Comparative Methods of CCSG Data Gathering: Biosketches, Publications, Grants

A look at the Past , Present, and Future of Data Collection

CCAF-IT 2017

Ben Busby, Mahendra Yatawara, & Susan Sharpe

NIH U.S. National Library of Medicine

CEDARS-SINAI

MOFFITT CANCER CENTER

NCBI

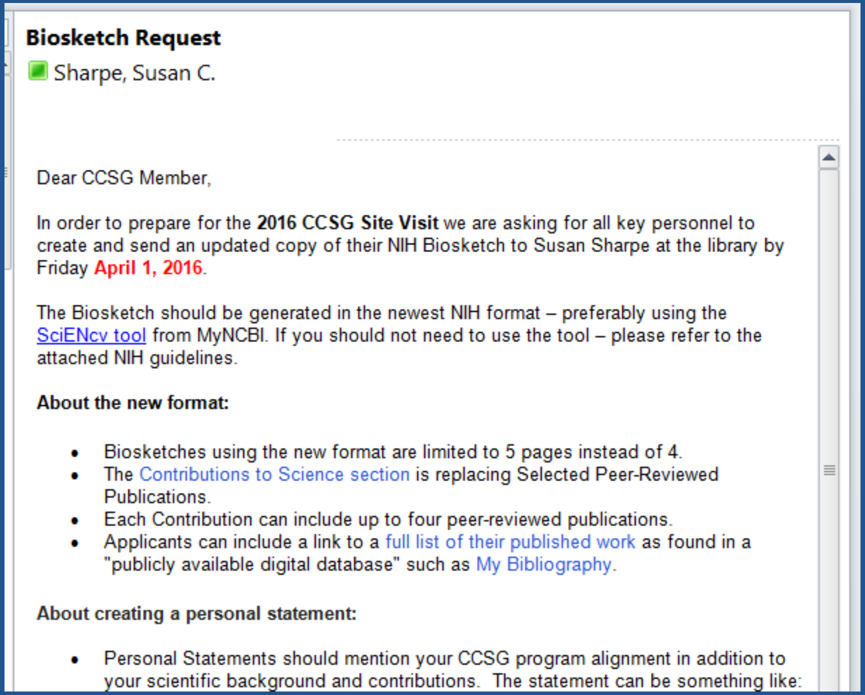
Biosketches:

sometimes member data collection is like herding cats…

Susan Sharpe, MA

Biosketches: The expectation

- Routine process, simple 5 page CV of relevant work & interests.
- SciENcv – latest and greatest template.
- Expected Gathering Process: Ask and receive.



Biosketch Request

🟩 Sharpe, Susan C.

Dear CCSG Member,

In order to prepare for the **2016 CCSG Site Visit** we are asking for all key personnel to create and send an updated copy of their NIH Biosketch to Susan Sharpe at the library by Friday **April 1, 2016**.

The Biosketch should be generated in the newest NIH format – preferably using the SciENcv tool from MyNCBI. If you should not need to use the tool – please refer to the attached NIH guidelines.

About the new format:

- Biosketches using the new format are limited to 5 pages instead of 4.
- The Contributions to Science section is replacing Selected Peer-Reviewed Publications.
- Each Contribution can include up to four peer-reviewed publications.
- Applicants can include a link to a full list of their published work as found in a "publicly available digital database" such as My Bibliography.

About creating a personal statement:

- Personal Statements should mention your CCSG program alignment in addition to your scientific background and contributions. The statement can be something like:

NIH U.S. National Library of Medicine

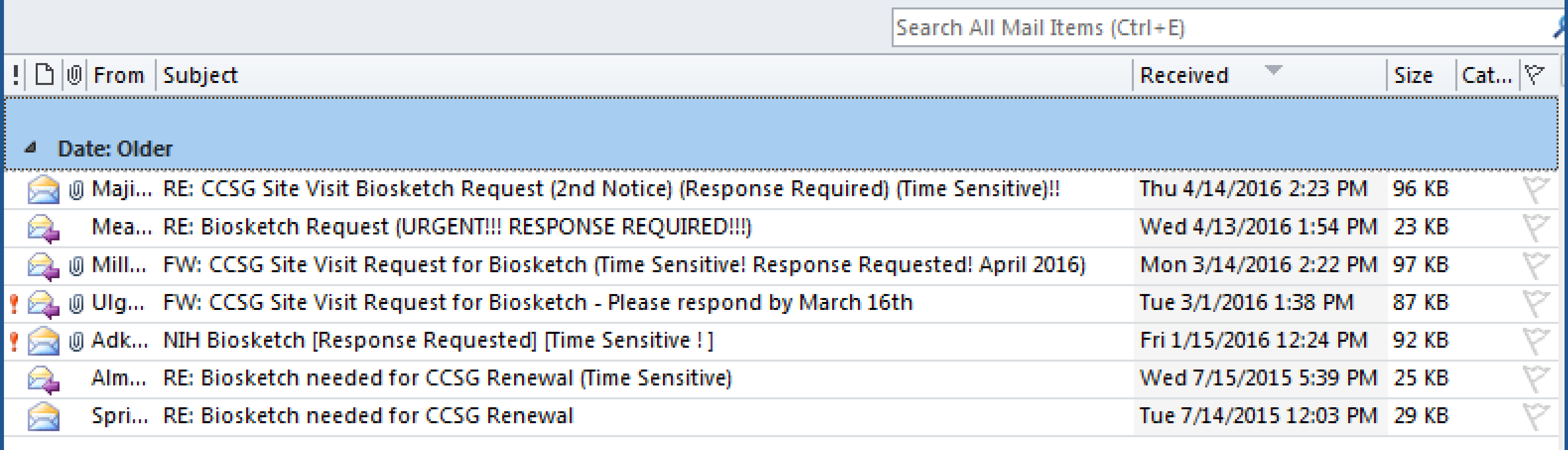
CEDARS-SINAI

MOFFITT CANCER CENTER

NCBI

The reality:
a process firmly rooted in the past

1. Email all required personnel request for Biosketch. Include links to SciENcv, provide Word Template, & latest instructions.
2. Wait. Some biosketches return. Edit. Store locally or send to shared drive.
3. Email personnel request for Biosketch Reminder. Add high priority message to email.
4. Wait. Some biosketches return. Edit. Store locally or send to shared drive.
5. Send messages to Faculty Leaders asking for support and encouragement.
6. Wait. Some biosketches return. Edit. Store locally or send to shared drive.
7. Rinse-Repeat x10 times.
8. Biosketches are gathered. Review and make final edits.



Search All Mail Items (Ctrl+E)

!	📎	From	Subject	Received	Size
			Date: Older		
	📎	Maji...	RE: CCSG Site Visit Biosketch Request (2nd Notice) (Response Required) (Time Sensitive)!!	Thu 4/14/2016 2:23 PM	96 KB
		Mea...	RE: Biosketch Request (URGENT!!! RESPONSE REQUIRED!!!)	Wed 4/13/2016 1:54 PM	23 KB
	📎	Mill...	FW: CCSG Site Visit Request for Biosketch (Time Sensitive! Response Requested! April 2016)	Mon 3/14/2016 2:22 PM	97 KB
!	📎	Ulg...	FW: CCSG Site Visit Request for Biosketch - Please respond by March 16th	Tue 3/1/2016 1:38 PM	87 KB
!	📎	Adk...	NIH Biosketch [Response Requested] [Time Sensitive !]	Fri 1/15/2016 12:24 PM	92 KB
		Alm...	RE: Biosketch needed for CCSG Renewal (Time Sensitive)	Wed 7/15/2015 5:39 PM	25 KB
		Spri...	RE: Biosketch needed for CCSG Renewal	Tue 7/14/2015 12:03 PM	29 KB

Names redacted to protect the guilty.

SciENcv:

What went right:

- Automatically puts information in new format
- Create multiple versions
- Share entry and upkeep responsibilities with delegates
- Create sharable URL
- Links to MyBibliography

What went wrong:

- URL version doesn't enable viewers to download.
- De-centralized management (PI-centric, instead of institutionally)
- No delivery mechanism:
  - PDFs & Emails can be lost, forgotten, deleted, etc.

How do members of CCAF gather Biosketches?

- 69% of respondents rely on Members to submit and maintain Biosketches
- 14% have homegrown systems that centralize and keep track of Biosketches
- 4% have some sort of vendor system

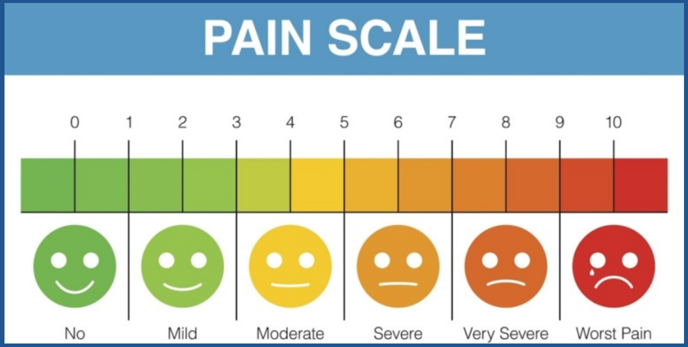
Biosketch Methods of Collection:	**# of Responses:**
We rely on **Members** to submit and maintain Biosketches.	34
We use a **homegrown system** to collect, create, manage, & store.	7
Other: Members write, we edit or provide templates.	4
Homegrown Other: Yes, we have a homegrown solution, but…	1
Vendor Other: Yes, we have a vendor solution, but…	1
We use a **vendor supported** system to collect, create, manage, & store.	1
We use existing **NIH** provided tools (NCBI, etc).	1
Grand Total	**49**

Are we happy?

Satisfaction	# of Responses:
Dissatisfied	16
OK	16
Satisfied	12
Very Dissatisfied	2
Very Satisfied	3
Grand Total	49

- 63% are pretty OK with current methods
- 37% are not

Who’s Happy Here:	#of Responses
OK	**16**
Other: Members write, we edit.	3
Vendor Other: Yes, we have a vendor solution, but…	1
We rely on Members to submit and maintain Biosketches.	12
Satisfied	**12**
We rely on Members to submit and maintain Biosketches.	7
We use a homegrown system to collect, create, manage, & store.	4
We use existing NIH provided tools (NCBI, etc).	1
Very Satisfied	**3**
We use a homegrown system to collect, create, manage, & store.	3
Grand Total	31



What's working?

Vendor Products

- Complion

Other:

- Centralized department devoted to entry & management
- Regular (Monthly!) updates

Homegrown Products

- Mission-Based Management
- Nexus
- Faculty Collaboration Database (FCD)
- Customized SciENcv Clone

The times they aren't changing....

Planning on Changing Anytime Soon?	**# of Responses:**
No.	36
We want to change, but have no plans.	1
We're curious about what others do.	1
We're looking towards our University to implement a process/product.	2
Yes. Attempting to choose between homegrown & vendor solutions.	1
Yes. We have plans to move to vendor solution.	4
Yes. We are buying a vendor solution.	1
Yes. We are working on a homegrown solution.	2
Grand Total	48

- 16% are planning on changing their methods
- 4% waiting for the next big thing

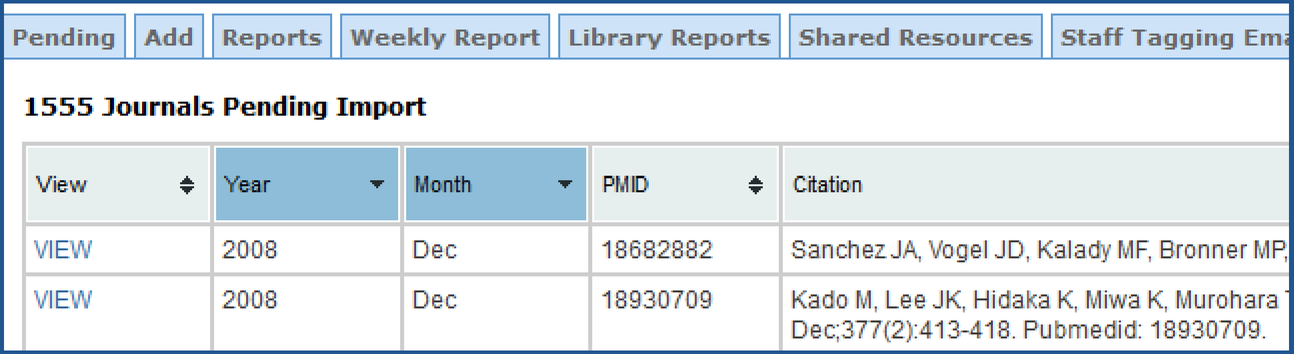
Vendors:

- Research Management System (RMS)
- Nexus
- Salesforce
- Mendix
- Café
- RES Forte

Publications

How Moffitt collects Pub Data:

- Nightly search of author names via API to MEDLINE
- Download into holding queue
  - Impact Factor automatically assigned
- Daily author verification screening by human



1555 Journals Pending Import

View	Year	Month	PMID	Citation
VIEW	2008	Dec	18682882	Sanchez JA, Vogel JD, Kalady MF, Bronner MP,
VIEW	2008	Dec	18930709	Kado M, Lee JK, Hidaka K, Miwa K, Murohara Dec;377(2):413-418. Pubmedid: 18930709.

What works for us, may not work for you:

Pros:

- Automated & customizable search algorithm
- Very little need for author input
- Standardized citation information
- Ability to pull corresponding data: Grant IDs, ORCiD, MeSH, IF

Cons:

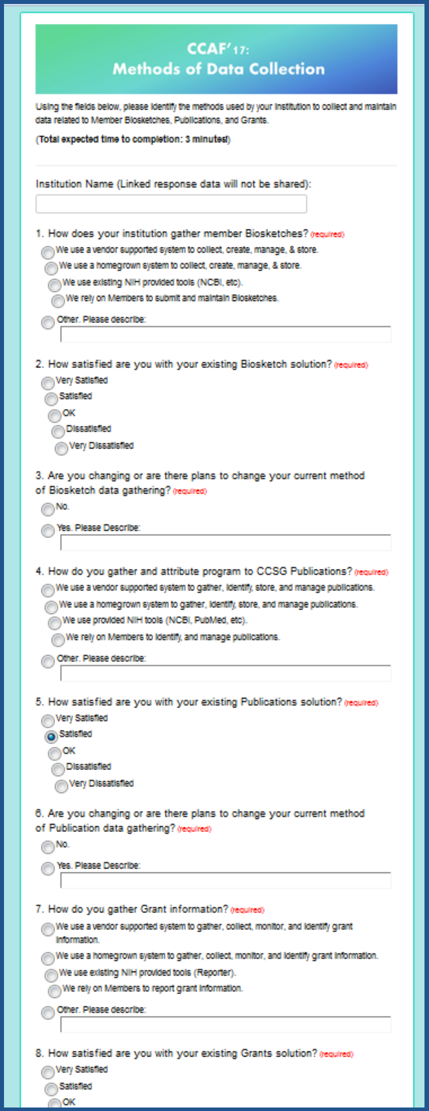
- Labor intensive
- Centralizing Screening process requires dedicated staff members
- Author Name Disambiguation remains a stumbling block

Publications: The current state and a look at our center's process

Mahendra Yatawara, MBA

CCAF-IT 2017 Survey

- http://moffitt.libsurveys.com/CCAFData
- Survey sent out April 20th
- **Survey active until May 2nd**
- Institutions responding: 44



CCAF'17: Methods of Data Collection

Using the fields below, please identify the methods used by your Institution to collect and maintain data related to Member Biosketches, Publications, and Grants.

(Total expected time to completion: 3 minutes)

Institution Name (Linked response data will not be shared):

1. How does your institution gather member Biosketches? (required)

○ We use a vendor supported system to collect, create, manage, & store.
○ We use a homegrown system to collect, create, manage, & store.
○ We use existing NIH provided tools (NCBI, etc).
○ We rely on Members to submit and maintain Biosketches.
○ Other. Please describe:

2. How satisfied are you with your existing Biosketch solution? (required)

○ Very Satisfied
○ Satisfied
○ OK
○ Dissatisfied
○ Very Dissatisfied

3. Are you changing or are there plans to change your current method of Biosketch data gathering? (required)

○ No.
○ Yes. Please Describe:

4. How do you gather and attribute program to CCSG Publications? (required)

○ We use a vendor supported system to gather, identify, store, and manage publications.
○ We use a homegrown system to gather, identify, store, and manage publications.
○ We use provided NIH tools (NCBI, PubMed, etc).
○ We rely on Members to identify, and manage publications.
○ Other. Please describe:

5. How satisfied are you with your existing Publications solution? (required)

○ Very Satisfied
● Satisfied
○ OK
○ Dissatisfied
○ Very Dissatisfied

6. Are you changing or are there plans to change your current method of Publication data gathering? (required)

○ No.
○ Yes. Please Describe:

7. How do you gather Grant information? (required)

○ We use a vendor supported system to gather, collect, monitor, and identify grant information.
○ We use a homegrown system to gather, collect, monitor, and identify grant information.
○ We use existing NIH provided tools (Reporter).
○ We rely on Members to report grant information.
○ Other. Please describe:

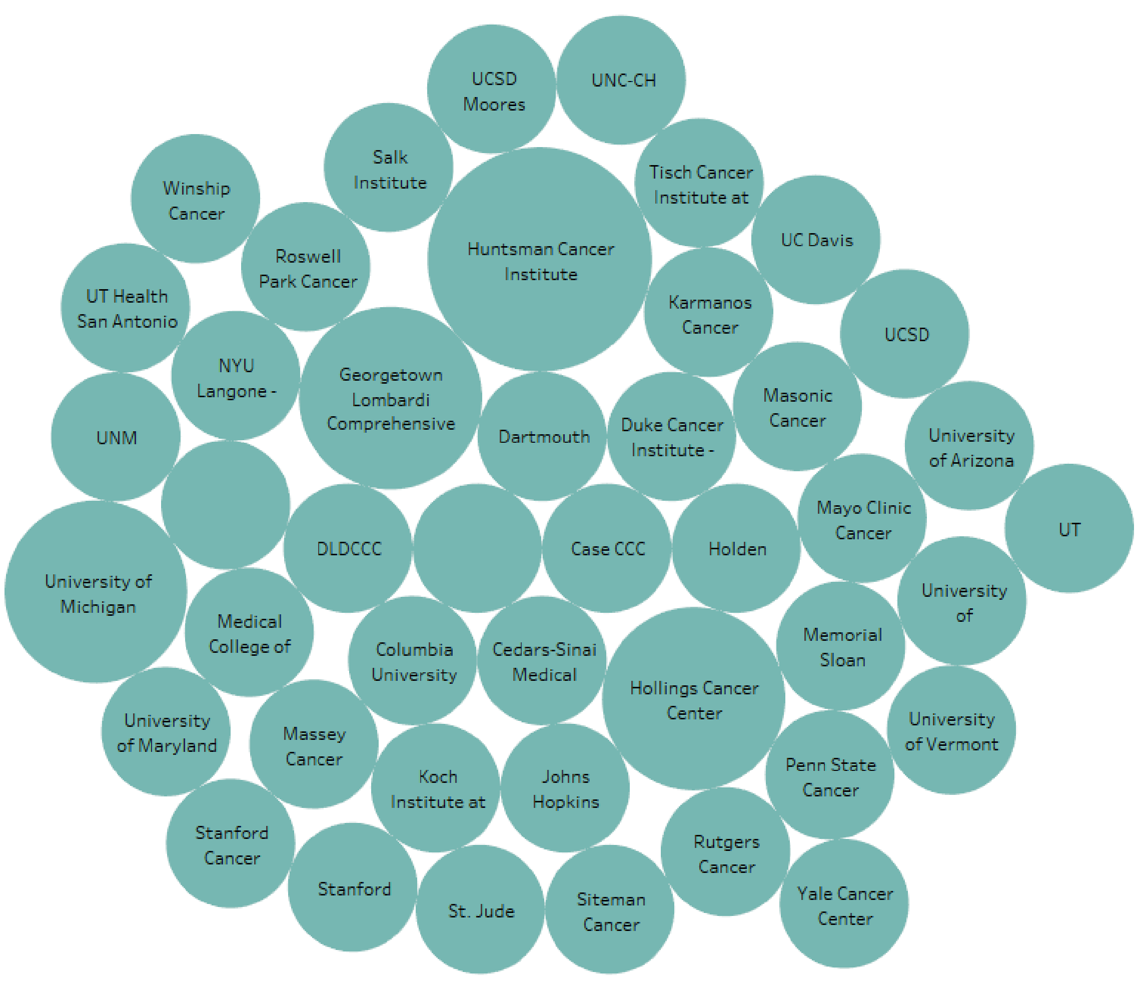
8. How satisfied are you with your existing Grants solution? (required)

○ Very Satisfied
○ Satisfied
○ OK

NIH U.S. National Library of Medicine

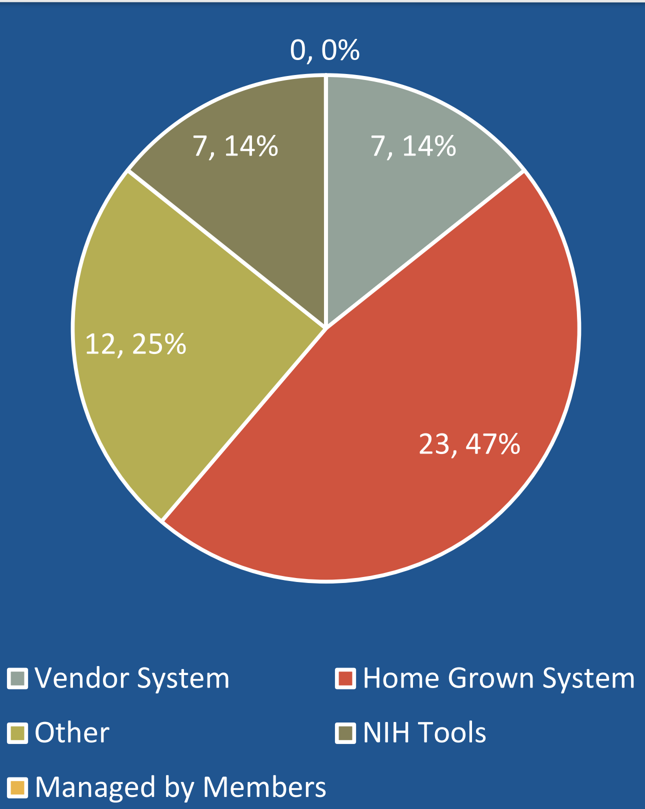
NCBI

Institutions Responding



How do Centers manage Pubs for CCSG?

Publications Systems



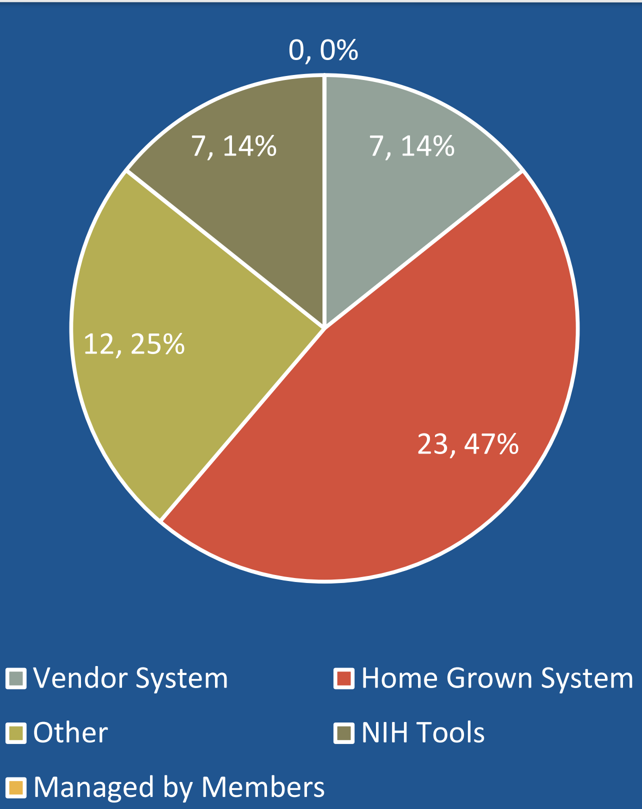
Vendor and Other

CAFÉ by USC	3
Opus/EVAL by Forte	2
Lattice Grid	2

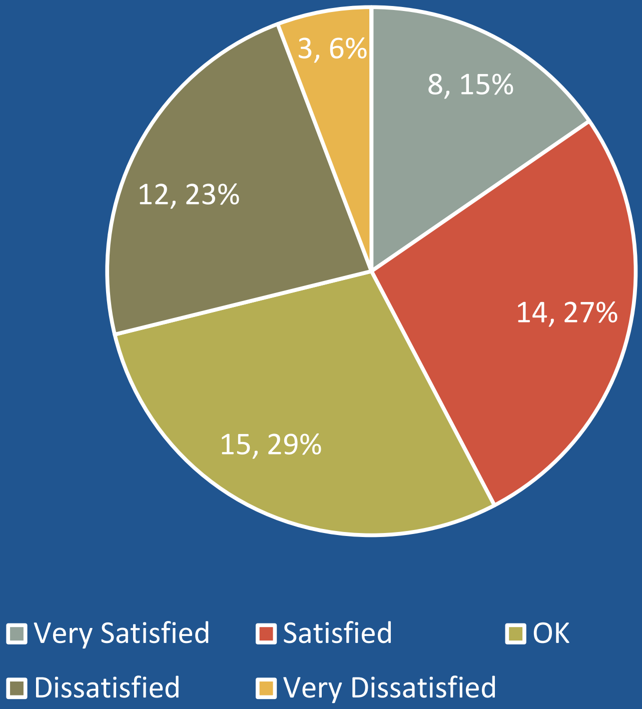
Homegrown & NIH
Nexus
Homegrown

Satisfaction with current Pubs System?

Publications Systems

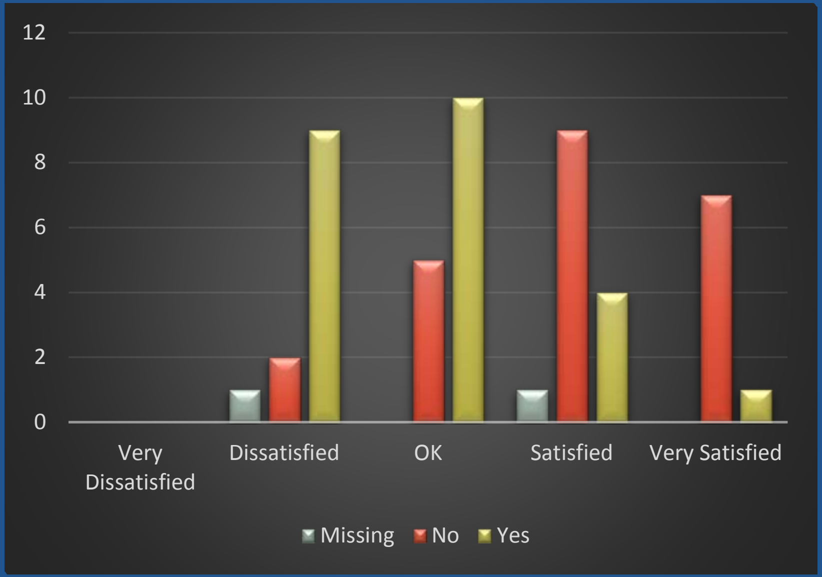


Satisfaction Level

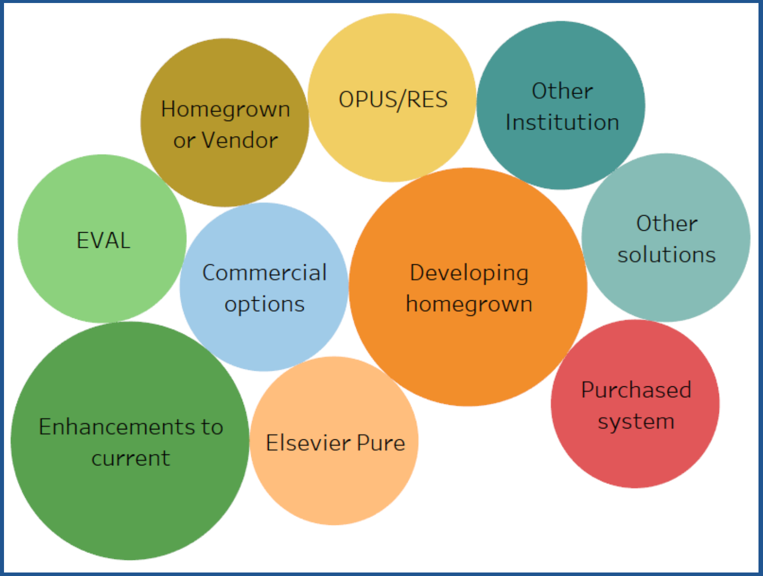


Plans to change Pubs Solution?

Satisfaction and Change

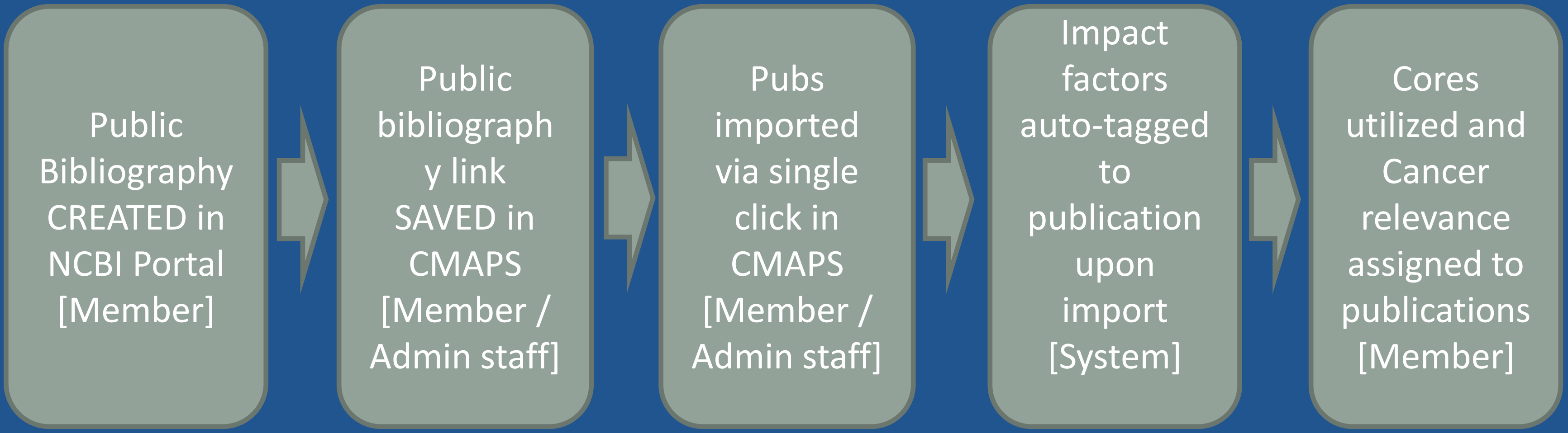


Possible alternatives



How Cedars-Sinai Collects Publications Data

Initial pubs import



Public Bibliography CREATED in NCBI Portal [Member]

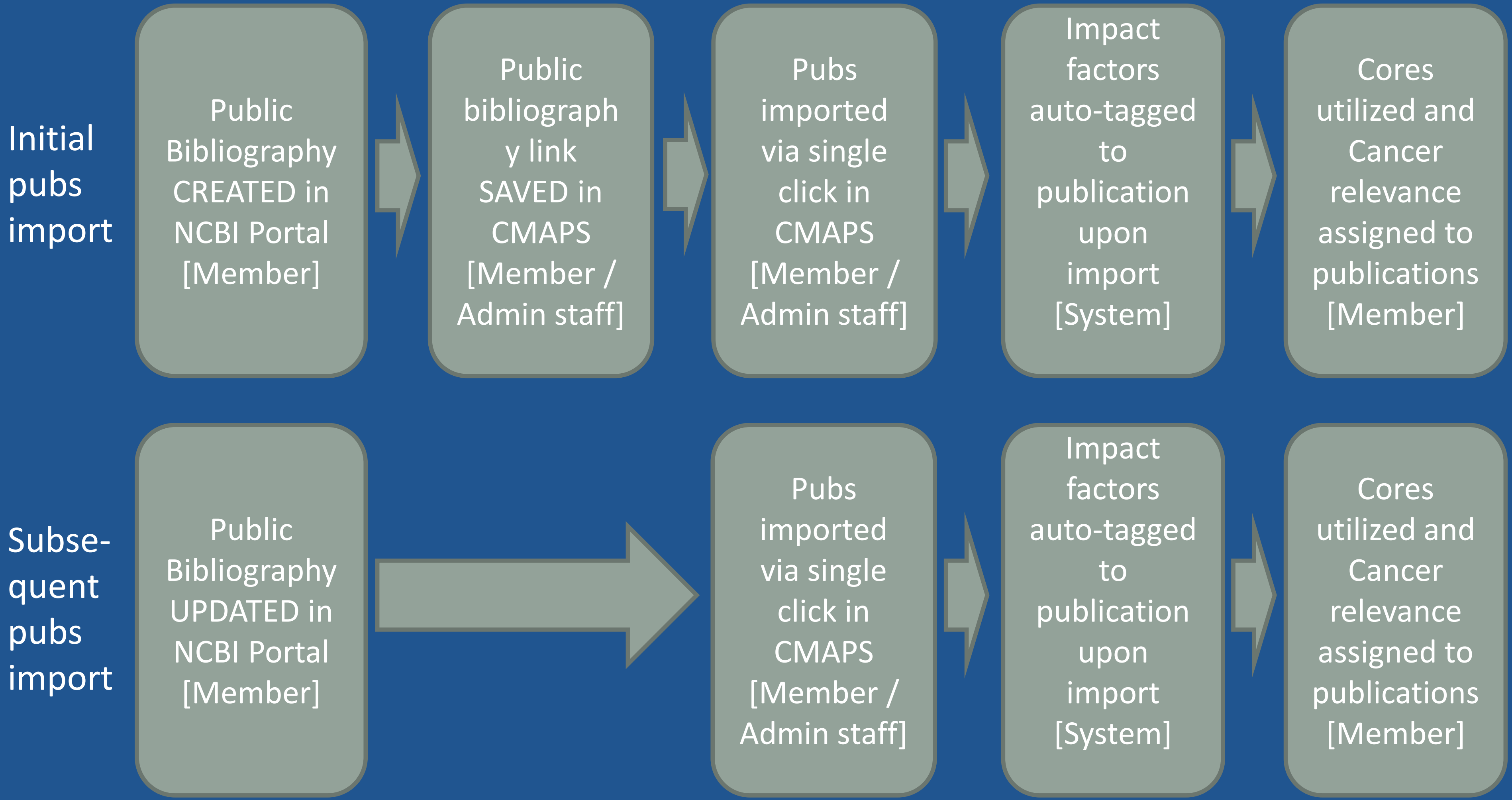
Public bibliography link SAVED in CMAPS [Member / Admin staff]

Pubs imported via single click in CMAPS [Member / Admin staff]

Impact factors auto-tagged to publication upon import [System]

Cores utilized and Cancer relevance assigned to publications [Member]

How Cedars-Sinai Collects Publications Data



Initial pubs import

Public Bibliography CREATED in NCBI Portal [Member]

Public bibliography link SAVED in CMAPS [Member / Admin staff]

Pubs imported via single click in CMAPS [Member / Admin staff]

Impact factors auto-tagged to publication upon import [System]

Cores utilized and Cancer relevance assigned to publications [Member]

Subse-quent pubs import

Public Bibliography UPDATED in NCBI Portal [Member]

Pubs imported via single click in CMAPS [Member / Admin staff]

Impact factors auto-tagged to publication upon import [System]

Cores utilized and Cancer relevance assigned to publications [Member]

How is this process working for us?

Key Advantages

- Reduction in non-value added work from CC Admin
- Members maintain in single location (NCBI portal)
- Auto-assignment of Impact Factor
- One-click reports

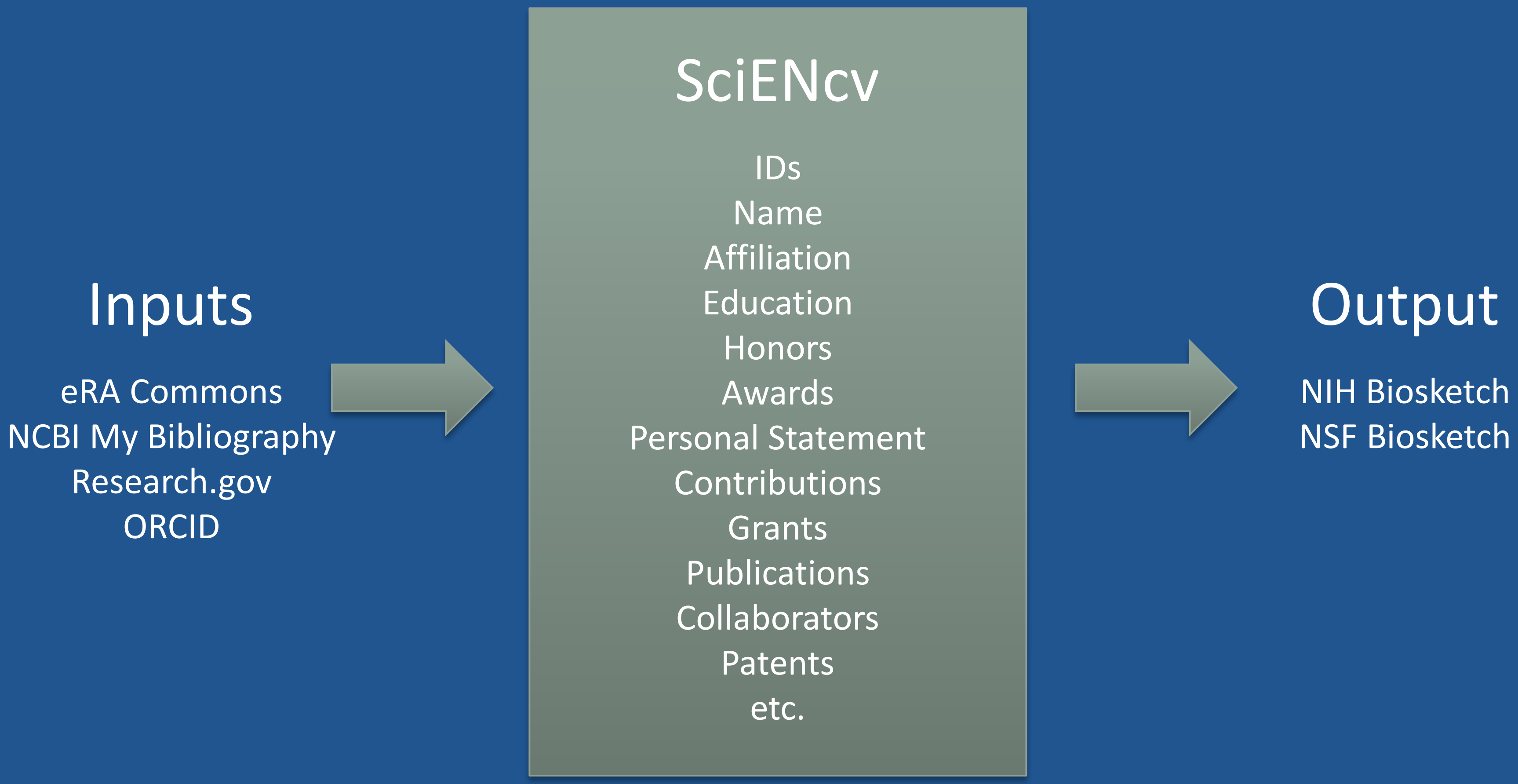
Limitations

- Reminders for Members to keep NCBI Bibliography up-to-date
- Reminders for Members to allocate Core usage and Cancer Relevance to pubs in CMAPS

The Futures: Biosketches, Grants, Pubs… and Data!

Ben Busby, NCBI

SciENcv



Inputs

eRA Commons
NCBI My Bibliography
Research.gov
ORCID

SciENcv

IDs
Name
Affiliation
Education
Honors
Awards
Personal Statement
Contributions
Grants
Publications
Collaborators
Patents
etc.

Output

NIH Biosketch
NSF Biosketch

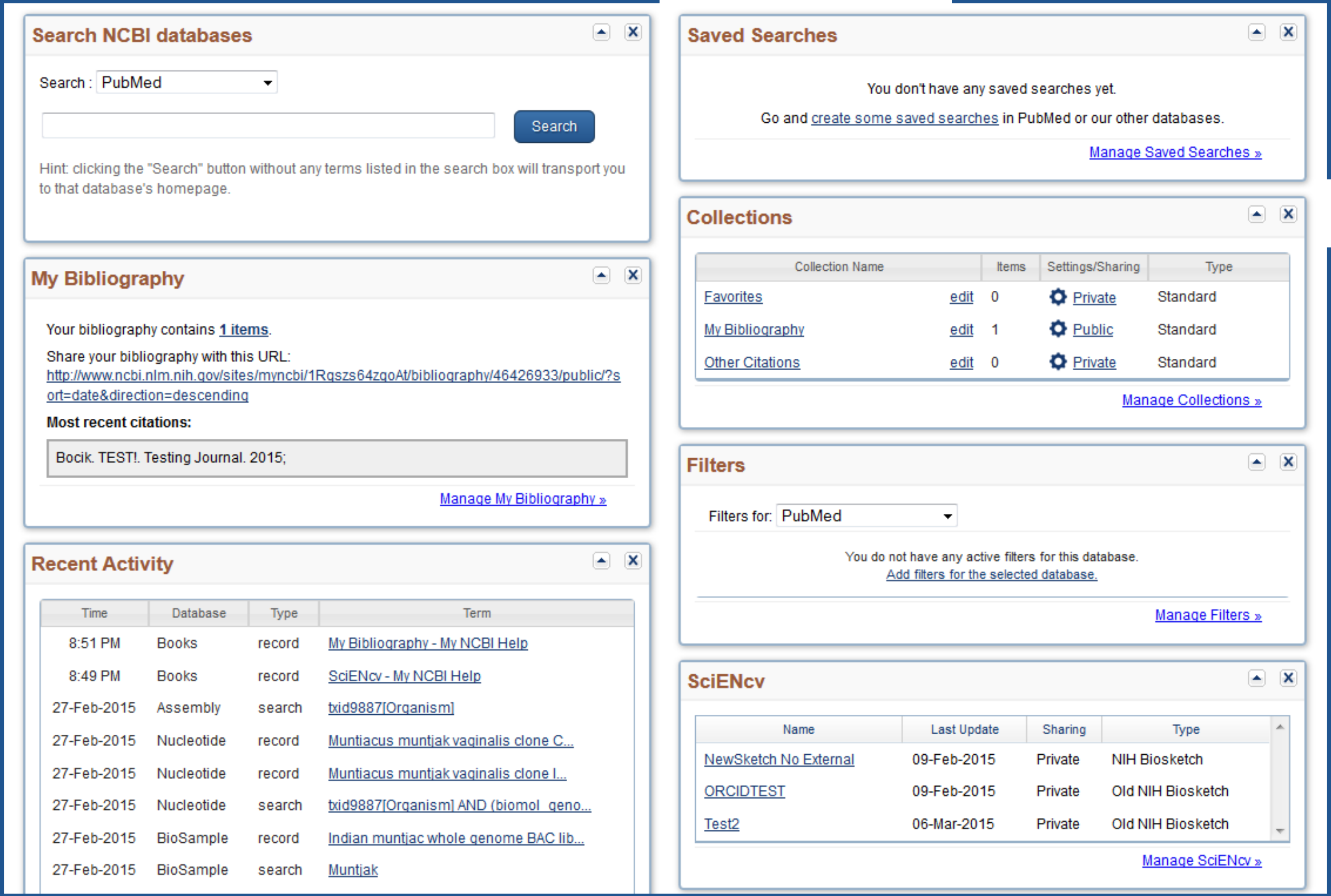
NCBI

Saved Searches

Collections

My Bibliography

SciENcv



Search NCBI databases

Search : PubMed

Search

Hint: clicking the "Search" button without any terms listed in the search box will transport you to that database's homepage.

My Bibliography

Your bibliography contains 1 items.

Share your bibliography with this URL:
http://www.ncbi.nlm.nih.gov/sites/myncbi/1Rgszs64zgoAt/bibliography/46426933/public/?sort=date&direction=descending

Most recent citations:

Bocik. TEST!. Testing Journal. 2015;

Manage My Bibliography »

Recent Activity

Time	Database	Type	Term
8:51 PM	Books	record	My Bibliography - My NCBI Help
8:49 PM	Books	record	SciENcv - My NCBI Help
27-Feb-2015	Assembly	search	txid9887[Organism]
27-Feb-2015	Nucleotide	record	Muntiacus muntjak vaginalis clone C...
27-Feb-2015	Nucleotide	record	Muntiacus muntjak vaginalis clone I...
27-Feb-2015	Nucleotide	search	txid9887[Organism] AND (biomol_geno...
27-Feb-2015	BioSample	record	Indian muntjac whole genome BAC lib...
27-Feb-2015	BioSample	search	Muntjak

Saved Searches

You don't have any saved searches yet.

Go and create some saved searches in PubMed or our other databases.

Manage Saved Searches »

Collections

Collection Name		Items	Settings/Sharing	Type
Favorites	edit	0	Private	Standard
My Bibliography	edit	1	Public	Standard
Other Citations	edit	0	Private	Standard

Manage Collections »

Filters

Filters for: PubMed

You do not have any active filters for this database.
Add filters for the selected database.

Manage Filters »

SciENcv

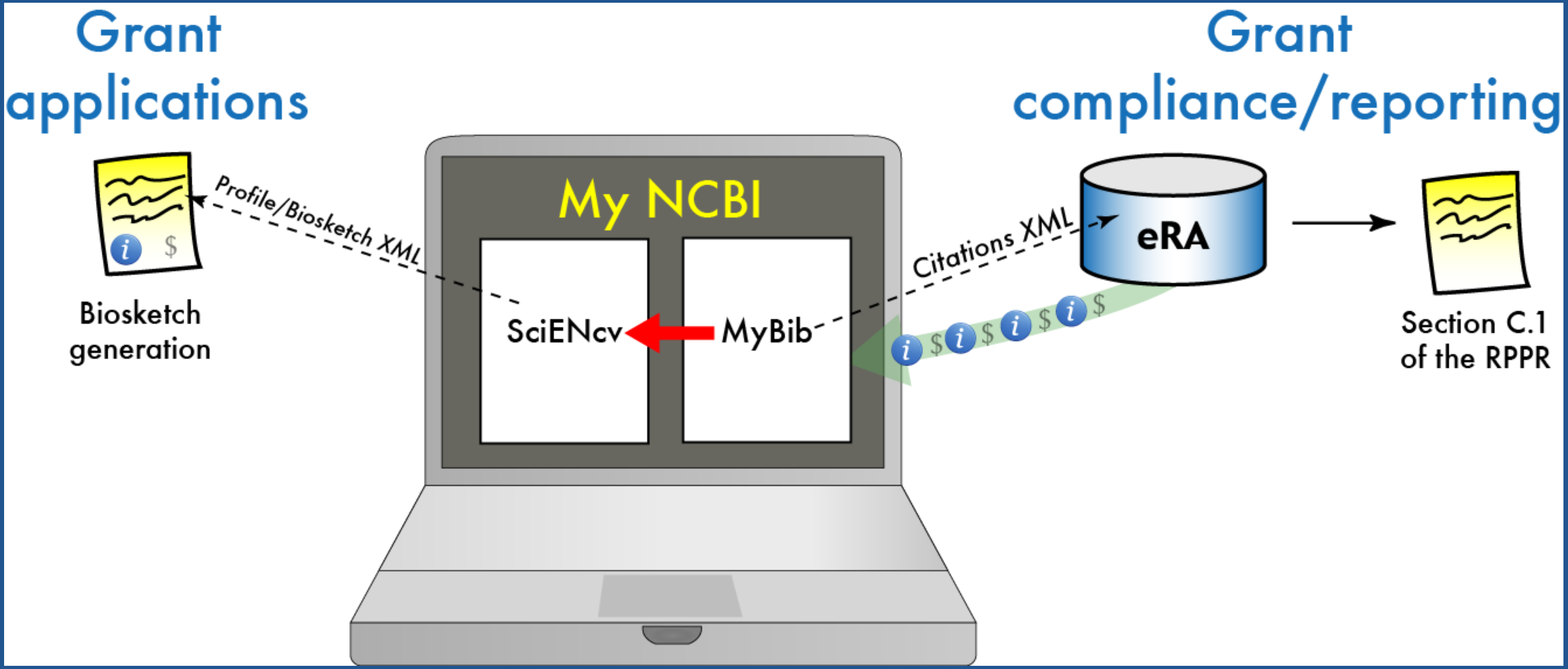
Name	Last Update	Sharing	Type
NewSketch No External	09-Feb-2015	Private	NIH Biosketch
ORCIDTEST	09-Feb-2015	Private	Old NIH Biosketch
Test2	06-Mar-2015	Private	Old NIH Biosketch

Manage SciENcv »

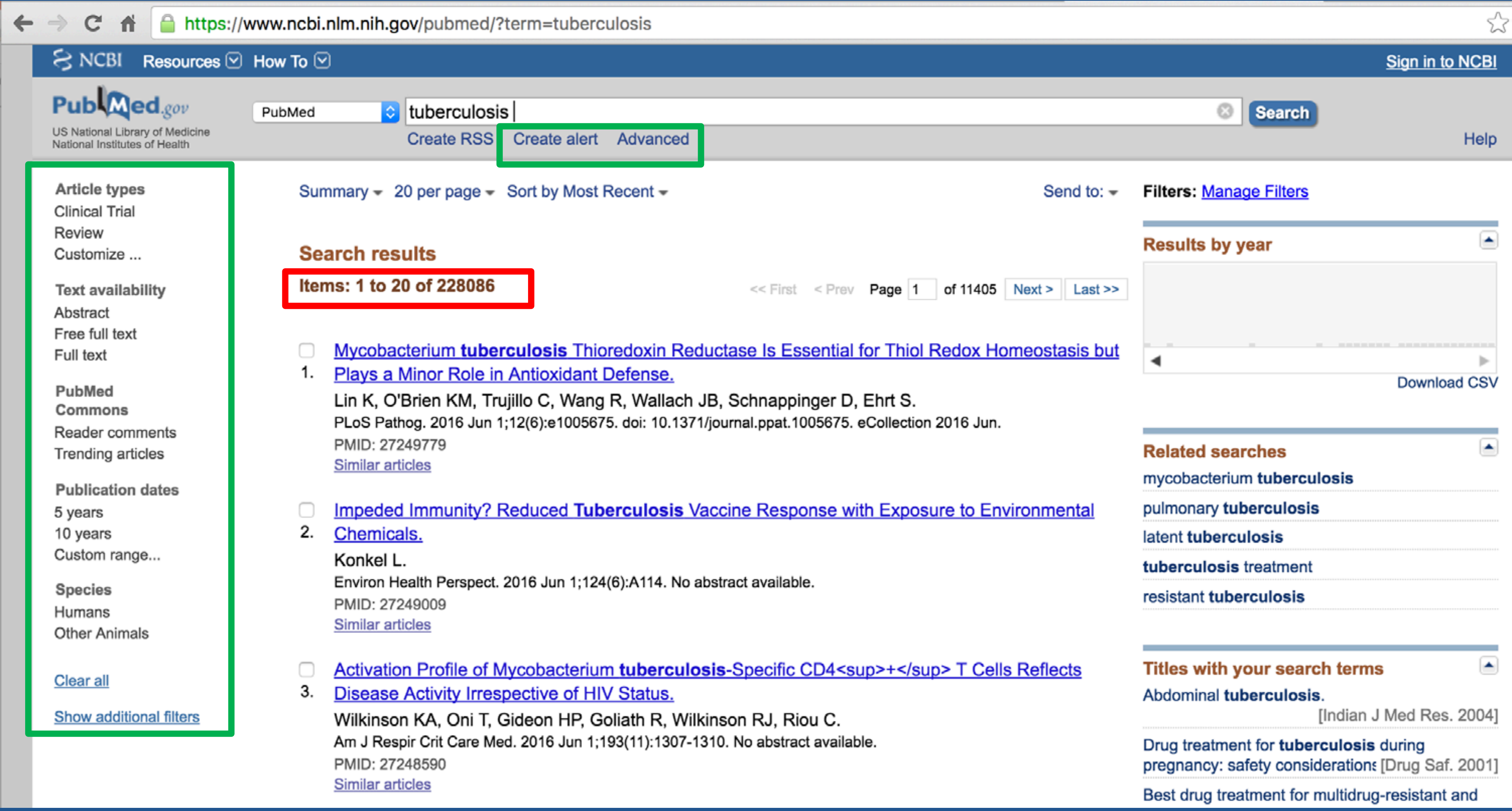
NIH U.S. National Library of Medicine

NCBI

NCBI


Grant applications
Biosketch generation
Profile/Biosketch XML
My NCBI
SciENcv
MyBib
Citations XML
eRA
Grant compliance/reporting
Section C.1 of the RPPR
i
$

Better PubMed Searches!


https://www.ncbi.nlm.nih.gov/pubmed/?term=tuberculosis
NCBI
Resources
How To
Sign in to NCBI
PubMed.gov
US National Library of Medicine
National Institutes of Health
PubMed
Search
Create RSS
Create alert
Advanced
Help

tuberculosis

Article types
Clinical Trial
Review
Customize ...

Text availability
Abstract
Free full text
Full text

PubMed Commons
Reader comments
Trending articles

Publication dates
5 years
10 years
Custom range...

Species
Humans
Other Animals

Clear all

Show additional filters

Summary ▾ 20 per page ▾ Sort by Most Recent ▾ Send to: ▾

Filters: Manage Filters

Search results

Items: 1 to 20 of 228086

<< First < Prev Page 1 of 11405 Next > Last >>

□ 1. Mycobacterium **tuberculosis** Thioredoxin Reductase Is Essential for Thiol Redox Homeostasis but Plays a Minor Role in Antioxidant Defense.
Lin K, O'Brien KM, Trujillo C, Wang R, Wallach JB, Schnappinger D, Ehrt S.
PLoS Pathog. 2016 Jun 1;12(6):e1005675. doi: 10.1371/journal.ppat.1005675. eCollection 2016 Jun.
PMID: 27249779
Similar articles

□ 2. Impeded Immunity? Reduced **Tuberculosis** Vaccine Response with Exposure to Environmental Chemicals.
Konkel L.
Environ Health Perspect. 2016 Jun 1;124(6):A114. No abstract available.
PMID: 27249009
Similar articles

□ 3. Activation Profile of Mycobacterium **tuberculosis**-Specific CD4⁺ T Cells Reflects Disease Activity Irrespective of HIV Status.
Wilkinson KA, Oni T, Gideon HP, Goliath R, Wilkinson RJ, Riou C.
Am J Respir Crit Care Med. 2016 Jun 1;193(11):1307-1310. No abstract available.
PMID: 27248590
Similar articles

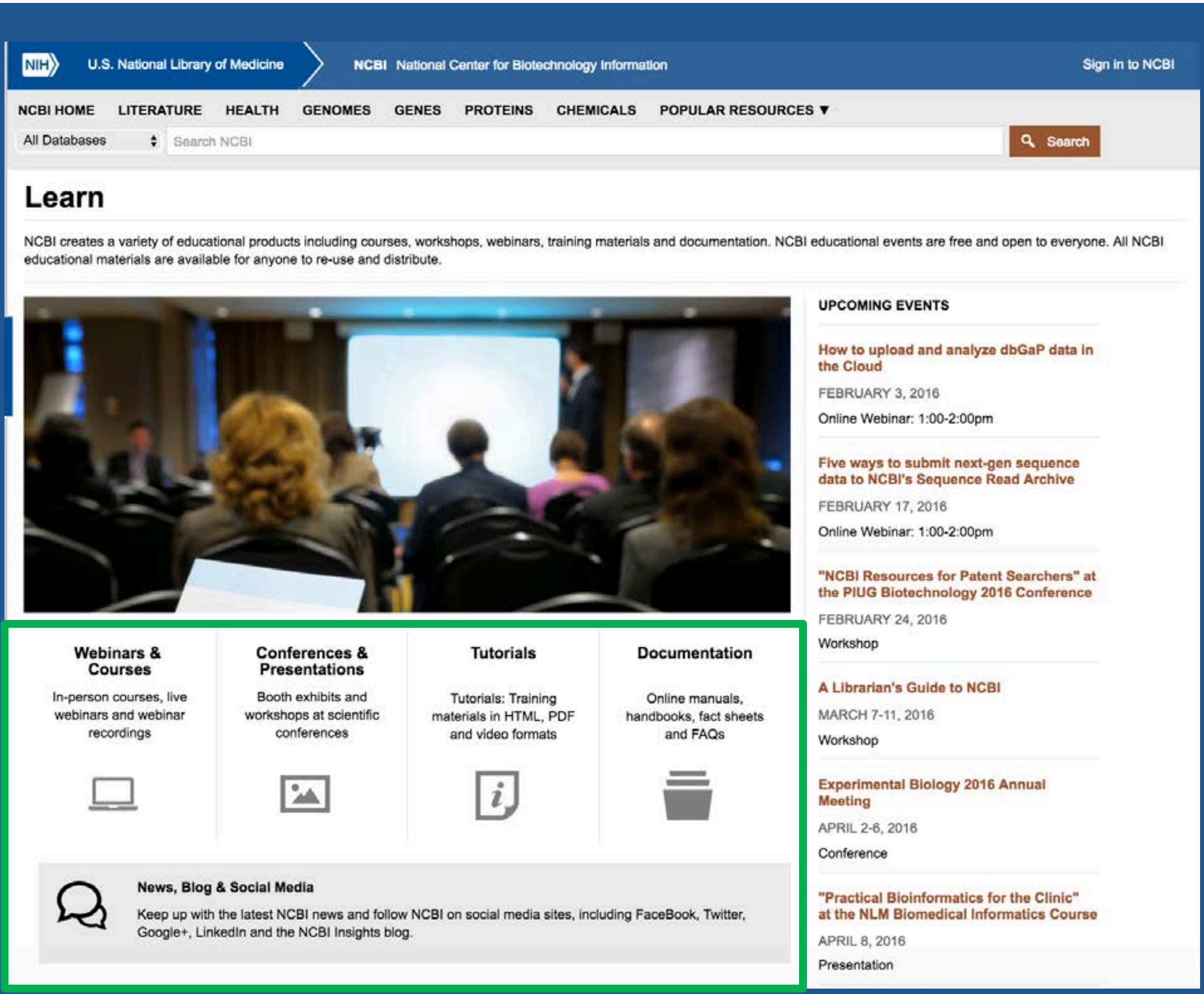
Results by year

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]
Drug treatment for **tuberculosis** during pregnancy: safety considerations [Drug Saf. 2001]
Best drug treatment for multidrug-resistant and

For more information go to: ncbi.nlm.nih.gov/learn



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.

Webinars & Courses

In-person courses, live webinars and webinar recordings

Conferences & Presentations

Booth exhibits and workshops at scientific conferences

Tutorials

Tutorials: Training materials in HTML, PDF and video formats

Documentation

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.

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

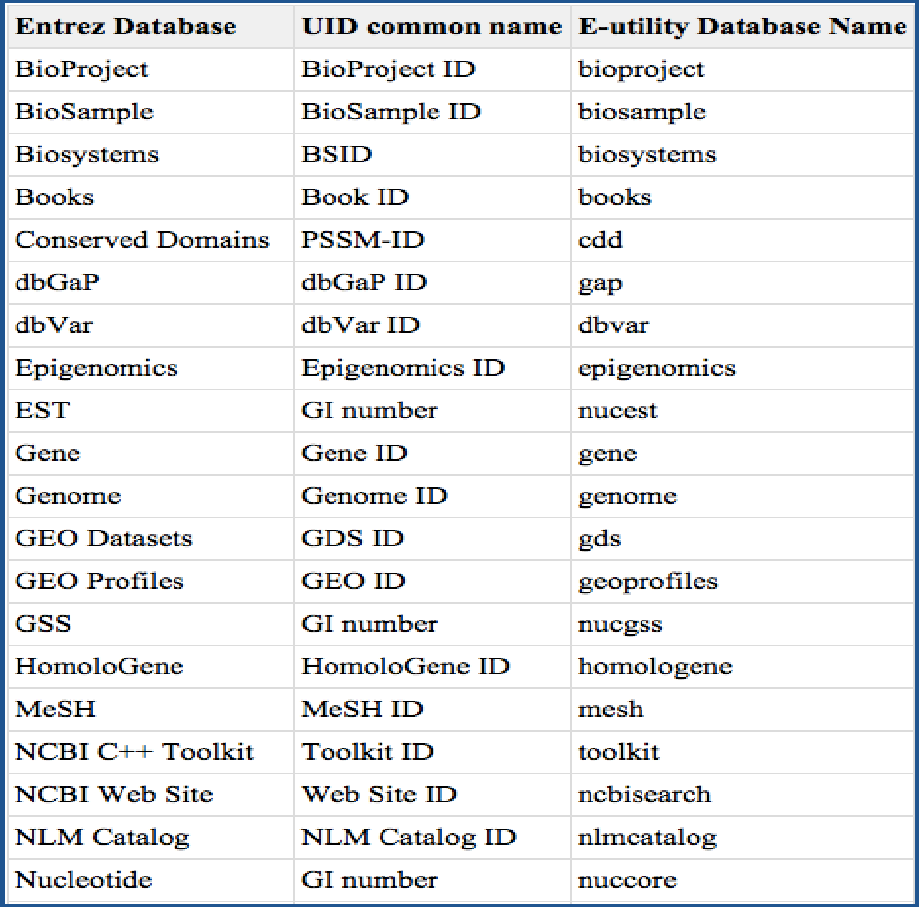
"NCBI Resources for Patent Searchers" at the PIUG Biotechnology 2016 Conference
FEBRUARY 24, 2016
Workshop

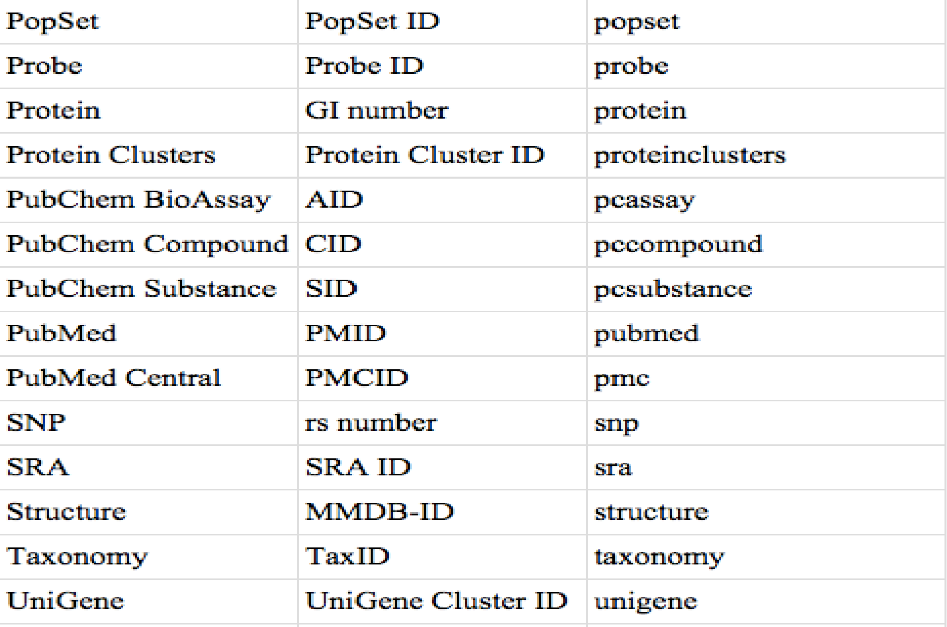
A Librarian's Guide to NCBI
MARCH 7-11, 2016
Workshop

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

E-Utilities (Eutils)





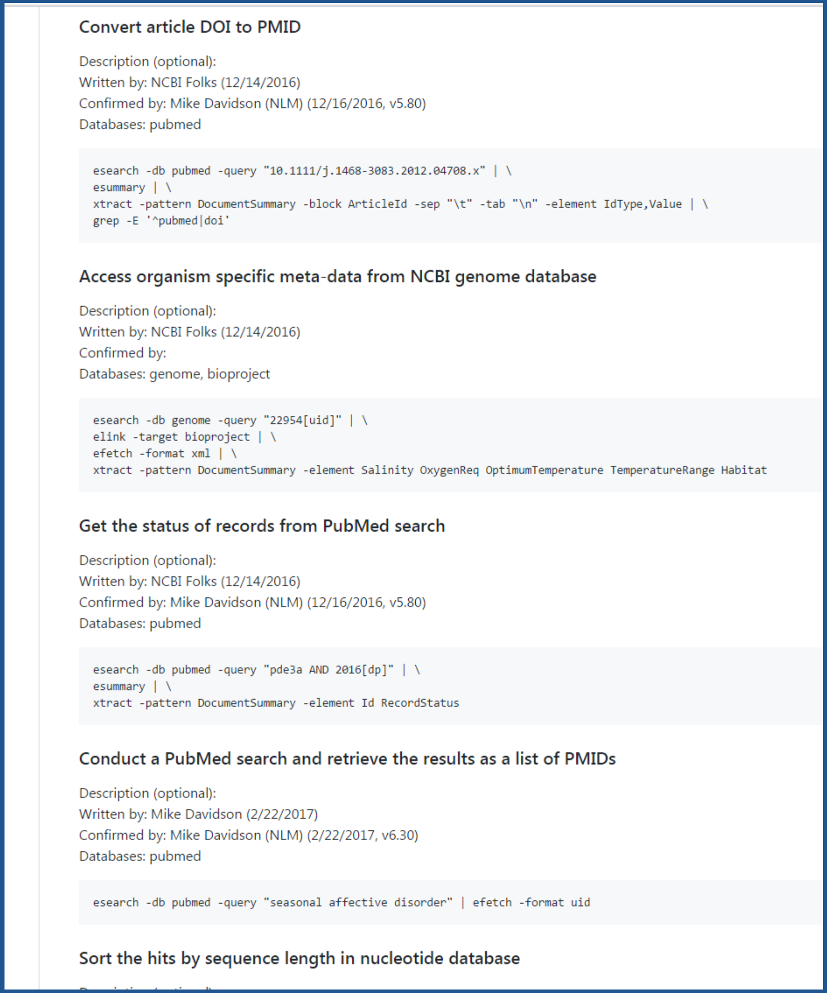
Entrez Database	UID common name	E-utility Database Name
BioProject	BioProject ID	bioproject
BioSample	BioSample ID	biosample
Biosystems	BSID	biosystems
Books	Book ID	books
Conserved Domains	PSSM-ID	cdd
dbGaP	dbGaP ID	gap
dbVar	dbVar ID	dbvar
Epigenomics	Epigenomics ID	epigenomics
EST	GI number	nucest
Gene	Gene ID	gene
Genome	Genome ID	genome
GEO Datasets	GDS ID	gds
GEO Profiles	GEO ID	geoprofiles
GSS	GI number	nucgss
HomoloGene	HomoloGene ID	homologene
MeSH	MeSH ID	mesh
NCBI C++ Toolkit	Toolkit ID	toolkit
NCBI Web Site	Web Site ID	ncbisearch
NLM Catalog	NLM Catalog ID	nlmcatalog
Nucleotide	GI number	nuccore
PopSet	PopSet ID	popset
Probe	Probe ID	probe
Protein	GI number	protein
Protein Clusters	Protein Cluster ID	proteinclusters
PubChem BioAssay	AID	pcassay
PubChem Compound	CID	pccompound
PubChem Substance	SID	pcsubstance
PubMed	PMID	pubmed
PubMed Central	PMCID	pmc
SNP	rs number	snp
SRA	SRA ID	sra
Structure	MMDB-ID	structure
Taxonomy	TaxID	taxonomy
UniGene	UniGene Cluster ID	unigene

Introducing… Entrez Direct
The E-utilities on the UNIX command line

```
esearch –db gene –query “foxp2[gene]
AND human[orgn]” | \

elink –target protein –name
gene_protein_refseq | \

efetch –format fasta
```

ftp.ncbi.nlm.nih.gov/entrez/entrezdirect/

The EDirect Cookbook!



Convert article DOI to PMID

Description (optional):
Written by: NCBI Folks (12/14/2016)
Confirmed by: Mike Davidson (NLM) (12/16/2016, v5.80)
Databases: pubmed

```
esearch -db pubmed -query "10.1111/j.1468-3083.2012.04708.x" | \
esummary | \
xtract -pattern DocumentSummary -block ArticleId -sep "\t" -tab "\n" -element IdType,Value | \
grep -E '^pubmed|doi'
```

Access organism specific meta-data from NCBI genome database

Description (optional):
Written by: NCBI Folks (12/14/2016)
Confirmed by:
Databases: genome, bioproject

```
esearch -db genome -query "22954[uid]" | \
elink -target bioproject | \
efetch -format xml | \
xtract -pattern DocumentSummary -element Salinity OxygenReq OptimumTemperature TemperatureRange Habitat
```

Get the status of records from PubMed search

Description (optional):
Written by: NCBI Folks (12/14/2016)
Confirmed by: Mike Davidson (NLM) (12/16/2016, v5.80)
Databases: pubmed

```
esearch -db pubmed -query "pde3a AND 2016[dp]" | \
esummary | \
xtract -pattern DocumentSummary -element Id RecordStatus
```

Conduct a PubMed search and retrieve the results as a list of PMIDs

Description (optional):
Written by: Mike Davidson (2/22/2017)
Confirmed by: Mike Davidson (NLM) (2/22/2017, v6.30)
Databases: pubmed

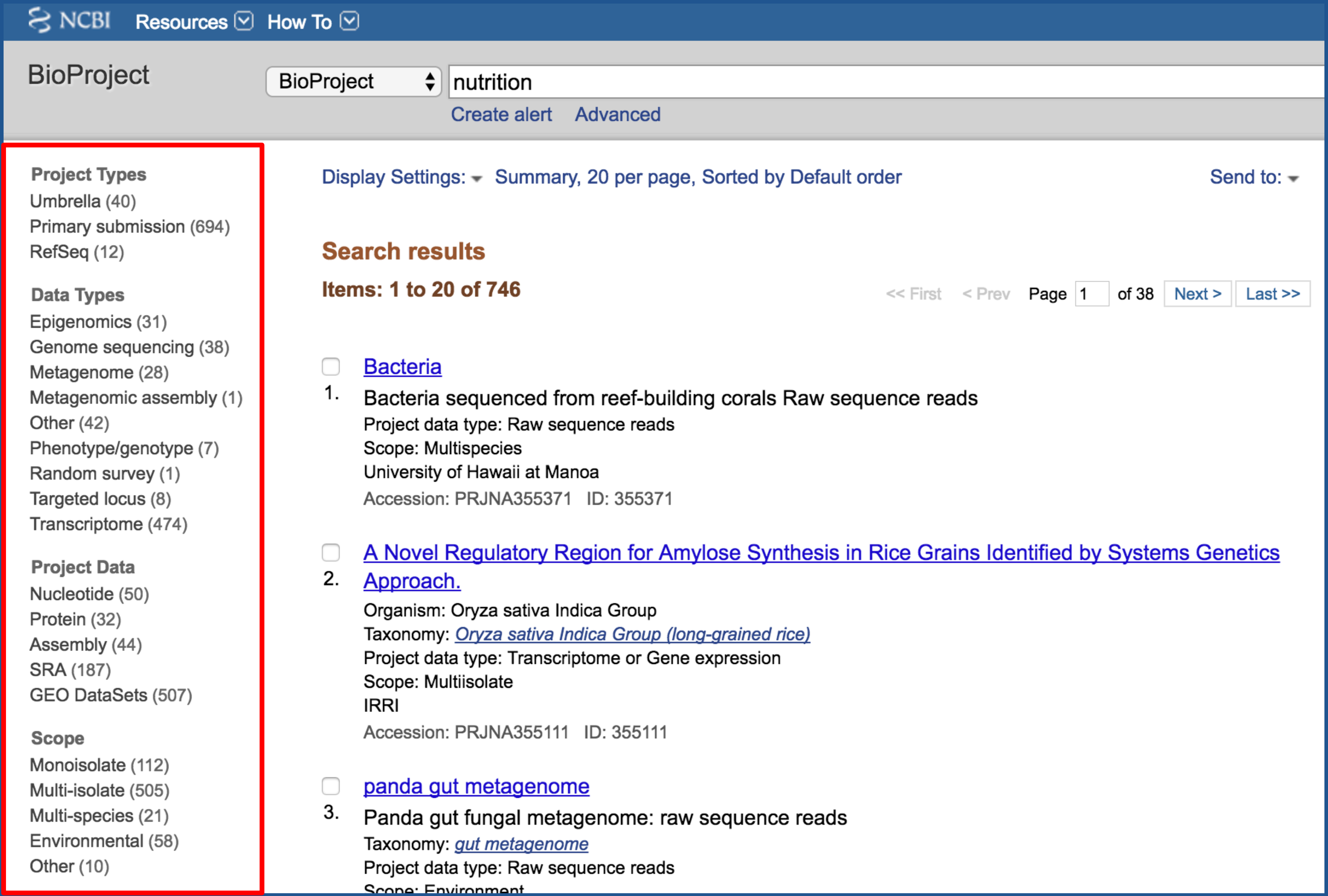
```
esearch -db pubmed -query "seasonal affective disorder" | efetch -format uid
```

Sort the hits by sequence length in nucleotide database

Description (optional):

Google for
EDirect Cookbook

BioProject



Summary, 20 per page, Sorted by Default order

Project Types
Umbrella (40)
Primary submission (694)
RefSeq (12)

Data Types
Epigenomics (31)
Genome sequencing (38)
Metagenome (28)
Metagenomic assembly (1)
Other (42)
Phenotype/genotype (7)
Random survey (1)
Targeted locus (8)
Transcriptome (474)

Project Data
Nucleotide (50)
Protein (32)
Assembly (44)
SRA (187)
GEO DataSets (507)

Scope
Monoisolate (112)
Multi-isolate (505)
Multi-species (21)
Environmental (58)
Other (10)

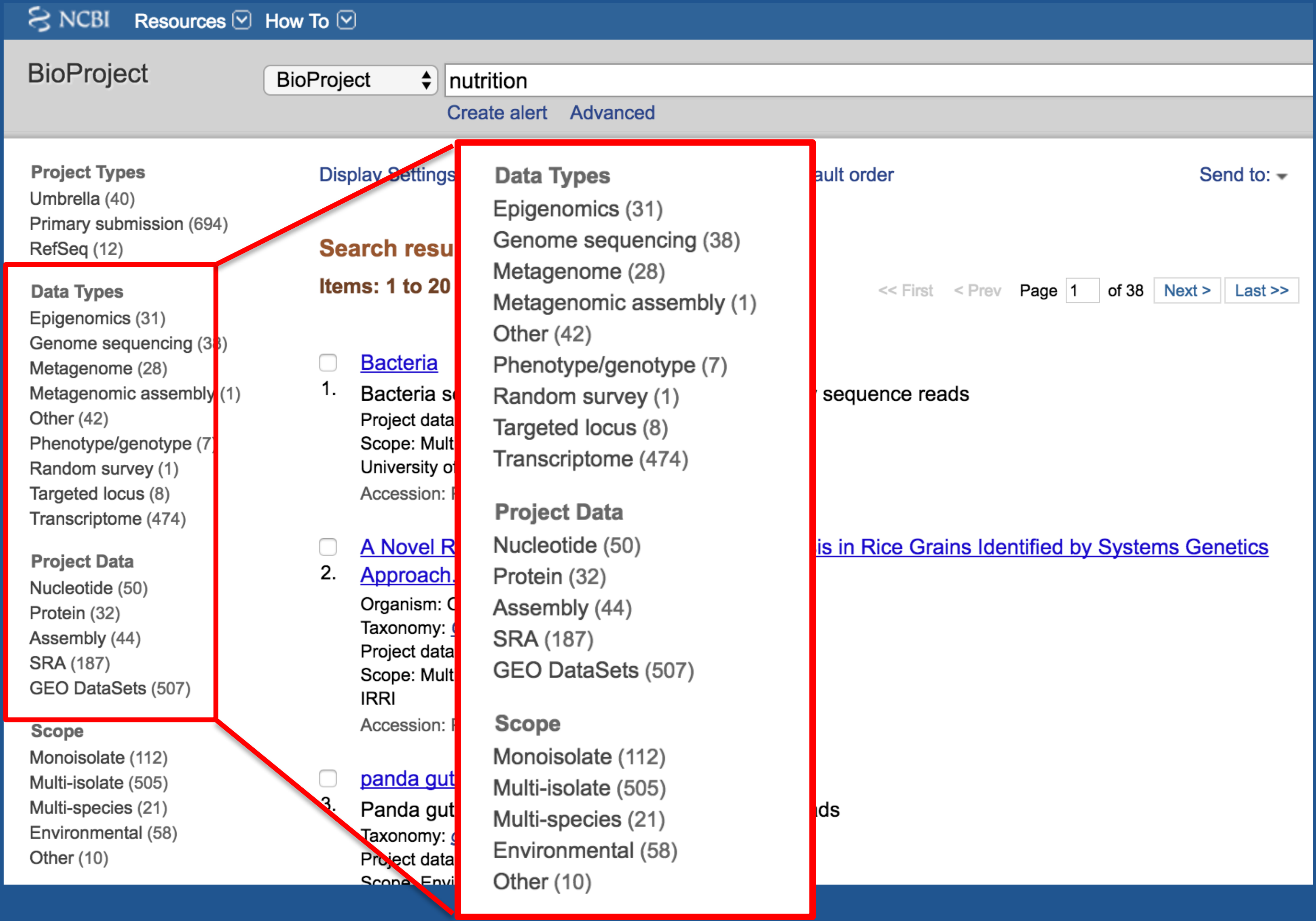
Search results

Items: 1 to 20 of 746

□ Bacteria
1. Bacteria sequenced from reef-building corals Raw sequence reads
Project data type: Raw sequence reads
Scope: Multispecies
University of Hawaii at Manoa
Accession: PRJNA355371 ID: 355371

□ A Novel Regulatory Region for Amylose Synthesis in Rice Grains Identified by Systems Genetics
2. Approach.
Organism: Oryza sativa Indica Group
Taxonomy: *Oryza sativa Indica Group (long-grained rice)*
Project data type: Transcriptome or Gene expression
Scope: Multiisolate
IRRI
Accession: PRJNA355111 ID: 355111

□ panda gut metagenome
3. Panda gut fungal metagenome: raw sequence reads
Taxonomy: *gut metagenome*
Project data type: Raw sequence reads
Scope: Environment



BioProject: nutrition

Page 1 of 38

Project Types
Umbrella (40)
Primary submission (694)
RefSeq (12)

Data Types
Epigenomics (31)
Genome sequencing (38)
Metagenome (28)
Metagenomic assembly (1)
Other (42)
Phenotype/genotype (7)
Random survey (1)
Targeted locus (8)
Transcriptome (474)

Project Data
Nucleotide (50)
Protein (32)
Assembly (44)
SRA (187)
GEO DataSets (507)

Scope
Monoisolate (112)
Multi-isolate (505)
Multi-species (21)
Environmental (58)
Other (10)

Search resu

Items: 1 to 20

1. Bacteria
Bacteria s ... sequence reads
Project data
Scope: Mult
University o
Accession: P

2. A Novel R ... is in Rice Grains Identified by Systems Genetics Approach.
Organism: O
Taxonomy: O
Project data
Scope: Mult
IRRI
Accession: P

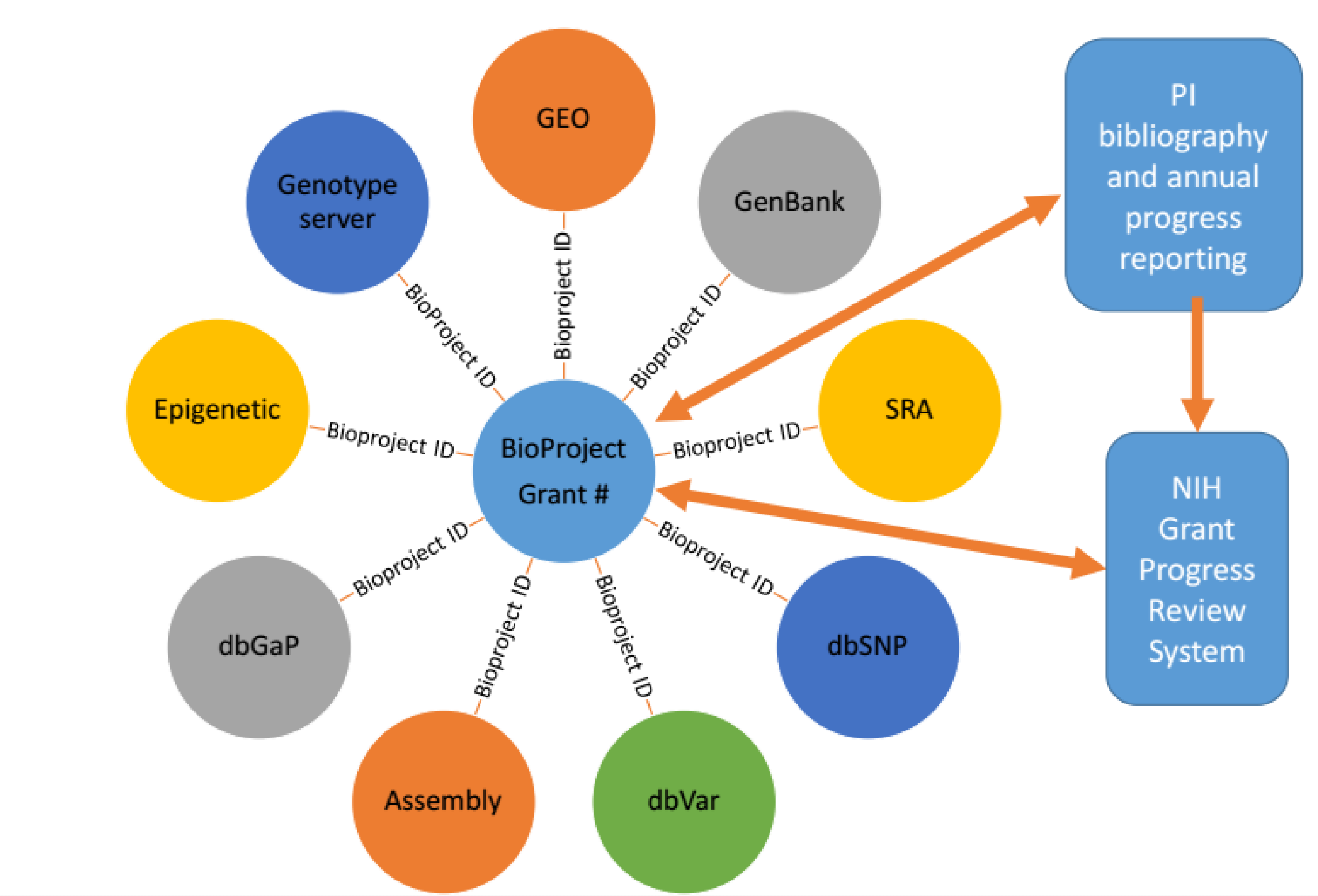
3. panda gut
Panda gut ... ds
Taxonomy: g
Project data
Scope: Envi

Data Types
Epigenomics (31)
Genome sequencing (38)
Metagenome (28)
Metagenomic assembly (1)
Other (42)
Phenotype/genotype (7)
Random survey (1)
Targeted locus (8)
Transcriptome (474)

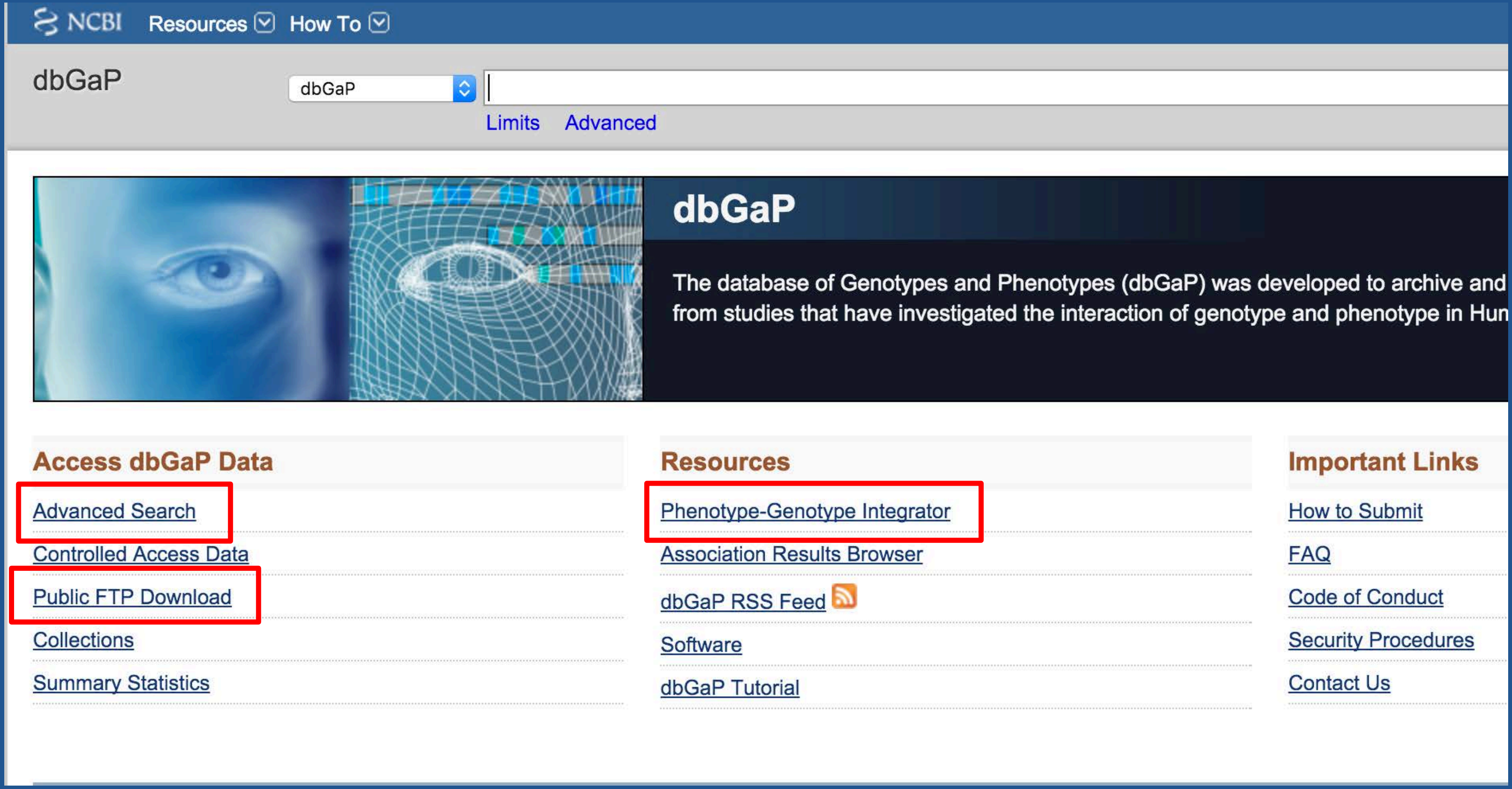
Project Data
Nucleotide (50)
Protein (32)
Assembly (44)
SRA (187)
GEO DataSets (507)

Scope
Monoisolate (112)
Multi-isolate (505)
Multi-species (21)
Environmental (58)
Other (10)

Reporting


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

dbGaP



dbGaP

The database of Genotypes and Phenotypes (dbGaP) was developed to archive and
from studies that have investigated the interaction of genotype and phenotype in Hun

Access dbGaP Data

- Advanced Search
- Controlled Access Data
- Public FTP Download
- Collections
- Summary Statistics

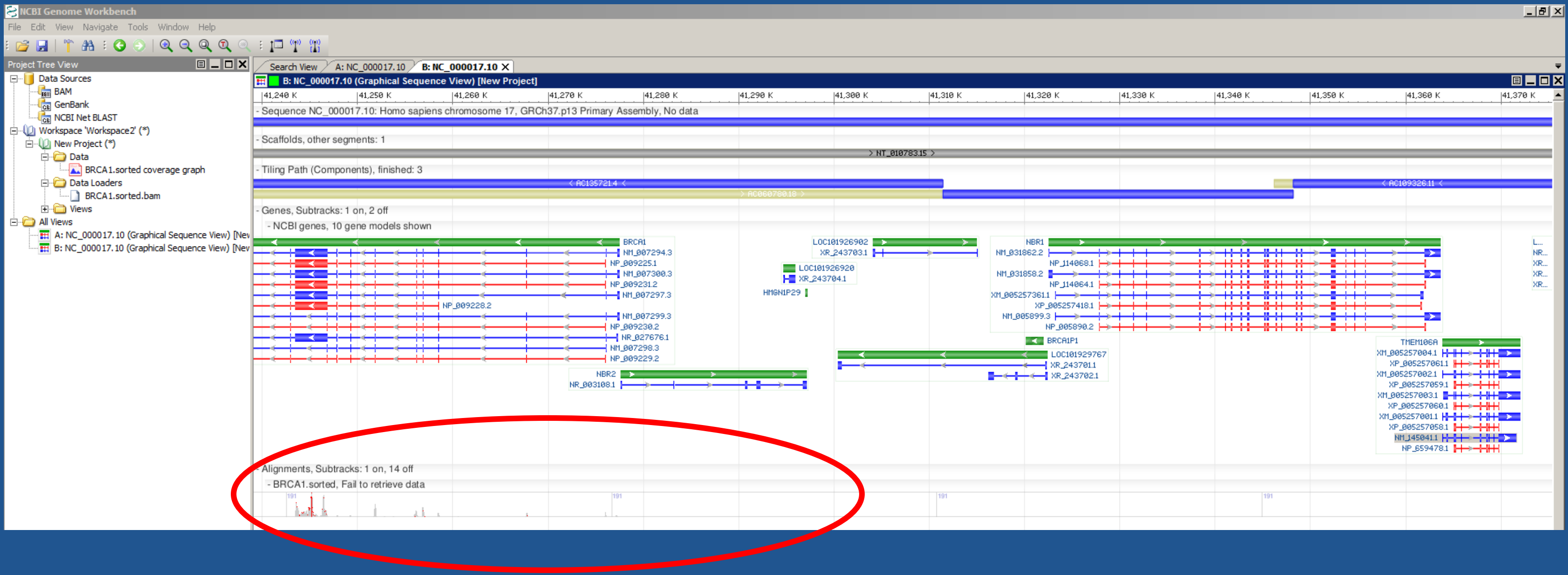
Resources

- Phenotype-Genotype Integrator
- Association Results Browser
- dbGaP RSS Feed
- Software
- dbGaP Tutorial

Important Links

- How to Submit
- FAQ
- Code of Conduct
- Security Procedures
- Contact Us

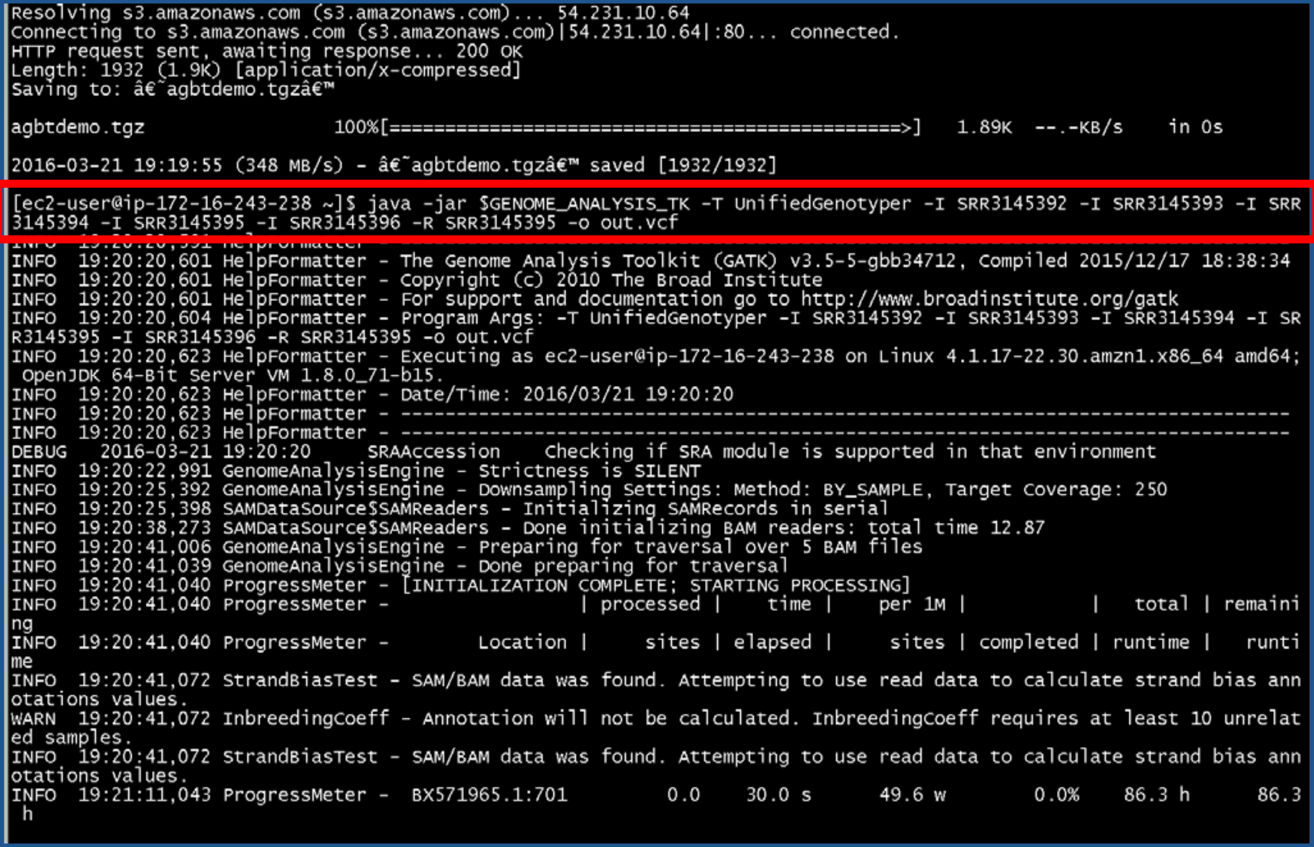
Minimizing Data Transfer



- Sequence NC_000017.10: Homo sapiens chromosome 17, GRCh37.p13 Primary Assembly, No data
- Scaffolds, other segments: 1
- Tiling Path (Components), finished: 3
- Genes, Subtracks: 1 on, 2 off
- NCBI genes, 10 gene models shown
- Alignments, Subtracks: 1 on, 14 off
- BRCA1.sorted, Fail to retrieve data

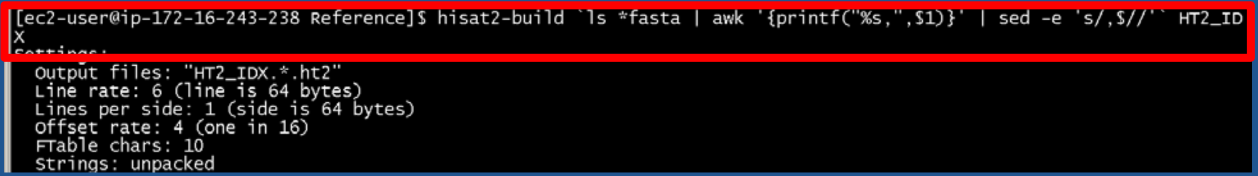
```
sam-dump.2.6.3 --aligned-region 17:41243452-41277500
SRR925743 > BRCA1.sam
```

Minimizing Data Transfer

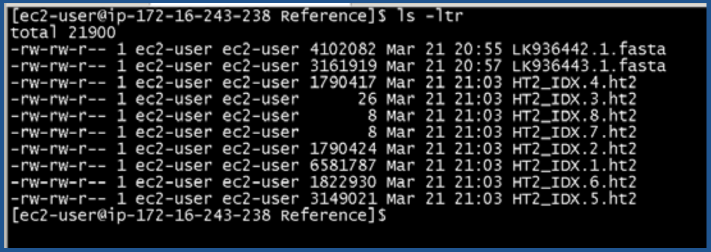


```
Resolving s3.amazonaws.com (s3.amazonaws.com)... 54.231.10.64
Connecting to s3.amazonaws.com (s3.amazonaws.com)|54.231.10.64|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 1932 (1.9K) [application/x-compressed]
Saving to: â€˜agbtdemo.tgzâ€™

agbtdemo.tgz                 100%[===============================================>]   1.89K  --.-KB/s    in 0s

2016-03-21 19:19:55 (348 MB/s) - â€˜agbtdemo.tgzâ€™ saved [1932/1932]

[ec2-user@ip-172-16-243-238 ~]$ java -jar $GENOME_ANALYSIS_TK -T UnifiedGenotyper -I SRR3145392 -I SRR3145393 -I SRR
3145394 -I SRR3145395 -I SRR3145396 -R SRR3145395 -o out.vcf
INFO  19:20:20,601 HelpFormatter - The Genome Analysis Toolkit (GATK) v3.5-5-gbb34712, Compiled 2015/12/17 18:38:34
INFO  19:20:20,601 HelpFormatter - Copyright (c) 2010 The Broad Institute
INFO  19:20:20,601 HelpFormatter - For support and documentation go to http://www.broadinstitute.org/gatk
INFO  19:20:20,604 HelpFormatter - Program Args: -T UnifiedGenotyper -I SRR3145392 -I SRR3145393 -I SRR3145394 -I SR
R3145395 -I SRR3145396 -R SRR3145395 -o out.vcf
INFO  19:20:20,623 HelpFormatter - Executing as ec2-user@ip-172-16-243-238 on Linux 4.1.17-22.30.amzn1.x86_64 amd64;
 OpenJDK 64-Bit Server VM 1.8.0_71-b15.
INFO  19:20:20,623 HelpFormatter - Date/Time: 2016/03/21 19:20:20
INFO  19:20:20,623 HelpFormatter - --------------------------------------------------------------------------------
INFO  19:20:20,623 HelpFormatter - --------------------------------------------------------------------------------
DEBUG   2016-03-21 19:20:20     SRAAccession    Checking if SRA module is supported in that environment
INFO  19:20:22,991 GenomeAnalysisEngine - Strictness is SILENT
INFO  19:20:25,392 GenomeAnalysisEngine - Downsampling Settings: Method: BY_SAMPLE, Target Coverage: 250
INFO  19:20:25,398 SAMDataSource$SAMReaders - Initializing SAMRecords in serial
INFO  19:20:38,273 SAMDataSource$SAMReaders - Done initializing BAM readers: total time 12.87
INFO  19:20:41,006 GenomeAnalysisEngine - Preparing for traversal over 5 BAM files
INFO  19:20:41,039 GenomeAnalysisEngine - Done preparing for traversal
INFO  19:20:41,040 ProgressMeter - [INITIALIZATION COMPLETE; STARTING PROCESSING]
INFO  19:20:41,040 ProgressMeter -                 | processed |    time |    per 1M |           |   total | remaini
ng
INFO  19:20:41,040 ProgressMeter -        Location |     sites | elapsed |     sites | completed | runtime |   runti
me
INFO  19:20:41,072 StrandBiasTest - SAM/BAM data was found. Attempting to use read data to calculate strand bias ann
otations values.
WARN  19:20:41,072 InbreedingCoeff - Annotation will not be calculated. InbreedingCoeff requires at least 10 unrelat
ed samples.
INFO  19:20:41,072 StrandBiasTest - SAM/BAM data was found. Attempting to use read data to calculate strand bias ann
otations values.
INFO  19:21:11,043 ProgressMeter -  BX571965.1:701        0.0    30.0 s      49.6 w        0.0%    86.3 h      86.3
 h
```

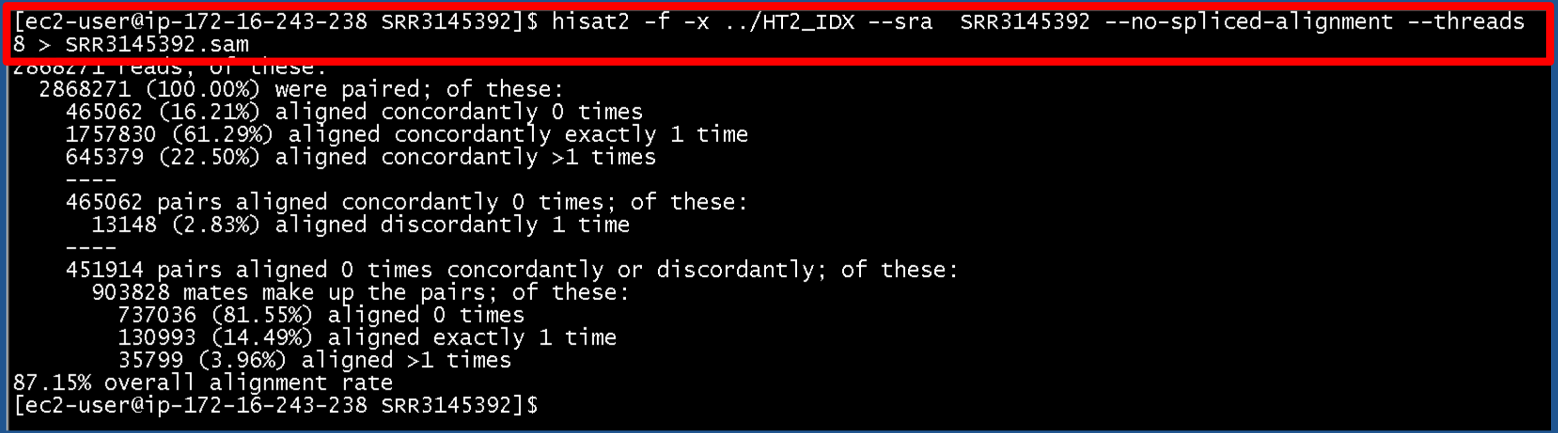
Minimizing Data Transfer



```
[ec2-user@ip-172-16-243-238 Reference]$ hisat2-build `ls *fasta | awk '{printf("%s,",$1)}' | sed -e 's/,$//'` HT2_ID
X
Settings:
  Output files: "HT2_IDX.*.ht2"
  Line rate: 6 (line is 64 bytes)
  Lines per side: 1 (side is 64 bytes)
  Offset rate: 4 (one in 16)
  FTable chars: 10
  Strings: unpacked
```

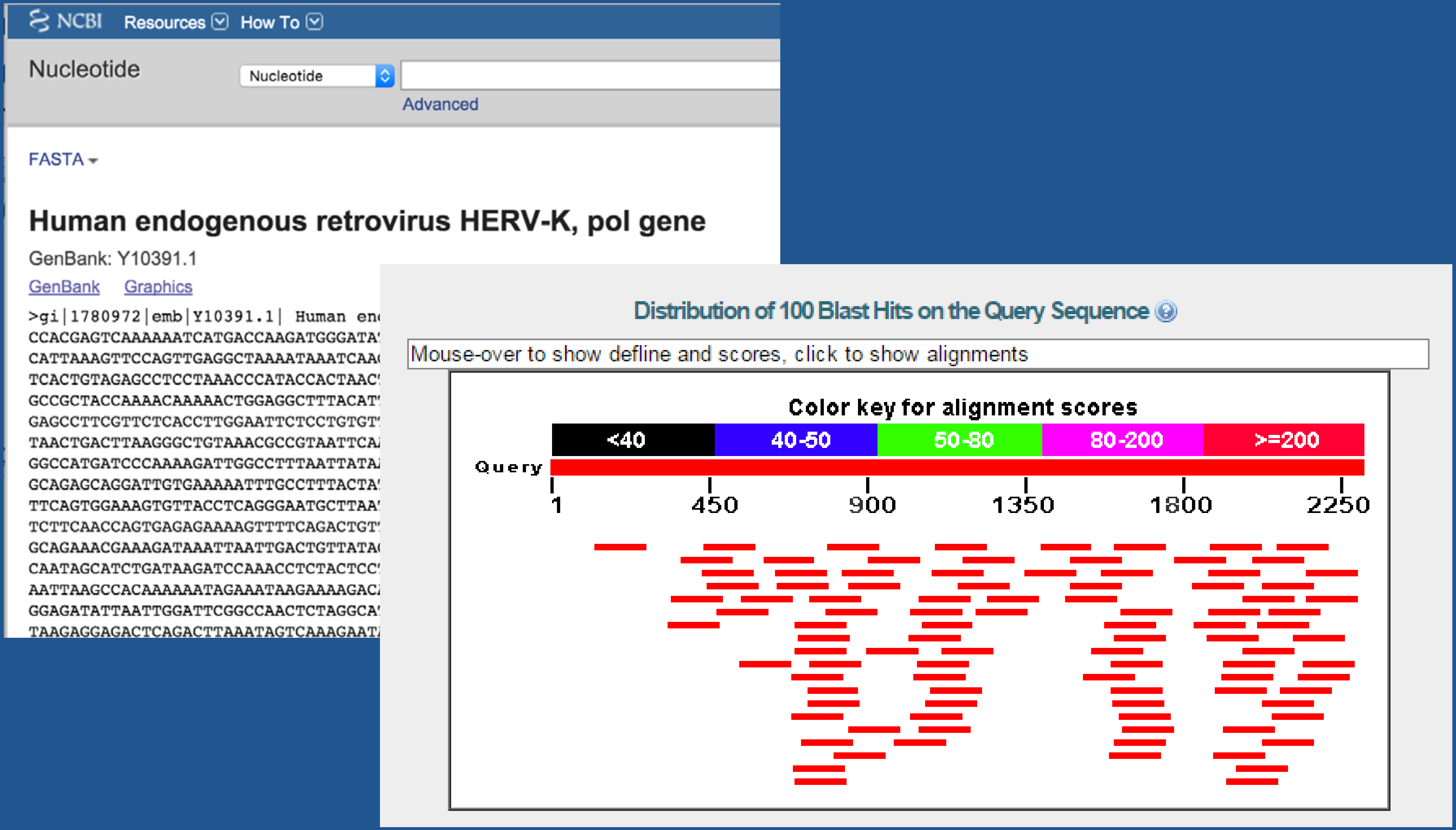


```
[ec2-user@ip-172-16-243-238 Reference]$ ls -ltr
total 21900
-rw-rw-r-- 1 ec2-user ec2-user 4102082 Mar 21 20:55 LK936442.1.fasta
-rw-rw-r-- 1 ec2-user ec2-user 3161919 Mar 21 20:57 LK936443.1.fasta
-rw-rw-r-- 1 ec2-user ec2-user 1790417 Mar 21 21:03 HT2_IDX.4.ht2
-rw-rw-r-- 1 ec2-user ec2-user      26 Mar 21 21:03 HT2_IDX.3.ht2
-rw-rw-r-- 1 ec2-user ec2-user       8 Mar 21 21:03 HT2_IDX.8.ht2
-rw-rw-r-- 1 ec2-user ec2-user       8 Mar 21 21:03 HT2_IDX.7.ht2
-rw-rw-r-- 1 ec2-user ec2-user 1790424 Mar 21 21:03 HT2_IDX.2.ht2
-rw-rw-r-- 1 ec2-user ec2-user 6581787 Mar 21 21:03 HT2_IDX.1.ht2
-rw-rw-r-- 1 ec2-user ec2-user 1822930 Mar 21 21:03 HT2_IDX.6.ht2
-rw-rw-r-- 1 ec2-user ec2-user 3149021 Mar 21 21:03 HT2_IDX.5.ht2
[ec2-user@ip-172-16-243-238 Reference]$
```



```
[ec2-user@ip-172-16-243-238 SRR3145392]$ hisat2 -f -x ../HT2_IDX --sra  SRR3145392 --no-spliced-alignment --threads
8 > SRR3145392.sam
2868271 reads; of these:
  2868271 (100.00%) were paired; of these:
    465062 (16.21%) aligned concordantly 0 times
    1757830 (61.29%) aligned concordantly exactly 1 time
    645379 (22.50%) aligned concordantly >1 times
    ----
    465062 pairs aligned concordantly 0 times; of these:
      13148 (2.83%) aligned discordantly 1 time
    ----
    451914 pairs aligned 0 times concordantly or discordantly; of these:
      903828 mates make up the pairs; of these:
        737036 (81.55%) aligned 0 times
        130993 (14.49%) aligned exactly 1 time
        35799 (3.96%) aligned >1 times
87.15% overall alignment rate
[ec2-user@ip-172-16-243-238 SRR3145392]$
```

Minimizing Data Transfer



Human endogenous retrovirus HERV-K, pol gene

GenBank: Y10391.1

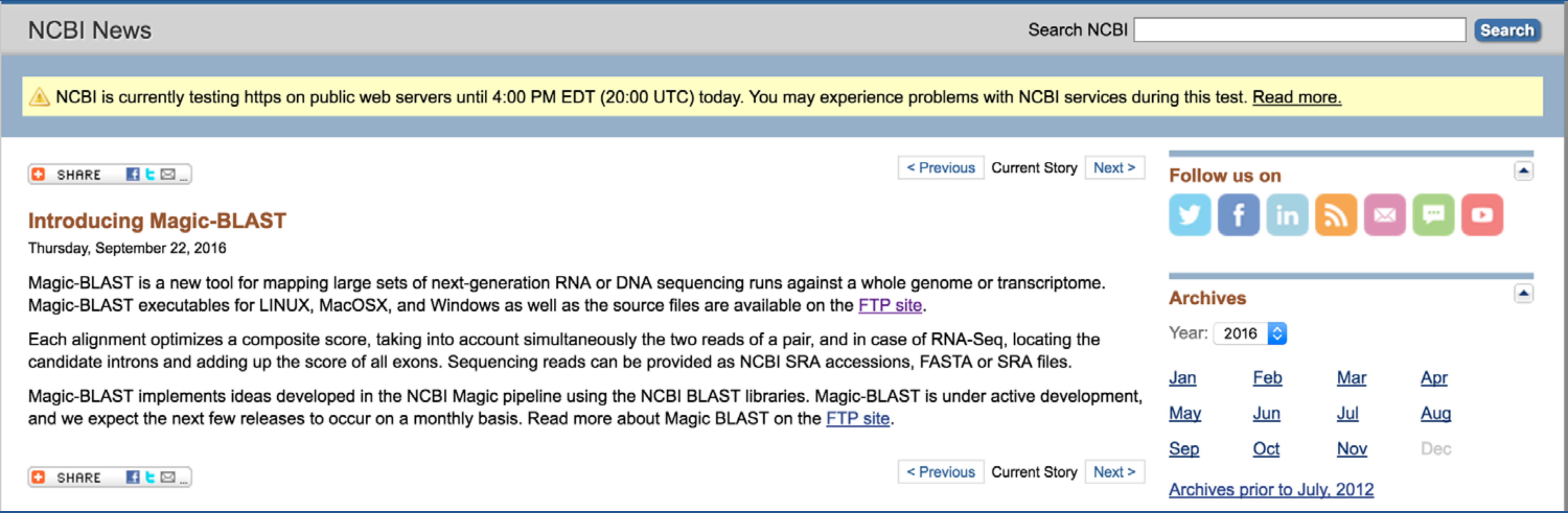
```
>gi|1780972|emb|Y10391.1| Human en
CCACGAGTCAAAAAATCATGACCAAGATGGGATA
CATTAAAGTTCCAGTTGAGGCTAAAATAAATCAA
TCACTGTAGAGCCTCCTAAACCCATACCACTAAC
GCCGCTACCAAAACAAAAACTGGAGGCTTTACAT
GAGCCTTCGTTCTCACCTTGGAATTCTCCTGTGT
TAACTGACTTAAGGGCTGTAAACGCCGTAATTCA
GGCCATGATCCCAAAAGATTGGCCTTTAATTATA
GCAGAGCAGGATTGTGAAAAATTTGCCTTTACTA
TTCAGTGGAAAGTGTTACCTCAGGGAATGCTTAA
TCTTCAACCAGTGAGAGAAAAGTTTTCAGACTGT
GCAGAAACGAAAGATAAATTAATTGACTGTTATA
CAATAGCATCTGATAAGATCCAAACCTCTACTCC
AATTAAGCCACAAAAAAATAGAAATAAGAAAAGAC
GGAGATATTAATTGGATTCGGCCAACTCTAGGCA
TAAGAGGAGACTCAGACTTAAATAGTCAAAGAAT
```

Distribution of 100 Blast Hits on the Query Sequence

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

Color key for alignment scores

Minimizing Data Transfer



NCBI News

Search NCBI Search

NCBI is currently testing https on public web servers until 4:00 PM EDT (20:00 UTC) today. You may experience problems with NCBI services during this test. Read more.

SHARE

< Previous Current Story Next >

Introducing Magic-BLAST

Thursday, September 22, 2016

Magic-BLAST is a new tool for mapping large sets of next-generation RNA or DNA sequencing runs against a whole genome or transcriptome. Magic-BLAST executables for LINUX, MacOSX, and Windows as well as the source files are available on the FTP site.

Each alignment optimizes a composite score, taking into account simultaneously the two reads of a pair, and in case of RNA-Seq, locating the candidate introns and adding up the score of all exons. Sequencing reads can be provided as NCBI SRA accessions, FASTA or SRA files.

Magic-BLAST implements ideas developed in the NCBI Magic pipeline using the NCBI BLAST libraries. Magic-BLAST is under active development, and we expect the next few releases to occur on a monthly basis. Read more about Magic BLAST on the FTP site.

SHARE

< Previous Current Story Next >

Follow us on

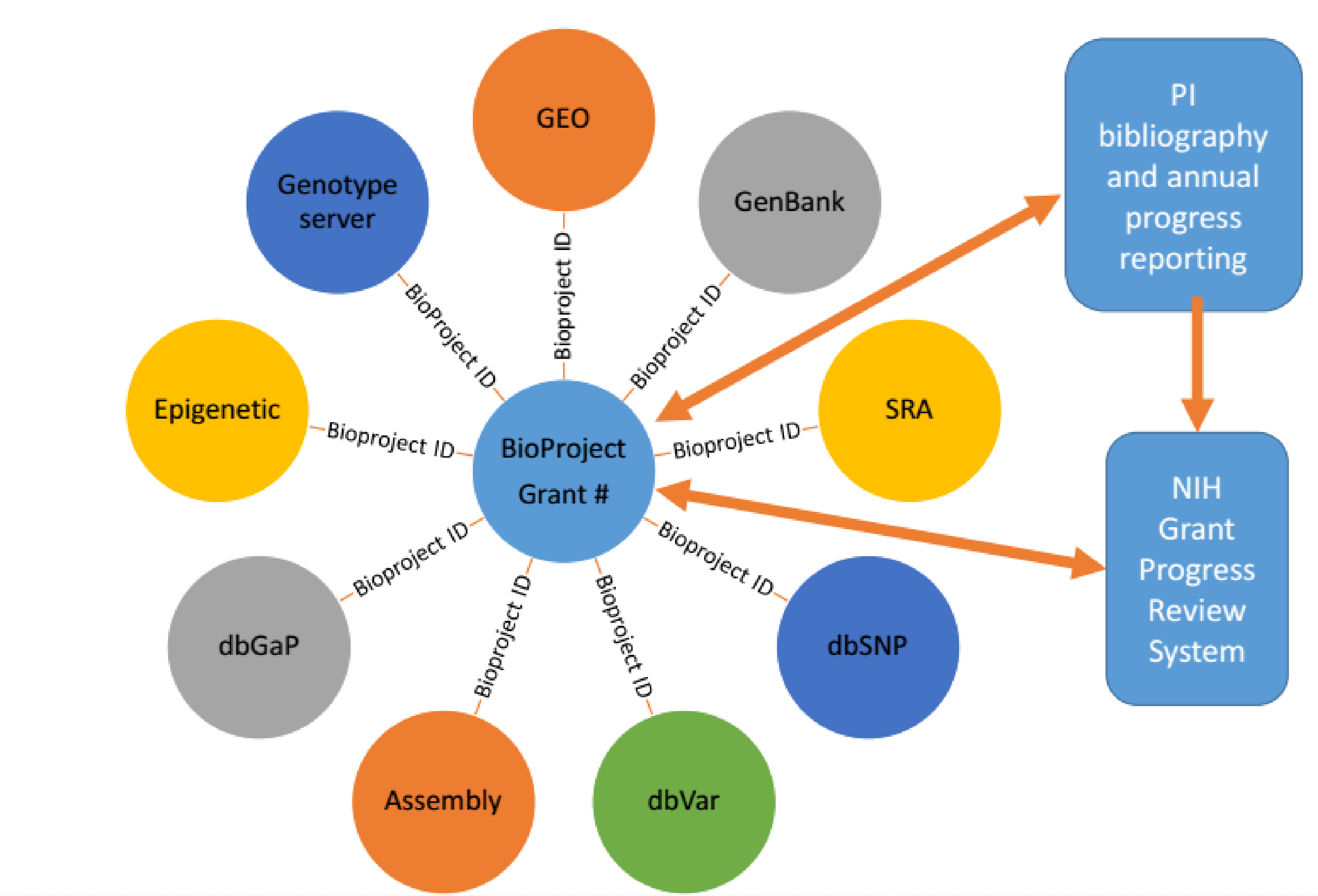
Archives

Year: 2016

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

Archives prior to July, 2012

Reporting


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

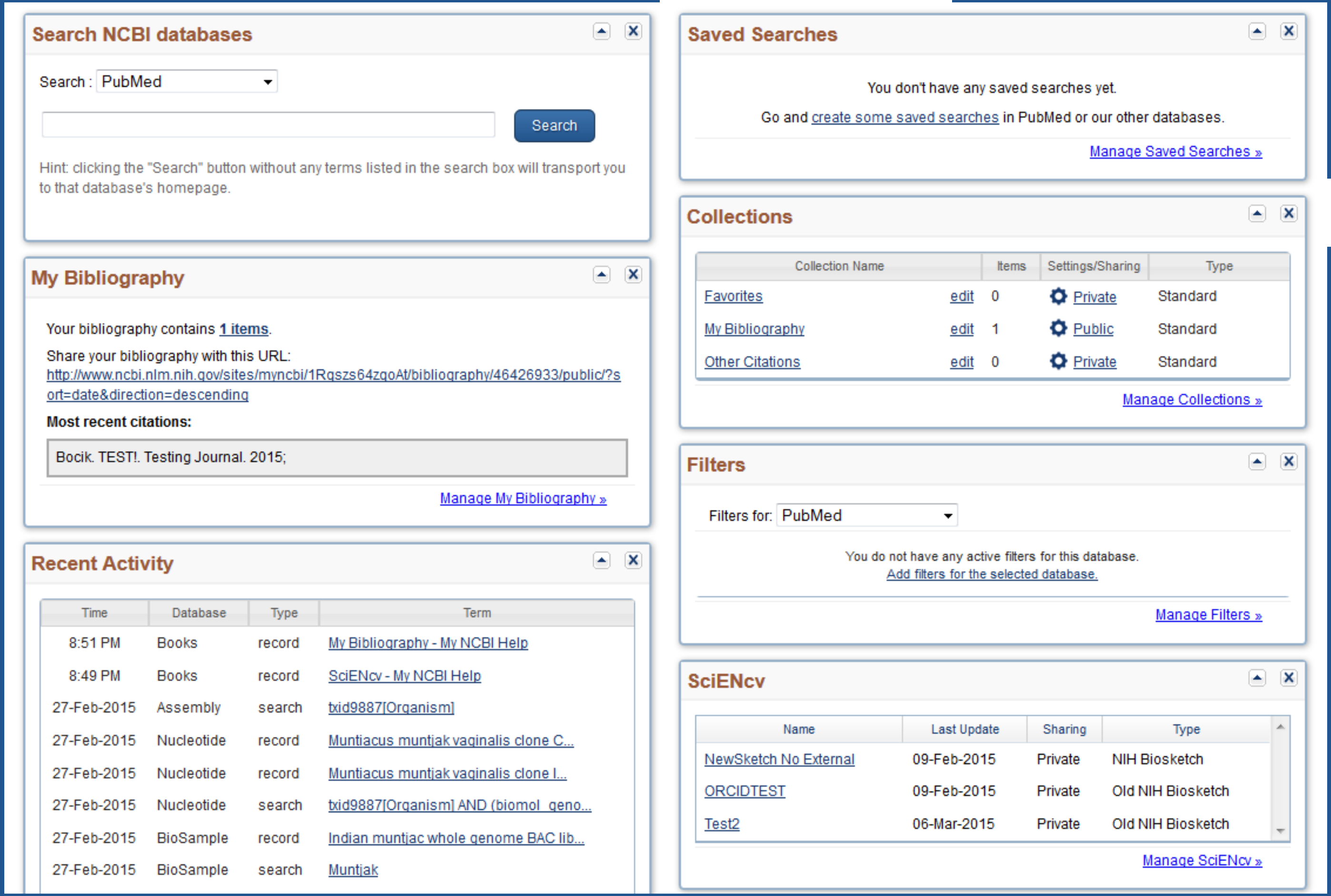
NCBI

Saved Searches

My Bibliography

Collections

SciENcv



Search NCBI databases

Search : PubMed

Search

Hint: clicking the "Search" button without any terms listed in the search box will transport you to that database's homepage.

My Bibliography

Your bibliography contains 1 items.

Share your bibliography with this URL:
http://www.ncbi.nlm.nih.gov/sites/myncbi/1Rqszs64zqoAt/bibliography/46426933/public/?sort=date&direction=descending

Most recent citations:

Bocik. TEST!. Testing Journal. 2015;

Manage My Bibliography »

Recent Activity

Time	Database	Type	Term
8:51 PM	Books	record	My Bibliography - My NCBI Help
8:49 PM	Books	record	SciENcv - My NCBI Help
27-Feb-2015	Assembly	search	txid9887[Organism]
27-Feb-2015	Nucleotide	record	Muntiacus muntjak vaginalis clone C...
27-Feb-2015	Nucleotide	record	Muntiacus muntjak vaginalis clone I...
27-Feb-2015	Nucleotide	search	txid9887[Organism] AND (biomol_geno...
27-Feb-2015	BioSample	record	Indian muntjac whole genome BAC lib...
27-Feb-2015	BioSample	search	Muntjak

Saved Searches

You don't have any saved searches yet.

Go and create some saved searches in PubMed or our other databases.

Manage Saved Searches »

Collections

Collection Name		Items	Settings/Sharing	Type
Favorites	edit	0	Private	Standard
My Bibliography	edit	1	Public	Standard
Other Citations	edit	0	Private	Standard

Manage Collections »

Filters

Filters for: PubMed

You do not have any active filters for this database.
Add filters for the selected database.

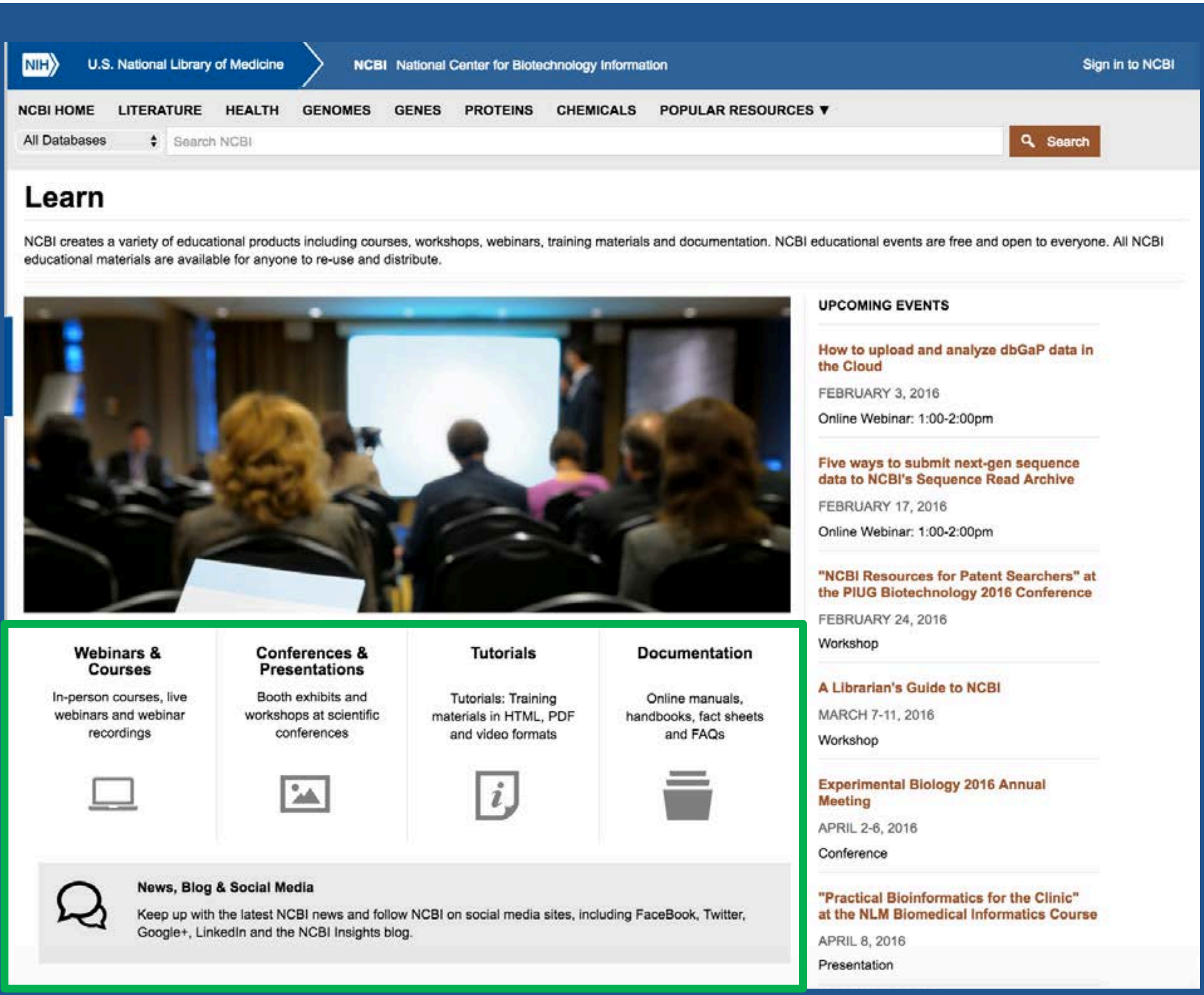
Manage Filters »

SciENcv

Name	Last Update	Sharing	Type
NewSketch No External	09-Feb-2015	Private	NIH Biosketch
ORCIDTEST	09-Feb-2015	Private	Old NIH Biosketch
Test2	06-Mar-2015	Private	Old NIH Biosketch

Manage SciENcv »

For more information go to: 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.

Webinars & Courses
In-person courses, live webinars and webinar recordings

Conferences & Presentations
Booth exhibits and workshops at scientific conferences

Tutorials
Tutorials: Training materials in HTML, PDF and video formats

Documentation
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.

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