



Open repositories

Fetch and manage open access datasets in R

Andigoni Malousi, PhD

Adjunct Lecturer in Bioinformatics
e-mail: andigoni@auth.gr
tel: 2310 999163
Medical School | AUTH



Data Import in R

Local files

These can be of different tabular formats, such as plain csv and statistical software files e.g., Excel, SPSS, STATA, SAS, Octave (foreign package).

Web files

Simply by setting the URL of the tabular data in the “read.csv” arguments.

Build-in package data/.RData

Datasets preloaded in R using the `data()` command and `.Rdata` files

SQL databases

Import data from relational databases using SQL server.

Public data repositories

Fetch public data (raw, processed) from online repositories.



How it started - The Reproducibility Crisis

Published: 27 January 2014

Policy: NIH plans to enhance reproducibility

Francis S. Collins & Lawrence A. Tabak [✉](#)

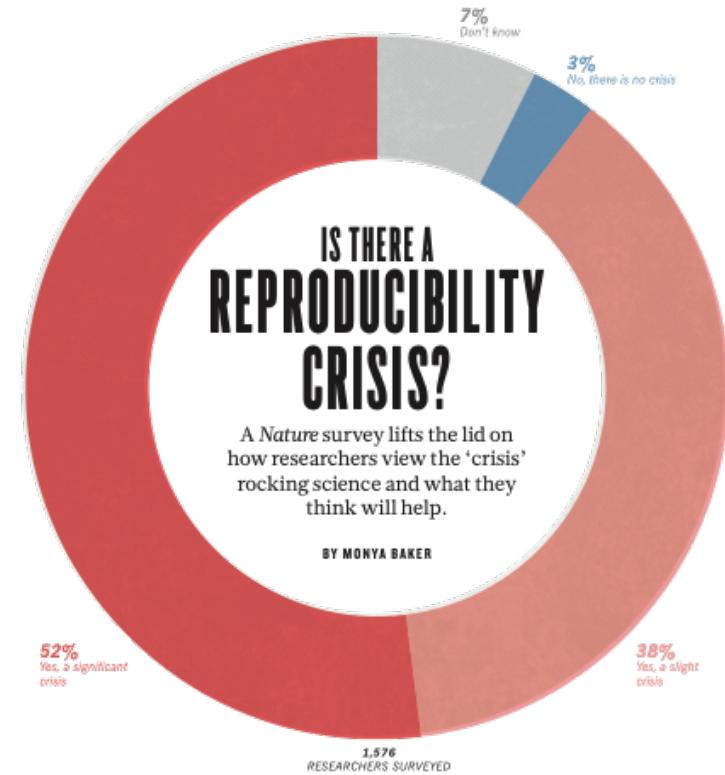
Nature 505, 612–613 (2014) | [Cite this article](#)

1094 Accesses | 768 Citations | 647 Altmetric | [Metrics](#)

Nature's poll of 1,576 scientists (2016)

<https://www.nature.com/articles/533452a>

- 70% failed to reproduce at least one other scientist's experiment (87% of chemists, 77% of biologists, 67% of medical researchers).
- 50% had failed to reproduce one of their own experiments.
- less than 20% had ever been contacted by another researcher unable to reproduce their work.



Miyakawa *Molecular Brain* (2020) 13:24
<https://doi.org/10.1186/s13041-020-0552-2>

Molecular Brain

EDITORIAL

Open Access

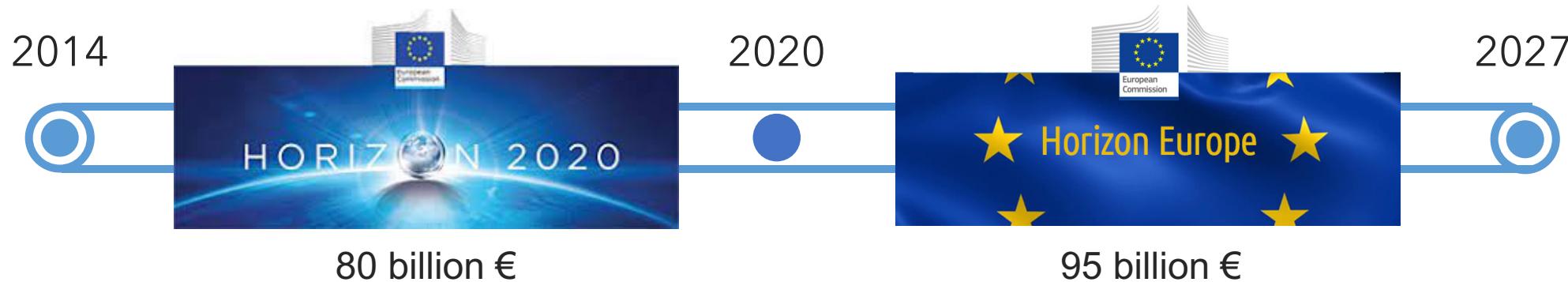
No raw data, no science: another possible source of the reproducibility crisis

Tsuyoshi Miyakawa



The Horizon 2020 Open Access Mandate

From Horizon 2020, the European Commission requires that all **peer-reviewed publications** resulting from project funding are **open access**.

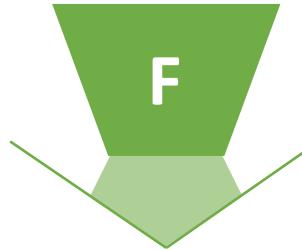


Open Science Practices

- Open access to research outputs, such as publications, data, software and workflows
- Early and open sharing of research, for example through preprints
- Use of open research repositories for data sharing
- Participation in open peer-review

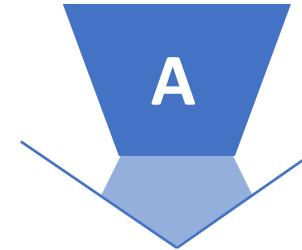
FAIR Principles

<https://www.go-fair.org/fair-principles/>



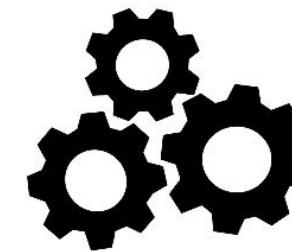
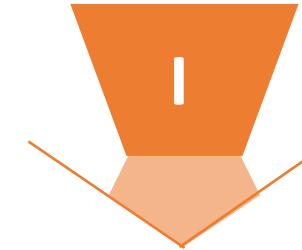
Findable

Globally unique and eternally **persistent DOI identifier** describing the data with rich metadata and making sure it is findable through disciplinary portals.



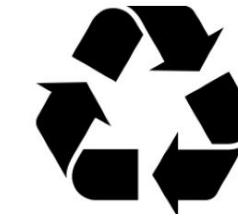
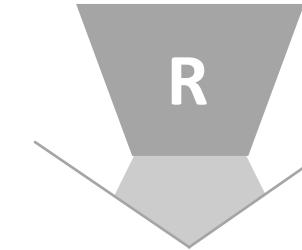
Accessible

Data and metadata should be retrievable in a **variety of formats** that are sensible to **humans and machines** using persistent identifiers.



Interoperable

The description of **metadata** elements should follow community guidelines that use an open, **well-defined vocabulary**.



Reusable

The data should maintain its initial richness to be reused in future research. It should also be **machine processable** and verifiable.



Open Access Repositories

What is a repository?

An open access repository stores and provides **free access** to a digital collection of **research data** and/or **research outputs**. For example, researchers often self-archive their publications in community-based repositories, such as **bioRxiv**, **medRxiv**, and **F1000Research**. A repository provides a globally unique '**persistent identifier**' (DOI, Digital Object Identifier), which allows easy discoverability.

F1000Research
Open for Science

bioRxiv
THE PREPRINT SERVER FOR BIOLOGY

medRxiv
THE PREPRINT SERVER FOR HEALTH SCIENCES

Where can I find open access repositories?

- <http://www.re3data.org>
- <https://www.osf.io>
- <https://datascience.nih.gov>

Global registry of data repositories
Open Science Foundation
NIH data repositories (Resources tab)



National Institutes
of Health

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

 OSF



- “Open Science” public repository, funded by OpenAIRE, CERN and the EU Horizon 2020 Program.
- Open to all research fields and types of data
- Free to use (file size up to 50GB)
- Archives datasets, software, publications, lessons, posters etc.

Special features

- **DOI versioning**: Enables data editing after publication, quoting a specific version of a dataset.
- **GitHub connection**: All new GitHub releases are automatically published as new versions.



November 10, 2020

Dataset Open Access

COVID-19 mortality correlation with cloudiness, sunlight, latitude in European countries

Iftime Adrian; Omer Secil; Burcea Victor

"COVID-19 mortality correlation with cloudiness, sunlight, latitude in European countries"

Dataset for article titled **Dataset description**

"COVID-19 mortality: positive correlation with cloudiness, sunlight and no correlation with latitude in Europe"

by SECIL OMER, ADRIANIFTIME, VICTOR BURCEA

Corresponding author: A. Iftime, University of Medicine and Pharmacy "Carol Davila", Biophysics Department, 8 Blvd. Eroii Sanitari, 050474 Bucharest, Romania. Email address: adrian.ifetime [at] umfcd.ro.

Article currently under review, "Epidemiology & Infection", ISSN: 0950-2688 (Print), 1469-4409 (Online)

=====

1.0.0.COVID-19_Mortality_Cloudiness_Insolation_EUROPE_March_August_2020.csv	30.6 kB	Preview Download
md5:f55d87f2f83e9d88a9a2a596af3ce7f3		?
1.0.0.Description_dataset_COVID_19.txt	3.3 kB	Download
md5:96e961d21274ce3b42ba3176a50e00e0		?
1.0.0.INFOGRAFIC_CloudFraction_vs_COVID-19_mortality_Europe_March-August_2020.png	147.5 kB	Preview Download
md5:791bba02f2cdc345c6b0f71bb51421c		?

Data files

DOI Assignment

Publication date:

November 10, 2020

DOI:

DOI [10.5281/zenodo.4266758](https://doi.org/10.5281/zenodo.4266758)

Keyword(s):

COVID-19 mortality Europe Cloudiness Insolation
Latitude

Published in:

Epidemiology & Infection (under review):

Communities:

Coronavirus Disease Research Community - COVID-19

Zenodo/zen4R package

The **zen4R** package offers an R interface to Zenodo. It supports creation of metadata, file uploads (including versioning), DOI assignment, record searches, file downloads.

```
library(zen4R)
#Create a Zenodo manager (always the starting point)
zen <- ZenodoManager$new()
#Interrogate and get a Zenodo record with its DOI or internal ID
my.zen <- zen$getRecordByDOI("10.5281/zenodo.4266758")
my.zen <- zen$getRecordById("4266758")
```

Fetch data through zen4R (with local file storage)

```
my.dir = "zen_downloads"
dir.create(my.dir) #Create a directory
files <- my.zen$listFiles() #List files in directory
my.zen$downloadFiles(path = my.dir) #Download files
input <- read.csv(<local path>)
```

Fetch data through zen4R (no local file storage)

```
files <- my.zen$listFiles() #List files in directory
input <- read.csv(file=files[1,4]) #Read in the first data file
```

Direct data import (no package required)

```
#Get the URL of the file of interest through the Zenodo interface by right-clicking the filename and “Copy link”
```

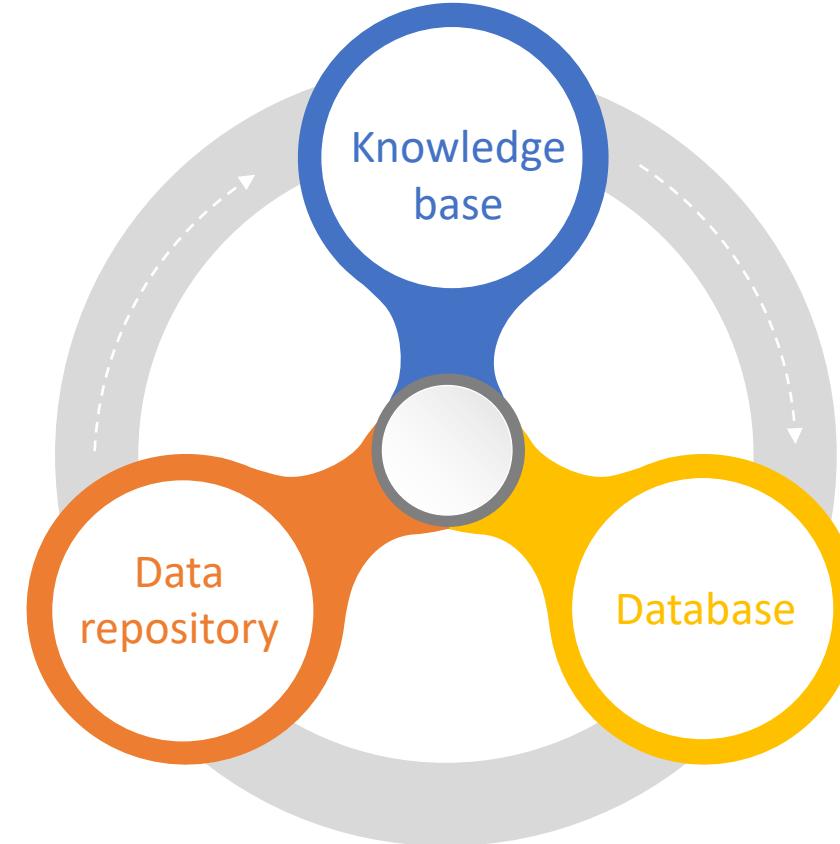
```
input <- read.csv(<URL>)
```

Data repository vs Database vs Knowledge base

Data can be stored in three main entities depending on the datatypes, organizational level, and querying capabilities.

Data repository

Accepts submission of relevant data from the community to store, organize, validate, archive, preserve and distribute the data, in compliance with the FAIR Data Principles.



Knowledge base

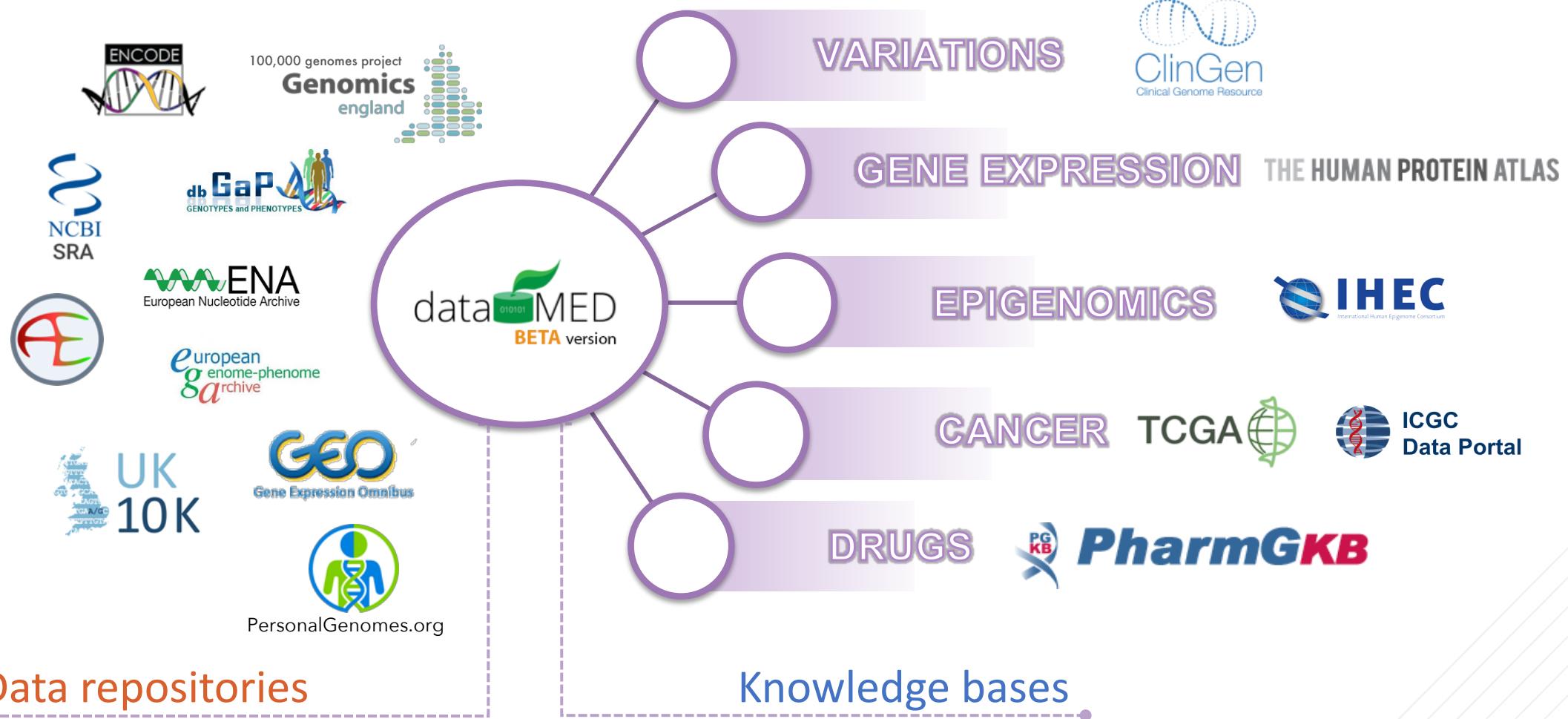
Extracts, accumulates, organizes, annotates and links the growing body of information to related data of external sources

Database

Is an organized collection of structured data, typically modeled in rows and columns in a series of tables. Data are easily accessed, modified, filtered. Example language: SQL

Example data repositories / knowledge bases

Mostly but not only -omics data



Data repositories

Knowledge bases

How big are the -omics “big” data?

PLOS BIOLOGY

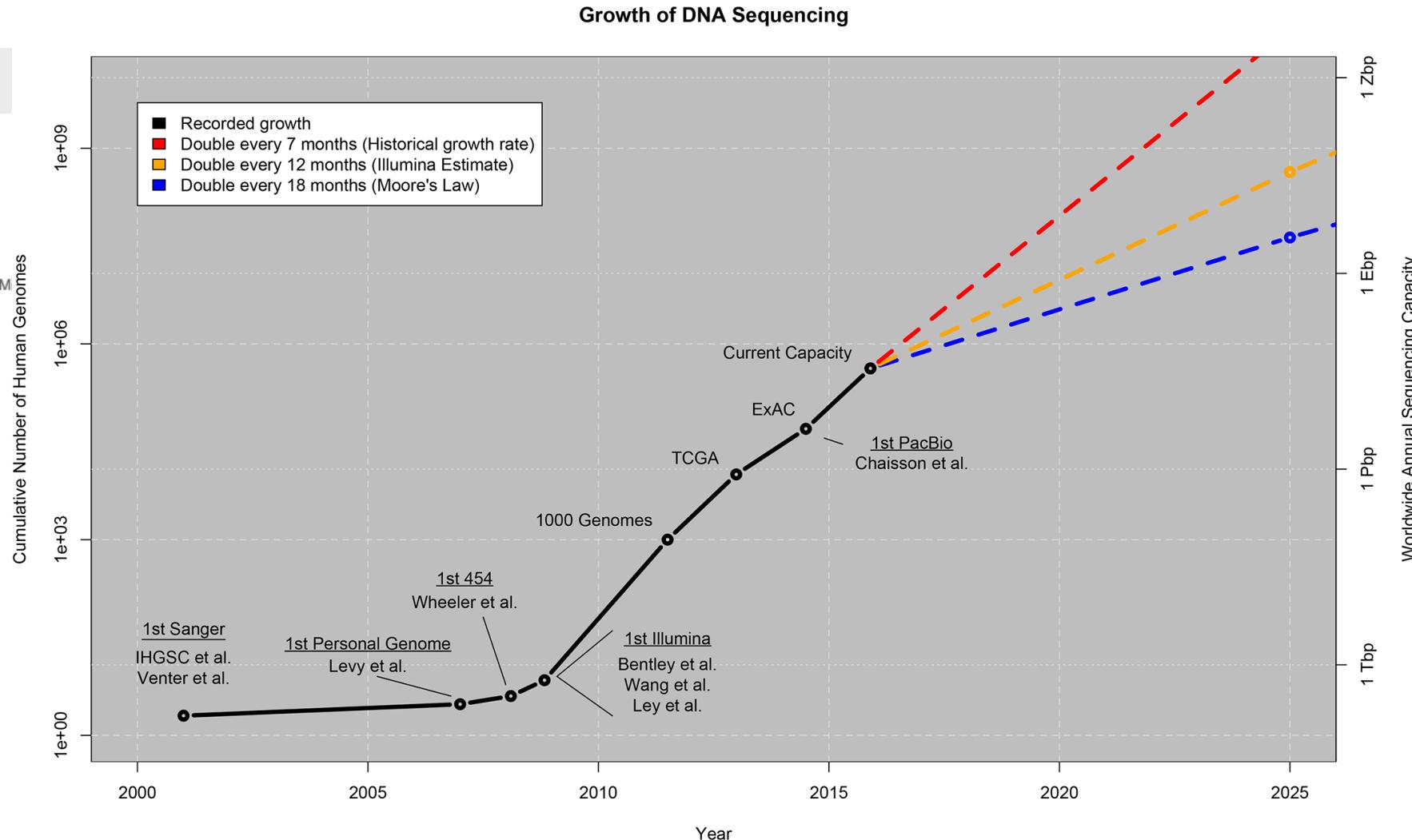
OPEN ACCESS

PERSPECTIVE

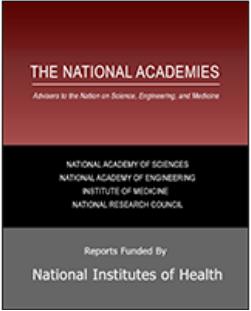
Big Data: Astronomical or Genomical?

Zachary D. Stephens, Skylar Y. Lee, Faraz Faghri, Roy H. Campbell, Chengxiang Zhai, Michael C. Schatz, Saurabh Sinha, Gene E. Robinson

Published: July 7, 2015 • <https://doi.org/10.1371/journal.pbio.1002195>



Toward Precision Medicine



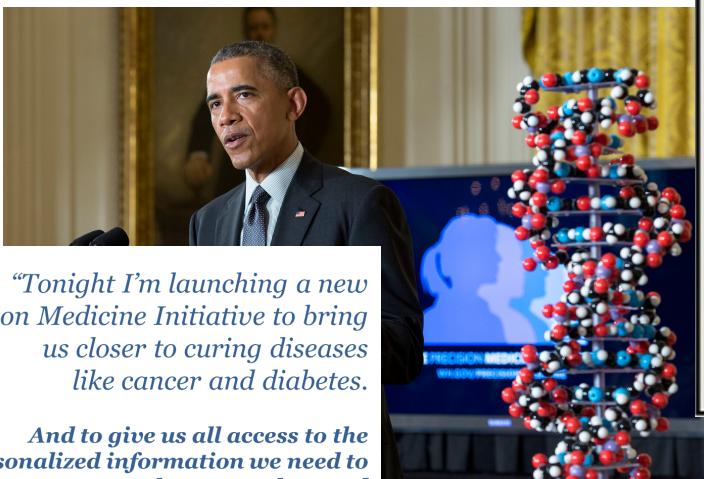
Toward Precision Medicine

Building a Knowledge Network for Biomedical Research and a New

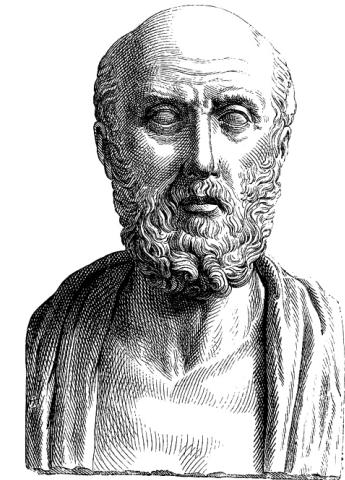
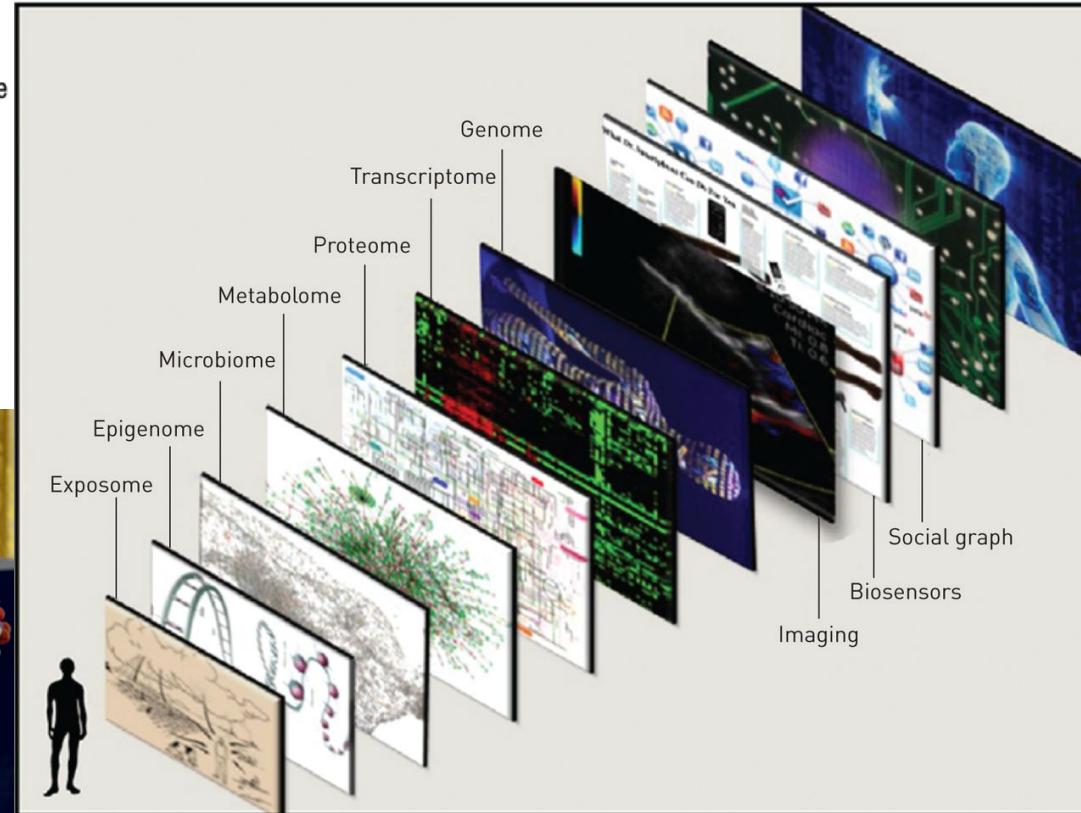
Taxonomy of Disease

National Research Council (US) Committee

Developing a New Taxonomy of Disease.



President Barack Obama
2015 State of the Union Address | January 20, 2015



It's far more important to know what person the disease has than what disease the person has.

(Hippocrates 460 – 370 B.C.)



The Cancer Genome Atlas (TCGA)

The Cancer Genome Atlas (TCGA) is a landmark cancer genomics program, aiming to molecularly characterize over 20,000 primary cancer and matched normal samples spanning 33 cancer types.

Collaboration of the NCI and the National Human Genome Research Institute.

Data are available through the Genomic Data Commons (GDC) portal (<https://gdc.cancer.gov>). GDC is a unified repository and knowledge base hosting data from 18 cancer research programs.

>4.000 pubmed abstracts in 2020 (keyword: TCGA)

NATIONAL CANCER INSTITUTE
THE CANCER GENOME ATLAS

TCGA BY THE NUMBERS

TCGA produced over



To put this into perspective, 1 petabyte of data is equal to



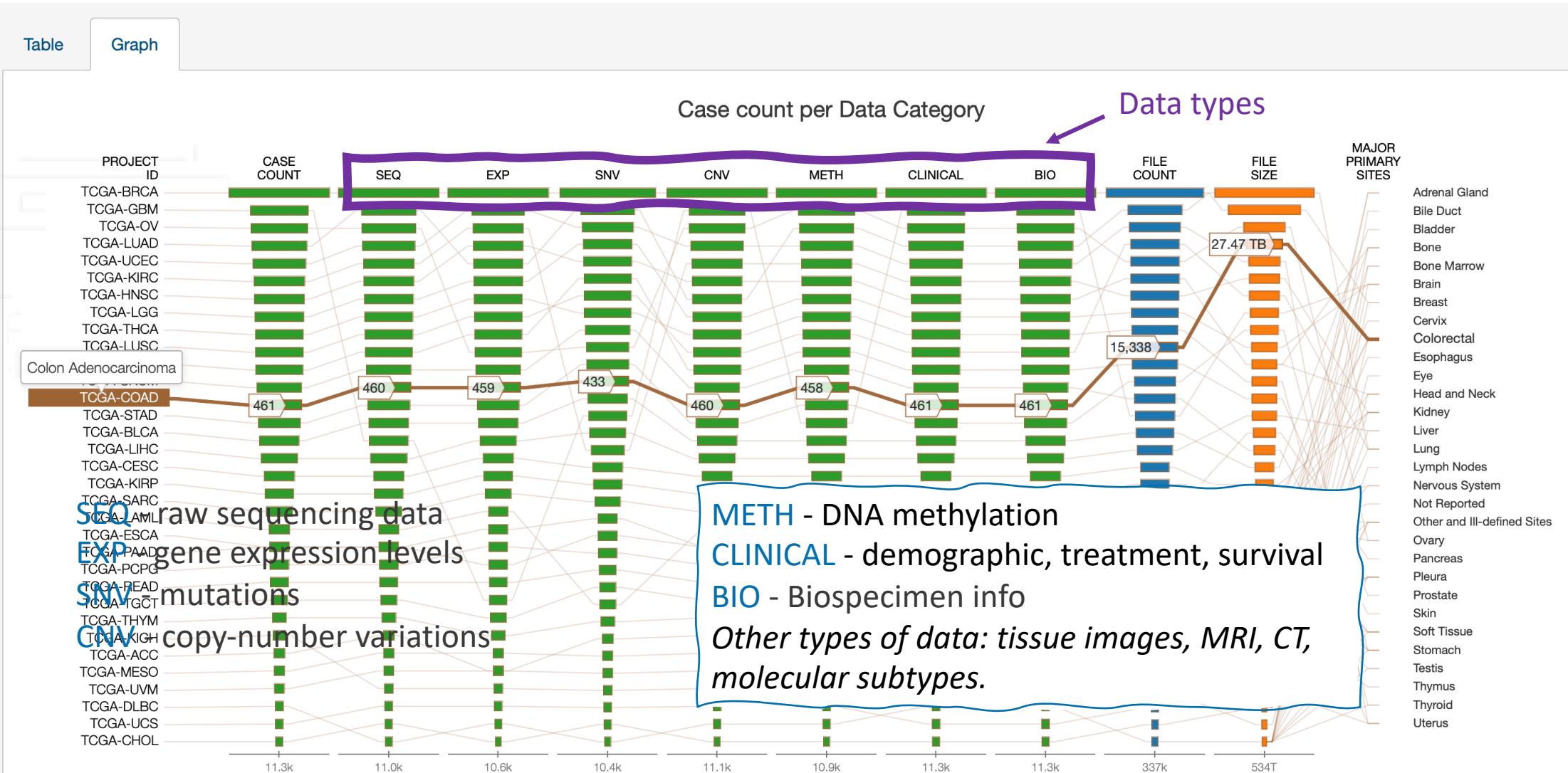
TCGA data describes



...based on paired tumor and normal tissue sets collected from

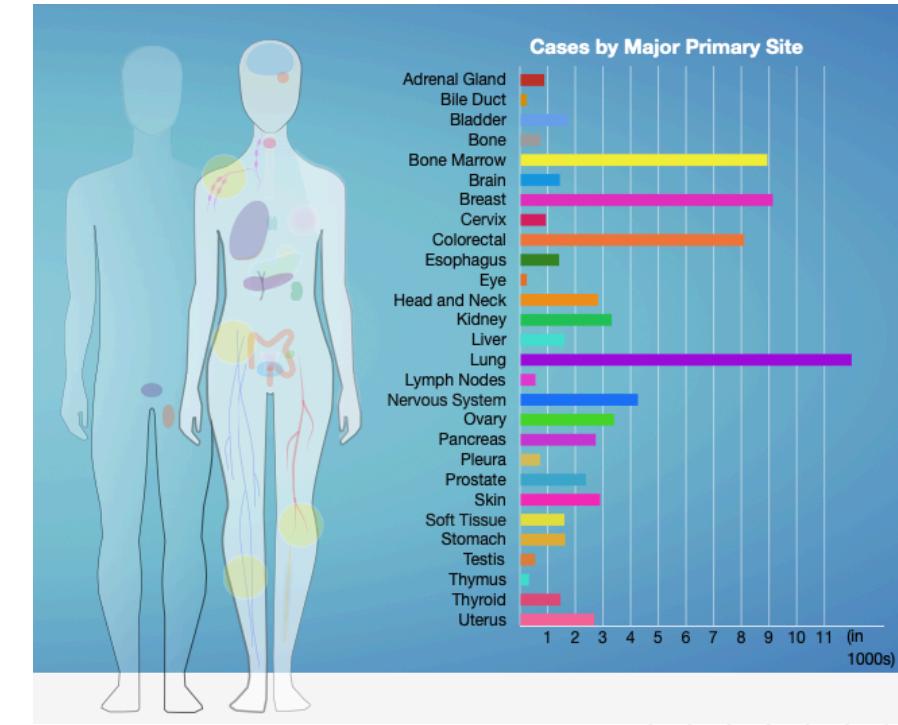
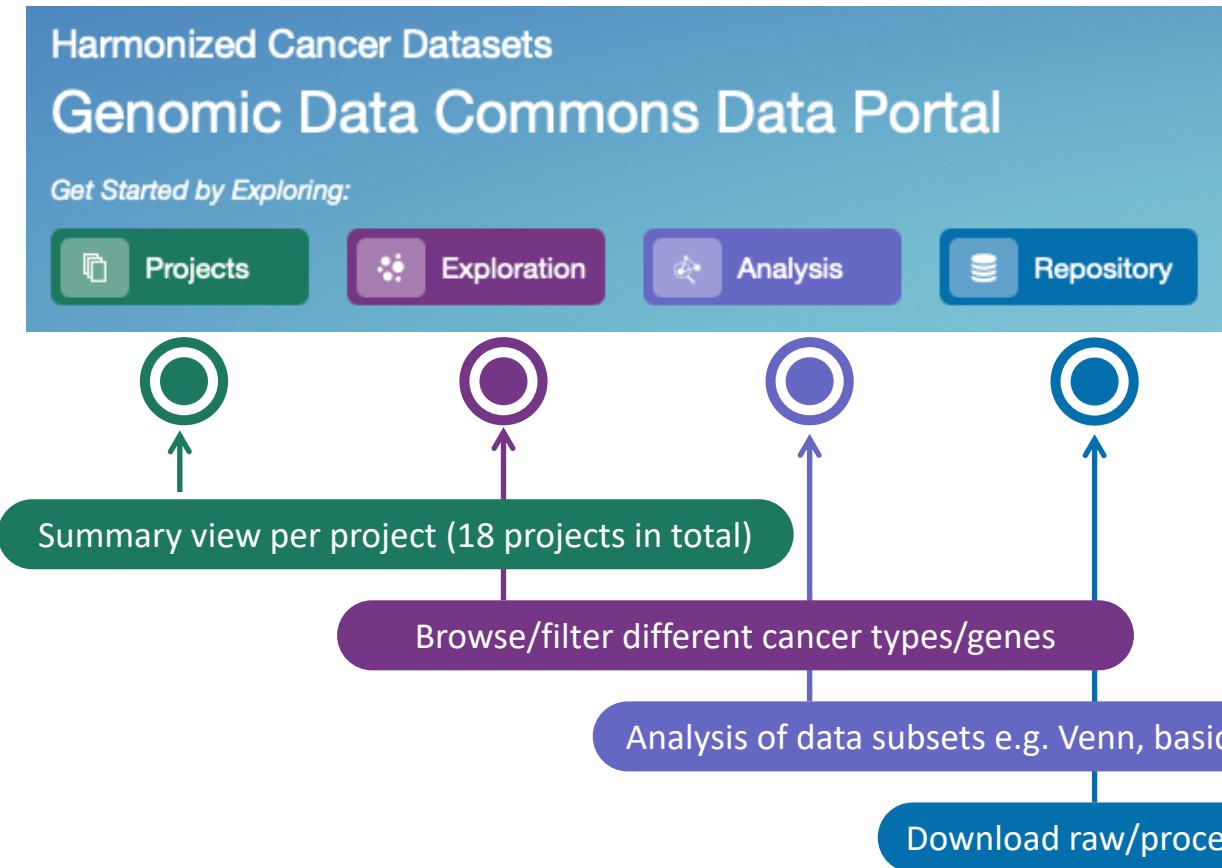


The Cancer Genome Atlas (TCGA)



The Cancer Genome Atlas (TCGA) - Web access

<https://gdc.cancer.gov>





The Cancer Genome Atlas (TCGA) - Programmatic access

[TCGAquery: Searching TCGA open-access data](#)

[TCGAquery: Searching TCGA open-access data for download](#)

[TCGAquery_Version: Retrieve versions information of the data in TCGA](#)

[TCGAquery_clinic & TCGAquery_clinicFilt: Working with clinical data.](#)

[TCGAquery_subtype: Working with molecular subtypes data.](#)

[TCGAquery_integrate: Summary of the common numbers of patient samples in different platforms](#)

[TCGAquery_investigate: Find most studied TFs in pubmed](#)

[TCGAquery_Social: Searching questions,answers and literature](#)

[TCGAdownload: Downloading open-access data](#)

[TCGAPrepare: Preparing the data](#)

[Preparing the data for other packages](#)

[TCGAanalyze: Analyze data from TCGA.](#)

[TCGAanalyze_Preprocessing: Preprocessing of Gene Expression data \(IlluminaHiSeq_RNASeqV2\).](#)

[TCGAanalyze DEA & TCGAanalyze_LevelTab: Differential expression analysis \(DEA\)](#)

[TCGAanalyze_EAcomplete & TCGAvisualize_EAbarplot: Enrichment Analysis](#)

[TCGAanalyze_survival: Survival Analysis: Cox Regression and dnet package](#)

[TCGAanalyze_DMR: Differentially methylated regions Analysis](#)

[TCGAvisualize: Visualize results from analysis functions with TCGA's data.](#)

[TCGAvisualize_PCA: Principal Component Analysis plot for differentially expressed genes](#)

[TCGAvisualize_SurvivalCoxNET: Survival Analysis: Cox Regression and dnet package](#)

[TCGAvisualize_meanMethylation: Sample Mean DNA Methylation Analysis](#)

[TCGAvisualize_starburst: Analyzing expression and methylation together](#)

TCGAbiolinks: An R/Bioconductor package for integrative analysis with GDC data

<https://doi.org/doi:10.18129/B9.bioc.TCGAbiolinks>



TCGAbiolinks is among the most popular packages. The complete list can be found through rseek.org.



`>browseVignettes("TCGAbiolinks")`
A quite informative list of tutorials for various usages of the package.

```
if (!requireNamespace("BiocManager",
quietly=TRUE))
install.packages("BiocManager")
BiocManager::install("TCGAbiolinks")
```



TCGAbiolinks: Example (Colon cancer, TCGA-COAD)

Gender - survival analysis

doi:10.1038/nature11252

```
#Load package
library(TCGAbiolinks)

#Summary of TCGA-COAD project
TCGAbiolinks:::getProjectSummary("TCGA-COAD")

#Get molecular subtype per patient
mol.subtype <- TCGAquery_subtype(tumor = "COAD")


```

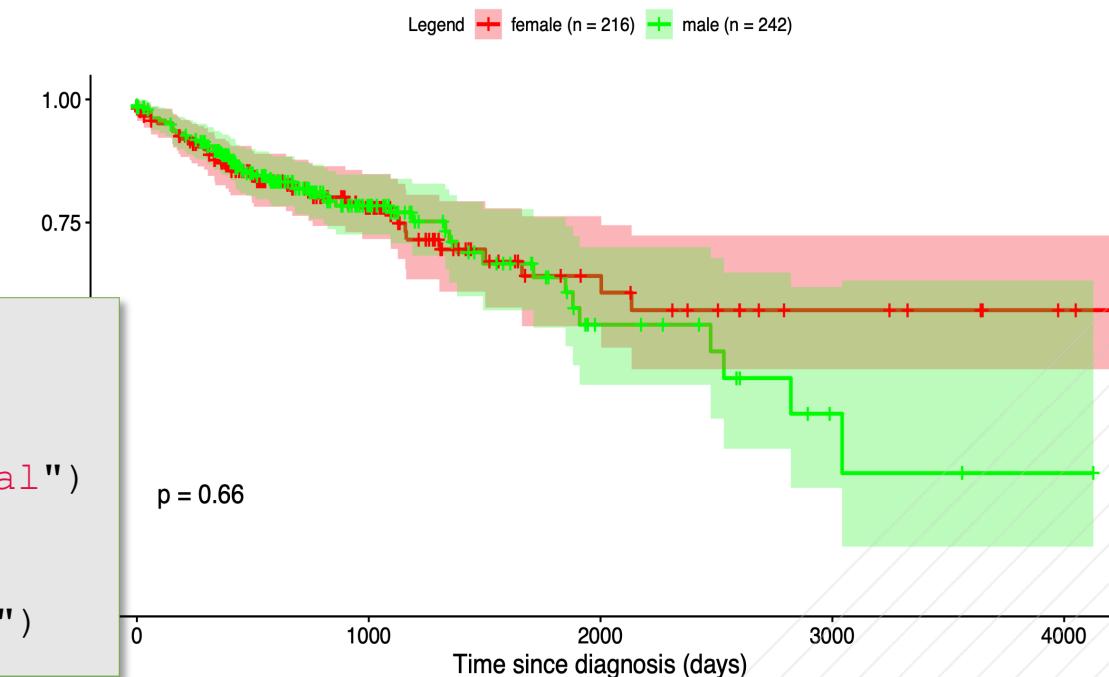


```
#Load package
library(survminer)
#Get clinical data per patient
clinical.coad <- GDCquery_clinic("TCGA-COAD", "clinical")
TCGAanalyze_survival(clinical.coad,
                      clusterCol = "gender",
                      main = "Gender survival analysis")
```

Comprehensive molecular characterization of human colon and rectal cancer

The Cancer Genome Atlas Network*

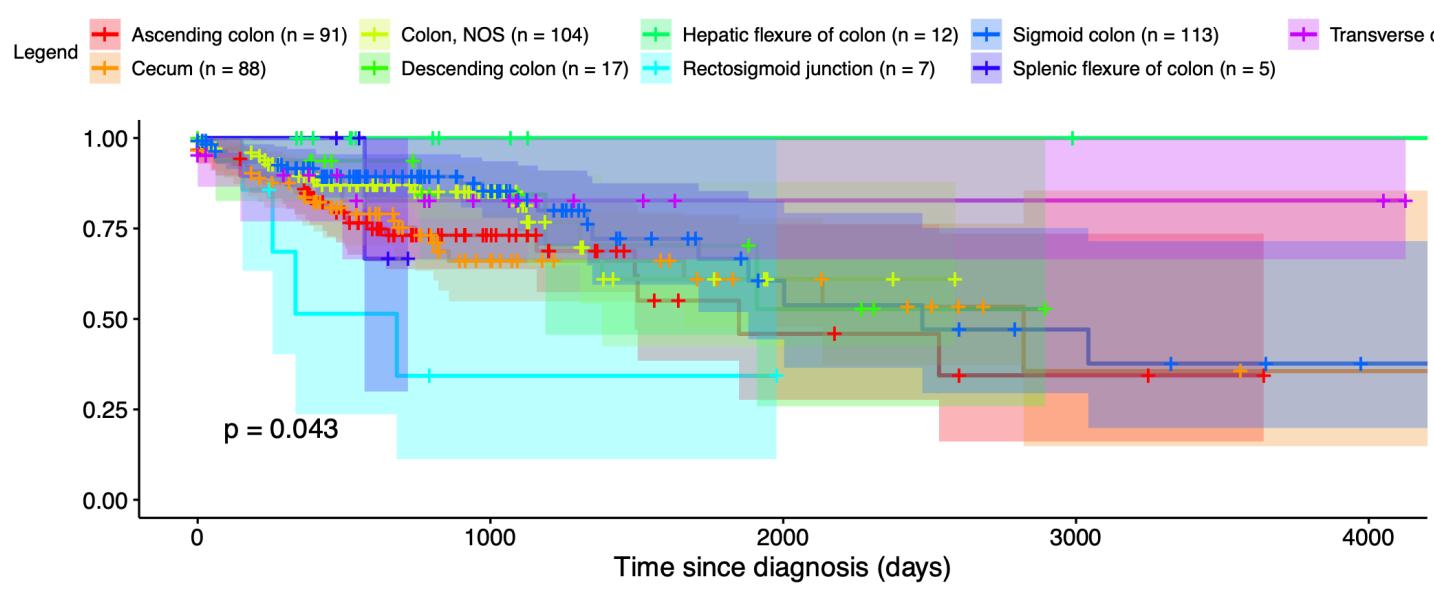
<https://www.nature.com/articles/nature11252>



TCGAbiolinks: Example (Colon cancer, TCGA-COAD)

Tissue of origin - survival analysis

```
#Get clinical data per patient
clinical.coad <- GDCquery_clinic("TCGA-COAD", "clinical")
TCGAanalyze_survival(clinical.coad,
                      clusterCol = "tissue_or_organ_of_origin",
                      main = "Tissue of origin survival analysis")
```



TCGAbiolinks: Example (Colon cancer, TCGA-COAD)

Tissue of origin - survival analysis

Mangone et al. BMC Public Health (2021) 21:906
<https://doi.org/10.1186/s12889-021-10746-4>

BMC Public Health

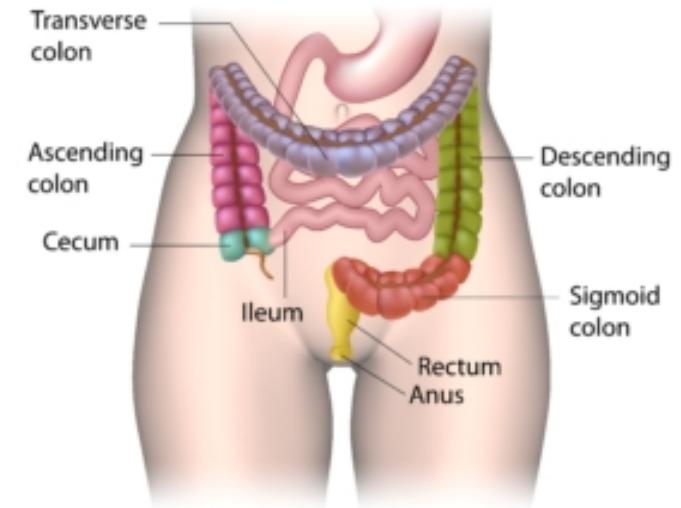
RESEARCH ARTICLE

Open Access

Colon cancer survival differs from right side to left side and lymph node harvest number matter



Lucia Mangone^{1*}, Carmine Pinto², Pamela Mancuso¹, Marta Ottone¹, Isabella Bisceglia¹, Giorgio Chiaranda³, Maria Michiara⁴, Massimo Vicentini¹, Giuliano Carrozza⁵, Stefano Ferretti⁶, Fabio Falcini⁷, Cesare Hassan⁸ and Paolo Giorgi Rossi¹



```
sub.clinical.coad <- subset(clinical.coad, tissue_or_organ_of_origin %in% c("Cecum",
"Ascending colon", "Descending colon", "Sigmoid colon"))

sub.clinical.coad$side <- ifelse(sub.clinical.coad$tissue_or_organ_of_origin %in% c("Cecum",
"Ascending colon"), "right side", "left side")

TCGAanalyze_survival(sub.clinical.coad, "side", main = "TCGA-COAD Survival analysis")
```

TCGAbiolinks: Example (Colon cancer, TCGA-COAD)

Tissue of origin - survival analysis

Mangone et al. BMC Public Health (2021) 21:906
<https://doi.org/10.1186/s12889-021-10746-4>

BMC Public Health

RESEARCH ARTICLE

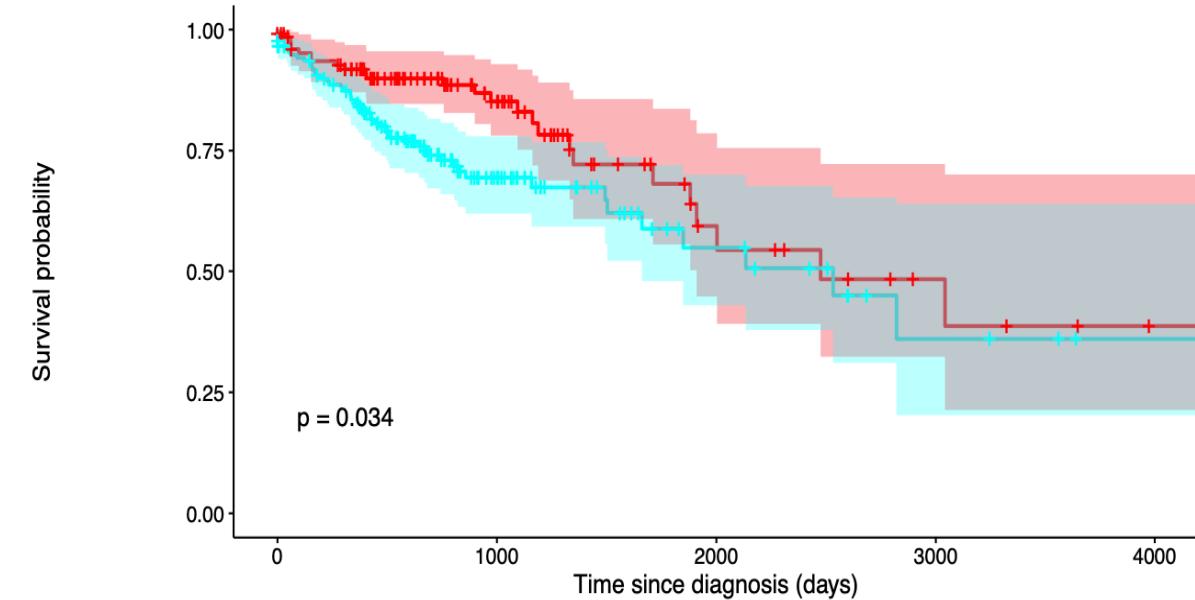
Open Access

Colon cancer survival differs from right side to left side and lymph node harvest number matter



Lucia Mangone^{1*}, Carmine Pinto², Pamela Mancuso¹, Marta Ottone¹, Isabella Bisceglia¹, Giorgio Chiaranda³, Maria Michiara⁴, Massimo Vicentini¹, Giuliano Carrozza⁵, Stefano Ferretti⁶, Fabio Falcini⁷, Cesare Hassan⁸ and Paolo Giorgi Rossi¹

Legend  left side (n = 130)  right side (n = 179)



Conclusions: This study confirms that right CRC has worse survival; the association is not due to screening status. An adequate removal of lymph nodes is associated with better survival, although the direction of the association in terms of causal links is not clear.

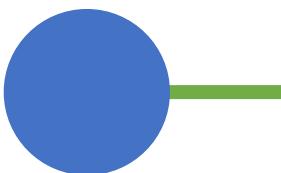
Gene Expression Omnibus (GEO)



GEO is an international public repository that archives and freely distributes microarray, next-generation sequencing, and other forms of high-throughput functional genomics data submitted by the research community.

Samples

A Sample record describes the conditions under which an individual Sample was handled, the manipulations it underwent



GSMxxx

Data series

A Series record links together a group of related Samples and provides a focal point and description of the whole study.



GSExxx

GDSxxx

Dataset

A Dataset represents a **curated** collection of biologically and statistically comparable GEO Samples and forms the basis of GEO's suite of data display and analysis tools.

GPLxxx



Platform

A Platform record is composed of a summary description of the array or sequencer and, for array-based Platforms.

Gene Expression Omnibus (GEO)



GEO2R is an interactive web tool that allows users to compare two or more groups of samples in a GEO Series in order to identify genes that are differentially expressed/methylated across experimental conditions.

Browse Content	Oct 2020
Repository Browser	
DataSets:	4348
Series: 	155226
Platforms:	22296
Samples:	4482088

Browse Content	Oct 2021
Repository Browser	
DataSets:	4348
Series: 	161995
Platforms:	22664
Samples:	4677806

Browse data series

GEO2R is only available for **microarray** gene expression and DNA methylation data.

GEO2R exports the R script for further analyses

Gene Expression Omnibus (GEO) - Example



Scope: Self Format: HTML Amount: Quick GEO accession: GSE106784

Series GSE106784 Query DataSets for GSE106784

Status	Public on Mar 01, 2021
Title	Whole genome gene microarray for diffuse gliomas
Organism	Homo sapiens
Experiment type	Expression profiling by array
Summary	We used whole genome microarray expression profiling as a discovery platform to identify high grade diffuse glioma associated differently expressed genes comparing with low grade diffuse glioma.
Overall design	We used seven high grade diffuse glioma patient samples and seven low grade diffuse glioma patient samples from Tianjin huanhu hospital during the same period to do the microarray.
Contributor(s)	Shu C, Wang Q, Wang J
Citation missing	Has this study been published? Please login to update or notify GEO .
Submission date	Nov 11, 2017
Last update date	Mar 03, 2021
Contact name	chang shu
E-mail(s)	nkchangshu@126.com
Organization name	Tianjin HuanHu hospital,School of medicine,Nankai university
Street address	94 weijin road
City	tianjin
ZIP/Postal code	300071
Country	China
Platform	GPL14550 Agilent-028004 SurePrint G3 Human GE 8x60K Microarray (Probe Name Version)
Samples (14)	GSM2850251 high grade diffuse glioma_1 GSM2850252 high grade diffuse glioma_2 GSM2850253 high grade diffuse glioma_3
More...	
Relations	BioProject PRJNA418009
Analyze with GEO2R	
Download family	Format
SOFT formatted family file(s)	SOFT [?]
MINIML formatted family file(s)	MINIML [?]
Series Matrix File(s)	TXT [?]
Data file (all samples)	
Supplementary file	Size Download File type/resource
GSE106784_RAW.tar	102.4 Mb (http://custom) TAR (of TXT)

Data series: GSE106784 (Mar 2021)

Description: Gene expression data from microarrays

Samples: 7 low-grade diffuse glioma

7 high-grade diffuse glioma

Platform: Agilent SurePrint G3 (GPL14550)

Gene targets: 42.545

Click on [Analyze with GEO2R](#) to run differential analysis.



Gene Expression Omnibus (GEO) - Example



GEO accession Set Whole genome gene microarray for diffuse gliomas

Samples		Define groups	
Group	Accession	Enter a group name:	List
high-grade DG	GSM2850251	<input type="text"/>	<input type="button" value="Cancel selection"/>
high-grade DG	GSM2850252	<input checked="" type="checkbox"/> high-grade DG (7 samples)	diffuse glioma_1 frozen tumor tissue
high-grade DG	GSM2850253	<input checked="" type="checkbox"/> low-grade DG (7 samples)	diffuse glioma_2 frozen tumor tissue
high-grade DG	GSM2850254		diffuse glioma_3 frozen tumor tissue
high-grade DG	GSM2850255		high grade diffuse glioma_4 frozen tumor tissue
high-grade DG	GSM2850256		high grade diffuse glioma_5 frozen tumor tissue
high-grade DG	GSM2850257		high grade diffuse glioma_6 frozen tumor tissue
low-grade DG	GSM2850258		high grade diffuse glioma_7 frozen tumor tissue
low-grade DG	GSM2850259		low grade diffuse glioma_1 frozen tumor tissue
			low grade diffuse glioma_2 frozen tumor tissue

- 1) Define groups: To create two groups (low-grade DG /high-grade DG)
- 2) Assign each sample to groups appropriately.
- 3) Press Analyze (GEO2R tab)
- 4) Upon completion navigate through the results to identify the most important genes.
- 5) Click on the R script tab and copy-paste the R code to Rstudio.

GEO analysis in R



Package dependencies

`GEOquery` (Bioconductor), `limma` (Bioconductor),
`umap` (CRAN)

Fetch data from GEO / ExpressionSet (gset)

```
gset <- getGEO("GSE106784", GSEMatrix =TRUE,  
AnnotGPL=FALSE)
```

log2 transformation / Fits a linear model / eBayes

Transforms expression level to log scale and fits a linear model for which various statistics are computed using an empirical Bayes method (`limma` package).

Differential analysis and visualizations

Creates the tT table that contains the top differentially expressed genes based on user-defined thresholds.

GEOquery: Is a Bioconductor package that acts as a bridge between GEO and R.

limma: Probably the most well-known package for differential analysis of microarray and RNA-seq studies.

umap: Dimensionality reduction method enabling non-linear projection of data.

Installation

To install this package, start R (version "4.1") and enter:

```
if (!requireNamespace("BiocManager", quietly = TRUE))  
  install.packages("BiocManager")  
  
BiocManager::install("GEOquery")
```

GEO analysis in R



Package dependencies

GEOquery (Bioconductor), limma (Bioconductor),
umap (CRAN)

Fetch data from GEO / ExpressionSet (gset)

```
gset <- getGEO("GSE106784", GSEMatrix =TRUE,  
AnnotGPL=FALSE)
```

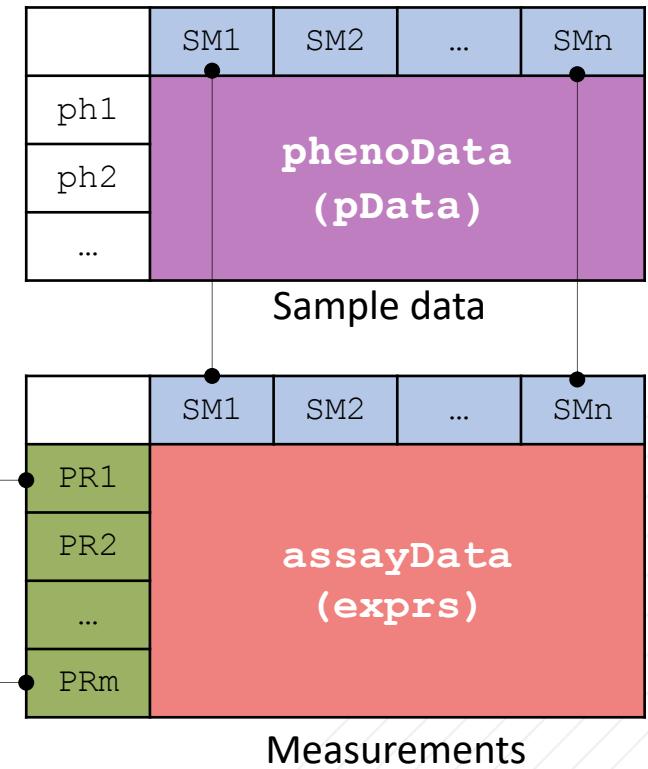
log2 transformation / Fits a linear model / eBayes

Transforms expression level to log scale and fits a linear model for which various statistics are computed using an empirical Bayes method (limma package).

Differential analysis and visualizations

Creates the tT table that contains the top differentially expressed genes based on user-defined thresholds.

ExpressionSet: Container for high-throughput assays and experimental metadata.



GEO analysis in R

Package dependencies

GEOquery (Bioconductor), limma (Bioconductor),
umap (CRAN)

Fetch data from GEO / ExpressionSet (gset)

```
gset <- getGEO("GSE106784", GSEMatrix =TRUE,  
AnnotGPL=FALSE)
```

log2 transformation / Fits a linear model / eBayes

Transforms expression level to log scale and fits a linear model for which various statistics are computed using an empirical Bayes method (limma package).

Differential analysis and visualizations

Creates the tT table that contains the top differentially expressed genes based on user-defined thresholds.

```
ex <- exprs(gset)  
exprs(gset) <- log2(ex) #log2 transformation  
#design is a matrix of 0 and 1 corresponding to group membership  
design <- model.matrix(~group + 0, gset)  
fit <- lmFit(gset, design) #fit linear model  
fit2 <- eBayes(fit2, 0.01) #calculate statistics
```

GEO analysis in R

Package dependencies

GEOquery (Bioconductor), limma (Bioconductor),
umap (CRAN)

Fetch data from GEO / ExpressionSet (gset)

```
gset <- getGEO("GSE106784", GSEMatrix =TRUE,  
AnnotGPL=FALSE)
```

log2 transformation / Fits a linear model / eBayes

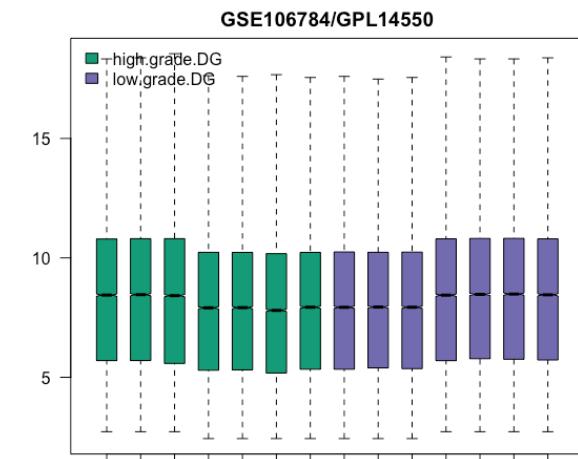
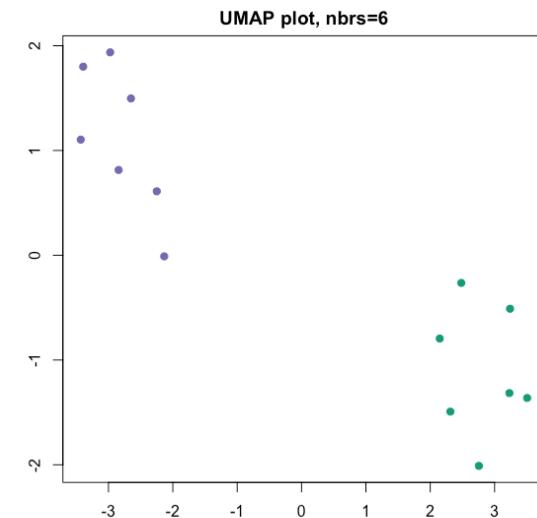
Transforms expression level to log scale and fits a linear model for which various statistics are computed using an empirical Bayes method (limma package).

Differential analysis and visualizations

Creates the tT table that contains the top differentially expressed genes based on user-defined thresholds.

#Top 250 differentially expressed genes

```
tT <- topTable(fit2, adjust="fdr", sort.by="B",  
number=250)
```



GEO analysis in R: A step further

Pathway analysis / pathfindR (CRAN package)

The goal of pathway analysis is to translate a list of genes that are differentially expressed across the given phenotypes into meaningful biological phenomena.



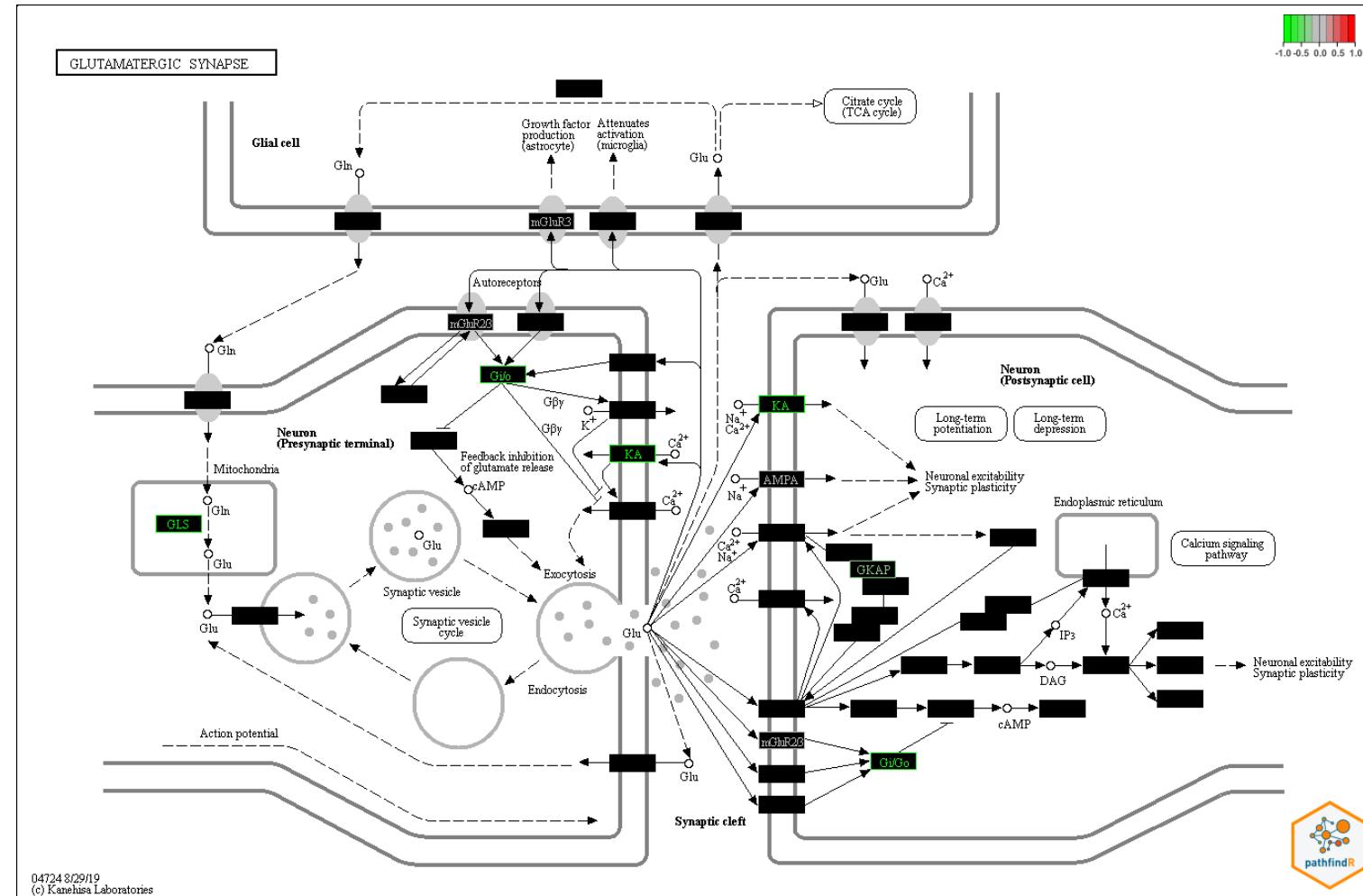
```
# Front. Genet. doi:10.3389/fgene.2019.00858
library(pathfindR)
```

```
#Three columns are only needed as input
path.input <- tT[, c("GENE_SYMBOL", "logFC",
"adj.P.Val"), ]
pathway_analysis <- run_pathfindR(path.input)
```

GEO analysis in R: A step further

The deregulated genes (in green) of the most significant pathway.

All genes are down-regulated in high-grade diffuse glioma vs low-grade.



GEO analysis in R: A step further...



ID	Term_Description	Fold_Enrichment	occurrence	support	lowest_p	highest_p	Up_regulated	Down_regulated
hsa04724	Glutamatergic synapse	5.677252	10	0.0091743	1.5e-05	1.5e-05		GLS, GRIK2, GRIA3, DLGAP1, GRM3, GNAI1

nature

Explore content ▾ Journal information ▾ Publish with us ▾

nature > articles > article

Article | Published: 18 September 2019

Glutamatergic synaptic input to glioma cells drives brain tumour progression

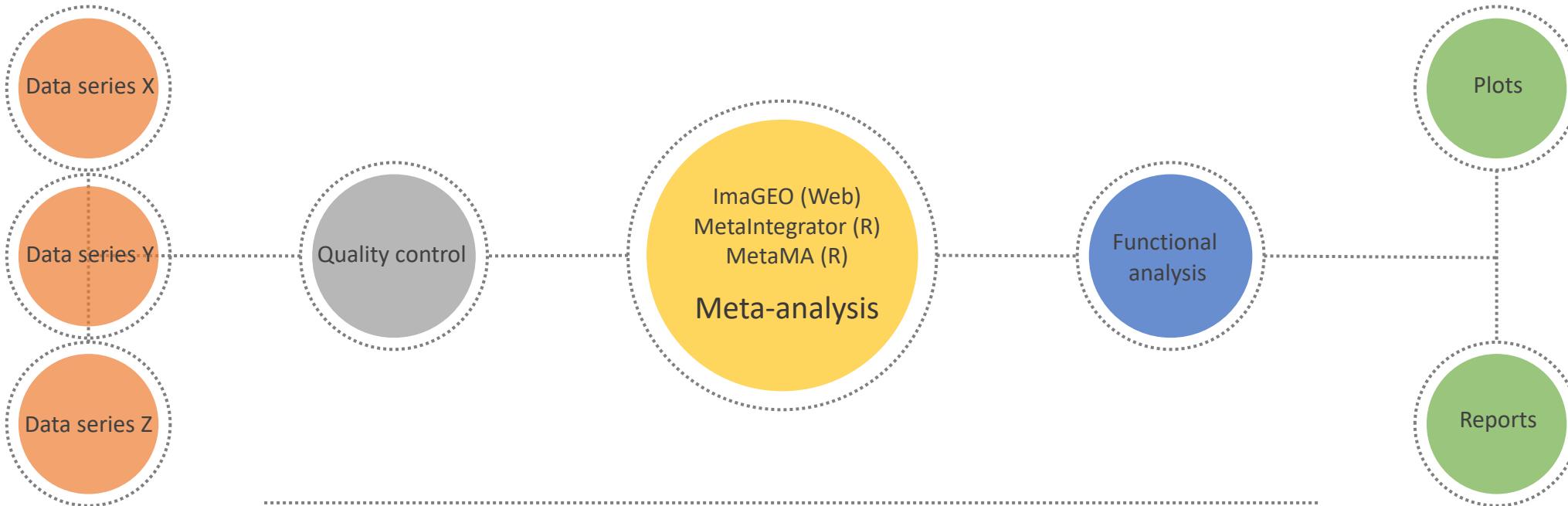
Varun Venkataramani✉, Dimitar Ivanov Tanev, [...]Thomas Kuner✉

Nature 573, 532–538 (2019) | Cite this article

41k Accesses | 172 Citations | 637 Altmetric | Metrics



Meta-analysis of GEO data series



Meta-analysis tools fetch GEO data and meta-data, perform quality checks and transformations and then apply meta-analysis methods such as effect size, Fisher's, Stouffer's, adaptively-weighted, sum or product of rank. The results are then combined and functionally characterized and compiled in user-friendly reports.



Open repositories

Fetch and manage open access datasets in R

Andigoni Malousi, PhD

Adjunct Lecturer in Bioinformatics
e-mail: andigoni@auth.gr
tel: 2310 999163
Medical School | AUTH

