Human Genes, Variation, and Medical Genetics Resources

Peter Cooper peter.cooper@nih.gov

Topics

- Variation Resources / Databases
- Medical Genetics Resources
- Web access to data
  - Using Entrez
  - Linking from Gene
  - Other tools and browsers
- Live Searches
Relevant Databases

- dbSNP
- dbVar
- dbGaP
- ClinVar
- MedGen
- GTR

Sequence Variation

1000 Genomes Project

Medical Genetics

Sequence Variation Databases
NCBI Variation Databases

- **dbSNP** small scale variants
  - Submitted (ss) and reference (rs)
- **dbVar** large scale variants
  - Variant calls (ssv) and regions (sv)
  - No reference variants
- **ClinVar** variant – phenotype assertions
  - Phenotypic assertions about small and large scale variants
  - Submitted from a large number of sources
- **dbGaP** phenotype – genotype association studies
  - Genome-wide association studies, medical sequencing, molecular diagnostic assays
  - Controlled access to individual level data

NCBI’s SNP Database

- Small Scale variations (< 50 bp)
- Rare variants and common polymorphisms
  - Single Nucleotide Variations (SNVs)
  - Simple repeats
  - Insertion-Deletion variations
- Primary records (submitted SNP) and derivative (RefSNP)
  - 165 million human RefSNPs (non-redundant)
Placing on Reference Sequences

A SNP record essentially consists of positions on Reference sequences and the variation.

- Clinical significance allele origin from submitter and ClinVar
- Validated by dbSNP
- Global MAF provided from 1000Genomes

dbVar: Large-scale Variations

Structural variants
Copy number variants (CNV)

4,692,250 variant regions
- copy number variation (2,944,557)
- insertion (1,423,247)
- mobile element insertion (225,496)
- translocation (34,474)
- inversion (19,975)
- novel sequence insertion (13,629)
- sequence alteration (13,544)
- complex substitution (8,132)
- tandem duplication (6,908)
- indel (2,288)

A dbVar record contains the break points on a Reference and the variation.
Variant Calls / Variant Regions

One Variant Region: nsv428141

Four Variant Calls

1000Genomes: survey of human variation
The 1000 Genomes Project

1000 Genomes
A Deep Catalog of Human Genetic Variation

- Multi-phase, multi-center project
- Catalog of human genomic variation
  - Short and structural variations
  - Large source of dbSNP and dbVar data
- Phenotypically normal individuals
- Genotypes public
- Data mirrored at EBI and NCBI

Samples
- 2,504 individuals
- 26 human populations
- 5 continental regions

Variants
- 84.7 million SNPs
- 3.6 million indels
- 60 K structural variant regions

www.1000genomes.org

ClinVar: Variations and Assertions
ClinVar

- Bridges variation and medical / clinical resources
- Accepts submission of small and large (copy number) variants and phenotypic assertions
- Provides variant-condition accessions
  - RCV – aggregate
  - SCV – individual submissions
- Represent variants using HGVS and reference sequences including RefSeqGene
- Curates and interprets information on top of dbGaP, dbSNP, and dbVar
- Provides Review Status (star rating)

06/06/2016

ClinVar Statistics

140,434 variant records

Clinical Significance
- Pathogenic (46,395)
- Uncertain significance (38,645)
- Benign (19,451)
- Likely benign (16,868)
- Likely pathogenic (8,602)
- Conflicting interpretations (3,492)
- Risk factor (843)

Review status (star rating 1-4)
- Practice guideline (23) 4-star
- Expert panel (3,626) 3-star
- Multiple submitters (8,362) 2-star
- Single submitter (59,522) 1-star
- Conflicting interpretations (3,120) 1-star
- No assertion / criteria (65,781) 0-stars

06/06/2016
**Review Status**

<table>
<thead>
<tr>
<th>Variation ID</th>
<th>Review status</th>
</tr>
</thead>
<tbody>
<tr>
<td>7106</td>
<td></td>
</tr>
<tr>
<td></td>
<td>* * * practice guideline</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variation ID</th>
<th>Review status</th>
</tr>
</thead>
<tbody>
<tr>
<td>38125</td>
<td></td>
</tr>
<tr>
<td></td>
<td>* * * reviewed by expert panel</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variation ID</th>
<th>Review status</th>
</tr>
</thead>
<tbody>
<tr>
<td>12379</td>
<td></td>
</tr>
<tr>
<td></td>
<td>* * * criteria provided, multiple submitters, no conflicts</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variation ID</th>
<th>Review status</th>
</tr>
</thead>
<tbody>
<tr>
<td>216017</td>
<td></td>
</tr>
<tr>
<td></td>
<td>* * * criteria provided, single submitter</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variation ID</th>
<th>Review status</th>
</tr>
</thead>
<tbody>
<tr>
<td>55376</td>
<td></td>
</tr>
<tr>
<td></td>
<td>* * * * (0/4) no assertion criteria provided</td>
</tr>
</tbody>
</table>

**ClinVar Data Relationships**

- ClinVar is submission pathway to dbSNP/dbVar for clinical SNPs
- ClinVar provides submitted interpretation not available in dbSNP/dbVar

**Testing laboratories, LSDB, Authors…**

**Expert review**
Medical Genetics Resources

- **MedGen**
  - Human disorders and other phenotypes having a genetic component
  - Provides several controlled vocabularies of phenotypic terms

- **Online Mendelian Inheritance in Man (OMIM)**
  - Now at [www.omim.org](http://www.omim.org)
  - Still searchable at NCBI
  - Articles about human disease genes, phenotypes, and selected variants

- **The Genetic Testing Registry**
  - Submitted genetic tests with purpose, methodology, validity, utility, lab contacts and credentials
**MedGen**

- NCBI portal to Medical Genetics
- Aggregates data from other sources
- Automated and curated data feeds
- Unifies vocabularies from GTR, OMIM, UMLS, GeneReviews, GTR, HPO, Orphanet, MeSH and other sources
- Links to GeneReviews, Genetics Home Reference
- Phenotypes using standard vocabularies
- Links to NCBI and outside resources

**Genetic Testing Registry**

- Submitted tests for Mendelian disorders (including pharmacogenetic tests)
- Provides searches by
  - Disorder
  - Test
  - Laboratory
- Condition pages with links to other resources (Gene, GeneReviews)
The Database of Genotypes and Phenotypes

- Human phenotype data and related individual-level molecular data / genotypes
- Data for more than than 1 million individuals across 645 studies
- Information about studies
  - individual level molecular and phenotype data
  - analysis results
  - medical images
  - general information about the study,
  - research protocols
  - questionnaires
- Molecular data types
  - Genotypes
  - Expression
  - Genomic Sequence
  - Epigenomic data
  - Somatic mutation
  - Microbiome
- Individual level data requires Controlled Access application and approval

Accessing Variation Data
The Entrez System

All Databases (Entrez)

Search NCBI databases

right ventricular cardiomyopathy

Results found in 28 databases for "right ventricular cardiomyopathy"

<table>
<thead>
<tr>
<th>Database</th>
<th>Results</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ClinVar</td>
<td>294</td>
<td>human variations of clinical significance</td>
</tr>
<tr>
<td>dbGaP</td>
<td>91</td>
<td>genotype/phenotype interaction studies</td>
</tr>
<tr>
<td>GTR</td>
<td>273</td>
<td>genetic testing registry</td>
</tr>
<tr>
<td>MedGen</td>
<td>24</td>
<td>medical genetics literature and links</td>
</tr>
<tr>
<td>OMIM</td>
<td>127</td>
<td>online mendelian inheritance in man</td>
</tr>
<tr>
<td>SNP</td>
<td>934</td>
<td>short genetic variations</td>
</tr>
<tr>
<td>dbVar</td>
<td>15</td>
<td>genome structural variation studies</td>
</tr>
</tbody>
</table>
Linking from Gene

PMP22 peripheral myelin protein 22 (Homo sapiens (human))
Gene ID: 6376, updated on 24 Apr 2016

Summary
- Official Symbol: PMP22 (provided by HGNC)
- Official Full Name: peripheral myelin protein 22 (provided by HGNC)
- Gene type: protein coding
- RefSeq status: REVIEWED
- Organism: Homo sapiens
- Lineage: Eukaryota; Metazoa; Chordata; Cnidaria; Vertebrata; Euteleostomi; Mammalia; Primates; Hominidae; Homo
- Also known as: Summary

This gene encodes an integral membrane protein that is a major component of the peripheral myelin. Studies suggest two alternatively-used promoters drive tissue-specific expression of the gene. The gene is expressed in neurons and may be involved in myelination of peripheral nerves. Alternative splicing results in multiple transcript variants.

Phenotypes: conditions

Find tests for this gene in the NIH Genetic Testing Registry (GTR).
Review e21, and phenotype association data in this region using PheGene.
Associated conditions

- Charcot-Marie-Tooth disease, type IA
- Charcot-Marie-Tooth disease, type IE

GTR Tests
GTR condition
PheGene Gene Query
GTR Lab comparison

MedGen Concept
OMIM Phenotype
Variation

Short and structural variants with clinical significance (Gene search in ClinVar).

All structural variants (Gene search in dbVar)

Gene region in Variation Viewer (current and previous assemblies)

Gene region in 1K Genomes Browser

Access to all variation resources, viewers and tools

<ncbi>/variation/
Variation Tools and Viewers

- **Variation Viewer**
  - Integrated genome browser for SNPs and structural variants with filter, download options and ability to upload variants
- **1000 Genomes Browser**
  - Dedicated genome browser with Sequence alignments and variant calls
- **Variation Reporter**
  - Identifies known variants and functional consequences in uploaded data
- **Clinical Remap**
  - Map variants onto RefSeqGene records
- **Phenotype Genotype Integrator (PheGenI)**
  - Browser and search tool that integrates data from NHGRI GWAS catalog, Gene, dbGaP, OMIM, GTEx and dbSNP


<ncbi>/variation/tools/

The Variation Viewer Components

- One of several dedicated genome browsers at NCBI
- Powerful search and navigation features
- Filterable / Downloadable Variant table with links to variation database
- Upload your own data

Variation Viewer

Search

Gene & Exon Navigator

Variant Filter

Variant Table

06/06/2016
1000 Genomes Browser

- Main resource for interactive access to 1K Genomes
- Genotype table with allele frequencies / counts
- Displays aligned next-gen reads from samples as tracks
- Provides download of VCF for displayed regions

<ncbi>/variation/tools/1000genomes/

Getting Help

- Learn: <ncbi>/home/learn.shtml
- Factsheets: <ftp>/pub/factsheets/
- NCBI YouTube Channel: (www.youtube.com/ncbinlm)
- NCBI Helpdesk: info@ncbi.nlm.nih.gov
Live Demonstrations