

# REPOSITORY

[RE-POZ-i-TIV], NOUN;

1. A POSITIVE EXPERIENCE OF ACCESSING  
HUMAN GENOMIC DATA REPOSITORIES



Charlotte Whicher  
[charlotte@repositive.io](mailto:charlotte@repositive.io)

An overview of (human) genomic data sources

# What is genomic data

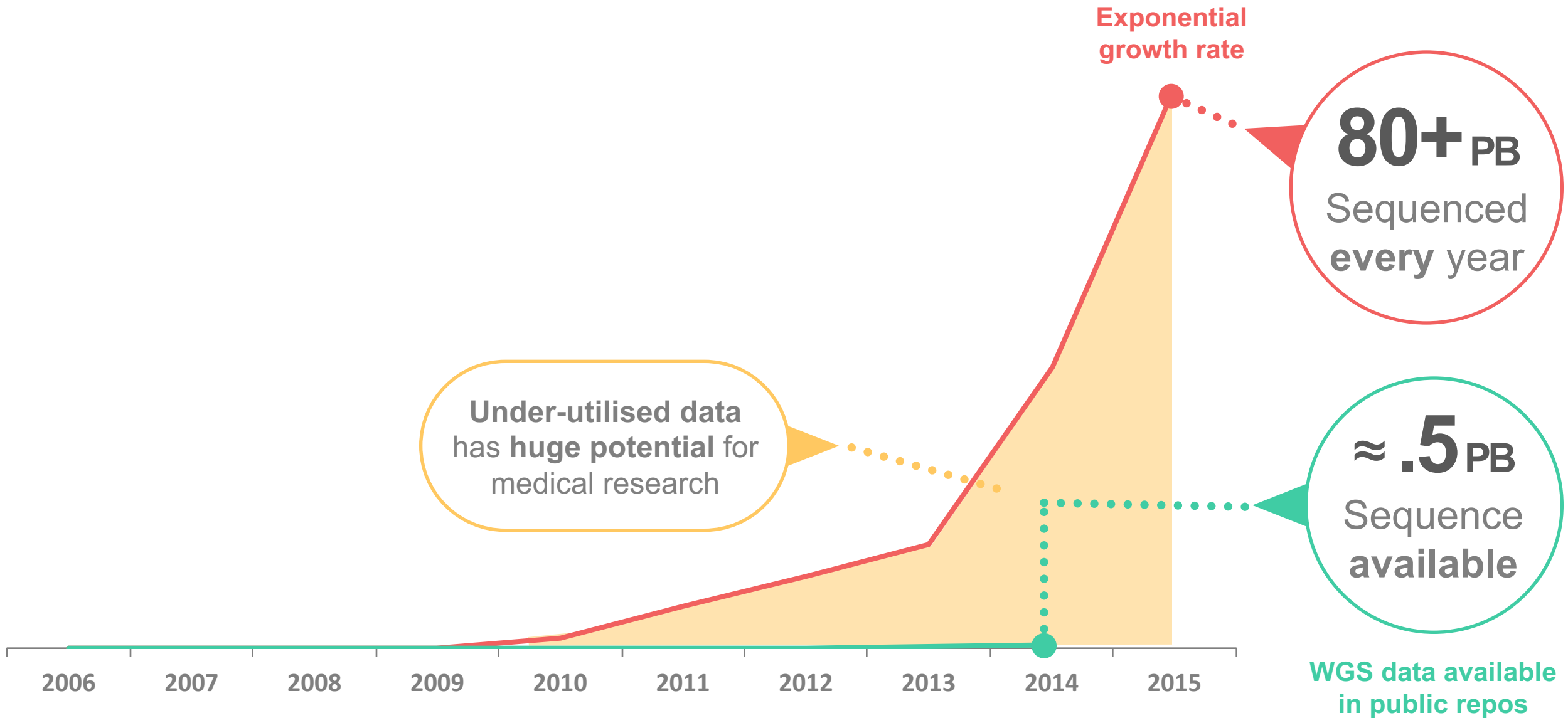
Genomics is a discipline in genetics.

Genomes are the complete set of DNA within a single cell of an organism.

Genomics applies array, next generation sequencing methods, and bioinformatics to sequence, assemble, and analyse the function and structure of genomes.



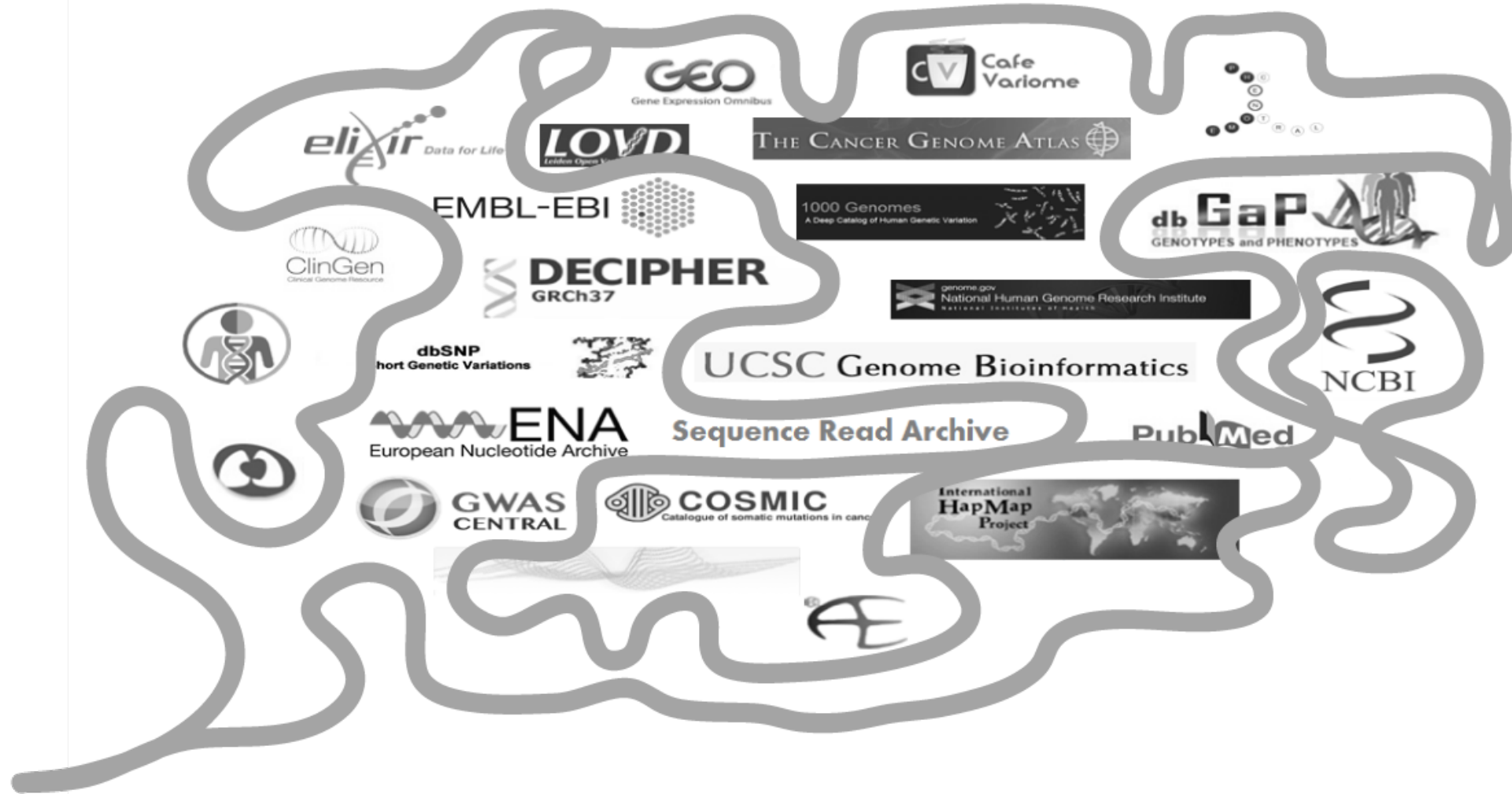
# Large amounts of Data, but not accessible



# Data is fragmented

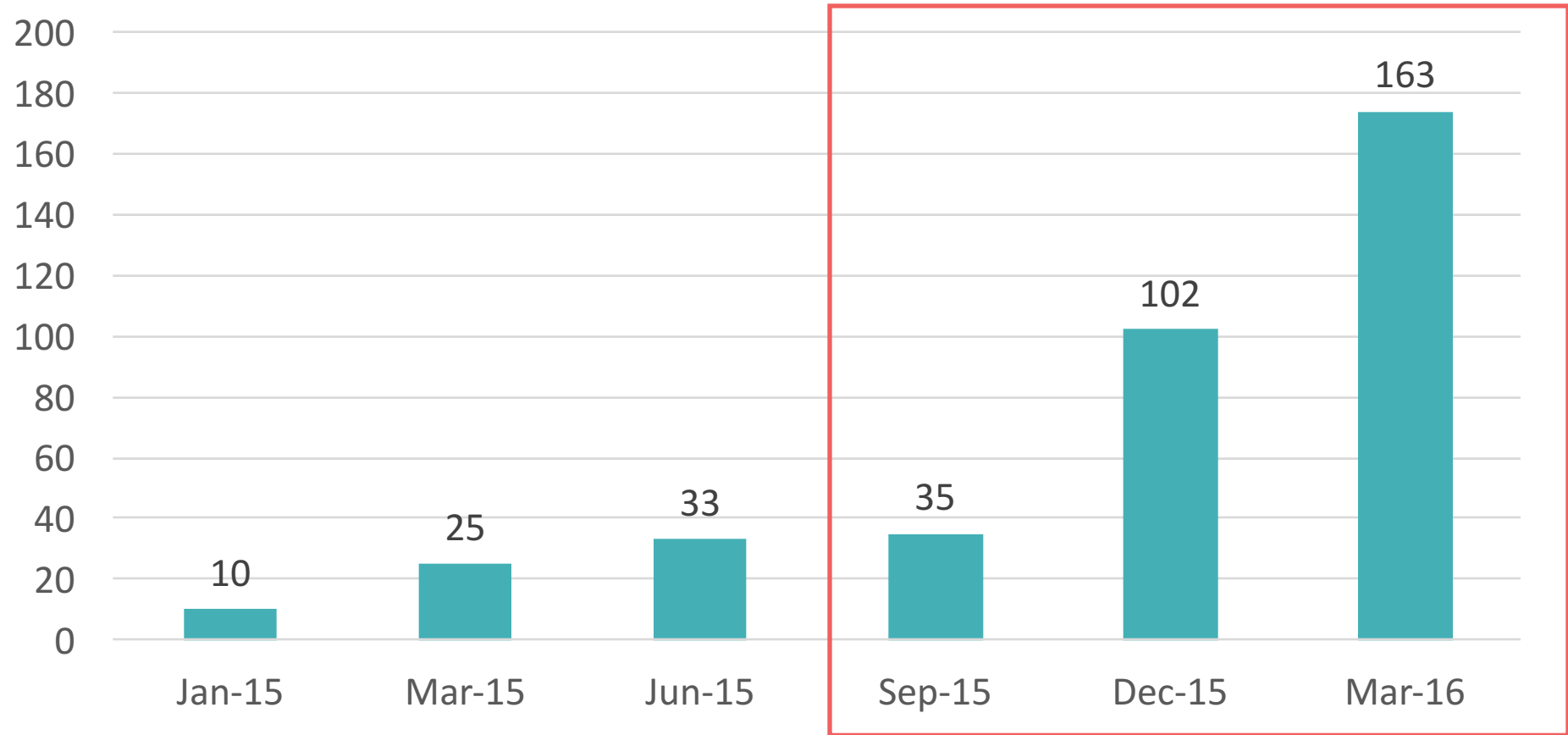


# It may be confusing



# Hundreds of data sources

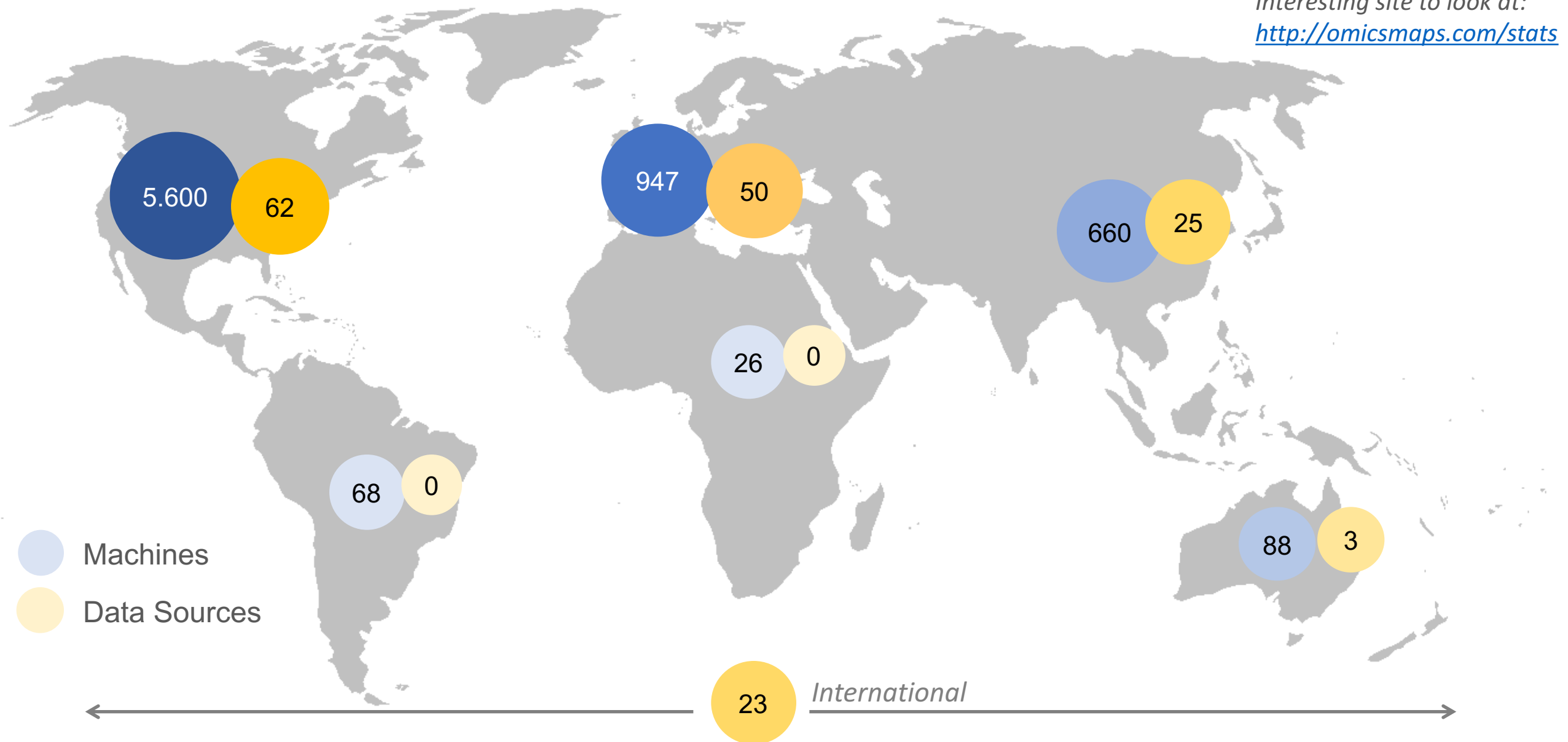
...but they aren't easy to find!



...at least 6 months work.

# Machines & Data sources

Interesting site to look at:  
<http://omicsmaps.com/stats>



# Main Data Sources (Repository) funders



EBI total 104 services, 19 repositories <http://www.ebi.ac.uk/services/all>

NCBI total 67 databases [http://www.ncbi.nlm.nih.gov/guide/all/#databases\\_](http://www.ncbi.nlm.nih.gov/guide/all/#databases_)



# Data accessibility

Has both Open and  
Restricted access data  
within one repository.

**Both**  
**17%**

**Restricted**

**38%**

Need to apply for  
access to the data.

■ Restricted

■ Unclear

■ Open

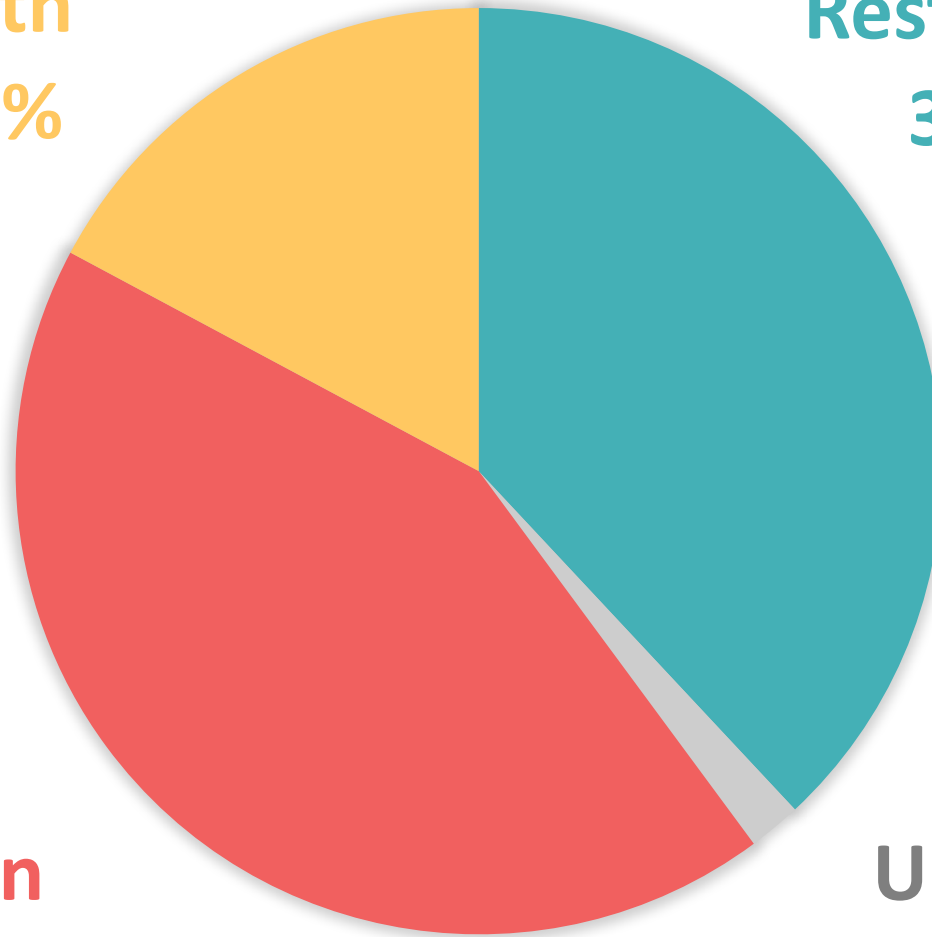
■ Both

Can download the  
data straight away  
or after logging in.

**Open**  
**43%**

**Unclear**

**2%**

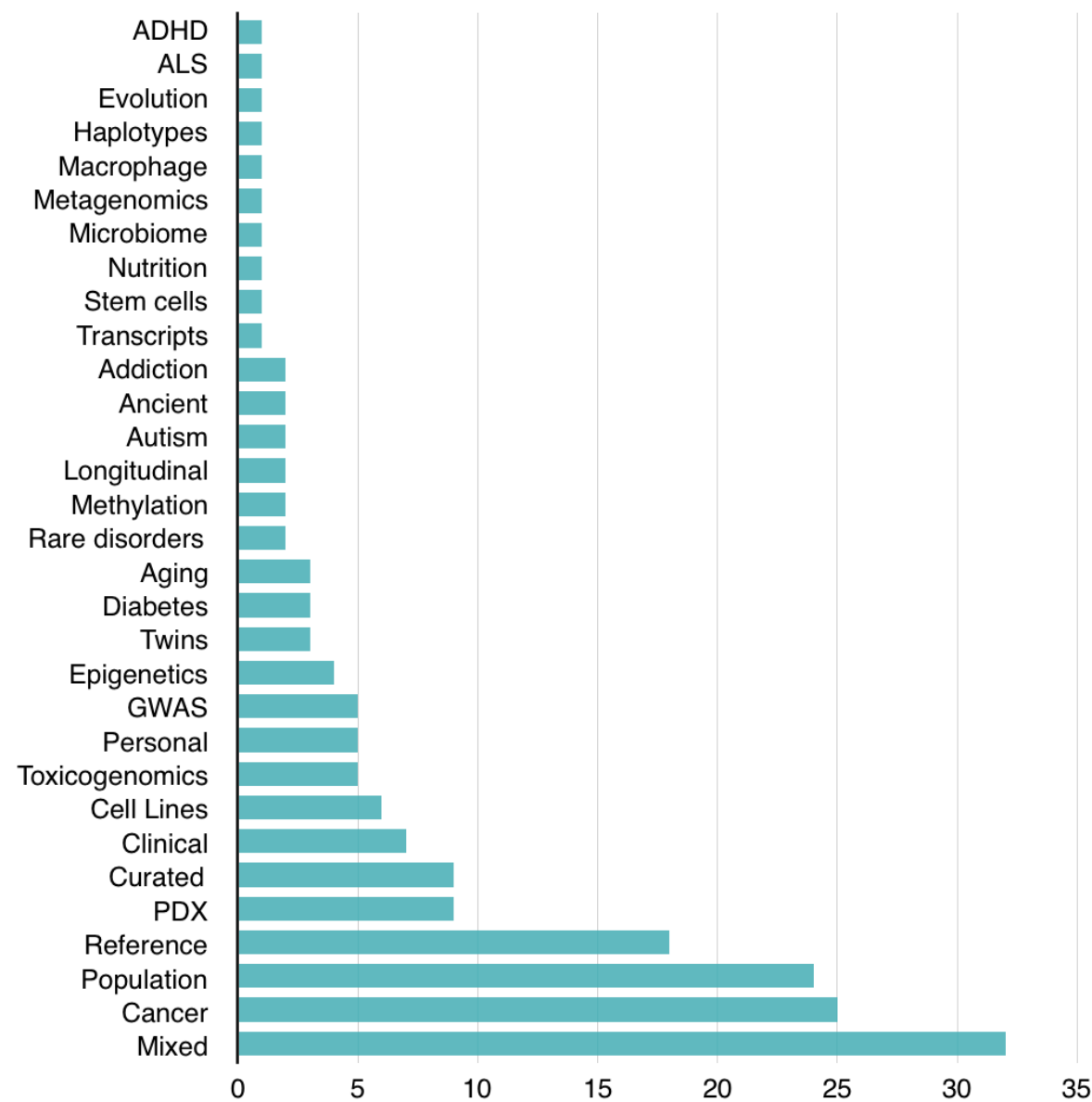


# Sequenced ethnicities

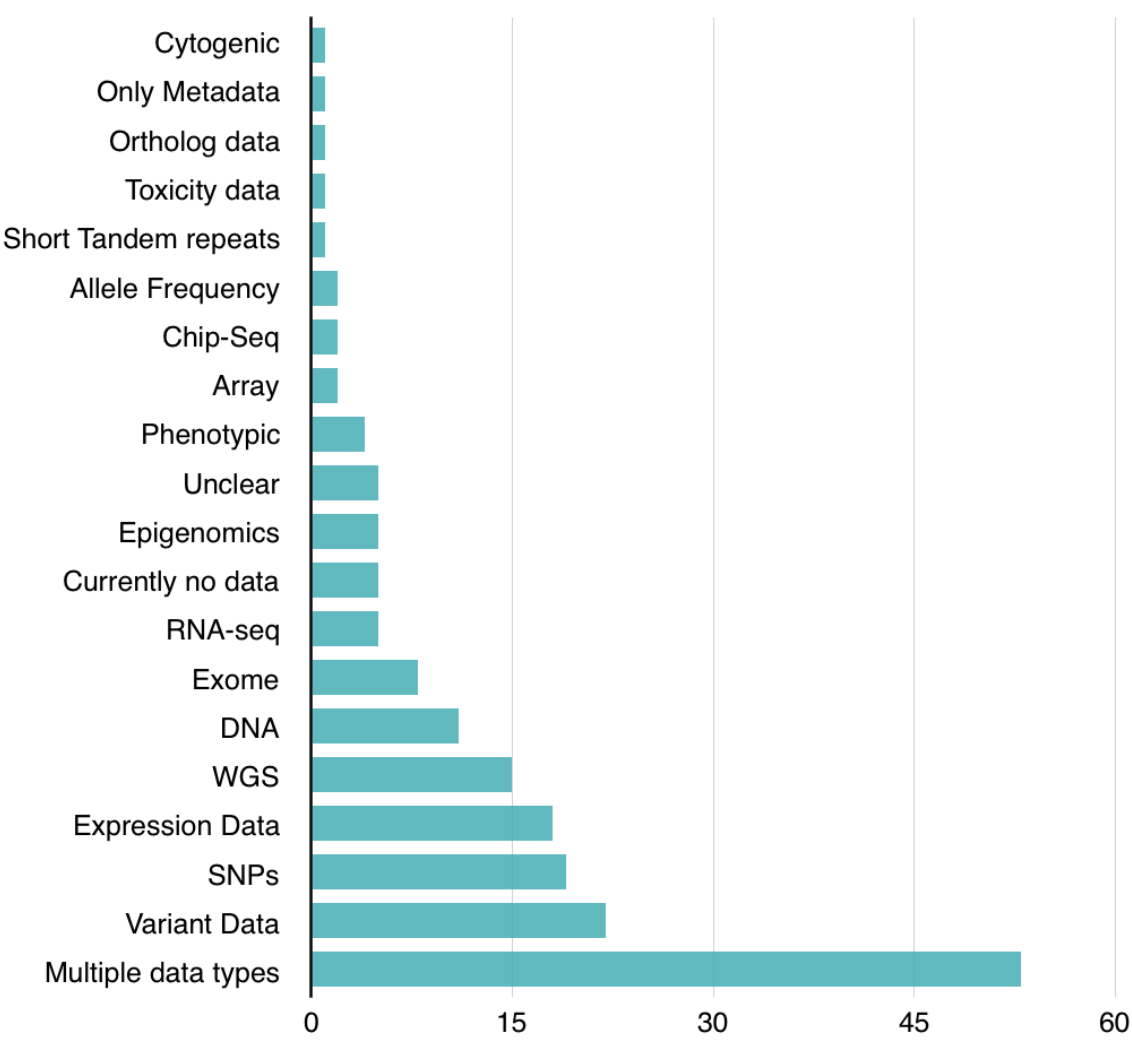


# Data source content

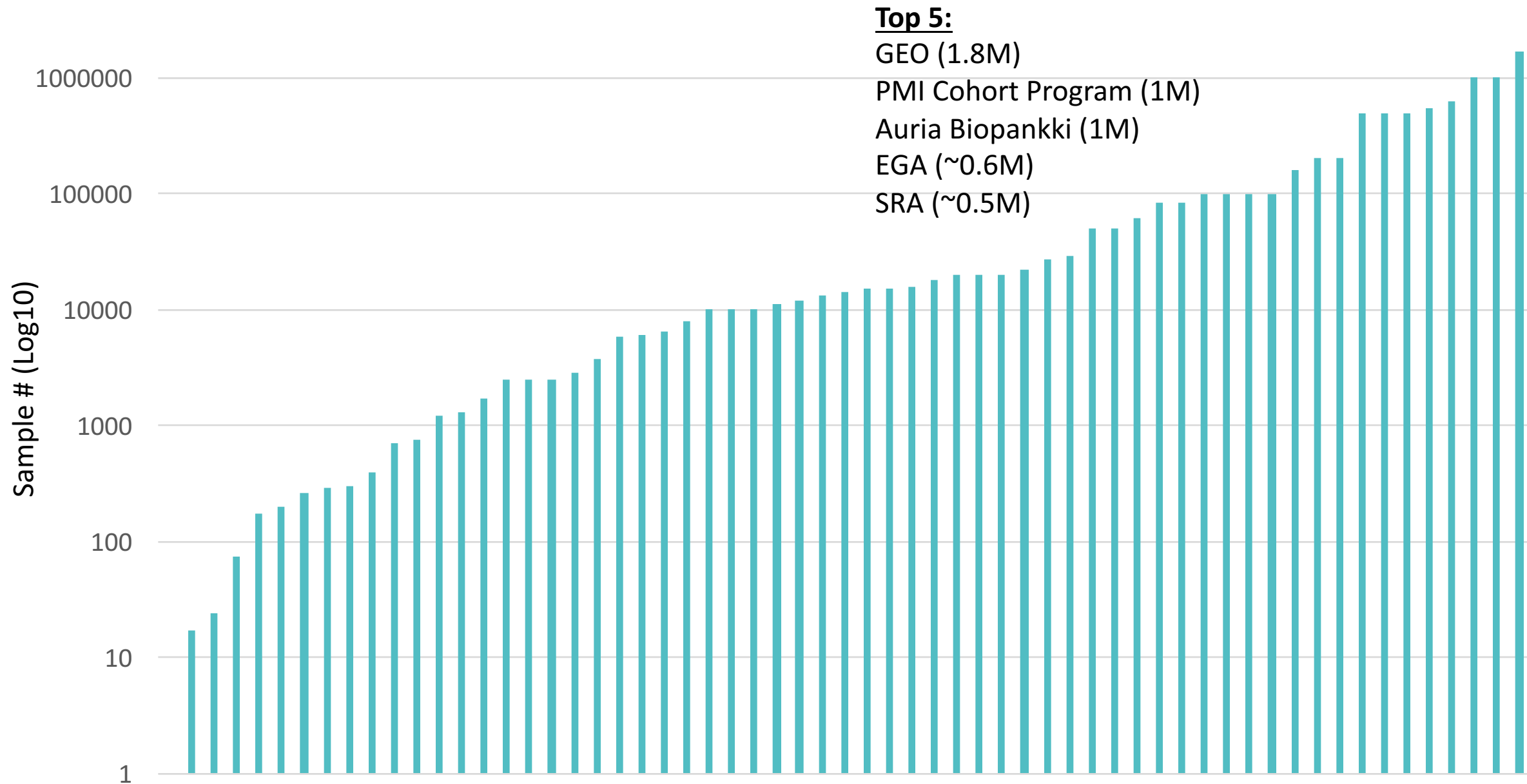
Dedicated to...



Assay Types



# Number of samples in Data sources



# Online Data source 'types'

University – Affiliated to a university. Often only members of that university can upload/download to/from it.

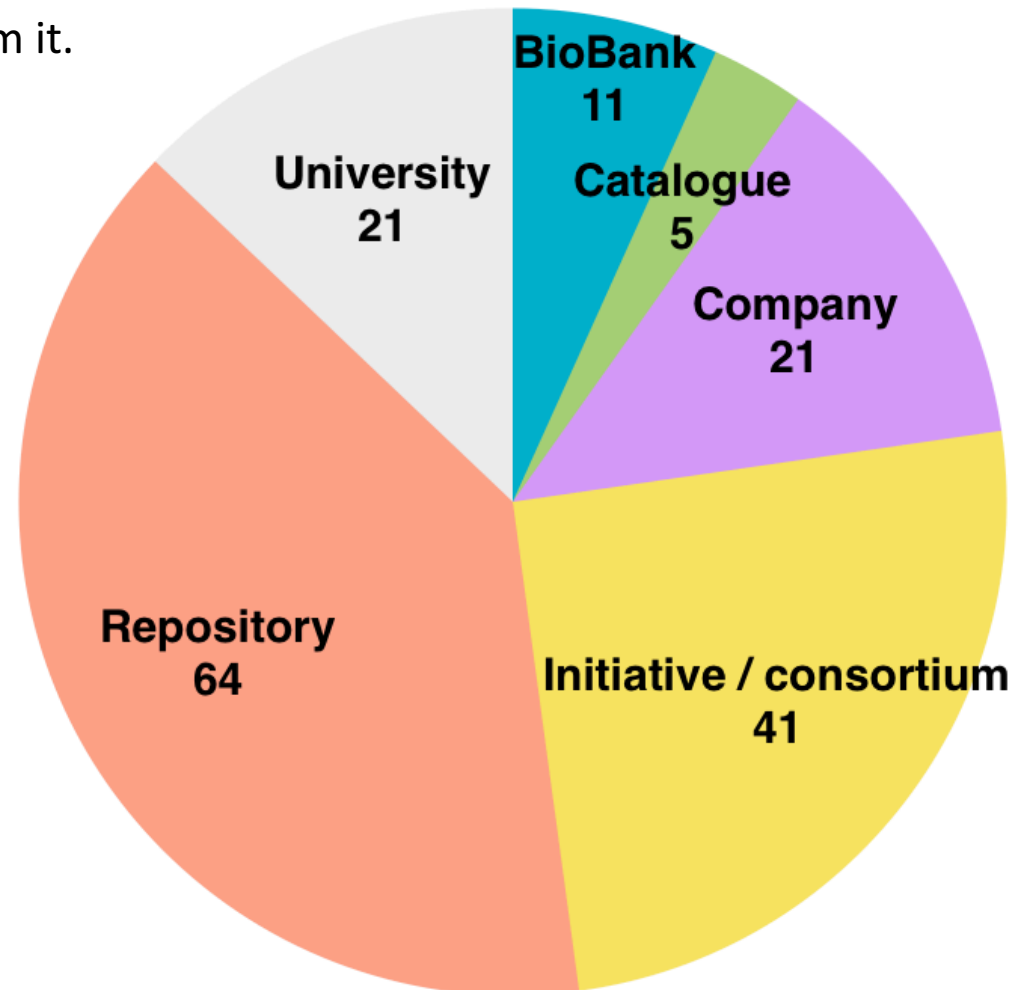
Biobank – many have sequence data of their biological samples.

Catalogue – doesn't have raw data but lists studies/datasets.

Company – For profit organisation. Listing data is not their main purpose.

Repository – Can download from, has data from multiple institutions. Often can also upload your own data there.

Initiative/Consortium – Has a specific purpose/aim. Often focussed on a question or disease.



# Access to NCBI data

The screenshot displays the NCBI (National Center for Biotechnology Information) homepage. At the top, the browser address bar shows [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov). The NCBI logo and navigation links (Resources, How To) are visible. A search bar is highlighted with a red box, containing the text "Genome" and "arthropoda or rodentia". The main content area features a "Welcome to NCBI" message and six primary service categories: Submit, Download, Learn, Develop, Analyze, and Research, each with a descriptive text and an icon. A left sidebar lists various resources like "Resource List (A-Z)", "Chemicals & Bioassays", and "DNA & RNA". On the right, there are sections for "Popular Resources" (including PubMed, Bookshelf, etc.) and "NCBI Announcements" with recent webinar dates. The footer includes the current page path, a link to the Help Desk, and navigation tabs for Popular, Featured, and NCBI Information.

NCBI National Center for Biotechnology Information

Search: Genome arthropoda or rodentia

**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](#) | [NCBI News](#)

**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 Announcements**  
April 8th webinar: "The NCBI Minute: Introducing MOLE-BLAST" Mar 25, 2015  
April 1st webinar: "A Practical Guide to Using NCBI BLAST on the Web" Mar 24, 2015  
dbSNP Build 143 Phase II now available Mar 17, 2015  
[More...](#)

You are here: NCBI > National Center for Biotechnology Information

POPULAR FEATURED NCBI INFORMATION

# Access to NCBI data

The screenshot shows the NCBI Genome database search results for the query "arthropoda or rodentia". The search results are displayed in a table with columns for organism name, overview, and details. The results are sorted by relevance, and the first four results are highlighted with a red box. The search results are as follows:

| Organism                           | Overview  | Details   |
|------------------------------------|---|---|
| <a href="#">Mus musculus</a>       | 1. The laboratory mouse is a major model organism for basic mammalian biology, human disease, and genome evolution, and its genome has been sequenced | Kingdom: Eukaryota; Subgroup: Mammals<br>Sequence data: genome assemblies:7<br>Haploid chromosomes: 21; Organelles: 1<br>Date: 2002/05/15<br>ID: 52 |
| <a href="#">Locusta migratoria</a> | 2. Locusta migratoria overview  | Kingdom: Eukaryota; Subgroup: Insects<br>Sequence data: genome assemblies:1<br>Organelles: 1<br>Date: 2013/12/18<br>ID: 10772                       |
| <a href="#">Anopheles gambiae</a>  | 3. The mosquito Anopheles gambiae is the principal vector of malaria  | Kingdom: Eukaryota; Subgroup: Insects<br>Sequence data: genome assemblies:3<br>Haploid chromosomes: 5; Organelles: 1<br>Date: 2002/03/22<br>ID: 46  |
| <a href="#">Marmota marmota</a>    | 4. Marmota marmota overview   | Kingdom: Eukaryota; Subgroup: Mammals<br>Sequence data: genome assemblies:1<br>Date: 2015/02/13<br>ID: 36409  |

The search results are displayed on page 1 of 9. The search criteria are "arthropoda or rodentia". The search results are sorted by relevance. The search results are displayed in a table with columns for organism name, overview, and details. The results are sorted by relevance, and the first four results are highlighted with a red box. The search results are as follows:

Results: 1 to 20 of 167

See also 1006 organelle- and plasmid-only records matching your search

Find related data

Database: Select

Find items

Search details

"Arthropoda"[Organism] OR "Rodentia"[Organism]

Search

See more...

Recent activity

Turn Off Clear

arthropoda OR rodentia (167)

Genome

solenopsis invicta AND ("biomol dna" [Properties] AND "strategy wg..." (30)

SRA

solenopsis invicta AND ("biomol dna" [Properties]) (42)

SRA

A new species of Mengenilla (Insecta, Strepsiptera) from Tunisia

Ladona fulva

BioProject

See more...

# Access to NCBI data

**Harpegnathos saltator genome sequencing project**

**Project Type:** Genome sequencing; **Locus Tag Prefix:** EAI

**Attributes:** Scope: Monoisolate; Material: Genome; Capture: Whole; Method type: Sequencing

**Project Data:**

| Resource Name      | Number of Links |
|--------------------|-----------------|
| SEQUENCE DATA      |                 |
| Nucleotide (total) | 8894            |
| WGS master         | 1               |
| SRA Experiments    | 43              |
| Protein Sequences  | 15127           |
| PUBLICATIONS       |                 |
| PubMed             | 1               |
| PMC                | 1               |
| OTHER DATASETS     |                 |
| BioSample          | 2               |
| Assembly           | 1               |

**Assembly details:**

| Assembly        | Level    | WGS           | BioSample    | Taxonomy              |
|-----------------|----------|---------------|--------------|-----------------------|
| GCA_000147195.1 | Scaffold | AEAC000000000 | SAMN00016742 | Harpegnathos saltator |

**SRA Data Details**

| Parameter           | Value |
|---------------------|-------|
| Data volume, Gbases | 101   |
| Data volume, Mbytes | 71887 |

**Related RefSeq Project**

PRJNA273397 : Harpegnathos saltator strain:R22 G/1 RefSeq Genome and RefSeq Transcriptome or Gene expression

**Publications:**

1. Bonasio R *et al.*, "Genomic comparison of the ants Camponotus floridanus and Harpegnathos saltator.", *Science*, 2010 Aug 27;329(5995):1068-71

**Lineage:** Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Vespoidea; Formicidae; Ponerinae; Ponerini; Harpegnathos; Harpegnathos saltator [Taxonomy ID: 610380]

**Submission:**

Registration date: 27-Aug-2010  
BGI-Shenzhen, Shenzhen 518083, China

**See Genome Information for Harpegnathos saltator**

**NAVIGATE ACROSS**  
2 additional projects are related by organism.

**BioProject**  
BioSample  
Full text in PMC  
Genome  
Nucleotide  
Protein  
PubMed  
Related Genes  
SRA  
Taxonomy  
WGS master

**LinkOut to external resources**

GOLDCARD: Gc01516 [Genomes On Line Database]

SILVA SSU Database [SILVA]

SILVA LSU Database [SILVA]

**Recent activity**

Harpegnathos saltator

BioProject Links for SRA (Select 24670) (1)

BioSample Links for SRA (Select 24670) (1)

Generic sample from Camponotus floridanus

ants AND ("biomol dna"[Properties] AND "strategy wgs"[Properties])... (114)

See more...



# Access to NCBI data

The screenshot shows the NCBI SRA search results page for the query 'arthropoda or rodentia'. The browser address bar shows 'www.ncbi.nlm.nih.gov/sra/?term=arthropoda+or+rodentia'. The search bar contains 'arthropoda or rodentia' and a 'Search' button. The results are displayed in a table with columns for 'Accession', 'Description', and 'Download'. The first result is 'WGS sequencing of D.melanogaster DGRP-584' with accession SRX155986. The second result is 'WGS sequencing of D.melanogaster' with accession SRX156019. The third result is 'WGS sequencing of D.melanogaster' with accession SRX155999. The fourth result is 'WGS sequencing of D.melanogaster' with accession SRX155991. The fifth result is 'Illumina Sequencing of Drosophila melanogaster DGRP-802' with accession SRX025318. The sixth result is 'Other Sequencing of fruit fly' with accession SRX021491. The seventh result is 'Other Sequencing of fruit fly' with accession SRX021402. The eighth result is 'Other Sequencing of fruit fly' with accession SRX021391. The page also includes filters for 'Access' (Public), 'Source' (DNA, metagenomic), and 'Type' (genome, aligned data). The 'Results by taxon' section lists top organisms like Mus musculus, Anopheles gambiae, and Drosophila melanogaster. The 'Top Biosamples' section lists early embryo, Drosophila melanogaster DGRP, and Vibrio sp. CAIM 1540T. The 'Top Bioprojects' section lists Drosophila melanogaster Vari, BDGP modENCODE D. melanogast, Epigenomics projects for the, Drosophila melanogaster Tran, Functional genomics project, Mouse ENCODE epigenomic data, Mouse ENCODE functional geno, Mouse ENCODE transcriptome d, and Rattus norvegicus transcript. The 'Search in related databases' section is also present.

arthropoda or rodentia - SRA

www.ncbi.nlm.nih.gov/sra/?term=arthropoda+or+rodentia

NCBI Resources How To Sign in to NCBI

SRA

Search

Save search Advanced

Help

Show additional filters

Clear all

Access

Public (24036)

Source

clear

WGS sequencing of D.melanogaster DGRP-584

1. Accession: SRX155986

Type

clear

WGS sequencing of D.melanogaster

2. 1 ILLUMINA (Illumina Genome Analyzer Ix) run: 43.7M spots, 9.8G bases, 4.8Gb downloads

Accession: SRX156019

Other

aligned data (18392)

Clear all

Show additional filters

Display Settings: Summary, 20 per page

Send to: Manage Filters

Results: 1 to 20 of 24038

Filters activated: DNA, genome. Clear all to show 123433 items.

Results by taxon

Top Organisms Tree

Mus musculus (9757)

Anopheles gambiae (8084)

Drosophila melanogaster (1936)

Anopheles arabiensis (837)

Rattus norvegicus (341)

All other taxa (3083)

More...

Top Biosamples

early embryo (replicate 1) (3)

Drosophila melanogaster DGRP... (1)

Vibrio sp. CAIM 1540T (1)

E18.5 pancreas islets (1)

KWASHIORGUT\_codes (1)

Temoin\_nigers (1)

Temoin\_Nigers (1)

Top Bioprojects

Drosophila melanogaster Vari... (2691)

BDGP modENCODE D. melanogast... (568)

Epigenomics projects for the... (530)

Drosophila melanogaster Tran... (310)

Functional genomics project ... (298)

Mouse ENCODE epigenomic data (248)

Mouse ENCODE functional geno... (213)

Mouse ENCODE transcriptome d... (102)

Rattus norvegicus transcript... (15)

Search in related databases

| Database | Access |            | all |
|----------|--------|------------|-----|
|          | public | controlled |     |
|          |        |            |     |

# Access to NCBI data

Camponotus floridanus wh x

www.ncbi.nlm.nih.gov/sra/SRX022799[accn]

NCBI Resources How To

SRA SRA Search

Advanced

Display Settings: Full

**SRX022799: Camponotus floridanus whole genome sequencing**  
1 ILLUMINA (Illumina Genome Analyzer II) run: 18.9M spots, 1.7G bases, 1.1Gb downloads

Accession: SRX022799  
Experiment design: Standard Solexa protocol  
Submitted by: BGI  
Study summary: [SRP002786](#) • Harpegnathos saltator genome sequencing project • [PRJNA50203](#) • [All experiments](#) • [Run Selector](#) (less...)  
Center Project: Harpegnathos saltator  
Sample: [SAMN00016771](#) • Generic sample from Camponotus floridanus (less...)  
Organism: [Camponotus floridanus](#)  
Library: [ANTdanD2BBDLAAPE](#) (less...)  
Strategy: WGS  
Source: GENOMIC  
Selection: PCR  
Layout: PAIRED, Nominal length: 5012  
Platform: [Illumina](#) (less...)  
Instrument model: Illumina Genome Analyzer II  
Spot descriptor:  
1 forward 45 reverse  
Experiment attributes:  
center\_name: BGI  
Total: 1 run, 18.9M spots, 1.7G bases, 1.1Gb

| #  | Run                       | # of Spots | # of Bases | Size                  | Published  |
|----|---------------------------|------------|------------|-----------------------|------------|
| 1. | <a href="#">SRR059191</a> | 18,938,689 | 1.7G       | <a href="#">1.1Gb</a> | 2010-08-27 |

ID: 24670

Send to:

Related information

- BioProject
- BioSample
- PMC
- PubMed
- Taxonomy
- WGS

Recent activity

- ants AND ("biomol dna"[f wgs][Properties]... (114)
- Sample from Papilio xuth
- BioSample Links for SRA
- (generic sample arthropo [orgn] AND ("biom... (8)
- generic sample arthropo [Properties] AND "stra...

# Access to NCBI data

The screenshot shows the NCBI BioSample database interface. The browser address bar displays the URL: [www.ncbi.nlm.nih.gov/biosample?LinkName=sra\\_biosample&from\\_uid=24670](http://www.ncbi.nlm.nih.gov/biosample?LinkName=sra_biosample&from_uid=24670). The page title is "Generic sample from Camponotus floridanus". The "Identifiers" section shows "BioSample: SAMN00016771; Sample name: **Camponotus\_floridanus\_C129**; SRA: SRS084746". The "Organism" section lists "Camponotus floridanus (Florida carpenter ant)" and its taxonomic lineage. The "Attributes" section shows "No attributes". The "BioProject" section shows "PRJNA50203 Harpegnathos saltator" and a link to "Retrieve all samples from this project". The "Submission" section shows "BGI; 2010-09-02". The "Accession" section shows "SAMN00016771 ID: 16771" and links to "BioProject" and "SRA". The right sidebar contains "Related information" (BioProject, SRA, Taxonomy) and "Recent activity" (BioSample Links for SRA (Select 24670) (1), Generic sample from Camponotus floridanus biosample, ants AND ("biomol dna"[Properties] AND "strategy wgs"[Properties])... (114) SRA, Sample from Papilio xuthus biosample, BioSample Links for SRA (Select 1235728) (1) BioSample). The footer contains navigation links for "GETTING STARTED", "RESOURCES", "POPULAR", "FEATURED", and "NCBI INFORMATION".

Generic sample from Camponotus floridanus

Identifiers BioSample: SAMN00016771; Sample name: **Camponotus\_floridanus\_C129**; SRA: SRS084746

Organism [Camponotus floridanus](#) (Florida carpenter ant)  
cellular organisms; Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Vespoidea; Formicidae; Formicinae; Camponotini; Camponotus

Attributes **No attributes**

BioProject [PRJNA50203](#) Harpegnathos saltator  
Retrieve [all samples](#) from this project

Submission BGI; 2010-09-02

Accession: SAMN00016771 ID: 16771  
[BioProject](#) [SRA](#)

Related information  
BioProject  
SRA  
Taxonomy

Recent activity  
[Turn Off](#) [Clear](#)  
BioSample Links for SRA (Select 24670) (1) BioSample  
Generic sample from Camponotus floridanus biosample  
ants AND ("biomol dna"[Properties] AND "strategy wgs"[Properties])... (114) SRA  
Sample from Papilio xuthus biosample  
BioSample Links for SRA (Select 1235728) (1) BioSample  
[See more...](#)

You are here: NCBI > DNA & RNA > BioSample

Write to the Help Desk

**GETTING STARTED**  
NCBI Education  
NCBI Help Manual  
NCBI Handbook  
Training & Tutorials

**RESOURCES**  
Chemicals & Bioassays  
Data & Software  
DNA & RNA  
Domains & Structures  
Genes & Expression  
Genetics & Medicine  
Genomes & Maps  
Homology  
Literature

**POPULAR**  
PubMed  
Bookshelf  
PubMed Central  
PubMed Health  
BLAST  
Nucleotide  
Genome  
SNP  
Gene

**FEATURED**  
Genetic Testing Registry  
PubMed Health  
GenBank  
Reference Sequences  
Gene Expression Omnibus  
Map Viewer  
Human Genome  
Mouse Genome  
Influenza Virus

**NCBI INFORMATION**  
About NCBI  
Research at NCBI  
NCBI News  
NCBI FTP Site  
NCBI on Facebook  
NCBI on Twitter  
NCBI on YouTube

# At Repositiv we vision is

**1-click to human genomic data access**

*to make **finding data** as easy as **finding a book**  
on Amazon, booking a **hotel** on Expedia!*

# A platform to make data accessible for research

**MAKE DATA  
DISCOVERABLE**

**ACCESS  
IT FAST**

**MAXIMIZE  
COMMUNITY**

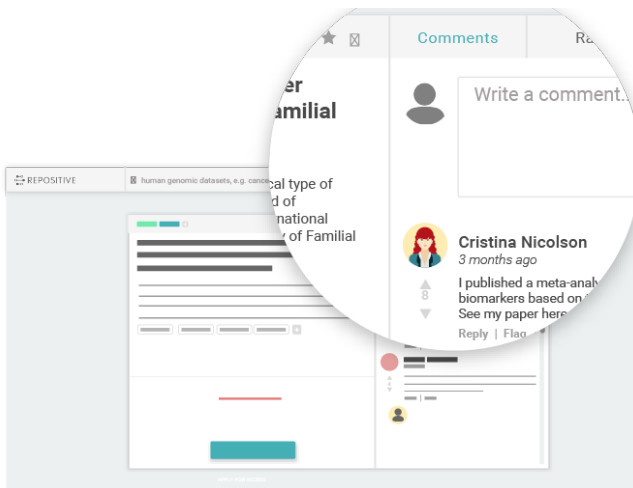
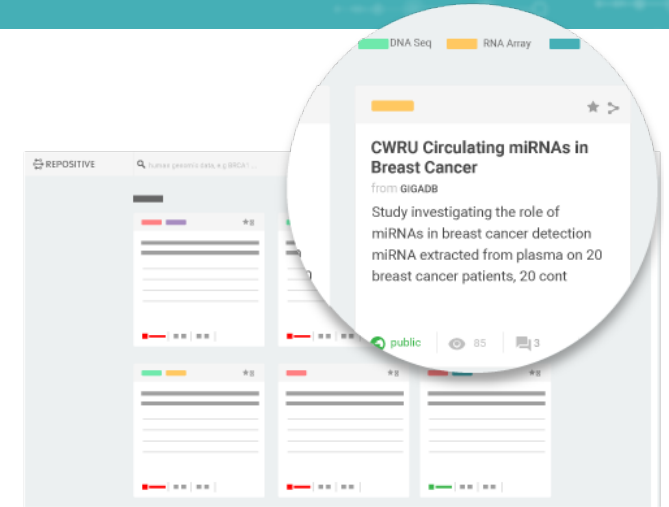


# The Repositive platform includes



**Search, see  
related results**

**Discover and  
access**



**Co-annotate data &  
community feedback**

**Find colleagues &  
their data interests**

