

THE DATA CITATION INDEX^S

NOW ON

THOMSON REUTERS
WEB OF KNOWLEDGESM

NIGEL ROBINSON
DIRECTOR, YORK OPERATIONS
7 DECEMBER 2012

INTERNATIONAL ASSOCIATION OF SCIENTIFIC, TECHNICAL & MEDICAL PUBLISHERS
The voice of research publishing since 1969

stm

DIGITAL UNIVERSE EXPANSION

1.8
ZETTABYTES

(1.8 trillion gigabytes) of information
will be created and replicated in 2011 -
and growing fast (it has grown by a factor of 9 in just five years)

<http://www.emc.com/leadership/programs/digital-universe.htm>

BENEFITS OF RESEARCH DATA SHARING



By tk-link

- Advancement of scholarship
- Verification of results
- Promotion of scholar's work

INFLUENCE OF FUNDING AGENCIES

2010 Mandate from the US National Science Foundation –

All funding proposals submitted or due on or after January 18, 2011, must include a “Data Management Plan” describing how the proposal will conform to NSF policy on the dissemination and sharing of research results.



IMPACT ON RESEARCH LIBRARIES

The image shows a composite of two web pages. The top page is titled "Data Management at Harvard" and features a navigation menu with items like "Overview", "Feedback", "Harvard Guidelines and Policies", "Funding Agencies Guidelines and Policies", "Publishing and Storing Your Data", "Security", and "Best Practices". The main content area is titled "Funding Agencies Guidelines and Policies" and specifically addresses the "National Science Foundation".

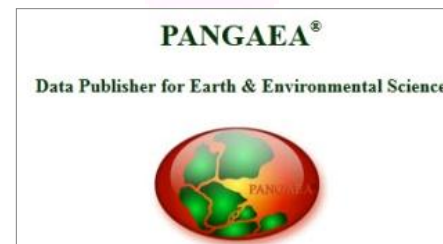
The bottom page is a screenshot of the Association of Research Libraries (ARL) website. The header includes the ARL logo and the text "ASSOCIATION OF RESEARCH LIBRARIES" with the URL "www.arl.org". The main article is titled "Transforming Research Libraries" and "E-Science", with a sub-heading "Guide for Research Libraries: The NSF Data Sharing Policy". The authors are listed as Patricia Hswe and Ann Holt. The article text states: "In Spring 2010, the National Science Foundation (NSF) announced that it would alter its data sharing policy to require data management plans (DMPs) in future grant proposals to the agency. The Association for Research Libraries has developed...". A sidebar on the left contains a "Planning and Visioning" section with links to "Resources for Planning and Visioning" and "Envisioning Research Library Futures: A Scenario Thinking Project". A "More on This Topic" section lists "Unpacking the NSF Requirement" and "A New Leadership Role for Libraries".

VISIBILITY OF RESEARCH DATA

Grant funding agencies

Journal publishers

Data repositories & registration agencies



WHERE DO WE START?

Discover data repositories, data studies & data sets in the context of traditional literature

Help researchers find data sets and studies and track the full impact of their research output

Data Citation IndexSM : Connecting Data to the Research it Informs

Search Results within the Data Citation Index present all of the powerful Web of Knowledge options for exploring a body of information.

The screenshot displays the Data Citation Index search results for the query "Murder Cases". The interface includes a navigation bar with "All Databases", "Select a Database", "Web of Science", and "Additional Resources". A search bar at the top left shows the query and search history. The results section shows "Results Title=(Murder Cases) Timespan=All Years" with a "Scientific WebPlus" logo and a "View Web Results >>" link. A note indicates that alternative forms of the search term may have been applied. The results are sorted by "Publication Date - newest to oldest" and show approximately 331,559 results. A "Refine Results" sidebar on the left allows filtering by Databases (WOS, BCI, CSCD), General Categories (Science & Technology, Social Sciences, Art & Humanities), Subject Areas, Document Types, Authors, Group Authors, Editors, Source Titles, Book Series, Conference Titles, Publication Years, Institutions, Funding Agencies, Languages, and Countries/Territories. The main results list includes:

- 1. Title: **Murder Cases in 33 Large Urban Counties in the United States, 1988.**
Author(s): United States Department of Justice; Bureau of Justice Statistics
Source: Inter-university Consortium for Political and Social Research DOI: <http://dx.doi.org/10.3886/ICPSR09907>
Times Cited: 0 (from All Databases)
[View abstract]
- 2. Title: **Femicide-Suicide A Cluster of 3 Cases in Los Angeles County**
Author(s): Scheinin, Lisa
Source: AMERICAN JOURNAL OF FORENSIC MEDICINE AND PATHOLOGY Volume: 366 Issue: 1584, Sp. Iss. SI Pages: 3545-3553 DOI: [10.1098/rstb.2011.0073](http://dx.doi.org/10.1098/rstb.2011.0073) Published: DEC 27 2011
Times Cited: 3 (from All Databases)
[Links] [Full Text] [View abstract]
- 3. Title: **White matter integrity and math performance in pediatric multiple sclerosis: a diffusion tensor imaging study**
Author(s): Till Christine; Deotto Angela; Tipu Vicentiu, et al.
Source: Neuroreport Volume: 22 Issue: 18 Pages: 1005-1009 DOI: [10.1097/WNR.0b013e32834dc301](http://dx.doi.org/10.1097/WNR.0b013e32834dc301) Published: DEC 21 2011
Times Cited: 1 (from All Databases)
[Links] [Full Text] [View abstract]
- 4. Title: **Generation of stable reporter breast cancer cell lines for the identification of ER subtype selective ligands**
Author(s): Shanle Erin K.; Hawse John R.; Xu Wei
Source: Biochemical Pharmacology Volume: 82 Issue: 12 Pages: 1940-1949 DOI: [10.1016/j.bcp.2011.08.026](http://dx.doi.org/10.1016/j.bcp.2011.08.026) Published: DEC 15 2011
Times Cited: 1 (from All Databases)
[Links] [Full Text] [View abstract]
- 5. Title: **Effect of using a new cyanoacrylate adhesive on the shear bond strength of orthodontic brackets.**
Author(s): Bishara S E.; VonWald L.; Laffoon J F.; et al.
Source: The Angle orthodontist Volume: 71 Issue: 6 Pages: 466-9 DOI: [10.1016/j.cnsns.2010.01.018](http://dx.doi.org/10.1016/j.cnsns.2010.01.018) Published: 2001-Dec
Times Cited: 1 (from All Databases)
[Links] [Full Text] [View abstract]
- 6. Title: **Monitoring animal diversity using acoustic indices: Implementation in a temperate woodland.**
Source: Ecological Indicators Volume: 13 Issue: 1 Pages: 46-54 DOI: [10.1016/j.cnsns.2010.01.030](http://dx.doi.org/10.1016/j.cnsns.2010.01.030) Published: February 2012
Times Cited: 6 (from All Databases)
[Links] [Full Text] [View abstract]

Data Citation IndexSM : Connecting Data to the Research it Informs



REUTERS/Aly Song

- ~80 repositories
- Comprising >2 million records
- Reciprocal links to/from DCI and WoS records

TYPES OF DATA BY DISCIPLINE

ART & HUMANITIES

CULTURAL
HERITAGE

LANGUAGE CORPUS

IMAGE
COLLECTIONS

RECORDINGS

SOCIAL SCIENCES

POLL DATA

ECONOMIC
STATISTICS

LONGITUDINAL DATA

NATIONAL CENSUS

PUBLIC OPINION
SURVEYS

SCIENCE & TECHNOLOGY

MAPS

ALGORITHMS

GENOMICS

SKY SURVEYS

ASTROPHYSICS

REMOTE SENSING

MUSEUM SPECIMENS

REPOSITORY SELECTION & EVALUATION



REUTERS/ Moham Rizza

REPOSITORY EVALUATION, SELECTION, AND COVERAGE POLICIES

FOR THE DATA CITATION INDEXSM WITHIN THOMSON REUTERS WEB OF KNOWLEDGESM

- Editorial Content
- Persistence and stability
- Thoroughness of descriptive information.
- Links from data to research literature.

Selection essay:

http://wokinfo.com/products_tools/multidisciplinary/dci/selection_essay/

CHALLENGES

Data availability

Data quality



INDEXING OF RESEARCH DATA REPOSITORIES

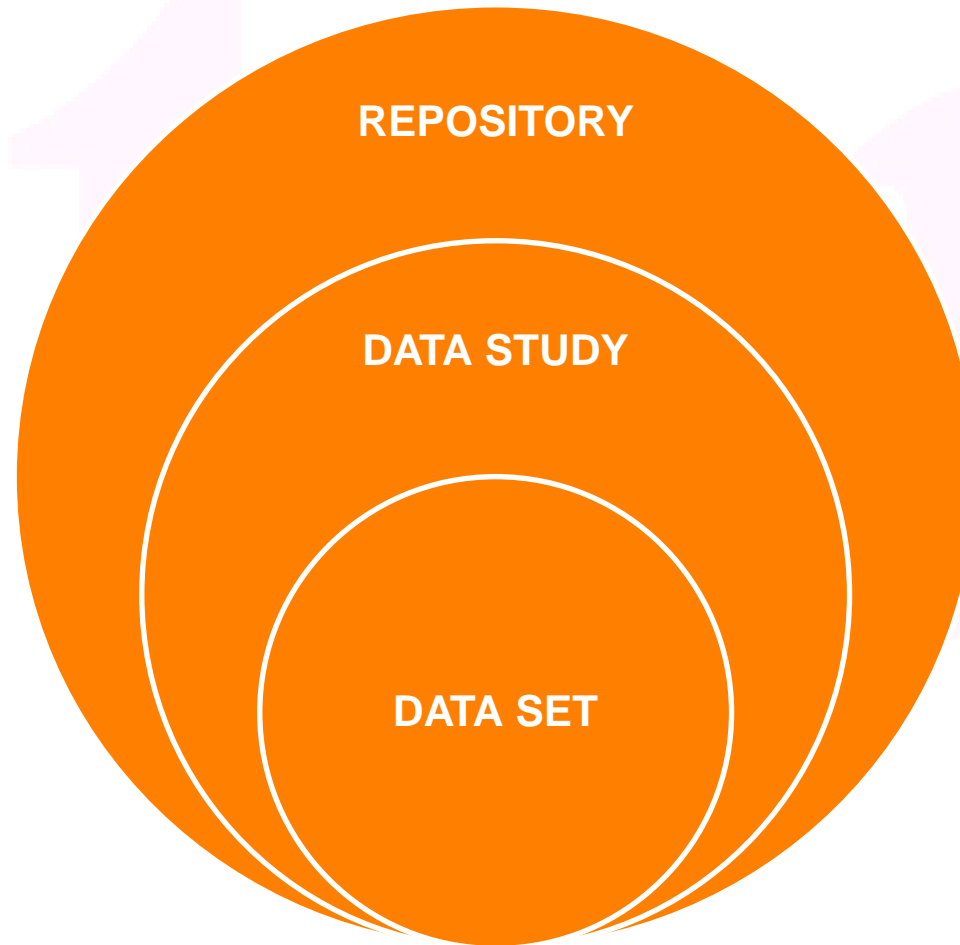
TR takes
descriptive
metadata feed
from repository

Repository
raw metadata
is analyzed by
TR

TR adds
metadata

TR DCI records:
- data repository
- data study
- data set

DATA CITATION INDEX DOCUMENT TYPES



U.S. Dept. of Justice, Bureau of Justice Statistics (1996): MURDER CASES IN 33 LARGE URBAN COUNTIES IN THE UNITED STATES, 1988. Version 1. Inter-university Consortium for Political and Social Research [distributor].

<http://dx.doi.org/10.3886/ICPSR09907.v1>

Our first data set is the Bureau of Justice Statistics "Murder Cases in 33 Large Urban Counties." This is a random sample of homicide cases drawn from prosecutors' files. The data set includes information on offender characteristics, victim characteristics and trial outcomes for 2800 murders. The 75 largest counties account for more than half of the murders in the U.S. each year. This data set brings together information on the crime, the offender, the victim, and the sentence. Such information cannot all be linked in other larger data sets such as the Uniform Crime Reporting (UCR) Data or the National Crime Victimization Survey (NCVS). Most crime

Data Citation IndexSM : Connecting Data to the Research it Informs

Search Results within the Data Citation Index present all of the powerful Web of Knowledge options for exploring a body of information.

The screenshot displays the Data Citation Index web interface. At the top, there are navigation tabs: "All Databases", "Select a Database", "Data Citation Index", and "Additional Resources". Below these are search options: "Search", "Cited Reference Search", "Advanced Search", and "Search History".

The main section is titled "Data Citation IndexSM". It shows search results for the query "Topic=(child* health*)". The results are displayed on page 1 of 13. A "Refine Results" sidebar is visible on the left, with red arrows pointing to the "Data Types" and "Institutions" sections. The "Data Types" section includes options like "STRUCTURED QUESTIONNAIRE (94)", "SEMI STRUCTURED QUESTIONNAIRE", "RNA (5)", "SURVEY DATA (5)", and "EXPRESSION PROFILING BY ARRAY (3)". The "Institutions" section includes "UNIVERSITY OF JYVASKYLA DEPARTMENT OF PSYCHOLOGY (12)" and "UNIVERSITY OF TAMPERE TAMPERE SCHOOL OF PUBLIC HEALTH (7)".

The search results list includes:

- 1. Title: **School Well-being Profile 2010-2011: Personnel.**
Author(s): Konu, Anne
Source: Finnish Social Science Data Archive
Source URL: <http://www.fsd.uta.fi/english/data/catalogue/>
Document Type: Data study Times Cited: 0 (from All Databases)

A dropdown menu for sorting options is shown, with "Times Cited -- highest to lowest" selected. Other options include "Year Published -- newest to oldest", "Year Published -- oldest to newest", "Processing Date -- newest to oldest", "Processing Date -- oldest to newest", "Relevance", "First Author -- A to Z", "First Author -- Z to A", "Source -- A to Z", and "Source -- Z to A".

Below the first result, there is a "View abstract" link. The interface also shows a "Save to" section with options for "ENDNOTE WEB", "ENDNOTE", and "RefWorks".

The bottom part of the screenshot shows another search result for "Title=(Murder Cases)". The results are displayed on page 1 of 10,000. The search results list includes:

- 1. Title: **Murder Cases in 33 Large Urban Counties in the United States, 1988.**
Author(s): United States Department of Justice; Bureau of Justice Statistics
Source: Inter-university Consortium for Political and Social Research DOI: <http://dx.doi.org/10.3886/ICPSR09907>
Times Cited: 0 (from All Databases)
- 2. Title: **Familicide: Suicide A Cluster of 3 Cases in Los Angeles County**
Author(s): Scheinin, Lisa
Source: AMERICAN JOURNAL OF FORENSIC MEDICINE AND PATHOLOGY Volume: 366 Issue: 1584, Sp. Iss. SI Pages: 3545-3553 DOI: <http://dx.doi.org/10.1098/ijsb.2011.0073> Published: DEC 27 2011
Times Cited: 3 (from All Databases)
- 3. Title: **White matter integrity and math performance in pediatric multiple sclerosis: a diffusion tensor imaging study**
Author(s): Tii Christine, Deotto Angela, Tipu Vicentiu, et al.
Source: Neuroreport Volume: 22 Issue: 18 Pages: 1005-1009 DOI: <http://dx.doi.org/10.1097/WNR.0b013e328284dc301> Published: DEC 21 2011
Times Cited: 1 (from All Databases)
- 5. Title: **Effect of using a new cyanoacrylate adhesive on the shear bond strength of orthodontic brackets.**
Author(s): Bahara S E, VonWald L, Laffoon J F, et al.
Source: The Angle orthodontist Volume: 74 Issue: 6 Pages: 855-9 DOI: <http://dx.doi.org/10.1016/j.jao.2010.04.018> Published: 2011 Dec

Data Citation IndexSM : Connecting Data to the Research it Informs

Data Citation IndexSM

“Repository” Record – Link to the citing literature.

<< Back to results list

Record 3 of 6

Record from Data Citation IndexSM

Save to: [ENDNOTE[®] WEB](#) [ENDNOTE[®]](#) [RefWorks](#) [ResearcherID](#) more options

UniProt Knowledgebase.

Editor(s): Uniprot Consortium

Source: UniProt Knowledgebase

Source URL: <http://www.uniprot.org/>

Cited References: 0

Abstract: The UniProt Knowledgebase accurate, consistent and rich annotated amino acid sequence, protein name or possible is added. This includes widely of the quality of annotation in the form

Document Type: **Repository**

Accession Number: DRCI:DATA201

Language: English

Funding:

Funding Agency

National Institutes of Health

European Commission SLING

NIH GO

Swiss Federal Office of Education an

GEN2PHEN

MICROME

NIH

Full Text Full Text Links NCBI Save to: ENDNOTE[®] WEB ENDNOTE[®] RefWorks ResearcherID more options

From protein sequences to 3D-structures and beyond: the example of the UniProt Knowledgebase

Author(s): Hinz, U (Hinz, Ursula)¹

Group Author(s): UniProt Consortium

Source: CELLULAR AND MOLECULAR LIFE SCIENCES Volume: 67 Issue: 7 Pages: 1049-1064 DOI: 10.1007/s00018-009-0229-6 Published: APR 2010

Times Cited: 5 (from Web of Science)

Cited References: 106 [[view related records](#)] [Citation Map](#)

Abstract: With the dramatic increase in the volume of experimental results in every domain of life sciences, assembling pertinent data and combining information from different fields has become a challenge. Information is dispersed over numerous specialized databases and is presented in many different formats. Rapid access to experiment-based information about well-characterized proteins helps predict the function of uncharacterized proteins identified by large-scale sequencing. In this context, universal knowledgebases play essential roles in providing access to data from complementary types of experiments and serving as hubs with cross-references to many specialized databases. This review outlines how the value of experimental data is optimized by combining high-quality protein sequences with complementary experimental results, including information derived from protein 3D-structures, using as an example the UniProt knowledgebase (UniProtKB) and the tools and links provided on its website (<http://www.uniprot.org/>). It also evokes precautions that are necessary for successful predictions and extrapolations.

Accession Number: WOS:000275419800003

Document Type: Review

Language: English

Author Keywords: Data flood; Annotation; Swiss-Prot; Knowledgebase; UniProtKB; Proteomics; Structural genomics; Protein 3D-structure

KeyWords Plus: DEPENDENT K+ CHANNEL; X-RAY-STRUCTURE; STRUCTURAL GENOMICS; CRYSTAL-STRUCTURE; UNIPROT; SWISS-PROT; DISORDERED PROTEINS; INTERACTOME NETWORK; ATOMIC

222886-2

5R01GM080646-04

0284GM088646-0188

Times Cited: 11

[Create Citation Alert](#)

This article has been cited 11 times in Web of Knowledge.

Hinz, Ursula. From protein sequences to 3D-structures and beyond: the example of the UniProt Knowledgebase. CELLULAR AND MOLECULAR LIFE SCIENCES, APR 2010.

Lima, Tania. HAMAP: a database of completely sequenced microbial proteome sets and manually curated microbial protein families in UniProtKB/Swiss-Prot. NUCLEIC ACIDS RESEARCH, JAN 2009.

Suzek, Baris E. UniRef: comprehensive and non-redundant UniProt reference clusters. BIOINFORMATICS, MAY 15 2007.

[\[view all 11 citing articles \]](#)

Cited References: 0

Additional information

[How to cite this Resource](#)

Times Cited: 5

[Create Citation Alert](#)

This article has been cited 5 times in Web of Knowledge.

Moro, Monica. Identification of New Hematopoietic Cell Subsets with a Polyclonal Antibody Library Specific for Neglected Proteins. PLOS ONE, APR 4 2012.

MacDonald, Justin A. Intrinsically Disordered N-Terminus of Calponin Homology-Associated Smooth Muscle Protein (CHASM) Interacts with the Calponin Homology Domain to Enable Tropomyosin Binding. BIOCHEMISTRY, APR 3 2012.

Bombarely, Aureliano. The Sol Genomics Network (solgenomics.net): growing tomatoes using Perl. NUCLEIC ACIDS RESEARCH, JAN 2011.

[\[view all 5 citing articles \]](#)

Related Records:

Find similar Web of Knowledge records based on shared

“Repository” Records – information presented may vary depending upon the characteristics of the repository.

STM

Data Citation IndexSM : Connecting Data to the Research it Informs

Data Citation IndexSM

<< Back to results list

Record 1 of 5,338

Record from Data Citation IndexSM

Save to: [ENDNOTE[®] WEB](#) [ENDNOTE[®]](#) [RefWorks](#) [ResearcherID](#) [more options](#)

GSE12195: Mutations of multiple genes deregulate the NF- κ B pathway in diffuse large B cell lymphoma.

From Repository: Gene Expression Omnibus.

Author(s): Pasqualucci, Laura; Basso, Katia

Source: Gene Expression Omnibus **Source URL:** <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12195> (Viewed Date: 12 Dec 2011)

Published Year: 2009

Cited References: 0

Abstract: Diffuse large B-cell lymphoma (DLBCL), the most common form of lymphoma in adulthood, comprises distinct subtypes including germinal center B cell-like (GCB) and activated B cell like (ABC) DLBCL. Gene expression of its most aggressive subtype, ABC-DLBCL, is associated with constitutive activation of the NF- κ B transcription factor. The fraction of cases, it remains unclear whether NF- κ B activation in these tumors represents an intrinsic program of pathogenetic event. Here we show that >50% of ABC-DLBCL and a smaller fraction of GCB-DLBCL carry somatic mutations including negative (TNFAIP3/A20) and positive (CARD11, TRAF2, TRAF5, MAP3K7/TAK1 and TNFRSF11A/RAA1) A20 gene, which encodes for a ubiquitin-modifying enzyme involved in termination of NF- κ B responses, in the majority of the patients displaying biallelic inactivation by mutations and/or deletions, suggesting a tumor suppressor role for TRAF2 and CARD11 produce molecules with significantly enhanced ability to activate NF- κ B. Thus, our results suggest that ABC-DLBCL is caused by genetic lesions affecting multiple genes, whose loss or activation may promote lymphomagenesis. We show that most ABC-DLBCL and a smaller fraction of GCB-DLBCL display germline mutations in NF- κ B pathway genes, with A20 representing the most frequently mutated gene.

Document Type: [Data study](#)

Data Type: Expression profiling by array

Accession Number: DRCL:DATA2012007000275875

Language: English

Author Keywords: Phenotypic characterization of human DLBCL

Addresses:

1. Columbia University, Institute for Cancer Genetics, 1130 St Nicholas Ave, New York, 10032, USA
2. Columbia University, Institute for Cancer Genetics, 1150 St. Nicholas Avenue, New York, 10032, USA

E-mail Address: lp171@columbia.edu; kb451@columbia.edu

Web of Science Category: Biochemistry & Molecular Biology; Genetics & Heredity

Subject Area: Biochemistry & Molecular Biology; Genetics & Heredity

Taxonomic Data:

SUPER TAXA	TAXA NOTES	Organism Classifier	Organism Name
Animalia, Chordata, Vertebrata, Mammalia, Primates	Animals, Chordates, Humans, Mammals, Primates, Vertebrates	Hominidae	Homo sapiens

Miscellaneous: Transcription; missense mutation; Gene Expression Profiling; phenotype; Tumor; B-Cell lymphoma; Germinal Center; genomics; Molecular Genetics

Associated Records: [\[View All \]](#)

GSM476306: Lymphoblastoid B cell line_IARC 304.	Data set	Link to External Source
GSM476290: Follicular lymphoma_FLE.14.	Data set	Link to External Source
GSM476291: Follicular lymphoma_FLE.15.	Data set	Link to External Source
GSM476279: Follicular lymphoma_FLE.03.	Data set	Link to External Source

Times Cited: 4

[Create Citation Alert](#)

This article has been cited 4 times in Web of Knowledge.

Pasqualucci, Laura. Inactivating mutations of acetyltransferase genes in B-cell lymphoma. NATURE, MAR 10 2011.

How to cite this Resource - Windows Internet Explorer

Thomson Reuters recommend citing this resource as:
Pasqualucci, Laura; Basso, Katia (2009): GSE12195: Mutations of multiple genes deregulate the NF- κ B pathway in diffuse large B cell lymphoma. Gene Expression Omnibus. <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12195>

Cited References: 0

[How to cite this Resource](#)

[Suggest a correction](#)

If you would like to improve the quality of the data in this record, please [suggest a correction](#).

This Data Study record links to multiple Data Set records.

STM

Data Citation IndexSM : Connecting Data to the Research it Informs

Data Citation IndexSM

<< Back to previous page

Record 1 of 1

Record from Data Citation IndexSM

Save to: [ENDNOTE[®] WEB](#) [ENDNOTE[®]](#) [RefWorks](#) [ResearcherID](#) more options

GSM476306: Lymphoblastoid B cell line_IARC 304.

From Data study: GSE12195: Mutations of multiple genes deregulate the NF-kB pathway in diffuse large B cell lymphoma.

Author(s): Pasqualucci, Laura; Basso, Katia

Source: Gene Expression Omnibus Source URL: <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM476306> (Viewed Date: 12 Dec 2011)

Published Year: 2010

Cited References: 0

Abstract: Gene expression data from lymphoblastoid cell line.

Document Type: Data set

Data Type: RNA

Accession Number: DR0

Language: English

Addresses:

1. Columbia University, Ins
2. Columbia University, Ins

E-mail Address: lp171@

Web of Science Catego

Subject Area: Biochemis

Taxonomic Data:

SUPER TAXA

Animalia, Chordata, Vert
Primates

Method: Platform:GPL57

NCBI GEO Gene Expression Omnibus

HOME SEARCH SITE MAP GEO Publications FAQ MIAME Email GEO

NCBI > GEO > Accession Display

Not logged in | Login

Scope: Self Format: HTML Amount: Quick GEO accession: GSM476306 GO

Sample GSM476306 Query DataSets for GSM476306

Status	Public on Feb 01, 2010
Title	Lymphoblastoid B cell line_IARC 304
Sample type	RNA
Source name	EBV-immortalized lymphoblastoid cell line
Organism	<i>Homo sapiens</i>
Characteristics	cell type: EBV-immortalized lymphoblastoid B cell cell line: IARC 304
Treatment protocol	Cells were lysated in Trizol Reagent (Invitrogen) and frozen
Extracted molecule	total RNA
Extraction protocol	Total RNA was extracted using Trizol Reagent (Invitrogen) following the manufacturer's indications.
Label	biotin
Label protocol	Biotinylated cRNA were prepared according to the standard Affymetrix protocol from 6 ug total RNA (Expression Analysis Technical Manual, 2001, Affymetrix).
Hybridization protocol	Following fragmentation, 15 ug of cRNA were hybridized for at least 16 hr at 45C on GeneChip Human Genome Array U133Plus_2. GeneChips were washed and stained in the Affymetrix Fluidics Station 450.

Times Cited: 0

Create Citation Alert

This article has been cited 0 times in Web of Knowledge.

Cited References: 0

Additional information


How to cite this Resource

Suggest a correction

If you would like to improve the quality of the data in this record, please suggest a correction.

STM

Time's Up!



About your speaker:

- Name: Nigel Robinson
- Company: Thomson Reuters
- Tel: +44 207 433 4577
- Email:
nigel.robinson@thomsonreuters.com