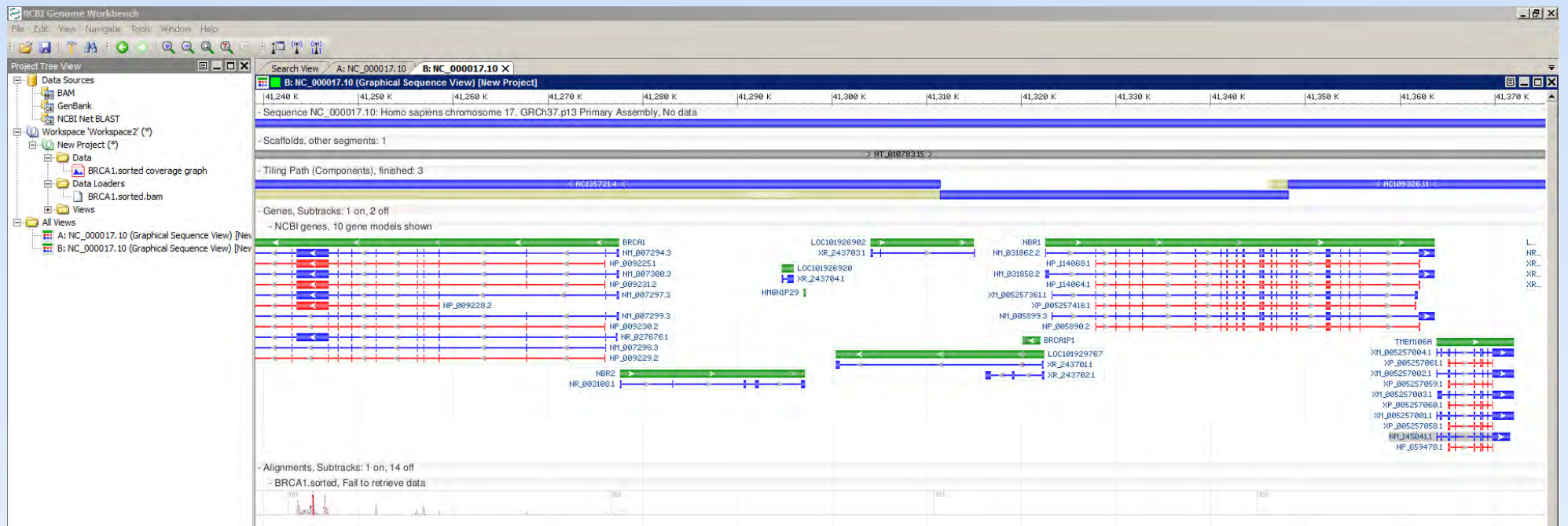


# Genomic Variation in the Rising Era of Individual Genome Sequence: Why we are Enabling Everyone to Share and Use Public Datasets

*August 2016*



# Better Pubmed (and Google) Searches

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

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

PubMed tuberculosis Search

RSS Save search Advanced Help

Show additional filters

Article types  
Clinical Trial  
Review  
More ...

Text availability  
Abstract  
Free full text  
Full text

PubMed Commons  
Reader comments

Publication dates  
5 years  
10 years  
Custom range...

Species  
Humans  
Other Animals

Clear all

Show additional filters

Display Settings: ☒ Summary, 20 per page, Sorted by Recently Added

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

Results: 1 to 20 of 212597 Page 1 of 10630 Next > Last >>

☐ [A Potential Protein Adjuvant Derived from Mycobacterium tuberculosis Rv0652 Enhances Dendritic Cells-Based Tumor Immunotherapy.](#)  
Lee SJ, Shin SJ, Lee MH, Lee MG, Kang TH, Park WS, Soh BY, Park JH, Shin YK, Kim HW, Yun CH, Jung ID, Park YM.  
PLoS One. 2014 Aug 7;9(8):e104351. doi: 10.1371/journal.pone.0104351. eCollection 2014.  
PMID: 25102137 [PubMed - as supplied by publisher]

☐ [The Fitness Landscape of HIV-1 Gag: Advanced Modeling Approaches and Validation of Model Predictions by In Vitro Testing.](#)  
Mann JK, Barton JP, Ferguson AL, Omarjee S, Walker BD, Chakraborty A, Ndung'u T.  
PLoS Comput Biol. 2014 Aug 7;10(8):e1003776. doi: 10.1371/journal.pcbi.1003776. eCollection 2014 Aug.  
PMID: 25102049 [PubMed - as supplied by publisher]

☐ [Utility of Gastric Aspirates for Diagnosing Tuberculosis in Children in a Low Prevalence Area: Predictors of Positive Cultures and Significance of Non-Tuberculous Mycobacteria.](#)  
Kordy F, Richardson SE, Stephens D, Lam R, Jamieson F, Kitai I.  
Pediatr Infect Dis J. 2014 Aug 6. [Epub ahead of print]  
PMID: 25101762 [PubMed - as supplied by publisher]

☐ [Visceral Leishmaniasis and HIV Co-infection in Bihar, India: Long-term Effectiveness and Treatment Outcomes with Liposomal Amphotericin B \(AmBisome\).](#)  
Burza S, Mahajan R, Sinha PK, van Griensven J, Pandey K, Lima MA, Sanz MG, Sunyoto T, Kumar S, Mitra G, Kumar R, Verma N, Das P.  
PLoS Negl Trop Dis. 2014 Aug 7;8(8):e3053. doi: 10.1371/journal.pntd.0003053. eCollection 2014 Aug.  
PMID: 25101665 [PubMed - as supplied by publisher]

New feature  
Try the new Display Settings option - Sort by Relevance

Results by year  
Download CSV

Related searches  
mycobacterium tuberculosis  
pulmonary tuberculosis  
latent tuberculosis  
tuberculosis treatment  
resistant tuberculosis

Titles with your search terms  
Abdominal tuberculosis.  
[Indian J Med Res. 2004]



# For more information go to:

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


NIH 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
			

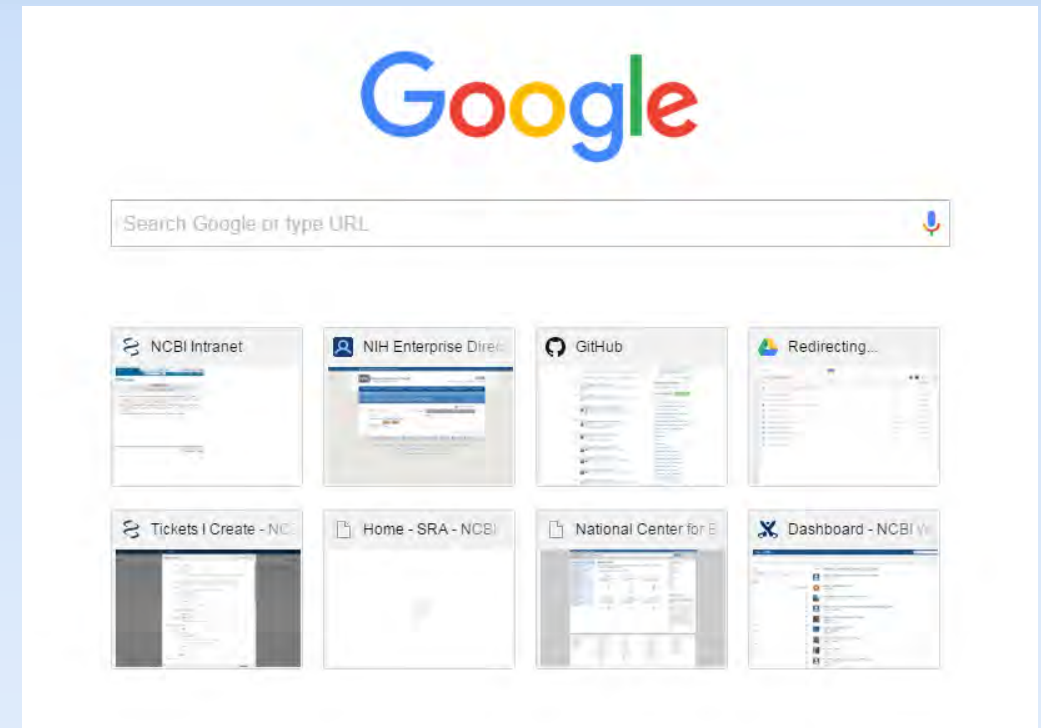
### News, Blog & Social Media

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

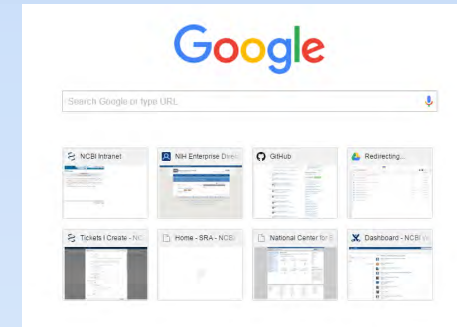


# The Two Things You Need to Know to Handle “Big Data”

```
##FORMAT<(ID=DP,Number=1,Type=Integer,Description="Read Depth")
##FORMAT<(ID=DPF,Number=1,Type=Integer,Description="Filtered Depth")
##FORMAT<(ID=DPT,Number=1,Type=Integer,Description="Total Depth")
##FORMAT<(ID=DS,Number=0,Type=Flag,Description="Were any of the samples downsampled?")
##FORMAT<(ID=Dels,Number=1,Type=Float,Description="Fraction of Reads Containing Spanning Deletions")
##FORMAT<(ID=FC,Number=1,Type=String,Description="Functional Consequence")
##FORMAT<(ID=FS,Number=1,Type=Float,Description="Phred-scaled p-value using Fisher's exact test to detect strand bias")
##FORMAT<(ID=GI,Number=1,Type=String,Description="Gene ID")
##FORMAT<(ID=GL,Number=3,Type=Float,Description="Log-scaled likelihoods for AA,AB,BB genotypes where A=ref and B=alt; not applicable if site is not biallelic")
##FORMAT<(ID=GQ,Number=1,Type=Float,Description="Genotype Quality")
##FORMAT<(ID=GT,Number=1,Type=String,Description="Genotype")
##FORMAT<(ID=H,Number=1,Type=Integer,Description="Largest Contiguous Homopolymer Run of Variant Allele In Either Direction")
##FORMAT<(ID=HS,Number=1,Type=Float,Description="Haplotype Score: Consistency of the site with two segregating haplotypes")
##FORMAT<(ID=IC,Number=1,Type=Float,Description="Inbreeding coefficient as estimated from the genotype likelihoods per-sample when compared against the Hardy-Weinberg expectation")
##FORMAT<(ID=MM,Number=2,Type=Float,Description="average # of mismatches per consensus indel-supporting read/per reference-supporting read")
##FORMAT<(ID=MQ,Number=2,Type=Float,Description="average mapping quality of consensus indel-supporting reads/reference-supporting reads")
##FORMAT<(ID=MQ0,Number=1,Type=Integer,Description="Total Mapping Quality Zero Reads")
##FORMAT<(ID=MQR,Number=1,Type=Float,Description="RMS Mapping Quality")
##FORMAT<(ID=MQRS,Number=1,Type=Float,Description="Mapping Quality Rank Sum: Z-score From Wilcoxon rank sum test of Alt vs. Ref read mapping qualities")
##FORMAT<(ID=NQSEQ,Number=2,Type=Float,Description="Within NQS window: average quality of bases from consensus indel-supporting reads/reference-supporting reads")
##FORMAT<(ID=NQSM,Number=2,Type=Float,Description="Within NQS window: fraction of mismatching bases in consensus indel-supporting reads/in reference-supporting reads")
##FORMAT<(ID=PL,Number=1,Type=Integer,Description="Normalized, Phred-scaled likelihoods for genotypes as defined in the VCF specification")
##FORMAT<(ID=QD,Number=1,Type=Float,Description="Variant Confidence/Quality by Depth")
##FORMAT<(ID=ROI,Number=1,Type=Integer,Description="Region of Interest")
##FORMAT<(ID=RPES,Number=1,Type=Float,Description="Read Pos Rank Sum: Z-score from Wilcoxon rank sum test of Alt vs. Ref read position bias")
##FORMAT<(ID=SB,Number=1,Type=Float,Description="Strand Bias")
##FORMAT<(ID=SC,Number=4,Type=Integer,Description="strandness: counts of forward-/reverse-aligned indel-supporting reads / forward-/reverse-aligned reference supporting reads")
##FORMAT<(ID=TI,Number=1,Type=String,Description="Transcript ID")
##FORMAT<(ID=VAL,Number=1,Type=Integer,Description="Bitmask field bits: (1) not assessed, (2) novel, (4) seen in public data set, (8) seen in other samples from submitting lab using this platform, (16) validated in a different sample using another platform, (32) validated in this sample using another platform")
##FORMAT<(ID=VLM,Number=1,Type=Character,Description="Variant Likely Mutation for variant in the mitochondrial DNA : (T) True, (F) False")
##INFO<(ID=DB,Number=0,Type=Flag,Description="dbSNP Membership")
##INFO<(ID=EXON,Number=0,Type=Flag,Description="Exon Region")
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT GIAB_V_2_10_High_Confidence_SNPs_Indels
1 58211 A G 3498 PASS GT:DP:GQ:PL:HR:VAL 1/1:150:250:3517:269:19:1:1
1 102769 C A 6319 PASS GT:DP:GQ:PL:HR:VAL 1/1:265:584:6319:584:0:0:1
1 126113 A A 3955 PASS GT:DP:GQ:PL:HR:VAL 1/1:487:1144:3955:1144:0:2:1
1 136048 C T 12719 PASS GT:DP:GQ:PL:HR:VAL 1/1:798:1380:12719:1380:0:0:1
1 714427 G A 17107 PASS GT:DP:GQ:PL:HR:VAL 1/1:663:1507:17107:1507:0:1:32
[buzbybr@lmem06 Giab_demo]$
```



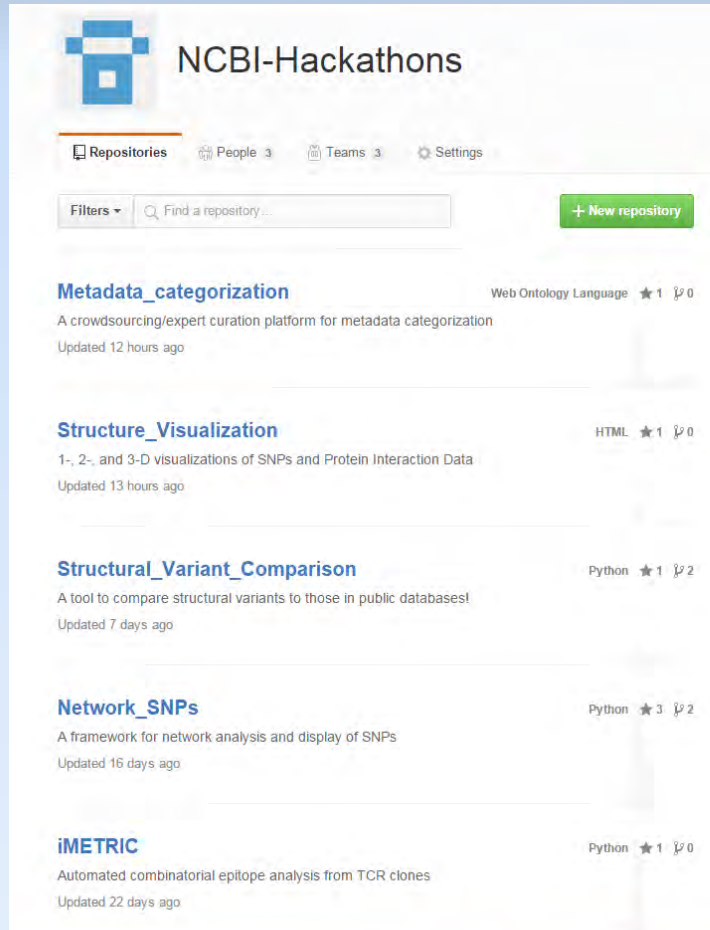
# The Two Things You Need to Know to Handle “Big Data”

[illegible]

- Software Carpentry
  - Stack Overflow
- Biostars and SEQanswers

# Oh yeah, and this...

github.com/NCBI-hackathons



The screenshot shows the GitHub repository page for NCBI-Hackathons. The repository has a blue header with the NCBI logo and the name "NCBI-Hackathons". Below the header, there are tabs for "Repositories", "People", "Teams", and "Settings". A search bar is present with the text "Find a repository...". A green button labeled "+ New repository" is on the right. The main content area lists several repositories with their names, languages, star counts, and fork counts. The repositories listed are: Metadata\_categorization (Web Ontology Language, 1 star, 0 forks), Structure\_Visualization (HTML, 1 star, 0 forks), Structural\_Variant\_Comparison (Python, 1 star, 2 forks), Network\_SNPs (Python, 3 stars, 2 forks), iMETRIC (Python, 1 star, 0 forks), Pharmacogenomics\_Prediction\_Pipeline\_P3 (R, 7 stars, 5 forks), Community\_Software\_Tools\_for\_NGS (no language specified, 9 stars, 5 forks), SRA2R (HTML, 3 stars, 1 fork), HASSL\_Homogeneous\_Analysis\_of\_SRA\_rnaSequencing\_Libraries (Python, 4 stars, 2 forks), Resources\_for\_Designing\_Hackathons (no language specified, 0 stars, 0 forks), Educational\_Resources (no language specified, 0 stars, 0 forks), RNA\_mapping (Python, 1 star, 6 forks), seqr-tokenizer (Java, 1 star, 0 forks), seqr (Java, 6 stars, 2 forks), DASH\_cell\_type (R, 0 stars, 2 forks), NCBI\_August\_Hackathon\_Push\_Button\_Genomics\_Solution (Python, 4 stars, 3 forks), movetest (no language specified, 0 stars, 0 forks), RNA-seq\_Comparison\_Pipeline (Python, 0 stars, 0 forks), and hackathon\_v001\_metagenomics (Shell, 8 stars, 2 forks).

**NCBI-Hackathons**

Repositories People Teams Settings

Filters Find a repository... + New repository

**Metadata\_categorization** Web Ontology Language ★ 1 🍴 0  
A crowdsourcing/expert curation platform for metadata categorization  
Updated 12 hours ago

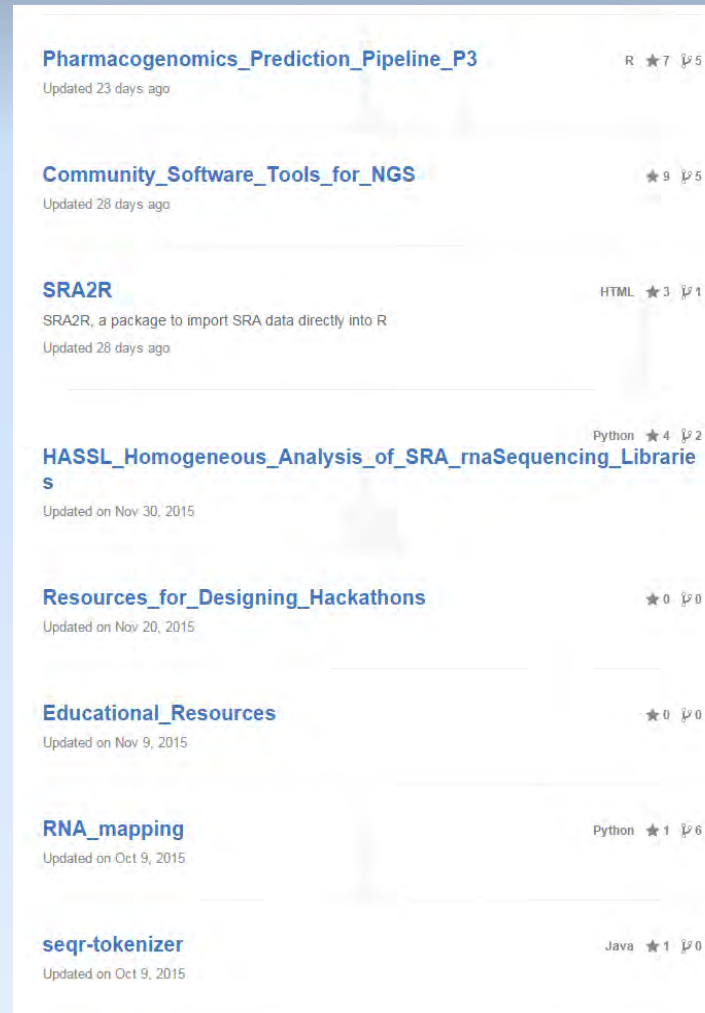
**Structure\_Visualization** HTML ★ 1 🍴 0  
1-, 2-, and 3-D visualizations of SNPs and Protein Interaction Data  
Updated 13 hours ago

**Structural\_Variant\_Comparison** Python ★ 1 🍴 2  
A tool to compare structural variants to those in public databases!  
Updated 7 days ago

**Network\_SNPs** Python ★ 3 🍴 2  
A framework for network analysis and display of SNPs  
Updated 16 days ago

**iMETRIC** Python ★ 1 🍴 0  
Automated combinatorial epitope analysis from TCR clones  
Updated 22 days ago

github.com/NCBI-hackathons



The screenshot shows a list of repositories from the NCBI-Hackathons GitHub page. The repositories listed are: Pharmacogenomics\_Prediction\_Pipeline\_P3 (R, 7 stars, 5 forks), Community\_Software\_Tools\_for\_NGS (no language specified, 9 stars, 5 forks), SRA2R (HTML, 3 stars, 1 fork), HASSL\_Homogeneous\_Analysis\_of\_SRA\_rnaSequencing\_Libraries (Python, 4 stars, 2 forks), Resources\_for\_Designing\_Hackathons (no language specified, 0 stars, 0 forks), Educational\_Resources (no language specified, 0 stars, 0 forks), RNA\_mapping (Python, 1 star, 6 forks), and seqr-tokenizer (Java, 1 star, 0 forks).

**Pharmacogenomics\_Prediction\_Pipeline\_P3** R ★ 7 🍴 5  
Updated 23 days ago

**Community\_Software\_Tools\_for\_NGS** ★ 9 🍴 5  
Updated 28 days ago

**SRA2R** HTML ★ 3 🍴 1  
SRA2R, a package to import SRA data directly into R  
Updated 28 days ago

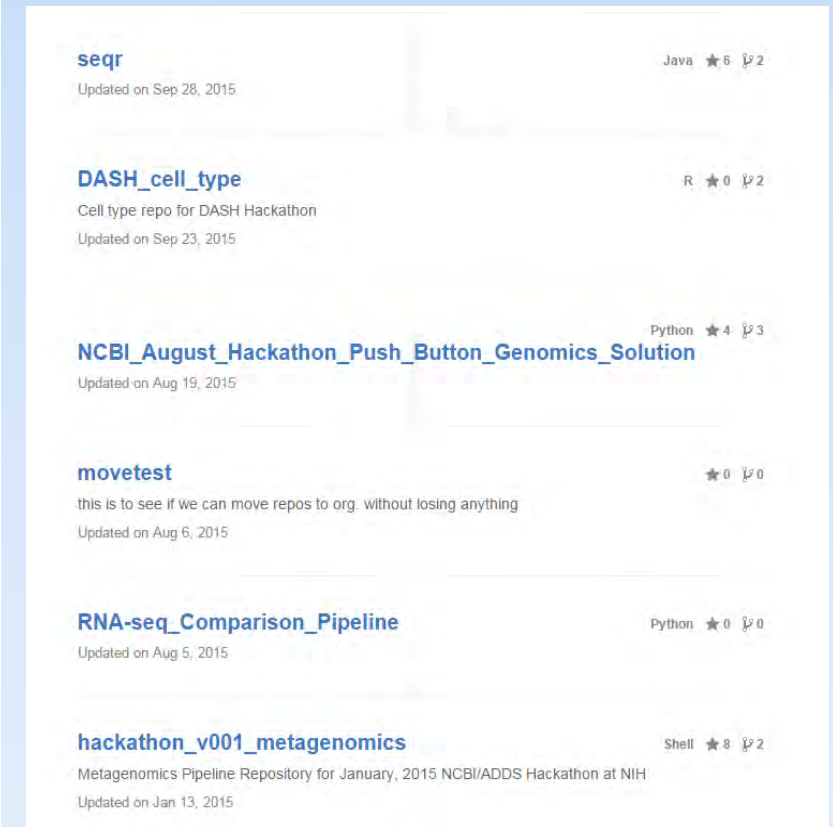
**HASSL\_Homogeneous\_Analysis\_of\_SRA\_rnaSequencing\_Libraries** Python ★ 4 🍴 2  
Updated on Nov 30, 2015

**Resources\_for\_Designing\_Hackathons** ★ 0 🍴 0  
Updated on Nov 20, 2015

**Educational\_Resources** ★ 0 🍴 0  
Updated on Nov 9, 2015

**RNA\_mapping** Python ★ 1 🍴 6  
Updated on Oct 9, 2015

**seqr-tokenizer** Java ★ 1 🍴 0  
Updated on Oct 9, 2015



The screenshot shows a list of repositories from the NCBI-Hackathons GitHub page. The repositories listed are: seqr (Java, 6 stars, 2 forks), DASH\_cell\_type (R, 0 stars, 2 forks), NCBI\_August\_Hackathon\_Push\_Button\_Genomics\_Solution (Python, 4 stars, 3 forks), movetest (no language specified, 0 stars, 0 forks), RNA-seq\_Comparison\_Pipeline (Python, 0 stars, 0 forks), and hackathon\_v001\_metagenomics (Shell, 8 stars, 2 forks).

**seqr** Java ★ 6 🍴 2  
Updated on Sep 28, 2015

**DASH\_cell\_type** R ★ 0 🍴 2  
Cell type repo for DASH Hackathon  
Updated on Sep 23, 2015

**NCBI\_August\_Hackathon\_Push\_Button\_Genomics\_Solution** Python ★ 4 🍴 3  
Updated on Aug 19, 2015

**movetest** ★ 0 🍴 0  
this is to see if we can move repos to org. without losing anything  
Updated on Aug 6, 2015

**RNA-seq\_Comparison\_Pipeline** Python ★ 0 🍴 0  
Updated on Aug 5, 2015

**hackathon\_v001\_metagenomics** Shell ★ 8 🍴 2  
Metagenomics Pipeline Repository for January, 2015 NCBI/ADDS Hackathon at NIH  
Updated on Jan 13, 2015



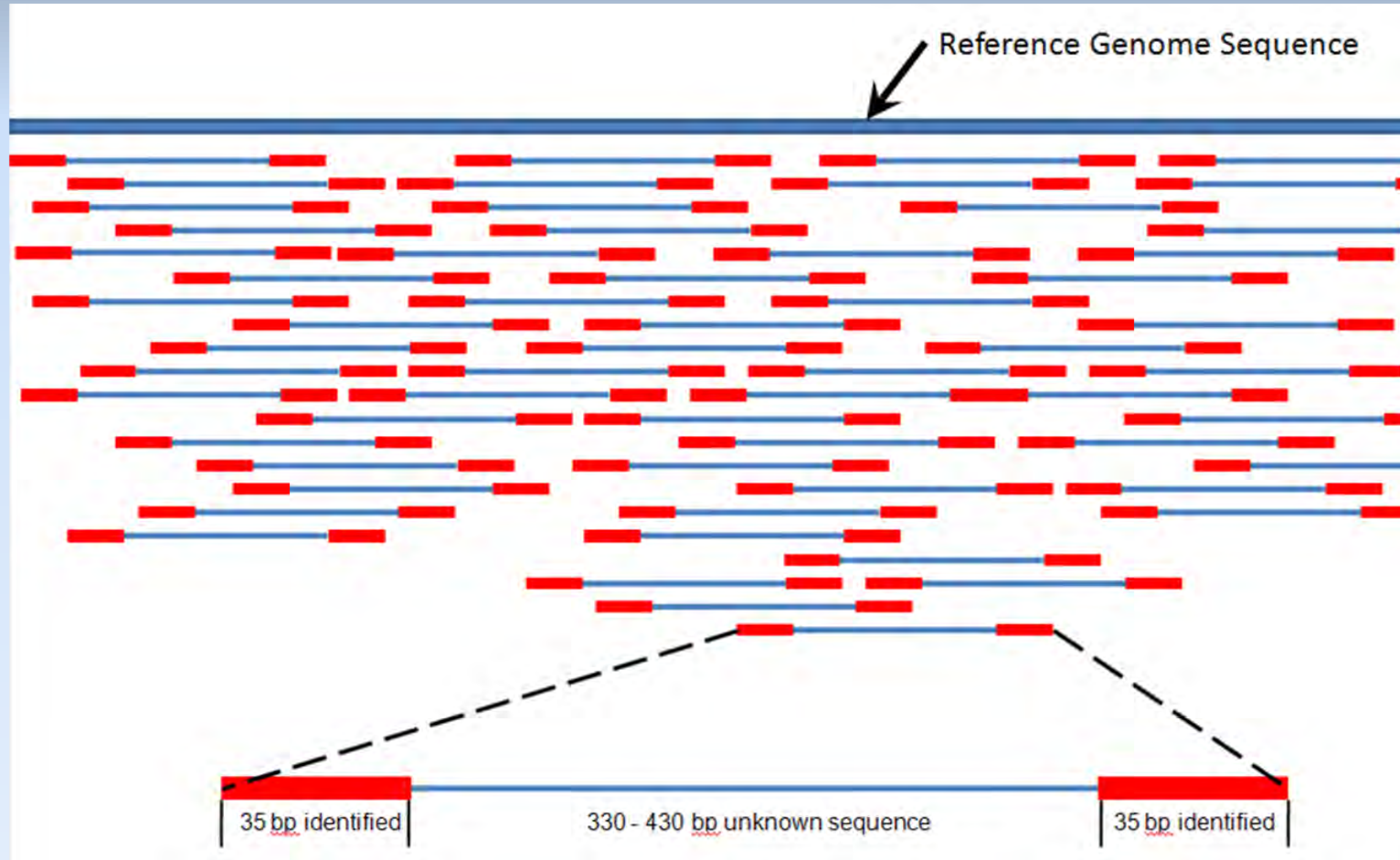
# The Five Things You Need to Know About “Big Data”

*examples from genomics*

- Check data quality
  - Current largest source of noise is library prep
- Setting parameters (or not)
- Remember multiple etiologies
  - “If it looks like a duck”
- Multiple testing correction
  - Not Bonferroni
- Validation can come in many forms (as long as its independent)

# Review of terminology and concepts

## Next Generation Sequencing

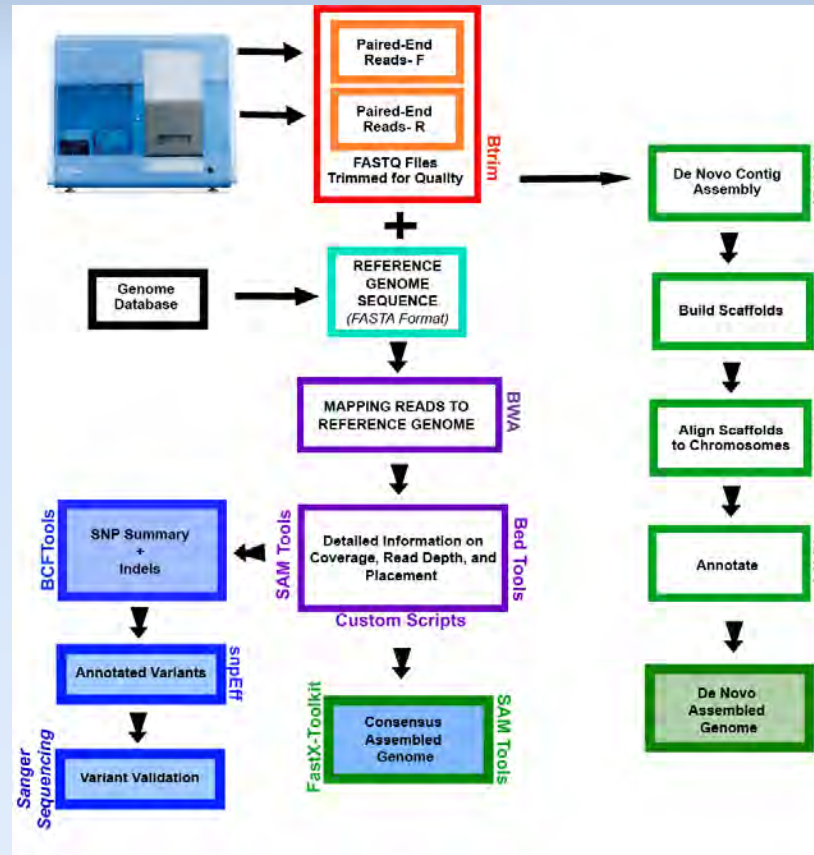


Graphic Credit:  
Spencer Martin, UBC



# 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>

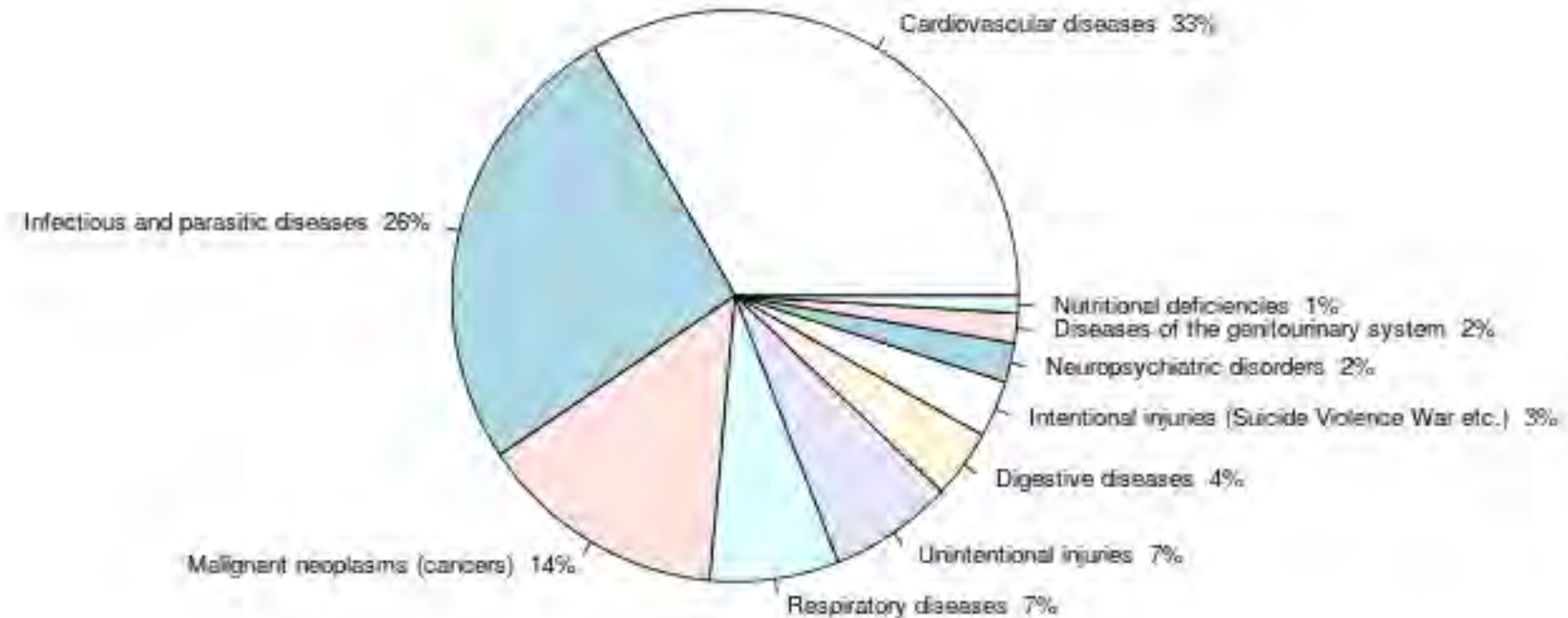
**how:** Data can be viewed at the gene level using variation viewer, or downloaded by gene by searching the dbVar database. A complete set of CNVs for many organisms can be downloaded from:  
<http://ftp.ncbi.nlm.nih.gov/pub/dbVar/data/>

downstream analysis

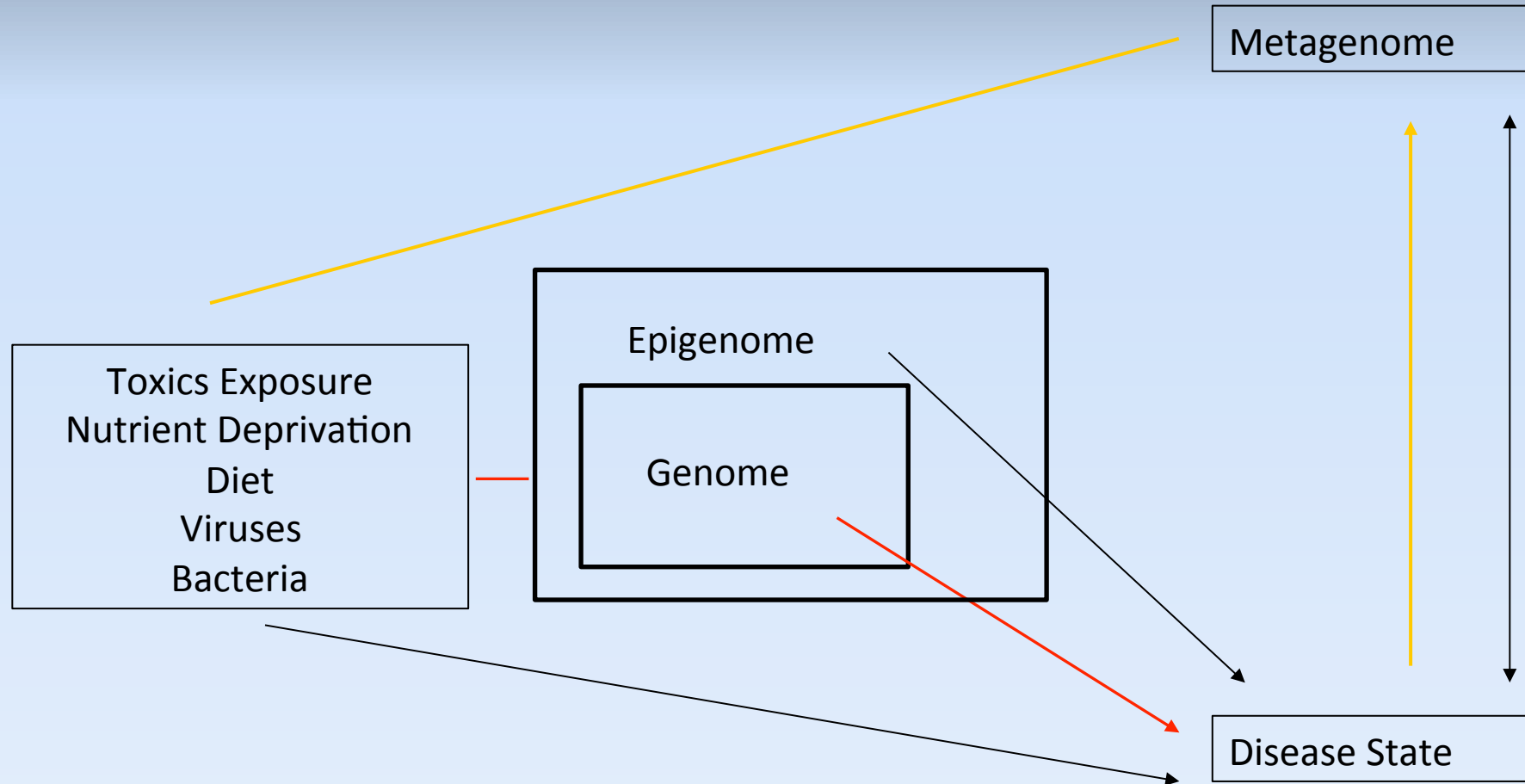
**NCBI NGS Online Workshop – Available on the  
NCBI YouTube Channel!**

# Some Supplemental Material for Those New to Bioinformatics and Genomics

*The opinions expressed here are those of the author.  
They do not necessarily  
reflect the views of HHS or any other federal agency.*

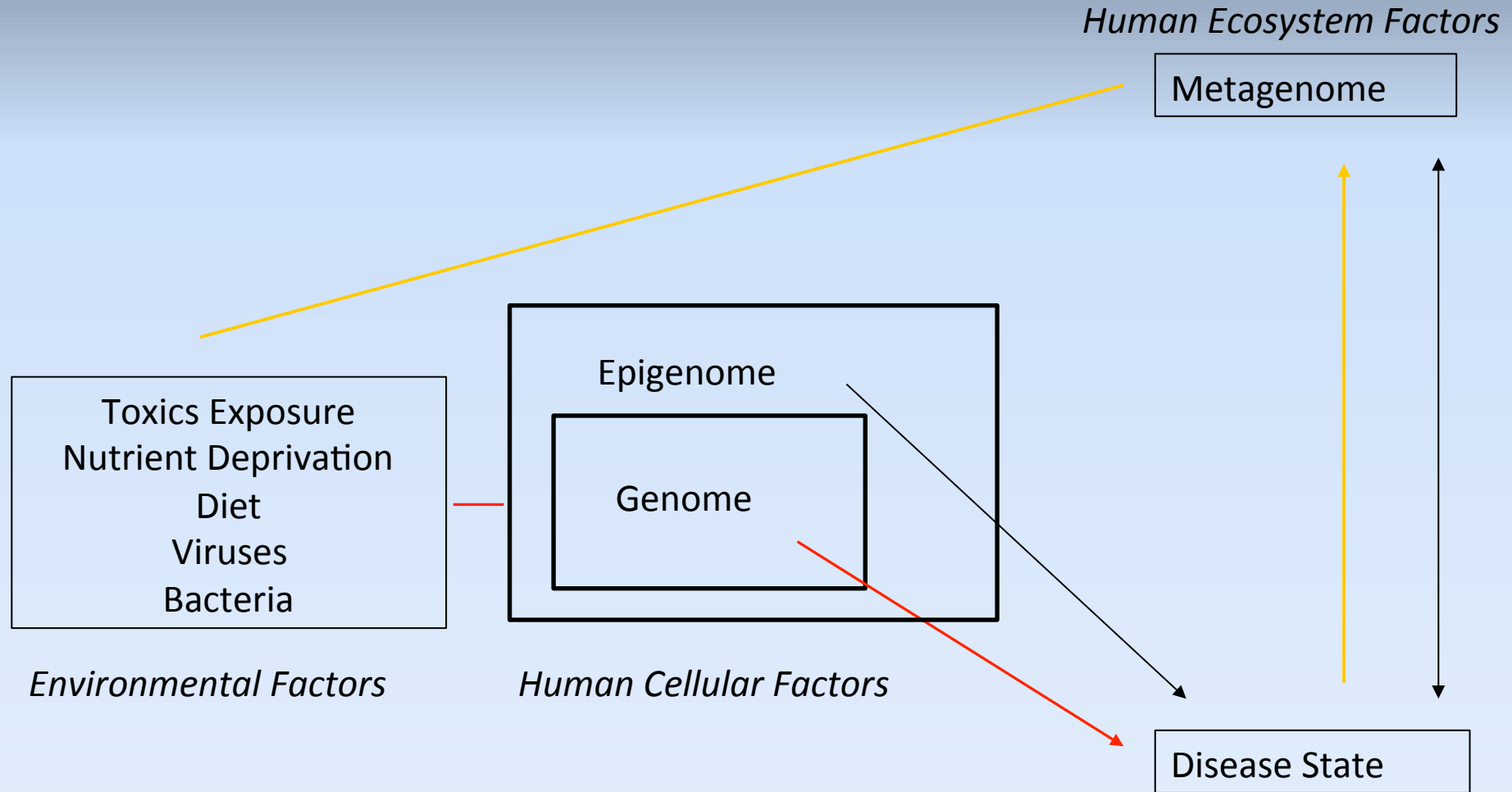


# Biological Information Flows and Disease States

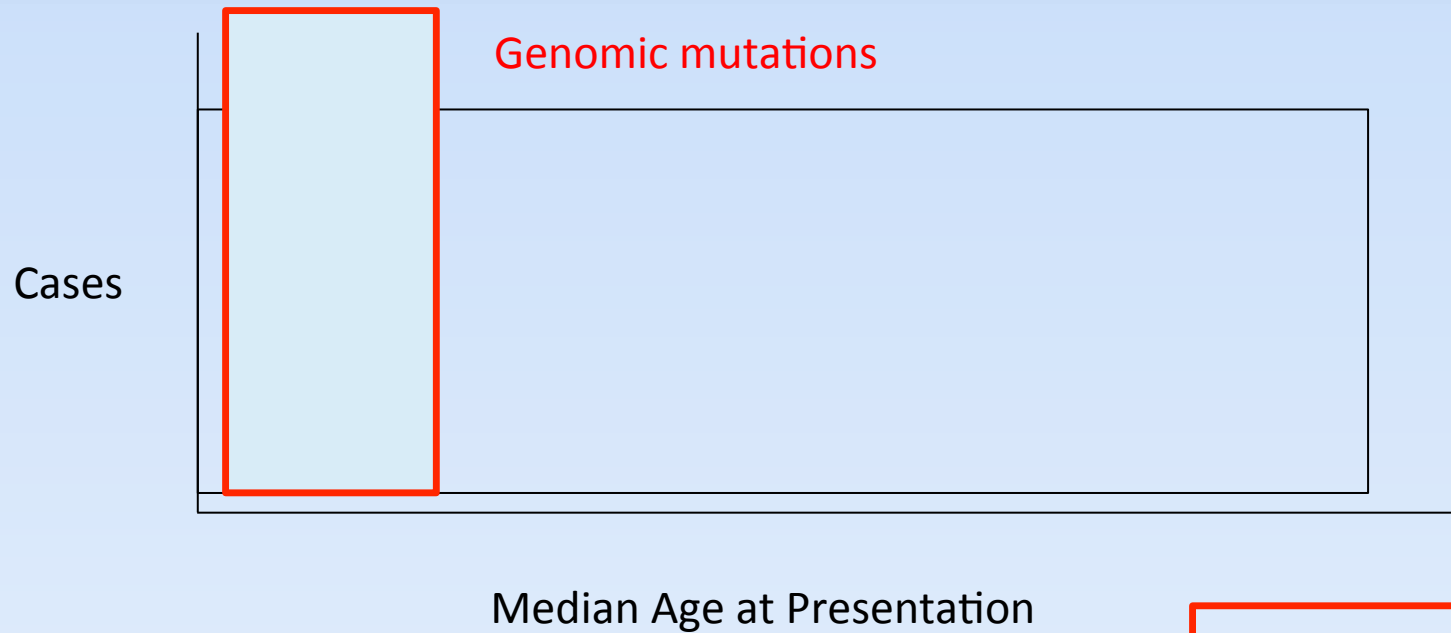




# Defining Terms



# Rethinking Disease

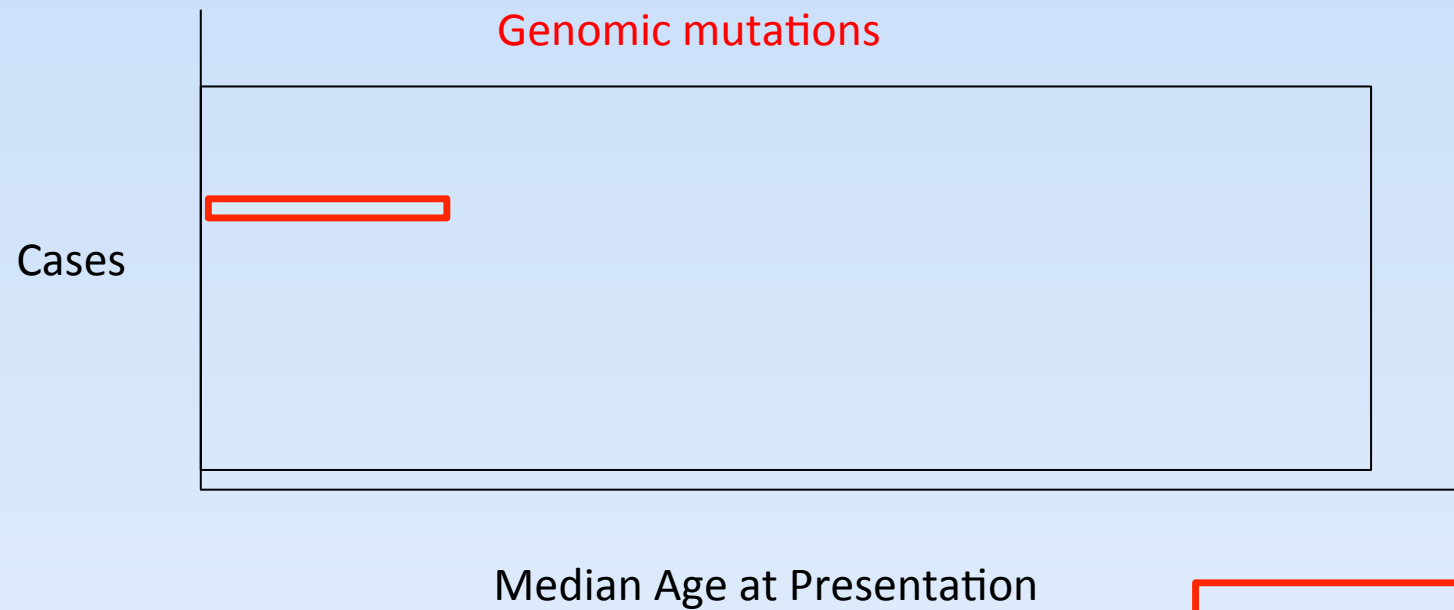


*“Etiology == Genomic Cause”*

*A completely hypothetical example to  
illustrate a point*

~~1 disease  
1 genetic  
cause~~

# Rethinking Disease

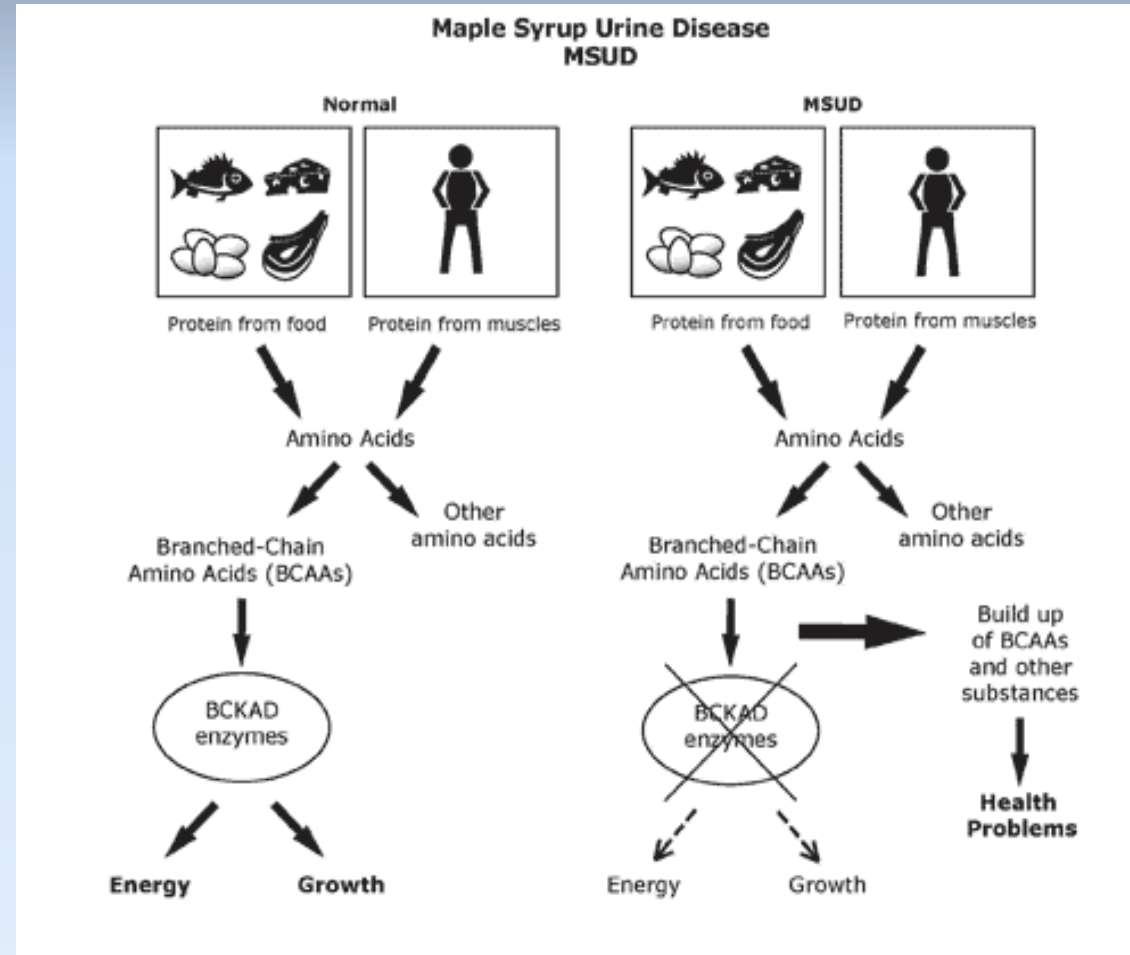


**“Etiology == Genomic Cause”**

*A completely hypothetical example to  
illustrate a point*

**~~1 disease  
1 genetic  
cause~~**

# Rare but Treatable Variants!





# Gene

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

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

Gene   [Advanced](#) [Search](#) [Help](#)

[Display Settings](#)  Full Report [Send to](#)  [Hide sidebar >>](#)

**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; Haplorrhini; 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)
- BioAssay by Target (Summary)



# ClinVar

NCBI

Resources

How To

busbybr@ncbi.nlm.nih.gov

My NCBI

Sign Out

ClinVar

ClinVar

Search ClinVar for gene symbols, HGVS expressions, conditions, and more

Search

Advanced

Help

Home

About

Data use and maintenance

Using the website



How to submit

Statistics

FTP site

**NM\_000056.3(BCKDHB):c.356T>G (p.Val119Gly)**

**NM\_000056.3(BCKDHB):c.356T>G (p.Val119Gly)**

Go to:  

Variant type: single nucleotide variant

Cytogenetic location: 6q14

Genomic location: Chr6:80167690 (on Assembly GRCh38)  
Chr6:80877407 (on Assembly GRCh37)

Protein change: V69G, V119G

HGVS: NG\_009775.1:g.66064T>G  
NM\_000056.3:c.356T>G  
NM\_183050.2:c.356T>G  
[...more](#)

Links: OMIM: 248611.0004  
dbSNP: 121965005





NCBI 1000 Genomes Browser: rs121965005

Molecular consequence: NM\_000056.3:c.356T>G: missense variant [Sequence  
Ontology [SO:0001583](#)]

**Clinical significance**

NM\_000056.3(BCKDHB):c.356T>G (p.Val119Gly) [Help](#)

Clinical significance: Pathogenic/Likely pathogenic

Review status:     (1/4)


Number of submission(s): 1


Condition(s)  
MAPLE SYRUP URINE DISEASE, CLASSIC, TYPE IB

See supporting ClinVar records

**1 Affected Gene**

branched chain keto acid dehydrogenase E1, beta polypeptide (BCKDHB)  
[Gene - OMIM - Variation viewer]

 Search ClinVar for variants within BCKDHB

 Search ClinVar for variants including BCKDHB

**Browser views**

RefSeqGene

Variation viewer [GRCh38 - GRCh37]

UCSC [GRCh38/hg38 - GRCh37/hg19]

**Related information**



dbSNP

Gene

PubMed

**Recent Activity**

Assertion and evidence details

Go to:  

Clinical Assertions

Evidence

Germine

[Help](#)

Clinical significance (Last evaluated)	Review status (Assertion method)	Collection method	Condition(s) (Mode of inheritance)	Origin	Citations	Submitter (Last submitted)	Submission accession
Pathogenic (Jun 28, 2007)	classified by single submitter (literature only)	literature only	MAPLE SYRUP URINE DISEASE, CLASSIC, TYPE IB	germline	<a href="#">PubMed</a> (1)	<a href="#">OMIM</a> (Dec 30, 2010)	SCV000032952

Last Updated: Oct 30, 2014





# Variation Viewer

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

YouTube

Pick Assembly

Search

Q: BCKDHB

Enter a location, gene name or phenotype

Genes Other features

Name	Location
BCKDHB	Chr6 80.11M - 80.35M
HDAC6	ChrX 48.80M - 48.82M
POLB	Chr8 42.34M - 42.37M
CUL3	Chr2 224.5M - 224.6M
BCKDHA	Chr19 41.40M - 41.43M
DBT	Chr1 100.2M - 100.2M
PPM1K	Chr4 88.26M - 88.28M

Your Data

History

Region Details

Features of Interest

Other sequence representations - None

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

Region: BCKDHB NM\_000056.3

Exons: click an exon above to zoom in

NC\_000006.12: 80M..80M (288Kbp)

Tools Tracks

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

IP5

LOC105377868 XR\_942719.1

ClinVar Short Variations based on dbSNP Build 146 (Homo sapiens Annotation Release 107), 201...

dbVar ClinVar Large Variations

nsv917239 nsv498057 nsv995553 nsv868928

nsv931882 nsv497995 nsv531333 nsv497995

dbSNP Build 146 (Homo sapiens Annotation Release 107) all data

Variation Data

Filter by

Download Edit columns

Items 1 - 30 of 14,684

Source database	Variant ID	Location	Variant type	Gene	Molecular consequences	Most severe clinical significance	1000G MAF	GO-ESP MAF	ExAC MAF	Publications
<input type="checkbox"/> dbSNP (14,286)	nsv429572	149,661 - 170,741,917	copy number variation	PTP4A1 and 1392 more						1
<input type="checkbox"/> dbVar (398)	esv3337429	32,687,594 - 131,679,681	copy number variation	PTP4A1 and 589 more						1
In ClinVar	nsv1143323	51,982,921 - 100,746,163	copy number variation	ZNF292 and 189 more						1
<input type="checkbox"/> Yes (76)	nsv1135963	51,982,938 - 100,746,164	copy number variation	ZNF292 and 189 more						1
<input type="checkbox"/> No (14,608)	nsv483033	62,690,096 - 87,290,282	copy number variation	ZNF292 and 85 more						1
Most severe clinical significance	nsv498057	64,549,655 - 83,426,791	copy number variation	IBTK and 63 more		Pathogenic				1
<input type="checkbox"/> Pathogenic (28)	nsv1017107	66,339,982 - 101,039,508	copy number variation	ZNF292 and 135 more						1
<input type="checkbox"/> Likely pathogenic (9)	esv3335765	67,672,127 - 108,257,931	copy number variation	ZNF292 and 180 more						1
	nsv1019964	69,662,932 - 89,756,313	copy number variation	ZNF292 and 100 more						1

# Clinical Applications

**GTR: GENETIC TESTING REGISTRY**

Conditions/Phenotypes

[GTR Home](#) > Conditions/Phenotypes > Maple syrup urine disease type 1B

**Maple syrup urine disease type 1B**

**Synonyms:** MAPLE SYRUP URINE DISEASE, TYPE 1B  
MSUD due to deficiency of  $\alpha$ 1-beta subunit of branched-chain  $\alpha$ -keto acid dehydrogenase complex, MSUD type 3 (formerly),  
MSUD type 1B

**Available tests** 25 tests are in the database for this condition. [Compare labs offering these tests.](#)  
Check [Associated genes](#) and [Related conditions](#) for additional relevant tests.

**Clinical tests** [\(25 available\)](#)

**Molecular Genetics Tests**

[Sequence analysis of the entire coding region \(17\)](#)  
[Deletion/duplication analysis \(7\)](#)  
[Targeted variant analysis \(5\)](#)

**Associated genes**

[BCKDHB](#) [see tests for this gene](#)

**Also known as:** RP1-279A18.1, E1B, dJ279A18.1, BCKDHB  
**Summary:** branched chain keto acid dehydrogenase E1, beta polypeptide

**Related conditions**

☐ Clinical test ☐ Research test ☐ OMIM ☐ GeneReviews

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<a href="#">Disorder of amino acid metabolism</a>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<a href="#">Argininosuccinate lyase deficiency</a>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<a href="#">Citrullinemia type I</a>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<a href="#">Homocystinuria</a>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<a href="#">Maple syrup urine disease</a>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<a href="#">Maple syrup urine disease type 2</a>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<a href="#">Maple syrup urine disease type 1A</a>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<b>Maple syrup urine disease type 1B</b>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<a href="#">Phenylketonuria</a>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<a href="#">Dihydropyridine reductase deficiency</a>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<a href="#">Maternal phenylketonuria</a>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<a href="#">Tyrosinemia type I</a>

**Clinical features**

[Imported from Human Phenotype Ontology \(HPO\)](#)

Hypertonia  
Intellectual disability  
Lactic acidosis  
Seizures  
Pancreatitis  
[Show all \(17\)](#)

[Go to complete MedGen record for Maple syrup urine disease type 1B](#)

**Reviews**

PubMed Clinical Queries  
[Reviews in PubMed](#)

**Clinical resources**

Clinicaltrials.gov

**Molecular resources**

RefSeqGene  
[View BCKDHB variations in ClinVar](#)

**Consumer resources**

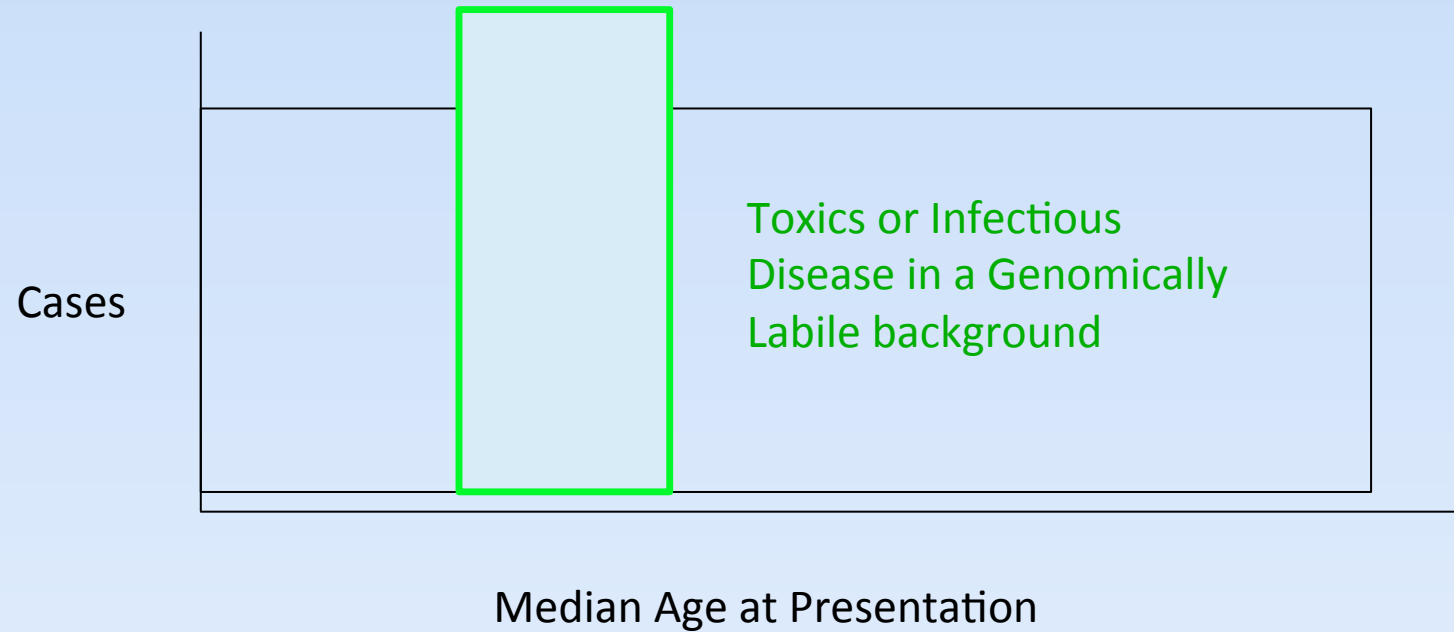
NCATS Office of Rare Diseases Research (GARD)  
[MedlinePlus](#)







# Rethinking Disease



*A completely hypothetical example to  
illustrate a point*

# Pharmacogenomics

[PR Newswire Services](#) | [Knowledge Center](#) | [Browse News Releases](#) | [Contact PR Newswire](#) | [Send a News Release](#)

See more news releases in [Health Care & Hospitals](#) | [Medical Pharmaceuticals](#) | [Licensing](#) | [New Products & Services](#)

## Thomson Reuters and NuMedii Launch Ground-Breaking Initiative to Identify Drugs for Repurposing

Companies leverage leverage content, Big Data analytics and expertise to improve success of drug discovery

Like

0

in

+1

More

Print

Email

Embed

PHILADELPHIA, Nov. 4, 2013 /PRNewswire/ — The IP & Science business of Thomson Reuters, the world's leading provider of intelligent information for businesses and professionals, today announced a strategic initiative with NuMedii, a biotech company dedicated to revolutionizing drug discovery, to identify new treatments for areas of unmet medical need. The partnership, which pairs NuMedii's technology with the authoritative, hand-curated drug and disease information and systems biology expertise of Thomson Reuters, will create an unprecedented repository of content and methodologies to systematically identify new applications for existing drug compounds.

NuMedii will use Thomson Reuters MetaCore and Integrity content, as well as Thomson Reuters Life Sciences Professional Services expertise in systems biology, in conjunction with its unique Big Data technology and databases, to find FDA-approved drugs or discontinued development compounds that are appropriate for repurposing. The companies will leverage this comprehensive collection of high-quality data, knowledge and predictive technologies to identify therapeutic candidates with the greatest probability for clinical success, in turn fulfilling NuMedii's mission to discover and de-risk drugs that target new pathologies.

"NuMedii is blending various innovative technologies and life science data to create a next generation drug discovery and development engine," said Gini Deshpande, Ph.D., founder and chief executive officer of NuMedii. "This unique partnership

### More by this Source

Thomson Reuters lanceert nieuwe, wereldwijde gegevensbank specifiek voor transferprijzen  
Nov 12, 2013, 17:43 ET

Thomson Reuters lance une nouvelle base de données mondiale des entreprises privées spécifiquement conçue pour les prix de transfert  
Nov 12, 2013, 15:00 ET

Thomson Reuters veröffentlicht neue weltweite, speziell für

# CANCER DISCOVERY

[Home](#) | [OnlineFirst](#) | [Current Issue](#) | [Past Issues](#) | [News](#) | [Subscriptions](#) | [Alerts](#) | [Feedback](#) | [AACR Publications](#) | [CME](#) | [AACR Home](#)

Search CD

Go

Advanced Search »

## A Drug Repositioning Approach Identifies Tricyclic Antidepressants as Inhibitors of Small Cell Lung Cancer and Other Neuroendocrine Tumors

Nadine S. Jahchan<sup>1,2</sup>, Joel T. Dudley<sup>1</sup>, Pawel K. Mazur<sup>1,2</sup>, Natasha Flores<sup>1,2</sup>, Dian Yang<sup>1,2</sup>, Alec Palmerton<sup>1,2</sup>, Anne-Flore Zmoos<sup>1,2</sup>, Dedeepya Vaka<sup>1,2</sup>, Kim Q.T. Tran<sup>1,2</sup>, Margaret Zhou<sup>1,2</sup>, Karolina Krasinska<sup>3</sup>, Jonathan W. Riess<sup>4</sup>, Joel W. Neal<sup>5</sup>, Purvesh Khatri<sup>1,2</sup>, Kwon S. Park<sup>1,2</sup>, Atul J. Butte<sup>1,2</sup> and Julien Sage<sup>1,2</sup>

Author Affiliations

Author Notes

Corresponding Authors:

Julien Sage, Stanford University, 265 Campus Drive, SIM1 G2078, Stanford, CA 94305-5457. Phone: 650-724-9246; Fax: 650-736-0195; E-mail: [julsage@stanford.edu](mailto:julsage@stanford.edu); and Atul J. Butte, Stanford University, 251 Campus Drive West, MSOB X1C63, Stanford, CA 94305-5479. Phone: 650-725-1337; Fax: 650-723-7070; [abutte@stanford.edu](mailto:abutte@stanford.edu)

J.T. Dudley and P.K. Mazur contributed equally to this work.

This Article

Published OnlineFirst September 28, 2013; doi: 10.1158/2159-8290.CD-13-0183

Abstract

Full Text (PDF)

Supplementary Data

Classifications

Research Articles

Services

Email this article to a colleague

Alert me when this article is cited

Alert me if a correction is posted

Similar articles in this journal

Similar articles in PubMed

Download to citation manager

Request Permissions

patient ACCESS

Google Scholar

PubMed

Social Bookmarking

Current Issue

November 2013, 3 (11)



Alert me to new issues of Cancer Discovery

News

Call for Papers

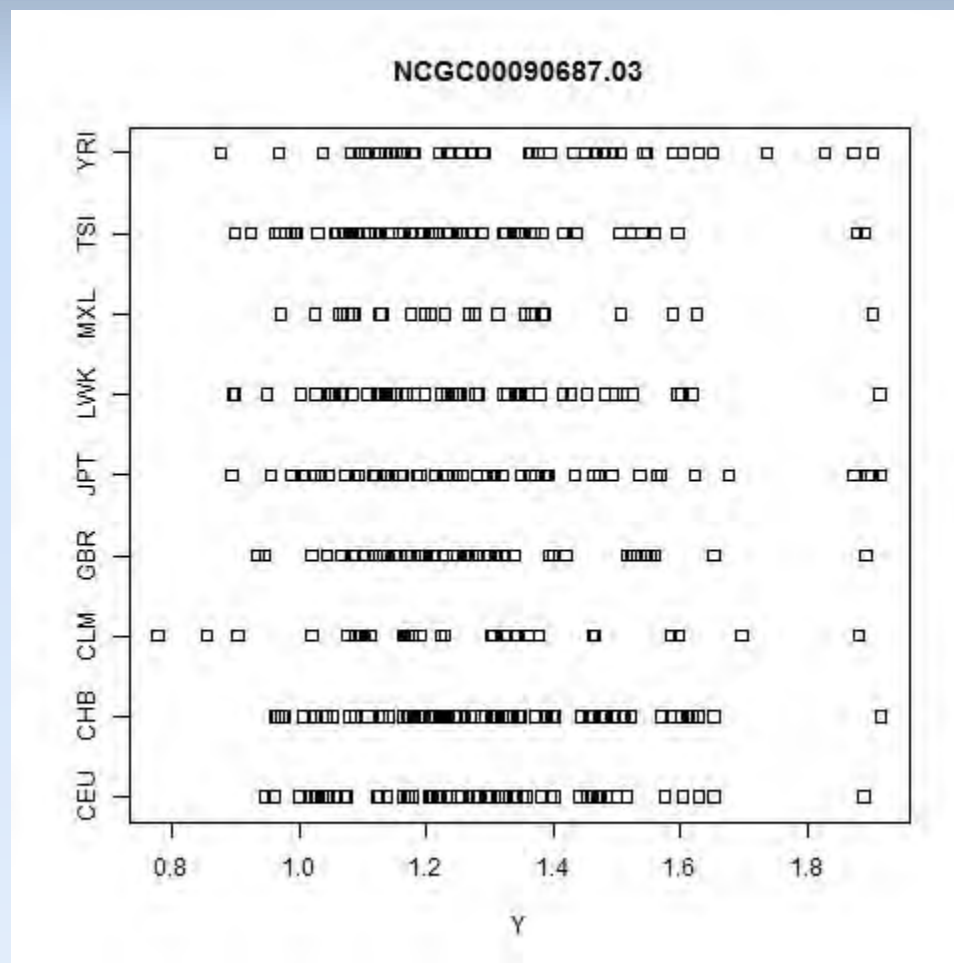
Instructions for Authors/Editorial Policies

Submit a Manuscript

Author Services Center

Journal Sections



# Toxicogenomics






# Bacterial and Viral Infections may Trigger Disease Incidence

← → ↻ 🏠 ⓘ snyderome.stanford.edu



## Integrated Personal Omics Profiling



department of  
GENETICS

Home Data

[G](#) [f](#) [t](#) [v](#) [+](#) 23

### Personal Omics Profiling Reveals Dynamic Molecular and Medical Phenotypes

(Cell. 2012 Mar 16; 148:1293-1307.)

# Bacterial and Viral Infections may Trigger Disease Incidence

Fig. S8

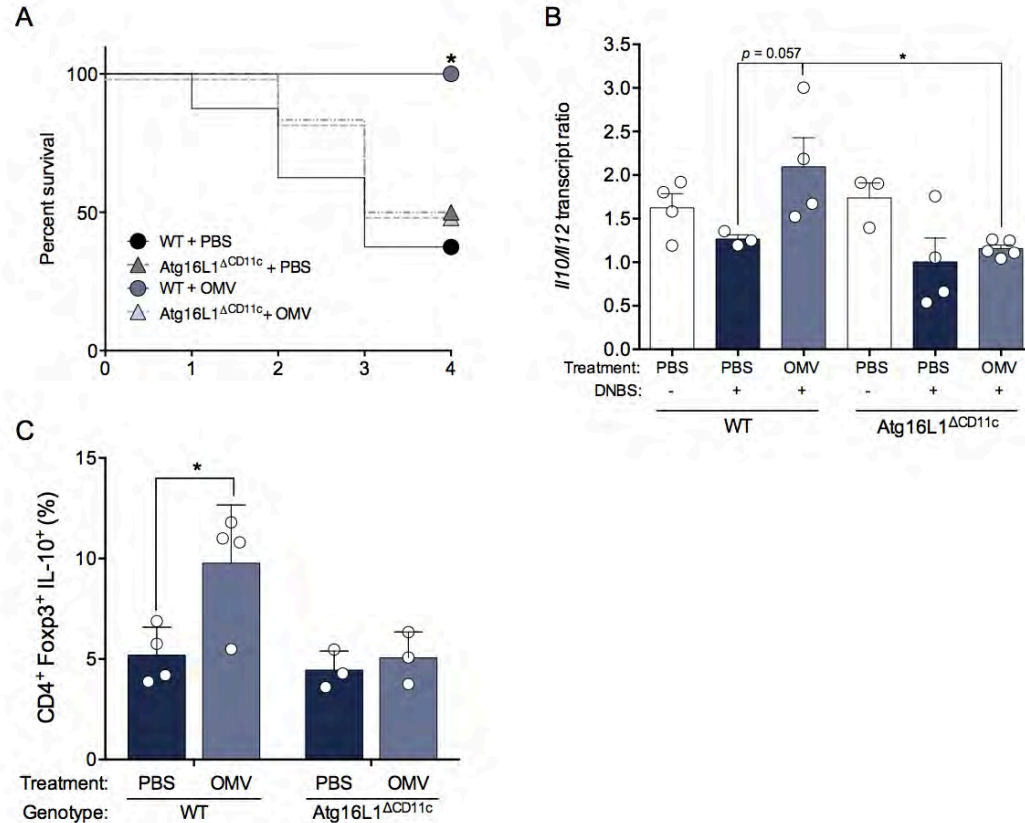


Fig. S8. *B. fragilis* OMV-mediated protection from colitis requires ATG16L1 in DCs.

NCBI Resources How To

PubMed US National Library of Medicine National Institutes of Health

Advanced

Format: Abstract

Send to

Science. 2016 May 27;352(6289):1116-20. doi: 10.1126/science.aad9948. Epub 2016 May 5.

**Gene-microbiota interactions contribute to the pathogenesis of inflammatory bowel disease.**

Chu H<sup>1</sup>, Khosravi A<sup>2</sup>, Kusumawardhani IP<sup>2</sup>, Kwon AH<sup>2</sup>, Vasconcelos AC<sup>3</sup>, Cunha LD<sup>4</sup>, Mayer AE<sup>5</sup>, Shen Y<sup>2</sup>, Wu WL<sup>2</sup>, Kambal A<sup>5</sup>, Targen SR<sup>6</sup>, Xavier RJ<sup>7</sup>, Ernst PB<sup>3</sup>, Green DR<sup>4</sup>, McGovern DP<sup>6</sup>, Virgin HW<sup>5</sup>, Mazmanian SK<sup>1</sup>.

Author information

**Abstract**

Inflammatory bowel disease (IBD) is associated with risk variants in the human genome and dysbiosis of the gut microbiome, though unifying principles for these findings remain largely undescribed. The human commensal *Bacteroides fragilis* delivers immunomodulatory molecules to immune cells via secretion of outer membrane vesicles (OMVs). We reveal that OMVs require IBD-associated genes, ATG16L1 and NOD2, to activate a noncanonical autophagy pathway during protection from colitis. ATG16L1-deficient dendritic cells do not induce regulatory T cells (T(regs)) to suppress mucosal inflammation. Immune cells from human subjects with a major risk variant in ATG16L1 are defective in T(reg) responses to OMVs. We propose that polymorphisms in susceptibility genes promote disease through defects in "sensing" protective signals from the microbiome, defining a potentially critical gene-environment etiology for IBD.

Copyright © 2016, American Association for the Advancement of Science.

PMID: 27230380 PMCID: PMC4996125 [Available on 2017-05-27] DOI: 10.1126/science.aad9948

[PubMed - Indexed for MEDLINE]

# Where is the data? NCBI SRA (and dbGaP)

The screenshot shows the NCBI SRA search results page. The top navigation bar includes the NCBI logo, 'Resources', and 'How To'. The search bar contains 'SRA' and 'personal omics profiling'. Below the search bar, there are links for 'Create alert' and 'Advanced'. The left sidebar shows filters for 'Access' (Public (30)), 'Source' (DNA (5), RNA (25)), 'Type' (genome (5)), and 'Other' (aligned data (2)). The main content area displays 'Search results' for 'Items: 1 to 20 of 30'. A list of six results is shown, each with a checkbox, a link to the study, and details about the run, including the number of spots, bases, and downloads, as well as the accession number.

NCBI Resources How To

SRA SRA personal omics profiling

Create alert Advanced

Access Public (30)

Source DNA (5) RNA (25)

Type genome (5)

Other aligned data (2)

Clear all

Show additional filters

Summary 20 per page Send to:

View results as an expanded interactive table using the RunSelector. [Send results to Run selector](#)

**Search results**

Items: 1 to 20 of 30

1. [GSM818587: microRNA-Seq of timepoint7](#)  
1 ILLUMINA (Illumina Genome Analyzer Ix) run: 25.5M spots, 917.4M bases, 537.7Mb downloads  
Accession: SRX101444

2. [GSM818586: microRNA-Seq of timepoint6](#)  
1 ILLUMINA (Illumina Genome Analyzer Ix) run: 23.6M spots, 848.2M bases, 481.8Mb downloads  
Accession: SRX101443

3. [GSM818585: microRNA-Seq of timepoint5](#)  
1 ILLUMINA (Illumina Genome Analyzer Ix) run: 31M spots, 1.1G bases, 702.7Mb downloads  
Accession: SRX101442

4. [GSM818584: microRNA-Seq of timepoint3](#)  
1 ILLUMINA (Illumina Genome Analyzer Ix) run: 31.2M spots, 1.1G bases, 705.7Mb downloads  
Accession: SRX101441

5. [GSM818583: microRNA-Seq of timepoint2](#)  
1 ILLUMINA (Illumina Genome Analyzer Ix) run: 29.4M spots, 1.1G bases, 681.5Mb downloads  
Accession: SRX101440

6. [GSM818582: RNA-Seq of timepoint 21](#)  
1 ILLUMINA (Illumina Genome Analyzer) run: 65.4M spots, 13.2G bases, 8.1Gb downloads  
Accession: SRX101439



# Blasting into SRA

NCBI BLAST/ blastn suite-SRA/ Formatting Results - 4JRYNZUA015

[Edit and Resubmit](#) [Save Search Strategies](#) [Formatting options](#) [Download](#) [YouTube](#) [How to read this page](#) [Blast report description](#)

gi|372099106:156199424-156200818 Mus musculus...

RID 4JRYNZUA015 (Expires on 10-25 04:06 am)

Query ID lc|30149

Description gi|372099106:156199424-156200818 Mus musculus strain C57BL/6J chromosome 4, GRCm38.p2 C57BL/6J

Molecule type nucleic acid

Query Length 910

Database Name SRA

Description See details

Program BLASTN 2.2.30+ [Citation](#)

Other reports: [Search Summary](#) [Distance tree of results](#)

**Graphic Summary**

Distribution of 100 Blast Hits on the Query Sequence

Mouse-over to show details and scores, click to show alignments

Color key for alignment scores

Query 1 150 300 450 600 750 900

<40 40-50 50-80 80-200 >=200

**Descriptions**

Sequences producing significant alignments:

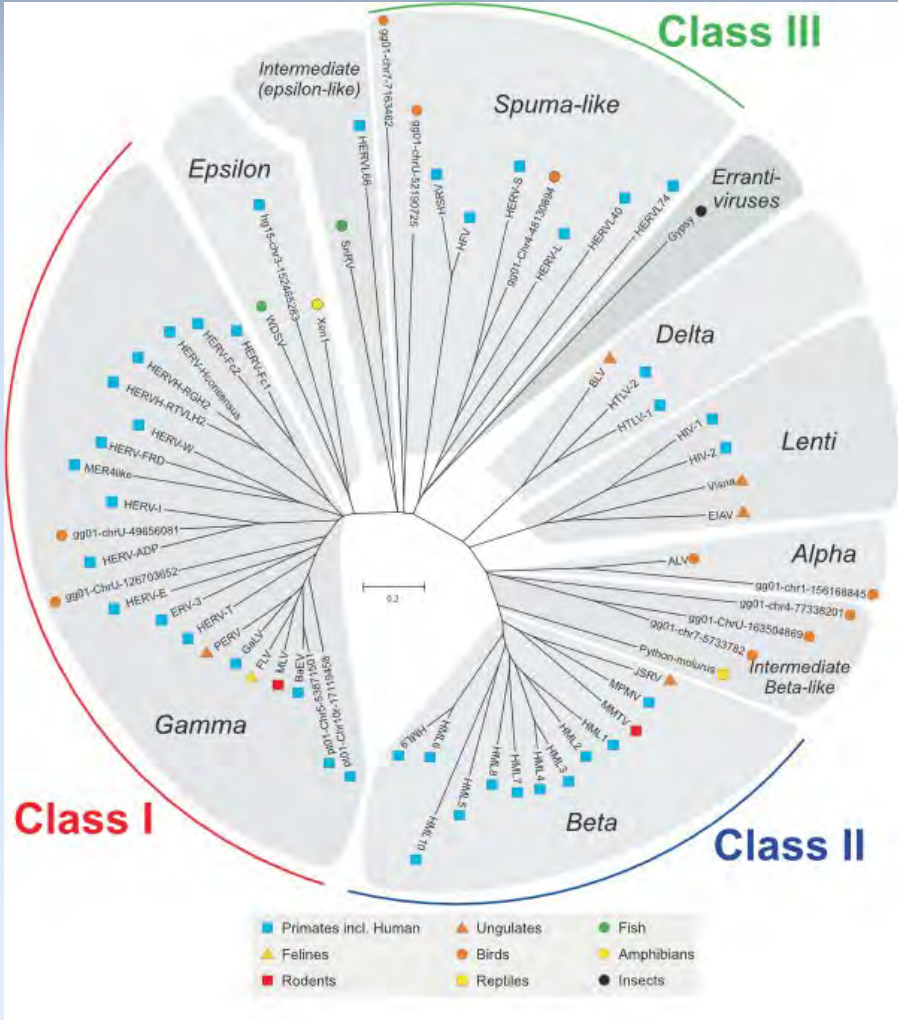
Select: All None Selected: 0

Alignments [Download](#) [Graphics](#) [Distance tree of results](#)

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	SRX335902	93.5	93.5	5%	3e-17	100%	<a href="#">SRA:SRR953224.434425.1</a>
<input type="checkbox"/>	SRX335902	93.5	93.5	5%	3e-17	100%	<a href="#">SRA:SRR953224.404386.1</a>
<input type="checkbox"/>	SRX335902	93.5	93.5	5%	3e-17	100%	<a href="#">SRA:SRR953224.352795.1</a>
<input type="checkbox"/>	SRX335902	93.5	93.5	5%	3e-17	100%	<a href="#">SRA:SRR953224.298156.1</a>
<input type="checkbox"/>	SRX335902	93.5	93.5	5%	3e-17	100%	<a href="#">SRA:SRR953224.277011.1</a>
<input type="checkbox"/>	SRX335902	93.5	93.5	5%	3e-17	100%	<a href="#">SRA:SRR953224.51202.1</a>
<input type="checkbox"/>	SRX335902	93.5	93.5	5%	3e-17	100%	<a href="#">SRA:SRR953223.3789925.1</a>
<input type="checkbox"/>	SRX335902	93.5	93.5	5%	3e-17	100%	<a href="#">SRA:SRR953223.3749502.1</a>
<input type="checkbox"/>	SRX335902	93.5	93.5	5%	3e-17	100%	<a href="#">SRA:SRR953223.3654842.1</a>
<input type="checkbox"/>	SRX335902	93.5	93.5	5%	3e-17	100%	<a href="#">SRA:SRR953223.3145970.1</a>
<input type="checkbox"/>	SRX335902	93.5	93.5	5%	3e-17	100%	<a href="#">SRA:SRR953223.3094680.1</a>





# Endogenous Retroviruses




Jern P, Sperber GO, Blomberg J

# NCBI Ebola Portal

 NCBI

Resources 

How To 

busbybr@ncbi.nlm.nih.gov

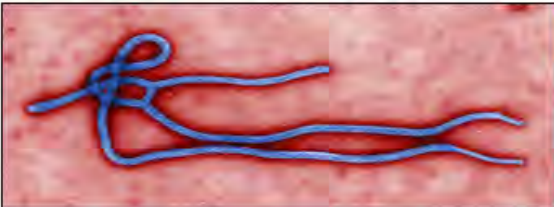
My NCBI

Sign Out

Virus Variation

Search NCBI

Search



## Ebolavirus Resource

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

### Ebolavirus database

- [Search nucleotide and protein sequences](#)
- [Help](#)
- [How to Cite Us](#)
- [Contact Us](#)

### Other NCBI Ebolavirus resources

- [Zaire ebolavirus reference genome](#)
- [Publications](#)
- [Genome browser](#)
- [Taxonomy](#)

### External Ebolavirus resources

- [Ebola Outbreak 2014: Information Resources](#)
- [Health Map](#)
- [UCSC Ebola Genome Portal](#)
- [ViralZone](#)
- [Virus Pathogen Resource](#)