PhenCode: Linking Human Mutations and Phenotype

www.bx.psu.edu/phencode

*Giardine B., Kent W.J., Miller W., Hardison R.C.
Center for Comparative Genomics and Bioinformatics, Pennsylvania State University, University Park, PA;
Center for Biomolecular Science and Engineering, University of California, Santa Cruz, CA
*E-mail: giardine@bx.psu.edu

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.

Data Sources

BGMUT
X-ALD
SRD5A2
OSTN1base
LMDp
KinMutBase
CLCN7base
IARC TP53
CA2base
ARdb
AD&FTDMD

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

PhenCode: Home | Introduction | How-to | Examples | FAQ | Mailing list | Contact us | Disclaimer

PhenCode Query Page

This form sends the chosen Locus Variants data from PhenCode to a genome browser, as one or more custom tracks. The tracks are genome-wide; "browser position" simply specifies the initial view.

BROWSER CHOICES

Destination: UCSC Genome Browser

Human genome assembly: hg18: March 2006

Browser position: GARS [e.g. "chr12:101,761,637-101,761,687" or "HBB"]

DATA CHOICES

Data sources:
- [x] Locus-Specific Databases
- [ ] UniProt (Swiss-Prot/TrEMBL)

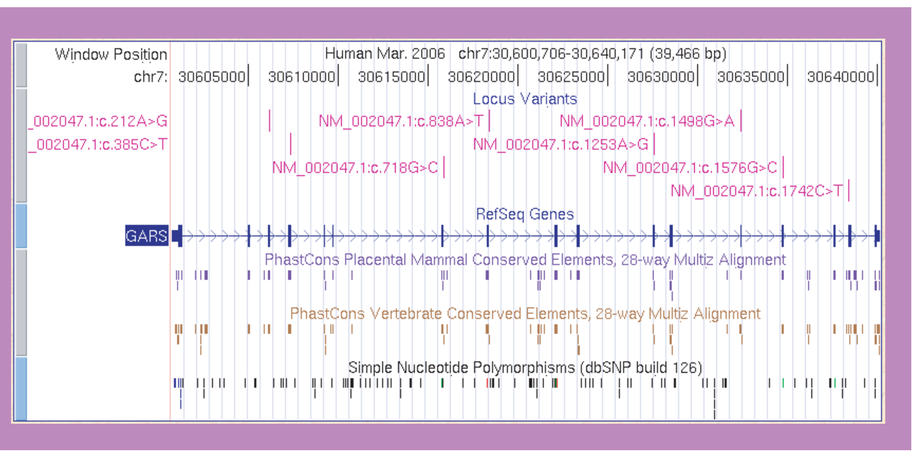
Mutation types:
- [x] substitutions
- [ ] deletions
- [ ] insertions
- [ ] duplications
- [ ] complex

Mutation locations:
- [x] exon
- [x] intron
- [x] 5′ UTR
- [x] 3′ UTR
- [x] not within known transcription unit

Estimated coordinates:
- [x] mutations whose coordinates are estimated

Go

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.

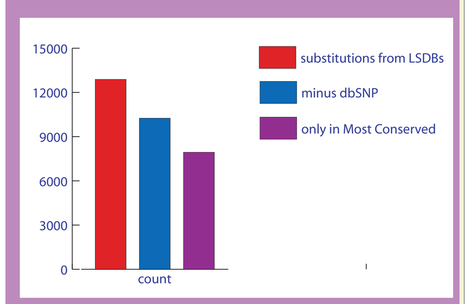


Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. For help in using this application see Using the Table Browser for a description of the controls in this form, the User's Guide for general information and sample queries, and the OpenHelix Table Browser tutorial for a narrated presentation of the software features and usage. For more complex queries, you may want to use Galaxy or our public MySQL server. Refer to the Credits page for the list of contributors and usage restrictions associated with these data.

clade: Vertebrate **genome:** Human **assembly:** Mar. 2006
group: All Tracks **track:** PhenCodeLSDB
table: ct_PhenCodeLSDB [remove custom track] [describe table schema]
region: genome position chr17:7520135-7520347 [lookup] [define regions]
identifiers (names/accessions): [paste list] [upload list]
filter: [create]
intersection with snp126: [edit] [clear]
correlation: [create]
output format: custom track Send output to Galaxy
output file: (leave blank to keep output in browser)
file type returned: plain text gzip compressed

Note: Intersection doesn't work with all fields or selected fields output.
[get output] [summary/statistics]

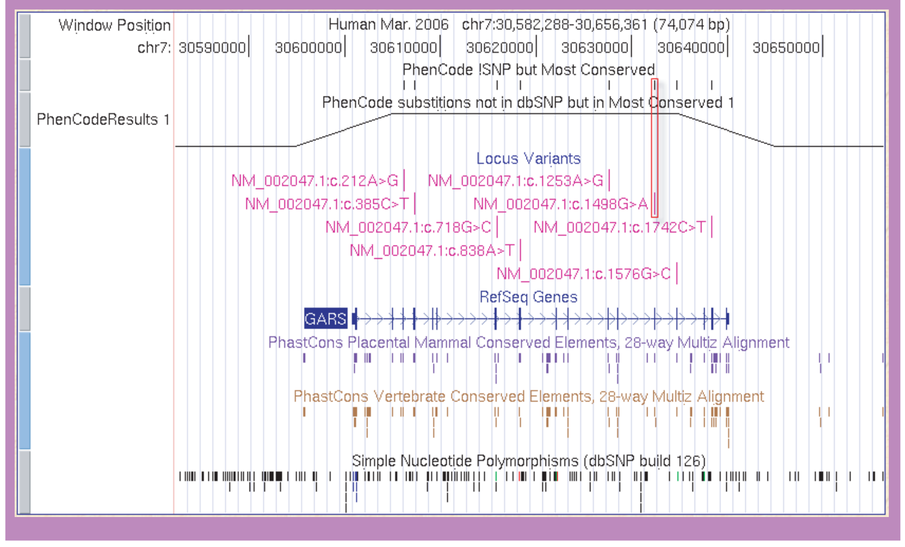
To reset **all** user cart settings (including custom tracks), click here.

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

Human Genome Graphs

clade: Vertebrate genome: Human assembly: Mar. 2006
graph PhenCodeResults 1 in purple , PhenCodeSubs 1 in red
[upload] [import] [configure] [correlate] significance threshold: 3.5 [browse regions] [sort genes]

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.

PhenCode Home | Introduction | How-to | Examples | FAQ | Disclaimer | Contact us

Mutation Database of Inherited Peripheral Neuropathies

Eva Nelis, VIB - Department of Molecular Genetics, University of Antwerp, Antwerpen, Belgium

NM_002047.1:c.1498G>A
base change type : substitution
location : exon
Human March 2006 (hg18)
chr7:30632420-30632421 +
Common name: c.1498G>A
Asp500Asn
External links: IPNMDB: 826
Phenotype common name: CMT2 / dHMN

Key:
Red is initial custom track from PhenCode.
Purple is the results of table browser queries.

Summary:
1. PhenCode data is public. It can be viewed on the UCSC Genome Browser, Ensembl Contig View or downloaded as text.
   www.bx.psu.edu/phencode
2. The depth of the LSDB data in PhenCode complements the broader coverage of generic SNP sources.

Acknowledgements:
Work was supported by NIH grants HG002238 (WM) and DK65806 (RH), and NHGRI grant 1P41HG02371 (WJK).