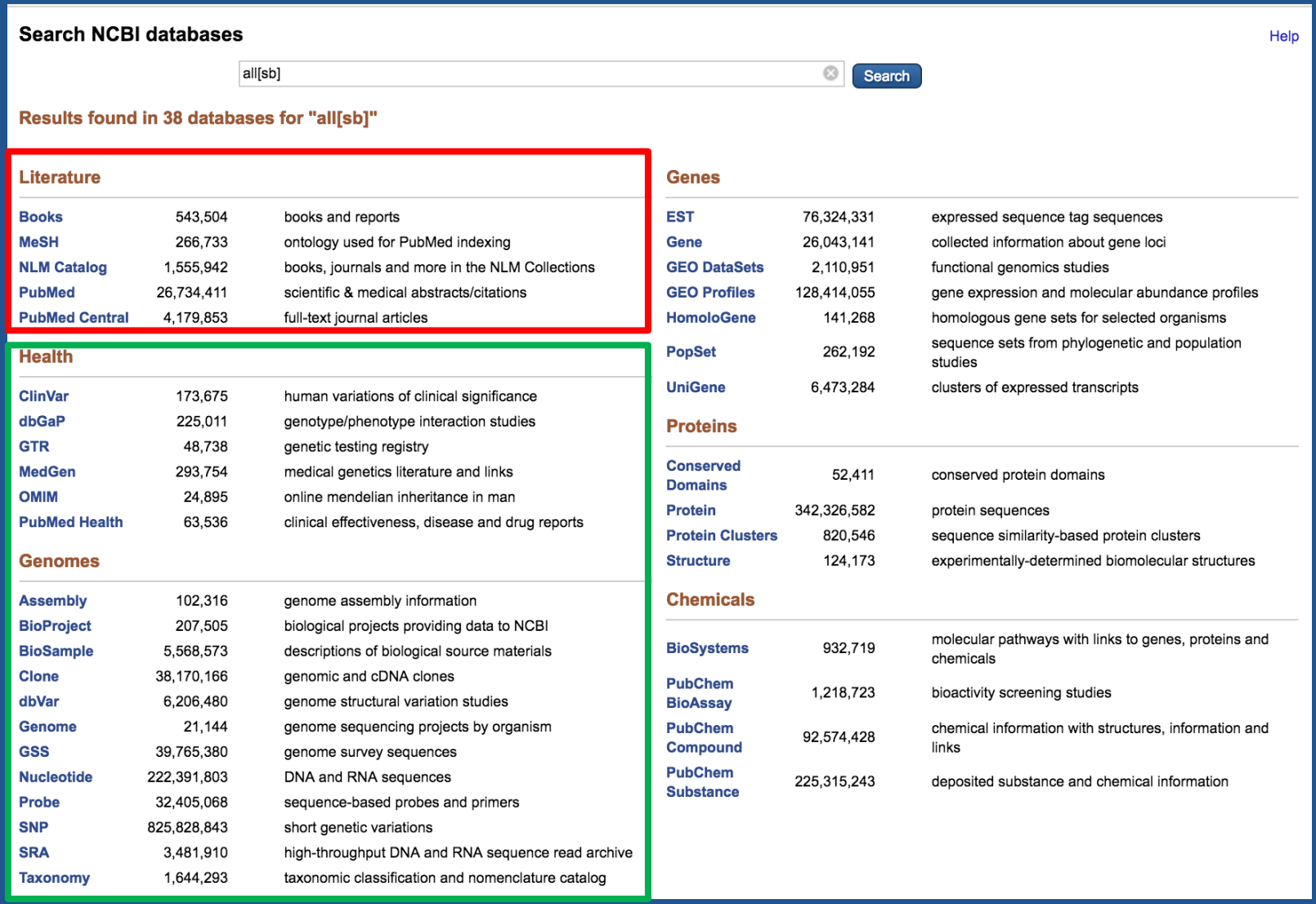
Exploring the Many Possible Futures of Data Science

Making the Transition from Sharing Data to Sharing Knowledge

Ben Busby, Ph.D.
Genomics Outreach Coordinator
NCBI
ben.busby@nih.gov

NCBI



Search NCBI databases Help

all[sb] Search

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

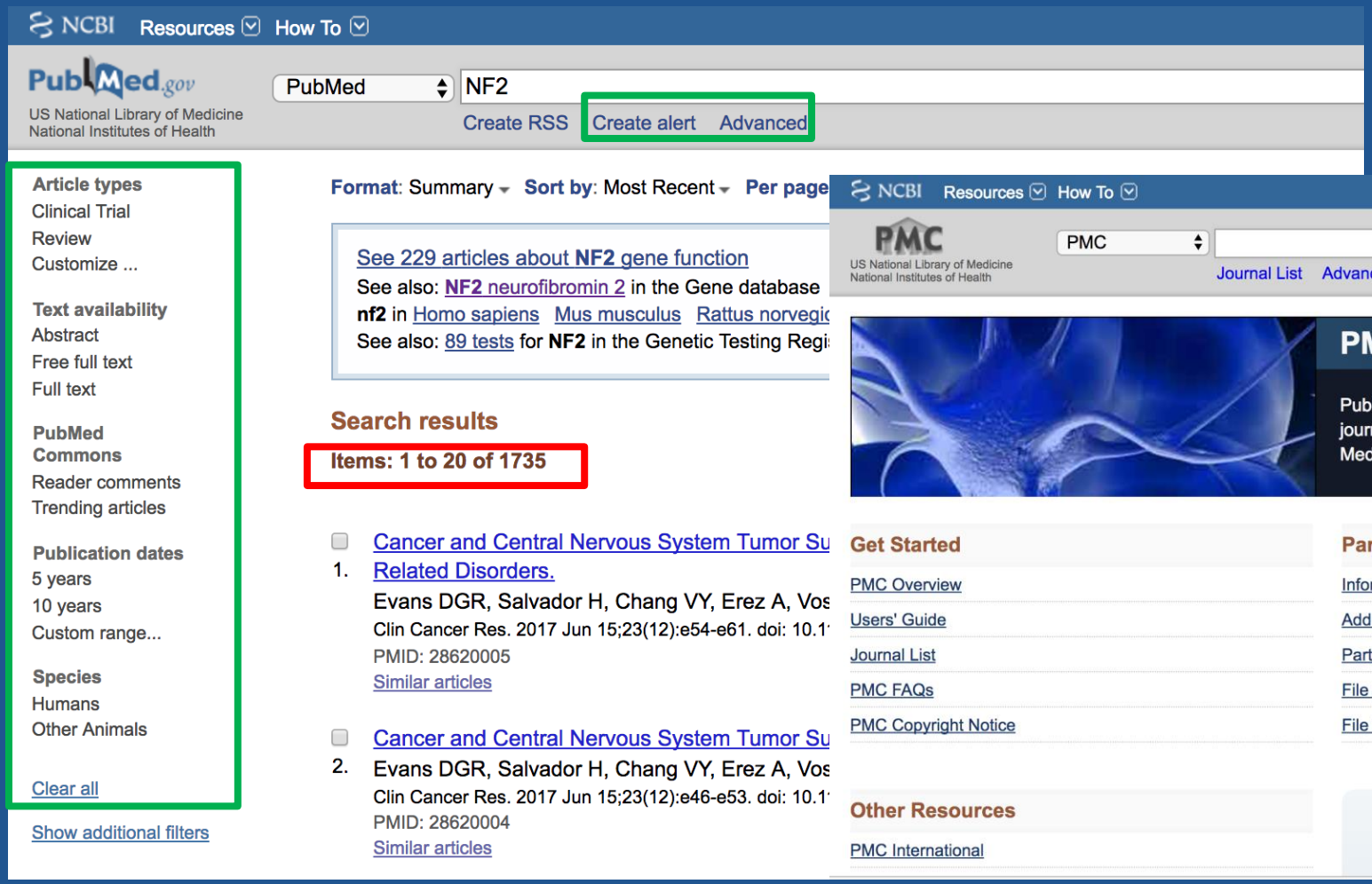
Proteins

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

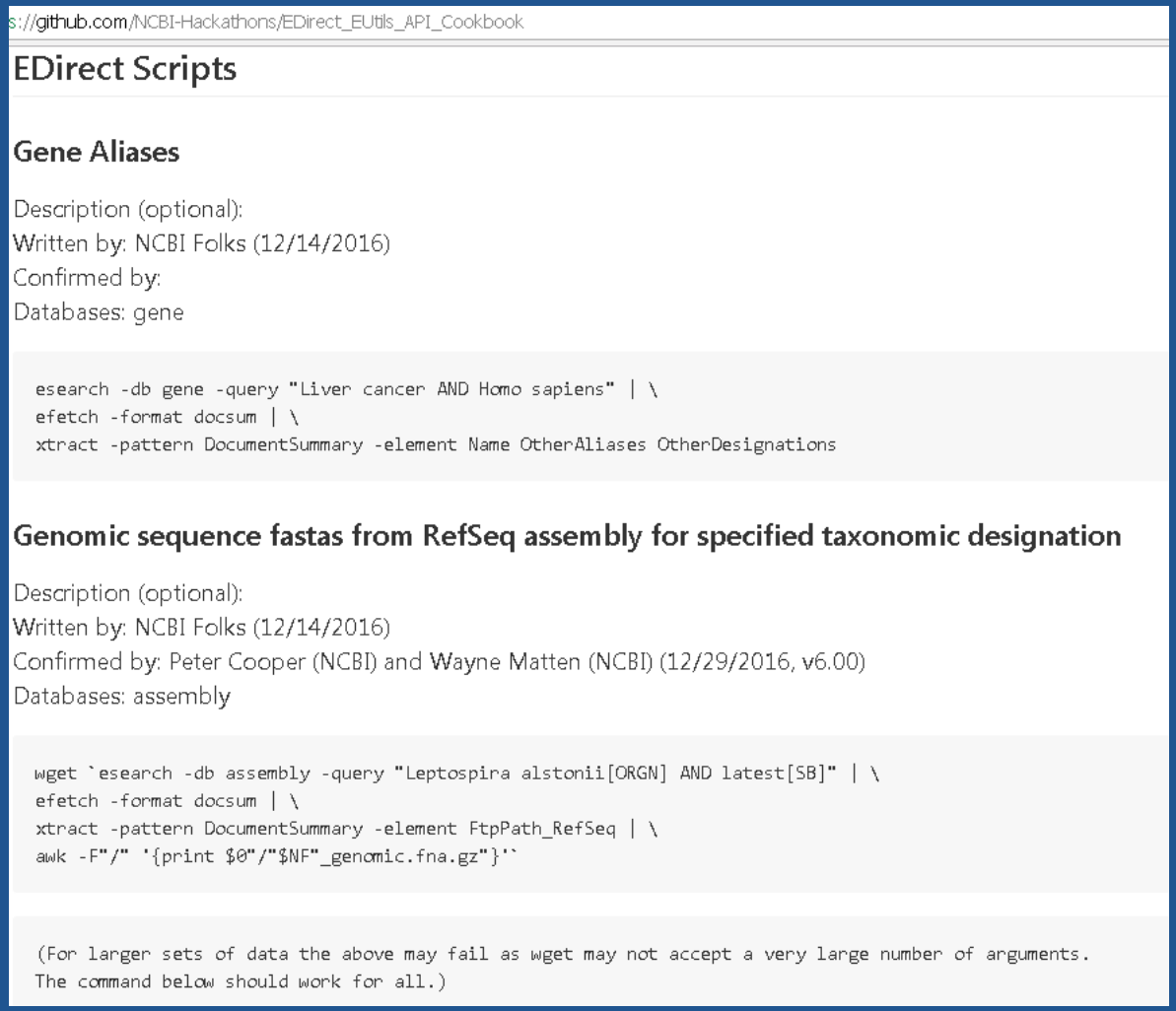
Chemicals

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!



EUtils (Search API) Command Line EDirect



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

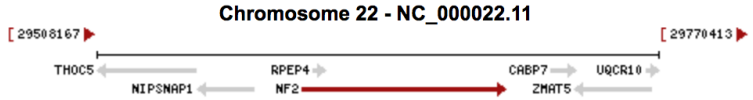
Summary

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

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)

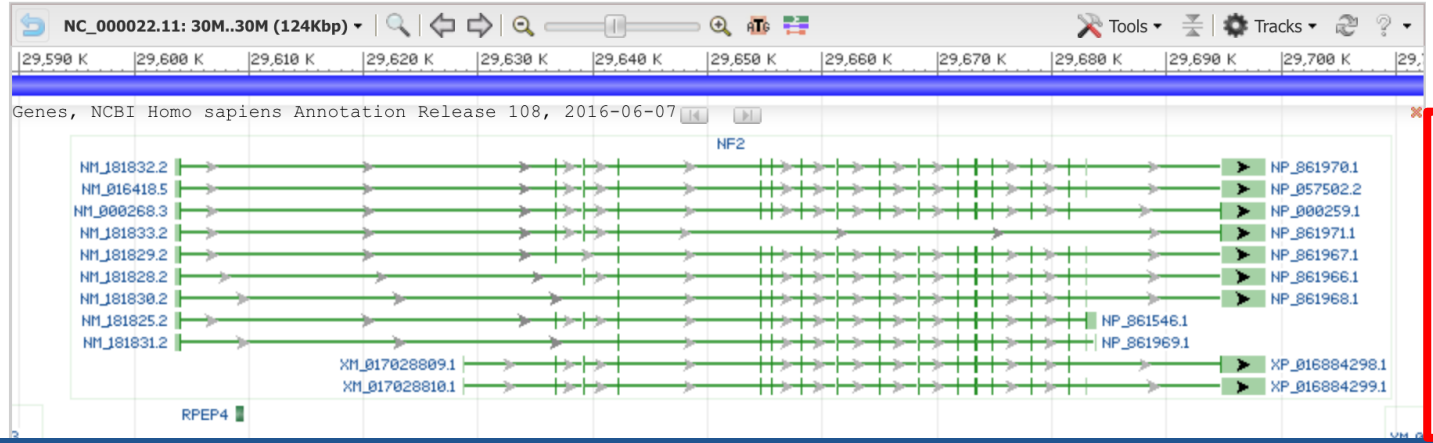


Genomic regions, transcripts, and products

Go to reference sequence details

Genomic Sequence: NC_000022.11 Chromosome 22 Reference GRCh38.p7 Primary Assembly

Go to nucleotide: Graphics FASTA GenBank

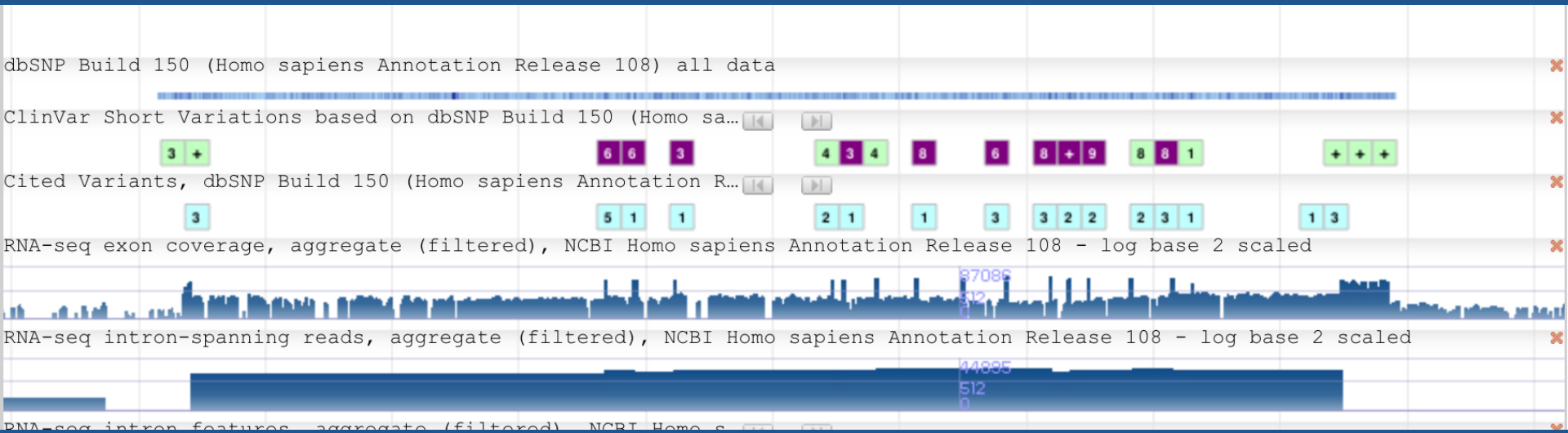


Variation Viewer (GRCh37.p13)
Variation Viewer (GRCh38)
1000 Genomes Browser (GRCh37.p13)
Ensembl
UCSC

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

Professional guidelines

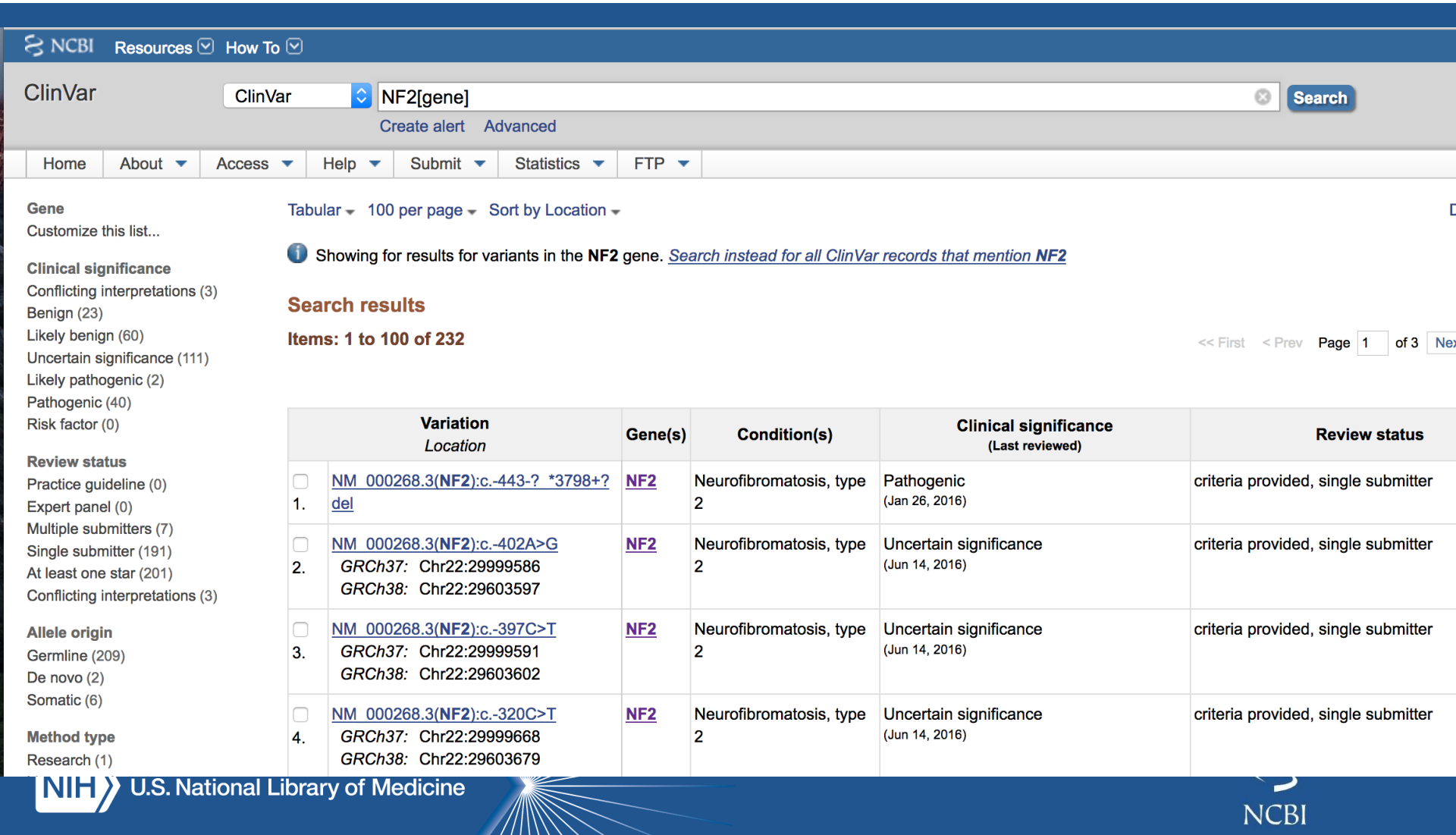
Description

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



NCBI Resources How To

ClinVar | ClinVar | NF2[gene] | Search

Create alert Advanced

Home | About | Access | Help | Submit | Statistics | FTP

Gene
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)

Tabular | 100 per page | Sort by Location

Showing for results for variants in the **NF2** gene. *Search instead for all ClinVar records that mention **NF2***

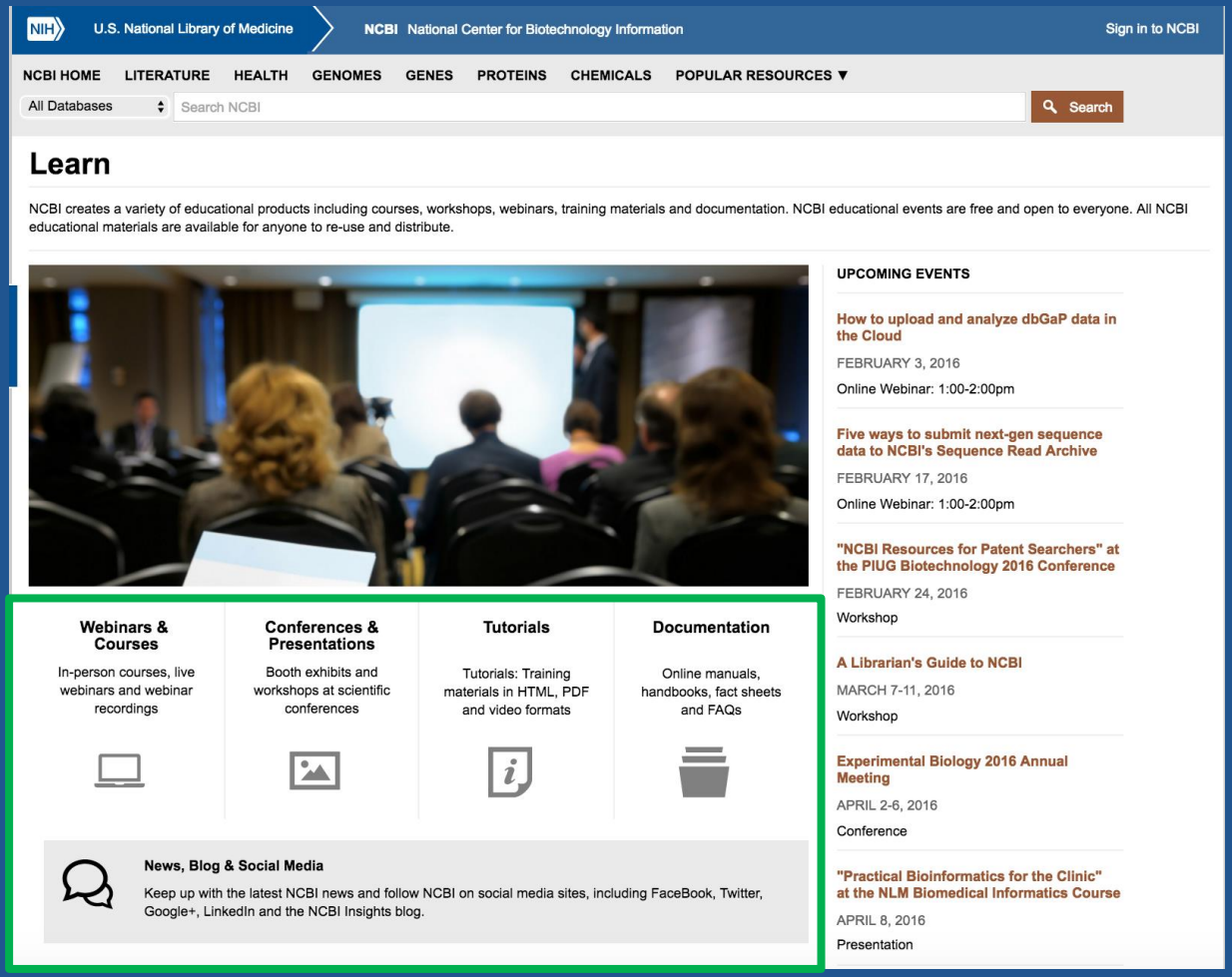
Search results

Items: 1 to 100 of 232

<< First < Prev Page 1 of 3 Nex

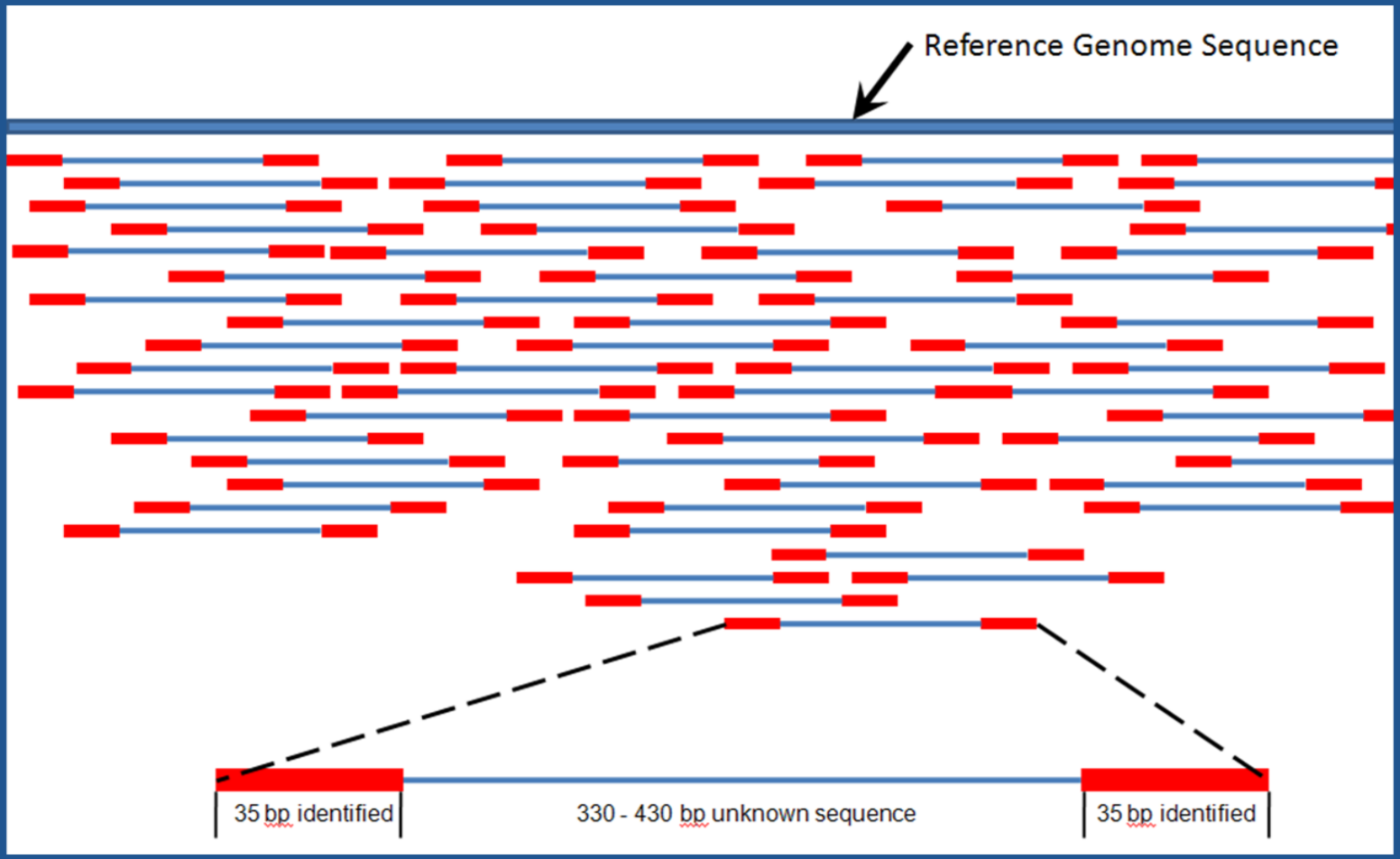
	Variation *Location*	Gene(s)	Condition(s)	Clinical significance (Last reviewed)	Review status
1.	NM_000268.3(**NF2**):c.-443-? *3798+? del	NF2	Neurofibromatosis, type 2	Pathogenic (Jan 26, 2016)	criteria provided, single submitter
2.	NM_000268.3(**NF2**):c.-402A>G *GRCh37:* Chr22:29999586 *GRCh38:* Chr22:29603597	NF2	Neurofibromatosis, type 2	Uncertain significance (Jun 14, 2016)	criteria provided, single submitter
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
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



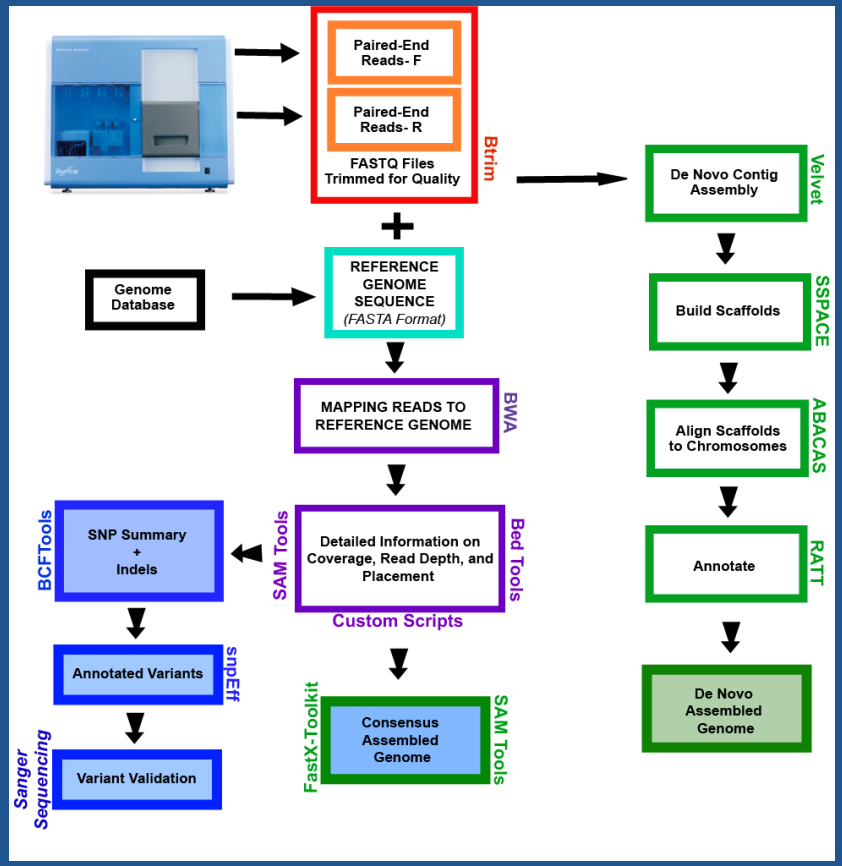
Review of terminology and concepts

Next Generation Sequencing



Review of terminology and concepts

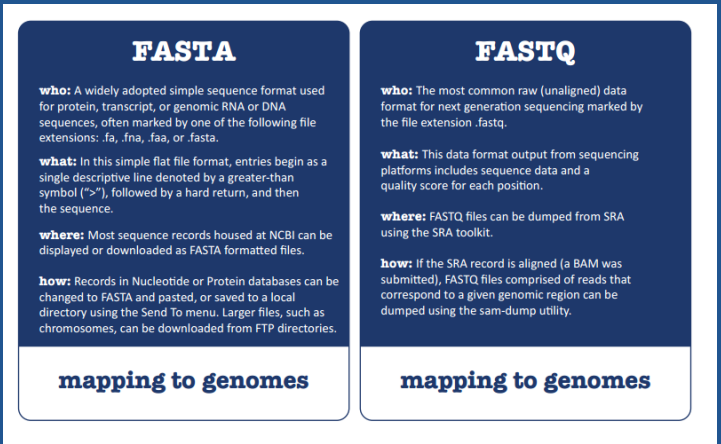
How Genomes are Mapped and Assembled

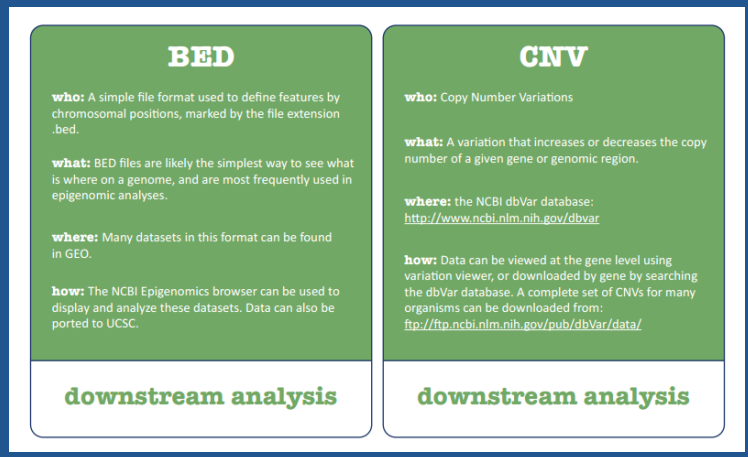


Review of terminology and concepts

How Genomes are Mapped and Assembled

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



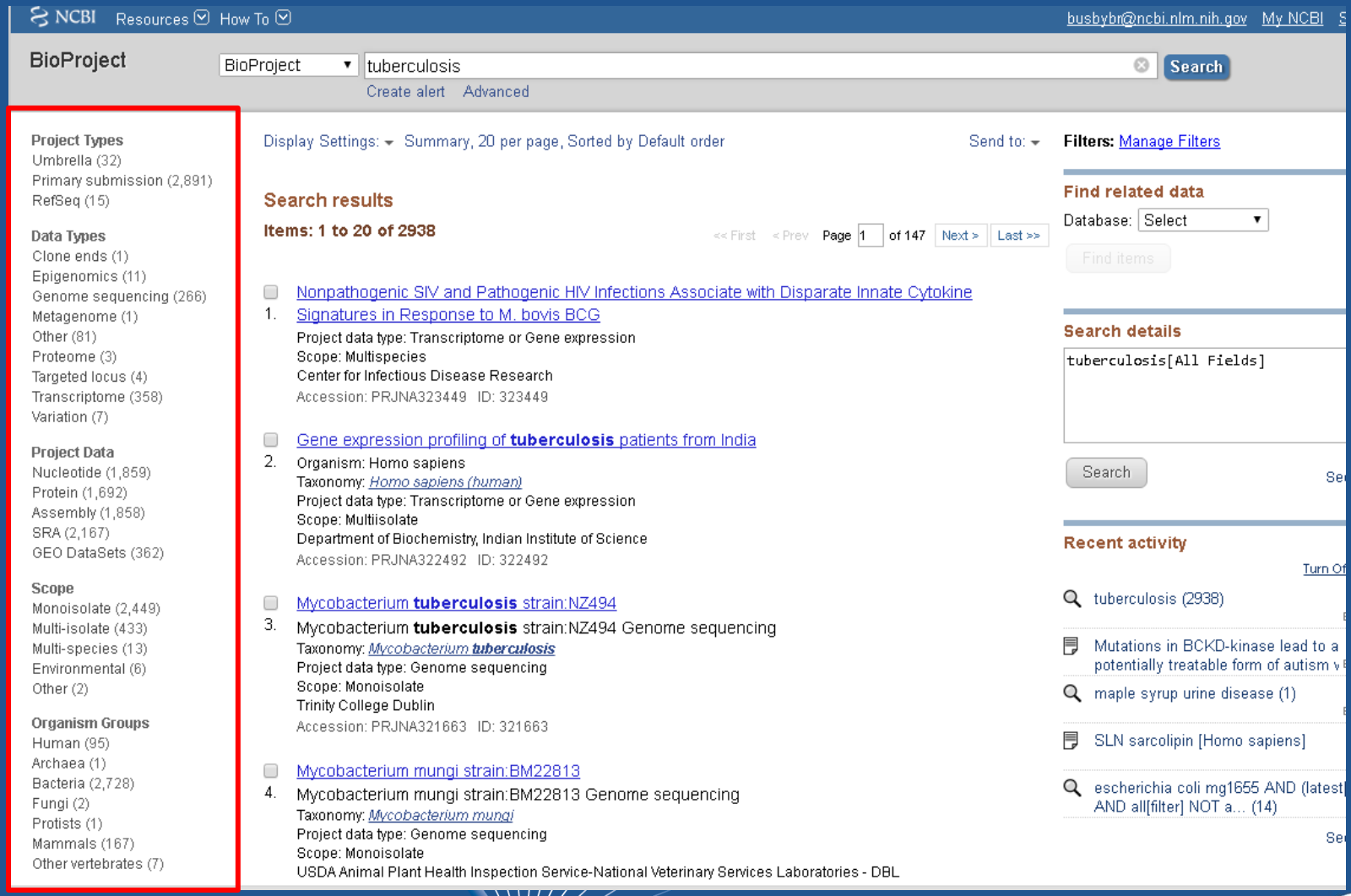


NCBI NGS Online Workshop – Available on the NCBI YouTube Channel!

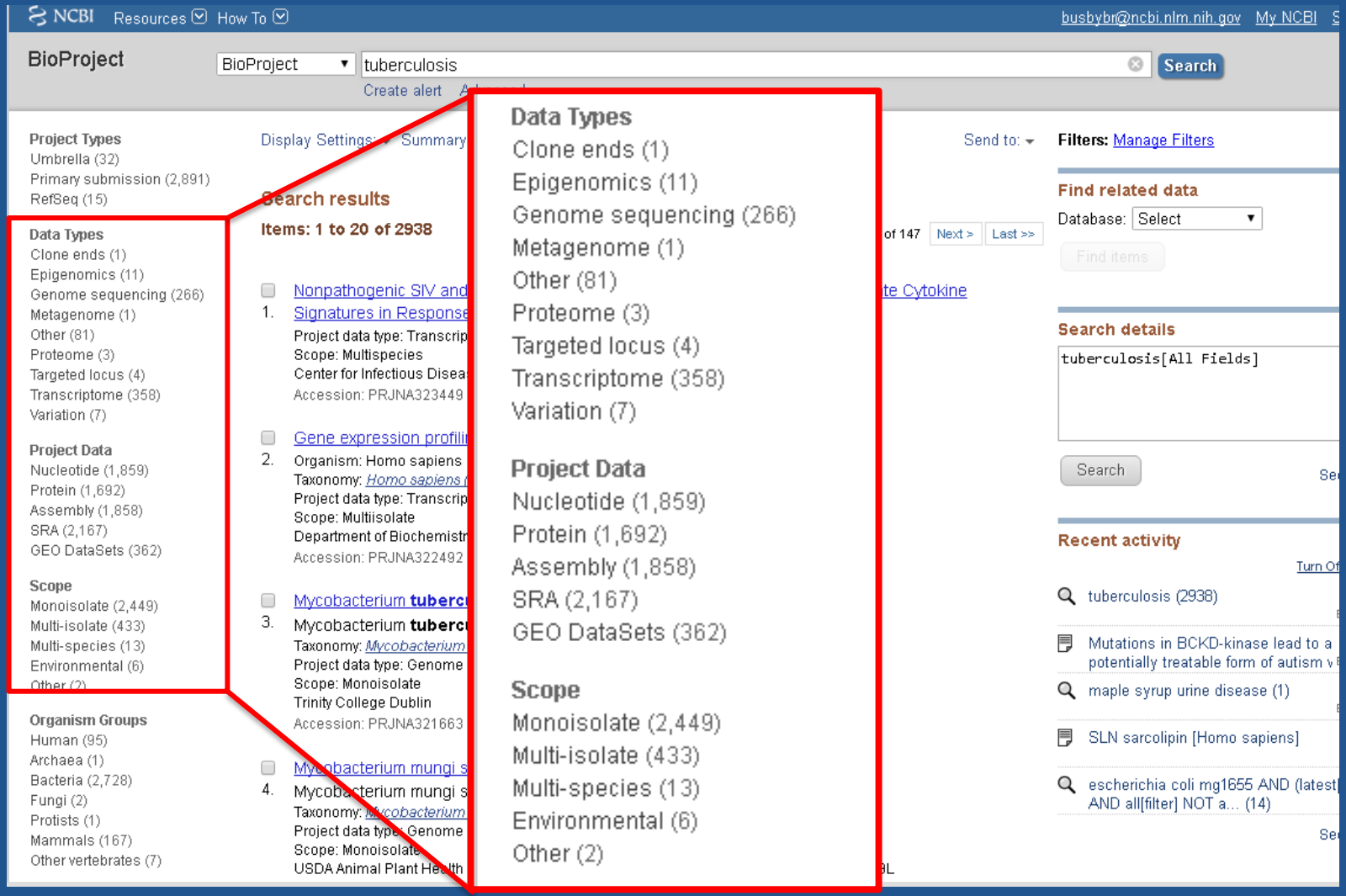
My View of Data Transfer Principles

- 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

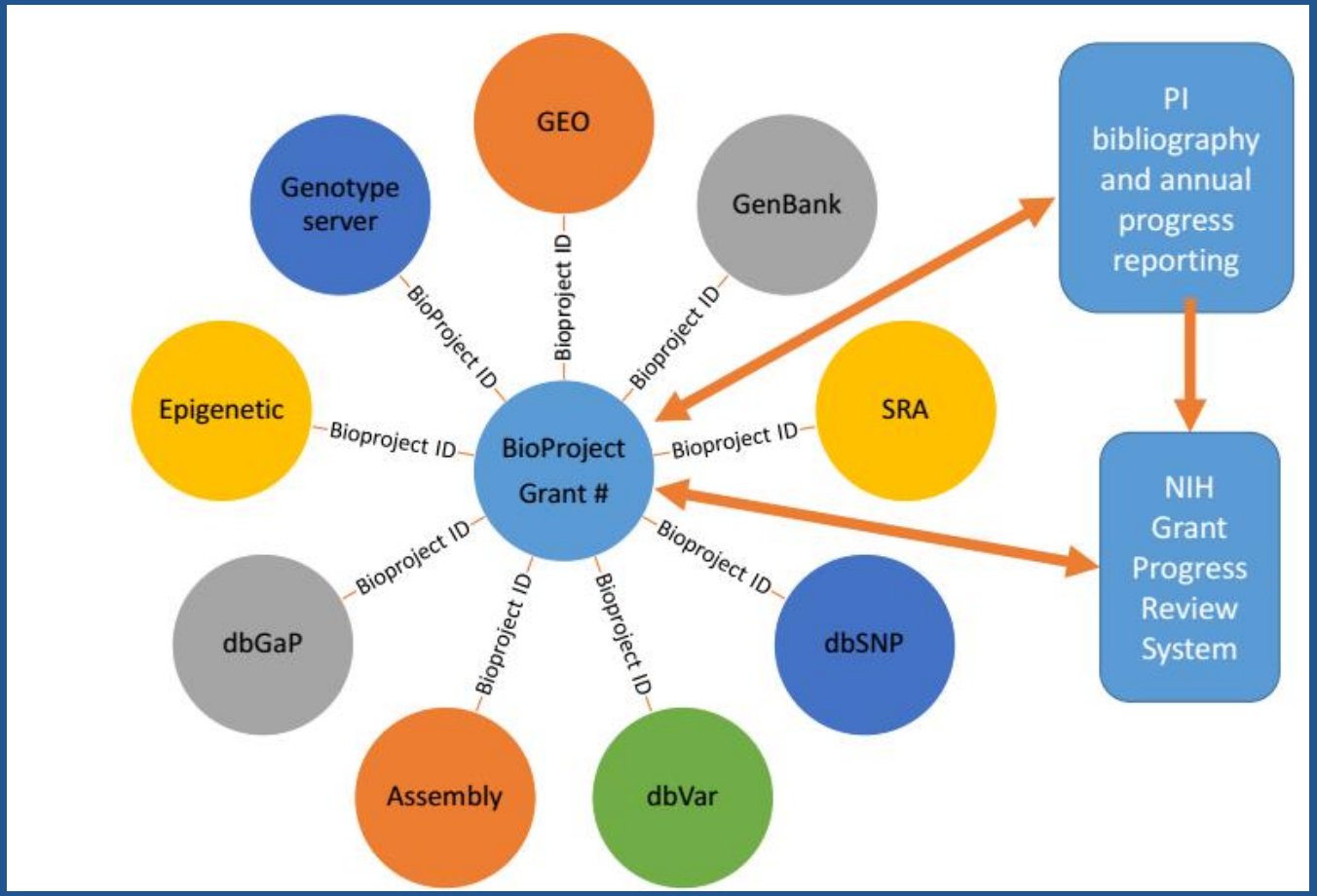
BioProject


NCBI Resources How To
busbybr@ncbi.nlm.nih.gov My NCBI
BioProject
BioProject
tuberculosis
Search
Create alert Advanced
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, 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: Multiisolate
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
Find related data
Database: Select
Find items
Search details
tuberculosis[All Fields]
Search
Recent activity
tuberculosis (2938)
Mutations in BCKD-kinase lead to a potentially treatable form of autism
maple syrup urine disease (1)
SLN sarcolipin [Homo sapiens]
escherichia coli mg1655 AND (latest AND all[filter] NOT a... (14)

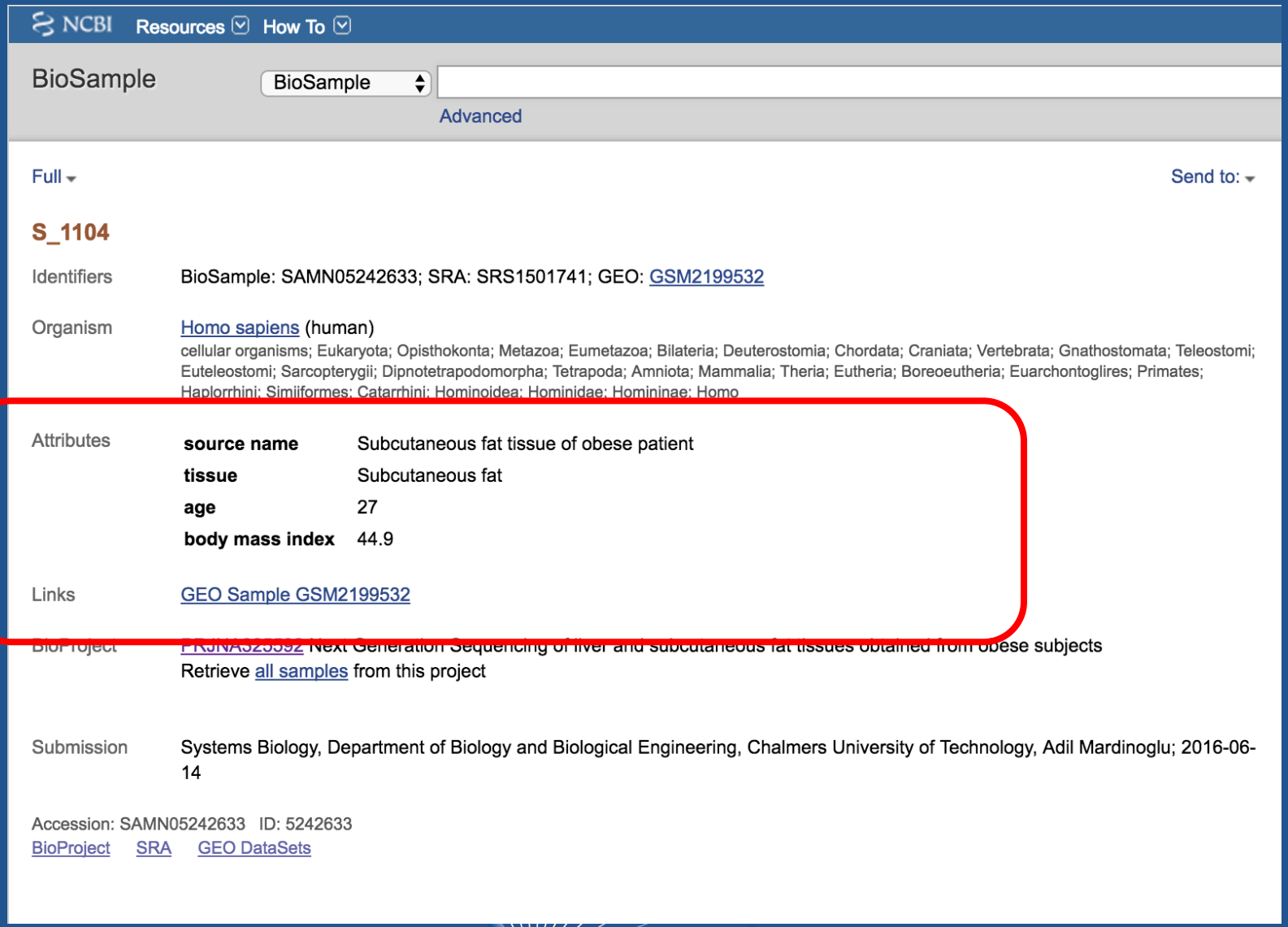
BioProject


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
Send to:
Search results
Items: 1 to 20 of 2938
of 147 Next > Last >>
1. Nonpathogenic SIV and
Signatures in Response
Project data type: Transcrip
Scope: Multispecies
Center for Infectious Disea
Accession: PRJNA323449
te Cytokine
2. Gene expression profili
Organism: Homo sapiens
Taxonomy: Homo sapiens
Project data type: Transcrip
Scope: Multiisolate
Department of Biochemistr
Accession: PRJNA322492
3. Mycobacterium tuberc
Mycobacterium tuberc
Taxonomy: Mycobacterium
Project data type: Genome
Scope: Monoisolate
Trinity College Dublin
Accession: PRJNA321663
4. Mycobacterium mungi s
Mycobacterium mungi s
Taxonomy: Mycobacterium
Project data type: Genome
Scope: Monoisolate
USDA Animal Plant Health
Filters: Manage Filters
Find related data
Database: Select
Find items
Search details
tuberculosis[All Fields]
Search
Recent activity
Turn Of
tuberculosis (2938)
Mutations in BCKD-kinase lead to a potentially treatable form of autism
maple syrup urine disease (1)
SLN sarcolipin [Homo sapiens]
escherichia coli mg1655 AND (latest AND all[filter] NOT a... (14)
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)

Reporting


GEO
Genotype server
GenBank
PI bibliography and annual progress reporting
Epigenetic
SRA
BioProject Grant #
NIH Grant Progress Review System
dbGaP
dbSNP
Assembly
dbVar
BioProject ID
Bioproject ID

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
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 MIxS, 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 MIxS 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 appro

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/.

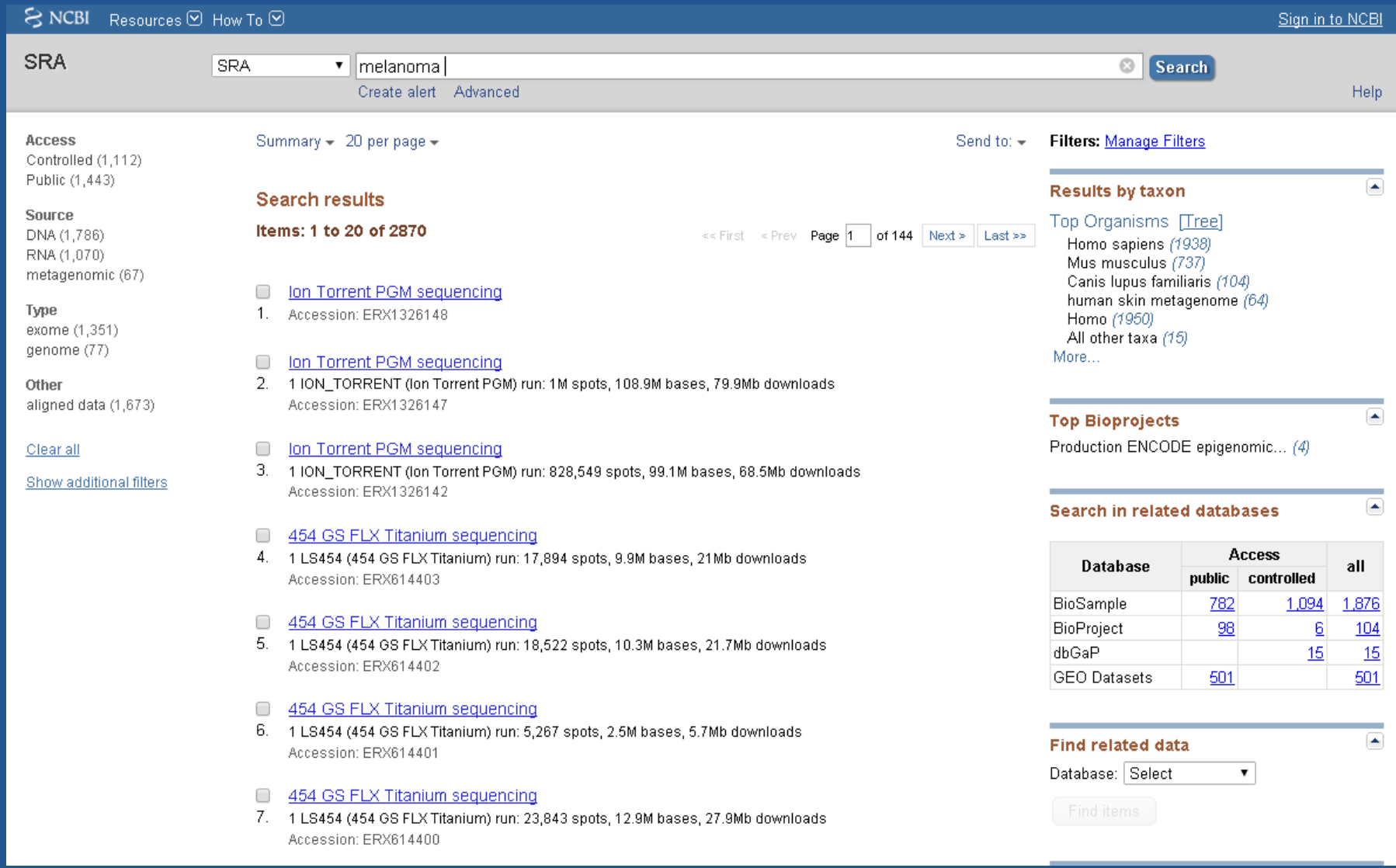
CAUTION: Be aware that Excel may automatically apply formatting to your data. In particular, take care with dates, incrementing autofills and special characters like / or -. Doublecheck that your text file is accurate be

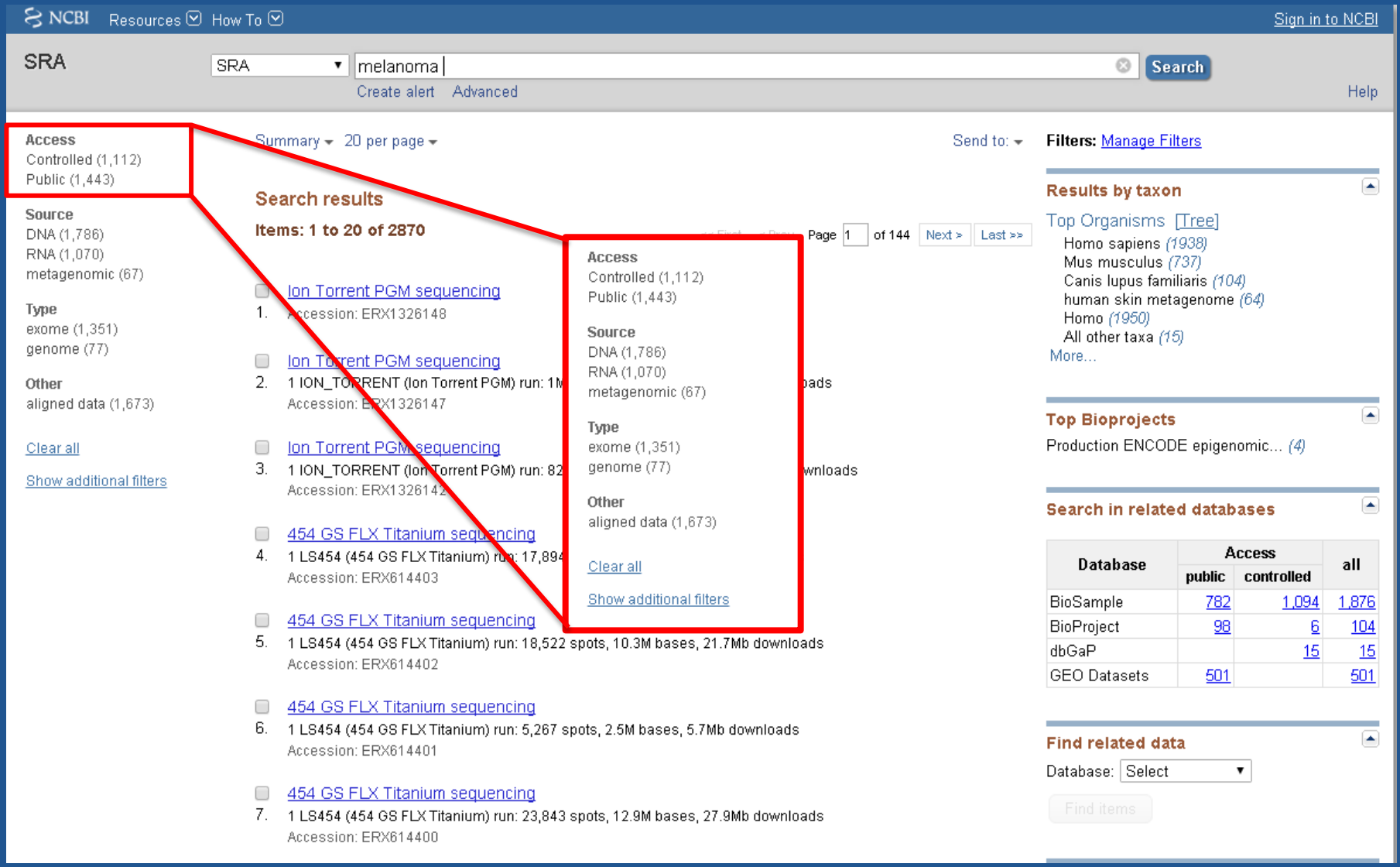
TO MAKE A SUBMISSION:

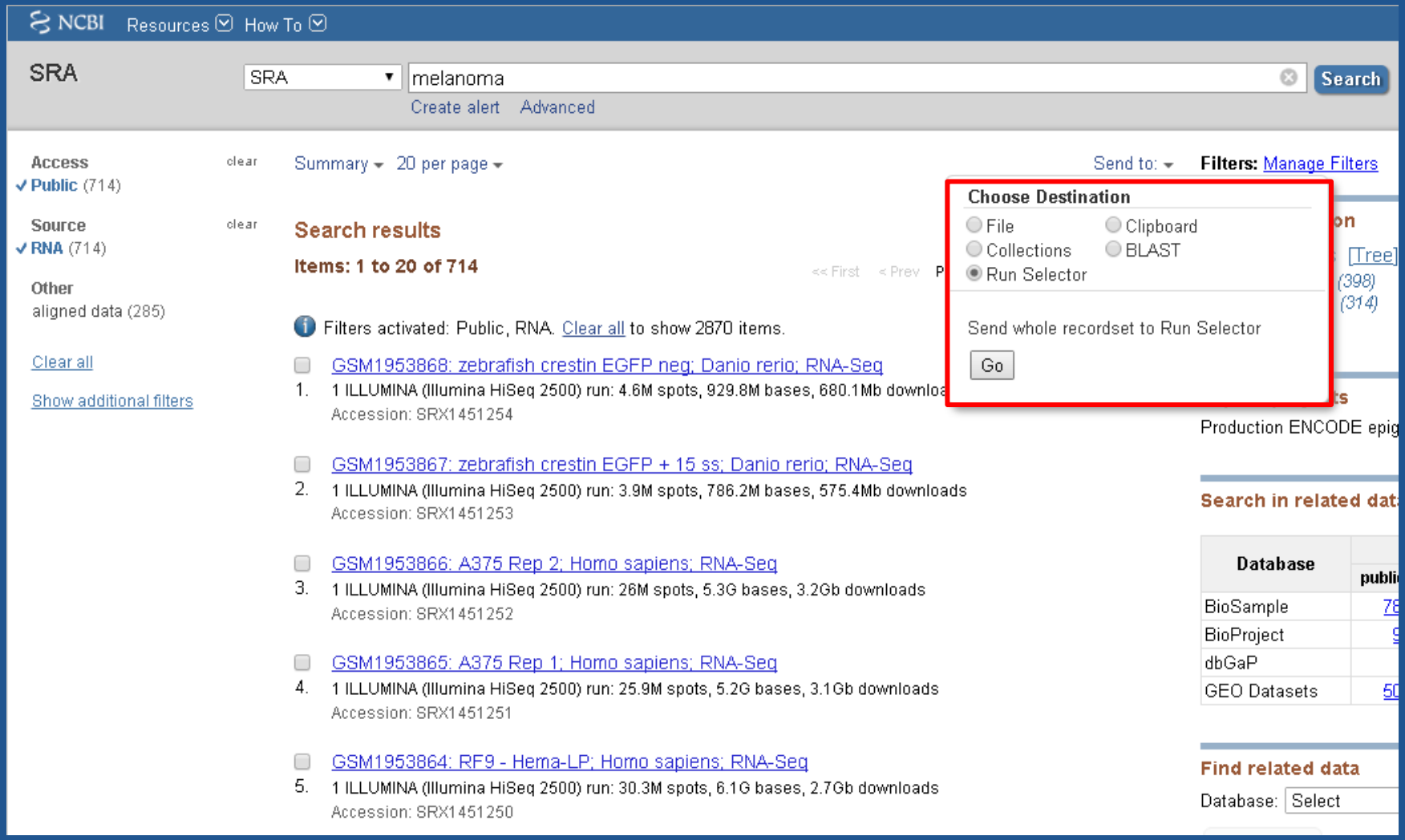
1. Complete this template table
2. Save the worksheet as a Text (Tab-delimited) file -- (use 'File, Save as, Save as type: Text (Tab-delimited)')
3. Upload the text file on the 'Attributes' tab of the BioSample Submission Portal at https://submit.ncbi.nlm.nih.gov/subs/biosample/.

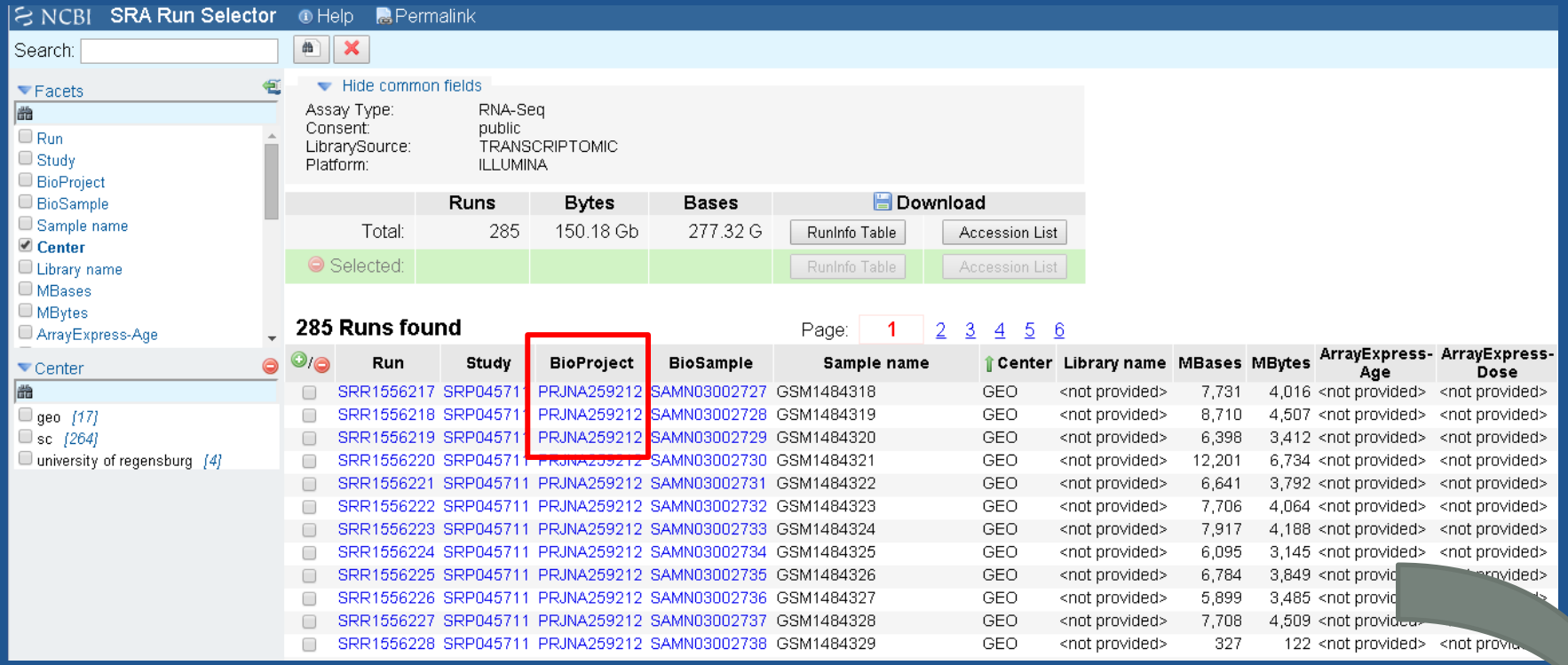
If you have any questions, please contact us at biosamplehelp@ncbi.nlm.nih.gov.

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


NCBI Resources How To
Sign in to NCBI
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:
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
Filters: Manage 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
public
controlled
all
BioSample 782 1,094 1,876
BioProject 98 6 104
dbGaP 15 15
GEO Datasets 501 501
Find related data
Database: Select
Find items


NCBI Resources How To
Sign in to NCBI
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:
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
Filters: Manage 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
public
controlled
all
BioSample 782 1,094 1,876
BioProject 98 6 104
dbGaP 15 15
GEO Datasets 501 501
Find related data
Database: Select
Find items

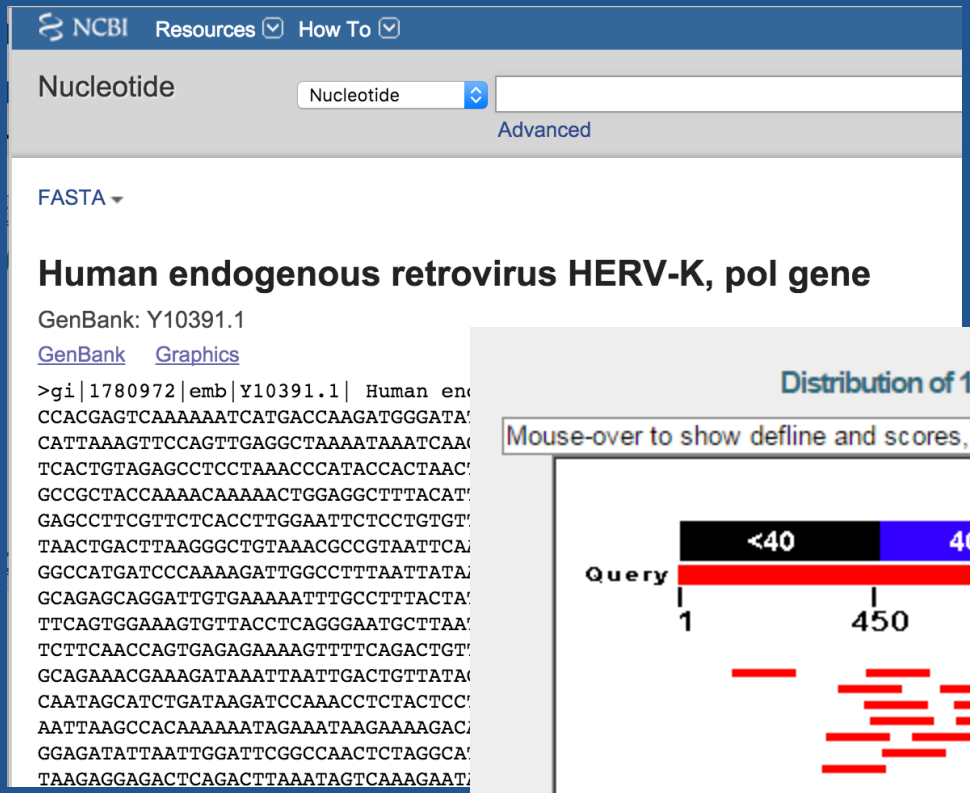

NCBI Resources How To
SRA
SRA
melanoma
Search
Create alert Advanced
Access clear
Public (714)
Source clear
RNA (714)
Other
aligned data (285)
Clear all
Show additional filters
Summary 20 per page
Send to:
Filters: Manage Filters
Search results
Items: 1 to 20 of 714
<< First < Prev
Choose Destination
File
Clipboard
Collections
BLAST
Run Selector
Send whole recordset to Run Selector
Go
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 downloa
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
[Tree]
(398)
(314)
Production ENCODE epig
Search in related dat
Database
BioSample
BioProject
dbGaP
GEO Datasets
Find related data
Database: Select

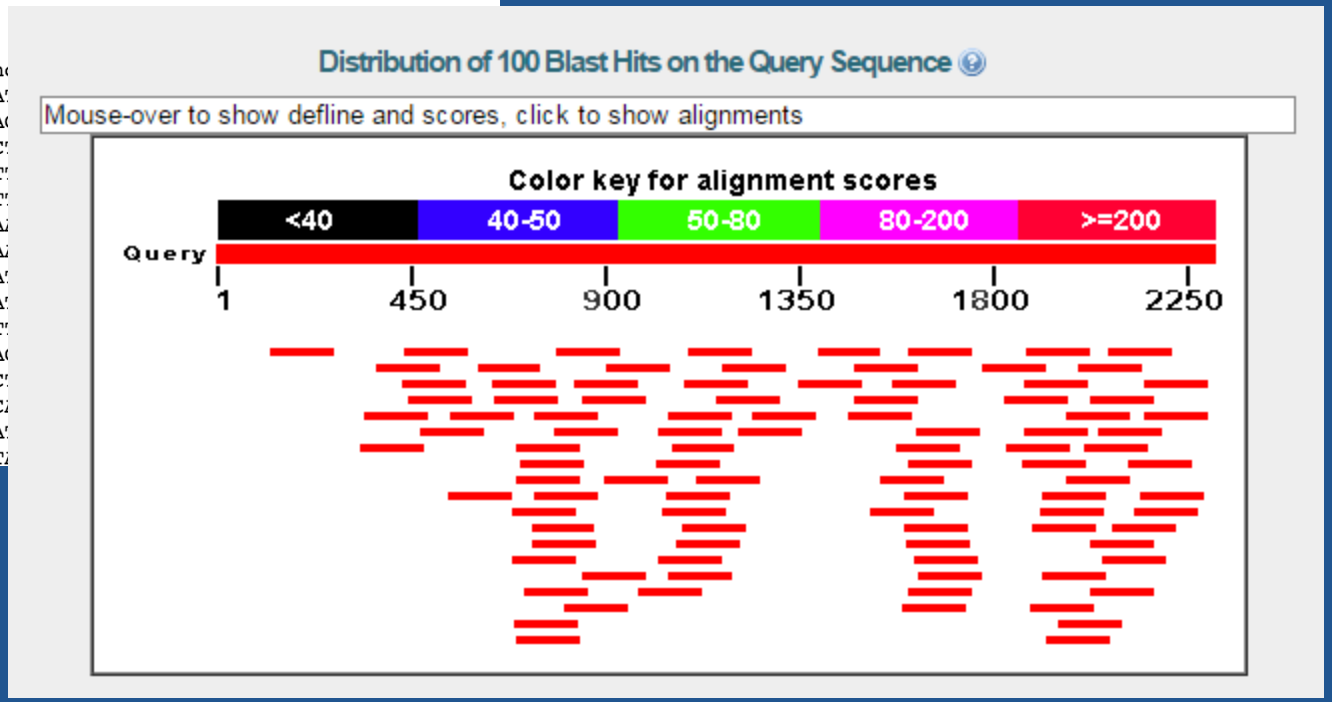


Run	Study	BioProject	BioSample	Sample name	Center	Library name	MBases	MBytes	ArrayExpress-Age	ArrayExpress-Dose
SRR1556217	SRP045711	PRJNA259212	SAMN03002727	GSM1484318	GEO	<not provided>	7,731	4,016	<not provided>	<not provided>
SRR1556218	SRP045711	PRJNA259212	SAMN03002728	GSM1484319	GEO	<not provided>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

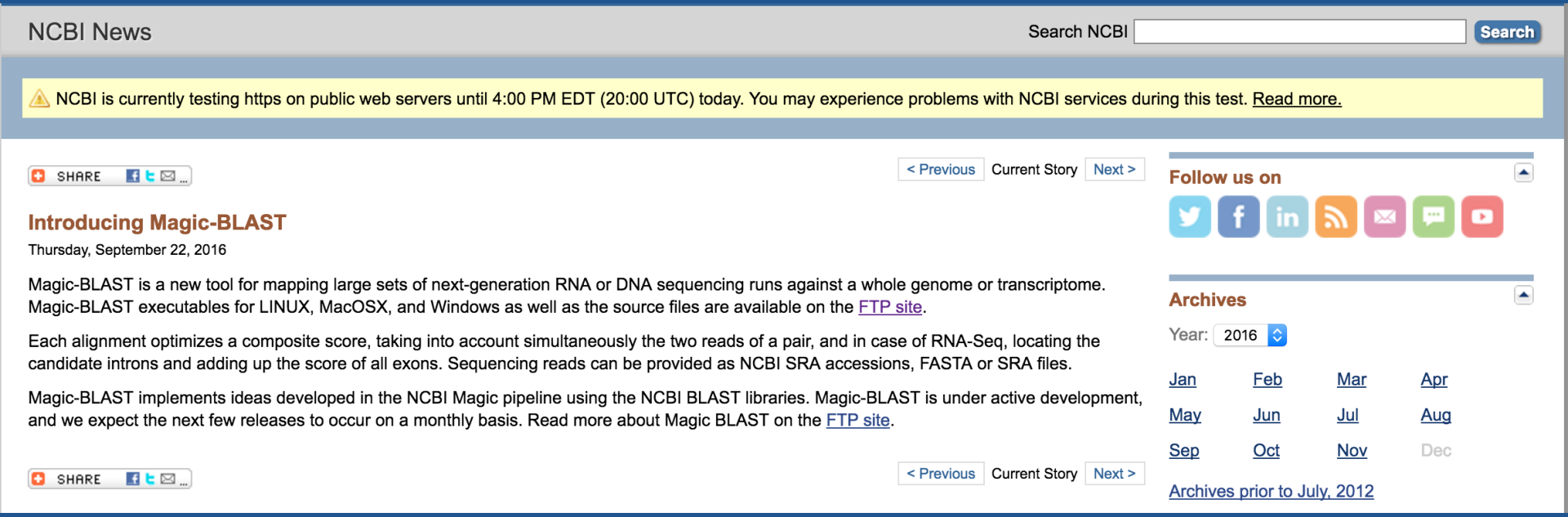
Experiment	InsertSize	LibraryLayout	LibrarySelection	LoadDate	Organism	ReleaseDate	SRA Sample	Sample Description	cell type	source name
SRX685301	0	SINGLE	cDNA	2014-08-22	Homo sapiens	2015-03-05	SRS688179	<not provided>	low passage primary melanoma cultures	low passage primary melanoma cultures
SRX685302	0	SINGLE	cDNA	2016-02-02	Homo sapiens	2015-03-05	SRS688181	<not provided>	low passage primary melanoma cultures	low passage primary melanoma cultures
SRX685303	0	SINGLE	cDNA	2016-02-02	Homo sapiens	2015-03-05	SRS688180	<not provided>	low passage primary melanoma cultures	low passage primary melanoma cultures
SRX685304	0	SINGLE	cDNA	2014-08-22	Homo sapiens	2015-03-05	SRS688182	<not provided>	low passage primary melanoma cultures	low passage primary melanoma cultures
SRX685305	0	SINGLE	cDNA	2014-09-08	Homo sapiens	2015-03-05	SRS688183	<not provided>	low passage primary melanoma cultures	low passage primary melanoma cultur
SRX685306	0	SINGLE	cDNA	2016-02-02	Homo sapiens	2015-03-05	SRS688184	<not provided>	low passage primary melanoma cultures	low passage primary melanoma cul
SRX685307	0	SINGLE	cDNA	2016-02-02	Homo sapiens	2015-03-05	SRS688185	<not provided>	low passage primary melanoma cultures	low passage primary melanoma
SRX685308	0	SINGLE	cDNA	2016-02-02	Homo sapiens	2015-03-05	SRS688186	<not provided>	low passage primary melanoma cultures	low passage primary melanoma c
SRX685309	0	SINGLE	cDNA	2016-02-02	Homo sapiens	2015-03-05	SRS688187	<not provided>	low passage primary melanoma cultures	low passage primary melanoma cultu
SRX685310	0	SINGLE	cDNA	2016-02-02	Homo sapiens	2015-03-05	SRS688188	<not provided>	low passage primary melanoma cultures	low passage primary melanoma cultures
SRX685311	0	SINGLE	cDNA	2014-08-22	Homo sapiens	2015-03-05	SRS688189	<not provided>	melanoma cell line	melanoma cell line
SRX685312	0	SINGLE	cDNA	2014-08-22	Homo sapiens	2015-03-05	SRS688190	<not provided>	low passage primary melanoma cultures	low passage primary melanoma cultures
SRX685313	0	SINGLE	cDNA	2014-08-22	Homo sapiens	2015-03-05	SRS688191	<not provided>	low passage primary melanoma cultures	low passage primary melanoma cultures
SRX685314	0	SINGLE	cDNA	2014-08-22	Homo sapiens	2015-03-05	SRS688192	<not provided>	low passage primary melanoma cultures	low passage primary melanoma cultures
SRX685315	0	SINGLE	cDNA	2014-08-22	Homo sapiens	2015-03-05	SRS688193	<not provided>	low passage primary melanoma cultures	low passage primary melanoma cultures
SRX685316	0	SINGLE	cDNA	2014-08-22	Homo sapiens	2015-03-05	SRS688194	<not provided>	low passage primary melanoma cultures	low passage primary melanoma cultures
SRX685317	0	SINGLE	cDNA	2014-08-22	Homo sapiens	2015-03-05	SRS688195	<not provided>	low passage primary melanoma cultures	low passage primary melanoma cultures

Investigation of NGS: SRA BLAST!





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

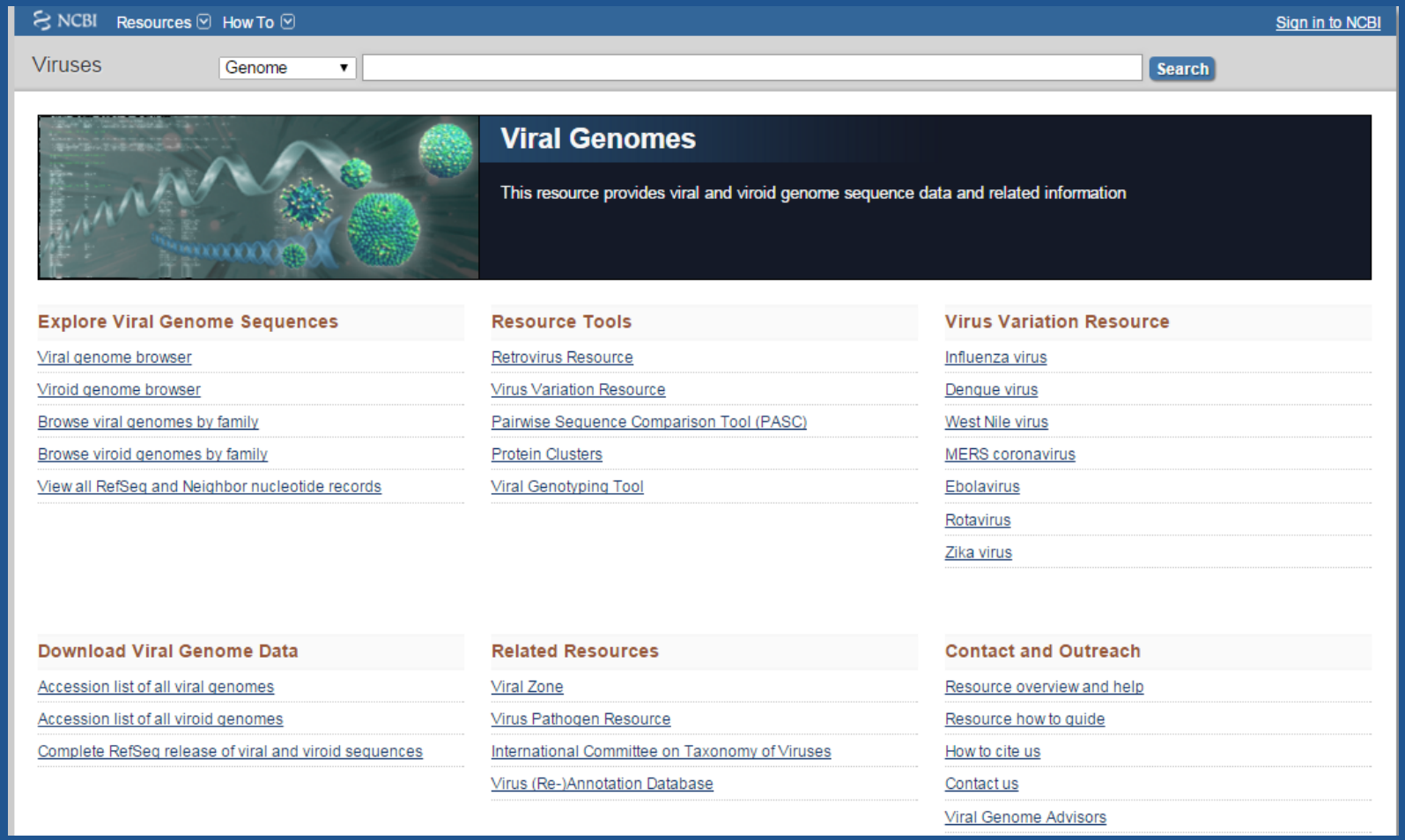
Jan Feb Mar Apr
May Jun Jul Aug
Sep Oct Nov Dec

Archives prior to July, 2012

NIH U.S. National Library of Medicine

NCBI

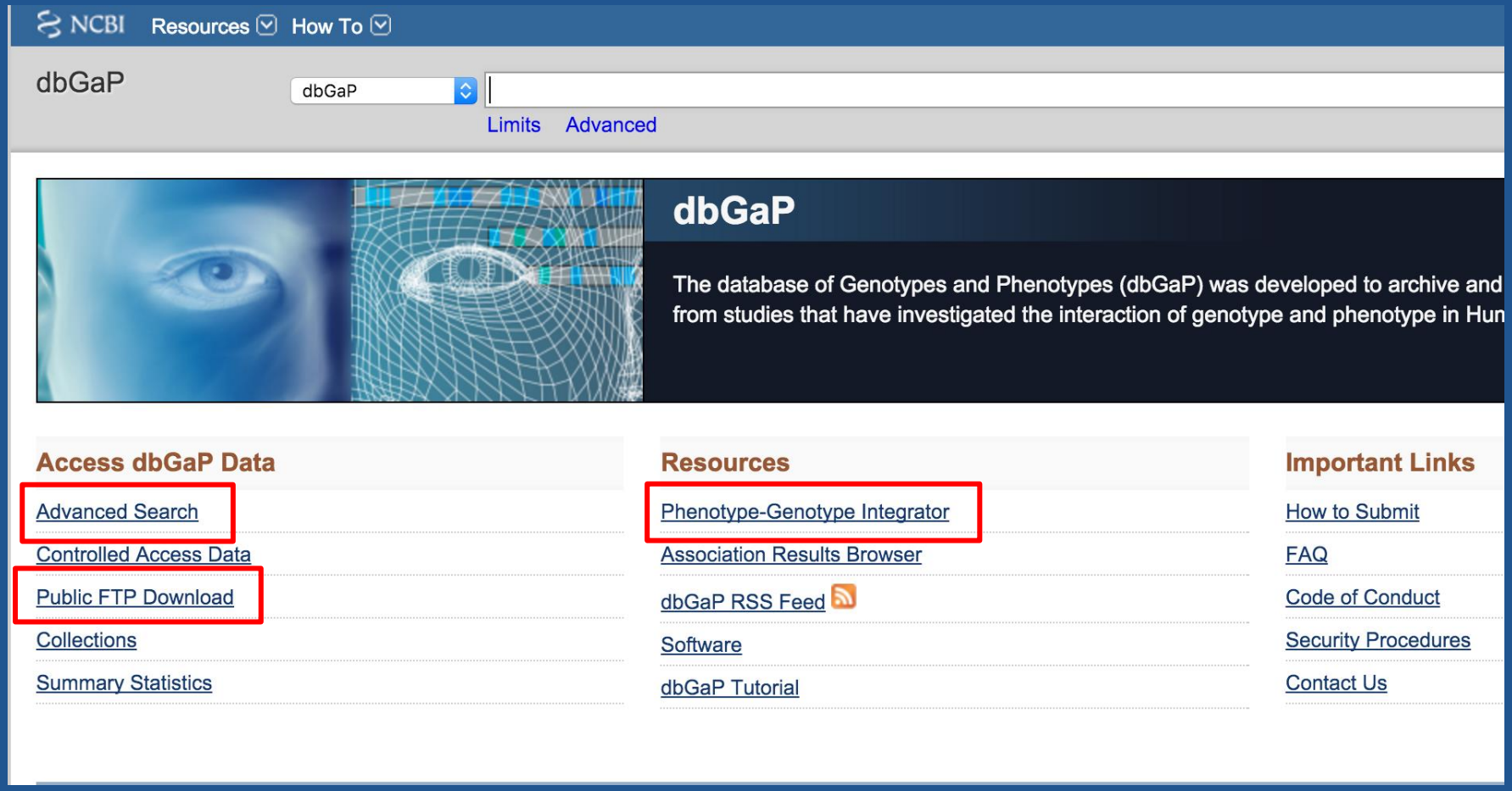
Viral Genomes



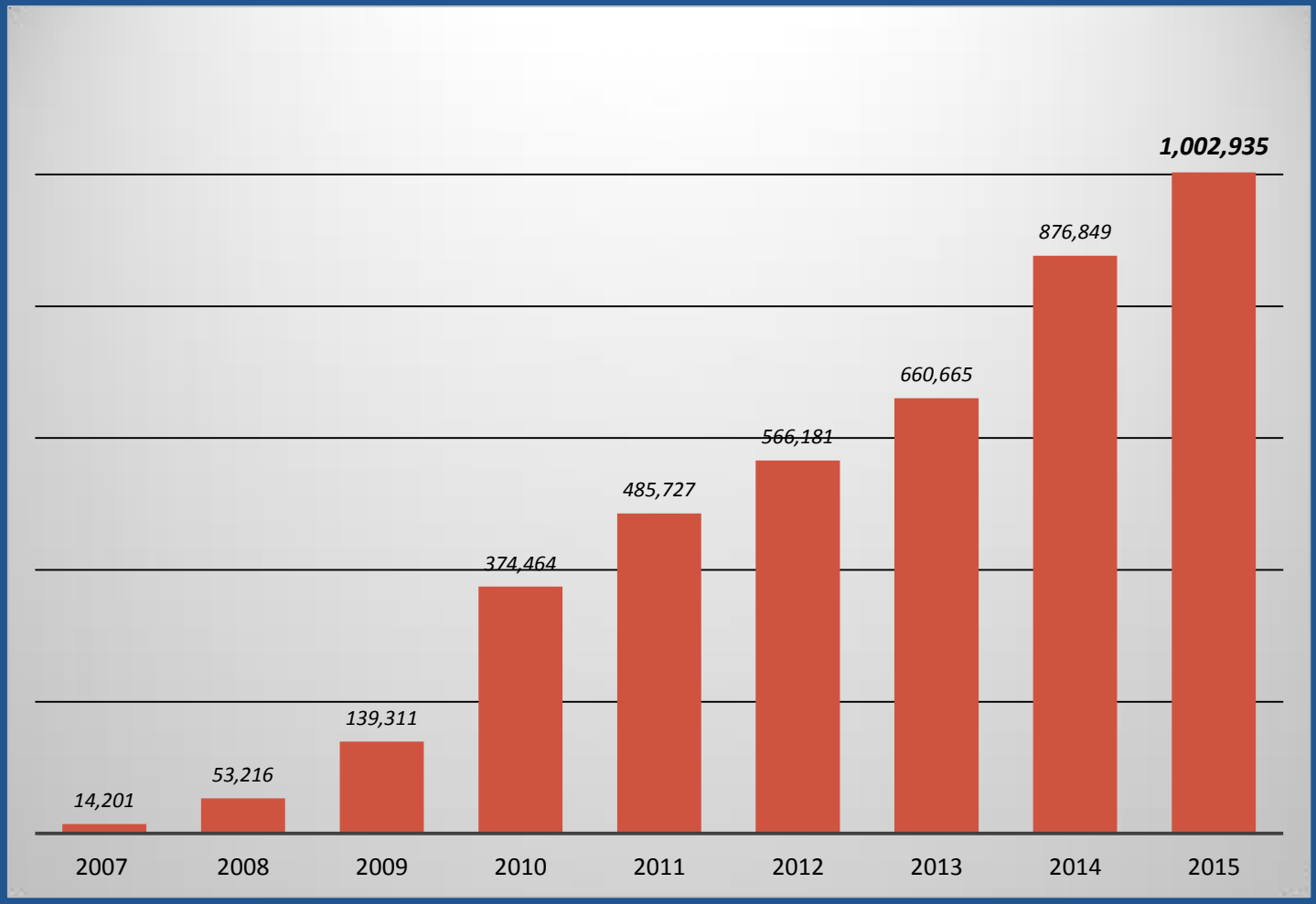
Virus Variation

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


NCBI
Resources
How To
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


14,201
53,216
139,311
374,464
485,727
566,181
660,665
876,849
1,002,935
2007
2008
2009
2010
2011
2012
2013
2014
2015

dbGaP Advanced Search

Study Disease/Focus IS Melanoma

Type Keyword or Phrase

Study Disease/Focus (326)

melanoma Sort By Alphabetical clear

- [x] 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 Markerset (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 Markerset	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 Markerset	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

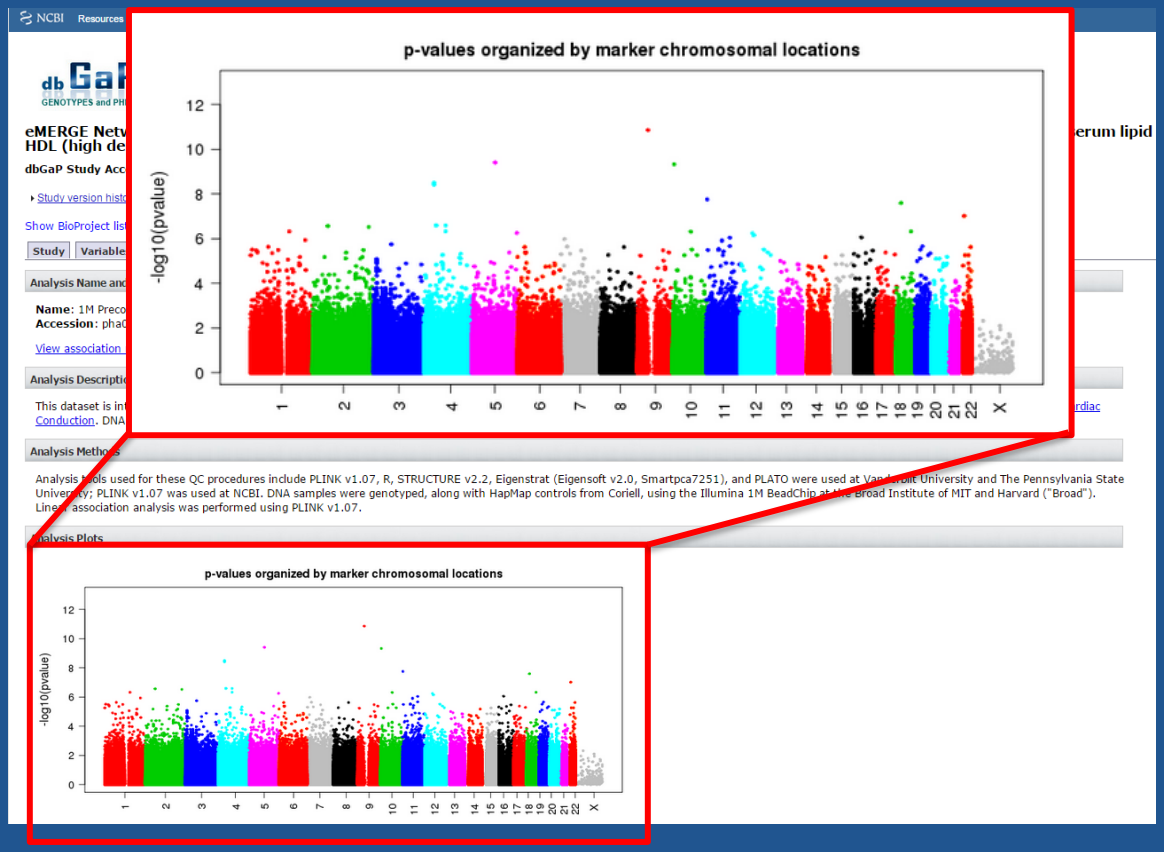
We conducted whole-genome sequencing of probands from several melanoma families, identifying one individual carrying a novel germline variant (c.G1075A, NM_000248.3; p.E318K, NP_000239.1; rs149617956) in the melanoma lineage-specific oncogene MITF. … While the variant cosegregated with melanoma in some,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

FileSelector RunSelector PubMed PMC MeSH BioProject BioSample SRA

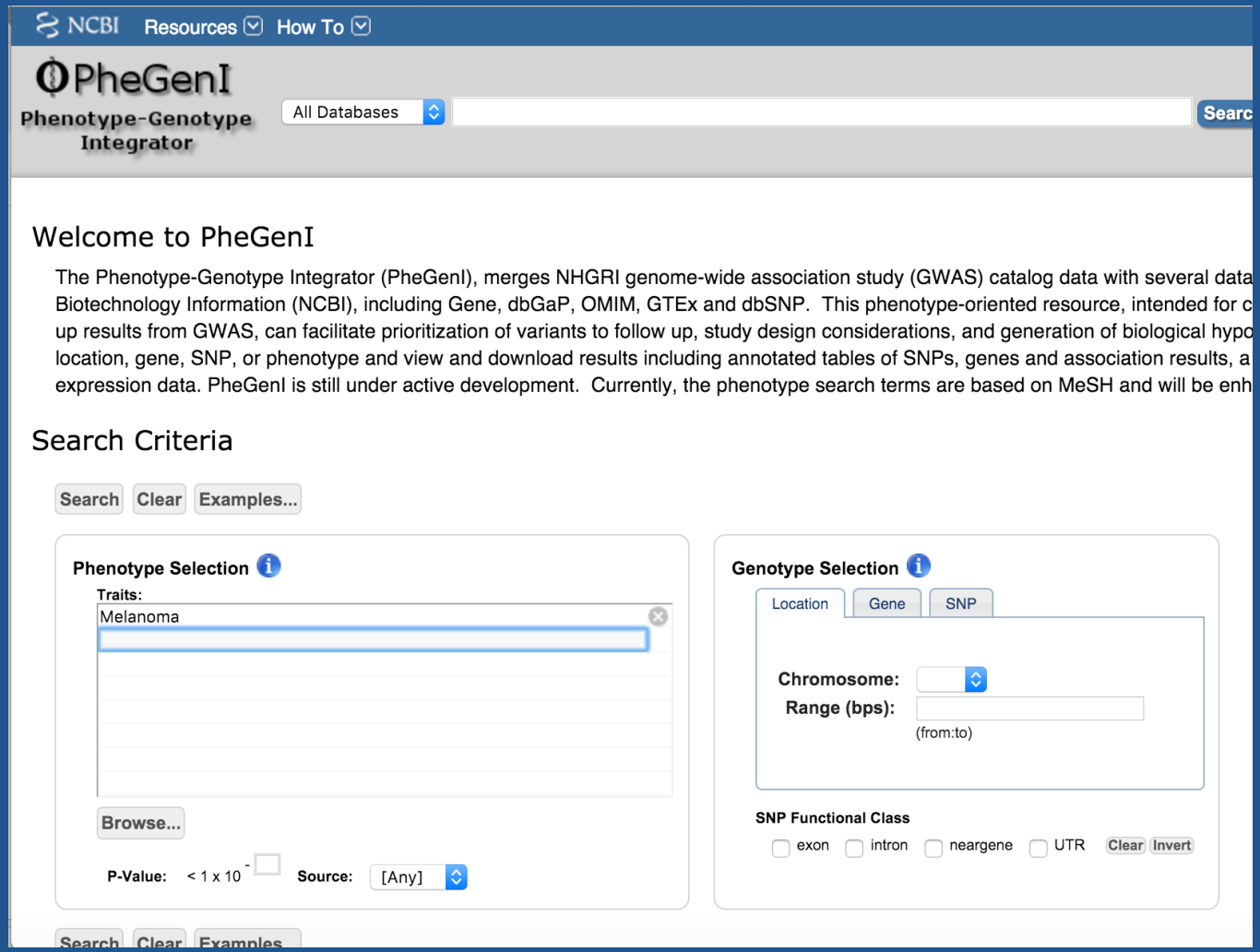
Melanoma Genome Sequencing Project

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

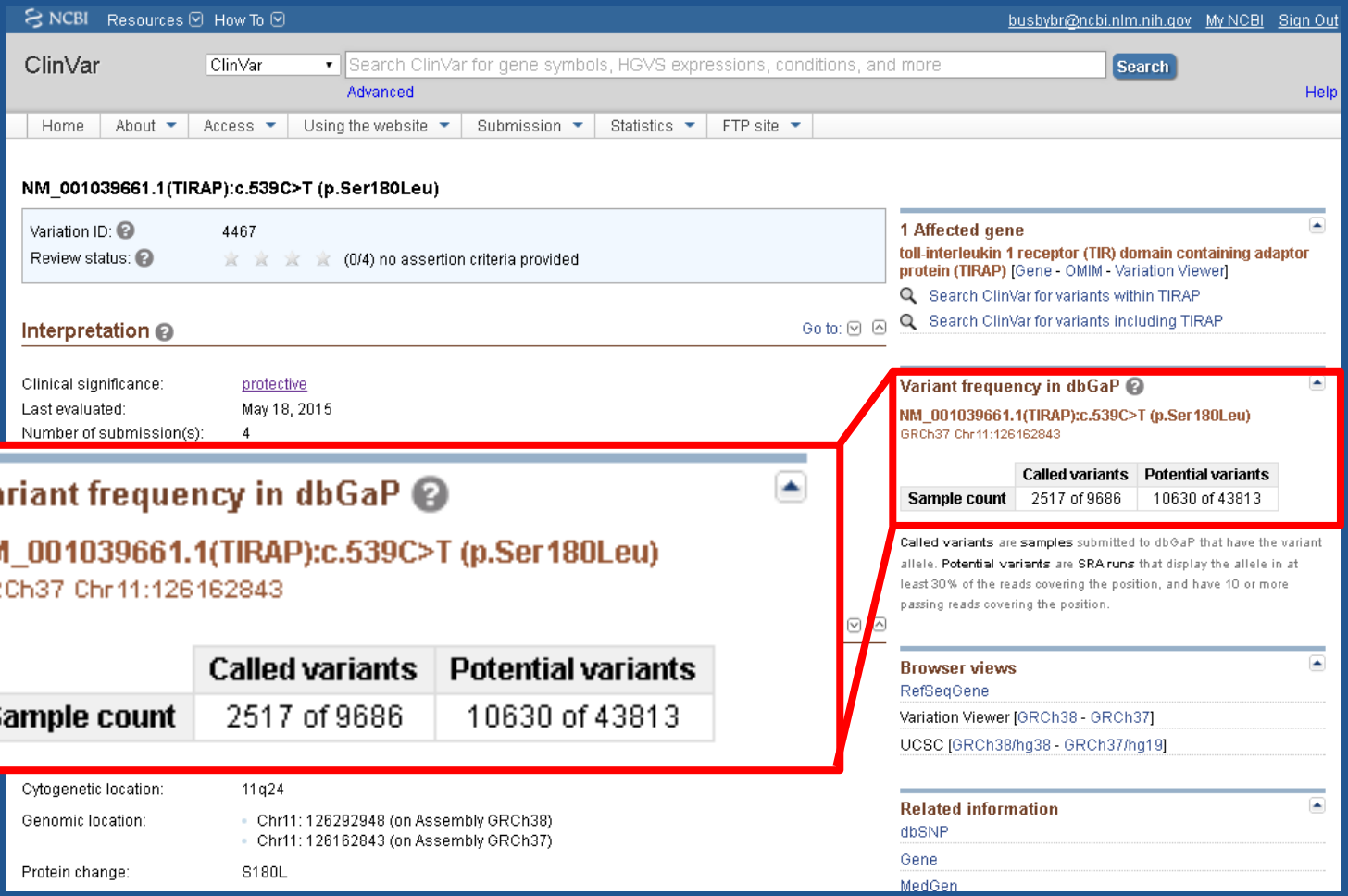
dbGaP – GWAS and PheGenI



dbGaP – GWAS and PheGenI



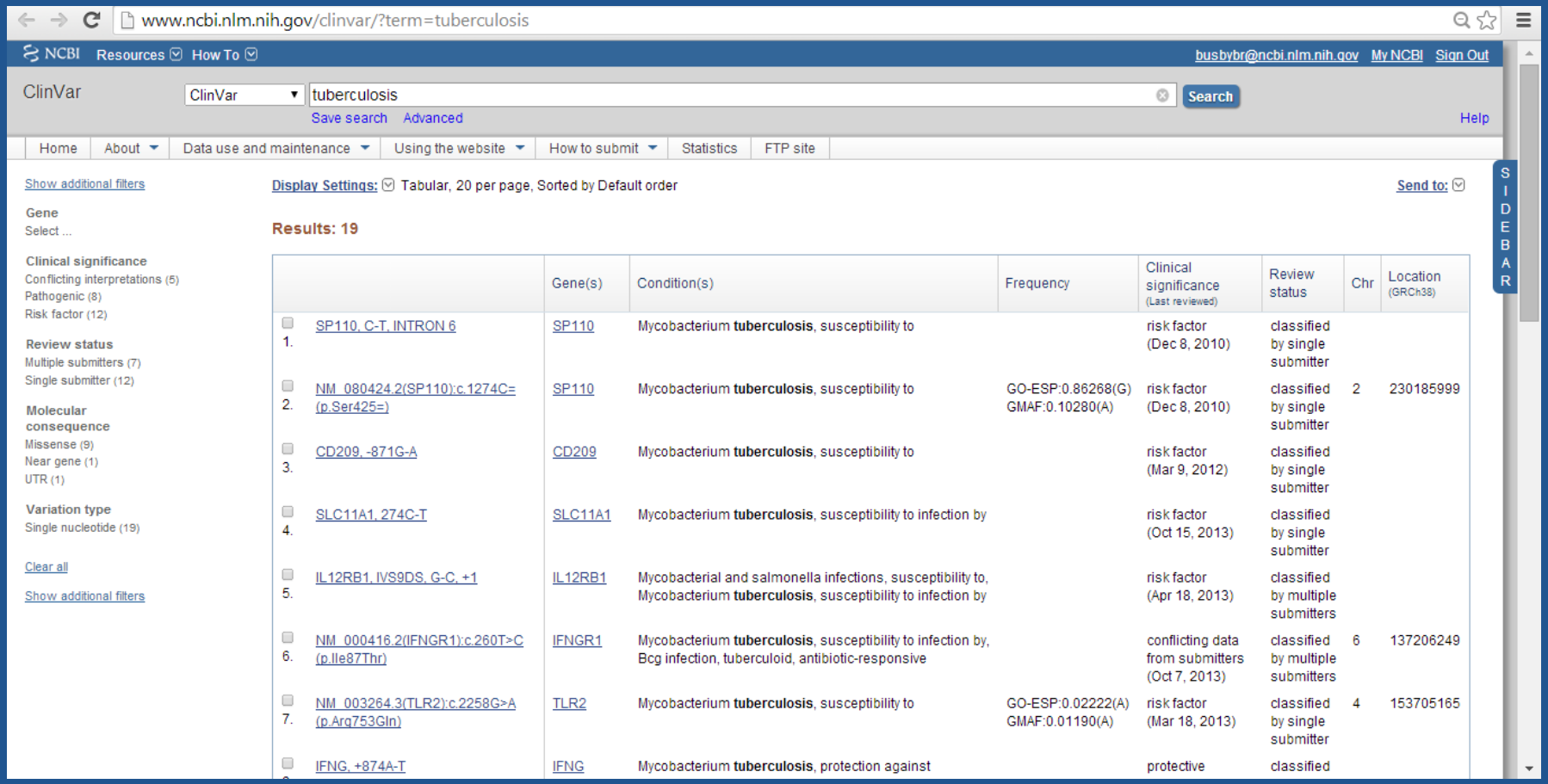
dbGaP – ClinVar



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

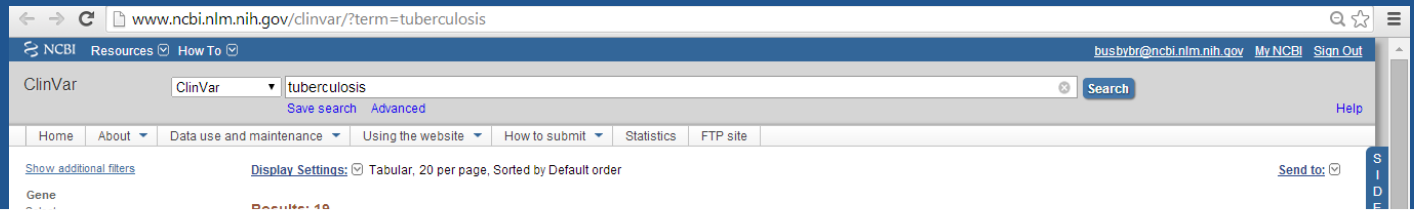
	Called variants	Potential variants
Sample count	2517 of 9686	10630 of 43813

ClinVar



		Gene(s)	Condition(s)	Frequency	Clinical significance (Last reviewed)	Review status	Chr	Location (GRCh38)
1.	SP110, C-T, INTRON 6	SP110	Mycobacterium **tuberculosis**, susceptibility to		risk factor (Dec 8, 2010)	classified by single submitter		
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
3.	CD209, -871G-A	CD209	Mycobacterium **tuberculosis**, susceptibility to		risk factor (Mar 9, 2012)	classified by single submitter		
4.	SLC11A1, 274C-T	SLC11A1	Mycobacterium **tuberculosis**, susceptibility to infection by		risk factor (Oct 15, 2013)	classified by single submitter		
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		
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
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
	IFNG, +874A-T	IFNG	Mycobacterium **tuberculosis**, protection against		protective	classified		

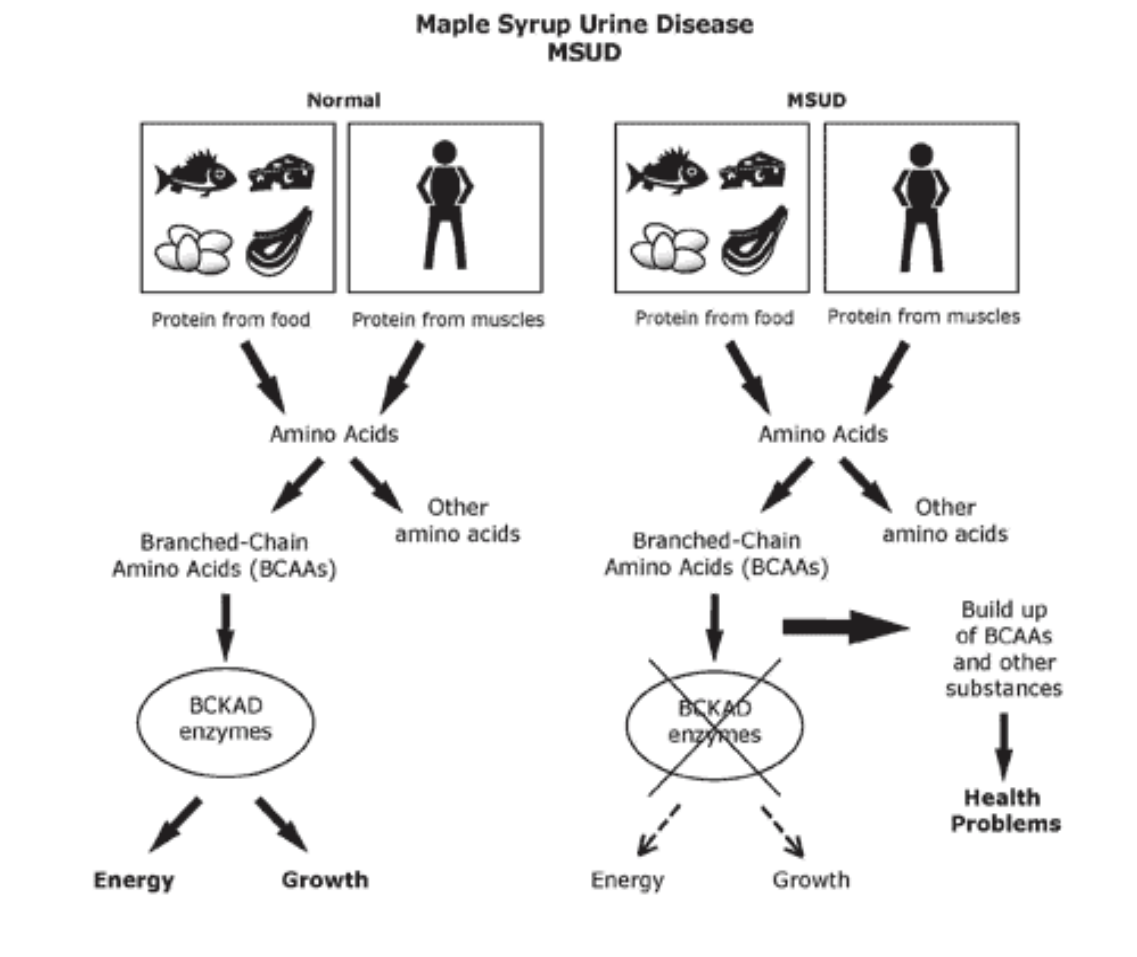
ClinVar



Results: 19

	Gene(s)	Condition(s)	Frequency	Clinical significance (Last reviewed)	Review status	Chr	Location (GRCh38)
1. SP110, C-T, INTRON 6	SP110	Mycobacterium **tuberculosis**, susceptibility to		risk factor (Dec 8, 2010)	classified by single submitter		
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
3. CD209, -871G-A	CD209	Mycobacterium **tuberculosis**, susceptibility to		risk factor (Mar 9, 2012)	classified by single submitter		
4. SLC11A1, 274C-T	SLC11A1	Mycobacterium **tuberculosis**, susceptibility to infection by		risk factor (Oct 15, 2013)	classified by single submitter		
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		

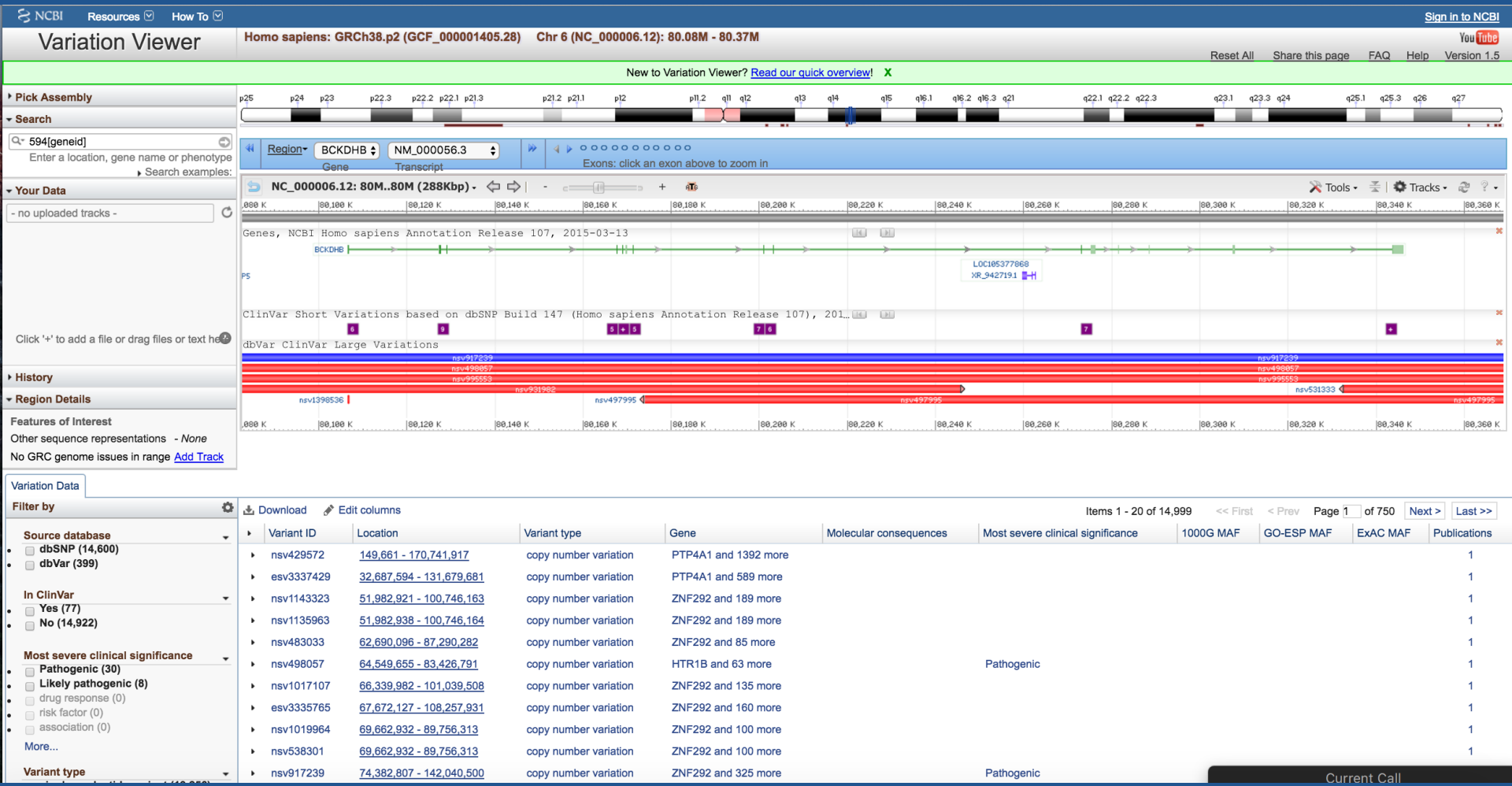
ClinVar – Why Should we Care?



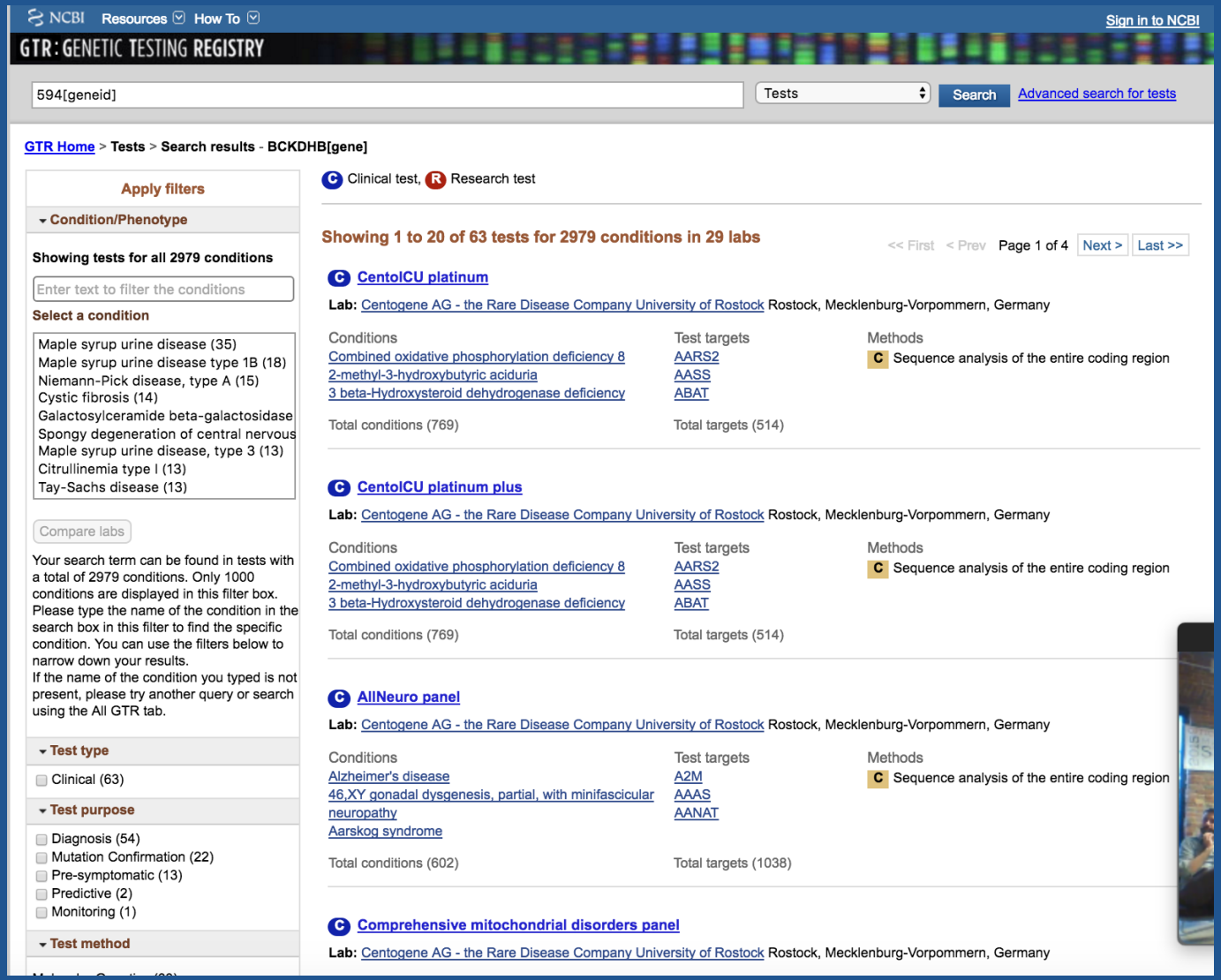
ClinVar – Why Should we Care?

ClinVar – Why Should we Care?

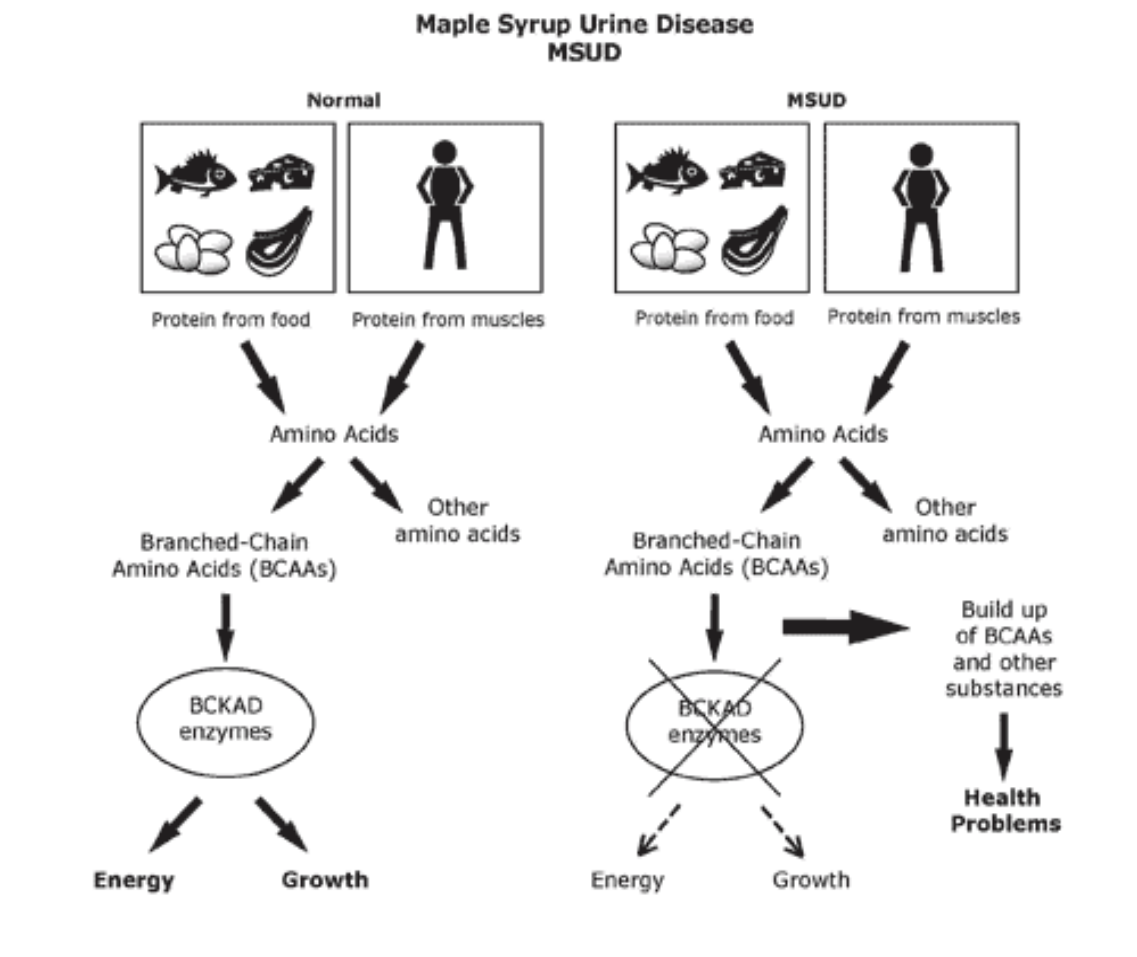
ClinVar – Why Should we Care?



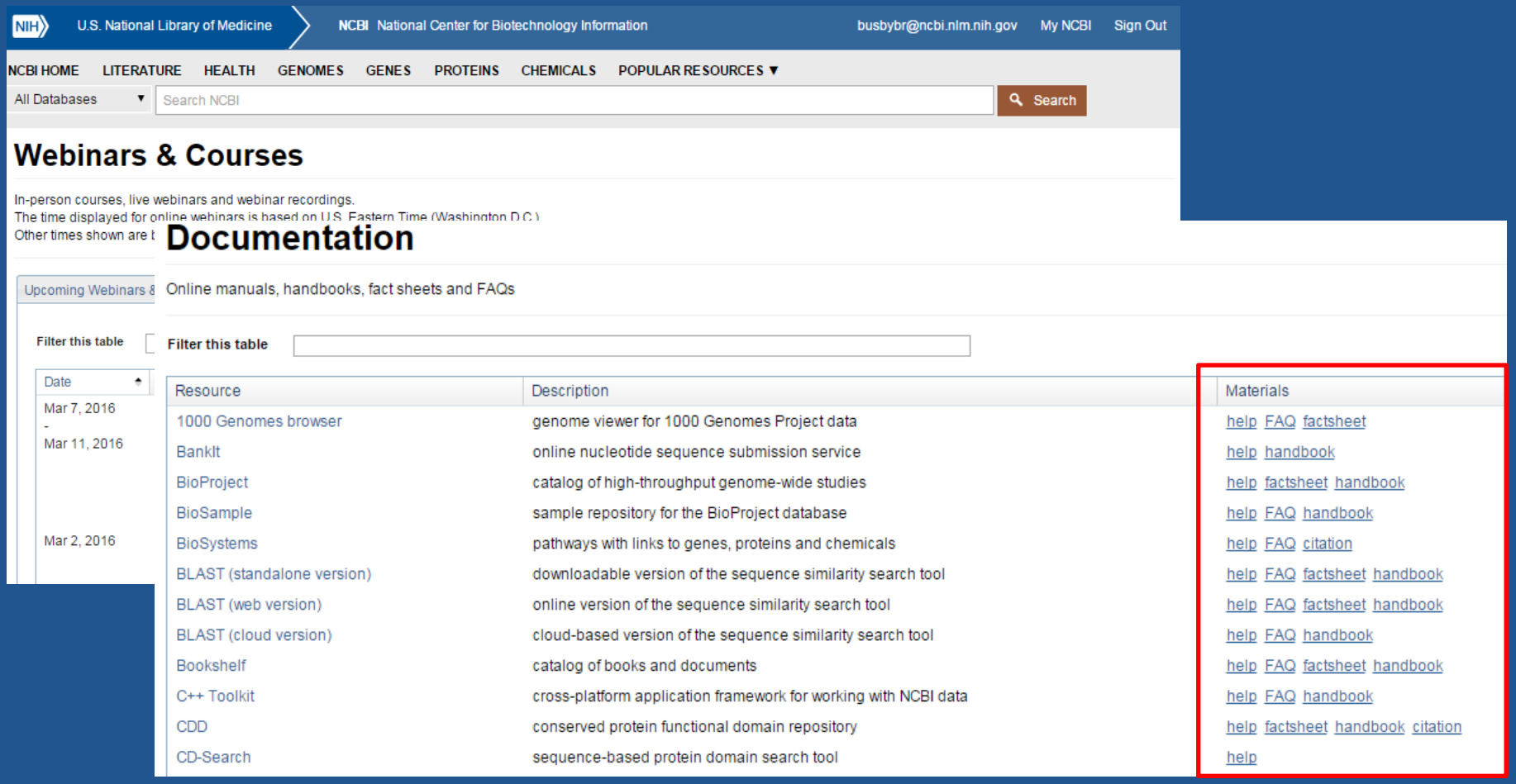
ClinVar – Why Should we Care?



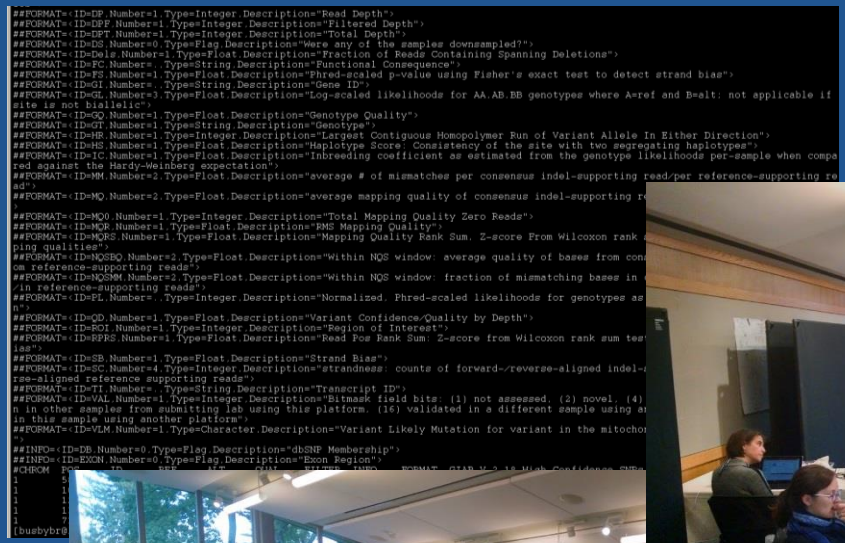
ClinVar – Why Should we Care?



Where to Get More Information!



Data Science Training!

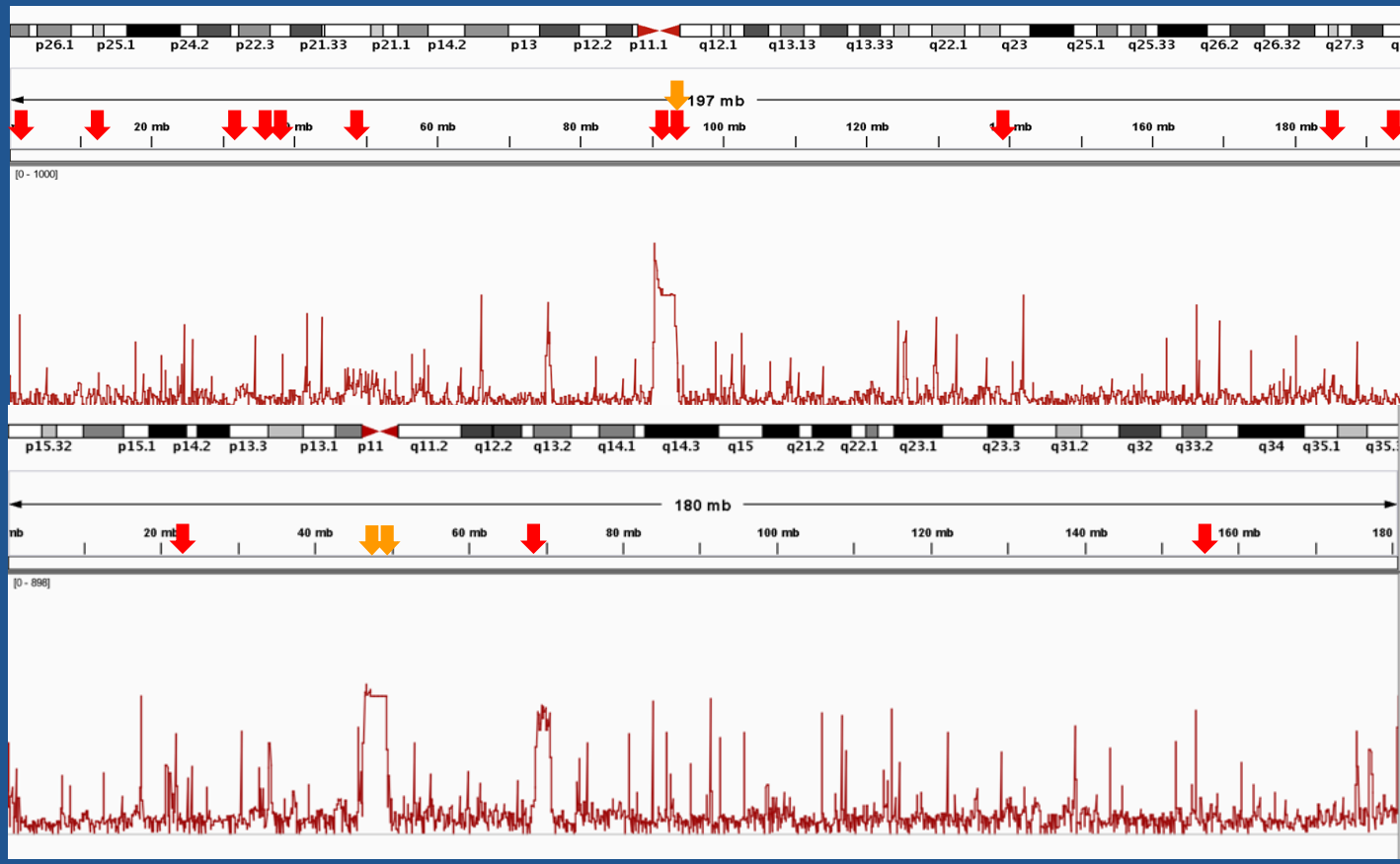


DangerTrack!

NCBI region list

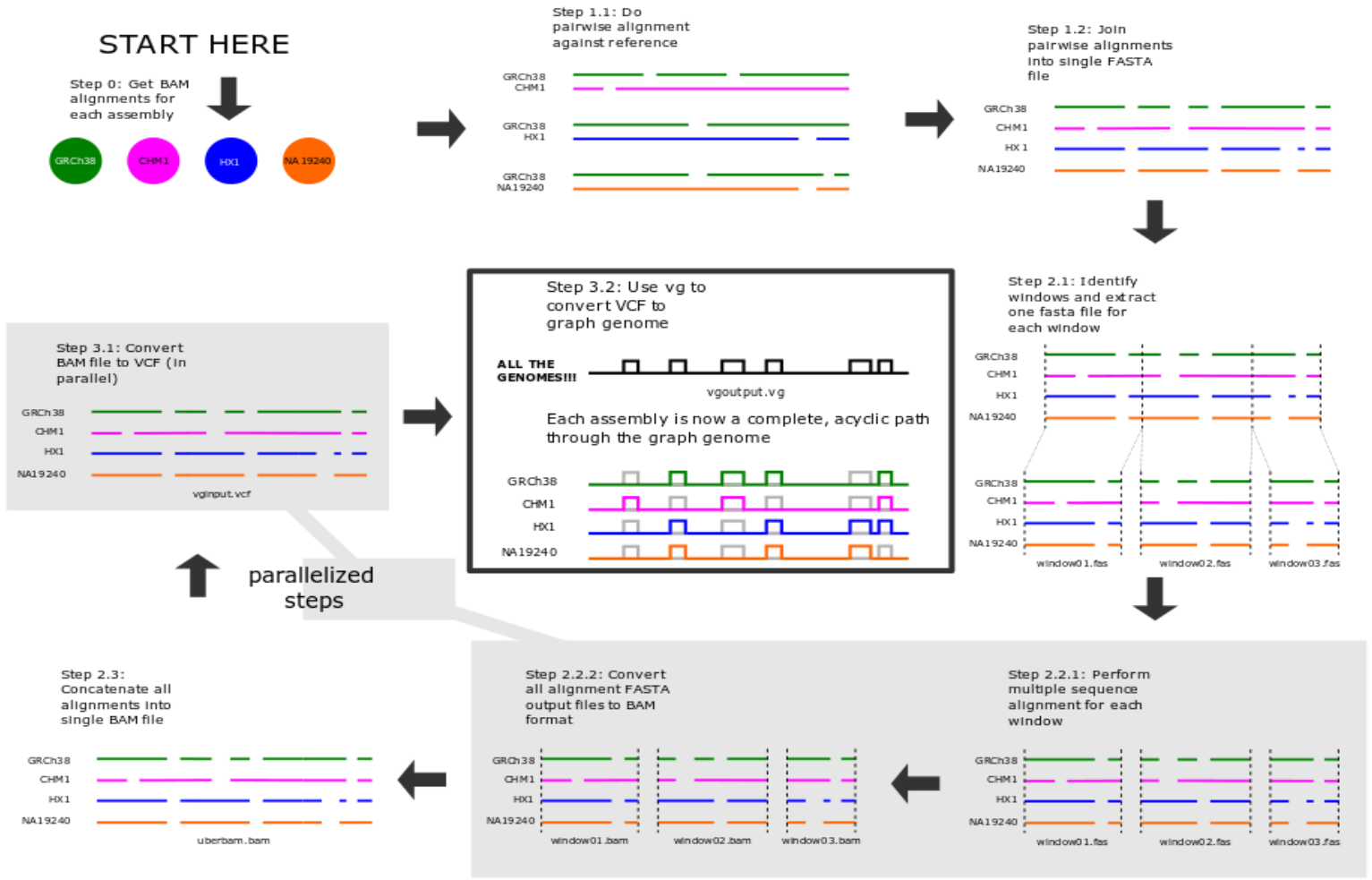
Encode blacklist

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

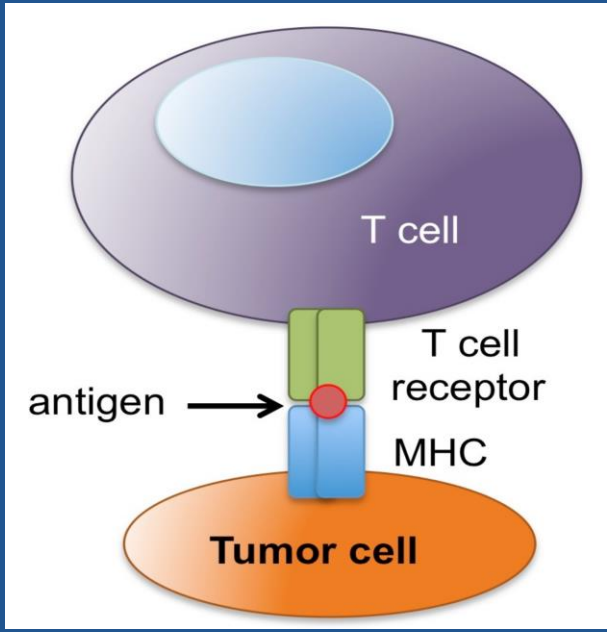


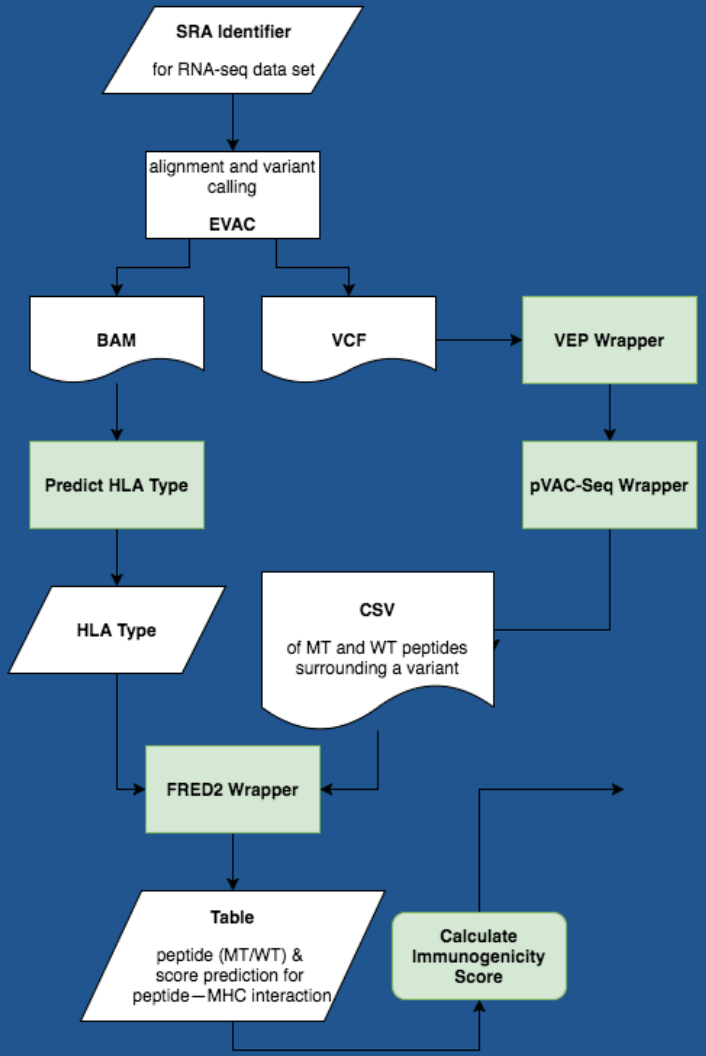
We've run 12 hackathons in the last two years. We continue to build on those projects .

Graph Genomes!!!



Finding immunogenic peptides from single RNA-seq samples

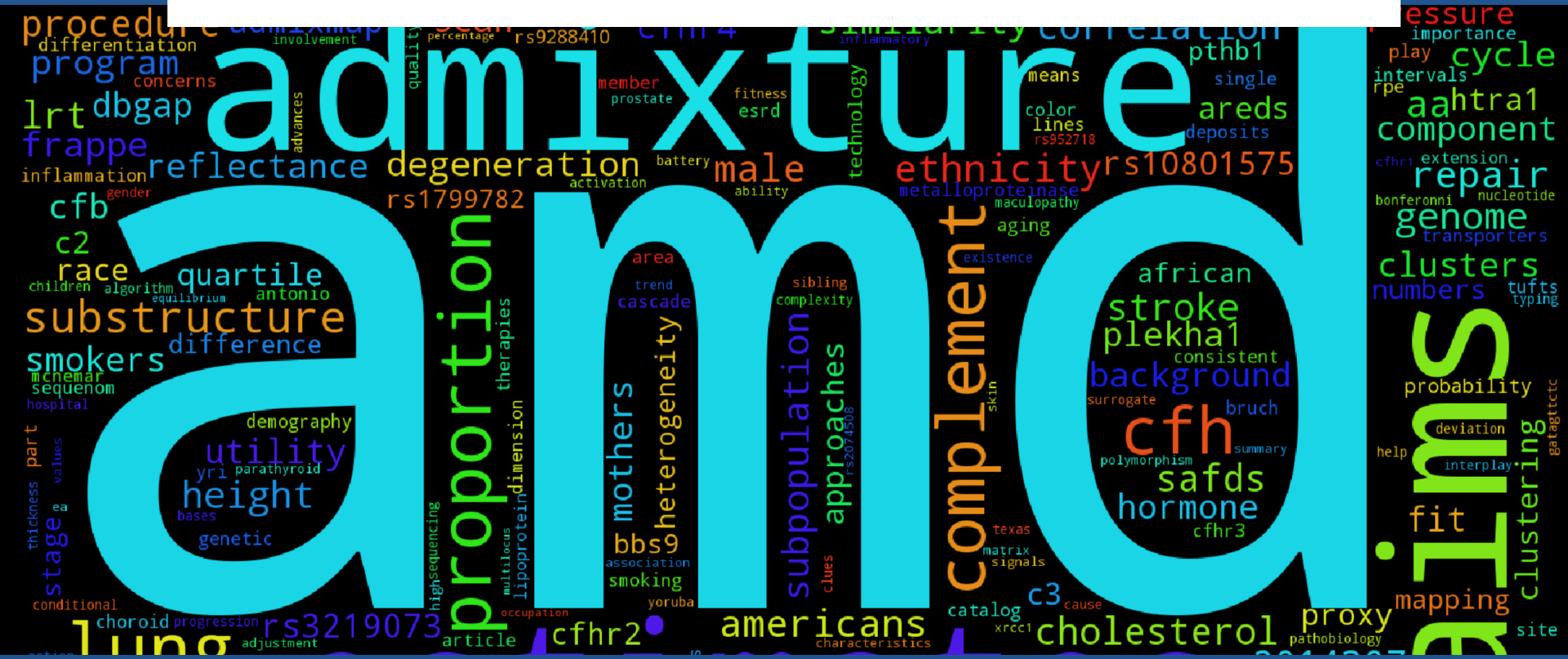




PhenVar

Input rsid numbers as integers seperated by spaces

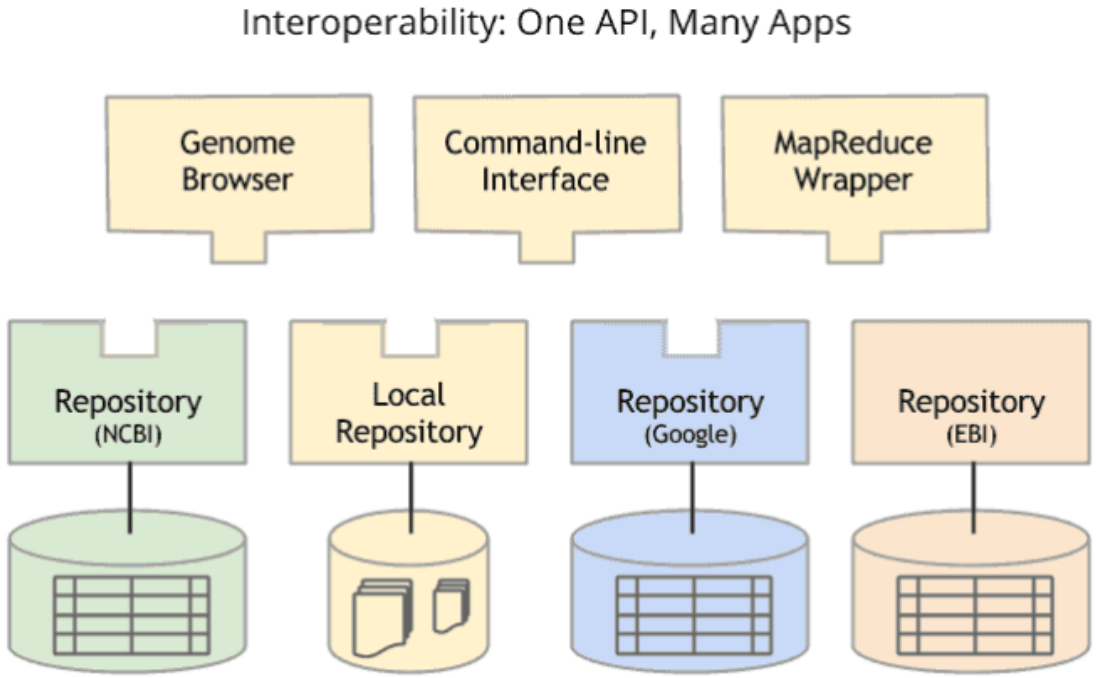
RSID(s): 6003 Search



Collaboration!

Global Alliance for Genomics and Health

Data Working Group

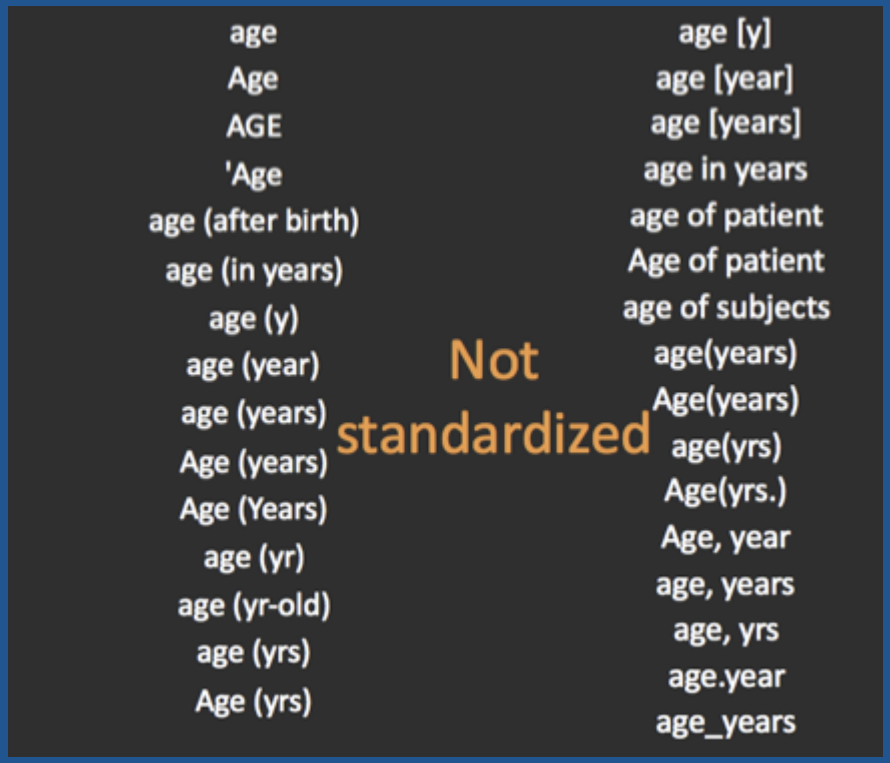


Genomics API

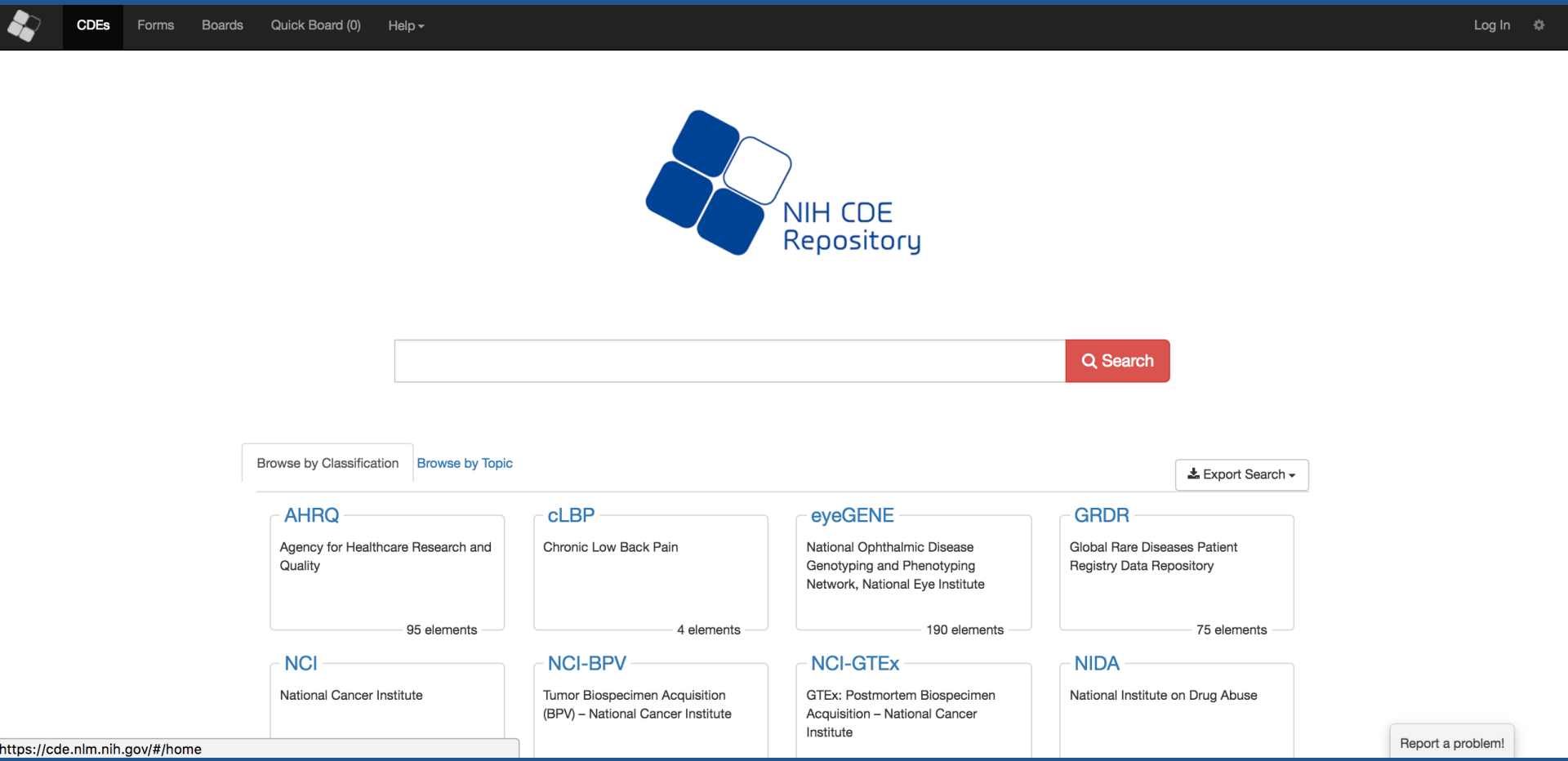
The Global Alliance for Genomics and Health (GA4GH) Genomics API will allow the interoperable exchange of genomic information across multiple organizations and on multiple platforms. This is a freely available open standard for interoperability, that uses common web protocols to support serving

The Future

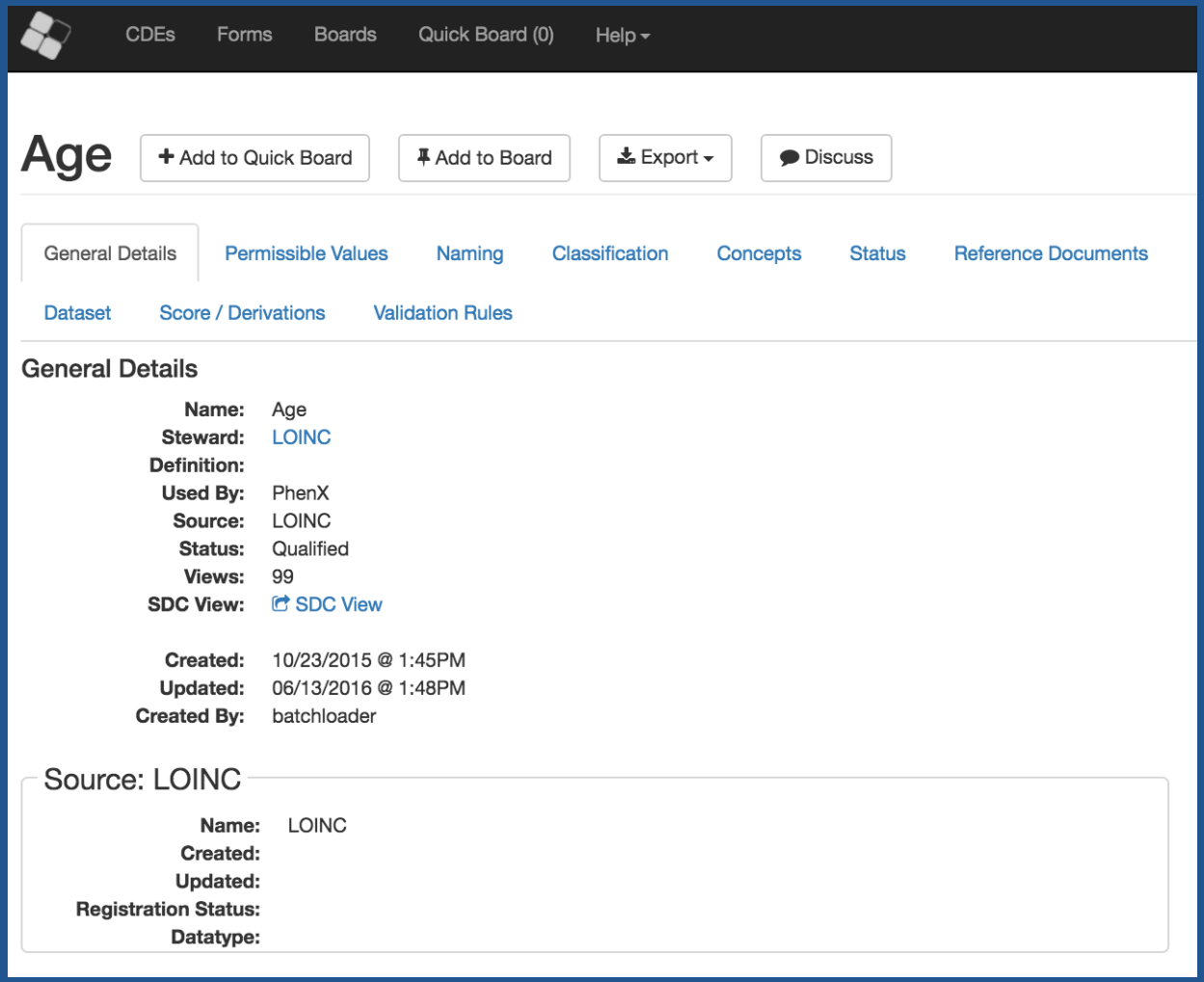
Ontological Standardization



Ontological Standardization



Ontological Standardization



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
provider.

Clicking on one of the links below will search Entrez databases for records with links to that
update their links, therefore links from Entrez records and the available resources are subjec

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 (amphiweb): 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

NCBI Resources How To

BioProject

BioProject | loprovGOLD[SB]

Create alert Advanced

Project Types
Umbrella (2)
Primary submission (47,313)
RefSeq (42)

Data Types
Assembly (4)
Clone ends (11)
Epigenomics (10)
Exome (27)
Genome sequencing (42,413)
Map (2)
Metagenome (336)
Metagenomic assembly (135)
Other (579)
Phenotype/genotype (8)
Random survey (14)
Targeted locus (33)
Transcriptome (629)
Variation (65)

Project Data
Nucleotide (34,925)
Protein (27,218)
Assembly (34,220)
SRA (24,932)
dbSNP (4)
dbVar (3)
GEO DataSets (12)

Scope
Monoisolate (43,657)
Multi-isolate (1,759)
Multi-species (49)
Environmental (570)
Synthetic (9)
Other (504)

Display Settings: Summary, 20 per page, Sorted by Default order

Search results

Items: 1 to 20 of 51251

<< First < Prev Page 1 of 2563 Next >

Bacteroidetes bacterium R-53146

1. Bacteroidetes sp. R-53146 genome sequencing and assembly
Taxonomy: *Bacteroidetes bacterium R-53146*
Project data type: Genome sequencing
Attributes : Scope: Monoisolate; Material: Genome; Capture: Whole; Method Type: Sequencing
Joint Genome Institute
Accession: PRJNA292838 ID: 292838

Endozoicomonas atrinae strain:WP70

2. Endozoicomonas atrinae strain:WP70 Genome sequencing and assembly
Taxonomy: *Endozoicomonas atrinae*
Project data type: Genome sequencing and assembly
Attributes : Scope: Monoisolate; Material: Genome; Capture: Whole; Method Type: Sequencing
Aarhus University
Accession: PRJNA292506 ID: 292506

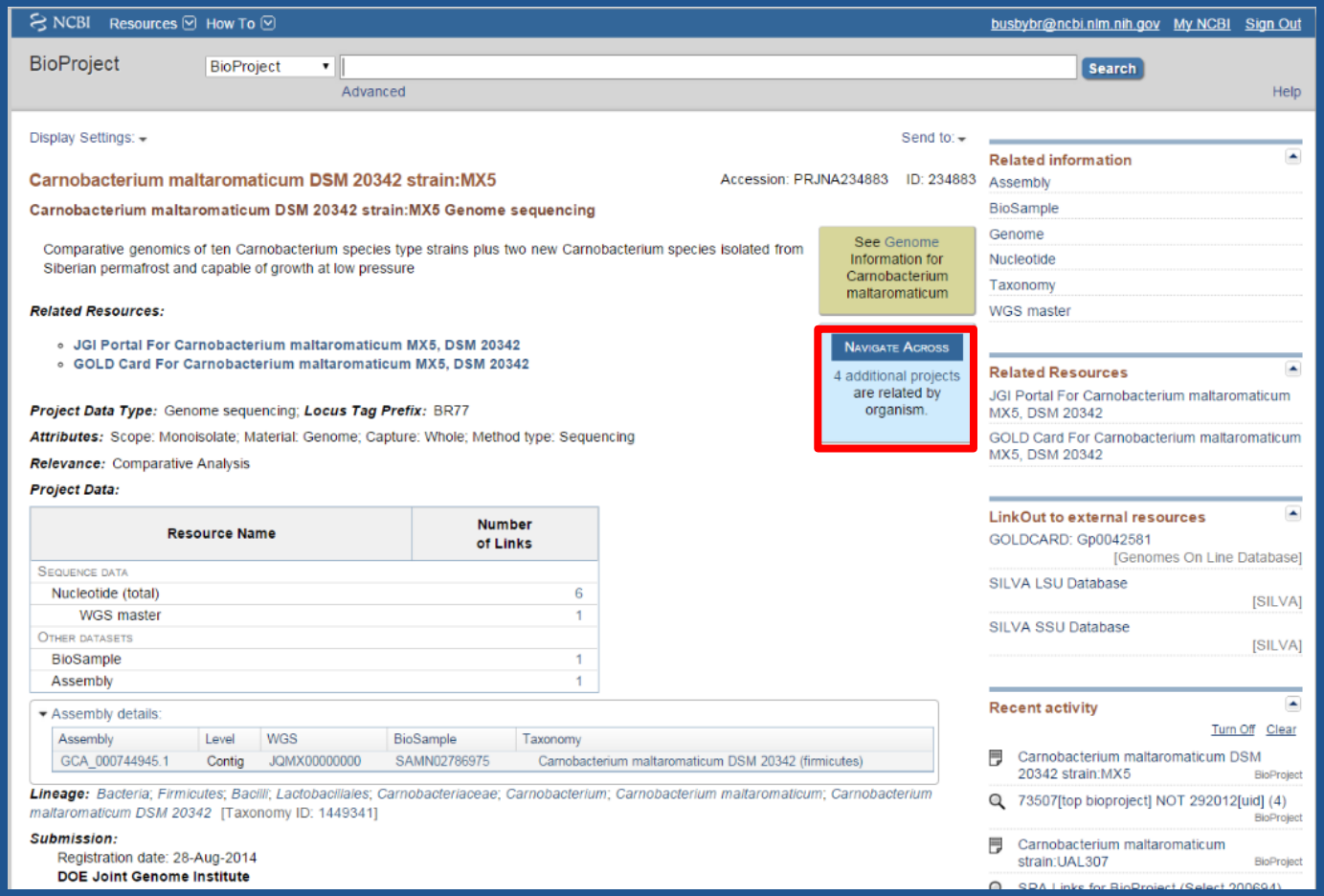
Carnobacterium maltaromaticum strain:UAL307

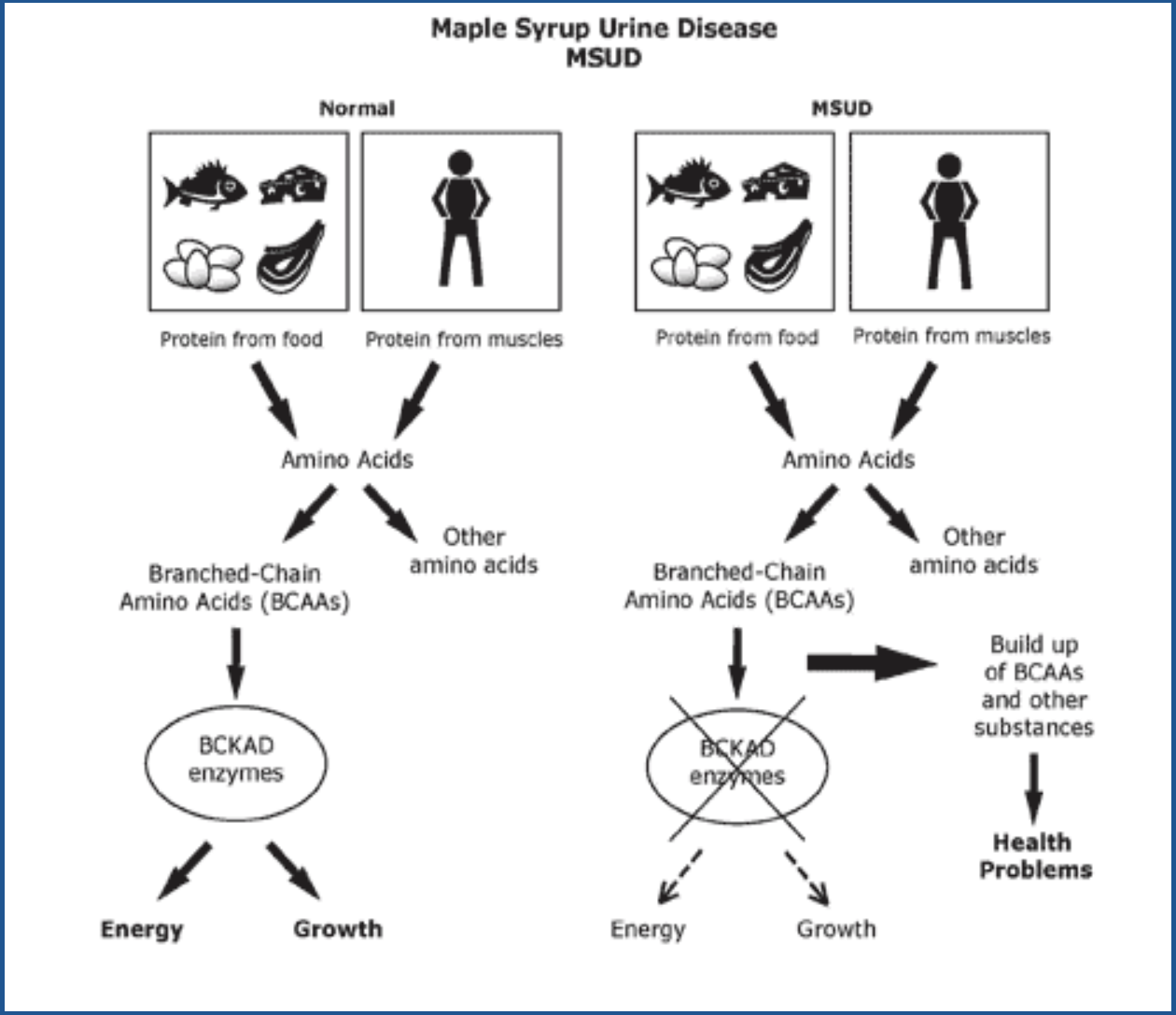
3. Carnobacterium maltaromaticum strain:UAL307 Genome sequencing and assembly
Taxonomy: *Carnobacterium maltaromaticum*
Project data type: Genome sequencing and assembly
Attributes : Scope: Monoisolate; Material: Genome; Capture: Whole; Method Type: Sequencing
University of Alberta
Accession: PRJNA292012 ID: 292012

Endozoicomonas sp. KASP37

4. Endozoicomonas ascidiicola KASP37 genome sequencing and assembly
Taxonomy: *Endozoicomonas sp. KASP37*
Project data type: Genome sequencing and assembly
Attributes : Scope: Monoisolate; Material: Genome; Capture: Whole; Method Type: Sequencing
Aarhus University

Integration into a Larger Data Discovery Framework Example: GOLD (JGI)




Maple Syrup Urine Disease
MSUD
Normal
MSUD
Protein from food
Protein from muscles
Protein from food
Protein from muscles
Amino Acids
Amino Acids
Other amino acids
Other amino acids
Branched-Chain Amino Acids (BCAAs)
Branched-Chain Amino Acids (BCAAs)
BCKAD enzymes
BCKAD enzymes
Build up of BCAAs and other substances
Health Problems
Energy
Growth
Energy
Growth