from the LSDBs in PhenCode.

In this example, we are seeking the substitutions from the Locus Variants that are not in dbSNP but are in the regions from the UCSC Genome Browser’s Most Conserved track (PhastCons). We then follow the links to view more details on a single variant.

1. Start by selecting the browser, assembly and initial view. Then select the mutations to include in the custom track.

2. The Locus Variants appear as a custom track. Adjust the browser to show other tracks of interest.

3. The Table Browser is used to do genome wide comparisons of custom tracks and resident tracks.

4. All tracks can be displayed in the Genome Graphs tool to show coverage. Clicking on the bars takes you back to the Browser.

5. Here are the custom tracks made using the Table Browser and Genome Graphs. The Locus Variants link back to PhenCode and the data sources.

6. The PhenCode page provides a summary of the data on the selected mutation and a link back to the data source.