

Exploring the Many Possible Futures of Data Science

Making the Transition from Sharing Data to Sharing Knowledge

Ben Busby, Ph.D.

Genomics Outreach Coordinator

NCBI

ben.busby@nih.gov



Search NCBI databases

[Help](#)

Results found in 38 databases for "all[sb]"

Literature

Books	543,504	books and reports
MeSH	266,733	ontology used for PubMed indexing
NLM Catalog	1,555,942	books, journals and more in the NLM Collections
PubMed	26,734,411	scientific & medical abstracts/citations
PubMed Central	4,179,853	full-text journal articles

Health

ClinVar	173,675	human variations of clinical significance
dbGaP	225,011	genotype/phenotype interaction studies
GTR	48,738	genetic testing registry
MedGen	293,754	medical genetics literature and links
OMIM	24,895	online mendelian inheritance in man
PubMed Health	63,536	clinical effectiveness, disease and drug reports

Genomes

Assembly	102,316	genome assembly information
BioProject	207,505	biological projects providing data to NCBI
BioSample	5,568,573	descriptions of biological source materials
Clone	38,170,166	genomic and cDNA clones
dbVar	6,206,480	genome structural variation studies
Genome	21,144	genome sequencing projects by organism
GSS	39,765,380	genome survey sequences
Nucleotide	222,391,803	DNA and RNA sequences
Probe	32,405,068	sequence-based probes and primers
SNP	825,828,843	short genetic variations
SRA	3,481,910	high-throughput DNA and RNA sequence read archive
Taxonomy	1,644,293	taxonomic classification and nomenclature catalog

Genes

EST	76,324,331	expressed sequence tag sequences
Gene	26,043,141	collected information about gene loci
GEO DataSets	2,110,951	functional genomics studies
GEO Profiles	128,414,055	gene expression and molecular abundance profiles
HomoloGene	141,268	homologous gene sets for selected organisms
PopSet	262,192	sequence sets from phylogenetic and population studies
UniGene	6,473,284	clusters of expressed transcripts

Proteins

Conserved Domains	52,411	conserved protein domains
Protein	342,326,582	protein sequences
Protein Clusters	820,546	sequence similarity-based protein clusters
Structure	124,173	experimentally-determined biomolecular structures

Chemicals

BioSystems	932,719	molecular pathways with links to genes, proteins and chemicals
PubChem BioAssay	1,218,723	bioactivity screening studies
PubChem Compound	92,574,428	chemical information with structures, information and links
PubChem Substance	225,315,243	deposited substance and chemical information

Scientific Literature, and links to relevant datasets!

NCBI

Resources

How To

PubMed.gov

US National Library of Medicine
National Institutes of Health

PubMed

NF2

Create RSS

Create alert

Advanced

Article types

Clinical Trial

Review

Customize ...

Text availability

Abstract

Free full text

Full text

PubMed Commons

Reader comments

Trending articles

Publication dates

5 years

10 years

Custom range...

Species

Humans

Other Animals

Clear all

Show additional filters

Format: Summary

Sort by: Most Recent

Per page

See 229 articles about **NF2** gene function

See also: **NF2** neurofibromin 2 in the Gene database

nf2 in [Homo sapiens](#) [Mus musculus](#) [Rattus norvegicus](#)

See also: [89 tests](#) for **NF2** in the Genetic Testing Registry

Search results

Items: 1 to 20 of 1735

☐

[Cancer and Central Nervous System Tumor Susceptibility Related Disorders.](#)

Evans DGR, Salvador H, Chang VY, Erez A, Vos
Clin Cancer Res. 2017 Jun 15;23(12):e54-e61. doi: 10.1158/1078-0432.CCR-17-0005
PMID: 28620005
[Similar articles](#)

☐

[Cancer and Central Nervous System Tumor Susceptibility Related Disorders.](#)

Evans DGR, Salvador H, Chang VY, Erez A, Vos
Clin Cancer Res. 2017 Jun 15;23(12):e46-e53. doi: 10.1158/1078-0432.CCR-17-0004
PMID: 28620004
[Similar articles](#)

NCBI

Resources

How To

PMC

US National Library of Medicine
National Institutes of Health

PMC

Journal List

Advanced

PMC

PubMed Central® (PMC) is a free full-text archive of journal literature at the U.S. National Institutes of Health (NIH/NLM).

Get Started

[PMC Overview](#)

[Users' Guide](#)

[Journal List](#)

[PMC FAQs](#)

[PMC Copyright Notice](#)

Other Resources

[PMC International](#)

Participate

[Information for Publishers](#)

[Add a Journal to PMC](#)

[Participation Agreements](#)

[File Submission Specifications](#)

[File Validation Tools](#)

4.3 MILLION Articles

are archived in PMC.

EUtils (Search API) Command Line

EDirect

[s://github.com/NCBI-Hackathons/EDirect_EUtils_API_Cookbook](https://github.com/NCBI-Hackathons/EDirect_EUtils_API_Cookbook)

EDirect Scripts

Gene Aliases

Description (optional):

Written by: NCBI Folks (12/14/2016)

Confirmed by:

Databases: gene

```
esearch -db gene -query "Liver cancer AND Homo sapiens" | \
efetch -format docsum | \
xtract -pattern DocumentSummary -element Name OtherAliases OtherDesignations
```

Genomic sequence fastas from RefSeq assembly for specified taxonomic designation

Description (optional):

Written by: NCBI Folks (12/14/2016)

Confirmed by: Peter Cooper (NCBI) and Wayne Matten (NCBI) (12/29/2016, v6.00)

Databases: assembly

```
wget `esearch -db assembly -query "Leptospira alstonii[ORGN] AND latest[SB]" | \
efetch -format docsum | \
xtract -pattern DocumentSummary -element FtpPath_RefSeq | \
awk -F"/" '{print $0/"$NF"_genomic.fna.gz"}``
```

(For larger sets of data the above may fail as wget may not accept a very large number of arguments.
The command below should work for all.)

Gene and Protein Information

NCBI Resources ▾ How To ▾

Gene

Gene ▾

Advanced

Full Report ▾

Send to: ▾

NF2 neurofibromin 2 [*Homo sapiens* (human)]

Gene ID: 4771, updated on 8-Jun-2017

Summary

Official Symbol	NF2 provided by HGNC
Official Full Name	neurofibromin 2 provided by HGNC
Primary source	HGNC:HGNC:7773
See related	Ensembl:ENSG00000186575 MIM:607379 ; Vega:OTTHUMG00000030727
Gene type	protein coding
RefSeq status	REVIEWED
Organism	Homo sapiens
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as	ACN; SCH; BANF
Summary	This gene encodes a protein that is similar to some members of the ERM (ezrin, radixin, moesin) family of proteins that are thought to link cytoskeletal components with proteins in the cell membrane. This gene product has been shown to interact with cell-surface proteins, proteins involved in cytoskeletal dynamics and proteins involved in regulating ion transport. This gene is expressed at high levels during embryonic development; in adults, significant expression is found in Schwann cells, meningeal cells, lens and nerve. Mutations in this gene are associated with neurofibromatosis type II which is characterized by nervous system and skin tumors and ocular abnormalities. Two predominant isoforms and a number of minor isoforms are produced by alternatively spliced transcripts. [provided by RefSeq, Jul 2008]
Orthologs	mouse all

Genomic context



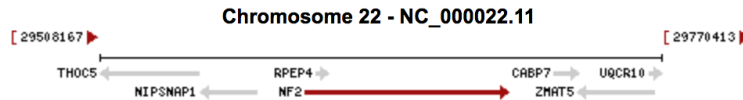
U.S. National Library of Medicine



Gene and Protein Information

Annotation release	Status	Assembly	Chr	Location
108	current	GRCh38.p7 (GCF_000001405.33)	22	NC_000022.11 (29603556..29698600)
105	previous assembly	GRCh37.p13 (GCF_000001405.25)	22	NC_000022.10 (29999545..30094589)

[Variation Viewer \(GRCh37.p13\)](#)
[Variation Viewer \(GRCh38\)](#)
[1000 Genomes Browser \(GRCh37.p13\)](#)
[Ensembl](#)
[UCSC](#)

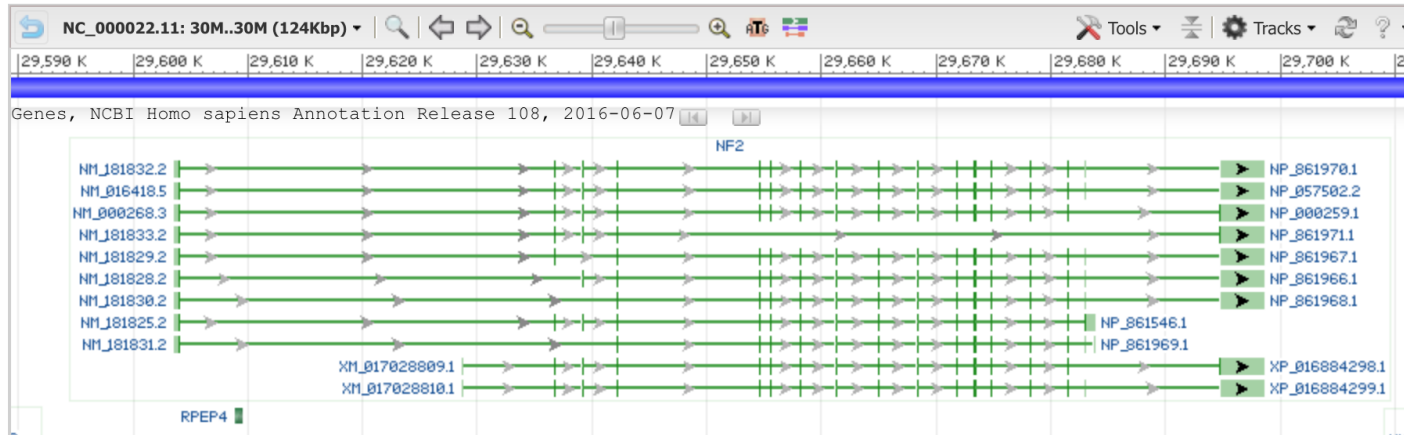


Genomic regions, transcripts, and products

Go to [reference sequence details](#)

Genomic Sequence:

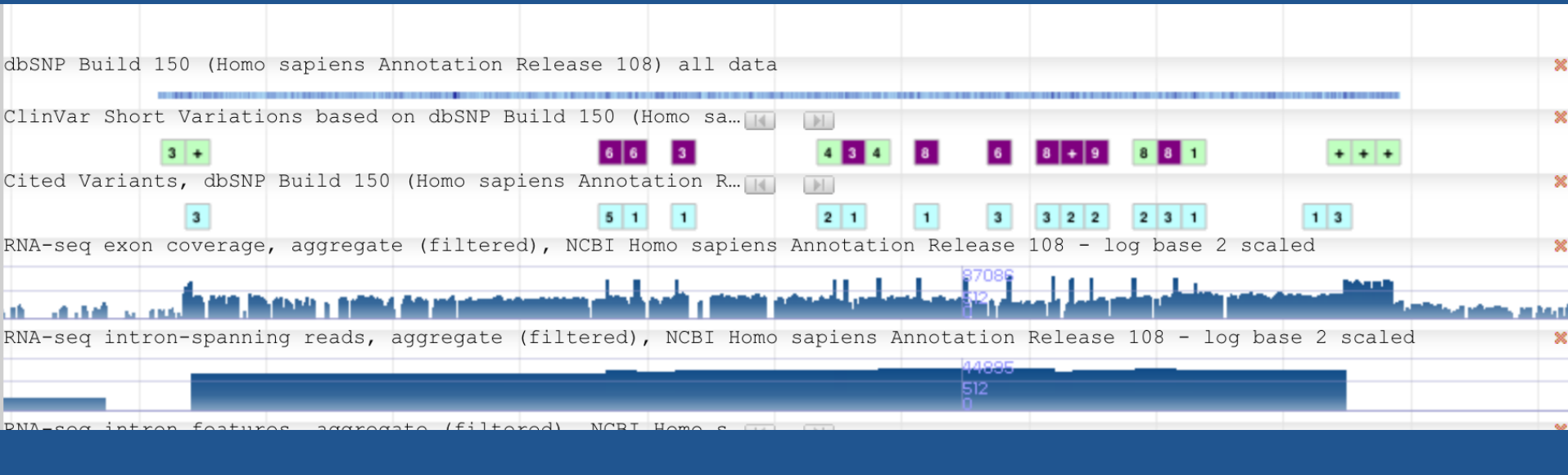
Go to nucleotide: [Graphics](#) [FASTA](#) [GenBank](#)



Related information

[Order cDNA clone](#)
[3D structures](#)
[BioAssay by Target \(List\)](#)
[BioAssay by Target \(Summary\)](#)
[BioAssay, by Gene target](#)
[BioAssays, RNAi Target, Active](#)
[BioAssays, RNAi Target, Tested](#)
[BioProjects](#)
[BioSystems](#)
[Books](#)
[CCDS](#)
[ClinVar](#)
[Conserved Domains](#)
[dbVar](#)

Gene and Protein Information



Phenotypes

[Find tests for this gene in the NIH Genetic Testing Registry \(GTR\)](#)

[Review eQTL and phenotype association data in this region using PheGenI](#)

Professional guidelines

Description

Professional guideline

ACMG 2013

The ACMG recommends that laboratories performing clinical sequencing seek and report mutations in NF2 that are pathogenic or expected to be pathogenic.

[Guideline](#), [PubMed](#)

Description

Professional guideline

ACMG 2013

The ACMG recommends that laboratories performing clinical sequencing seek and report mutations in NF2 that are pathogenic or expected to be pathogenic.

[Guideline](#), [PubMed](#)

ClinVar

ClinVar

NF2[gene]

×

Search

Create alert Advanced

Gene Tabular 100 per page Sort by Location

Customize this list...

Clinical significance

Conflicting interpretations (3)

Benign (23)

Likely benign (60)

Uncertain significance (111)

Likely pathogenic (2)

Pathogenic (40)

Risk factor (0)

Review status

Practice guideline (0)

Expert panel (0)

Multiple submitters (7)

Single submitter (191)

At least one star (201)

Conflicting interpretations (3)

Allele origin

Germline (209)

De novo (2)

Somatic (6)

Method type

Research (1)

Showing for results for variants in the NF2 gene. [Search instead for all ClinVar records that mention NF2](#)


Search results

Items: 1 to 100 of 232

	Variation Location	Gene(s)	Condition(s)	Clinical significance (Last reviewed)	Review status
<input type="checkbox"/> 1.	NM_000268.3(NF2):c.-443-? *3798+? del	NF2	Neurofibromatosis, type 2	Pathogenic (Jan 26, 2016)	criteria provided, single submitter
<input type="checkbox"/> 2.	NM_000268.3(NF2):c.-402A>G GRCh37: Chr22:29999586 GRCh38: Chr22:29603597	NF2	Neurofibromatosis, type 2	Uncertain significance (Jun 14, 2016)	criteria provided, single submitter
<input type="checkbox"/> 3.	NM_000268.3(NF2):c.-397C>T GRCh37: Chr22:29999591 GRCh38: Chr22:29603602	NF2	Neurofibromatosis, type 2	Uncertain significance (Jun 14, 2016)	criteria provided, single submitter
<input type="checkbox"/> 4.	NM_000268.3(NF2):c.-320C>T GRCh37: Chr22:29999668 GRCh38: Chr22:29603679	NF2	Neurofibromatosis, type 2	Uncertain significance (Jun 14, 2016)	criteria provided, single submitter

For more information go to:

ncbi.nlm.nih.gov/learn

 U.S. National Library of Medicine

NCBI National Center for Biotechnology Information


Sign in to NCBI

NCBI HOME LITERATURE HEALTH GENOMES GENES PROTEINS CHEMICALS POPULAR RESOURCES ▼

All Databases Search NCBI Search

Learn

NCBI creates a variety of educational products including courses, workshops, webinars, training materials and documentation. NCBI educational events are free and open to everyone. All NCBI educational materials are available for anyone to re-use and distribute.



UPCOMING EVENTS

How to upload and analyze dbGaP data in the Cloud
FEBRUARY 3, 2016
Online Webinar: 1:00-2:00pm





Five ways to submit next-gen sequence data to NCBI's Sequence Read Archive
FEBRUARY 17, 2016
Online Webinar: 1:00-2:00pm


"NCBI Resources for Patent Searchers" at the PIUG Biotechnology 2016 Conference
FEBRUARY 24, 2016
Workshop

A Librarian's Guide to NCBI
MARCH 7-11, 2016
Workshop

Experimental Biology 2016 Annual Meeting
APRIL 2-6, 2016
Conference

"Practical Bioinformatics for the Clinic" at the NLM Biomedical Informatics Course
APRIL 8, 2016
Presentation

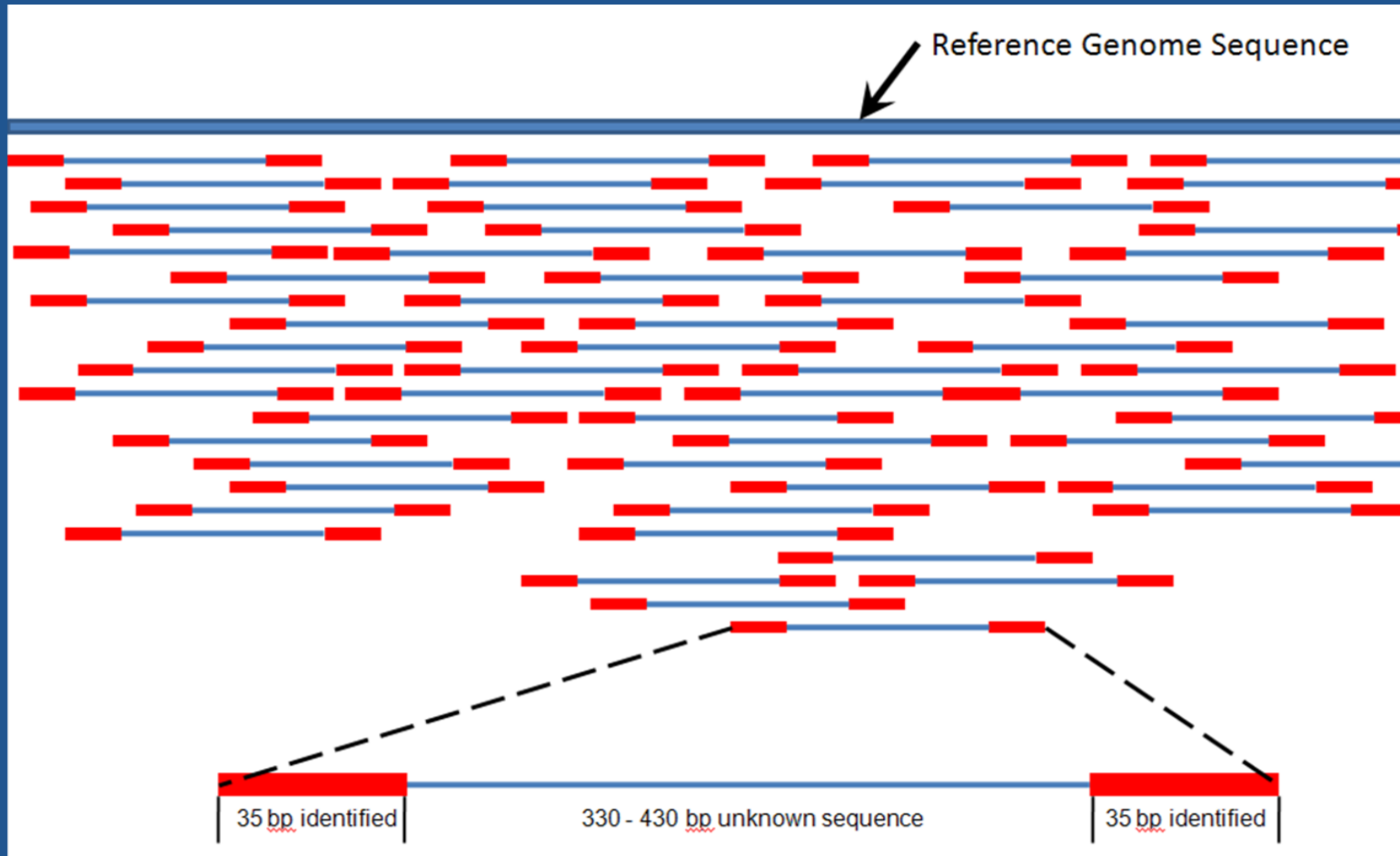
Webinars & Courses	Conferences & Presentations	Tutorials	Documentation
In-person courses, live webinars and webinar recordings	Booth exhibits and workshops at scientific conferences	Tutorials: Training materials in HTML, PDF and video formats	Online manuals, handbooks, fact sheets and FAQs
			

**News, Blog & Social Media**

Keep up with the latest NCBI news and follow NCBI on social media sites, including FaceBook, Twitter, Google+, LinkedIn and the NCBI Insights blog.

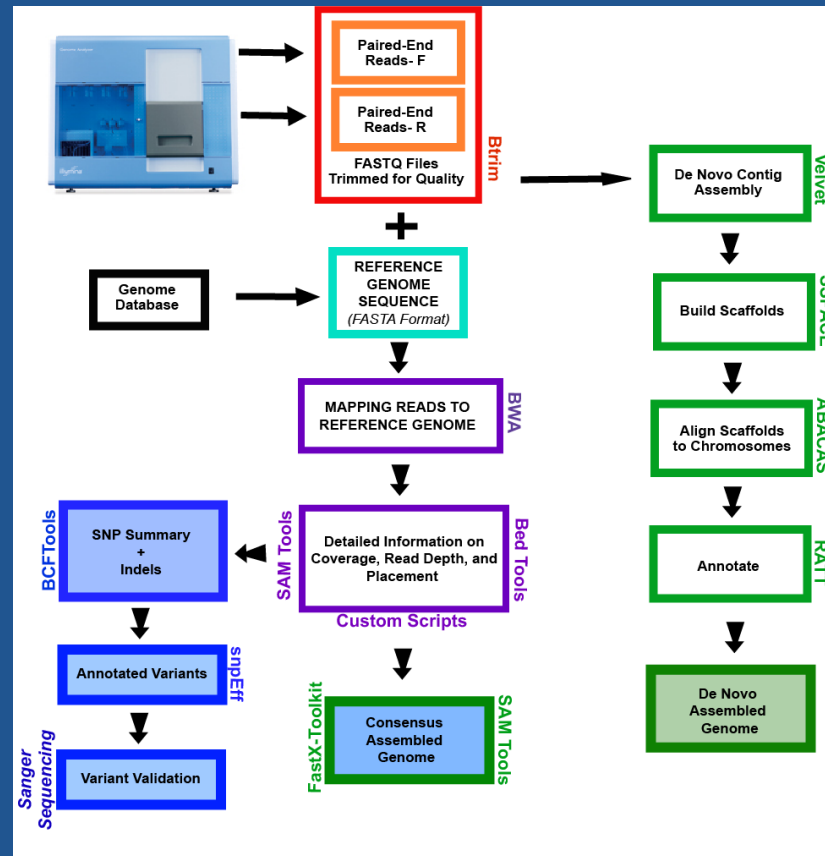
Review of terminology and concepts

Next Generation Sequencing



Review of terminology and concepts

How Genomes are Mapped and Assembled



© Martine Zilversmit 2013

Review of terminology and concepts

How Genomes are Mapped and Assembled

<http://1.usa.gov/1J1xmYs>

FASTA

who: A widely adopted simple sequence format used for protein, transcript, or genomic RNA or DNA sequences, often marked by one of the following file extensions: .fa, .fna, .faa, or .fasta.

what: In this simple flat file format, entries begin as a single descriptive line denoted by a greater-than symbol (">"), followed by a hard return, and then the sequence.

where: Most sequence records housed at NCBI can be displayed or downloaded as FASTA formatted files.

how: Records in Nucleotide or Protein databases can be changed to FASTA and pasted, or saved to a local directory using the Send To menu. Larger files, such as chromosomes, can be downloaded from FTP directories.

mapping to genomes

FASTQ

who: The most common raw (unaligned) data format for next generation sequencing marked by the file extension .fastq.

what: This data format output from sequencing platforms includes sequence data and a quality score for each position.

where: FASTQ files can be dumped from SRA using the SRA toolkit.

how: If the SRA record is aligned (a BAM was submitted), FASTQ files comprised of reads that correspond to a given genomic region can be dumped using the sam-dump utility.

mapping to genomes

BED

who: A simple file format used to define features by chromosomal positions, marked by the file extension .bed.

what: BED files are likely the simplest way to see what is where on a genome, and are most frequently used in epigenomic analyses.

where: Many datasets in this format can be found in GEO.

how: The NCBI Epigenomics browser can be used to display and analyze these datasets. Data can also be ported to UCSC.

downstream analysis

CNV

who: Copy Number Variations

what: A variation that increases or decreases the copy number of a given gene or genomic region.

where: the NCBI dbVar database:
<http://www.ncbi.nlm.nih.gov/dbvar>

how: Data can be viewed at the gene level using variation viewer, or downloaded by gene by searching the dbVar database. A complete set of CNVs for many organisms can be downloaded from:
<ftp://ftp.ncbi.nlm.nih.gov/pub/dbVar/data/>

downstream analysis

**NCBI NGS Online Workshop – Available on the
NCBI YouTube Channel!**



My View of Data Transfer Principles

- **Metadata Search**
 - **Rapid NoSQL (for now)**
 - **Integration**
 - **Non-ambiguous identifiers**
- **Transferring Small amounts of Data**
 - **Data still gets transferred in the cloud**
 - **Underlying structure**
 - **Finding specific data from validated formats**
- **Democratization of Data**
 - **Rapid comparison by domain experts**
- **Reporting**
 - **Metrics to report data upload and [unique IP] download of datasets**
 - **Post-publication User Review**



BioProject

NCBI Resources How To busbybr@ncbi.nlm.nih.gov My NCBI

BioProject BioProject tuberculosis Search

Create alert Advanced

Display Settings: Summary, 20 per page, Sorted by Default order Send to: Filters: [Manage Filters](#)

Search results

Items: 1 to 20 of 2938 << First < Prev Page 1 of 147 Next > Last >>

☐ [Nonpathogenic SIV and Pathogenic HIV Infections Associate with Disparate Innate Cytokine Signatures in Response to M. bovis BCG](#)

1. Project data type: Transcriptome or Gene expression
Scope: Multispecies
Center for Infectious Disease Research
Accession: PRJNA323449 ID: 323449

☐ [Gene expression profiling of tuberculosis patients from India](#)

2. Organism: Homo sapiens
Taxonomy: *Homo sapiens (human)*
Project data type: Transcriptome or Gene expression
Scope: Multispecies
Department of Biochemistry, Indian Institute of Science
Accession: PRJNA322492 ID: 322492

☐ [Mycobacterium tuberculosis strain: NZ494](#)

3. Mycobacterium tuberculosis strain: NZ494 Genome sequencing
Taxonomy: *Mycobacterium tuberculosis*
Project data type: Genome sequencing
Scope: Monoisolate
Trinity College Dublin
Accession: PRJNA321663 ID: 321663

☐ [Mycobacterium mungi strain: BM22813](#)

4. Mycobacterium mungi strain: BM22813 Genome sequencing
Taxonomy: *Mycobacterium mungi*
Project data type: Genome sequencing
Scope: Monoisolate
USDA Animal Plant Health Inspection Service-National Veterinary Services Laboratories - DBL

Project Types
Umbrella (32)
Primary submission (2,891)
RefSeq (15)

Data Types
Clone ends (1)
Epigenomics (11)
Genome sequencing (266)
Metagenome (1)
Other (81)
Proteome (3)
Targeted locus (4)
Transcriptome (358)
Variation (7)

Project Data
Nucleotide (1,859)
Protein (1,692)
Assembly (1,858)
SRA (2,167)
GEO DataSets (362)

Scope
Monoisolate (2,449)
Multi-isolate (433)
Multi-species (13)
Environmental (6)
Other (2)

Organism Groups
Human (95)
Archaea (1)
Bacteria (2,728)
Fungi (2)
Protists (1)
Mammals (167)
Other vertebrates (7)

Find related data
Database: Select
Find items

Search details
tuberculosis[All Fields]
Search

Recent activity
Turn Off
tuberculosis (2938)
Mutations in BCKD-kinase lead to a potentially treatable form of autism v
maple syrup urine disease (1)
SLN sarcosin [Homo sapiens]
escherichia coli mg1655 AND (latest AND all[filter] NOT a... (14)

BioProject

NCBI Resources How To busbybr@ncbi.nlm.nih.gov My NCBI

BioProject BioProject tuberculosis Search

Create alert

Project Types
Umbrella (32)
Primary submission (2,891)
RefSeq (15)

Data Types
Clone ends (1)
Epigenomics (11)
Genome sequencing (266)
Metagenome (1)
Other (81)
Proteome (3)
Targeted locus (4)
Transcriptome (358)
Variation (7)

Project Data
Nucleotide (1,859)
Protein (1,692)
Assembly (1,858)
SRA (2,167)
GEO DataSets (362)

Scope
Monoisolate (2,449)
Multi-isolate (433)
Multi-species (13)
Environmental (6)
Other (2)

Organism Groups
Human (95)
Archaea (1)
Bacteria (2,728)
Fungi (2)
Protists (1)
Mammals (167)
Other vertebrates (7)

Display Settings: Summary

Search results
Items: 1 to 20 of 2938

☐ [Nonpathogenic SIV and Signatures in Response](#)
Project data type: Transcriptome
Scope: Multispecies
Center for Infectious Disease
Accession: PRJNA323449

☐ [Gene expression profiling](#)
Organism: Homo sapiens
Taxonomy: [Homo sapiens](#)
Project data type: Transcriptome
Scope: Multiisolate
Department of Biochemistry
Accession: PRJNA322492

☐ [Mycobacterium tuberculosis](#)
Mycobacterium tuberculosis
Taxonomy: [Mycobacterium tuberculosis](#)
Project data type: Genome
Scope: Monoisolate
Trinity College Dublin
Accession: PRJNA321663

☐ [Mycobacterium mungii](#)
Mycobacterium mungii
Taxonomy: [Mycobacterium mungii](#)
Project data type: Genome
Scope: Monoisolate
USDA Animal Plant Health

Data Types
Clone ends (1)
Epigenomics (11)
Genome sequencing (266)
Metagenome (1)
Other (81)
Proteome (3)
Targeted locus (4)
Transcriptome (358)
Variation (7)

Project Data
Nucleotide (1,859)
Protein (1,692)
Assembly (1,858)
SRA (2,167)
GEO DataSets (362)

Scope
Monoisolate (2,449)
Multi-isolate (433)
Multi-species (13)
Environmental (6)
Other (2)

of 147 Next > Last >>

Send to: Filters: [Manage Filters](#)

Find related data
Database: Select
Find items

Search details
tuberculosis[All Fields]
Search

Recent activity
Turn On

tuberculosis (2938)

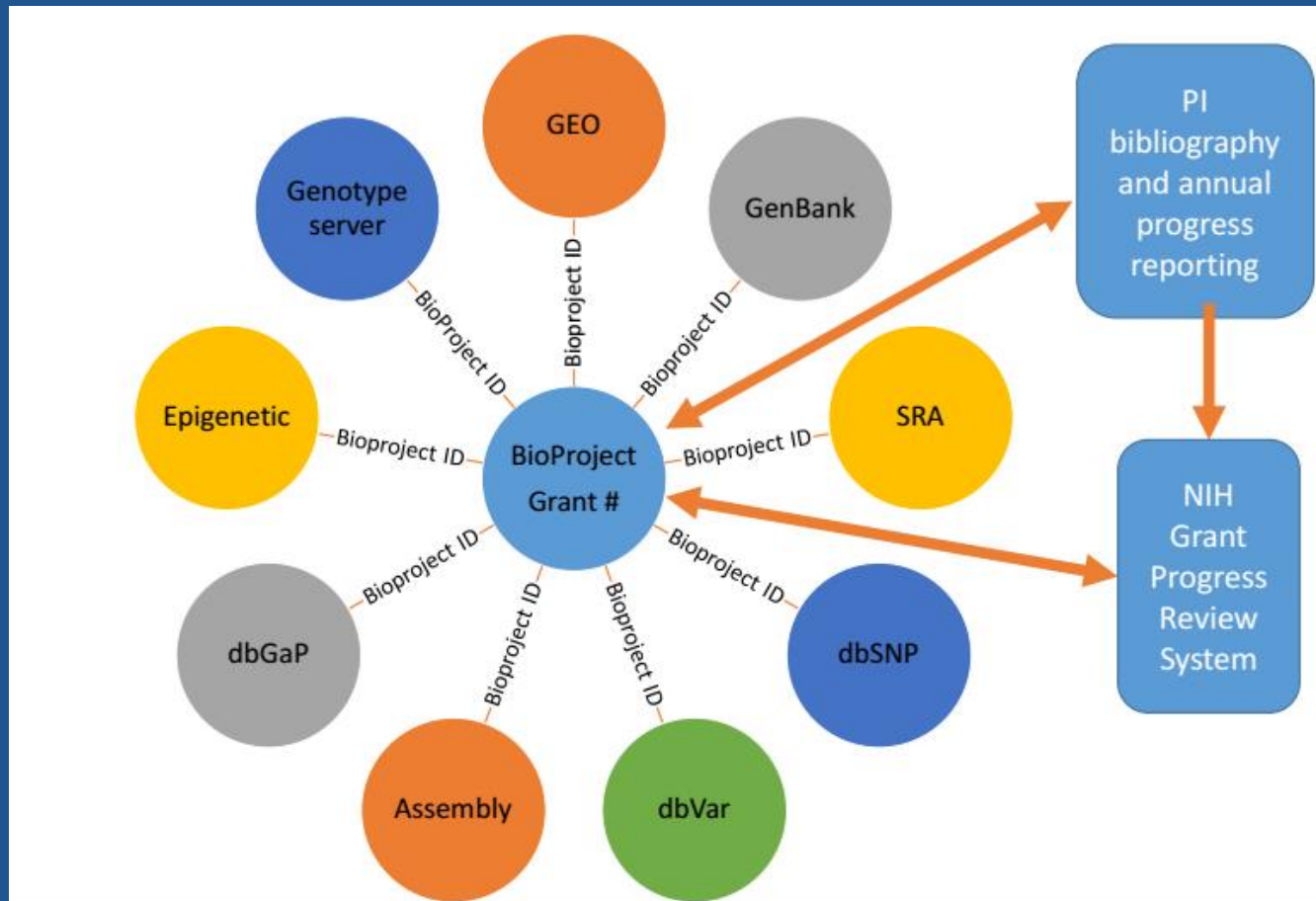
Mutations in BCKD-kinase lead to a potentially treatable form of autism v

maple syrup urine disease (1)

SLN sarcosin [Homo sapiens]

escherichia coli mg1655 AND (latest AND all[filter] NOT a... (14)

Reporting



BioSample

NCBI Resources How To

BioSample

BioSample

Advanced

Full

Send to

S_1104

Identifiers BioSample: SAMN05242633; SRA: SRS1501741; GEO: [GSM2199532](#)

Organism [Homo sapiens](#) (human)
cellular organisms; Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Dipnotetrapodomorpha; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Boreoeutheria; Euarchontoglires; Primates; Haplorrhini; Simiiformes; Catarrhini; Hominoidea; Hominidae; Homininae; Homo

Attributes

source name	Subcutaneous fat tissue of obese patient
tissue	Subcutaneous fat
age	27
body mass index	44.9

Links [GEO Sample GSM2199532](#)

BioProject [PRJNA523592](#) Next-Generation Sequencing of liver and subcutaneous fat tissues obtained from obese subjects
Retrieve [all samples](#) from this project

Submission Systems Biology, Department of Biology and Biological Engineering, Chalmers University of Technology, Adil Mardinoglu; 2016-06-14

Accession: SAMN05242633 ID: 5242633

[BioProject](#) [SRA](#) [GEO DataSets](#)

BioSample

Submission Portal

Preview BioSample types and attributes

This page provides a preview of the sample attributes that submitters are asked to supply during the submission process. After selecting the relevant Sample type, use the **Download** button to download the submission template, or the **Definition** button to review the attribute definitions and formats.

✳ Select the package that best describes your samples:

☐ **Pathogen affecting public health**

Use for pathogen samples that are relevant to public health. Required attributes include those considered useful for the rapid analysis and trace back of pathogens.

☐ **Microbe**

Use for bacteria or other unicellular microbes when it is not appropriate or advantageous to use MlxS, Pathogen or Virus packages.

☐ **Model organism or animal sample**

Use for multicellular samples or cell lines derived from common laboratory model organisms, e.g., mouse, rat, Drosophila, worm, fish, frog, or large mammals including zoo and farm animals.

☐ **Metagenome or environmental sample**

Use for metagenomic and environmental samples when it is not appropriate or advantageous to use MlxS packages.

BioSample

This is a submission template for batch deposit of 'Human; version 1.0' samples to the NCBI BioSample database (<http://www.ncbi.nlm.nih.gov/biosample/>).

GREEN fields are mandatory. Your submission will fail if any mandatory fields are not completed. If information is unavailable for any mandatory field, please enter 'not collected', 'not applicable' or 'missing' as appropriate.

YELLOW fields are optional. Leave optional fields empty (or delete them) if no information is available.

You can add any number of custom fields to fully describe your BioSamples, simply include them in the table.

Hover over field name to view definition, or see <http://www.ncbi.nlm.nih.gov/biosample/docs/attributes/>.

CAUTION: Be aware that Excel may automatically apply formatting to your data. In particular, take care with dates, incrementing autofills and special characters like / or -. Doublecheck that your text file is accurate before submitting.

TO MAKE A SUBMISSION:

1. Complete this template table
2. Save the worksheet as a Text (Tab-delimited) file -- (use 'File, Save as, Save as type: Text (Tab-delimited)')
3. Upload the text file on the 'Attributes' tab of the BioSample Submission Portal at <https://submit.ncbi.nlm.nih.gov/subs/biosample/>.

If you have any questions, please contact us at biosamplehelp@ncbi.nlm.nih.gov.

*sample_name	sample_title	bioproject_accession	*organism	*isolate	*age	*biomaterial_provider	*sex	*tissue	cell_line	cell_subtype	cell_type
--------------	--------------	----------------------	-----------	----------	------	-----------------------	------	---------	-----------	--------------	-----------



SRA

SRA ▾

melanoma



Search

[Create alert](#) [Advanced](#)[Help](#)

Access

Controlled (1,112)

Public (1,443)

Source

DNA (1,786)

RNA (1,070)

metagenomic (67)

Type

exome (1,351)

genome (77)

Other

aligned data (1,673)

[Clear all](#)[Show additional filters](#)

Summary ▾ 20 per page ▾

Send to: ▾

Filters: [Manage Filters](#)

Search results

Items: 1 to 20 of 2870

<< First < Prev Page 1 of 144 Next > Last >>

☐ [Ion Torrent PGM sequencing](#)

1. Accession: ERX1326148

☐ [Ion Torrent PGM sequencing](#)

2. 1 ION_TORRENT (Ion Torrent PGM) run: 1M spots, 108.9M bases, 79.9Mb downloads

Accession: ERX1326147

☐ [Ion Torrent PGM sequencing](#)

3. 1 ION_TORRENT (Ion Torrent PGM) run: 828,549 spots, 99.1M bases, 68.5Mb downloads

Accession: ERX1326142

☐ [454 GS FLX Titanium sequencing](#)

4. 1 LS454 (454 GS FLX Titanium) run: 17,894 spots, 9.9M bases, 21Mb downloads

Accession: ERX614403

☐ [454 GS FLX Titanium sequencing](#)

5. 1 LS454 (454 GS FLX Titanium) run: 18,522 spots, 10.3M bases, 21.7Mb downloads

Accession: ERX614402

☐ [454 GS FLX Titanium sequencing](#)

6. 1 LS454 (454 GS FLX Titanium) run: 5,267 spots, 2.5M bases, 5.7Mb downloads

Accession: ERX614401

☐ [454 GS FLX Titanium sequencing](#)

7. 1 LS454 (454 GS FLX Titanium) run: 23,843 spots, 12.9M bases, 27.9Mb downloads

Accession: ERX614400

Results by taxon

Top Organisms [\[Tree\]](#)

Homo sapiens (1938)

Mus musculus (737)

Canis lupus familiaris (104)

human skin metagenome (64)

Homo (1950)

All other taxa (15)

[More...](#)

Top Bioprojects

Production ENCODE epigenomic... (4)

Search in related databases

Database	Access		all
	public	controlled	
BioSample	782	1,094	1,876
BioProject	98	6	104
dbGaP		15	15
GEO Datasets	501		501

Find related data

Database: [Select](#) ▾[Find items](#)

SRA

SRA ▾

melanoma



Search

Create alert Advanced

Help

Access

Controlled (1,112)
Public (1,443)

Source

DNA (1,786)
RNA (1,070)
metagenomic (67)

Type

exome (1,351)
genome (77)

Other

aligned data (1,673)

[Clear all](#)

[Show additional filters](#)

Summary ▾ 20 per page ▾

Send to: ▾

Filters: [Manage Filters](#)

Search results

Items: 1 to 20 of 2870

Page 1 of 144

[Next >](#)

[Last >>](#)

☐ [Ion Torrent PGM sequencing](#)

1. Accession: ERX1326148

☐ [Ion Torrent PGM sequencing](#)

2. 1 ION_TORRENT (Ion Torrent PGM) run: 1M
Accession: ERX1326147

☐ [Ion Torrent PGM sequencing](#)

3. 1 ION_TORRENT (Ion Torrent PGM) run: 82
Accession: ERX1326142

☐ [454 GS FLX Titanium sequencing](#)

4. 1 LS454 (454 GS FLX Titanium) run: 17,894
Accession: ERX614403

☐ [454 GS FLX Titanium sequencing](#)

5. 1 LS454 (454 GS FLX Titanium) run: 18,522 spots, 10.3M bases, 21.7Mb downloads
Accession: ERX614402

☐ [454 GS FLX Titanium sequencing](#)

6. 1 LS454 (454 GS FLX Titanium) run: 5,267 spots, 2.5M bases, 5.7Mb downloads
Accession: ERX614401

☐ [454 GS FLX Titanium sequencing](#)

7. 1 LS454 (454 GS FLX Titanium) run: 23,843 spots, 12.9M bases, 27.9Mb downloads
Accession: ERX614400

Access

Controlled (1,112)
Public (1,443)

Source

DNA (1,786)
RNA (1,070)
metagenomic (67)

Type

exome (1,351)
genome (77)

Other

aligned data (1,673)

[Clear all](#)

[Show additional filters](#)

Results by taxon

Top Organisms [\[Tree\]](#)

Homo sapiens (1938)
Mus musculus (737)
Canis lupus familiaris (104)
human skin metagenome (64)
Homo (1950)
All other taxa (15)

[More...](#)

Top Bioprojects

Production ENCODE epigenomic... (4)

Search in related databases


Database	Access		all
	public	controlled	
BioSample	782	1,094	1,876
BioProject	98	6	104
dbGaP		15	15
GEO Datasets	501		501

Find related data

Database: [Select](#) ▾

[Find items](#)

SRA

SRA melanoma 

Search

[Create alert](#) [Advanced](#)

Access

 Public (714)

Source

 RNA (714)

Other

aligned data (285)

[Clear all](#)[Show additional filters](#)

clear

Summary  20 per page Send to: Filters: [Manage Filters](#)

Search results

Items: 1 to 20 of 714

<< First < Prev P

 Filters activated: Public, RNA. [Clear all](#) to show 2870 items.

- ☐ [GSM1953868: zebrafish crestin EGFP neg; Danio rerio; RNA-Seq](#)
1. 1 ILLUMINA (Illumina HiSeq 2500) run: 4.6M spots, 929.8M bases, 680.1Mb downloads
Accession: SRX1451254
- ☐ [GSM1953867: zebrafish crestin EGFP + 15 ss; Danio rerio; RNA-Seq](#)
2. 1 ILLUMINA (Illumina HiSeq 2500) run: 3.9M spots, 786.2M bases, 575.4Mb downloads
Accession: SRX1451253
- ☐ [GSM1953866: A375 Rep 2; Homo sapiens; RNA-Seq](#)
3. 1 ILLUMINA (Illumina HiSeq 2500) run: 26M spots, 5.3G bases, 3.2Gb downloads
Accession: SRX1451252
- ☐ [GSM1953865: A375 Rep 1; Homo sapiens; RNA-Seq](#)
4. 1 ILLUMINA (Illumina HiSeq 2500) run: 25.9M spots, 5.2G bases, 3.1Gb downloads
Accession: SRX1451251
- ☐ [GSM1953864: RF9 - Hema-LP; Homo sapiens; RNA-Seq](#)
5. 1 ILLUMINA (Illumina HiSeq 2500) run: 30.3M spots, 6.1G bases, 2.7Gb downloads
Accession: SRX1451250

Choose Destination

- ☐ File
- ☐ Clipboard
- ☐ Collections
- ☐ BLAST
- ☒ Run Selector

Send whole recordset to Run Selector


Go

Production ENCODE epig

Search in related data

Database	public
BioSample	76
BioProject	9
dbGaP	
GEO Datasets	50

Find related data

Database:  Select

Search:

Facets

- ☐ Run
- ☐ Study
- ☐ BioProject
- ☐ BioSample
- ☐ Sample name
- ☒ Center
- ☐ Library name
- ☐ MBases
- ☐ MBytes
- ☐ ArrayExpress-Age

Center

- ☒ geo [17]
- ☒ sc [264]
- ☐ university of regensburg [4]

Hide common fields

Assay Type: RNA-Seq
Consent: public
LibrarySource: TRANSCRIPTOMIC
Platform: ILLUMINA

	Runs	Bytes	Bases	Download	
Total:	285	150.18 Gb	277.32 G	RunInfo Table	Accession List
Selected:				RunInfo Table	Accession List

285 Runs found

Page: 1 2 3 4 5 6

Run	Study	BioProject	BioSample	Sample name	Center	Library name	MBases	MBytes	ArrayExpress-Age	ArrayExpress-Dose
<input type="checkbox"/> SRR1556217	SRP045711	PRJNA259212	SAMN03002727	GSM1484318	GEO	<not provided>	7,731	4,016	<not provided>	<not provided>
<input type="checkbox"/> SRR1556218	SRP045711	PRJNA259212	SAMN03002728	GSM1484319	GEO	<not provided>	8,710	4,507	<not provided>	<not provided>
<input type="checkbox"/> SRR1556219	SRP045711	PRJNA259212	SAMN03002729	GSM1484320	GEO	<not provided>	6,398	3,412	<not provided>	<not provided>
<input type="checkbox"/> SRR1556220	SRP045711	PRJNA259212	SAMN03002730	GSM1484321	GEO	<not provided>	12,201	6,734	<not provided>	<not provided>
<input type="checkbox"/> SRR1556221	SRP045711	PRJNA259212	SAMN03002731	GSM1484322	GEO	<not provided>	6,641	3,792	<not provided>	<not provided>
<input type="checkbox"/> SRR1556222	SRP045711	PRJNA259212	SAMN03002732	GSM1484323	GEO	<not provided>	7,706	4,064	<not provided>	<not provided>
<input type="checkbox"/> SRR1556223	SRP045711	PRJNA259212	SAMN03002733	GSM1484324	GEO	<not provided>	7,917	4,188	<not provided>	<not provided>
<input type="checkbox"/> SRR1556224	SRP045711	PRJNA259212	SAMN03002734	GSM1484325	GEO	<not provided>	6,095	3,145	<not provided>	<not provided>
<input type="checkbox"/> SRR1556225	SRP045711	PRJNA259212	SAMN03002735	GSM1484326	GEO	<not provided>	6,784	3,849	<not provided>	<not provided>
<input type="checkbox"/> SRR1556226	SRP045711	PRJNA259212	SAMN03002736	GSM1484327	GEO	<not provided>	5,899	3,485	<not provided>	<not provided>
<input type="checkbox"/> SRR1556227	SRP045711	PRJNA259212	SAMN03002737	GSM1484328	GEO	<not provided>	7,708	4,509	<not provided>	<not provided>
<input type="checkbox"/> SRR1556228	SRP045711	PRJNA259212	SAMN03002738	GSM1484329	GEO	<not provided>	327	122	<not provided>	<not provided>

Experiment	InsertSize	LibraryLayout	LibrarySelection	LoadDate	Organism	ReleaseDate	SRA Sample	Sample Description	cell type	source name
SRX685301	0 SINGLE	cDNA		2014-08-22	Homo sapiens	2015-03-05	SRX685301	<not provided>	low passage primary melanoma cultures	low passage primary melanoma cultures
SRX685302	0 SINGLE	cDNA		2016-02-02	Homo sapiens	2015-03-05	SRX685302	<not provided>	low passage primary melanoma cultures	low passage primary melanoma cultures
SRX685303	0 SINGLE	cDNA		2016-02-02	Homo sapiens	2015-03-05	SRX685303	<not provided>	low passage primary melanoma cultures	low passage primary melanoma cultures
SRX685304	0 SINGLE	cDNA		2014-08-22	Homo sapiens	2015-03-05	SRX685304	<not provided>	low passage primary melanoma cultures	low passage primary melanoma cultures
SRX685305	0 SINGLE	cDNA		2014-09-08	Homo sapiens	2015-03-05	SRX685305	<not provided>	low passage primary melanoma cultures	low passage primary melanoma cultures
SRX685306	0 SINGLE	cDNA		2016-02-02	Homo sapiens	2015-03-05	SRX685306	<not provided>	low passage primary melanoma cultures	low passage primary melanoma cultures
SRX685307	0 SINGLE	cDNA		2016-02-02	Homo sapiens	2015-03-05	SRX685307	<not provided>	low passage primary melanoma cultures	low passage primary melanoma cultures
SRX685308	0 SINGLE	cDNA		2016-02-02	Homo sapiens	2015-03-05	SRX685308	<not provided>	low passage primary melanoma cultures	low passage primary melanoma cultures
SRX685309	0 SINGLE	cDNA		2016-02-02	Homo sapiens	2015-03-05	SRX685309	<not provided>	low passage primary melanoma cultures	low passage primary melanoma cultures
SRX685310	0 SINGLE	cDNA		2016-02-02	Homo sapiens	2015-03-05	SRX685310	<not provided>	low passage primary melanoma cultures	low passage primary melanoma cultures
SRX685311	0 SINGLE	cDNA		2014-08-22	Homo sapiens	2015-03-05	SRX685311	<not provided>	melanoma cell line	melanoma cell line
SRX685312	0 SINGLE	cDNA		2014-08-22	Homo sapiens	2015-03-05	SRX685312	<not provided>	low passage primary melanoma cultures	low passage primary melanoma cultures
SRX685313	0 SINGLE	cDNA		2014-08-22	Homo sapiens	2015-03-05	SRX685313	<not provided>	low passage primary melanoma cultures	low passage primary melanoma cultures
SRX685314	0 SINGLE	cDNA		2014-08-22	Homo sapiens	2015-03-05	SRX685314	<not provided>	low passage primary melanoma cultures	low passage primary melanoma cultures
SRX685315	0 SINGLE	cDNA		2014-08-22	Homo sapiens	2015-03-05	SRX685315	<not provided>	low passage primary melanoma cultures	low passage primary melanoma cultures
SRX685316	0 SINGLE	cDNA		2014-08-22	Homo sapiens	2015-03-05	SRX685316	<not provided>	low passage primary melanoma cultures	low passage primary melanoma cultures
SRX685317	0 SINGLE	cDNA		2014-08-22	Homo sapiens	2015-03-05	SRX685317	<not provided>	low passage primary melanoma cultures	low passage primary melanoma cultures

Investigation of NGS: SRA BLAST!

NCBI Resources How To

Nucleotide Nucleotide Advanced

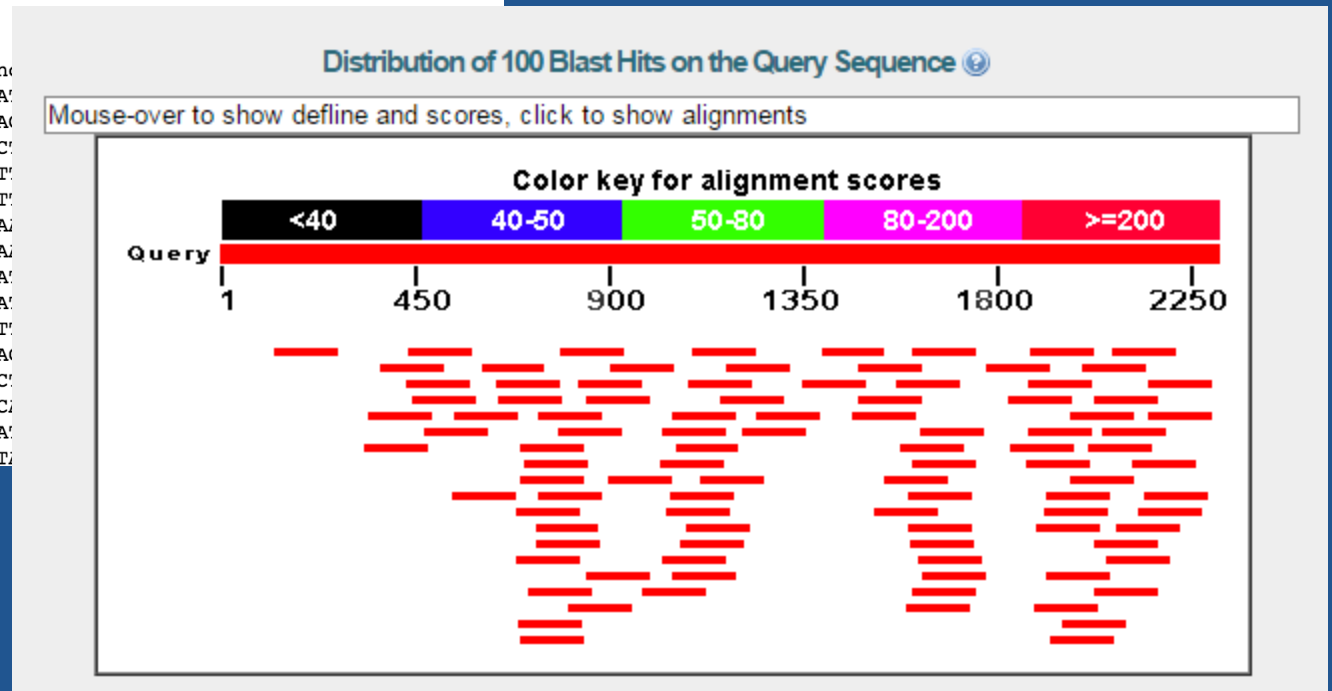
FASTA

Human endogenous retrovirus HERV-K, pol gene

GenBank: Y10391.1

[GenBank](#) [Graphics](#)

```
>gi|1780972|emb|Y10391.1| Human en
CCACGAGTCAAAAAATCATGACCAAGATGGGATA
CATTAAGTTCCAGTTGAGGCTAAAATAAATCAA
TCACTGTAGAGCCTCCTAAACCCATACCACTAAC
GCCGCTACCAAAACAAAACTGGAGGCTTTACAT
GAGCCTTCGTTCTCACCTTGAATTCTCCTGTGT
TAACTGACTTAAGGGCTGTAAACGCCGTAATTCA
GGCCATGATCCCAAAAGATTGGCCTTTAATTATA
GCAGAGCAGGATTGTGAAAAATTTGCCTTTACTA
TTCAGTGGAAGTGTTACCTCAGGGAATGCTTAA
TCTTCAACCAAGTGAGAGAAAAGTTTCAGACTGT
GCAGAAACGAAAGATAAATTAATTGACTGTATA
CAATAGCATCTGATAAGATCCAAACCTCTACTCC
AATTAAGCCACAAAAATAGAAATAAGAAAAGAC
GGAGATATTAATTGGATTTCGGCCAACCTCTAGGC
TAAGAGGAGACTCAGACTTAAATAGTCAAAGAAT
```



Investigation of NGS: MagicBLAST!

NCBI News

Search NCBI

Search

⚠ NCBI is currently testing https on public web servers until 4:00 PM EDT (20:00 UTC) today. You may experience problems with NCBI services during this test. [Read more.](#)

SHARE

< Previous

Current Story

Next >

Introducing Magic-BLAST

Thursday, September 22, 2016

Magic-BLAST is a new tool for mapping large sets of next-generation RNA or DNA sequencing runs against a whole genome or transcriptome. Magic-BLAST executables for LINUX, MacOSX, and Windows as well as the source files are available on the [FTP site](#).

Each alignment optimizes a composite score, taking into account simultaneously the two reads of a pair, and in case of RNA-Seq, locating the candidate introns and adding up the score of all exons. Sequencing reads can be provided as NCBI SRA accessions, FASTA or SRA files.

Magic-BLAST implements ideas developed in the NCBI Magic pipeline using the NCBI BLAST libraries. Magic-BLAST is under active development, and we expect the next few releases to occur on a monthly basis. Read more about Magic BLAST on the [FTP site](#).

SHARE

< Previous

Current Story

Next >

Follow us on






Archives

Year: 2016

[Jan](#) [Feb](#) [Mar](#) [Apr](#)
[May](#) [Jun](#) [Jul](#) [Aug](#)
[Sep](#) [Oct](#) [Nov](#) [Dec](#)

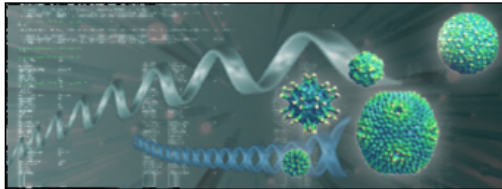
[Archives prior to July, 2012](#)

Viral Genomes

 [Resources](#)  [How To](#) 

[Sign in to NCBI](#)

Viruses



Viral Genomes

This resource provides viral and viroid genome sequence data and related information

Explore Viral Genome Sequences

- [Viral genome browser](#)
- [Viroid genome browser](#)
- [Browse viral genomes by family](#)
- [Browse viroid genomes by family](#)
- [View all RefSeq and Neighbor nucleotide records](#)

Resource Tools

- [Retrovirus Resource](#)
- [Virus Variation Resource](#)
- [Pairwise Sequence Comparison Tool \(PASC\)](#)
- [Protein Clusters](#)
- [Viral Genotyping Tool](#)

Virus Variation Resource

- [Influenza virus](#)
- [Denque virus](#)
- [West Nile virus](#)
- [MERS coronavirus](#)
- [Ebola virus](#)
- [Rotavirus](#)
- [Zika virus](#)

Download Viral Genome Data

- [Accession list of all viral genomes](#)
- [Accession list of all viroid genomes](#)
- [Complete RefSeq release of viral and viroid sequences](#)

Related Resources

- [Viral Zone](#)
- [Virus Pathogen Resource](#)
- [International Committee on Taxonomy of Viruses](#)
- [Virus \(Re-\)Annotation Database](#)

Contact and Outreach

- [Resource overview and help](#)
- [Resource how to guide](#)
- [How to cite us](#)
- [Contact us](#)
- [Viral Genome Advisors](#)



Virus Variation

 [Resources](#)  [How To](#) 

[Sign in to NCBI](#)

Virus Variation

Search NCBI

Search



Zika Virus Resource

Retrieve, view, and download Zika virus nucleotide and protein sequences from a value added database using a specialized search interface.

Zika virus sequences

- [Zika virus nucleotide sequences](#)
- [Zika virus protein sequences](#)
- [How to cite us](#)
- [Contact us](#)

Other NCBI Zika virus resources

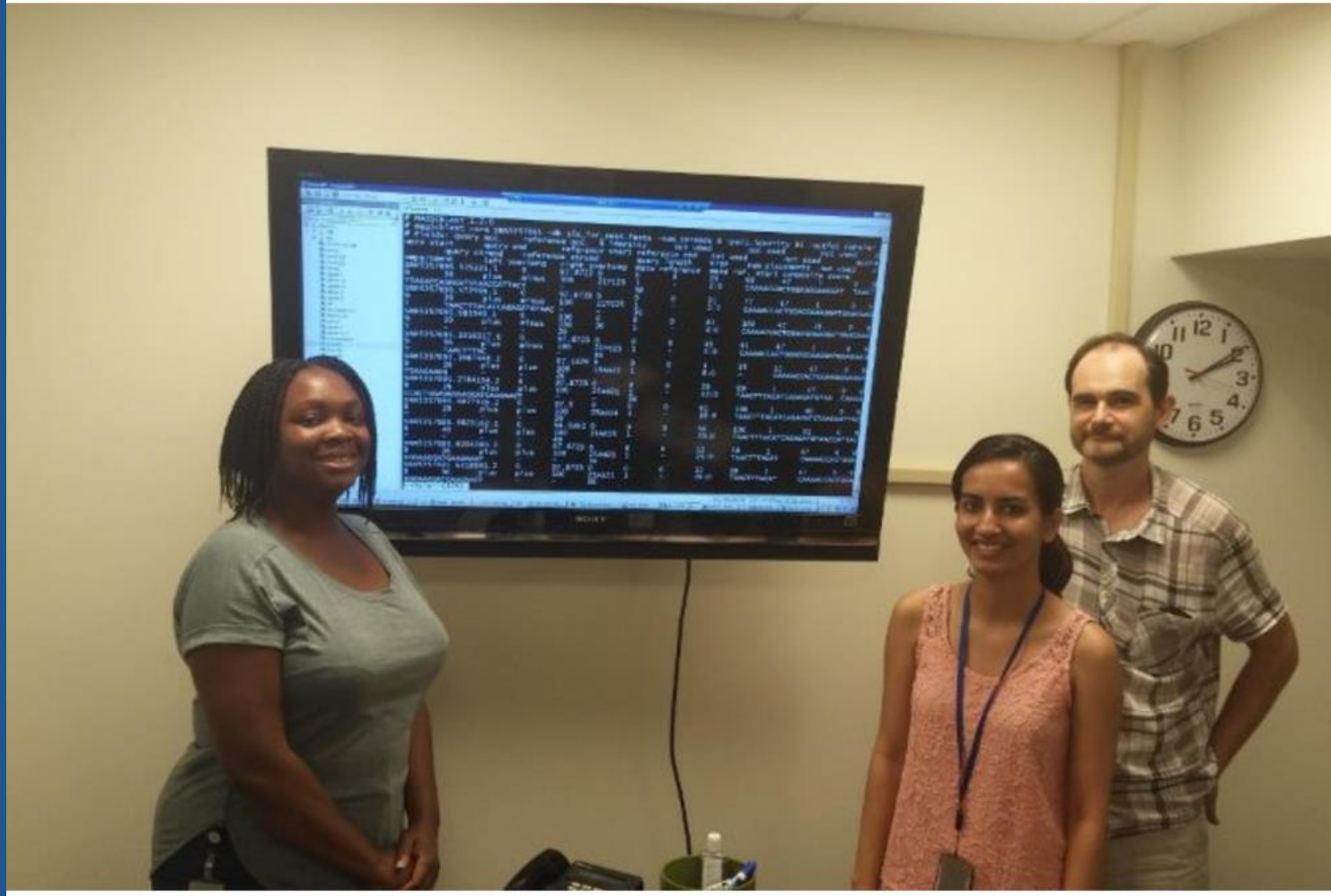
- [Zika virus reference genome](#)
- [Publications](#)
- [Genome browser](#)
- [Taxonomy](#)

External Zika virus resources

- [Zika virus health information resources](#)
- [HealthMap](#)
- [CDC](#)
- [WHO](#)
- [ViralZone](#)



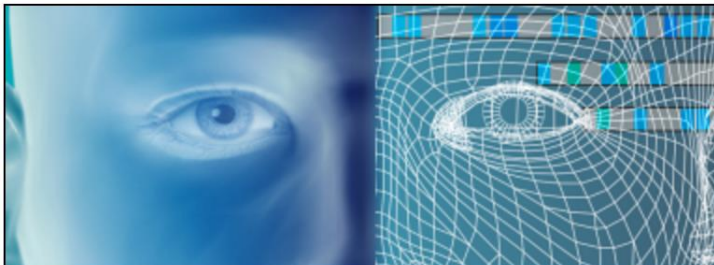
My team (plus Greg from BLAST team) @NCBI just figured out how to call variants on the fly from any NGS dataset using magicBLAST! **Chipo Mashayamombe** **Greg Boratyn** **Anmol Vohra**



dbGaP

dbGaP

Limits Advanced



dbGaP

The database of Genotypes and Phenotypes (dbGaP) was developed to archive and from studies that have investigated the interaction of genotype and phenotype in Hun

Access dbGaP Data

[Advanced Search](#)

[Controlled Access Data](#)

[Public FTP Download](#)

[Collections](#)

[Summary Statistics](#)

Resources

[Phenotype-Genotype Integrator](#)

[Association Results Browser](#)

[dbGaP RSS Feed](#) 

[Software](#)

[dbGaP Tutorial](#)

Important Links

[How to Submit](#)

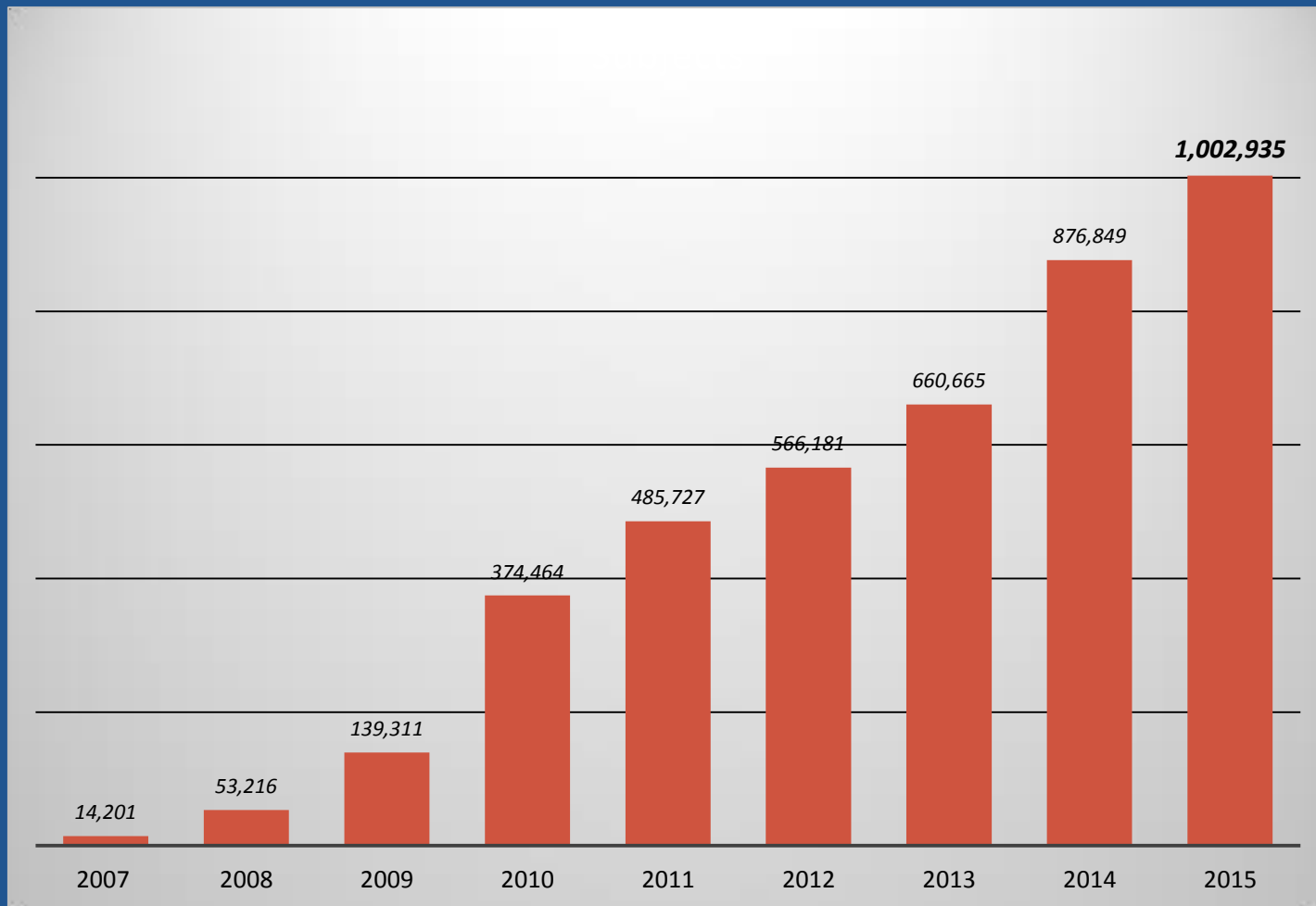
[FAQ](#)

[Code of Conduct](#)

[Security Procedures](#)

[Contact Us](#)

dbGaP



dbGaP Advanced Search

Study Disease/Focus IS **Melanoma**

Type Keyword or Phrase

Study Disease/Focus (326)

melanoma

Sort By Alphabetical clear

☒ Melanoma (7)

Study Design (3)

Sort By Decreasing

- ☐ Case Set (3)
- ☐ Case-Control (2)
- ☐ Tumor vs. Matched-Normal (2)

Study Molecular Data Type (7)

Sort By Decreasing

- ☐ Whole Exome (NGS) (4)
- ☐ SNP Genotypes (Array) (3)
- ☐ Whole Genome (NGS) (3)
- ☐ SNV Aggregate (.MAF) (2)
- ☐ RNA_Seq (NGS) (1)
- ☐ SNP Genotypes (imputed) (1)

Study Markers (5)

Sort By Alphabetical

- ☐ Genome-Wide_Human_SNP_Array_6_0 (1)
- ☐ HumanOmni1-Quad_v1-0_B (1)
- ☐ HumanOmniExpressExome-8v1_A (1)
- ☐ maf_grc37 (2)
- ☐ Not Provided (3)

NIH Institute (2)

Studies (7) Variables (440) Phenotype Datasets (26) Documents (7) Molecular Datasets (7) Analyses (0) 1/1

High Density SNP Association Analysis of Melanoma

Accession phs000187.v1.p1
 Study Disease/Focus Melanoma
 Study Design Case-Control
 Study Markers HumanOmni1-Quad_v1-0_B
 Study Molecular Data Type SNP Genotypes (Array), SNP Genotypes (imputed)
 Study Content 4 dataset(s) , 38 variable(s) , 7 document(s) , 3 genotype(s)
 NIH Institute NCI
 Study Consent GRU --- General research use
 Subject Count 3101
 Release Date 2010-05-18
 Embargo Release Date 2011-05-18

This research builds upon an extensive resource of melanoma cases and hospital based controls collected over several years at the U.T. M.D. Anderson Cancer Center. The goal of this ... research is to identify novel susceptibility and outcome-related genes for melanoma using a systematic genome-wide association-based approach. Our...

[FileSelector](#) [PubMed](#) [PMC](#) [MeSH](#) [BioProject](#) [BioSample](#)

A Novel Recurrent Mutation in MITF Predisposes to Familial and Sporadic Melanoma

Accession phs000419.v1.p1
 Study Disease/Focus Melanoma
 Study Design Case-Control
 Study Markers Not Provided
 Study Molecular Data Type Whole Genome (NGS)
 Study Content 4 dataset(s) , 15 variable(s) , SRA
 NIH Institute NCI
 Study Consent CRO --- Cancer research only
 Subject Count 1
 Release Date 2012-02-03
 Embargo Release Date 2012-02-03

We conducted whole-genome sequencing of probands from several melanoma families, identifying one individual carrying a novel germline variant (c.G1075A, NM_000248.3; p.E318K, NP_000239.1; rs149617956) in the melanoma lineage-specific oncogene MITF. ... While the variant cosegregated with melanoma in some, but not all cases in the family, linkage analysis of 31 families subsequently identified to...

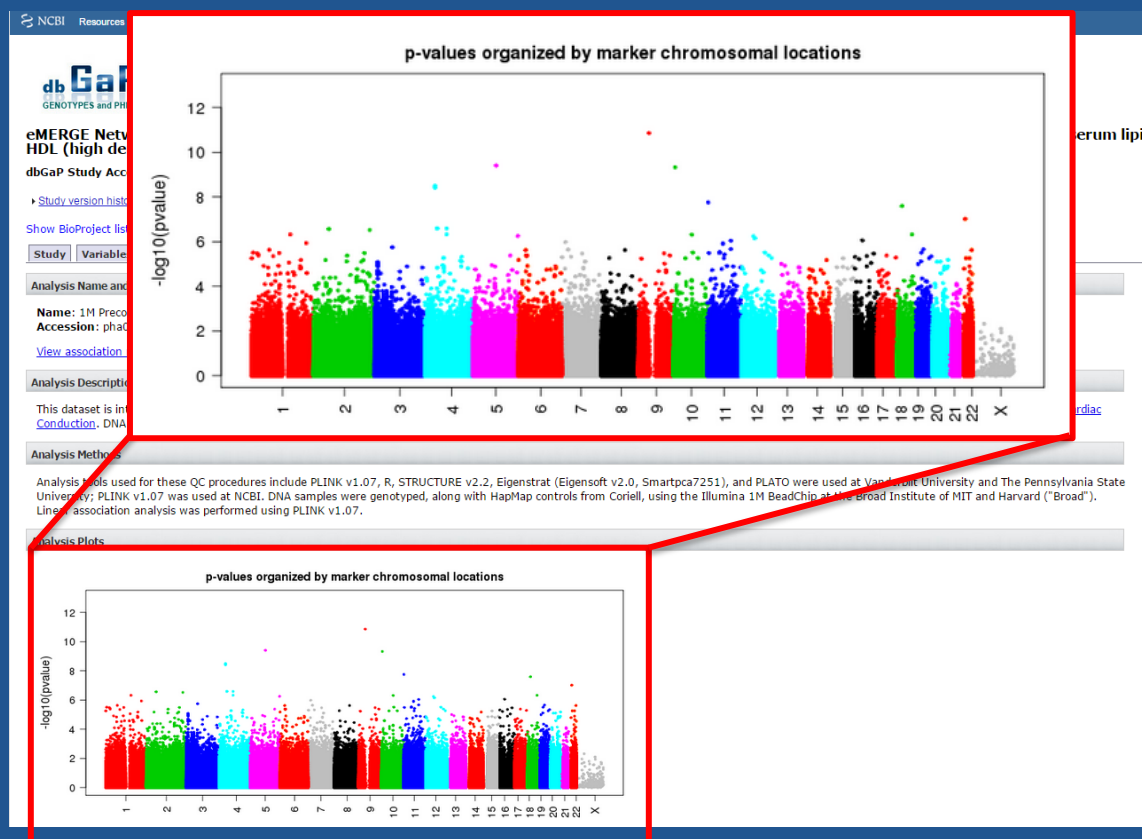
[FileSelector](#) [RunSelector](#) [PubMed](#) [PMC](#) [MeSH](#) [BioProject](#) [BioSample](#) [SRA](#)

Melanoma Genome Sequencing Project


Accession phs000452.v2.p1
 Study Disease/Focus Melanoma
 Study Design Case Set
 Study Markers maf_grc37, Genome-Wide_Human_SNP_Array_6_0




dbGaP – GWAS and PheGenI



dbGaP – GWAS and PheGenI

 NCBI Resources ▾ How To ▾


Phenotype-Genotype Integrator

All Databases ▾ Search

Welcome to PheGenI

The Phenotype-Genotype Integrator (PheGenI), merges NHGRI genome-wide association study (GWAS) catalog data with several data Biotechnology Information (NCBI), including Gene, dbGaP, OMIM, GTEx and dbSNP. This phenotype-oriented resource, intended for c up results from GWAS, can facilitate prioritization of variants to follow up, study design considerations, and generation of biological hypo location, gene, SNP, or phenotype and view and download results including annotated tables of SNPs, genes and association results, a expression data. PheGenI is still under active development. Currently, the phenotype search terms are based on MeSH and will be enh

Search Criteria

Search Clear Examples...

Phenotype Selection ⓘ

Traits:

P-Value: Source:

Genotype Selection ⓘ

Location

Gene

SNP

Chromosome:

Range (bps):

SNP Functional Class

☐ exon ☐ intron ☐ neargene ☐ UTR

Search Clear Examples...

dbGaP – ClinVar

NCBI Resources How To [bushybr@ncbi.nlm.nih.gov](#) [My NCBI](#) [Sign Out](#)

ClinVar ClinVar Search ClinVar for gene symbols, HGVS expressions, conditions, and more [Advanced](#) [Help](#)

Home About Access Using the website Submission Statistics FTP site

NM_001039661.1(TIRAP):c.539C>T (p.Ser180Leu)

Variation ID: 4467
Review status: (0/4) no assertion criteria provided

Interpretation Go to: [Clinical significance](#) [Last evaluated](#) [Number of submission\(s\)](#)

Clinical significance: [protective](#)
Last evaluated: May 18, 2015
Number of submission(s): 4

Variant frequency in dbGaP

NM_001039661.1(TIRAP):c.539C>T (p.Ser180Leu)
GRCh37 Chr11:126162843

	Called variants	Potential variants
Sample count	2517 of 9686	10630 of 43813

Called variants are samples submitted to dbGaP that have the variant allele. Potential variants are SRA runs that display the allele in at least 30% of the reads covering the position, and have 10 or more passing reads covering the position.

Browser views

[RefSeqGene](#)
[Variation Viewer \[GRCh38 - GRCh37\]](#)
[UCSC \[GRCh38/hg38 - GRCh37/hg19\]](#)

Related information

[dbSNP](#)
[Gene](#)
[MedGen](#)

Cytogenetic location: 11q24
Genomic location:

- Chr11: 126292948 (on Assembly GRCh38)
- Chr11: 126162843 (on Assembly GRCh37)

Protein change: S180L

1 Affected gene

toll-interleukin 1 receptor (TIR) domain containing adaptor protein (TIRAP) [Gene - OMIM - Variation Viewer]

[Search ClinVar for variants within TIRAP](#)
[Search ClinVar for variants including TIRAP](#)

Variant frequency in dbGaP

NM_001039661.1(TIRAP):c.539C>T (p.Ser180Leu)
GRCh37 Chr11:126162843

	Called variants	Potential variants
Sample count	2517 of 9686	10630 of 43813

Called variants are samples submitted to dbGaP that have the variant allele. Potential variants are SRA runs that display the allele in at least 30% of the reads covering the position, and have 10 or more passing reads covering the position.

Browser views

[RefSeqGene](#)
[Variation Viewer \[GRCh38 - GRCh37\]](#)
[UCSC \[GRCh38/hg38 - GRCh37/hg19\]](#)

Related information

[dbSNP](#)
[Gene](#)
[MedGen](#)

ClinVar

← → ↺ www.ncbi.nlm.nih.gov/clinvar/?term=tuberculosis 🔍 ☆ ☰

NCBI Resources [x] How To [x] [busbybr@ncbi.nlm.nih.gov](#) [My NCBI](#) [Sign Out](#)

ClinVar ClinVar tuberculosis [x] [Search](#)

[Save search](#) [Advanced](#) [Help](#)

Home About [v] Data use and maintenance [v] Using the website [v] How to submit [v] Statistics FTP site

[Show additional filters](#) [Display Settings:](#) [x] Tabular, 20 per page, Sorted by Default order [Send to:](#) [x]

Gene
Select ...

Clinical significance
Conflicting interpretations (5)
Pathogenic (8)
Risk factor (12)

Review status
Multiple submitters (7)
Single submitter (12)

Molecular consequence
Missense (9)
Near gene (1)
UTR (1)

Variation type
Single nucleotide (19)

[Clear all](#)
[Show additional filters](#)

Results: 19

	Gene(s)	Condition(s)	Frequency	Clinical significance (Last reviewed)	Review status	Chr	Location (GRCh38)
<input type="checkbox"/> 1. SP110, C-T, INTRON 6	SP110	Mycobacterium tuberculosis, susceptibility to		risk factor (Dec 8, 2010)	classified by single submitter		
<input type="checkbox"/> 2. NM_080424.2(SP110):c.1274C=>p.Ser425=	SP110	Mycobacterium tuberculosis, susceptibility to	GO-ESP:0.86268(G) GMAF:0.10280(A)	risk factor (Dec 8, 2010)	classified by single submitter	2	230185999
<input type="checkbox"/> 3. CD209, -871G-A	CD209	Mycobacterium tuberculosis, susceptibility to		risk factor (Mar 9, 2012)	classified by single submitter		
<input type="checkbox"/> 4. SLC11A1, 274C-T	SLC11A1	Mycobacterium tuberculosis, susceptibility to infection by		risk factor (Oct 15, 2013)	classified by single submitter		
<input type="checkbox"/> 5. IL12RB1, IVS9DS, G-C, +1	IL12RB1	Mycobacterial and salmonella infections, susceptibility to, Mycobacterium tuberculosis, susceptibility to infection by		risk factor (Apr 18, 2013)	classified by multiple submitters		
<input type="checkbox"/> 6. NM_000416.2(IFNGR1):c.260T>C(p.Ile87Thr)	IFNGR1	Mycobacterium tuberculosis, susceptibility to infection by, Bcg infection, tuberculoid, antibiotic-responsive		conflicting data from submitters (Oct 7, 2013)	classified by multiple submitters	6	137206249
<input type="checkbox"/> 7. NM_003264.3(TLR2):c.2258G>A(p.Arg753Gln)	TLR2	Mycobacterium tuberculosis, susceptibility to	GO-ESP:0.02222(A) GMAF:0.01190(A)	risk factor (Mar 18, 2013)	classified by single submitter	4	153705165
<input type="checkbox"/> 8. IFNG, +874A-T	IFNG	Mycobacterium tuberculosis, protection against		protective	classified		

SIDEBAR

ClinVar

www.ncbi.nlm.nih.gov/clinvar/?term=tuberculosis

NCBI Resources How To

busbybr@ncbi.nlm.nih.gov My NCBI Sign Out

ClinVar ClinVar tuberculosis Search

Save search Advanced Help

Home About Data use and maintenance Using the website How to submit Statistics FTP site

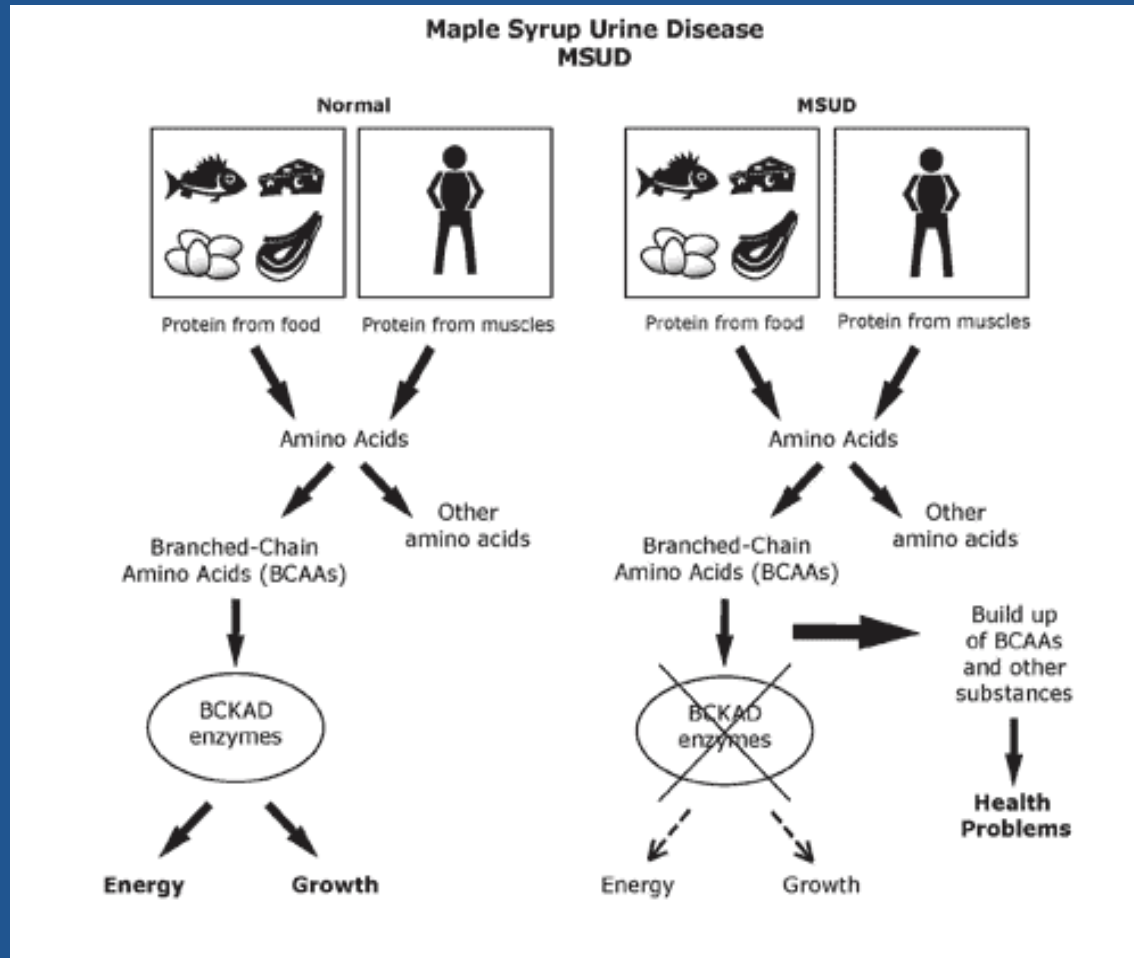
Show additional filters Display Settings: Tabular, 20 per page, Sorted by Default order Send to: SID

Results: 19

	Gene(s)	Condition(s)	Frequency	Clinical significance (Last reviewed)	Review status	Chr	Location (GRCh38)
<input type="checkbox"/> 1. SP110, C-T, INTRON 6	SP110	Mycobacterium tuberculosis , susceptibility to		risk factor (Dec 8, 2010)	classified by single submitter		
<input type="checkbox"/> 2. NM_080424.2(SP110):c.1274C=>T(p.Ser425=)	SP110	Mycobacterium tuberculosis , susceptibility to	GO-ESP:0.86268(G) GMAF:0.10280(A)	risk factor (Dec 8, 2010)	classified by single submitter	2	230185999
<input type="checkbox"/> 3. CD209, -871G-A	CD209	Mycobacterium tuberculosis , susceptibility to		risk factor (Mar 9, 2012)	classified by single submitter		
<input type="checkbox"/> 4. SLC11A1, 274C-T	SLC11A1	Mycobacterium tuberculosis , susceptibility to infection by		risk factor (Oct 15, 2013)	classified by single submitter		
<input type="checkbox"/> 5. IL12RB1, IVS9DS, G-C, +1	IL12RB1	Mycobacterial and salmonella infections, susceptibility to, Mycobacterium tuberculosis , susceptibility to infection by		risk factor (Apr 18, 2013)	classified by multiple submitters		



ClinVar – Why Should we Care?



ClinVar – Why Should we Care?

NCBI Resources How To busbybr@ncbi.nlm.nih.gov My NCBI Sign Out

Gene Gene Advanced Help

Display Settings: ☒ Full Report Send to:

BCKDHB branched chain keto acid dehydrogenase E1, beta polypeptide [*Homo sapiens* (human)]
Gene ID: 594, updated on 2-Nov-2014

Summary

Official Symbol BCKDHB provided by [HGNC](#)
Official Full Name branched chain keto acid dehydrogenase E1, beta polypeptide provided by [HGNC](#)
Primary source [HGNC:HGNC:987](#)
Locus tag RP1-279A18.1
See related [Ensembl:ENSG00000083123](#); [HPRD:02011](#); [MIM:248611](#); [Vega:OTTHUMG00000016430](#)
Gene type protein coding
RefSeq status REVIEWED
Organism [Homo sapiens](#)
Lineage Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorhini; Catarrhini; Hominidae; Homo
Also known as E1B; dJ279A18.1
Summary Branched-chain keto acid dehydrogenase is a multienzyme complex associated with the inner membrane of mitochondria, and functions in the catabolism of branched-chain amino acids. The complex consists of multiple copies of 3 components: branched-chain alpha-keto acid decarboxylase (E1), lipoamide acyltransferase (E2) and lipoamide dehydrogenase (E3). This gene encodes the E1 beta subunit, and mutations therein have been associated with maple syrup urine disease (MSUD), type 1B, a disease characterized by a maple syrup odor to the urine in addition to mental and physical retardation, and feeding problems. Alternative splicing at this locus results in transcript variants with different 3' non-coding regions, but encoding the same isoform. [provided by RefSeq, Jul 2008]

Genomic context

Location: 6q14.1 See BCKDHB in [Epigenomics](#), [MapViewer](#)

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

Related information

- Order cDNA clone
- 3D structures
- BioAssay
- BioAssay by Target (List)
- Protocols by Target (List)



ClinVar – Why Should we Care?



ClinVar – Why Should we Care?

NCBI Resources How To Sign in to NCBI

Variation Viewer

Homo sapiens: GRCh38.p2 (GCF_000001405.28) Chr 6 (NC_000006.12): 80.08M - 80.37M

New to Variation Viewer? [Read our quick overview!](#)

Reset All Share this page FAQ Help Version 1.5

Pick Assembly

Search

594[genid]

Enter a location, gene name or phenotype
Search examples:

Your Data

- no uploaded tracks -

Click '+' to add a file or drag files or text here

History

Region Details

NC_000006.12: 80M..80M (288Kbp)

Genes, NCBI Homo sapiens Annotation Release 107, 2015-03-13

ClinVar Short Variations based on dbSNP Build 147 (Homo sapiens Annotation Release 107), 201...

dbVar ClinVar Large Variations

Features of Interest

Other sequence representations - None

No GRC genome issues in range [Add Track](#)

Variation Data

Filter by

Source database

- dbSNP (14,600)
- dbVar (399)

In ClinVar

- Yes (77)
- No (14,922)

Most severe clinical significance

- Pathogenic (30)
- Likely pathogenic (8)
- drug response (0)
- risk factor (0)
- association (0)

More...

Variant type

Download Edit columns

Variant ID	Location	Variant type	Gene	Molecular consequences	Most severe clinical significance	1000G MAF	GO-ESP MAF	ExAC MAF	Publications
nsv429572	149,661 - 170,741,917	copy number variation	PTP4A1 and 1392 more						1
esv3337429	32,687,594 - 131,679,681	copy number variation	PTP4A1 and 589 more						1
nsv1143323	51,982,921 - 100,746,163	copy number variation	ZNF292 and 189 more						1
nsv1135963	51,982,938 - 100,746,164	copy number variation	ZNF292 and 189 more						1
nsv483033	62,690,096 - 87,290,282	copy number variation	ZNF292 and 85 more						1
nsv498057	64,549,655 - 83,426,791	copy number variation	HTR1B and 63 more		Pathogenic				1
nsv1017107	66,339,982 - 101,039,508	copy number variation	ZNF292 and 135 more						1
esv3335765	67,672,127 - 108,257,931	copy number variation	ZNF292 and 160 more						1
nsv1019964	69,662,932 - 89,756,313	copy number variation	ZNF292 and 100 more						1
nsv538301	69,662,932 - 89,756,313	copy number variation	ZNF292 and 100 more						1
nsv917239	74,382,807 - 142,040,500	copy number variation	ZNF292 and 325 more		Pathogenic				1

Items 1 - 20 of 14,999 << First < Prev Page 1 of 750 Next > Last >>

Current Call

ClinVar – Why Should we Care?

NCBI Resources How To Sign in to NCBI

GTR: GENETIC TESTING REGISTRY

594[geneid] Tests Search Advanced search for tests

[GTR Home](#) > Tests > Search results - BCKDHB[gene]

Apply filters

▼ Condition/Phenotype

Showing tests for all 2979 conditions

Enter text to filter the conditions

Select a condition

- Maple syrup urine disease (35)
- Maple syrup urine disease type 1B (18)
- Niemann-Pick disease, type A (15)
- Cystic fibrosis (14)
- Galactosylceramide beta-galactosidase
- Spongy degeneration of central nervous
- Maple syrup urine disease, type 3 (13)
- Citrullinemia type I (13)
- Tay-Sachs disease (13)

Compare labs

Your search term can be found in tests with a total of 2979 conditions. Only 1000 conditions are displayed in this filter box. Please type the name of the condition in the search box in this filter to find the specific condition. You can use the filters below to narrow down your results. If the name of the condition you typed is not present, please try another query or search using the All GTR tab.

▼ Test type

- ☐ Clinical (63)

▼ Test purpose

- ☐ Diagnosis (54)
- ☐ Mutation Confirmation (22)
- ☐ Pre-symptomatic (13)
- ☐ Predictive (2)
- ☐ Monitoring (1)

▼ Test method

▼ Clinical test, Research test

Showing 1 to 20 of 63 tests for 2979 conditions in 29 labs

<< First < Prev Page 1 of 4 Next > Last >>

C CentolCU platinum

Lab: [Centogene AG - the Rare Disease Company University of Rostock](#) Rostock, Mecklenburg-Vorpommern, Germany

Conditions	Test targets	Methods
Combined oxidative phosphorylation deficiency 8	AARS2	C Sequence analysis of the entire coding region
2-methyl-3-hydroxybutyric aciduria	AASS	
3 beta-Hydroxysteroid dehydrogenase deficiency	ABAT	
Total conditions (769)	Total targets (514)	

C CentolCU platinum plus

Lab: [Centogene AG - the Rare Disease Company University of Rostock](#) Rostock, Mecklenburg-Vorpommern, Germany

Conditions	Test targets	Methods
Combined oxidative phosphorylation deficiency 8	AARS2	C Sequence analysis of the entire coding region
2-methyl-3-hydroxybutyric aciduria	AASS	
3 beta-Hydroxysteroid dehydrogenase deficiency	ABAT	
Total conditions (769)	Total targets (514)	

C AllNeuro panel

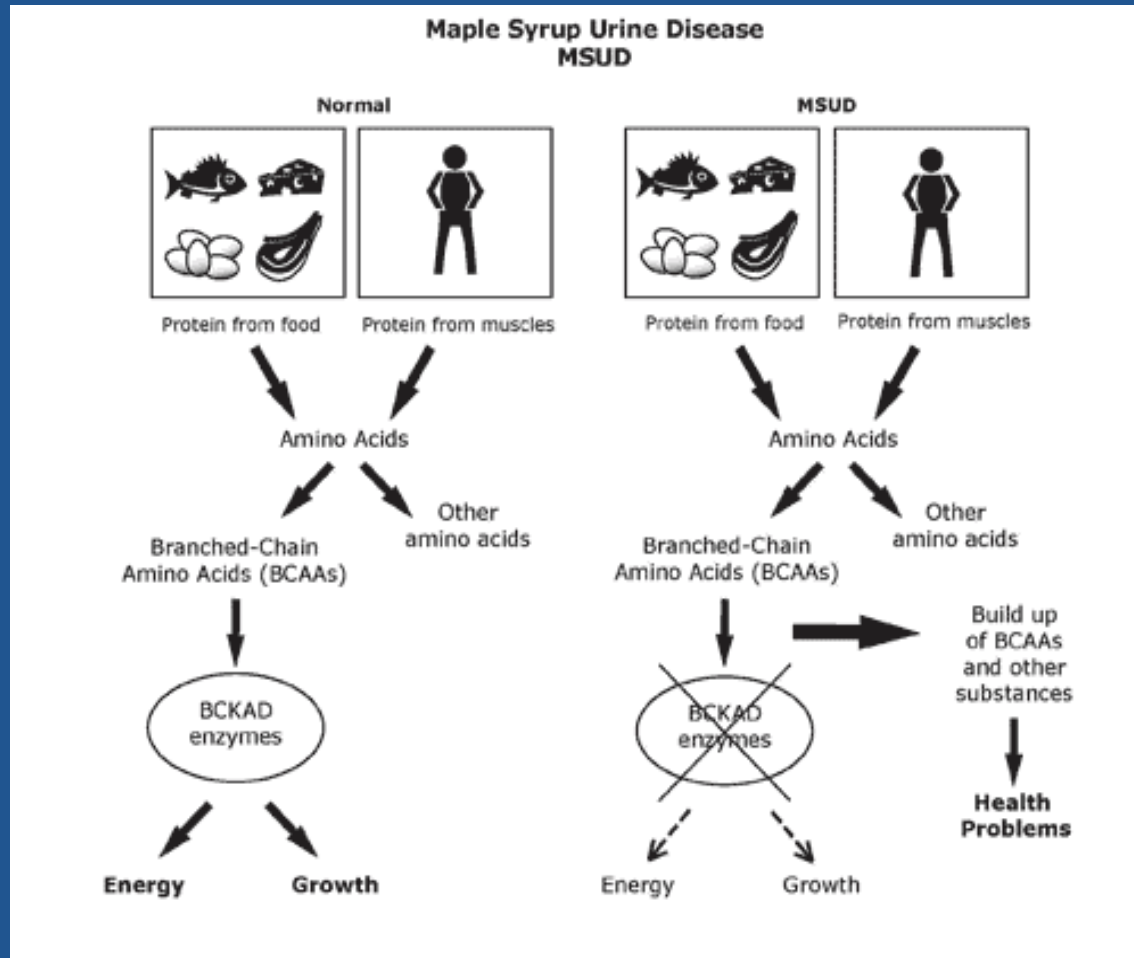
Lab: [Centogene AG - the Rare Disease Company University of Rostock](#) Rostock, Mecklenburg-Vorpommern, Germany

Conditions	Test targets	Methods
Alzheimer's disease	A2M	C Sequence analysis of the entire coding region
46,XY gonadal dysgenesis, partial, with minifascicular neuropathy	AAS	
Aarskog syndrome	AANAT	
Total conditions (602)	Total targets (1038)	

C Comprehensive mitochondrial disorders panel

Lab: [Centogene AG - the Rare Disease Company University of Rostock](#) Rostock, Mecklenburg-Vorpommern, Germany

ClinVar – Why Should we Care?



Where to Get More Information!

Webinars & Courses

In-person courses, live webinars and webinar recordings.
The time displayed for online webinars is based on U.S. Eastern Time (Washington D.C.)
Other times shown are t

Documentation

Upcoming Webinars 8 Online manuals, handbooks, fact sheets and FAQs

Filter this table

Filter this table

Date	Resource	Description	Materials
Mar 7, 2016	1000 Genomes browser	genome viewer for 1000 Genomes Project data	help FAQ factsheet
Mar 11, 2016	BankIt	online nucleotide sequence submission service	help handbook
	BioProject	catalog of high-throughput genome-wide studies	help factsheet handbook
	BioSample	sample repository for the BioProject database	help FAQ handbook
	BioSystems	pathways with links to genes, proteins and chemicals	help FAQ citation
Mar 2, 2016	BLAST (standalone version)	downloadable version of the sequence similarity search tool	help FAQ factsheet handbook
	BLAST (web version)	online version of the sequence similarity search tool	help FAQ factsheet handbook
	BLAST (cloud version)	cloud-based version of the sequence similarity search tool	help FAQ handbook
	Bookshelf	catalog of books and documents	help FAQ factsheet handbook
	C++ Toolkit	cross-platform application framework for working with NCBI data	help FAQ handbook
	CDD	conserved protein functional domain repository	help factsheet handbook citation
	CD-Search	sequence-based protein domain search tool	help

```
#FORMAT=ID:DP,Number1,Type=Integer,Description="Read Depth"
#FORMAT=ID:DPF,Number1,Type=Integer,Description="Filtered Depth"
#FORMAT=ID:DT,Number1,Type=Integer,Description="Total Depth"
#FORMAT=ID:DQ,Number0,Type=Flag,Description="Were any of the samples downsampled?"
#FORMAT=ID:Dole,Number1,Type=Float,Description="Fraction of Reads Containing Spanning Deletions"
#FORMAT=ID:FC,Number0,Type=String,Description="Functional Consequence"
#FORMAT=ID:FC,Number1,Type=Flag,Description="Phred-scaled P-value using Fisher's exact test to detect strand bias"
#FORMAT=ID:GQ,Number0,Type=String,Description="Gene ID"
#FORMAT=ID:GQ,Number1,Type=Flag,Description="Log-scaled likelihoods for AA,AB,EB genotypes where Aref and Balt: not applicable if site is not biallelic"
#FORMAT=ID:GQ,Number1,Type=Float,Description="Genotype Quality"
#FORMAT=ID:GQ,Number1,Type=String,Description="Genotype"
#FORMAT=ID:HM,Number1,Type=Integer,Description="Largest Contiguous Homopolymer Run of Variant Allele In Either Direction"
#FORMAT=ID:HM,Number1,Type=Flag,Description="Haplotype Score. Consistency of the site with two segregating haplotypes"
#FORMAT=ID:IC,Number0,Type=Flag,Description="Inbreeding coefficient as estimated from the genotype likelihoods per-sample when compared against the Hardy-Weinberg expectation"
#FORMAT=ID:MM,Number2,Type=Float,Description="average # of mismatches per consensus indel-supporting read-per-reference-supporting read"
#FORMAT=ID:MQ,Number2,Type=Float,Description="average mapping quality of consensus indel-supporting read-per-reference-supporting read"
#FORMAT=ID:MQ,Number1,Type=Integer,Description="Total Mapping Quality Zero Reads"
#FORMAT=ID:MQR,Number1,Type=Float,Description="RMS Mapping Quality"
#FORMAT=ID:MQS,Number1,Type=Float,Description="Mapping Quality Rank Sum, Z-score From Wilcoxon rank-sum test"
#FORMAT=ID:MQS,Number2,Type=Float,Description="Within MQS window, average quality of bases from contigs supporting read"
#FORMAT=ID:MQSM,Number2,Type=Float,Description="Within MQS window, fraction of mismatching bases in reference-supporting reads"
#FORMAT=ID:NR,Number0,Type=Integer,Description="Normalized, Phred-scaled likelihoods for genotypes as N"
#FORMAT=ID:OQ,Number1,Type=Float,Description="Variant Confidence, Quality by Depth"
#FORMAT=ID:QI,Number1,Type=Flag,Description="Region of Interest"
#FORMAT=ID:RP,Number1,Type=Flag,Description="Read Pos Rank Sum, Z-score from Wilcoxon rank sum test"
#FORMAT=ID:SD,Number1,Type=Flag,Description="Strand Bias"
#FORMAT=ID:SD,Number4,Type=Integer,Description="strandness: counts of forward-reverse-aligned indels-aligning reference supporting reads"
#FORMAT=ID:T,Number0,Type=String,Description="Transcript ID"
#FORMAT=ID:VAL,Number1,Type=Integer,Description="Bitmask field bits: (1) not assessed, (2) novel, (4) in use from existing job using this platform, (16) validated in a different sample using as in this sample using another platform"
#FORMAT=ID:VLM,Number1,Type=Character,Description="Variant Likely Mutation for variant in the mitochon"

##INFO=ID:EN,Number0,Type=Flag,Description="EBNF Membership"
##INFO=ID:EX,Number0,Type=Flag,Description="Exon Region"
##INFO=ID:FL,Number0,Type=Flag,Description="Flag"
##INFO=ID:FL,Number1,Type=Flag,Description="Flag"
##INFO=ID:FL,Number2,Type=Flag,Description="Flag"
##INFO=ID:FL,Number3,Type=Flag,Description="Flag"
##INFO=ID:FL,Number4,Type=Flag,Description="Flag"
##INFO=ID:FL,Number5,Type=Flag,Description="Flag"
##INFO=ID:FL,Number6,Type=Flag,Description="Flag"
##INFO=ID:FL,Number7,Type=Flag,Description="Flag"
##INFO=ID:FL,Number8,Type=Flag,Description="Flag"
##INFO=ID:FL,Number9,Type=Flag,Description="Flag"
##INFO=ID:FL,Number10,Type=Flag,Description="Flag"
##INFO=ID:FL,Number11,Type=Flag,Description="Flag"
##INFO=ID:FL,Number12,Type=Flag,Description="Flag"
##INFO=ID:FL,Number13,Type=Flag,Description="Flag"
##INFO=ID:FL,Number14,Type=Flag,Description="Flag"
##INFO=ID:FL,Number15,Type=Flag,Description="Flag"
##INFO=ID:FL,Number16,Type=Flag,Description="Flag"
##INFO=ID:FL,Number17,Type=Flag,Description="Flag"
##INFO=ID:FL,Number18,Type=Flag,Description="Flag"
##INFO=ID:FL,Number19,Type=Flag,Description="Flag"
##INFO=ID:FL,Number20,Type=Flag,Description="Flag"
##INFO=ID:FL,Number21,Type=Flag,Description="Flag"
##INFO=ID:FL,Number22,Type=Flag,Description="Flag"
##INFO=ID:FL,Number23,Type=Flag,Description="Flag"
##INFO=ID:FL,Number24,Type=Flag,Description="Flag"
##INFO=ID:FL,Number25,Type=Flag,Description="Flag"
##INFO=ID:FL,Number26,Type=Flag,Description="Flag"
##INFO=ID:FL,Number27,Type=Flag,Description="Flag"
##INFO=ID:FL,Number28,Type=Flag,Description="Flag"
##INFO=ID:FL,Number29,Type=Flag,Description="Flag"
##INFO=ID:FL,Number30,Type=Flag,Description="Flag"
##INFO=ID:FL,Number31,Type=Flag,Description="Flag"
##INFO=ID:FL,Number32,Type=Flag,Description="Flag"
##INFO=ID:FL,Number33,Type=Flag,Description="Flag"
##INFO=ID:FL,Number34,Type=Flag,Description="Flag"
##INFO=ID:FL,Number35,Type=Flag,Description="Flag"
##INFO=ID:FL,Number36,Type=Flag,Description="Flag"
##INFO=ID:FL,Number37,Type=Flag,Description="Flag"
##INFO=ID:FL,Number38,Type=Flag,Description="Flag"
##INFO=ID:FL,Number39,Type=Flag,Description="Flag"
##INFO=ID:FL,Number40,Type=Flag,Description="Flag"
##INFO=ID:FL,Number41,Type=Flag,Description="Flag"
##INFO=ID:FL,Number42,Type=Flag,Description="Flag"
##INFO=ID:FL,Number43,Type=Flag,Description="Flag"
##INFO=ID:FL,Number44,Type=Flag,Description="Flag"
##INFO=ID:FL,Number45,Type=Flag,Description="Flag"
##INFO=ID:FL,Number46,Type=Flag,Description="Flag"
##INFO=ID:FL,Number47,Type=Flag,Description="Flag"
##INFO=ID:FL,Number48,Type=Flag,Description="Flag"
##INFO=ID:FL,Number49,Type=Flag,Description="Flag"
##INFO=ID:FL,Number50,Type=Flag,Description="Flag"
##INFO=ID:FL,Number51,Type=Flag,Description="Flag"
##INFO=ID:FL,Number52,Type=Flag,Description="Flag"
##INFO=ID:FL,Number53,Type=Flag,Description="Flag"
##INFO=ID:FL,Number54,Type=Flag,Description="Flag"
##INFO=ID:FL,Number55,Type=Flag,Description="Flag"
##INFO=ID:FL,Number56,Type=Flag,Description="Flag"
##INFO=ID:FL,Number57,Type=Flag,Description="Flag"
##INFO=ID:FL,Number58,Type=Flag,Description="Flag"
##INFO=ID:FL,Number59,Type=Flag,Description="Flag"
##INFO=ID:FL,Number60,Type=Flag,Description="Flag"
##INFO=ID:FL,Number61,Type=Flag,Description="Flag"
##INFO=ID:FL,Number62,Type=Flag,Description="Flag"
##INFO=ID:FL,Number63,Type=Flag,Description="Flag"
##INFO=ID:FL,Number64,Type=Flag,Description="Flag"
##INFO=ID:FL,Number65,Type=Flag,Description="Flag"
##INFO=ID:FL,Number66,Type=Flag,Description="Flag"
##INFO=ID:FL,Number67,Type=Flag,Description="Flag"
##INFO=ID:FL,Number68,Type=Flag,Description="Flag"
##INFO=ID:FL,Number69,Type=Flag,Description="Flag"
##INFO=ID:FL,Number70,Type=Flag,Description="Flag"
##INFO=ID:FL,Number71,Type=Flag,Description="Flag"
##INFO=ID:FL,Number72,Type=Flag,Description="Flag"
##INFO=ID:FL,Number73,Type=Flag,Description="Flag"
##INFO=ID:FL,Number74,Type=Flag,Description="Flag"
##INFO=ID:FL,Number75,Type=Flag,Description="Flag"
##INFO=ID:FL,Number76,Type=Flag,Description="Flag"
##INFO=ID:FL,Number77,Type=Flag,Description="Flag"
##INFO=ID:FL,Number78,Type=Flag,Description="Flag"
##INFO=ID:FL,Number79,Type=Flag,Description="Flag"
##INFO=ID:FL,Number80,Type=Flag,Description="Flag"
##INFO=ID:FL,Number81,Type=Flag,Description="Flag"
##INFO=ID:FL,Number82,Type=Flag,Description="Flag"
##INFO=ID:FL,Number83,Type=Flag,Description="Flag"
##INFO=ID:FL,Number84,Type=Flag,Description="Flag"
##INFO=ID:FL,Number85,Type=Flag,Description="Flag"
##INFO=ID:FL,Number86,Type=Flag,Description="Flag"
##INFO=ID:FL,Number87,Type=Flag,Description="Flag"
##INFO=ID:FL,Number88,Type=Flag,Description="Flag"
##INFO=ID:FL,Number89,Type=Flag,Description="Flag"
##INFO=ID:FL,Number90,Type=Flag,Description="Flag"
##INFO=ID:FL,Number91,Type=Flag,Description="Flag"
##INFO=ID:FL,Number92,Type=Flag,Description="Flag"
##INFO=ID:FL,Number93,Type=Flag,Description="Flag"
##INFO=ID:FL,Number94,Type=Flag,Description="Flag"
##INFO=ID:FL,Number95,Type=Flag,Description="Flag"
##INFO=ID:FL,Number96,Type=Flag,Description="Flag"
##INFO=ID:FL,Number97,Type=Flag,Description="Flag"
##INFO=ID:FL,Number98,Type=Flag,Description="Flag"
##INFO=ID:FL,Number99,Type=Flag,Description="Flag"
##INFO=ID:FL,Number100,Type=Flag,Description="Flag"
##INFO=ID:FL,Number101,Type=Flag,Description="Flag"
##INFO=ID:FL,Number102,Type=Flag,Description="Flag"
##INFO=ID:FL,Number103,Type=Flag,Description="Flag"
##INFO=ID:FL,Number104,Type=Flag,Description="Flag"
##INFO=ID:FL,Number105,Type=Flag,Description="Flag"
##INFO=ID:FL,Number106,Type=Flag,Description="Flag"
##INFO=ID:FL,Number107,Type=Flag,Description="Flag"
##INFO=ID:FL,Number108,Type=Flag,Description="Flag"
##INFO=ID:FL,Number109,Type=Flag,Description="Flag"
##INFO=ID:FL,Number110,Type=Flag,Description="Flag"
##INFO=ID:FL,Number111,Type=Flag,Description="Flag"
##INFO=ID:FL,Number112,Type=Flag,Description="Flag"
##INFO=ID:FL,Number113,Type=Flag,Description="Flag"
##INFO=ID:FL,Number114,Type=Flag,Description="Flag"
##INFO=ID:FL,Number115,Type=Flag,Description="Flag"
##INFO=ID:FL,Number116,Type=Flag,Description="Flag"
##INFO=ID:FL,Number117,Type=Flag,Description="Flag"
##INFO=ID:FL,Number118,Type=Flag,Description="Flag"
##INFO=ID:FL,Number119,Type=Flag,Description="Flag"
##INFO=ID:FL,Number120,Type=Flag,Description="Flag"
##INFO=ID:FL,Number121,Type=Flag,Description="Flag"
##INFO=ID:FL,Number122,Type=Flag,Description="Flag"
##INFO=ID:FL,Number123,Type=Flag,Description="Flag"
##INFO=ID:FL,Number124,Type=Flag,Description="Flag"
##INFO=ID:FL,Number125,Type=Flag,Description="Flag"
##INFO=ID:FL,Number126,Type=Flag,Description="Flag"
##INFO=ID:FL,Number127,Type=Flag,Description="Flag"
##INFO=ID:FL,Number128,Type=Flag,Description="Flag"
##INFO=ID:FL,Number129,Type=Flag,Description="Flag"
##INFO=ID:FL,Number130,Type=Flag,Description="Flag"
##INFO=ID:FL,Number131,Type=Flag,Description="Flag"
##INFO=ID:FL,Number132,Type=Flag,Description="Flag"
##INFO=ID:FL,Number133,Type=Flag,Description="Flag"
##INFO=ID:FL,Number134,Type=Flag,Description="Flag"
##INFO=ID:FL,Number135,Type=Flag,Description="Flag"
##INFO=ID:FL,Number136,Type=Flag,Description="Flag"
##INFO=ID:FL,Number137,Type=Flag,Description="Flag"
##INFO=ID:FL,Number138,Type=Flag,Description="Flag"
##INFO=ID:FL,Number139,Type=Flag,Description="Flag"
##INFO=ID:FL,Number140,Type=Flag,Description="Flag"
##INFO=ID:FL,Number141,Type=Flag,Description="Flag"
##INFO=ID:FL,Number142,Type=Flag,Description="Flag"
##INFO=ID:FL,Number143,Type=Flag,Description="Flag"
##INFO=ID:FL,Number144,Type=Flag,Description="Flag"
##INFO=ID:FL,Number145,Type=Flag,Description="Flag"
##INFO=ID:FL,Number146,Type=Flag,Description="Flag"
##INFO=ID:FL,Number147,Type=Flag,Description="Flag"
##INFO=ID:FL,Number148,Type=Flag,Description="Flag"
##INFO=ID:FL,Number149,Type=Flag,Description="Flag"
##INFO=ID:FL,Number150,Type=Flag,Description="Flag"
##INFO=ID:FL,Number151,Type=Flag,Description="Flag"
##INFO=ID:FL,Number152,Type=Flag,Description="Flag"
##INFO=ID:FL,Number153,Type=Flag,Description="Flag"
##INFO=ID:FL,Number154,Type=Flag,Description="Flag"
##INFO=ID:FL,Number155,Type=Flag,Description="Flag"
##INFO=ID:FL,Number156,Type
```



DangerTrack!

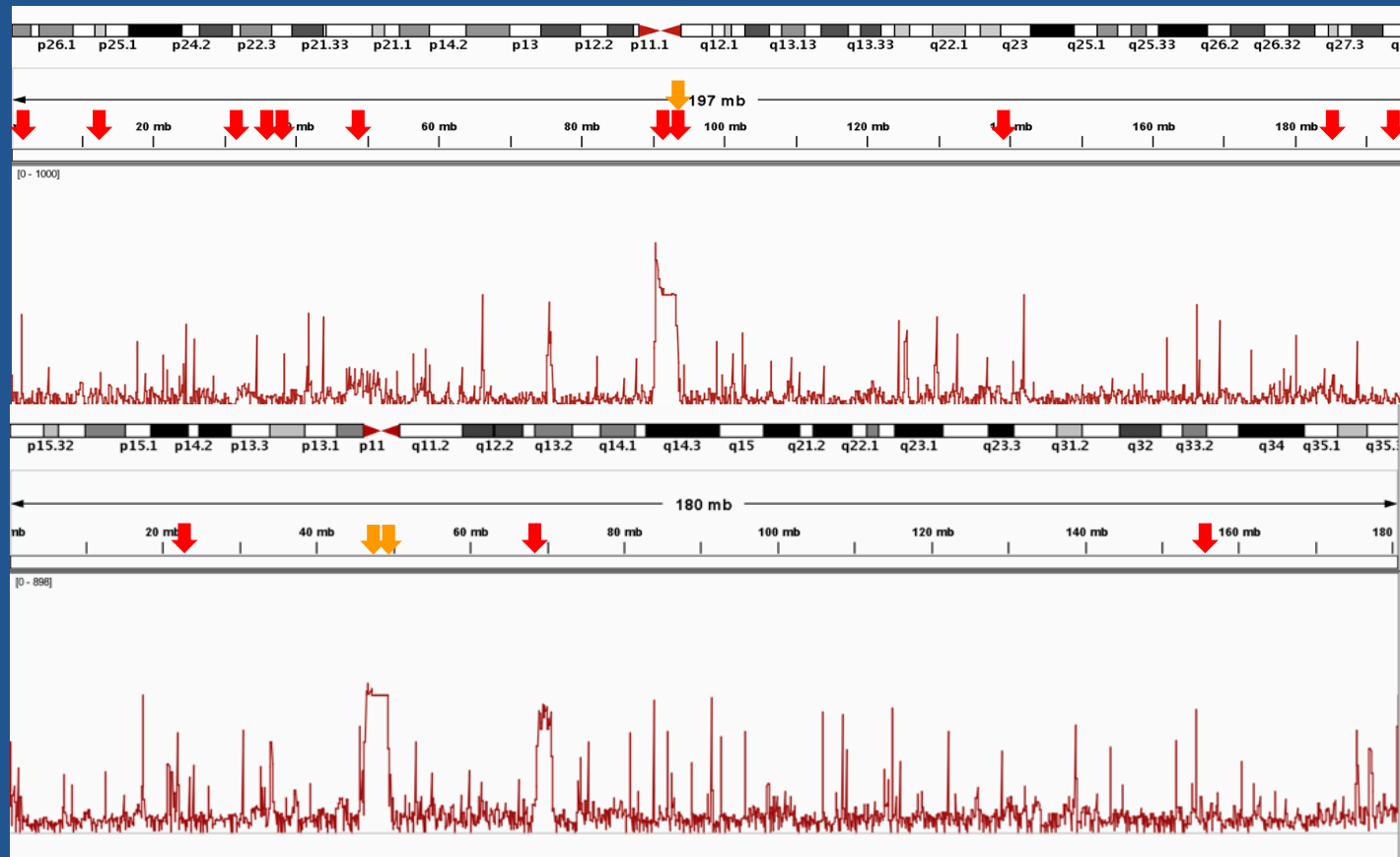
NCBI region list



Encode blacklist



Combined score is
the average of SVs,
mappability, GC..



**We've run 12 hackathons in
the last two years. We
continue to build on
those projects .**



Graph Genomes!!!

START HERE

Step 0: Get BAM alignments for each assembly



Step 1.1: Do pairwise alignment against reference



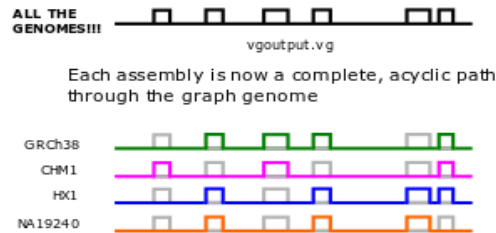
Step 1.2: Join pairwise alignments into single FASTA file



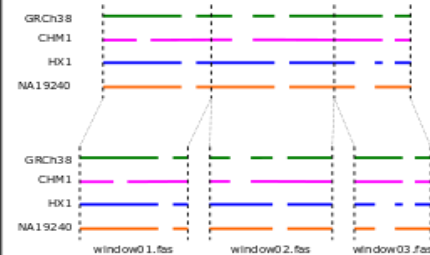
Step 3.1: Convert BAM file to VCF (in parallel)



Step 3.2: Use vg to convert VCF to graph genome



Step 2.1: Identify windows and extract one fasta file for each window

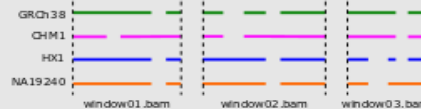


parallelized steps

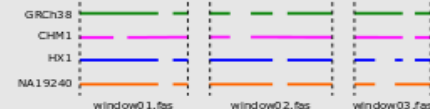
Step 2.3: Concatenate all alignments into single BAM file



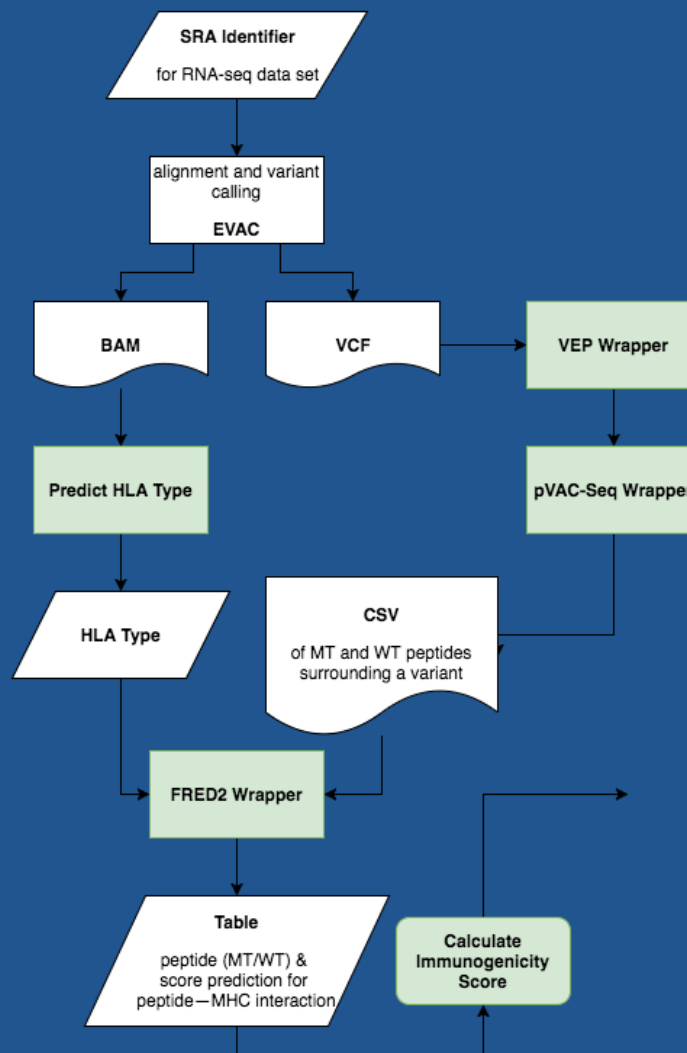
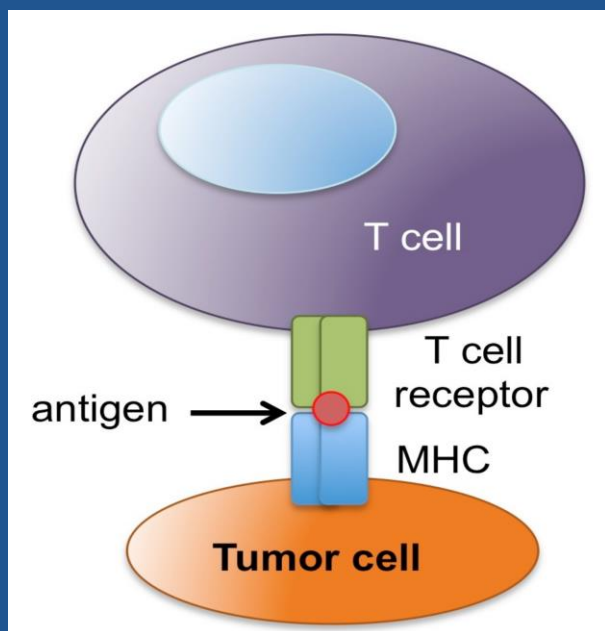
Step 2.2.2: Convert all alignment FASTA output files to BAM format



Step 2.2.1: Perform multiple sequence alignment for each window



Finding immunogenic peptides from single RNA-seq samples



Search

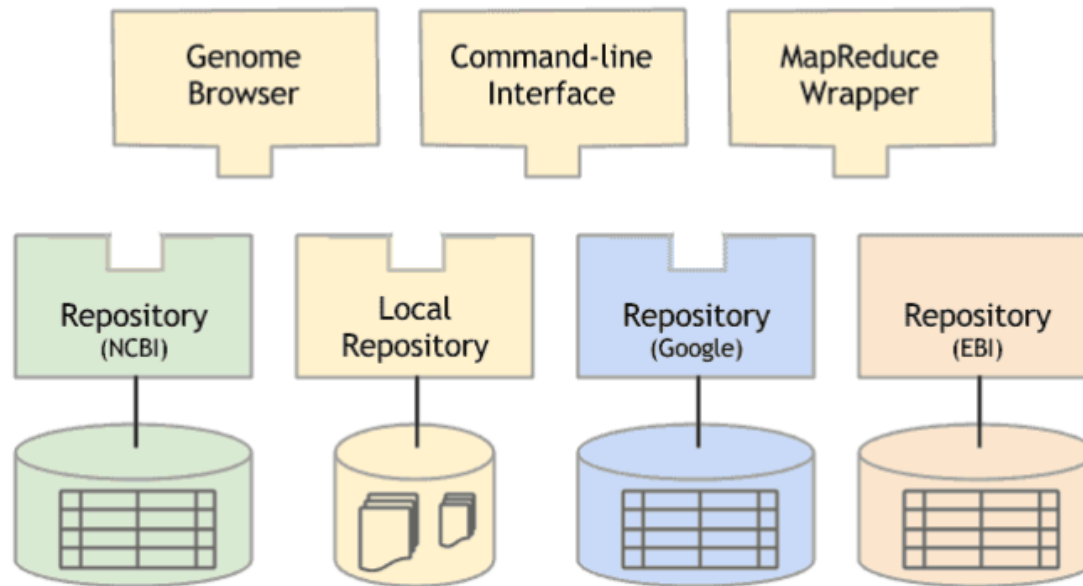


Collaboration!

Global Alliance for Genomics and Health

Data Working Group

Interoperability: One API, Many Apps



Genomics API

The [Global Alliance for Genomics and Health \(GA4GH\) Genomics API](#) will allow the interoperable exchange of genomic information across multiple organizations and on multiple platforms. This is a freely available open standard for interoperability, that uses common web protocols to support serving

The Future



Ontological Standardization

age	age [y]
Age	age [year]
AGE	age [years]
'Age	age in years
age (after birth)	age of patient
age (in years)	Age of patient
age (y)	age of subjects
age (year)	age(years)
age (years)	Age(years)
Age (years)	age(yrs)
Age (Years)	Age(yrs.)
age (yr)	Age, year
age (yr-old)	age, years
age (yrs)	age, yrs
Age (yrs)	age.year
	age_years

Not
standardized



Ontological Standardization

[CDEs](#)[Forms](#)[Boards](#)[Quick Board \(0\)](#)[Help](#)[Log In](#)[Q Search](#)[Browse by Classification](#)[Browse by Topic](#)[Export Search](#)

AHRQ

Agency for Healthcare Research and Quality

95 elements

cLBP

Chronic Low Back Pain

4 elements

eyeGENE

National Ophthalmic Disease Genotyping and Phenotyping Network, National Eye Institute

190 elements

GRDR

Global Rare Diseases Patient Registry Data Repository

75 elements

NCI

National Cancer Institute

NCI-BPV

Tumor Biospecimen Acquisition (BPV) – National Cancer Institute

NCI-GTEx

GTEx: Postmortem Biospecimen Acquisition – National Cancer Institute

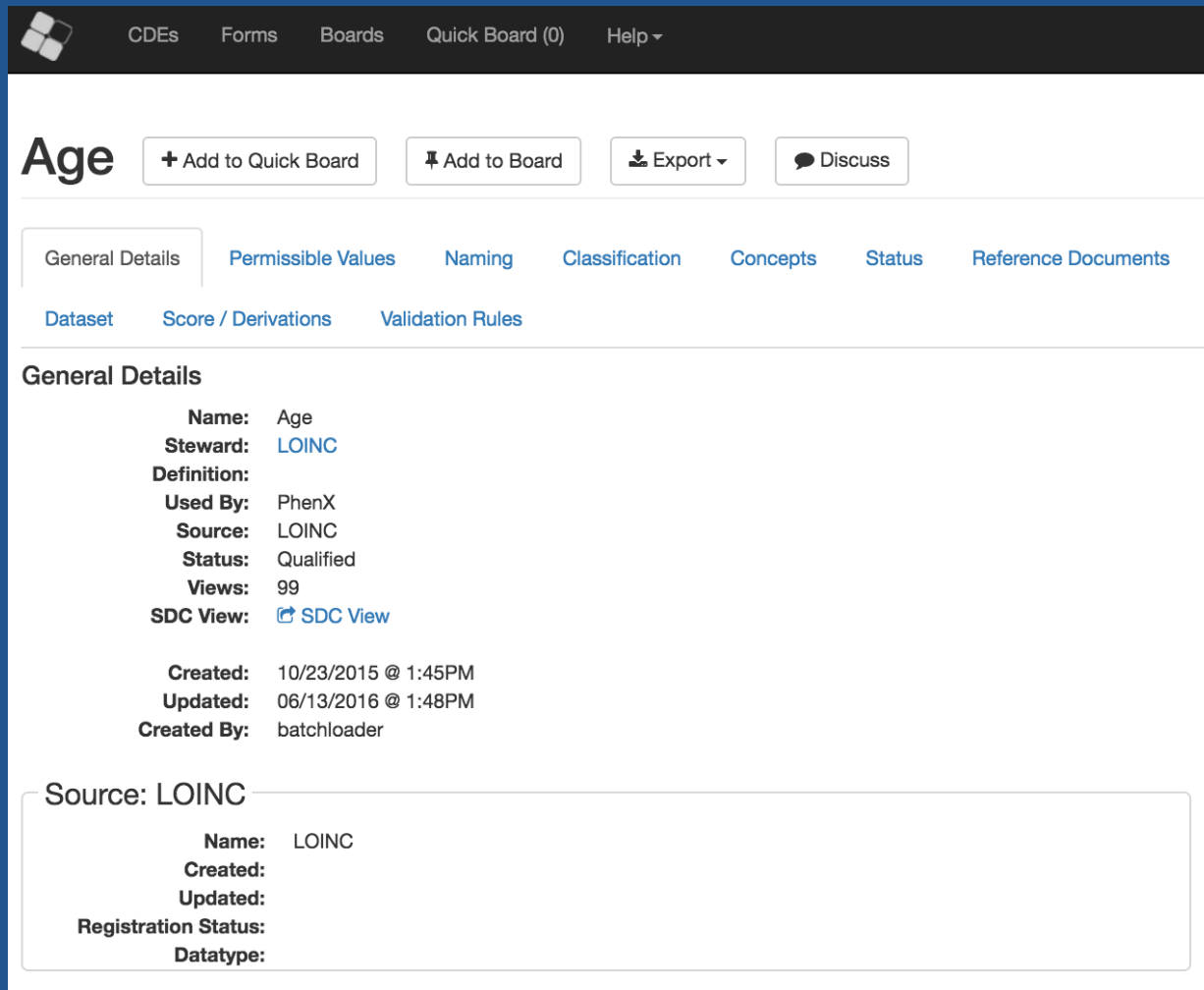
NIDA

National Institute on Drug Abuse

<https://cde.nlm.nih.gov/#/home>

[Report a problem!](#)

Ontological Standardization



The screenshot shows the LOINC (Logical Observation Identifiers Names and Codes) interface for the 'Age' concept. The top navigation bar includes links for CDEs, Forms, Boards, Quick Board (0), and Help. Below the navigation bar, the 'Age' concept is displayed with buttons for '+ Add to Quick Board', '+ Add to Board', 'Export', and 'Discuss'. A tabbed interface shows 'General Details' as the active tab, with other tabs for 'Permissible Values', 'Naming', 'Classification', 'Concepts', 'Status', and 'Reference Documents'. Under 'General Details', there are sub-tabs for 'Dataset', 'Score / Derivations', and 'Validation Rules'. The 'General Details' section lists the following information: Name: Age, Steward: LOINC, Definition: (empty), Used By: PhenX, Source: LOINC, Status: Qualified, Views: 99, and SDC View: SDC View. Below this, the creation and update history is shown: Created: 10/23/2015 @ 1:45PM, Updated: 06/13/2016 @ 1:48PM, and Created By: batchloader. At the bottom, the 'Source: LOINC' section is expanded, showing details for the LOINC source: Name: LOINC, Created: (empty), Updated: (empty), Registration Status: (empty), and Datatype: (empty).

Age + Add to Quick Board + Add to Board Export Discuss

General Details Permissible Values Naming Classification Concepts Status Reference Documents

Dataset Score / Derivations Validation Rules

General Details

Name: Age
Steward: LOINC
Definition:
Used By: PhenX
Source: LOINC
Status: Qualified
Views: 99
SDC View: SDC View

Created: 10/23/2015 @ 1:45PM
Updated: 06/13/2016 @ 1:48PM
Created By: batchloader

Source: LOINC

Name: LOINC
Created:
Updated:
Registration Status:
Datatype:

Integration into a Larger Data Discovery Framework

- a. A user-friendly means by which researchers search for relevant data. This could include both pull and push models, such as a web site to initiate searches and a mechanism to alert users about (new versions of) a data set of interest.
- b. A cyber-infrastructure effort informed by the community to efficiently and at scale enable interoperability between systems for creating, managing and using data. This would include:
 - i. A means to assign and describe metadata and potentially a central locator of multiple copies of the same data, as well as identification of the proper version.
 - ii. A means by which methods for crediting researchers for production and reuse of data can be obtained.
 - iii. A means by which NIH can track use of community-based metadata and data standards.
 - iv. A means by which NIH can track compliance with data sharing mandates and measure the impact of data sharing practices.
- c. A community-driven process for identifying and implementing existing standards, to improve discovery, understandability, re-use, and citation of datasets; for fostering the development of new standards where needed; for harmonization of existing standards; and for promoting adoption of standards.



Integration into a Larger Data Discovery Framework

Non-Bibliographic LinkOut Providers

The following LinkOut providers supply links to non-bibliographic web-accessible resources from PubMed and other Entrez database. A list of all [LinkOut providers](#) is also available.

Some providers may require user registration or subscription fees, or have other stipulations from the provider.

Clicking on one of the links below will search Entrez databases for records with links to that provider. Therefore links from Entrez records and the available resources are subject to change.

As of September 07, 2015, there are 252 resources.

- (NIF) (NIF): [PubMed](#)
- Addgene Non-profit plasmid repository (addgene): [Gene PubMed](#)
- ALFRED: The Allele Frequency Database (ALFRED): [SNP](#)
- AlgaeBase (AlgaeBase): [Taxonomy](#)
- All Catfish Species Inventory (catfish): [Taxonomy](#)
- Allen Brain Atlas (ABA): [Gene](#)
- American Academy of Family Physicians (AAFP): [PubMed](#)
- American Phytopathological Society (APPS): [Taxonomy](#)
- American Type Culture Collection (ATCC): [Nucleotide PubMed](#)
- Amphibian Species of the World (ASW): [Taxonomy](#)
- AmphibiaWeb (amphibweb): [Taxonomy](#)
- Angiosperm Phylogeny Website (APweb): [Taxonomy](#)
- Animal Diversity Web (ADW): [Taxonomy](#)
- AnimalBase (AnimalBase): [Taxonomy](#)
- antibodies-online (ABO): [PubMed](#)
- AntWeb (AntWeb): [Nucleotide Taxonomy](#)
- Arctos Specimen Database (Arctos): [Nucleotide Taxonomy](#)
- Area de Conservacion Guanacaste (ACG): [Taxonomy](#)
- Barcodes of Life (BoLD): [Nucleotide Taxonomy](#)
- Bee Genera of the World (BeeGen): [Taxonomy](#)
- Beetles of Florida (ColeoFla): [Taxonomy](#)
- BindingDB (BindingDB): [PubMed](#)
- Bio-Analytic Resource (bioarray): [Gene Nucleotide Protein](#)
- BioCyc (BioCyc): [Protein](#)
- BioGPS (BioGPS): [Gene](#)
- Bioimages.vanderbilt.edu (Bioimages): [Taxonomy](#)
- Bioinformatics.ca Links Directory (BIOCA): [PubMed](#)
- Biospecimen Research Database (ncibrd): [PubMed](#)
- Brain Operation Database (bodb): [PubMed](#)
- Breast Cancer TissueBank Bioinformatics Portal (bctbbp): [Gene](#)
- Broadly Neutralizing Antibody Electronic Resource (bnaber): [Nucleotide](#)
- Bugwood Network (University of Georgia) (Bugwood): [Taxonomy](#)
- Butterflies and Moths of the World (ButMoth): [Taxonomy](#)

The screenshot shows the NCBI BioProject search results for the query 'loprovGOLD[SB]'. The page is titled 'BioProject' and includes a search bar with the query and a 'Create alert' button. The results are displayed in a table with columns for 'Project Types', 'Data Types', 'Project Data', 'Scope', and 'Search results'. The search results are listed as follows:

- 1. [Bacteroidetes bacterium R-53146](#)
Bacteroidetes sp. R-53146 genome sequencing and assembly
Taxonomy: [Bacteroidetes bacterium R-53146](#)
Project data type: Genome sequencing
Attributes: Scope: Monoculture; Material: Genome; Capture: Whole; Method Type: Sequencing
Joint Genome Institute
Accession: PRJNA292838 ID: 292838
- 2. [Endozoicomonas atrinae strain:WP70](#)
Endozoicomonas atrinae strain:WP70 Genome sequencing and assembly
Taxonomy: [Endozoicomonas atrinae](#)
Project data type: Genome sequencing and assembly
Attributes: Scope: Monoculture; Material: Genome; Capture: Whole; Method Type: Sequencing
Aarhus University
Accession: PRJNA292506 ID: 292506
- 3. [Camobacterium maltaromaticum strain:UAL307](#)
Camobacterium maltaromaticum strain:UAL307 Genome sequencing and assembly
Taxonomy: [Camobacterium maltaromaticum](#)
Project data type: Genome sequencing and assembly
Attributes: Scope: Monoculture; Material: Genome; Capture: Whole; Method Type: Sequencing
University of Alberta
Accession: PRJNA292012 ID: 292012
- 4. [Endozoicomonas sp. KASP37](#)
Endozoicomonas ascidiicola KASP37 genome sequencing and assembly
Taxonomy: [Endozoicomonas sp. KASP37](#)
Project data type: Genome sequencing and assembly
Attributes: Scope: Monoculture; Material: Genome; Capture: Whole; Method Type: Sequencing
Aarhus University

Integration into a Larger Data Discovery Framework

Example: GOLD (JGI)

NCBI Resources How To busbybr@ncbi.nlm.nih.gov My NCBI Sign Out

BioProject BioProject Search Advanced Help

Display Settings: Send to:

Carnobacterium maltaromaticum DSM 20342 strain:MX5 Accession: PRJNA234883 ID: 234883

Carnobacterium maltaromaticum DSM 20342 strain:MX5 Genome sequencing

Comparative genomics of ten *Carnobacterium* species type strains plus two new *Carnobacterium* species isolated from Siberian permafrost and capable of growth at low pressure

Related Resources:

- JGI Portal For *Carnobacterium maltaromaticum* MX5, DSM 20342
- GOLD Card For *Carnobacterium maltaromaticum* MX5, DSM 20342

Project Data Type: Genome sequencing; **Locus Tag Prefix:** BR77

Attributes: Scope: Monoisolate; Material: Genome; Capture: Whole; Method type: Sequencing

Relevance: Comparative Analysis

Project Data:

Resource Name	Number of Links
SEQUENCE DATA	
Nucleotide (total)	6
WGS master	1
OTHER DATASETS	
BioSample	1
Assembly	1

▼ Assembly details:

Assembly	Level	WGS	BioSample	Taxonomy
GCA_000744945.1	Contig	JQMX000000000	SAMN02786975	Carnobacterium maltaromaticum DSM 20342 (firmicutes)

Lineage: Bacteria; Firmicutes; Bacilli; Lactobacillales; Carnobacteriaceae; Carnobacterium; Carnobacterium maltaromaticum; Carnobacterium maltaromaticum DSM 20342 [Taxonomy ID: 1449341]

Submission:
Registration date: 28-Aug-2014
DOE Joint Genome Institute

See Genome Information for Carnobacterium maltaromaticum

NAVIGATE ACROSS
4 additional projects are related by organism.

Related information

- Assembly
- BioSample
- Genome
- Nucleotide
- Taxonomy
- WGS master

Related Resources

- JGI Portal For *Carnobacterium maltaromaticum* MX5, DSM 20342
- GOLD Card For *Carnobacterium maltaromaticum* MX5, DSM 20342

LinkOut to external resources

- GOLDCARD: Gp0042581 [Genomes On Line Database]
- SILVA LSU Database [SILVA]
- SILVA SSU Database [SILVA]

Recent activity

- Carnobacterium maltaromaticum DSM 20342 strain:MX5 BioProject
- 73507[top bioproject] NOT 292012[uid] (4) BioProject
- Carnobacterium maltaromaticum strain:UAL307 BioProject
- SPA Links for BioProject (Select 200694)

Maple Syrup Urine Disease MSUD

