

Data From the National Library of Medicine

UNT Open Access Symposium 2017

Daniel E. Burgard

University of North Texas Health Science Center at Fort Worth

Regional Medical Library, South Central Region, National Library of Medicine

May 19, 2017

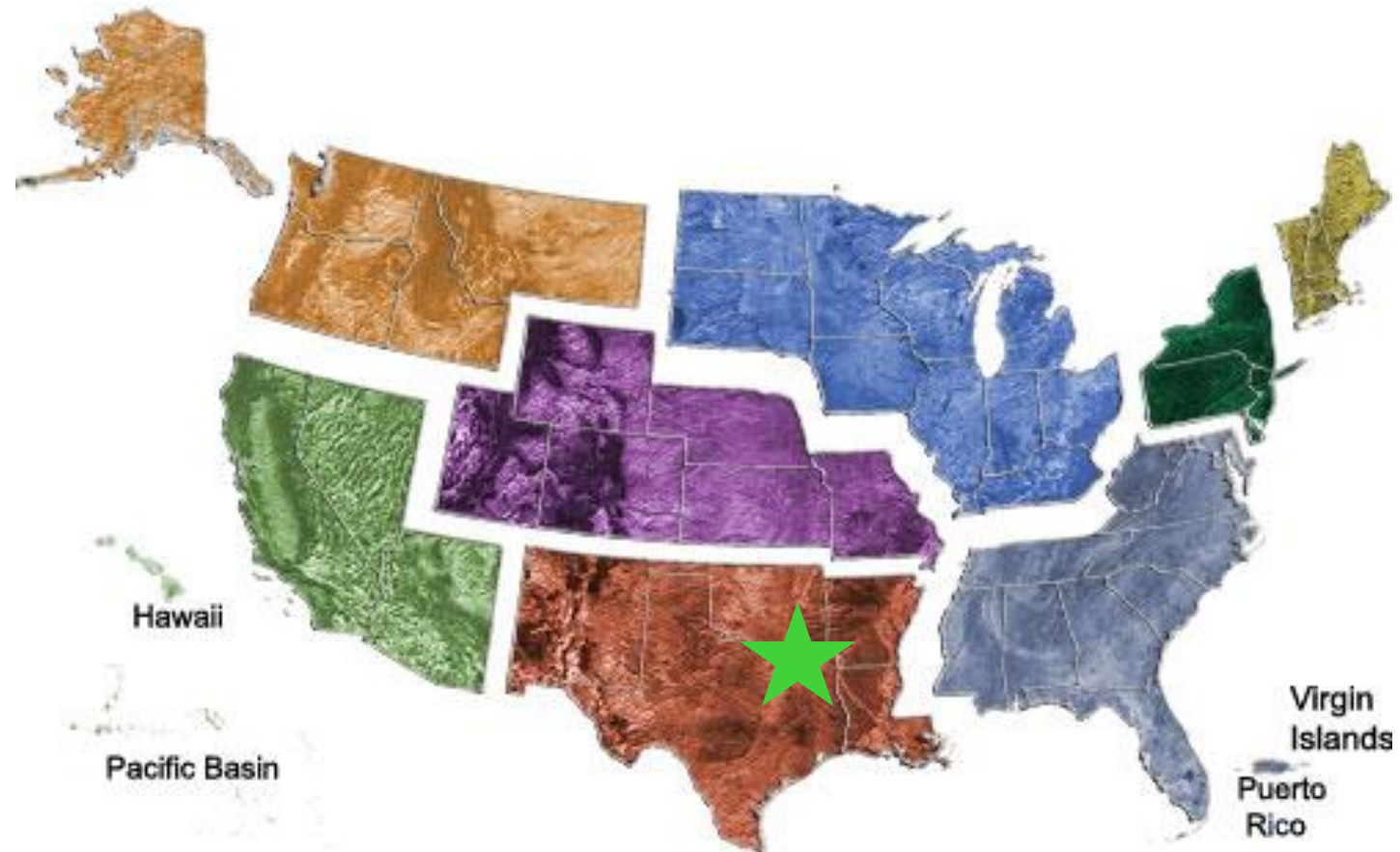


- “Serving as a leading global resource for building, curating and providing sophisticated access to molecular biology and genomic information, including those from the [Human Genome Project](#) and the [NIH Common Fund](#);”
- “Creating high-quality information services relevant to toxicology and environmental health, health services research, and public health;”

NLM Regional Medical Libraries 2016 - 2021

Host Libraries:

Univ of Massachusetts
Univ of Pittsburgh
Univ of Maryland
Univ Iowa
Univ of Utah
UNTHSC
Univ of Washington
UCLA



Data Types

- Bibliographic
- Toxicological
- Chemical
- Health
- Genetic

Audiences

- Basic scientists / researchers
- Healthcare professionals
- General public
- All ages
- Many languages

- Data and systems produced by NLM & NIH employees
- Data produced by NIH awardees

From Bench to Bedside





US National Library of Medicine
National Institutes of Health

PMC

Search

[Journal List](#) [Advanced](#)

[Help](#)



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).

PubReader



A whole new way to read scientific literature at PubMed Central

Get Started

[PMC Overview](#)

[Users' Guide](#)

[Journal List](#)

[PMC FAQs](#)

[PMC Copyright Notice](#)

Participate

[Information for Publishers](#)

[Add a Journal to PMC](#)

[Participation Agreements](#)

[File Submission Specifications](#)

[File Validation Tools](#)

Keep Up to Date

[New in PMC](#) | [RSS](#)

[PMC Announce Mail List](#)

[Utilities Announce Mail List](#)

[Tagging Guidelines Mail List](#)

Other Resources

[PMC International](#)

[Text Mining Collections](#)

[Developer Resources](#)

[NLM LitArch](#)

4.3 MILLION Articles

are archived in PMC.

Content provided in part by:

2007

326

4247

Public Access

[Funders and PMC](#)

[How Papers Get Into PMC](#)

[NIH Manuscript Submission System](#)

[My Bibliography](#)



Search MedlinePlus

GO

[About MedlinePlus](#) [Site Map](#) [FAQs](#) [Customer Support](#)

[Health Topics](#)

[Drugs & Supplements](#)

[Videos & Tools](#)

[Español](#)



Health Topics

Find information on health, wellness, disorders and conditions



Drugs & Supplements

Learn about prescription drugs, over-the-counter medicines, herbs, and supplements



Videos & Tools

Discover tutorials, health and surgery videos, games, and quizzes



Medical Encyclopedia

Articles and images for diseases, symptoms, tests

Share MedlinePlus



How can you get a good night's sleep?

Find out with our new topic: [Healthy Sleep](#).



Today's Health News

[Could a Weight-Loss Surgery Lead to Alcohol Abuse?](#)

[Is the AIDS Epidemic Winding Down in The U.S.?](#)

[It's Often Family to The Rescue During](#)

Stay Connected

Sign up for MedlinePlus email updates

Enter email address

GO





[About MedlinePlus](#) [Site Map](#) [FAQs](#) [Customer Support](#)
[Health Topics](#)
[Drugs & Supplements](#)
[Videos & Tools](#)
[Español](#)
[Home](#) → [Health Topics](#) → [Genes and Gene Therapy](#)

Genes and Gene Therapy



On this page

Basics

- [Summary](#)
- [Start Here](#)
- [Latest News](#)

Learn More

- [Related Issues](#)
- [Specifics](#)

See, Play and Learn

- [Games](#)

Research

- [Statistics and Research](#)
- [Clinical Trials](#)
- [Journal Articles](#)

Resources

- [Reference Desk](#)
- [Find an Expert](#)

For You

- [Children](#)
- [Teenagers](#)



Summary

Genes are the building blocks of inheritance. Passed from parent to child, they contain instructions for making proteins. If genes don't produce the right proteins or don't produce them correctly, a child can have a [genetic disorder](#).

Get Genes and Gene Therapy updates by email

Tox Town

ENVIRONMENTAL HEALTH CONCERNS AND TOXIC
CHEMICALS WHERE YOU LIVE, WORK, AND PLAY

[Home](#)[Text Version](#)[Neighborhoods](#)[Locations](#)[Chemicals](#)[For Educators](#)[Español](#)

City

View



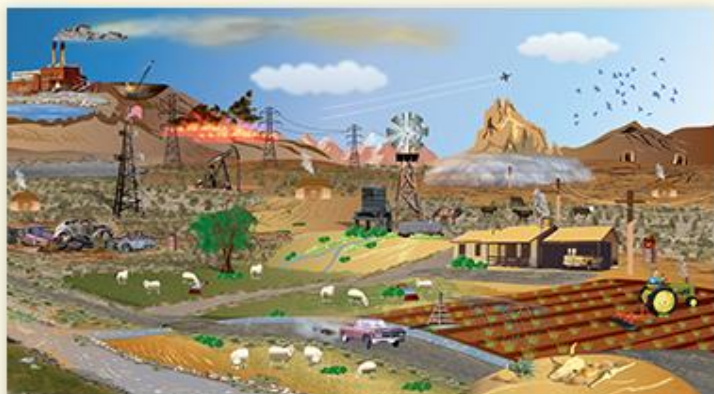
Farm

View



Port

View



What's New

- **May is Asthma and Allergy Awareness Month.**
 - Visit the **Indoor Air**, **Outdoor Air**, and **Homes** location pages.
- Click on **Drinking Water** in honor of **National Drinking Water Week, May 7 - 13, 2017.**
- Show how you care about the air during **Air**

- NCBI Home
- Resource List (A-Z)
- All Resources
- Chemicals & Bioassays
- Data & Software
- DNA & RNA
- Domains & Structures
- Genes & Expression
- Genetics & Medicine
- Genomes & Maps
- Homology
- Literature
- Proteins
- Sequence Analysis
- Taxonomy
- Training & Tutorials
- Variation

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

[About the NCBI](#) | [Mission](#) | [Organization](#) | [Blog](#)

Submit

Deposit data or manuscripts into NCBI databases



Download

Transfer NCBI data to your computer



Learn

Find help documents, attend a class or watch a tutorial



Develop

Use NCBI APIs and code libraries to build applications



Analyze

Identify an NCBI tool for your data analysis task



Research

Explore NCBI research and collaborative projects



Popular Resources

- PubMed
- Bookshelf
- PubMed Central
- PubMed Health
- BLAST
- Nucleotide
- Genome
- SNP
- Gene
- Protein
- PubChem

NCBI News & Blog

Retiring and replacing the BLink protein similarity service
17 May 2017

NCBI is discontinuing the BLink protein similarity service effective immediately

QuickBLASTP adds pre-processing to BLAST search
17 May 2017

QuickBLASTP, an accelerated version of BLASTP, adds a new pre-processing step

Search NCBI databases

[Help](#)

Search

Literature

Books	books and reports
MeSH	ontology used for PubMed indexing
NLM Catalog	books, journals and more in the NLM Collections
PubMed	scientific & medical abstracts/citations
PubMed Central	full-text journal articles

Health

ClinVar	human variations of clinical significance
dbGaP	genotype/phenotype interaction studies
GTR	genetic testing registry
MedGen	medical genetics literature and links
OMIM	online mendelian inheritance in man
PubMed Health	clinical effectiveness, disease and drug reports

Genomes

Assembly	genome assembly information
BioProject	biological projects providing data to NCBI
BioSample	descriptions of biological source materials
Clone	genomic and cDNA clones
dbVar	genome structural variation studies
Genome	genome sequencing projects by organism
GSS	genome survey sequences
Nucleotide	DNA and RNA sequences
Probe	sequence-based probes and primers
SNP	short genetic variations
SRA	high-throughput DNA and RNA sequence read archive
Taxonomy	taxonomic classification and nomenclature catalog

Genes

EST	expressed sequence tag sequences
Gene	collected information about gene loci
GEO DataSets	functional genomics studies
GEO Profiles	gene expression and molecular abundance profiles
HomoloGene	homologous gene sets for selected organisms
PopSet	sequence sets from phylogenetic and population studies
UniGene	clusters of expressed transcripts

Proteins

Conserved Domains	conserved protein domains
Protein	protein sequences
Protein Clusters	sequence similarity-based protein clusters
Structure	experimentally-determined biomolecular structures

Chemicals

BioSystems	molecular pathways with links to genes, proteins and chemicals
PubChem BioAssay	bioactivity screening studies
PubChem Compound	chemical information with structures, information and links
PubChem Substance	deposited substance and chemical information

Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. [Learn more](#)

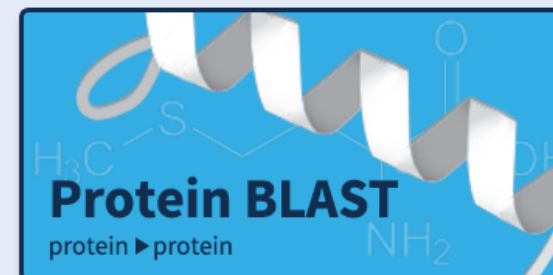
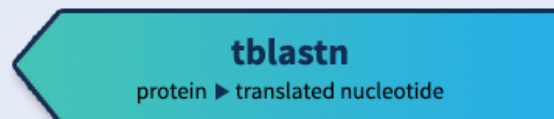
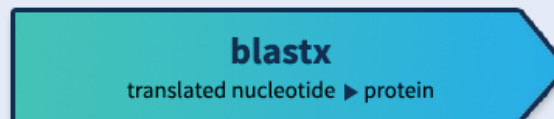
NEWS

Magic-BLAST 1.2.0 released

A new version of the BLAST RNA-seq mapping tool is now available.
Mon, 27 Feb 2017 14:00:00 EST

[More BLAST news...](#)

Web BLAST



BLAST Genomes

Search[Human](#)[Mouse](#)[Rat](#)[Microbes](#)

Standalone and API BLAST

**Download BLAST**

Get BLAST databases and executables

**Use BLAST API**

Call BLAST from your application

**Use BLAST in the cloud**

Start an instance at a cloud provider



Genome data download made easy!



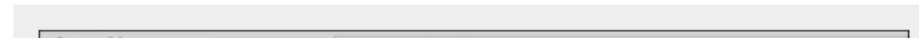
Posted on [May 8, 2017](#)

★★★★★ ⓘ 1 Vote

This blog post is directed toward Assembly users.

A new "Download assemblies" button is now available in the [Assembly](#) database. This makes it easy to download data for multiple genomes without having to write scripts.

For example, you can run a search in Assembly and use check boxes (see left side of screenshot below) to refine the set of genome assemblies of interest. Then, just open the "Download assemblies" menu, choose the source database ([GenBank](#) or [RefSeq](#)), choose the file type, and start the download. An archive file will be saved to your computer that can be expanded into a folder containing your selected genome data files.



- [August 2014](#)
- [June 2014](#)
- [May 2014](#)
- [April 2014](#)
- [March 2014](#)
- [January 2014](#)
- [December 2013](#)
- [November 2013](#)
- [October 2013](#)
- [September 2013](#)
- [July 2013](#)
- [June 2013](#)
- [May 2013](#)
- [April 2013](#)
- [March 2013](#)
- [February 2013](#)
- [January 2013](#)

NIH Data

Pi Day @ NIH - Celebrating the Intersection Between the Mathematical & Biomedical Sciences

LEARN MORE



PI DAY @NIH

Celebrating the Intersection Between the Mathematical & Biomedical Sciences



DataScience@NIH

The NIH is dedicated to harnessing the potential of the computational and quantitative sciences to elevate the impact and efficiency of biomedical research. NIH efforts in the integration of Data Science with the biomedical sciences are coordinated by the NIH Scientific Data Council and DataScience@NIH, under the

News

- Open Science Prize announces nextstrain.org as Grand Prize Winner

Congratulations to the nextstrain.org development team led by Trevor Bedford, PhD, of the... [read more](#)



NIH Data Sharing Repositories

This table lists NIH-supported data repositories that make data accessible for reuse. Most accept submissions of appropriate data from NIH-funded investigators (and others), but some restrict data submission to only those researchers involved in a specific research network. Also included are resources that aggregate information about biomedical data and information sharing systems. The table can be sorted according to name and by NIH Institute or Center and may be searched using keywords so that you can find repositories more relevant to your data. Links are provided to information about submitting data to and accessing data from the listed repositories. Additional information about the repositories and points-of-contact for further information or inquiries can be found on the websites of the individual repositories. Are we missing a data sharing repository? [Contact us](#).

Show entriesSearch:

IC	Repository Name	Repository Description	Data Submission Policy	Access to Data
NCI	Cancer Nanotechnology Laboratory (caNanoLab)	caNanoLab is a data sharing portal designed to facilitate information sharing in the biomedical nanotechnology research community to expedite and validate the use of nanotechnology in biomedicine. caNanoLab provides support for the annotation of nanomaterials with characterizations resulting from physico-chemical, in vitro, and in vivo assays and the sharing of these characterizations and associated nanotechnology protocols in a secure fashion.	How to submit your data to caNanoLab	How to access caNanoLab data
NCI	The Cancer Imaging Archive (TCIA)	The image data in The Cancer Imaging Archive (TCIA) is organized into purpose-built collections of subjects. The subjects typically have a cancer type and/or anatomical site (lung, brain, etc.) in common.	How to submit data to TCIA	How to access TCIA data
NCI (NHGRI, NIGMS)	PeptideAtlas	PeptideAtlas is a multi-organism, publicly accessible compendium of peptides identified in a large set of tandem mass spectrometry proteomics experiments. Mass spectrometer output files are collected for human, mouse, yeast, and several other organisms, and searched using the latest search engines and protein sequences.	How to submit data to PeptideAtlas	How to access Peptide Atlas data
NEI	EyeGENE®	The eyeGENE® Biorepository and corresponding Database contain family history and clinical eye exam data from subjects enrolled in eyeGENE® Program coupled to clinical grade DNA samples. This data and samples are submitted by collaborators throughout the US and Canada and the data is available on a controlled access basis to researchers world-wide.	How to submit data to EyeGENE®	How to access EyeGENE® data
NHGRI	FlyBase: A Drosophila Genomic and Genetic Database	Drosophila Genomic and Genetic database that includes proteomics data, microarrays and Tiling BAC's.	How to submit data to Flybase	How to access Flybase data
NHGRI	The Zebrafish Model Organism Database (ZFIN)	ZFIN serves as the zebrafish model organism database. It aims to: a) be the community database resource for the laboratory use of zebrafish, b) develop and support integrated zebrafish genetic, genomic and developmental information, c) maintain the definitive reference data sets of zebrafish research information, d) to link this information extensively to corresponding data in other model organism and human databases, e)	How to submit data to ZFIN	How to access ZFIN data



NIH Data Sharing Policies

This table lists data sharing policies in effect at NIH. It includes policies at the NIH, IC, division, and program levels that apply to broad sets of investigators and data. Individual requests for applications (RFAs) and program announcements (PA) may specify other requirements or expectations for data sharing that apply to specific projects.

IC	Data Sharing Policy Name	Description of Data Sharing Policy	Repositories
NIH	NIH Data Sharing Policy	Expects investigators seeking more than \$500K in direct support in any given year to submit a data sharing plan with their application or to indicate why data sharing is not possible.	No specific repository listed
NIH	NIH Policy on Deposit of Atomic Coordinates into Structural Databases	NIH policy requires that atomic coordinates from X-ray crystallographic and nuclear magnetic resonance experiments that were supported by NIH grants be deposited into the appropriate structural database at the time of submission of a research article drawing conclusions from these data.	Protein Data Bank
NHGRI	ENCODE Consortia Data Release, Data Use, and Publication Policies	Requires resource producers to release primary data along with an initial interpretation, in the form of genome features, to the appropriate public databases as soon as the data is verified. Consortia members will also identify validation standards that will be applied in subsequent analyses of the data or with additional experimentation where appropriate. All data will be deposited to public databases, such as GenBank or the ENCODE/modENCODE Data Coordination Centers (DCCs) and these pre-publication data will be available for all to use.	ENCODE
NIH	Genomic Data Sharing Policy	Expects that large-scale genomic research data from NIH-funded studies involving human specimens, as well as non-human and model organisms, will be shared through a publicly available data repository. All studies with human genomic data should be registered in dbGaP , and the data should be submitted to an NIH-designated data repository . Non-human data may be submitted to any widely used data repository.	dbGaP (for registration) NIH-designated data repository (for data)
NIH	NIH Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS)	Expects all investigators who receive NIH support to conduct genome-wide analysis of genetic variation in a study population to submit to the NIH GWAS data repository descriptive information about their studies for inclusion in an open access portion of the NIH GWAS data repository. Strongly encourages the submission of curated and coded phenotype, exposure, genotype, and pedigree data, as appropriate, to the NIH GWAS data repository as soon as quality control procedures have been completed at the local institution. These detailed data will be made available through a controlled access process according to the GWAS Data Access procedures. *This policy is superseded by the Genomic Data Sharing Policy, which will apply to grant applications due on or after January 25, 2015, or for contracts submitted on or after January 25, 2015, and for intramural research projects generating genomic data on or after January 25, 2015.	dbGaP
NHLBI	NHLBI Policy for Data Sharing from Clinical Trials and Epidemiological Studies	Encourages all applicants to include a plan to address data sharing or to state why data sharing is not possible. For studies that meet the the following criteria, applicants are required to provide a data sharing plan, which will be reviewed and approved by the relevant NHLBI program official: a) research applications/proposals requesting \$500000 direct costs; b) research studies that have 500 or more participants c) ancillary studies based on NHLBI-funded parent studies d) applications/proposals submitted in response to FOAs that specify inclusion of data sharing plans; or e) other research studies deemed appropriate for data sharing by NHLBI program official investigators.	NHLBI data repository , BioLINCC
NIA	Alzheimer's Disease Genetics Sharing Plan	NIA policy in the area of human Alzheimer's disease genetics applies to all NIA funded research in this area regardless of cost. NIA follows the NIH GWAS Policy and extends NIA's existing policy on sharing data on Alzheimer's disease genetics to include secondary analysis of data resulting from a genome wide association study. It is the policy of the NIA that useful specimens and Associated Phenotypic Data for the genetics of late onset Alzheimer's disease be deposited at the National Cell Repository for Alzheimer's Disease (NCRAD) whenever possible. It is the policy of the NIA that all Genetic Data derived from NIA funded studies for the genetics of late onset Alzheimer's disease be deposited at the National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site (NIAGADS) or another NIA approved site or both whenever possible. It is the policy of the NIA that all GWAS data, including secondary analysis data, derived from NIA funded studies for the genetics of late onset Alzheimer's disease be deposited at the NIH GWAS data repository (dbGaP) or another NIA approved site or both, wherever possible.	NCRAD , NIAGADS , dbGaP

Future of NLM/NIH Data & Access

- Federally funded research data sharing
 - Legislative activity
 - Mandated deposit
 - NLM/NIH managed repositories
- Issues

Data From the NLM

- Daniel Burgard
 - daniel.burgard@unthsc.edu
 - 817-735-2589
- Bench to bedside
- URL's from presentation
 - <https://www.nlm.nih.gov/>
 - <https://nnlm.gov/scr>
 - <https://www.ncbi.nlm.nih.gov/gquery/>
 - https://www.nlm.nih.gov/databases/download/data_distrib_main.html
 - <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
 - <https://ncbiinsights.ncbi.nlm.nih.gov/>
 - <https://datascience.nih.gov/index>
 - https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html
 - https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_policies.html