

Human Genes, Variation, and Medical Genetics Resources

Peter Cooper peter.cooper@nih.gov

06/06/2016



1

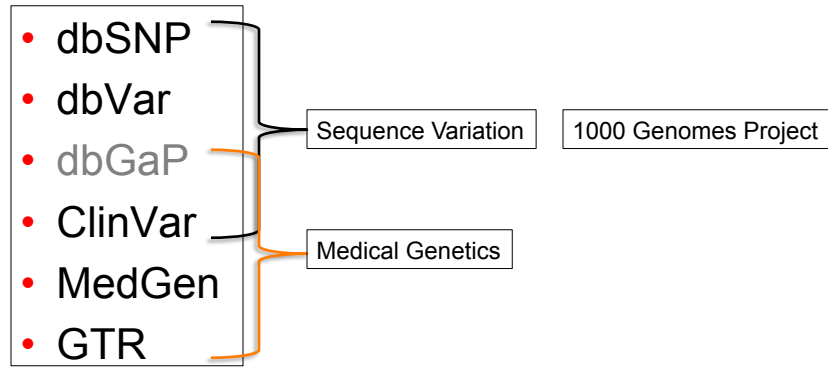
Topics

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

06/06/2016

2

Relevant Databases



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

06/06/2016

5

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)

06/06/2016

6

Placing on Reference Sequences

rs429358 [Homo sapiens]

GCTGGGCGGGACATGGAGGACGTG [C/T] GCGGCCGCCTGGTGCAGTACCGCGG

Chromosome: 19:44908684
 Gene: APOE (GeneView)
 Functional Consequence: missense
 Allele Origin: T(germline)/C(germline)
 Clinical significance: Pathogenic
 Validated: by 1000G,by 2hit 2allele,by cluster,by frequency,by hapmap
 Global MAF: C=0.1506/754
 HGVS: NC_000019.10:g.44908684T>C, NC_000019.9:g.45411941T>C, NG_007084.2:g.7903T>C, NM_000041.3:c.388T>C, NM_001302688.1:c.466T>C, NM_001302689.1:c.388T>C, NM_001302690.1:c.388T>C, NP_000032.1:p.Cys130Arg, NP_000032.1:p.Cys130Arg

- 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

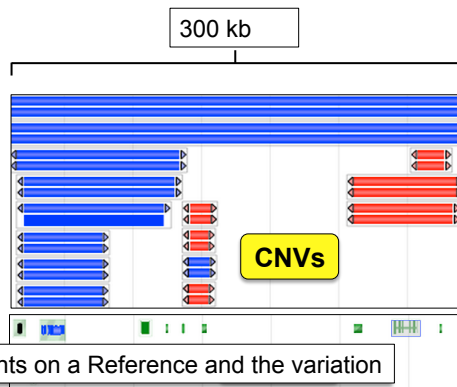
06/06/2016

7

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)



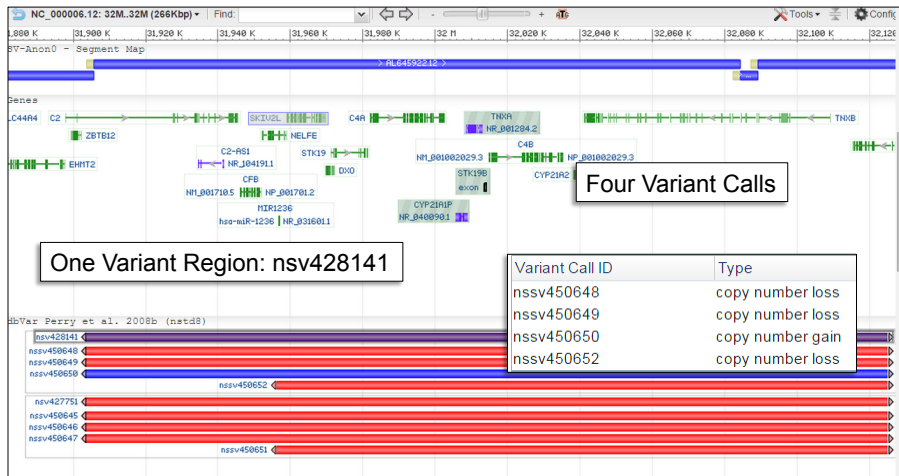
A dbVar record contains the break points on a Reference and the variation

HGVS NC_000006.12:g.(?_31903735)_(32125066_?)dup

06/06/2016

8

Variant Calls / Variant Regions



06/06/2016

9

1000Genomes: survey of human variation

06/06/2016

10

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

<http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/>

www.1000genomes.org

06/06/2016

11

ClinVar: Variations and Assertions

06/06/2016

12

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)

[<ncbi>/clinvar/](https://ncbi.nlm.nih.gov/clinvar/)

06/06/2016

13

ClinVar Statistics

140,434 variant records

The screenshot shows the ClinVar website interface. At the top, there is a navigation menu with links for Home, About, Access, Using the website, Submission, Statistics, and FTP site. A search bar is located on the right. Below the navigation, there is a section for variant records with a list of nucleotide sequences. Two callout boxes are overlaid on the page:

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

14

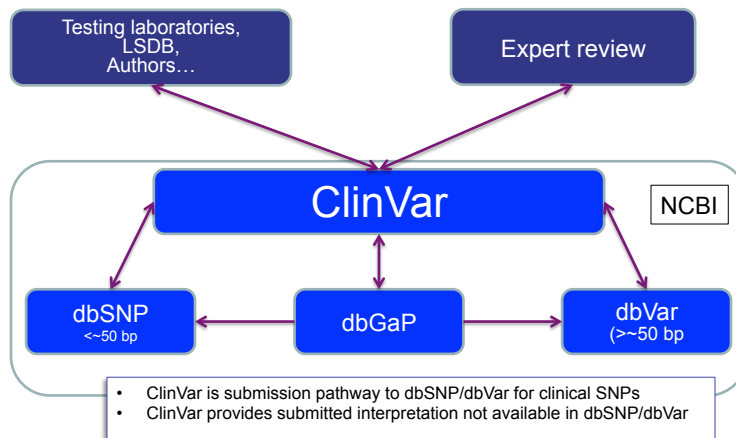
Review Status

NM_000492.3(CFTR):c.1519_1521delATC (p.Ile507del)	
Variation ID: ?	7106
Review status: ?	★★★★ practice guideline
NM_000059.3(BRCA2):c.7878G>C (p.Trp2626Cys)	
Variation ID: ?	38125
Review status: ?	★★★★★ reviewed by expert panel
NM_000546.5(TP53):c.1010G>A (p.Arg337His)	
Variation ID: ?	12379
Review status: ?	★★★ criteria provided, multiple submitters, no conflicts
NM_000038.5(APC):c.423G>T (p.Arg141Ser)	
Variation ID: ?	216017
Review status: ?	★★★ criteria provided, single submitter
NM_007294.3(BRCA1):c.5074G>T (p.Asp1692Tyr)	
Variation ID: ?	55376
Review status: ?	★☆☆ (0/4) no assertion criteria provided

06/06/2016

15

ClinVar Data Relationships



06/06/2016

16

Medical Genetics Resources

06/06/2016

17

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
 - 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

06/06/2016

18

MedGen

MedGen

MedGen

Search

Limits Advanced Help

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

ncbi.nlm.nih.gov/medgen

06/06/2016

19

Genetic Testing Registry

GTR: GENETIC TESTING REGISTRY

All GTR Tests Conditions/Phenotypes Genes Labs GeneReviews

Advanced search for tests

arrf Search All GTR

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

ncbi.nlm.nih.gov/gtr

06/06/2016

20

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
https://dbgap.ncbi.nlm.nih.gov/aa/dbgap_request_process.pdf

06/06/2016

21

Accessing Variation Data

06/06/2016

22

The Entrez System

06/06/2016

23

All Databases (Entrez)

Search NCBI databases

right ventricular cardiomyopathy

Results found in 28 databases for "right ventricular cardiomyopathy"

Health		
ClinVar	294	human variations of clinical significance
dbGaP	91	genotype/phenotype interaction studies
GTR	273	genetic testing registry
MedGen	24	medical genetics literature and links
OMIM	127	online mendelian inheritance in man
Genomes		
SNP	934	short genetic variations
dbVar	15	genome structural variation studies

06/06/2016

24

Linking from Gene

See 123 articles about **PMP22** gene function
 See also: **PMP22** peripheral myelin protein 22 in the Gene database
pmp22 in *Homo sapiens* *Mus musculus* *Rattus norvegicus* All 183 Gene records
 See also: 99 tests for **PMP22** in the Genetic Testing Registry

PubMed Gene Sensor

PMP22 peripheral myelin protein 22 [*Homo sapiens* (human)]
 Gene ID: 5376, updated on 24-Apr-2016 **Gene Record**

Table of contents

- Summary
- Genomic context
- Genomic regions, transcripts, and products
- Bibliography
- Phenotypes
- Variation
- Pathways from BioSystems
- Interactions
- General gene information
 - Markers, Clone Names, Homology, Gene Ontology
- General protein information
- NCBI Reference Sequences (RefSeq)
- Related sequences
- Additional links
 - Locus-specific Databases

Summary

Official Symbol PMP22 provided by HGNC
Official Full Name peripheral myelin protein 22 provided by HGNC
Primary source HGNC:HGNC:9118
See related Ensembl:ENSG00000109099; HPRD:03059; MIM:601097; Vega:OTTH
Gene type protein coding
RefSeq status REVIEWED
Organism *Homo sapiens*
Lineage Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Placentalia; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as DSS; GAS3; HNPP; CMT1A; CMT1E; GAS-3; Sp110; HMSNIA
Summary This gene encodes an integral membrane protein that is a major component of the peripheral myelin sheath. Studies suggest two alternately used promoters drive tissue-specific expression of this gene. Mutations in this gene are causes of Charcot-Marie-Tooth disease Type IA, Dejerine-Sottoliability to pressure palsies. Alternative splicing results in multiple transcripts.

06/06/2016 25

Phenotypes: conditions

Find tests for this gene in the NIH Genetic Testing Registry (GTR) **GTR Tests**
 Review eQTL and phenotype association data in this region using PheGenI **PheGenI Gene Query**

Associated conditions

Description	Tests
Charcot-Marie-Tooth disease, type IA MedGen: C0270911, OMIM: 118220, GeneReviews: Charcot-Marie-Tooth Neuropathy Type 1	Compare labs
Charcot-Marie-Tooth disease, type IE MedGen: C0270912, OMIM: 118220, GeneReviews: Charcot-Marie-Tooth Neuropathy Type 1	Compare labs
Dejerine-Sottoliability to pressure palsies MedGen: C0011195, OMIM: 145900, GeneReviews: Not available	Compare labs
Guillain-Barre syndrome, familial MedGen: C1841700, OMIM: 139393, GeneReviews: Not available	Compare labs
Hereditary liability to pressure palsies MedGen: C0393814, OMIM: 162500, GeneReviews: Hereditary Neuropathy with Liability to Pressure Palsies	Compare labs
Roussy-Levy syndrome MedGen: C0205713, OMIM: 180800, GeneReviews: Not available	Compare labs

MedGen Concept **OMIM Phenotype** **GTR Lab comparison**

06/06/2016 26

Variation

Variation
[See variants in ClinVar](#) → Short and structural variants with clinical significance (Gene search in ClinVar).
[See studies and variants in dbVar](#) → All structural variants (Gene search in dbVar)
[See Variation Viewer \(GRCh37.p13\)](#)
[See Variation Viewer \(GRCh38\)](#) } Gene region in Variation Viewer (current and previous assemblies)
 Genotypes
[See SNP Geneview Report](#)
[See 1000 Genomes Browser \(GRCh37.p13\)](#) → Gene region in 1K Genomes Browser

06/06/2016

27

The Variation Portal

Variation Search NCBI Search
Variation
 Access NCBI's variation resources
 Access to all variation resources, viewers and tools
Getting Started
[How to submit variants: dbSNP](#)
[How to submit variants: dbVar](#)
[How to submit controlled access data: dbGaP](#)
[How to submit your clinical data](#)
[Definitions](#)
[FAQ](#)
Variation Tools
[Variation Viewer **NEW!**](#)
[Variation Reporter](#)
[Clinical Remap](#)
[Phenotype-Genotype Integrator](#)
[1000 Genomes Browser](#)
Variation Databases
[dbSNP](#)
[dbVar](#)
[dbGaP](#)
[ClinVar](#)
[GTR](#)
[ncbi/variation/](http://ncbi.nlm.nih.gov/variation/)

06/06/2016

28

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

Genome Browsers playlist
(YouTube): <http://bit.ly/1MhKOT9>

[ncbi/variation/tools/](http://ncbi.nlm.nih.gov/variation/tools/)

06/06/2016

29

The Variation Viewer Components

The screenshot displays the NCBI Variation Viewer interface for Homo sapiens. Key components are highlighted with red boxes:

- Search:** A search bar with the placeholder text "Enter a location, gene name or phenotype" and search examples.
- Gene & Exon Navigator:** A navigation bar showing the current region (CFH) and transcript (NM_000186.3).
- Variant Filter:** A sidebar with filters for source database (dbSNP, dbVar), ClinVar status, and clinical significance (Pathogenic).
- Variant Table:** A table listing variants with columns for Variant ID, Location, Variant type, Gene, Molecular consequences, Worst clinical significance, 1000G MAF, ESP, and Publications.

- One of several dedicated genome browsers at NCBI
YouTube playlist on Genome Browsers: <http://bit.ly/1MhKOT9>
- Powerful search and navigation features
- Filterable / Downloadable Variant table with links to variation database
- Upload your own data

06/06/2016

30

1000 Genomes Browser

1000 Genomes Browser

Homo sapiens: GRCh37.p13 (GCF_000001405.25) Chr 1 (NC_000001.10): 196.7M - 196.7M

Phase 3

ATTENTION: You are browsing the alignment and genotype data from the Phase 3 May 2013 call set. Data from an earlier release is also available. [Click here to browse data from the Phase 1 March 2012 call set.](#)

Graphic pane

Genotype Table

- 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

06/06/2016 [ncbi>/variation/tools/1000genomes/](http://ncbi.nlm.nih.gov/variation/tools/1000genomes/) 31

Getting Help

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

Live Demonstrations

www.ncbi.nlm.nih.gov