NCBI Webinars

Human Genes, Variation, and Medical Genetics Resources

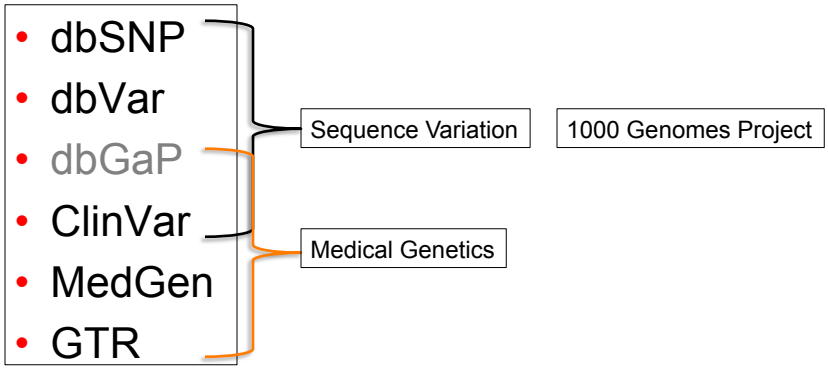
Peter Cooper peter.cooper@nih.gov

NCBI Webinars

Topics

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

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

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

NCBI Webinars

rs429358 *[Homo sapiens]*

GCTGGGCGCGGACATGGAGGACGTG[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_000041.3:c.388T>C, NP_000032.1:p.Cys130Arg, NP_[illegible]30Arg,

HGVS - NM_000041.3:c.388T>C

- 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

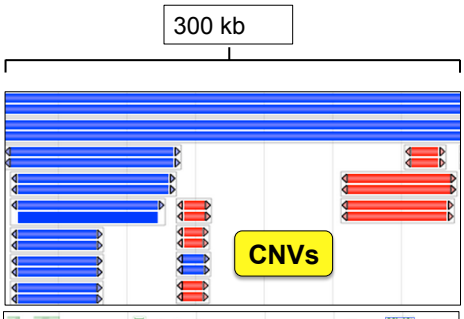
NCBI Webinars

Structural variants

Copy number variants (CNV)

4,692,250 variant regions

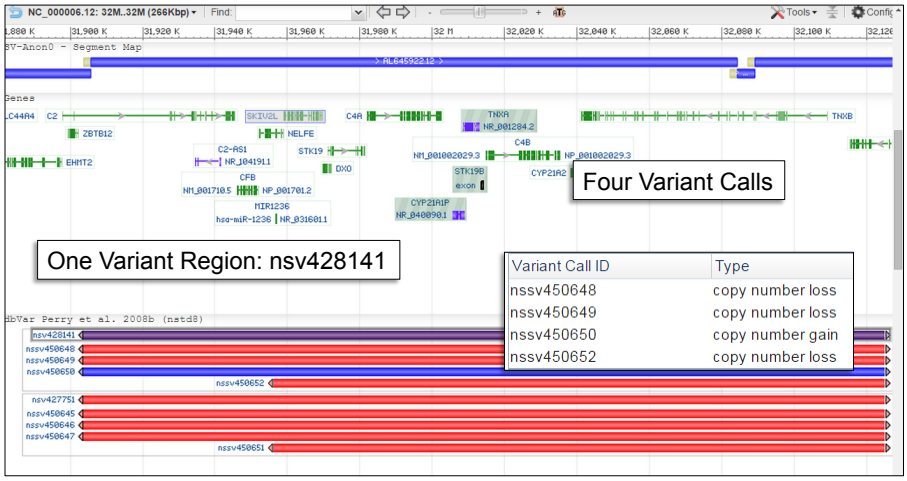
- copy number variation (2,944,557)
- insertion (1,423,247)
- mobile element insertion (225,496)
- translocation (34,474)
- inversion (19,975)
- novel sequence insertion(13,629)
- sequence alteration (13,544)
- complex substitution (8,132)
- tandem duplication(6,908)
- indel (2,288)



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

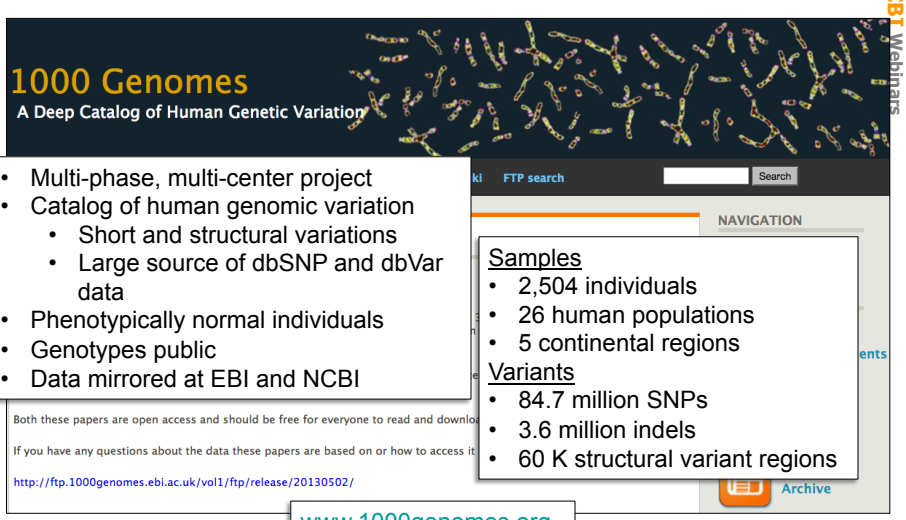
HGVS NC_000006.12:g.(?_31903735)_(32125066_?)dup

Variant Calls / Variant Regions



1000Genomes: survey of human variation

The 1000 Genomes Project



- 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

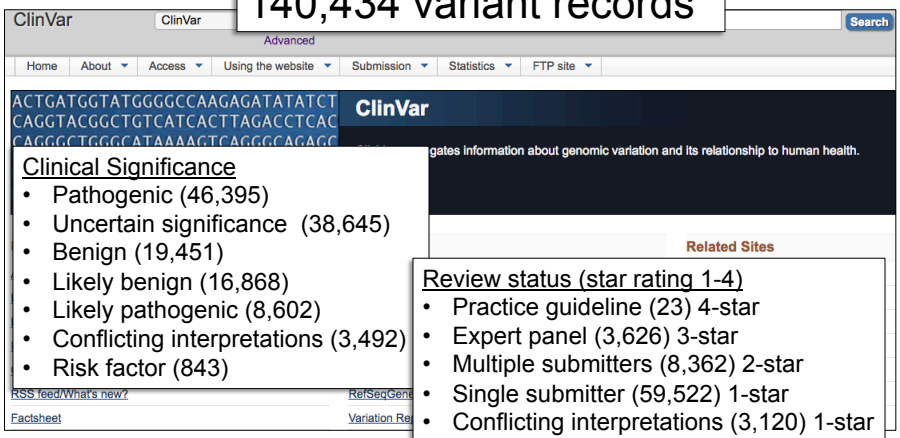
- Bridges variation and medical / clinical resources
- Accepts submission of small and large (copy number) variants and phenotypic assertions
- Provides variant-condition accessions
  - RCV – aggregate
  - SCV – individual submissions
- Represent variants using HGVS and reference sequences including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

<ncbi>/clinvar/

NCBI Webinars

ClinVar Statistics

140,434 variant records

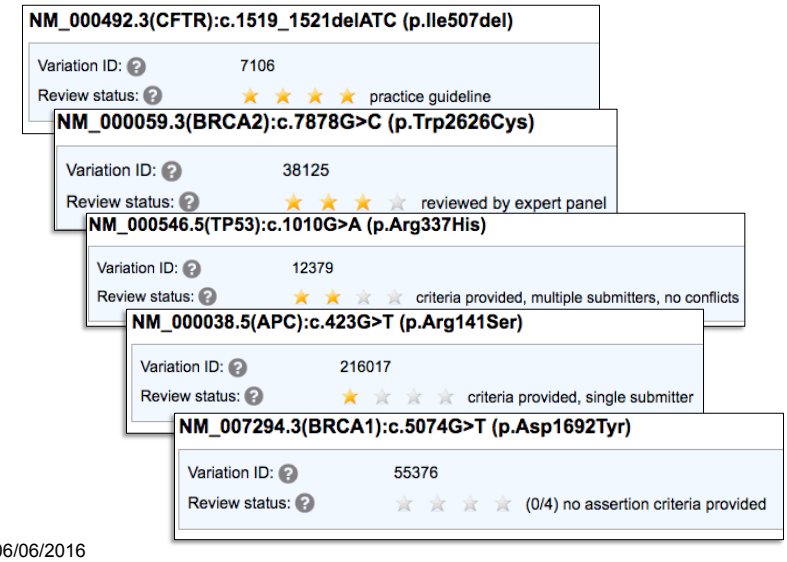


Clinical Significance
- Pathogenic (46,395)
- Uncertain significance (38,645)
- Benign (19,451)
- Likely benign (16,868)
- Likely pathogenic (8,602)
- Conflicting interpretations (3,492)
- Risk factor (843)

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

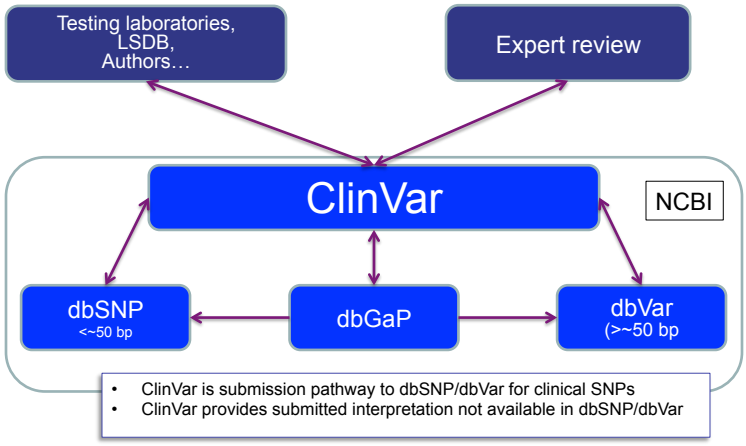
NCBI Webinars

Review Status



NCBI Webinars

ClinVar Data Relationships



NCBI Webinars

NCBI Webinars

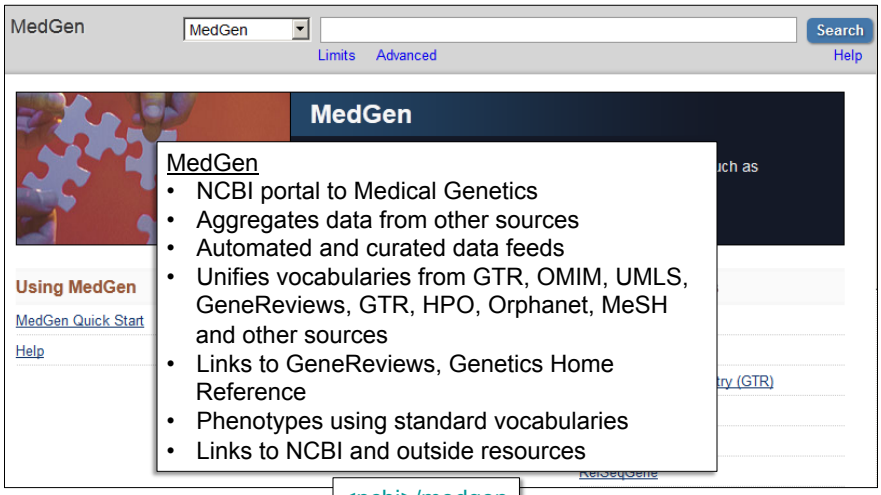
Medical Genetics Resources

Medical Genetics Resources

NCBI Webinars

- 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

MedGen

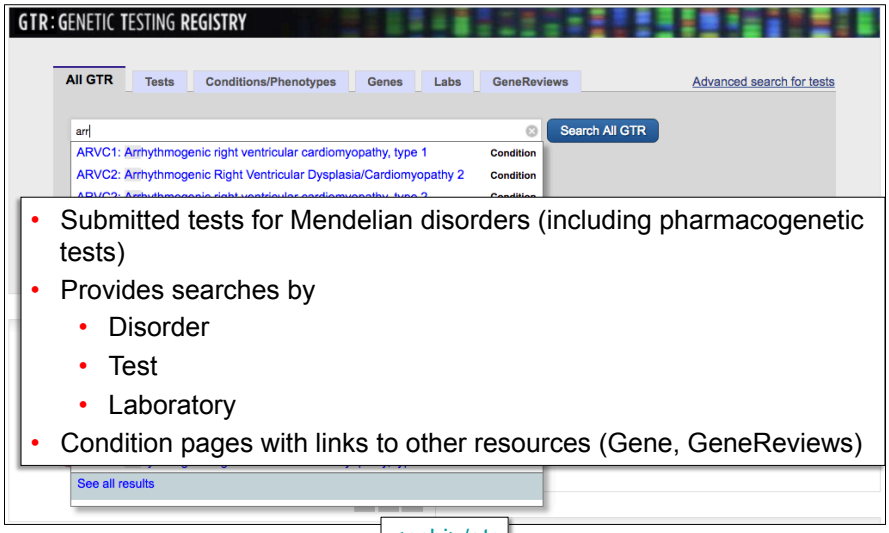


MedGen
- NCBI portal to Medical Genetics
- Aggregates data from other sources
- Automated and curated data feeds
- Unifies vocabularies from GTR, OMIM, UMLS, GeneReviews, GTR, HPO, Orphanet, MeSH and other sources
- Links to GeneReviews, Genetics Home Reference
- Phenotypes using standard vocabularies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

<ncbi>/medgen

NCBI Webinars

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)

<ncbi>/gtr

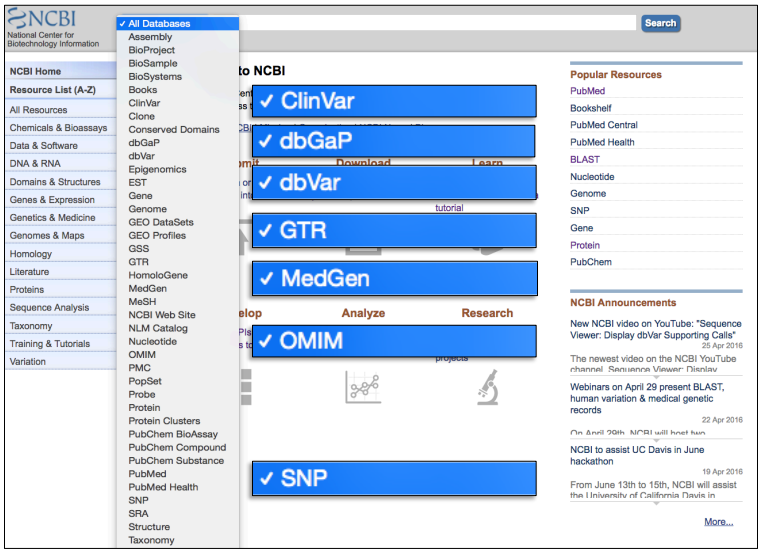
NCBI Webinars

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

Accessing Variation Data

The Entrez System



All Databases (Entrez)

Search NCBI databases

right ventricular cardiomyopathy Search

Results found in 28 databases for "right ventricular cardiomyopathy"

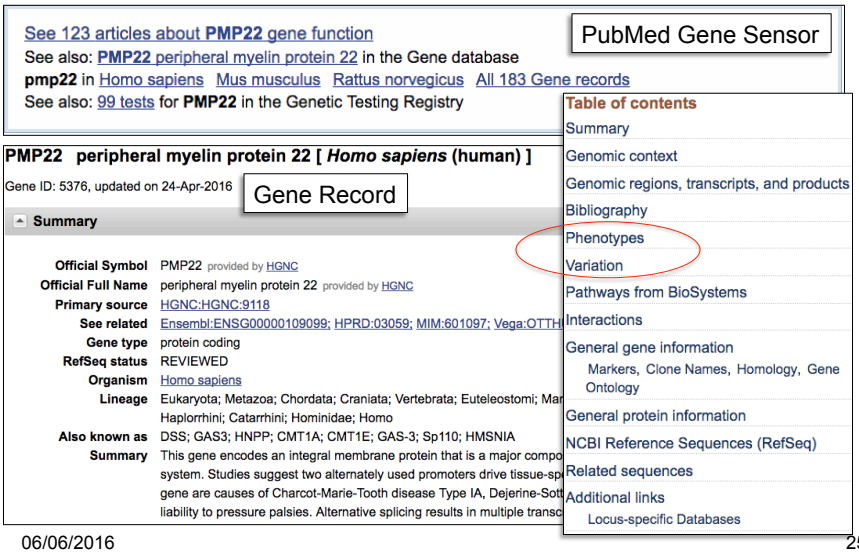
Health

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

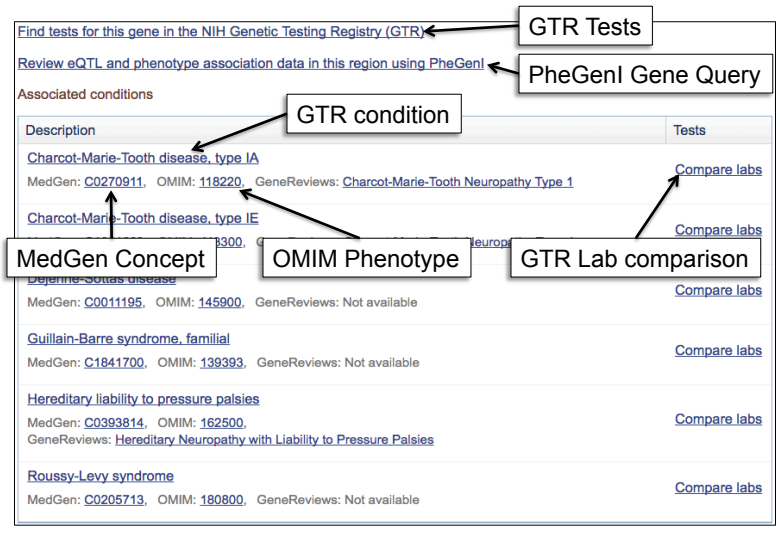
Genomes

SNP	934	short genetic variations
dbVar	15	genome structural variation studies

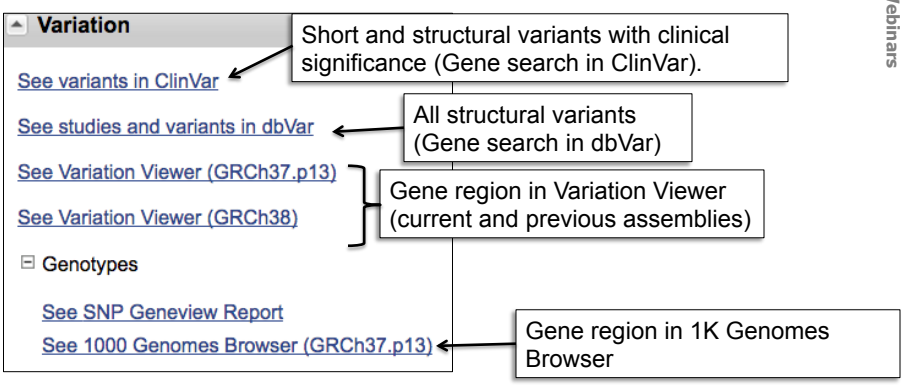
Linking from Gene



Phenotypes: conditions

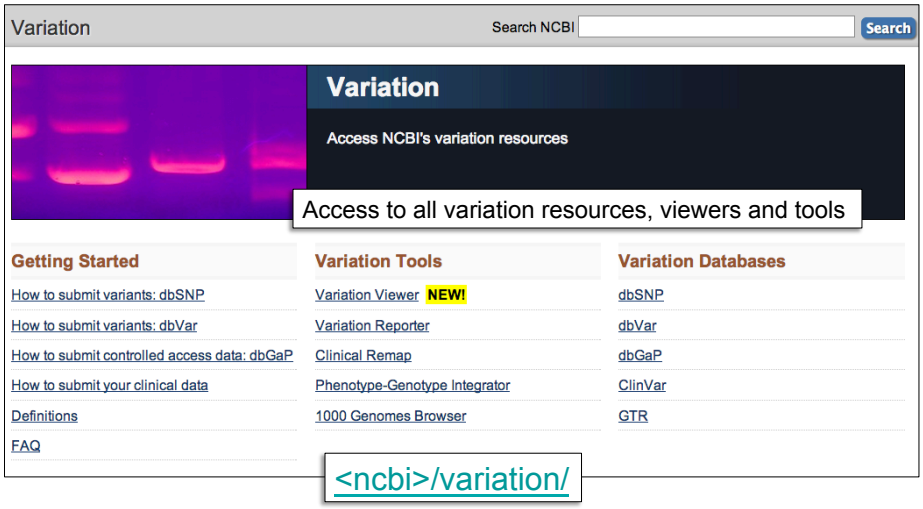


Variation



NCBI Webinars

The Variation Portal



NCBI Webinars

Variation Tools and Viewers

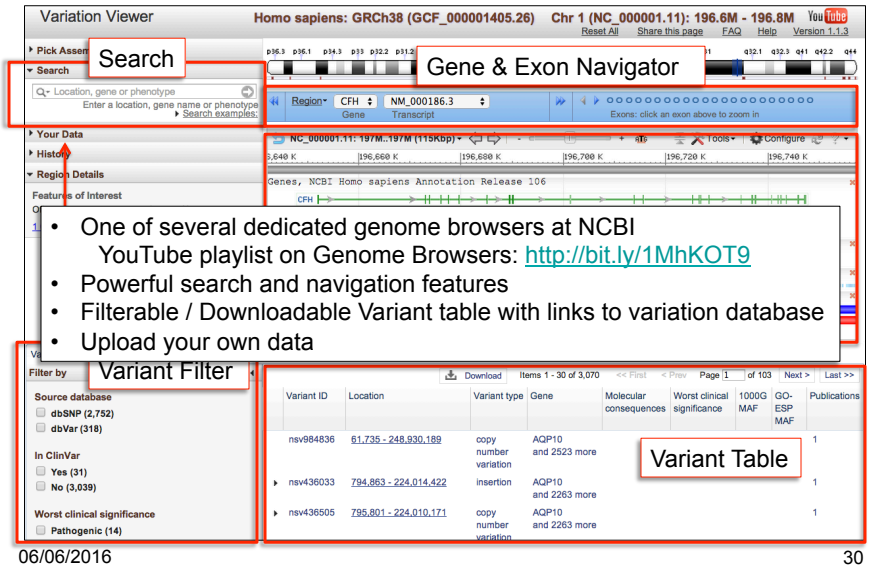
- Variation Viewer
  - Integrated genome browser for SNPs and structural variants with filter, download options and ability to upload variants
- 1000 Genomes Browser
  - Dedicated genome browser with Sequence alignments and variant calls
- Variation Reporter
  - Identifies known variants and functional 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/

NCBI Webinars

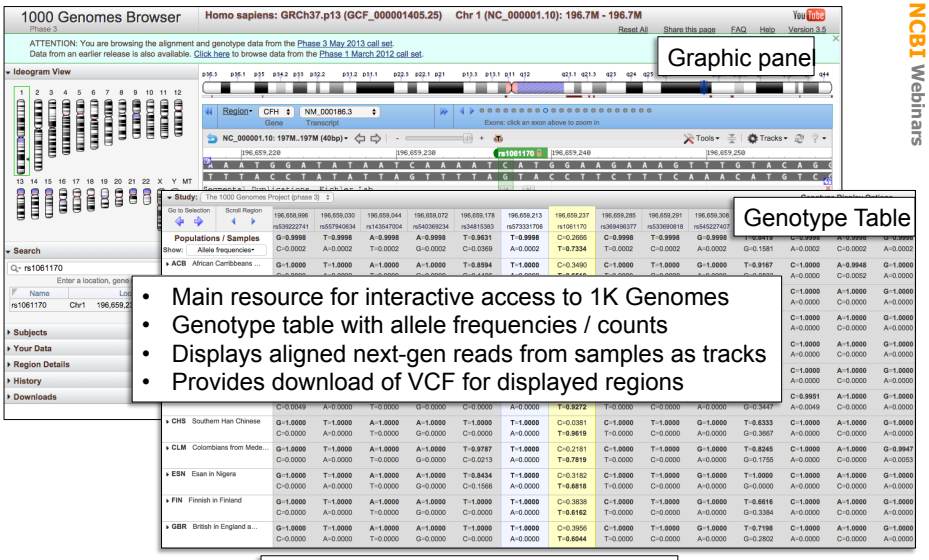
The Variation Viewer Components



- 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

NCBI Webinars

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

NCBI Webinars

www.ncbi.nlm.nih.gov