

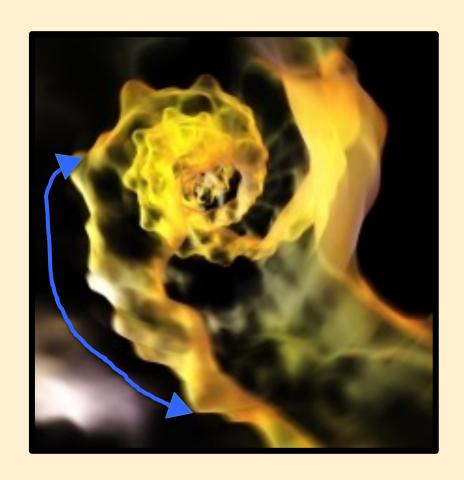
# Through the Looking Glass: Genetic Information Resources for You, Health Care Professionals, and Researchers

# Let's go on a journey...



# using x-ray crystallographic diffraction!

# What is it?



Spiral IV – X-ray crystallographic data was used from real DNA molecules to paint a unique portrait of the double helix.

The image omits the chemical bonds that crisscross the center of the molecule, so that the structural features of the helix, such as the major and minor grooves, can be seen more easily.

Credit: Kenneth Eward, BioGrafx Scientific & Medical Images, Ovid Michigan

# Who are they?



J. Craig Venter, Ph.D.

founder Celera Genomics,

The Institute for Genomic Research, &

J. Craig Venter Institute



Francis Collins, M.D.

director (in the year 2000) of
National Center for Human
Genome Research

# The Human Genome Project

Identified and mapped all genes in human genome

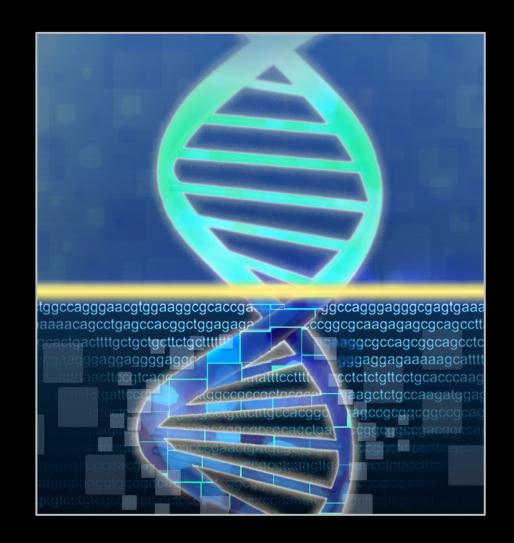
Largest international collaborative effort ever

Dr. Collins directed NIH's public project

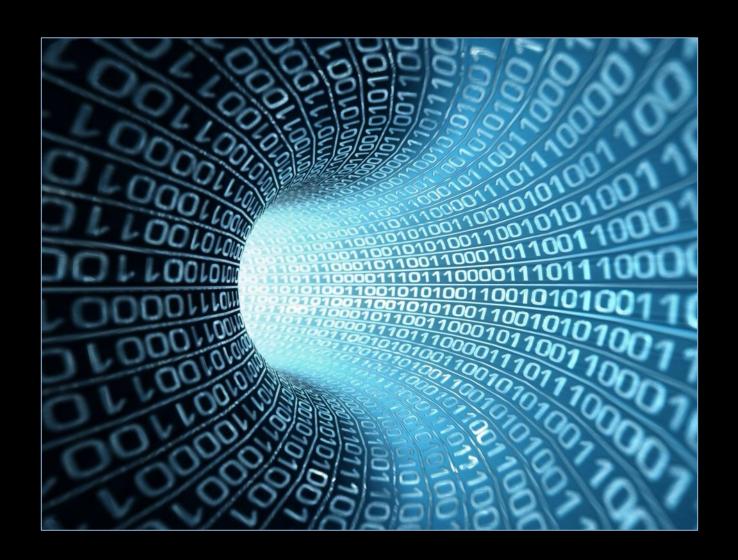
Dr. Venter directed

Celera's private project

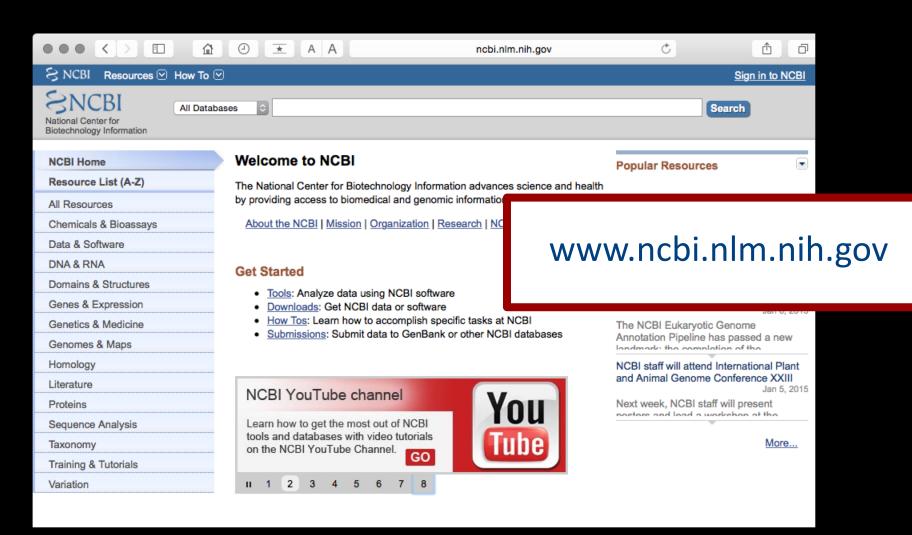
Completed in February 2001



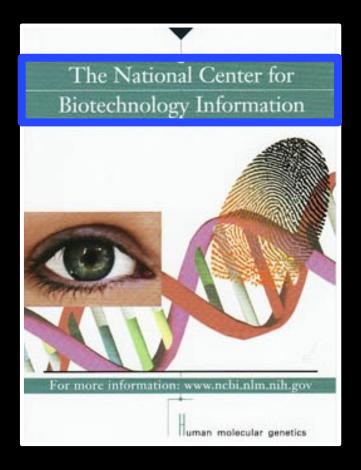
# Genetics Research Produces Big Data!!

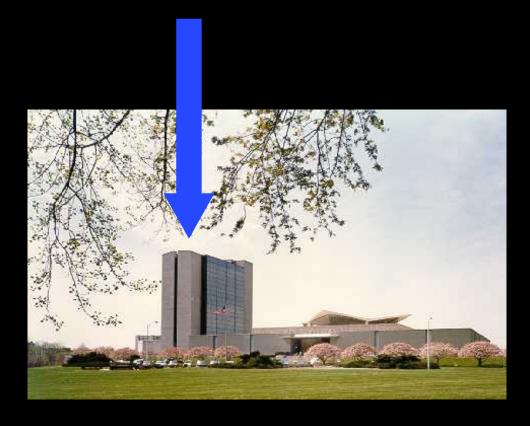


# NCBI's Databases are Freely Available for Anyone!



# Produced by





at the National Library of Medicine at the National Institutes of Health

But I'm Not a Scientist or a Doctor or ...

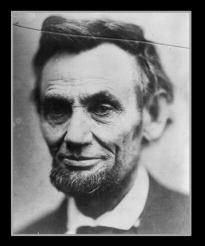
Why is this important to me???



http://www.research.archives.go

Angelina Jolie

**Breast Cancer** 



www.loc.go

Abraham Lincoln

Marfan's Syndrome

# Genetic predispositions or inherited diseases!



http://www.directorsblog.nih.gov

Sam Berns

Progeria

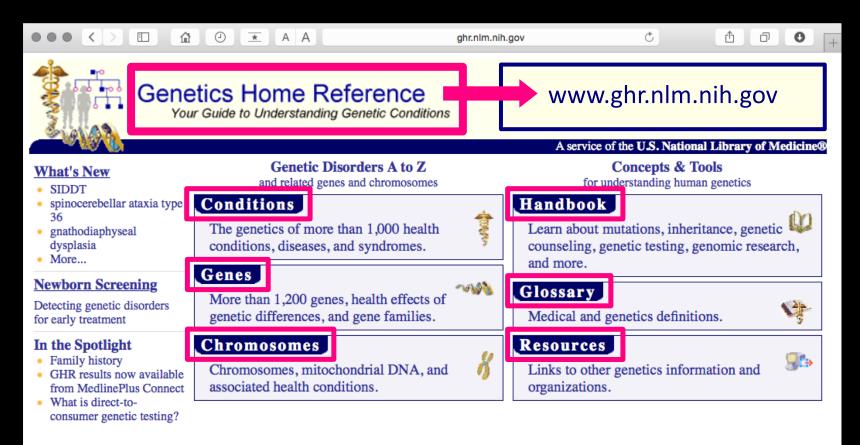


http://www.blogs.america.gov

Sonia Sotomayor

Type 1 Diabetes

# Free Genetic Information Resources for Everyone!



Genetics Home Reference provides consumer-friendly information about the effects of genetic variations on human health.

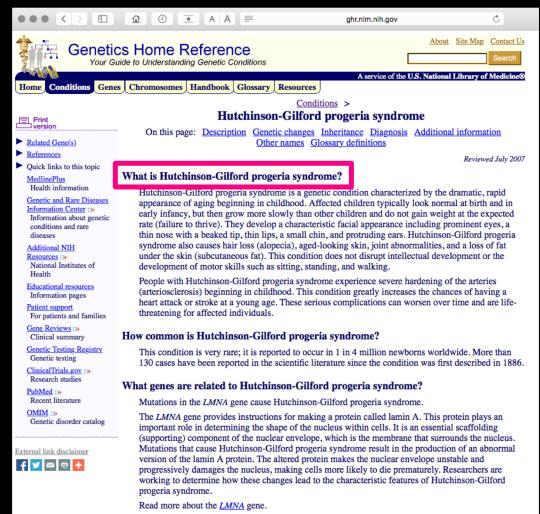
The resources on this site should not be used as a substitute for professional medical care or advice. Users seeking information about a personal genetic disease, syndrome, or condition should consult with a qualified healthcare professional. See <a href="How can I find a genetics">How can I find a genetics</a> <a href="professional">professional in my area?</a> in the Handbook.

Published: December 29, 2014

# What is Progeria?



Sam Berns



# What is a gene?

Definition(s)

Genetics Home Reference  Your Guide to Understanding Genetic Conditions	About Site Map Contact Us  Search
	A service of the U.S. National Library of Medicine®
Home Conditions Genes Chromosomes Handbook Glossary	Resources
	Glossary
$\underline{0-9} \mid \underline{\mathbf{A}} \mid \underline{\mathbf{B}} \mid \underline{\mathbf{C}} \mid \underline{\mathbf{D}} \mid \underline{\mathbf{E}} \mid \underline{\mathbf{F}} \mid \underline{\mathbf{G}} \mid \underline{\mathbf{H}} \mid \underline{\mathbf{I}} \mid \underline{\mathbf{J}} \mid$	$ \underline{K} \mid \underline{L} \mid \underline{M} \mid \underline{N} \mid \underline{O} \mid \underline{P} \mid \underline{Q} \cdot \underline{R} \mid \underline{S} \mid \underline{T} \mid \underline{U} \mid \underline{V} \mid \underline{W} \mid \underline{X} \mid \underline{Y} \cdot \underline{Z}$
Gene	

The functional and physical unit of heredity passed from parent to offspring. Genes are pieces of DNA, and most
genes contain the information for making a specific protein.

Definition from: Physician Data Query via <u>Unified Medical Language System</u> ⇒ at the National Library of Medicine

The basic unit of heredity, consisting of a segment of DNA arranged in a linear manner along a chromosome, which
codes for a specific protein or segment of protein leading to a particular characteristic or function

Definition from: GeneReviews from the University of Washington and the National Center for Biotechnology Information

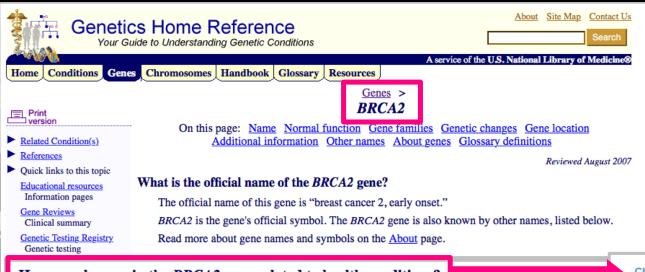
The gene is the basic physical unit of inheritance. Genes are passed from parents to offspring and contain the
information needed to specify traits. Genes are arranged, one after another, on structures called chromosomes. A
chromosome contains a single, long DNA molecule, only a portion of which corresponds to a single gene. Humans
have approximately 23,000 genes arranged on their chromosomes.

Definition from: Talking Glossary of Genetic Terms is from the National Human Genome Research Institute

The fundamental physical and functional unit of heredity. A gene is an ordered sequence of nucleotides located in a
particular position on a particular chromosome that encodes a specific functional product (i.e., a protein or RNA
molecule).

Definition from: <u>Human Genome Project Information</u> : at the U.S. Department of Energy

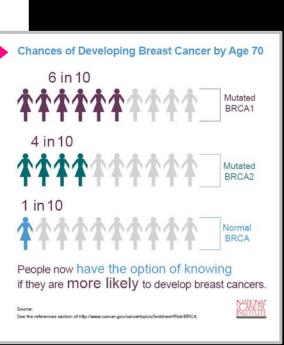
# My Doctor says I have a BRCA2 gene variant -



## How are changes in the BRCA2 gene related to health conditions?

breast cancer - increased risk from variations of the BRCA2 gene

Researchers have identified more than 800 mutations in the *BRCA2* gene, many of which are associated with an increased risk of breast cancer. Many *BRCA2* mutations insert or delete a small number of DNA building blocks (nucleotides) in the gene. Most of these genetic changes disrupt protein production from one copy of the gene in each cell, resulting in an abnormally small, nonfunctional version of the BRCA2 protein. Researchers believe that the defective BRCA2 protein is unable to help repair damaged DNA or fix mutations that occur in other genes. As these defects accumulate, they can allow cells to grow and divide uncontrollably and form a tumor.



# Which chromosome is related to Down Syndrome?



## Down syndrome

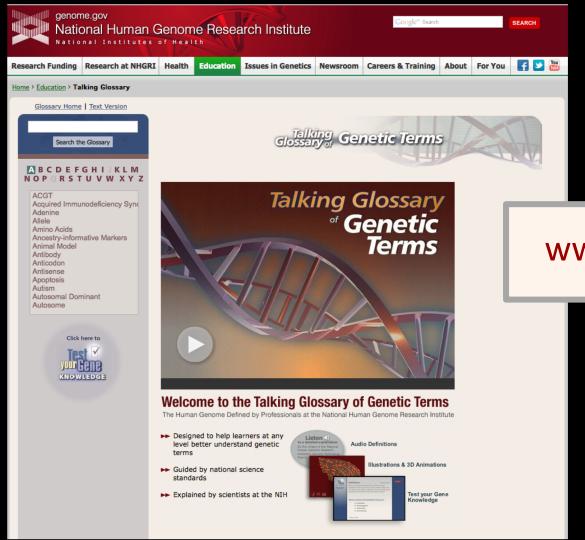
Down syndrome is a chromosomal condition that is associated with intellectual disability, a characteristic facial appearance, and weak muscle tone (hypotonia) in infancy. This condition is most often coused by trisomy 21. Trisomy 21 means that each cell in the body has three copies of chromosome 21 in stead of the usual two copies.

Less commonly, Down syndrome occurs when part of chromosome 21 becomes attached (translocated) to another chromosome during the formation of reproductive cells (eggs and sperm) or very early in fetal development. Affected people have two copies of chromosome 21 plus extra material from chromosome 21 attached to another chromosome, resulting in three copies of genetic material from chromosome 21. Affected individuals with this genetic change are said to have translocation Down syndrome.

In a very small percentage of cases, Down syndrome results from an extra copy of chromosome 21 in only some of the body's cells. In these people, the condition is called mosaic Down syndrome.

Researchers believe that having extra copies of genes on chromosome 21 disrupts the course of normal development, causing the characteristic features of Down syndrome and the increased risk of health problems associated with this condition.

# Talking Glossary of Genetic Terms



www.genome.gov

Research Funding Research at NHGRI Health Education Issues in Genetics Newsroom Careers & Training About For You



Home > Education > Talking Glossary

Glossary Home | Text Version

Search the Glossary

ABCDEFGHIJKL NOPORSTUVWX

Deletion

Diabetes (Diabetes Mellitus) Diploid

DNA (Deoxyribonucleic Acid) DNA Fingerprinting

**DNA Replication** 

**DNA Sequencing** Dominant

Double Helix Down Syndrome (Trisomy 21

Duplication

Click here to





Deletion

M Pronunciation

Deletion is a type of mutation involving the loss of genetic material. It can be small, involving a single missing DNA base pair, or large, involving a piece of a chromosome.

**★**) Listen Maximilian Muenke, M.D. defines Deletion

En Espanol



Maximilian Muenke, M.D.

Chief and Senior Investigator, Medical Genetics Branch; Head, Human Development Section; Director, Medical Genetics Residency and Fellowship Training Programs and Combined Pediatrics and Medical Genetics Residency Training Program

Dr. Muenke's research program seeks to improve knowledge about the formation of the central nervous system and to elucidate the origin of developmental disabilities and mental retardation. Specifically, his laboratory investigates birth defects that affect normal embryonic development and lead to neurological impairment. His two major areas of focus are holoprosencephaly (HPE) and attention deficit hyperactivity disorder (ADHD). Holoprosencephaly is a common birth defect characterized by the failure of the brain to divide properly into left and right hemispheres during early development. Children born with HPE show various degrees of developmental disabilities and mental retardation.



Send this term to a friend How to cite this term for research papers

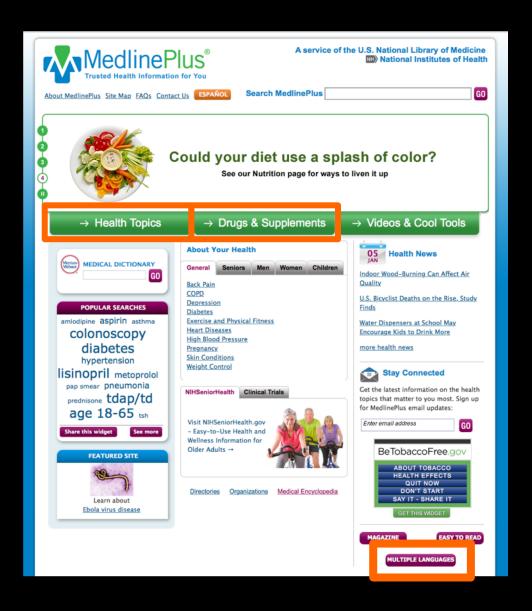
## Related Terms

- Base Pair
- Chromosome
- DNA (Deoxyribonucleic Acid)
- Gene
- Mutation

## Illustrations

- Download PDF
- Download PPT

# I want to know more about health ...



www.MedlinePlus.gov

from National Library of Medicine at N.I.H.

free, reliable information

easy to understand

links to more resources

many different languages

I'm a Clinician or a Nurse or a Therapist...

Why is this important to me???

# Web-Based Resources for Clinical Bioinformatics Anthony M. Joshua and Paul C. Boutros

Methods in Molecular Medicine. 2008; 141:309-29

PMID: 18453097

In the post-Human Genome Project era, awareness of the resources available through the internet is essential to both molecular biologists and clinicians.

... (it) is important to understand the principles upon which hypotheses are generated, experiments are based and conclusions reached.

... (knowing about) resources often facilitates their use and adoption

# Have you had a lot of flu patients?





# Influenza (Flu)







Share

The most recent FluView report for the 2014-2015 flu season shows that flu season in the United States has begun and about half the country is experiencing high levels of flu activity. Reports of flu illnesses, hospitalizations and deaths are elevated. Activity is expected to continue for several weeks, especially in parts of the country that have not yet seen significant activity.

# Data from NIAID Influenza Genome Sequencing Project and GenBank



data essential to identification of the genetic determinants of influenza pathogenicity

tools for flu sequence analysis

# DNA & clinical data related to human Major Histocompatibility Complex

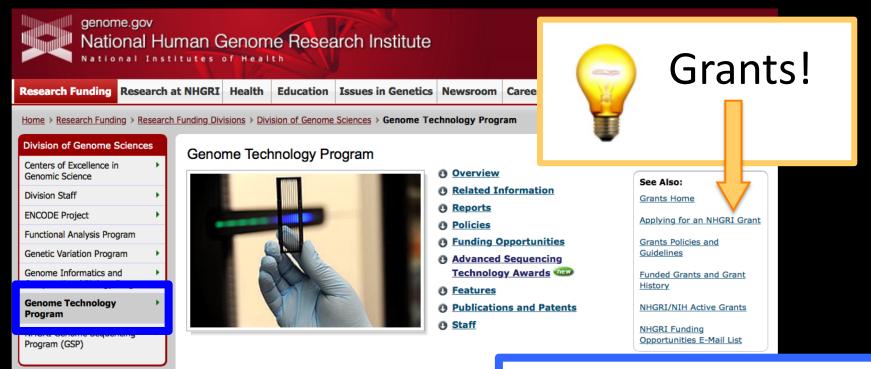


MHC: array of genes that play a central role in



the success of organ transplants

# National Human Genome Research Institute



## Genome Technology Program Overview

The Genome Technology program supports research to develop new low-cost determination of DNA sequence, SNP genotyping (Genetic 1 defined) experiments (Functional Analysis Program). Priorities include efficiency and decrease cost while maintaining or improving data que to achieve orders-of-magnitude improvement. Integration of process

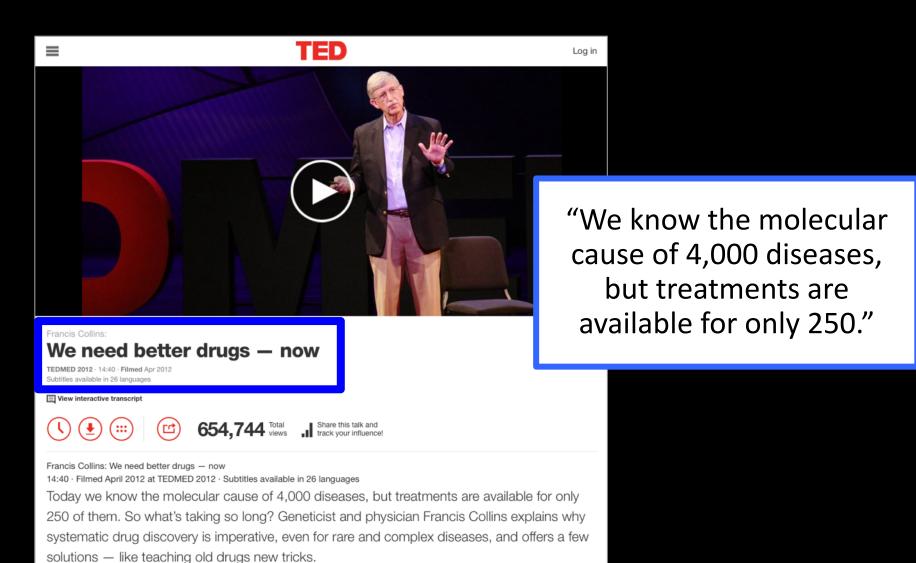
The program also supports and coordinates transfer of technology fr multidisciplinary programs that closely integrate research projects a

The National Human Genome Research Institute (NHGRI) participate bioengineering research. The development of integrated concepts ar research, methods and approaches. NHGRI supports bioengineering program announcements and requests for application, and through announcements, such as the Bioengineering Research Grants (BRG) and other Funding Opportunities that are listed below.

# NHGRI GOAL:

Sequencing a genome for only \$1,000.00!

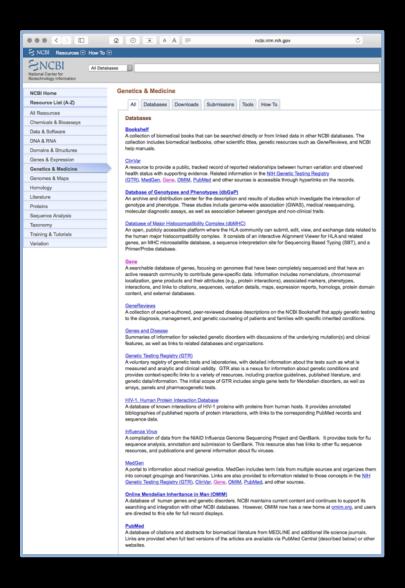
# NIH Director: Francis Collins, M.D.



# I'm a Research Scientist...

# What do the NCBI Genetics Databases include???

# **Growing Collection of Publicly Available Databases**



www.ncbi.nlm.nih.gov

# NCBI's Molecular Biology Databases



Genome assembly organization and additional information



The EST database is a collection of short single-read transcript sequences from GenBank. These sequences provide a resource to evaluate gene expression, find potential variation, and annotate genes.



Organizes information related to human medical genetics, such as attributes of conditions with a genetic contribution.



## RefSeqGene

RefSeqGene defines genomic sequences to be used as reference standards for well-characterized genes and is part of the Loous Refer Genomic (I RG) Project



A BioProject is a collection of biological data related to a single initiative, originating from a single organization or from a consortium. A BioProject record provides users a single place to find links to the diverse data types generated for that project.





OMIM

including GenBank, RefSeq, TPA and PDB. Genome, gene and transcript



Information about retroviruses and specialized tools for the analysis of



systems and their component genee, proteins, and small molecules, as well as literature describing those biosystems and other related data throughout

CCDS



OMM is a comprehensive, authoritative compendium of human genes and genetic phenotypes that is freely available and updated dely. OMM is authored and edited at the McKatck hithman Iretatus of Genetic Medicine, Johns Hopkins University School of Medicine, under the direction of Dr. Ada Hamosh. Its official home is omim.org.





Last week Top 10 Arrivals (12/21/2014 - 12/27/2014)

evolutionary relatedness of a population. The population could originate from different members of the same species, or from organisms from different spe-





## CCDS Database EBI . HGNC . MGI . NCBI . UCSC . WTSI







Database of single nucleotide polymorphisms (SNPs) and multiple small-scale variations that include insertions/deletions, microsatellites, and non-

SARS Coronavirus Genome Alignment Tool Help Contact us

6.422



ClinVar aggregates information about sequence variation and its relationship to human health.

The Conserved Domain Database is a resource for the annotation of functional units in proteins. Its collection of domain models includes a set

curated by NCBI, which utilizes 3D structure to provide insights into



Genome Reference Consortium

The Protein database is a collection of sequences from several sources, including translations from annotated coding regions in GenBank, RefSeq and TPA, as well as records from SwissProt, PIR, PRF, and POB. Protein sequences are the fundamental determinants of biological structure and function.





Clone DB is a database that integrates information about clones and libraries, including sequence data, map positions and distributor information. It replaces the former NCBI Clone Registry.



compliant data submissions. Array- and sequence-based data are

accepted. Tools are provided to help users query and download

Gene Expression Omnibus

This database stores curated gene expression DataSets, as well as original Series and Platform records in the Gene Expression Omnibus (GEO) reposito Enter search terms to locate experiments of interest. DataSet records contain

GEO



10000

0000

## PubChem BioAssay

Protein Clusters

The Publichem BioAssay Database comains bloactivity screens of chemici substances described in PubChem Substance. It provides searchable descriptions of each bioassay, including descriptions of the conditions and readouts specific to that screening procedure.



hree dimensional structures provide a wealth of information on the biological function and the evolutionary history of macromolecules. They can be used to examine sequence-structure-function relationships, interactions, active sites,



## MMMMMMXX



The database of Genotypes and Phenotypes (dbGaP) was developed to archive and distribute the results of studies that have investigated the





Information provided to describe substances in PubChem Substance. Structures stored within PubChem Compounds are pre-clustered and cross referenced by identity and similarity groups.



ACGTCGGTAC CTCT TC ACGTCGGTAC UniGene computationally identifies transcripts from the same locus; analyzes expression tissue, age, and health status; and reports related proteins (protEST) and clone resources GTGT



dbMHC Home



An automated system for constructing putative homology groups from the complete gene sets of a wide range of eukaryotic species.





## **Virus Variation**

Retrieve, analyze, and download viral sequences from a value added database using a specialized search interface.



Explore, view, and download genome-wide maps of DNA and histone modifications from our diverse collection of epigenomic data sets

Accounts External Links Contact Us

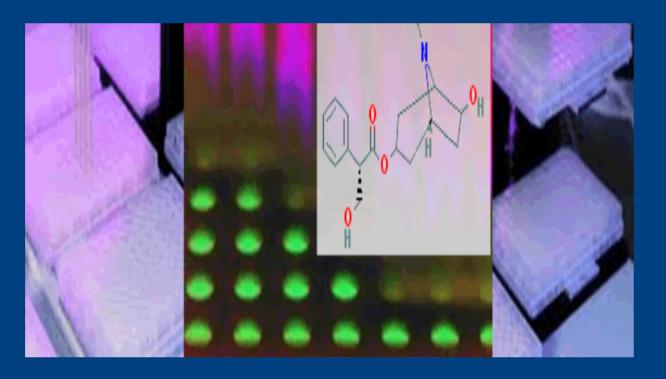




## A comprehensive, integrated, non-redundant, well-annotated set of reference sequences including genomic, transcript, and protein.

This resource provides viral and viroid genome sequence data and related

# **Chemical and Bioassays**



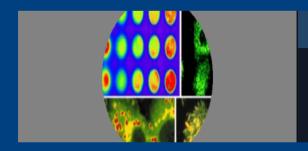




## PubChem BioAssay



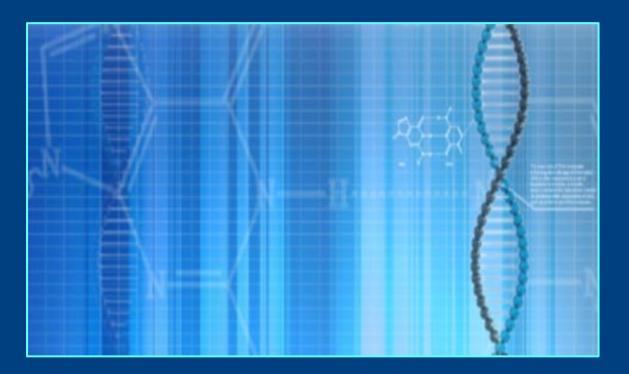




# **PubChem BioAssay**

The PubChem BioAssay Database contains bioactivity screens of chemical substances described in PubChem Substance. It provides searchable descriptions of each bioassay, including descriptions of the conditions and readouts specific to that screening procedure.

# **DNA** and RNA





















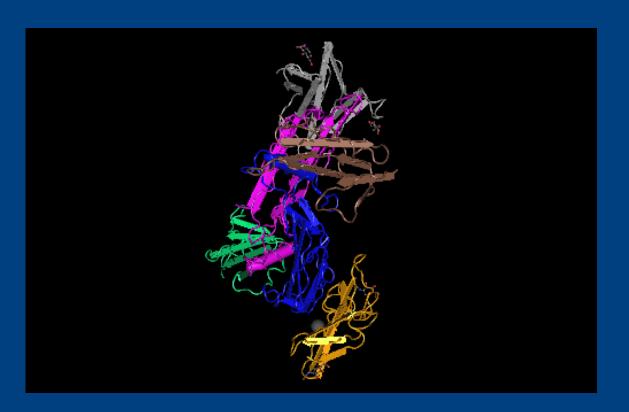
RefSeq: NCBI Reference Sequence
Database



# **Nucleotide**

The Nucleotide database is a collection of sequences from several sources, including GenBank, RefSeq, TPA and PDB. Genome, gene and transcript sequence data provide the foundation for biomedical research and discovery.

# **Domains and Structures**



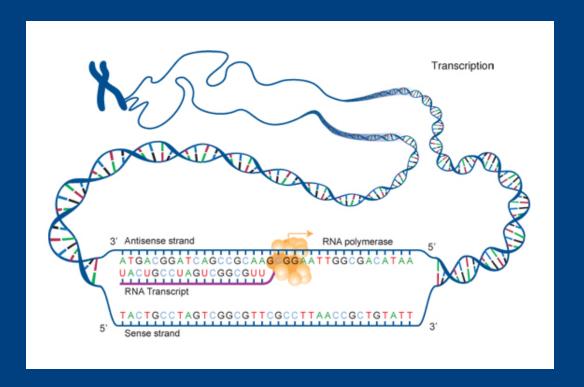




# **Structure**

Three dimensional structures provide a wealth of information on the biological function and the evolutionary history of macromolecules. They can be used to examine sequence-structure-function relationships, interactions, active sites, and more.

# Genes and Expression





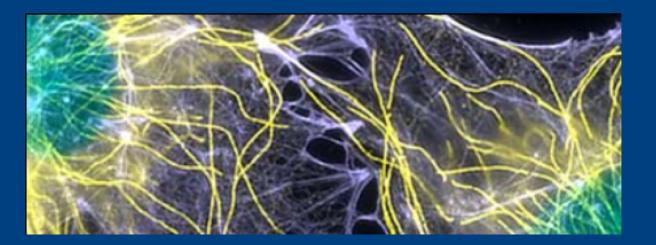


## Gene

Gene integrates information from a wide range of species. A record may include nomenclature, Reference Sequences (RefSeqs), maps, pathways, variations, phenotypes, and links to genome-, phenotype-, and locus-specific resources worldwide.



# **Genomes and Maps**







# Genome

This resource organizes information on genomes including sequences, maps, chromosomes, assemblies, and annotations.

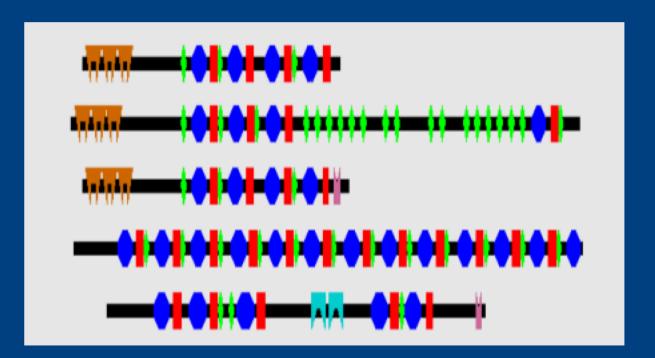




Virus Variation

Retrieve, analyze, and download viral sequences from a value added database using a specialized search interface.

# Homology











## **Protein Clusters**

This collection of related protein sequences (clusters) consists of proteins derived from the annotations of whole genomes, organelles and plasmids. It currently limited to Archaea, Bacteria, Plants, Fungi, Protozoans, and Viruses

# Literature







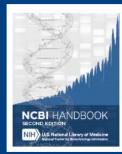
## PubMed

PubMed comprises more than 24 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites.



## **PMC**

PubMed Central® (PMC) is a free full-text archive of biomedical and life sciences journal literature at the U.S. National Institutes of Health's National Library of Medicine (NIH/NLM).



## The NCBI Handbook, 2nd edition

Bethesda (MD): National Center for Biotechnology Information (US); 2013-.

Copyright and Permissions

Search this book



## GeneReviews<sup>®</sup>

Edited by Roberta A Pagon, Editor-in-chief, Margaret P Adam, Holly H Ardinger, Thomas D Bird, Cynthia R Dolan, Chin-To Fong, Richard JH Smith, and Karen Stephens.

Seattle (WA): <u>University of Washington, Seattle</u>; 1993-2014. ISSN: 2372-0697

Copyright and Permissions

Search GeneReviews

GeneReviews Advanced Search Help

# **Proteins**







## **BioSystems**









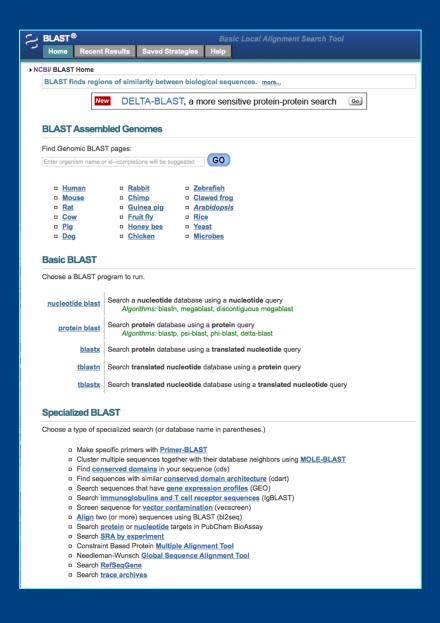
## RefSeq: NCBI Reference Sequence Database



# **Protein**

The Protein database is a collection of sequences from several sources, including translations from annotated coding regions in GenBank, RefSeq and TPA, as well as records from SwissProt, PIR, PRF, and PDB. Protein sequences are the fundamental determinants of biological structure and function.

### **Sequence Analysis**



### Taxonomy





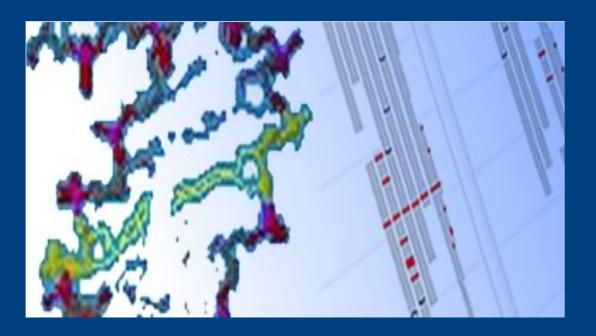




### **OMIM**

OMIM is a comprehensive, authoritative compendium of human genes and genetic phenotypes that is freely available and updated daily. OMIM is authored and edited at the McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, under the direction of Dr. Ada Hamosh. Its official home is omim.org.

### **Variation**













ACTGATGGTATGGGGCCAAGAGATATATCT CAGGTACGGCTGTCATCACTTAGACCTCAC CAGGGCTGGGCATAAAAGTCAGGGCAGAGC CCATGGTGCATCTGACTCCTGAGGAGAAGT GCAGGTTGGTATCAAGGTTACAAGACAGGT GGCACTGACTCTCTCTGCCTATTGGTCTAT

### ClinVar

ClinVar aggregates information about sequence variation and its relationship to human health.

### More Information

### Database resources of the National Center for **Biotechnology Information**

Nucleic Acids Research. 2010 Jan;38(Database issue):D13-D25

PMID: 19910364

Eric W. Sayers<sup>1,\*</sup>, Tanya Barrett<sup>1</sup>, Dennis A. Benson<sup>1</sup>, Evan Bolton<sup>1</sup>, Stephen H. Bryant<sup>1</sup>, Kathi Canese<sup>1</sup>, Vyacheslav Chetvernin<sup>1</sup>, Deanna M. Church<sup>1</sup>, Michael DiCuccio<sup>1</sup>, Scott Federhen<sup>1</sup>, Michael Feolo<sup>1</sup>, Ian M. Fingerman<sup>1</sup>, Lewis Y. Geer<sup>1</sup>, Wolfgang Helmberg<sup>2</sup>, Yuri Kapustin<sup>1</sup>, Sergey Krasnov<sup>1</sup>, David Landsman<sup>1</sup>, David J. Lipman<sup>1</sup>, Zhiyong Lu<sup>1</sup>, Thomas L. Madden<sup>1</sup>, Tom Madej<sup>1</sup>, Donna R. Maglott<sup>1</sup>, Aron Marchler-Bauer<sup>1</sup>, Vadim Miller<sup>1</sup>, Ilene Karsch-Mizrachi<sup>1</sup>, James Ostell<sup>1</sup>, Anna Panchenko<sup>1</sup>, Lon Phan<sup>1</sup>, Kim D. Pruitt<sup>1</sup>, Gregory D. Schuler<sup>1</sup>, Edwin Sequeira<sup>1</sup>, Stephen T. Sherry<sup>1</sup>, Martin Shumway<sup>1</sup>, Karl Sirotkin<sup>1</sup>, Douglas Slotta<sup>1</sup>, Alexandre Souvorov<sup>1</sup>, Grigory Starchenko<sup>1</sup>, Tatiana A. Tatusova<sup>1</sup>, Lukas Wagner<sup>1</sup>, Yanli Wang<sup>1</sup>, W. John Wilbur<sup>1</sup>, Eugene Yaschenko<sup>1</sup> and Jian Ye<sup>1</sup>

<sup>1</sup>National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Building 38A, 8600 Rockville Pike, Bethesda, MD 20894, USA and <sup>2</sup>University Clinic of Blood Group Serology and Transfusion Medicine, Medical University of Graz, Auenbruggerplatz 3, A-8036 Graz, Austria

Received September 30, 2011: Revised and Accepted November 14, 2011

acid sequence database, the National Center for Biotechnology Information (NCBI) provides analysis and retrieval resources for the data in GenBank and other biological data made available through the NCBI Website, NCBI resources include Entrez, the Entrez Programming Utilities, MyNCBI, PubMed, PubMed Central (PMC), Gene, the NCBI Taxonomy Browser, BLAST, BLAST Link (BLink), Primer-BLAST, COBALT, Splign, RefSeq, UniGene, HomoloGene, ProtEST, dbMHC, dbSNP, dbVar, Epigenomics, Genome and related tools, the Map Viewer, Model Maker, Evidence Viewer, Trace Archive, Sequence Read Archive, BioProject, BioSample, Retroviral Genotyping Tools, HIV-1/ Human Protein Interaction Database, Gene Expression Omnibus (GEO), Probe, Online Mendelian Inheritance in Animals (OMIA), the Molecular Modeling Database (MMDB), the Conserved Domain Database (CDD), the Conserved Domain Architecture Retrieval Tool (CDART), Biosystems,

molecule databases. Augmenting many of the Web applications are custom implementations of the BLAST program optimized to search specialized data sets. All of these resources can be accessed through the NCBI home page at www.ncbi.nlm.nih

PMID: 19910364

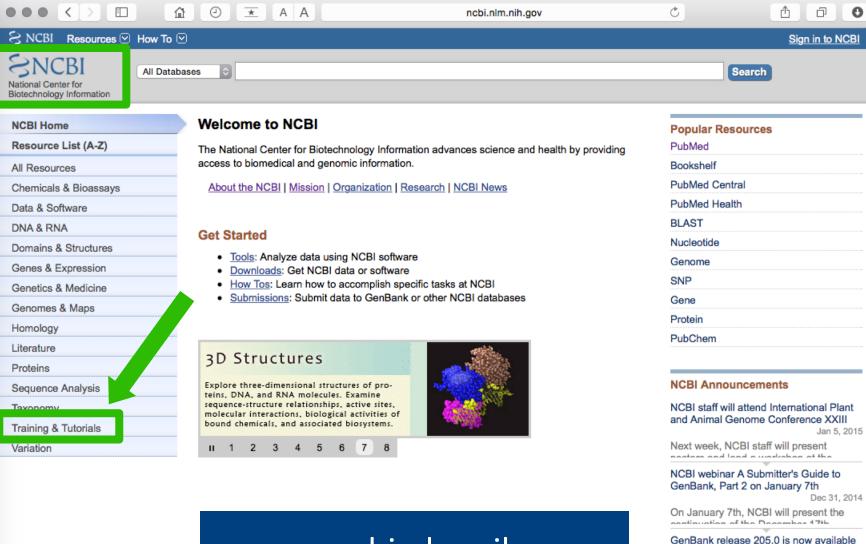
In addition to maintaining the GenBank® nucleic The National Center for Biotechnology Information NCBI Resources 
 How To 
 Sign in to NCBI (NCBI) at the National Institutes of Health was created in 1988 to develop information systems for molecular biology. In addition to maintaining the GenBank® (1) PubMed nucleic acid sequence database, which receives data Advanced through the international collaboration with DDBJ and EMBL-Bank as well as from the scientific community, NCBI provides data retrieval systems and computational resources for the analysis of GenBank data and many PubMed Commons other kinds of biological data. For the purposes of this article, after a summary of recent developments and an introduction to the Entrez system, the NCBI suite of PubMed comprises more than 24 million citations for resources is grouped into 10 broad categories based on biomedical literature from MEDLINE, life science journals, those in the NCBI guide. All resources discussed are availand online books. Citations may include links to full-text able from the NCBI guide at www.ncbi.nlm.nih.gov and Regulating ribosome recruitment? I Shatsky content from PubMed Central and publisher web sites. Protein Clusters and the PubChem suite of small can also be located using the Site Search database. In most critiques proposed RNA regulon mechanism. 1.usa.gov/1zfZekD

<sup>\*</sup>To whom correspondence should be addressed. Tel: 301 496 2475; Fax: 301 480 9241; Email: savers@ncbi.nlm.nih.gov

Published by Oxford University Press 2011.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecomm by-nc/3.0), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

## How do I learn to search the NCBI Genetics Databases???



www.ncbi.nlm.nih.gov

More...

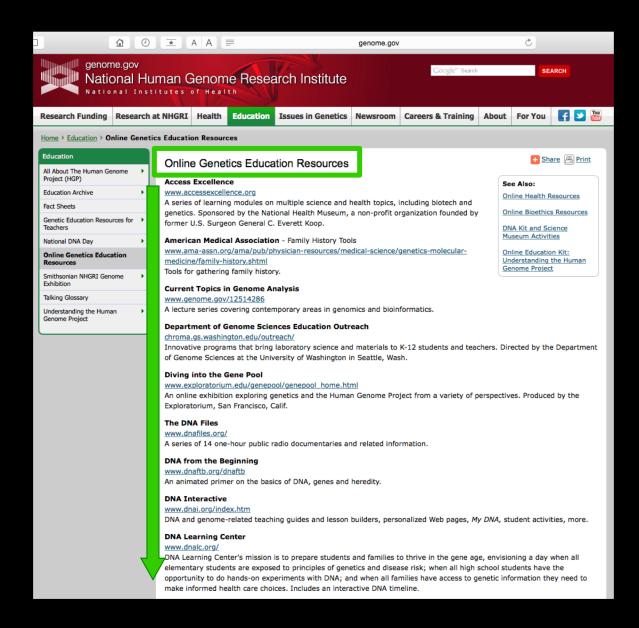
Dec 16, 2014

via FTP

Release 205.0 (12/12/2014) has

# What are some websites where I can learn more about genetics?

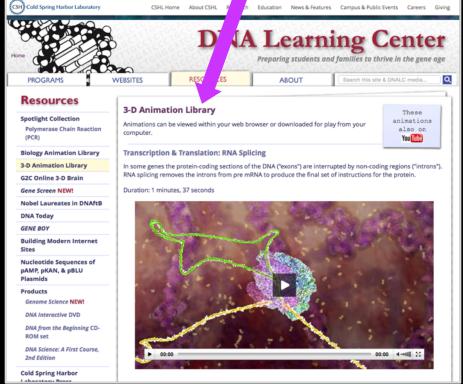
### www.genome.gov



### www.dnalc.org



### includes



### The Future of Medicine



diagnosis and treatment based on your genome!

### We live in the age of information.





http://www.research.archives.gov



http://www.iipdigital.usembassy.gov



http://www.in.go



http://www.nihrecord.nih.gov



http://www.directorsblog.nih.gov



http://www.blogs.america.gov



http://www.nichd.nih.gov



http://www.ncbi.nlm.nih.gov



### The End