Developing Original Cloud-based Bioinformatics Software Applications

Daniel Clarke, Alexander Lachmann, Avi Ma'ayan

Department of Pharmacological Sciences
Mount Sinai Center for Bioinformatics

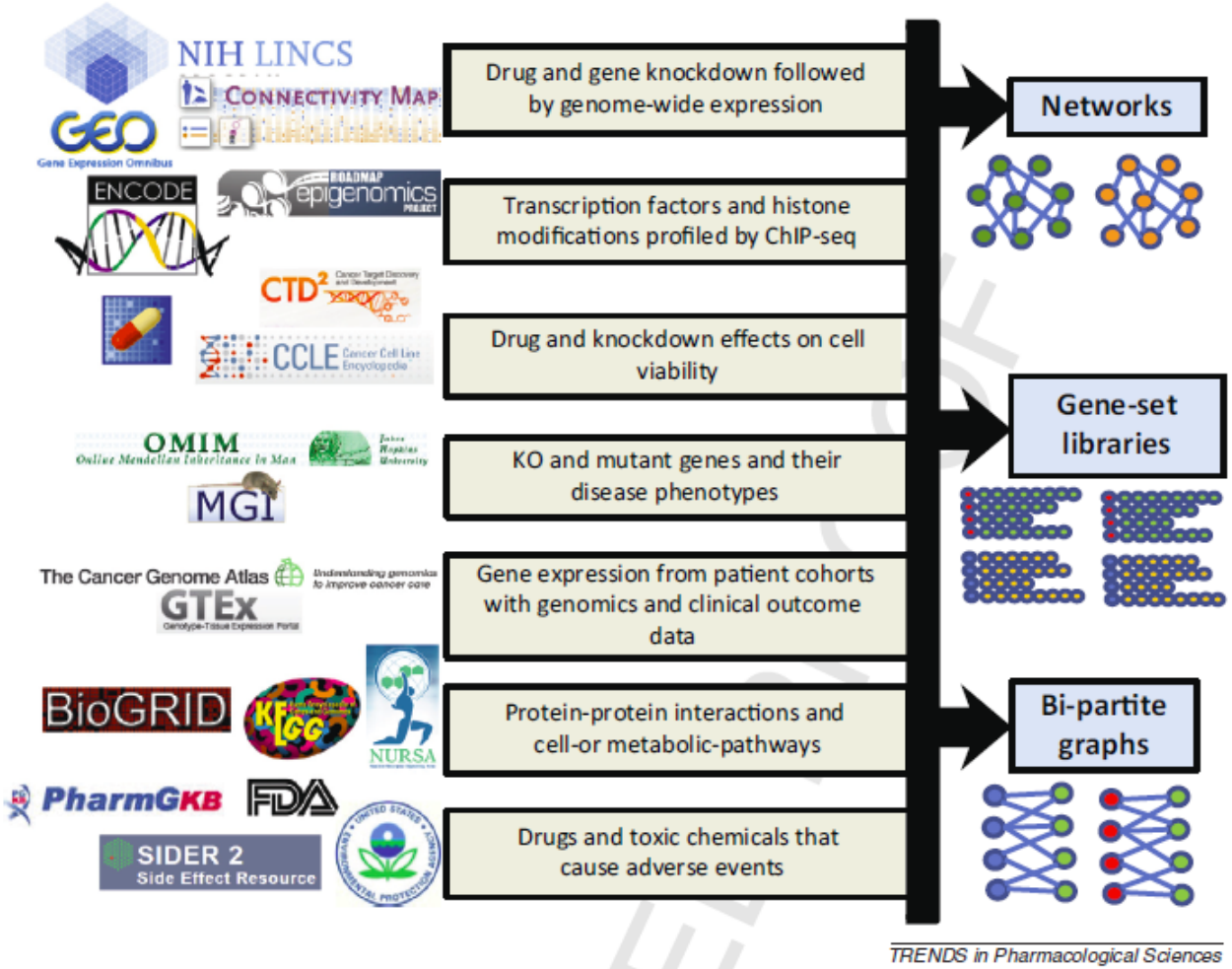
Icahn School of Medicine at Mount Sinai
New York, NY USA

BD2K-LINCS
DATA COORDINATION AND
INTEGRATION CENTER

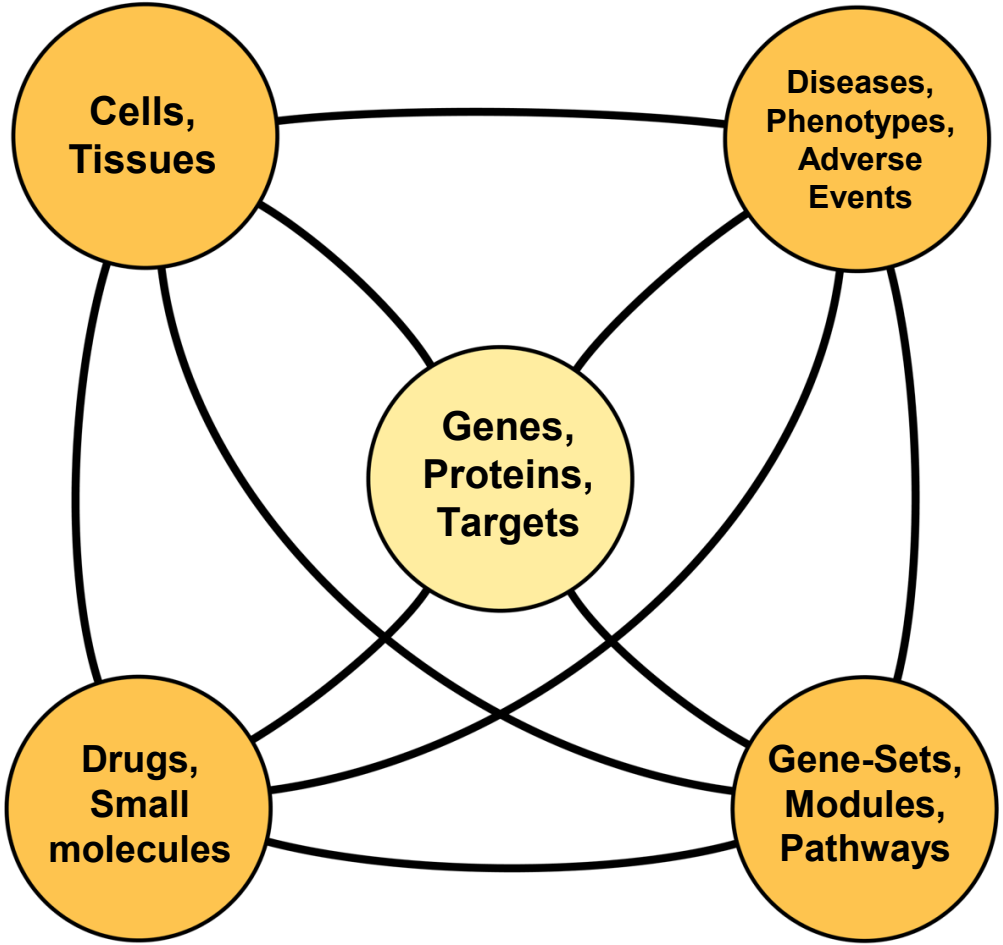
COMMON
FUND
DATA
ECOSYSTEM

IDG

LYME MIND



Ma'ayan et al. Trends Pharmacol Sci. 2014 Sep;35(9):450-60.


Cells, Tissues
Diseases, Phenotypes, Adverse Events
Genes, Proteins, Targets
Drugs, Small molecules
Gene-Sets, Modules, Pathways

Enrichr

Login | Register

32,668,268 lists analyzed
339,127 terms
171 libraries

Analyze | What's new? | Libraries | Gene search | Term search | About | Help

Input data

Choose an input file to upload. Either in BED format or a list of genes.

Try an example BED file.

Choose File No file chosen

Paste a list of valid Entrez gene symbols on each row in the text-box below. Try a gene set example.

Nsun3
Polrmt
Nlrx1
Sfxn5
Zc3h12c
Slc25a39
Arsg
Defb29
Ndufb6
Zfand1

375 gene(s) entered

Sample gene list

Contribute your list so it can be searched by others

Submit

Please acknowledge Enrichr in your publications by citing the following references:
Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics. 2013;128(14).*

Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research. 2016; gkw377* .

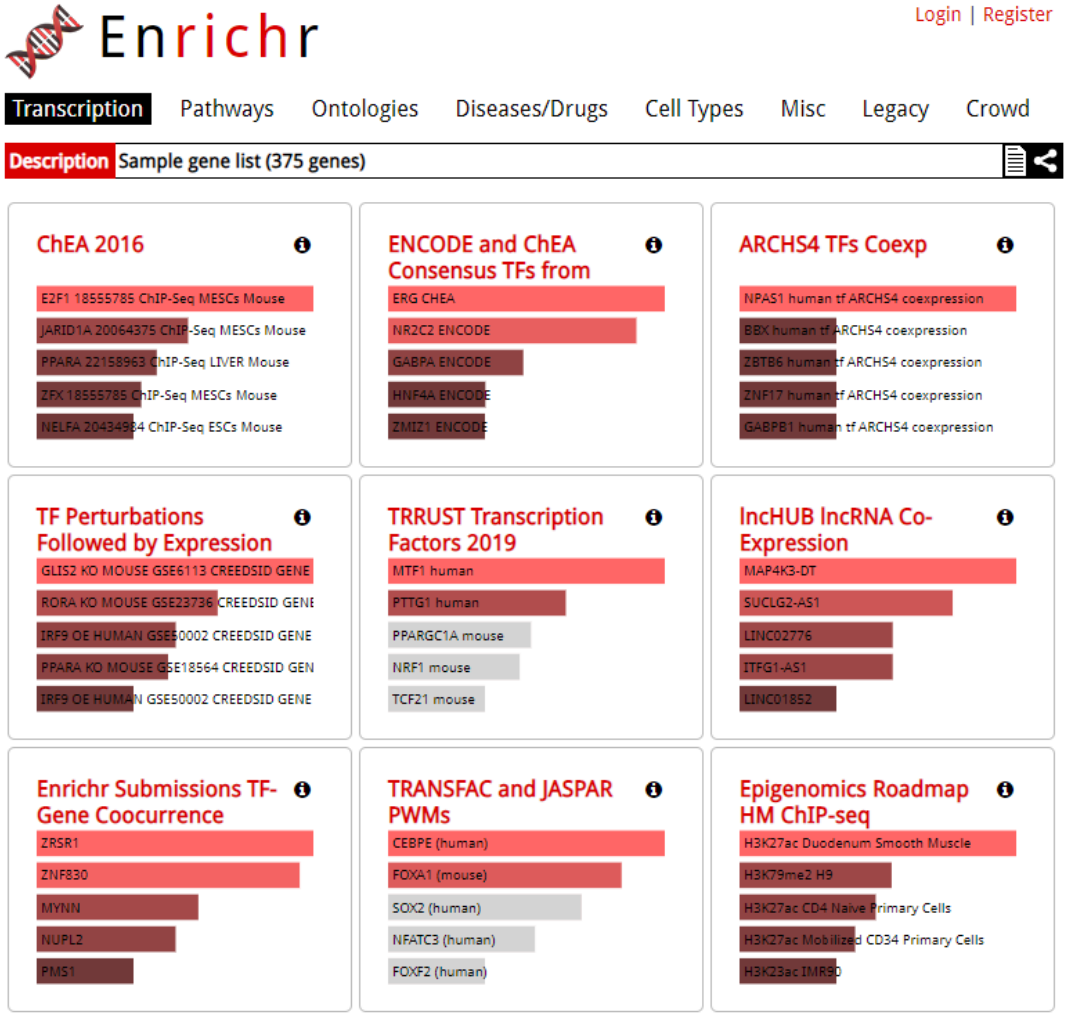
modEnrichr
A suite of gene list enrichment analysis tools

FlyEnrichr | YeastEnrichr | WormEnrichr | FishEnrichr

Click here to raise an issue on GitHub

To help us ensure continued support for Enrichr please upload a letter of support

https://maayanlab.cloud/Enrichr/



https://maayanlab.cloud/Enrichr/

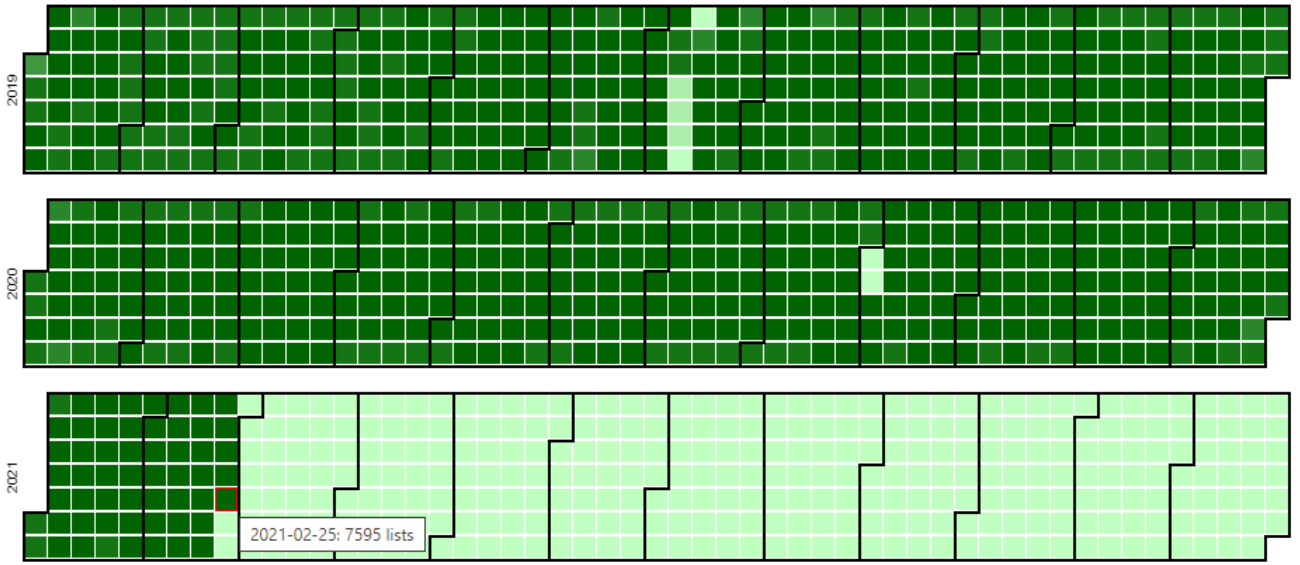
Enrichr

Login | Register

32,668,287 lists analyzed
339,127 terms
171 libraries

Analyze | What's new? | Libraries | Gene search | Term search | About | Help

Input data

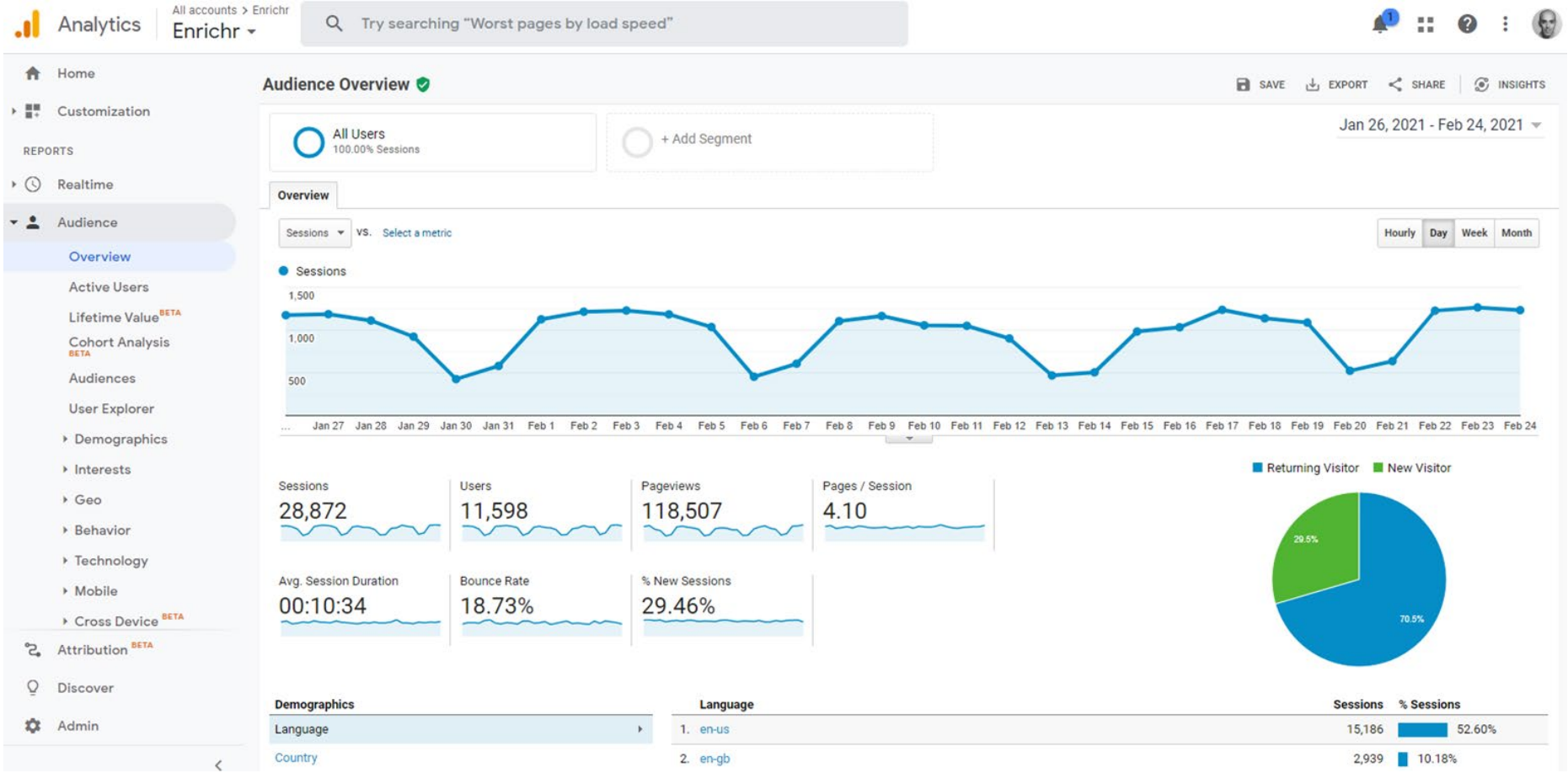


modEnrichr
A suite of gene list enrichment analysis tools

FlyEnrichr | YeastEnrichr | WormEnrichr | FishEnrichr

Click here to raise an issue on GitHub

To help us ensure continued support for Enrichr please upload a letter of support


Analytics
All accounts > Enrichr
Enrichr
Try searching "Worst pages by load speed"
Home
Customization
REPORTS
Realtime
Audience
Overview
Active Users
Lifetime Value BETA
Cohort Analysis BETA
Audiences
User Explorer
Demographics
Interests
Geo
Behavior
Technology
Mobile
Cross Device BETA
Attribution BETA
Discover
Admin
Audience Overview
SAVE
EXPORT
SHARE
INSIGHTS
All Users
100.00% Sessions
+ Add Segment
Jan 26, 2021 - Feb 24, 2021
Overview
Sessions
vs.
Select a metric
Hourly
Day
Week
Month
Sessions
1,500
1,000
500
Jan 27
Jan 28
Jan 29
Jan 30
Jan 31
Feb 1
Feb 2
Feb 3
Feb 4
Feb 5
Feb 6
Feb 7
Feb 8
Feb 9
Feb 10
Feb 11
Feb 12
Feb 13
Feb 14
Feb 15
Feb 16
Feb 17
Feb 18
Feb 19
Feb 20
Feb 21
Feb 22
Feb 23
Feb 24
Sessions
28,872
Users
11,598
Pageviews
118,507
Pages / Session
4.10
Avg. Session Duration
00:10:34
Bounce Rate
18.73%
% New Sessions
29.46%
Returning Visitor
New Visitor
29.5%
70.5%
Demographics
Language
Country
Language
Sessions
% Sessions
1. en-us
15,186
52.60%
2. en-gb
2,939
10.18%

Harmonizome

Search for genes or proteins and their functional terms extracted and organized from over a hundred publicly available resources. **Learn more**.

All

Example searches

achilles STAT3 breast cancer

Please acknowledge the Harmonizome in your publications by citing the following reference:

Rouillard AD, Gundersen GW, Fernandez NF, Wang Z, Monteiro CD, McDermott MG, Ma'ayan A. *The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins*. Database (Oxford). 2016 Jul 3;2016. pii: baw100.

Funding

BD2K-LINCS Data Coordination and Integration Center
Illuminating the Druggable Genome, Knowledge Management Center

https://maayanlab.cloud/Harmonizome/

Functional Associations

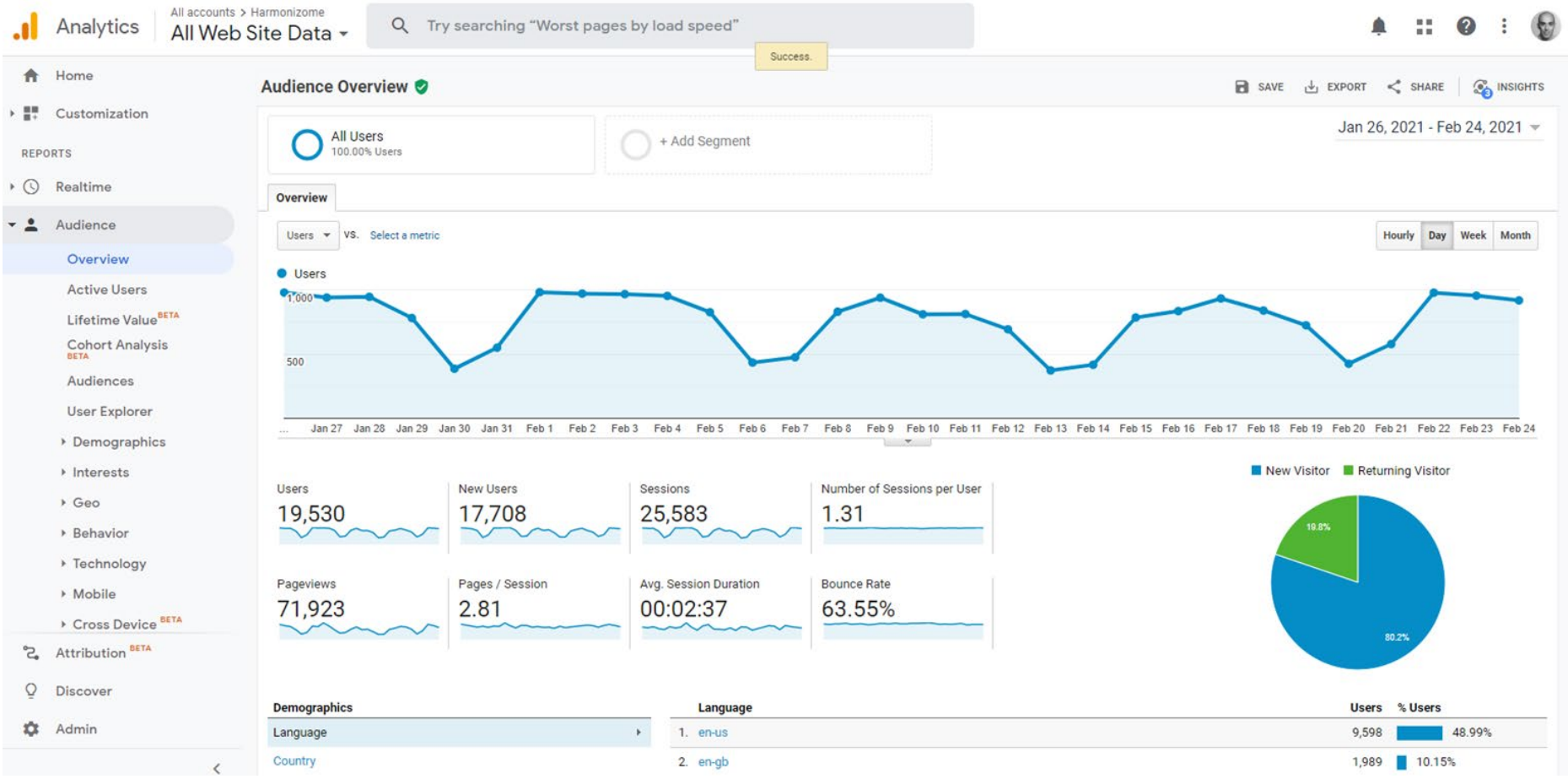
STAT3 has 14,374 functional associations with biological entities spanning 8 categories (molecular profile, organism, functional term, phrase or reference, disease, phenotype or trait, chemical, structural feature, cell line, cell type or tissue, gene, protein or microRNA) extracted from 100 datasets.

Click the + buttons to view associations for STAT3 from the datasets below.

*If available, associations are ranked by **standardized value***

	Dataset	Summary
+	Achilles Cell Line Gene Essentiality Profiles	Cell lines with fitness changed by STAT3 gene knockdown relative to other cell lines from the Achilles Cell Line Gene Essentiality Profiles dataset.
+	Allen Brain Atlas Adult Human Brain Tissue Gene Expression Profiles	Tissues with high or low expression of STAT3 gene relative to other tissues from the Allen Brain Atlas Adult Human Brain Tissue Gene Expression Profiles dataset.
+	Allen Brain Atlas Adult Mouse Brain Tissue Gene Expression Profiles	Tissues with high or low expression of STAT3 gene relative to other tissues from the Allen Brain Atlas Adult Mouse Brain Tissue Gene Expression Profiles dataset.
+	Allen Brain Atlas Developing Human Brain Tissue Gene Expression Profiles by Microarray	Tissue samples with high or low expression of STAT3 gene relative to other tissue samples from the Allen Brain Atlas Developing Human Brain Tissue Gene Expression Profiles by Microarray dataset.
+	Allen Brain Atlas Developing Human Brain Tissue Gene Expression Profiles by RNA-seq	Tissue samples with high or low expression of STAT3 gene relative to other tissue samples from the Allen Brain Atlas Developing Human Brain Tissue Gene Expression Profiles by RNA-seq dataset.
+	Allen Brain Atlas Prenatal Human Brain Tissue Gene Expression Profiles	Tissues with high or low expression of STAT3 gene relative to other tissues from the Allen Brain Atlas Prenatal Human Brain Tissue Gene Expression Profiles dataset.
+	Biocarta Pathways	Pathways involving STAT3 protein from the Biocarta Pathways dataset.
+	BioGPS Cell Line Gene Expression Profiles	Cell lines with high or low expression of STAT3 gene relative to other cell lines from the BioGPS Cell Line Gene Expression Profiles dataset.
+	BioGPS Human Cell Type and Tissue Gene Expression Profiles	Cell types and tissues with high or low expression of STAT3 gene relative to other cell types and tissues from the BioGPS Human Cell Type and Tissue Gene Expression Profiles dataset.
+	BioGPS Mouse Cell Type and Tissue Gene Expression Profiles	Cell types and tissues with high or low expression of STAT3 gene relative to other cell types and tissues from the BioGPS Mouse Cell Type and Tissue Gene Expression Profiles dataset.
+	CCLE Cell Line Gene CNV Profiles	Cell lines with high or low copy number of STAT3 gene relative to other cell lines from the CCLE Cell Line Gene CNV Profiles dataset.
+	CCLE Cell Line Gene Expression Profiles	Cell lines with high or low expression of STAT3 gene relative to other cell lines from the CCLE Cell Line Gene Expression Profiles dataset.

https://maayanlab.cloud/Harmonizome/gene/STAT3


Analytics
All accounts > Harmonizome
All Web Site Data
Try searching "Worst pages by load speed"
Success.
Home
Customization
REPORTS
Realtime
Audience
Overview
Active Users
Lifetime Value BETA
Cohort Analysis BETA
Audiences
User Explorer
Demographics
Interests
Geo
Behavior
Technology
Mobile
Cross Device BETA
Attribution BETA
Discover
Admin
Audience Overview
SAVE
EXPORT
SHARE
INSIGHTS
All Users
100.00% Users
+ Add Segment
Jan 26, 2021 - Feb 24, 2021
Overview
Users
vs.
Select a metric
Hourly
Day
Week
Month
Users
1,000
500
Jan 27 Jan 28 Jan 29 Jan 30 Jan 31 Feb 1 Feb 2 Feb 3 Feb 4 Feb 5 Feb 6 Feb 7 Feb 8 Feb 9 Feb 10 Feb 11 Feb 12 Feb 13 Feb 14 Feb 15 Feb 16 Feb 17 Feb 18 Feb 19 Feb 20 Feb 21 Feb 22 Feb 23 Feb 24
Users
19,530
New Users
17,708
Sessions
25,583
Number of Sessions per User
1.31
Pageviews
71,923
Pages / Session
2.81
Avg. Session Duration
00:02:37
Bounce Rate
63.55%
New Visitor
Returning Visitor
19.8%
80.2%
Demographics
Language
Country
Language
Users
% Users
1. en-us
9,598
48.99%
2. en-gb
1,989
10.15%

Impact of the Ma’ayan Lab Tools

>41 published bioinformatics tools and databases
>1.2 million unique users across all tools
>3,000 unique users per day
>30,000 unique users per month
>3,000 papers that cite the tools in 2020

L1000CDS2
BD2K-LINCS DCIC
An ultra-fast LINCS L1000 Characteristic Direction signature search engine

Drug-Pathway Browser
HMS LINCS
Interactive map of key signal transduction pathways and drug-target data

iLINCS
BD2K-LINCS DCIC
An integrative web platform for analysis of LINCS data and signatures

Drug Gene Budger
BD2K-LINCS DCIC
Identifies drugs & small molecules to regulate expression of target genes

Omics Integrator
NeuroLINCS
Discovers networks linking proteomic, transcriptomic, and epigenomic data

GEO2Enrichr
BD2K-LINCS DCIC
A web app and browser extension to extract and analyze signatures from GEO

RTK Profile Browser
HMS LINCS
Online tool for browsing breast cancer cell line RTK profiles

Repurposing App
LINCS Transcriptomics
Tool to explore repurposing collection of ~5000 tool compounds and drugs

Drug/Cell-line Browser
BD2K-LINCS DCIC
DCB provides interactive visualization of cancer cell-line viability data

Enrichr
BD2K-LINCS DCIC
An intuitive web-based gene list enrichment analysis tool with 133 libraries

Harmonizome
BD2K-LINCS DCIC
Web portal with a collection of 114 datasets abstracted into gene function tables

CREEDS
BD2K-LINCS DCIC
Collections of processed gene, drug and disease signatures from GEO

TieDIE
HMS LINCS
Software for computing regulatory subnetworks from mutation and transcript data

LINCS Project Mobile
BD2K-LINCS DCIC
A mobile application to explore LINCS centers and resources

Query App
LINCS Transcriptomics
Connections to user-defined signatures

LINCS Data Portal
BD2K-LINCS DCIC
Provides unified interface for searching all LINCS dataset packages and entities

CLUE Platform
LINCS Transcriptomics
Computational environment to interface with the L1000 data

Network2Canvas
BD2K-LINCS DCIC
Network visualization on a canvas with enrichment analysis

SEP L1000
BD2K-LINCS DCIC
Web portal for searching and browsing predictive small-molecule/ADR connections

Breast Cancer Browser
HMS LINCS
Online tool for browsing multiple datasets relevant to breast cancer biology

clarion
MEP LINCS
Millions of microenvironment perturbations at a click

L1000FWD
BD2K-LINCS DCIC
Large-scale visualization of drug-induced transcriptomic signatures

Drug Response Browser
HMS LINCS
Online tool for browsing breast cancer cell line drug dose-response data

LINCS Data Explorer
MEP LINCS
Visualization of the MEMA data from the MEP LINCS Center

HMS LINCS Database
HMS LINCS
Online database for access to HMS LINCS Center datasets and reagent information

LINCS Canvas Browser
BD2K-LINCS DCIC
LCB provides interactive clustering with enrichment of L1000 signatures

Touchstone App
LINCS Transcriptomics
Collection of CMap reference signatures

LJP-BCNB
BD2K-LINCS DCIC
LINCS Joint Project-Breast Cancer Network Browser is an interactive viz of 2344 signatures

eXpression2Kinases
BD2K-LINCS DCIC
Linking expression signatures to upstream cell signaling networks

PAEA
BD2K-LINCS DCIC
Enrichment analysis tool implementing the principal angle method

Adaptation Browser
HMS LINCS
Online tool for browsing melanoma cell line adaptive resistance mechanisms

Slicr
BD2K-LINCS DCIC
A metadata search engine that provides easy access to L1000 GEO data

cynetworkbrowser
BD2K-LINCS DCIC
An interactive tool for generating and viewing protein interaction networks

SynergySeq
BD2K-LINCS DCIC
Identify synergistic drug combinations

Datasets2Tools
BD2K-LINCS DCIC
A platform for the discovery and evaluation of biomedical digital objects

PIUMet
NeuroLINCS
Tool for discovering pathways of peaks from untargeted metabolomics data

Morpheus App
LINCS Transcriptomics
An interactive version of the ICV that lets you manipulate and annotate an existing dataset

ICV App
LINCS Transcriptomics
A matrix-based interactive heatmap to explore relationships within the data

GEN3VA
BD2K-LINCS DCIC
Aggregates and analyzes gene expression signatures extracted from GEO

Cell Dynamics Browser
HMS LINCS
Online tool for live-cell image data visualization

AChroMap
NeuroLINCS
A data integration tool for transcriptomic and epigenomic data

Panorama
LINCS PCCSE
Proteomics repository that provides access to the PCCSE P100 and GCP data

GF Response Browser
HMS LINCS
Online tool for browsing growth factor-stimulated cell signaling profiles

BioJupies
BD2K-LINCS DCIC
Automatically generates RNA-seq data analysis notebooks

GUIdock
BD2K-LINCS DCIC
Docker package containing a computational environment to run apps with a GUI

GR Browser
HMS LINCS
BD2K-LINCS DCIC
Online tool for calculating dose-response metrics and browsing response data

MEPmosaic
BD2K-LINCS DCIC
Mosaic visualization of high-content imaging assays of three cancer cell types

slinky
R package to query the L1000 metadata via the CLUE.io REST API

PiNET
BD2K-LINCS DCIC
Used to annotate, map and analyze a set of peptide moieties

OMERO
HMS LINCS
Client-server software for image data visualization, management, and analysis

piLINCS
BD2K-LINCS DCIC
Provides interface and access to LINCS proteomics datasets available in Panorama

Phosphosite Plus
BD2K-LINCS DCIC
Comprehensive resource for the study of protein post-translational modifications

ImageRail
HMS LINCS
Desktop software for formatting and analyzing high-throughput image data

ARCHS4
BD2K-LINCS DCIC
Massive mining of publicly available RNA-seq data from human and mouse

DToxS Portal
DToxS
Web portal to provide access to the DToxS Center's data, metadata and analyses

GREIN
BD2K-LINCS DCIC
Interactive web platform for re-analyzing GEO RNA-seq data

https://lincsproject.org/LINCS/tools

BioJupies

Search GEO datasets (e.g. cancer, p53)

Chrome Extension Help Sign In

Step 1.
Upload or Fetch RNA-seq Data

- Upload your raw or processed RNA-seq data
- Fetch >8,000 public RNA-seq datasets published in the Gene Expression Omnibus

Step 2.
Select Data Analysis Tools

- Select from multiple state-of-the-art RNA-seq data analysis tools
- Contribute your computational tool as a plugin

Step 3.
Generate Your Notebook

- Access and share your results through a permanent URL
- Download, rerun and customize your notebook using Docker

BioJupies Automatically Generates RNA-seq Data Analysis Notebooks

With BioJupies you can produce in seconds a customized, reusable, and interactive report from your own raw or processed RNA-seq data through a simple user interface

Get Started

To acknowledge BioJupies in your publications, please use the following reference: Torre, D., Lachmann, A., and Ma'ayan, A. (2018). BioJupies: Automated Generation of Interactive Notebooks for RNA-Seq Data Analysis in the Cloud. Cell Systems.

BioJupies now supports user accounts! Sign in from the top right corner of the page for access to unlimited private notebooks, RNA-seq datasets and alignment jobs.

https://maayanlab.cloud/biojupies/

ARCHS4 Search Visualize Download Chrome Extension Help Enter gene symbol... Search

Species

Human Mouse Sample Gene

Search

Metadata Signature Enrichment

Metadata Search

Human examples:
GSE81547, GSM2679484, Macrophage, Brain

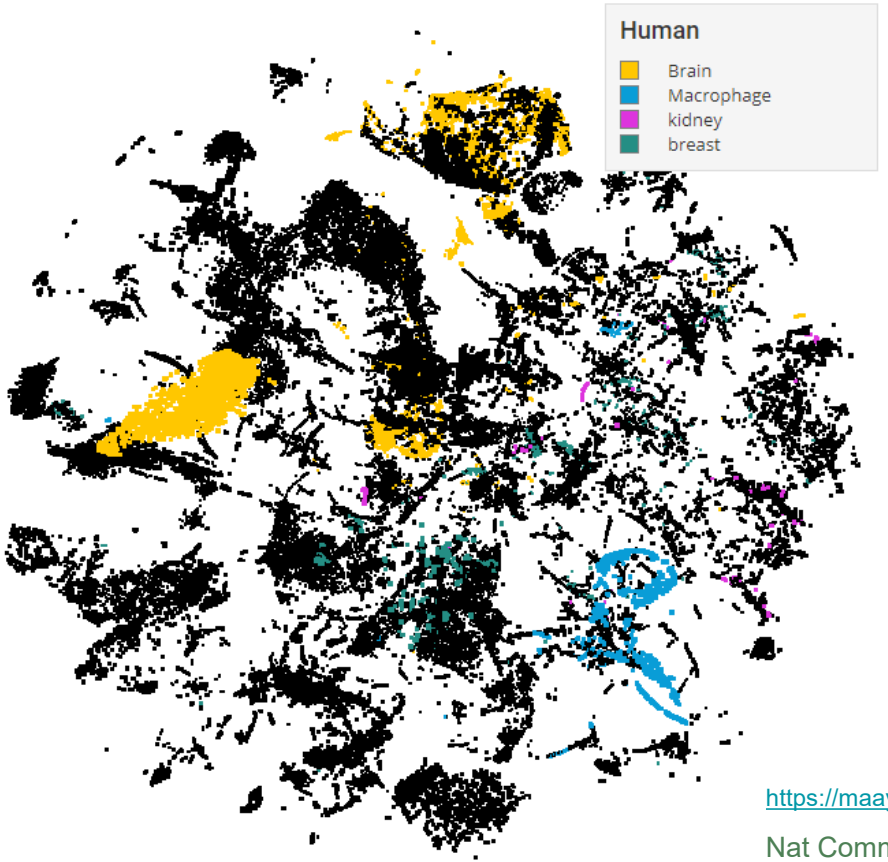
breast Search

Tissue Types

- Cardiovascular System
- Connective Tissue
- Digestive System
- Immune System
- Integumentary System
- Muscular System
- Nervous System
- Respiratory System
- Urogenital System

Cell Lines

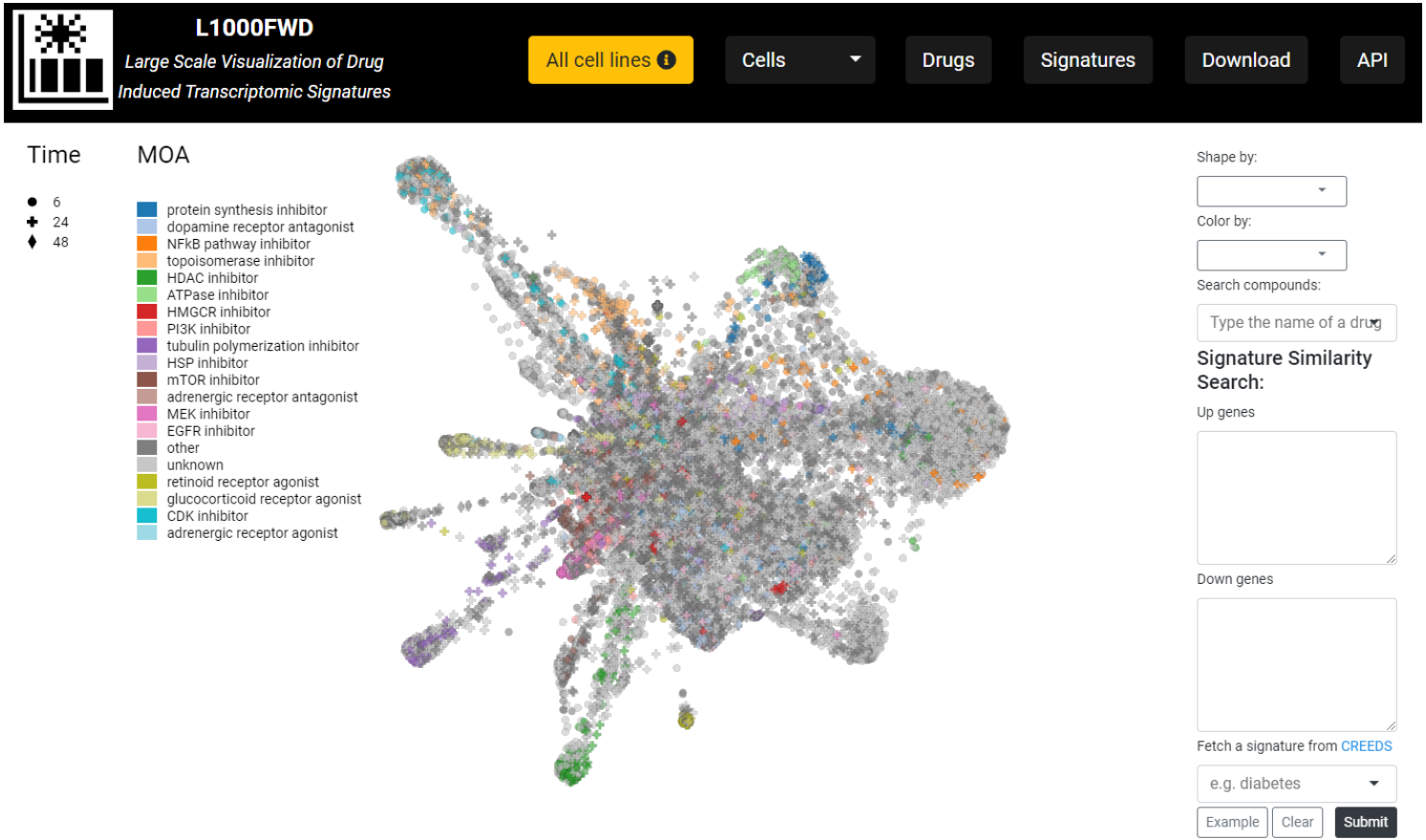
t-SNE view



https://maayanlab.cloud/archs4/data.html

Nat Commun. 2018 Apr 10;9(1):1366.

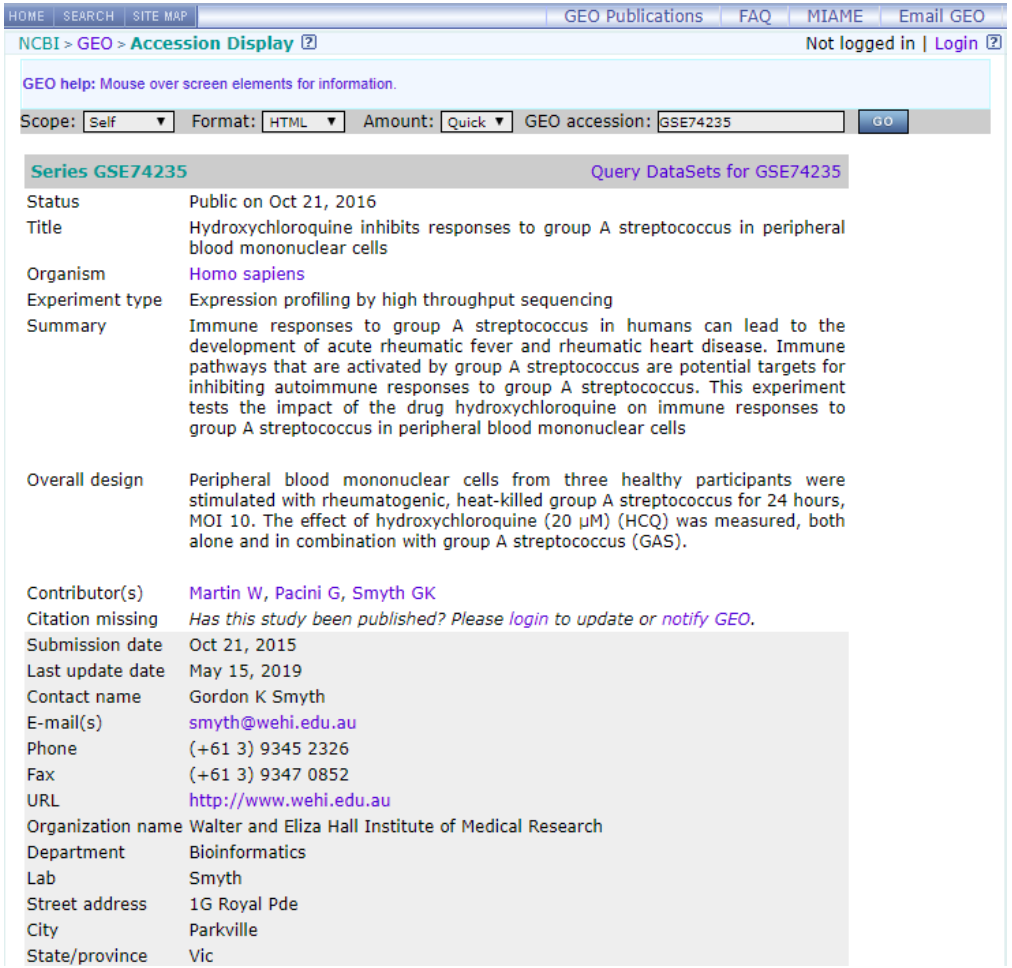
The L1000FWD Map - ~17K Signatures, ~5K Drugs



https://maayanlab.cloud/L1000FWD/main

Bioinformatics. 2018 Jun 15;34(12):2150-2152

What are the molecular effects of hydroxychloroquine on human cells?



HOME | SEARCH | SITE MAP | GEO Publications | FAQ | MIAME | Email GEO

NCBI > GEO > **Accession Display** | Not logged in | Login

GEO help: Mouse over screen elements for information.

Scope: Self | Format: HTML | Amount: Quick | GEO accession: GSE74235 | GO

Series GSE74235 — Query DataSets for GSE74235

Status	Public on Oct 21, 2016
Title	Hydroxychloroquine inhibits responses to group A streptococcus in peripheral blood mononuclear cells
Organism	Homo sapiens
Experiment type	Expression profiling by high throughput sequencing
Summary	Immune responses to group A streptococcus in humans can lead to the development of acute rheumatic fever and rheumatic heart disease. Immune pathways that are activated by group A streptococcus are potential targets for inhibiting autoimmune responses to group A streptococcus. This experiment tests the impact of the drug hydroxychloroquine on immune responses to group A streptococcus in peripheral blood mononuclear cells
Overall design	Peripheral blood mononuclear cells from three healthy participants were stimulated with rheumatogenic, heat-killed group A streptococcus for 24 hours, MOI 10. The effect of hydroxychloroquine (20 µM) (HCQ) was measured, both alone and in combination with group A streptococcus (GAS).
Contributor(s)	Martin W, Pacini G, Smyth GK
Citation missing	*Has this study been published? Please login to update or notify GEO.*
Submission date	Oct 21, 2015
Last update date	May 15, 2019
Contact name	Gordon K Smyth
E-mail(s)	smyth@wehi.edu.au
Phone	(+61 3) 9345 2326
Fax	(+61 3) 9347 0852
URL	http://www.wehi.edu.au
Organization name	Walter and Eliza Hall Institute of Medical Research
Department	Bioinformatics
Lab	Smyth
Street address	1G Royal Pde
City	Parkville
State/province	Vic

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE74235

BioJupies

Search GEO datasets (e.g. cancer, p53)

Chrome Extension Help Sign In

BioJupies Automatically Generates RNA-seq Data Analysis Notebooks

With BioJupies you can produce in seconds a customized, reusable, and interactive report from your own raw or processed RNA-seq data through a simple user interface

Get Started

https://maayanlab.cloud/biojupies/

Cell Systems 2018 Nov 28;7(5):556-561.e3.

BioJupies

Search GEO datasets (e.g. cancer, p53)

Chrome Extension Help Sign In

Which dataset would you like to analyze?

Use the form below to **search 9,145 publicly available datasets** published in the Gene Expression Omnibus database and processed by ARCHS4.

hydroxychloroquine

Displaying 1-1 of 1 results | Organism: All | Sort by: Newest | Samples: 6 35 70+

Hydroxychloroquine inhibits responses to group A streptococcus in peripheral blood mononuclear cells

GSE74235 *15 samples* Published *October 2016*

Analyze >

More Info

<< < 1 > >>

BioJupies

Search GEO datasets (e.g. cancer, p53)

Chrome Extension Help Sign In

Which analyses would you like to perform?

Use the form below to **add or remove data analysis and visualization tools** to your notebook. These tools will analyze the selected dataset and embed interactive results in your notebook. Once you have selected the desired tools, click **Continue** to proceed.

‹ Back Continue ›

Exploratory Data Analysis

These tools assist in visually exploring global patterns within the dataset.

PCA

Linear dimensionality reduction technique to visualize similarity between samples

Remove − More Info ▾

Clustergrammer

Interactive hierarchical clustering heatmap visualization

Remove − More Info ▾

Library Size Analysis

Analysis of readcount distribution for the samples within the dataset

Remove − More Info ▾

Differential Expression Analysis

These tools allow for calculation and exploration of differential gene expression between two groups of samples.

Differential Expression Table

Differential expression analysis between two groups of samples

Add + More Info ▾

Volcano Plot

Add +

Which samples would you like to compare?

One or more of the selected tools require generating a **gene expression signature**. To generate one, you must **define two groups of samples** whose gene expression you wish to compare by using the form below. Once you have defined the desired groups, click **Continue** to proceed.

‹ Back | Continue ›

What are the names of the groups?

First, **name the groups of samples** you wish to compare using the text boxes below. It is recommended to use names descriptive of the groups' experimental condition: for example, *WT, DMSO, Vehicle* for the **Control** group; *Metastatic, p53 KO, Dasatinib-6h* for the **Perturbation** group.

Untreated *6 samples* vs Perturbation *3 samples* Predict Groups

Which groups do the samples belong to?

Second, **select the samples** you wish to assign to the groups by using the dropdown menus below. To improve the statistical validity of the gene expression signature, each group must have **at least three samples**.

Group	Accession	Sample Title	Length Of Culture	Subject Id#	Treated With
Untreated	GSM1915090	Untreated 0h Patient 2	0 hours	2	none (untreated)
Perturbation	GSM1915091	HCQ 24h Patient 2	24 hours	2	20 μM hydroxychloroquine (HCQ)
Untreated	GSM1915092	Untreated 24h Patient 1	24 hours	1	none (untreated)
Perturbation	GSM1915093	HCQ 24h Patient 1	24 hours	1	20 μM hydroxychloroquine (HCQ)
Untreated	GSM1915094	Untreated 0h Patient 1	0 hours	1	none (untreated)

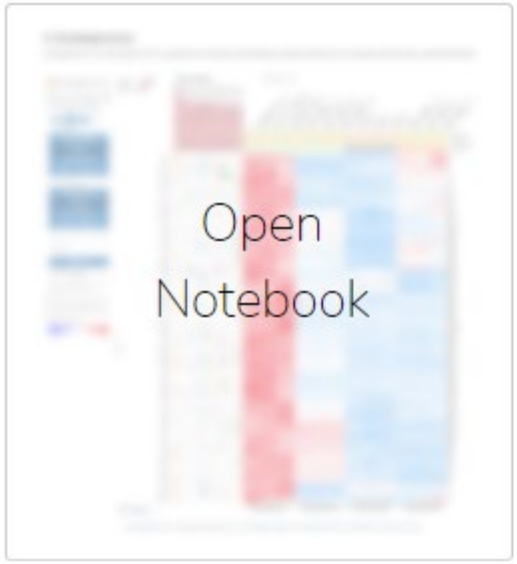
BioJupies

Search GEO datasets (e.g. cancer, p53)

Chrome Extension Help Sign In

Results

New Notebook Open Notebook



Untreated vs Perturbation Analysis Notebook | BioJupies

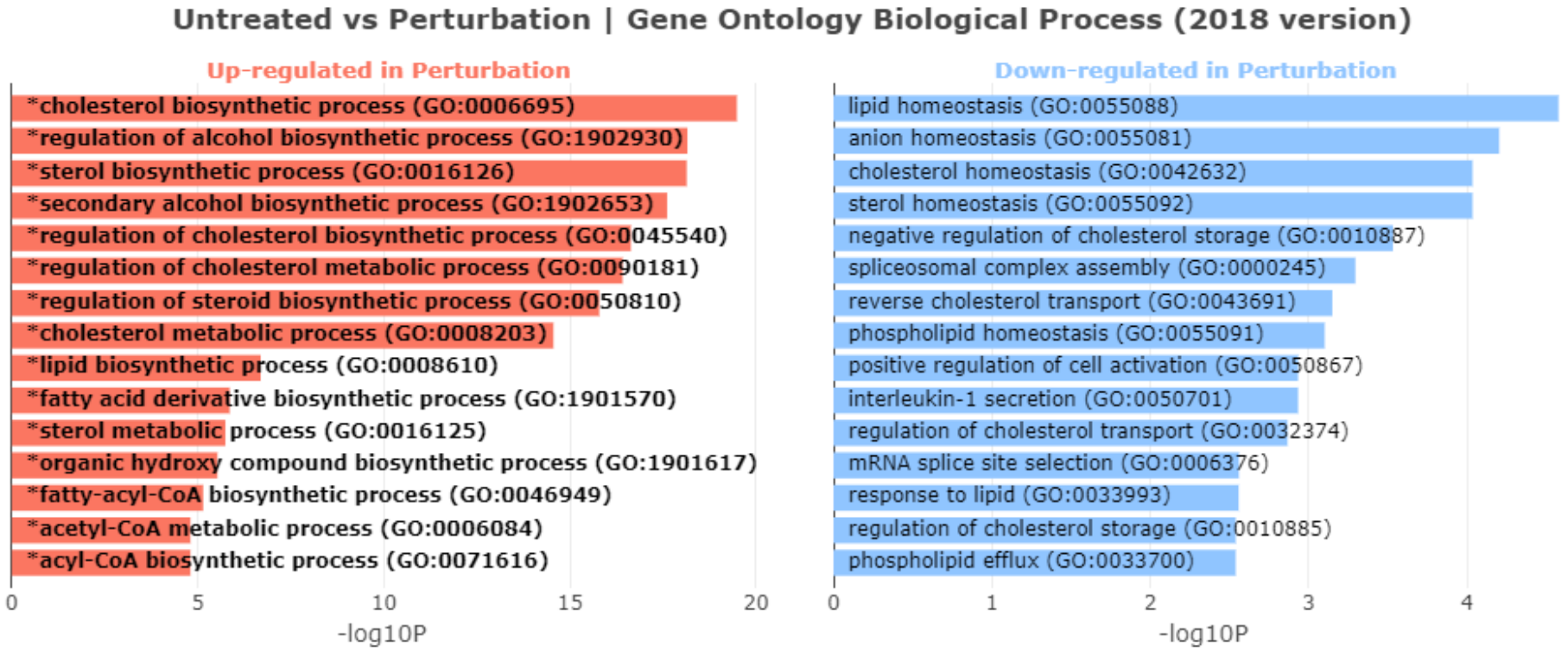
Dataset: *GSE74235*

Signature: *Untreated* vs *Perturbation*

Analysis Tools: PCA | Clustergrammer | Library Size Analysis | Differential Expression Table | Enrichr Links | Gene Ontology Enrichment Analysis | Pathway Enrichment Analysis | L1000FWD Query

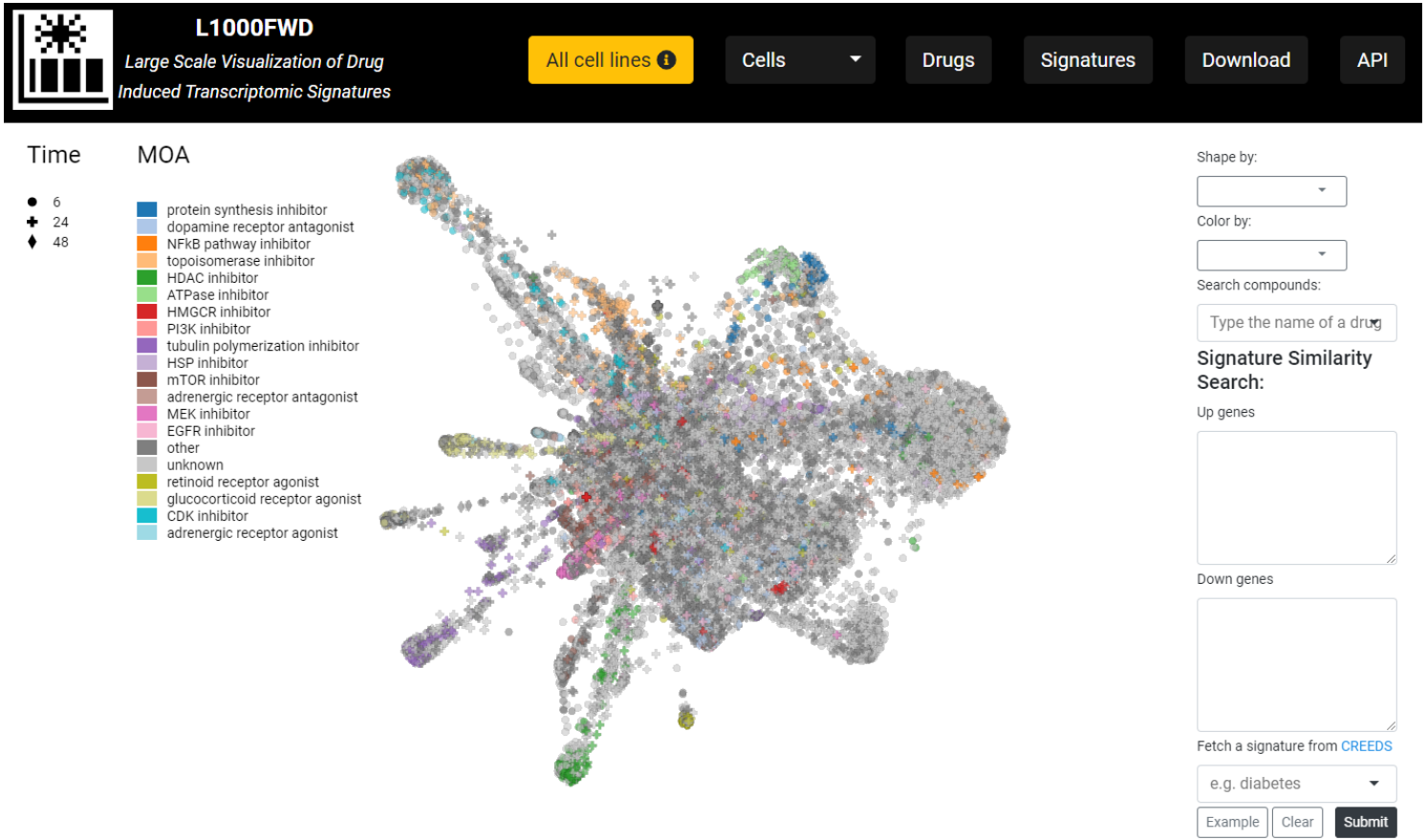
Tweet E-mail Copy Link

Effects of Hydroxychloroquine on Uninfected PBMCs



https://amp.pharm.mssm.edu/biojupies/notebook/ZxbvpHAn5

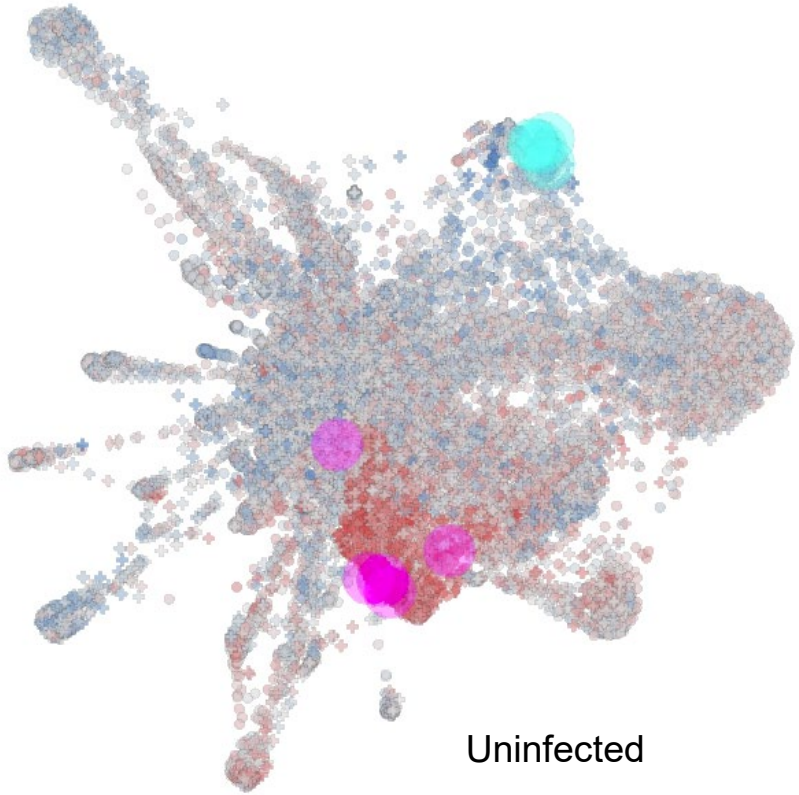
The L1000FWD Map - ~17K Signatures, ~5K Drugs



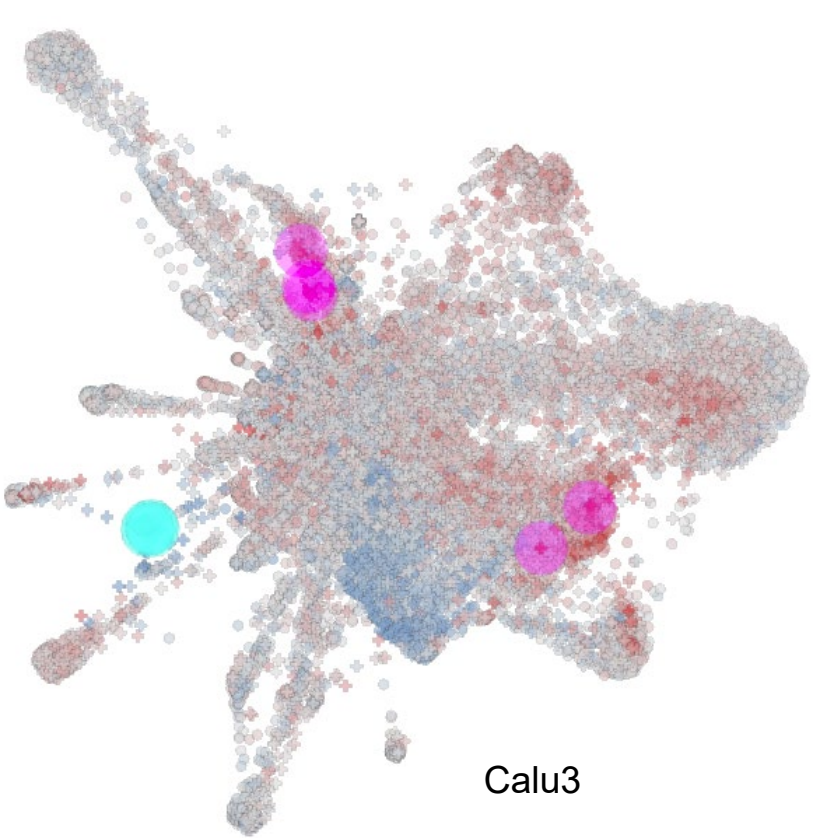
https://maayanlab.cloud/L1000FWD/main

Bioinformatics. 2018 Jun 15;34(12):2150-2152

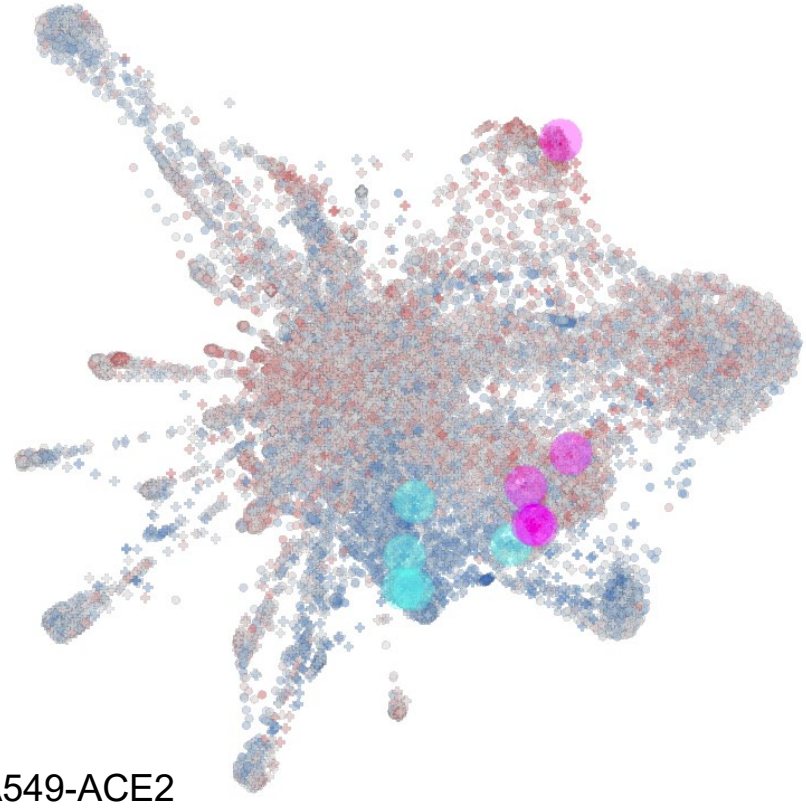
Projecting the Hydroxychloroquine Signatures onto the L1000FWD Map



Projecting the SARS-CoV-2 Signatures onto the L1000FWD Map



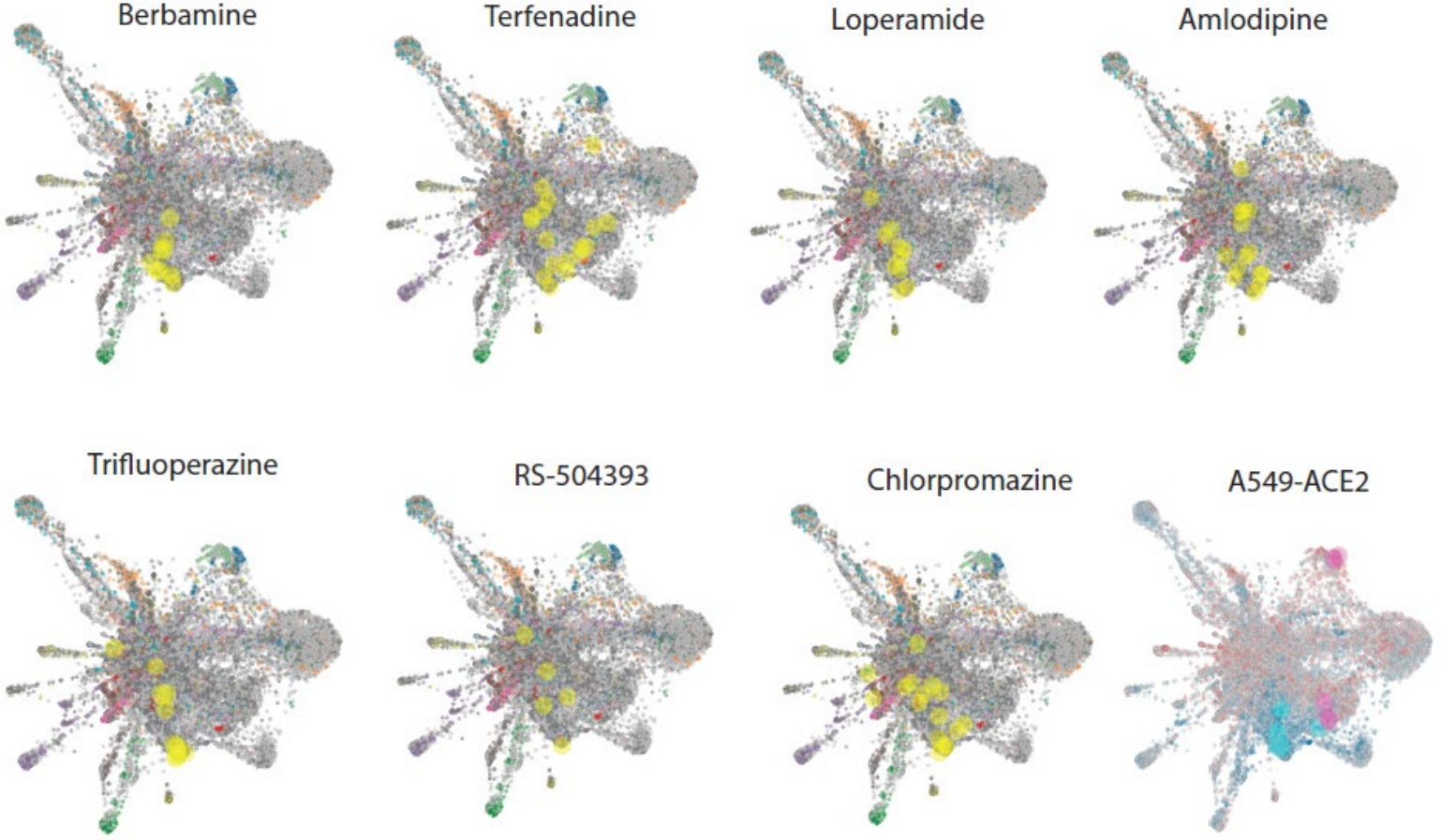
Calu3



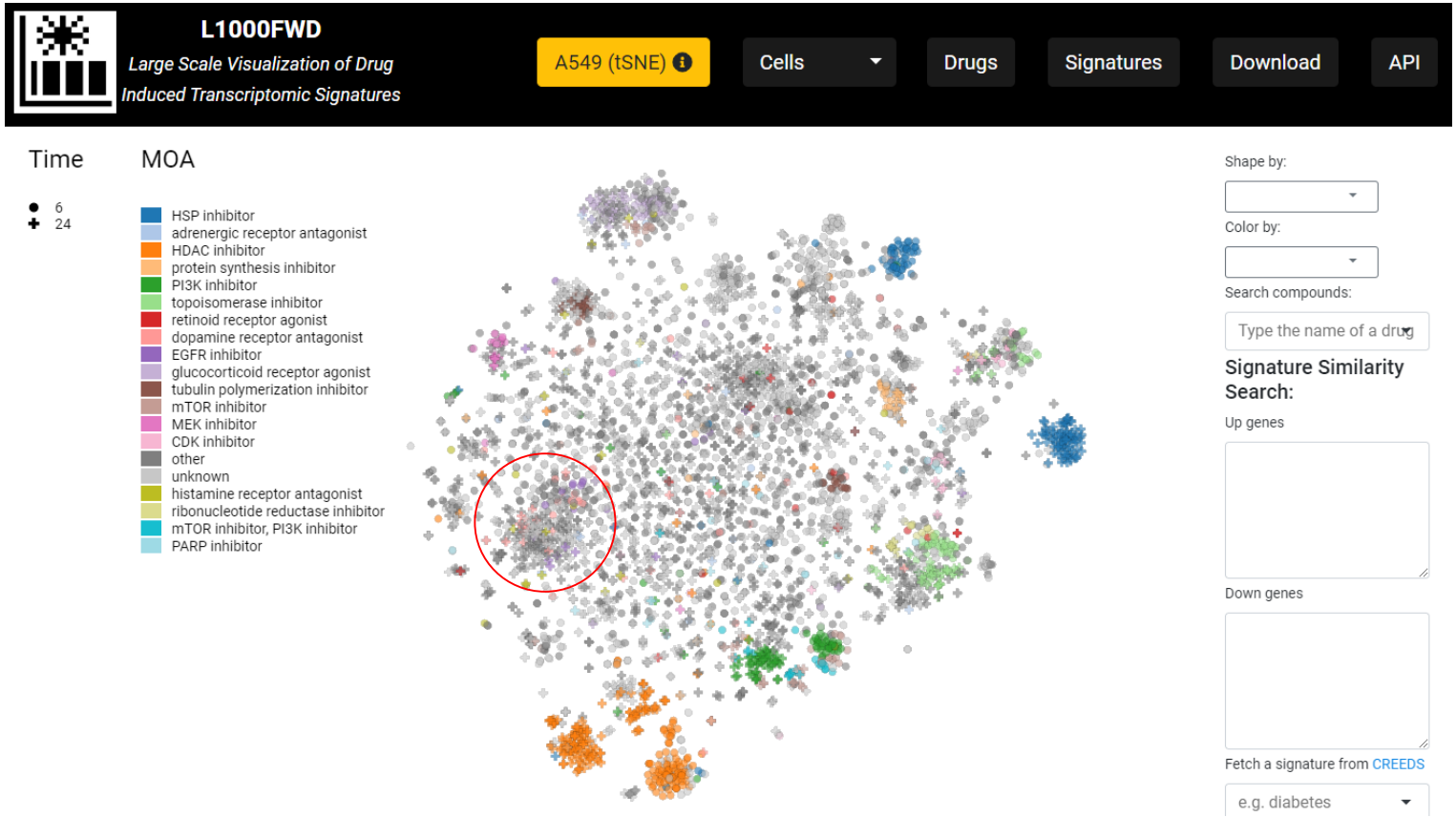
A549-ACE2

Cell. 2020 May 28;181(5):1036-1045.e9.

Drugs Hitting the Same Region in Expression Space



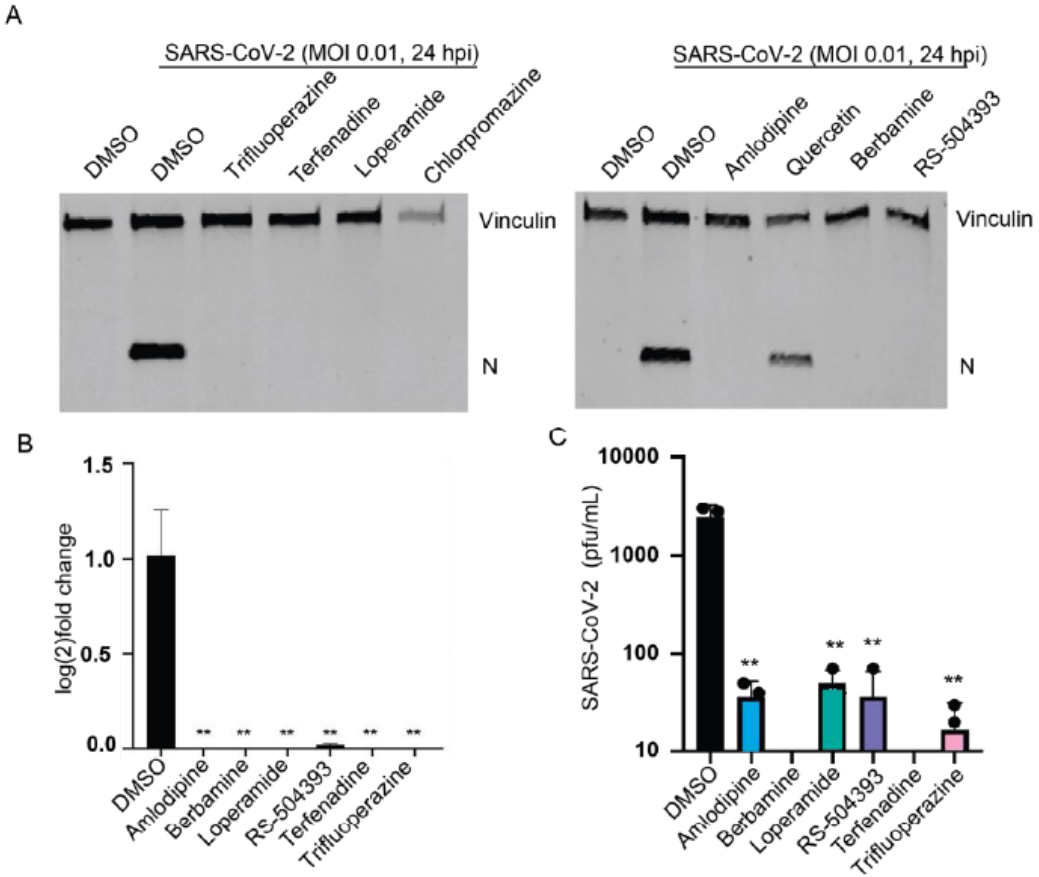
The L1000FWD Map for the A549 Cell-Line



https://maayanlab.cloud/L1000FWD/graph_page/A549-tSNE_layout.csv

Bioinformatics. 2018 Jun 15;34(12):2150-2152

TenOever Lab (Mount Sinai)



Daisy A. Hoagland, Daniel J.B. Clarke, Rasmus Møller, Yuling Han, Liuliu Yang, Megan L. Wojciechowicz, Alexander Lachmann, Kasopefoluwa Y. Oguntuyo, Christian Stevens, Benhur Lee, Shuibing Chen, Avi Ma'ayan, Benjamin R tenOever. **Modulating the transcriptional landscape of SARS-CoV-2 as an effective method for developing antiviral compounds.** bioRxiv 2020.07.12.199687

The COVID-19 Drug and Gene Set Library

A collection of drug and gene sets related to COVID-19 research contributed by the community

355 drug sets | 2003 unique drugs | 575 gene sets | 21818 unique genes

16 in-vitro drug screen drug sets

In-vitro drug screens to identify compounds that block SARS-CoV-2 replication in cell-based reporter assays.

106 differential expression signatures gene sets

Genome-wide differentially expressed genes in cells and tissues due to SARS-CoV-2 infection profiled with RNA-seq.

17 CRISPR screen gene sets

Genome-wide CRISPR screens to identify hits that increase or decrease SARS-CoV-2 proliferation in cell-based assays.

Drug sets | Gene sets

Experimental drug sets | Computational drug sets | Twitter drug sets | All drug sets

Draw a Venn diagram

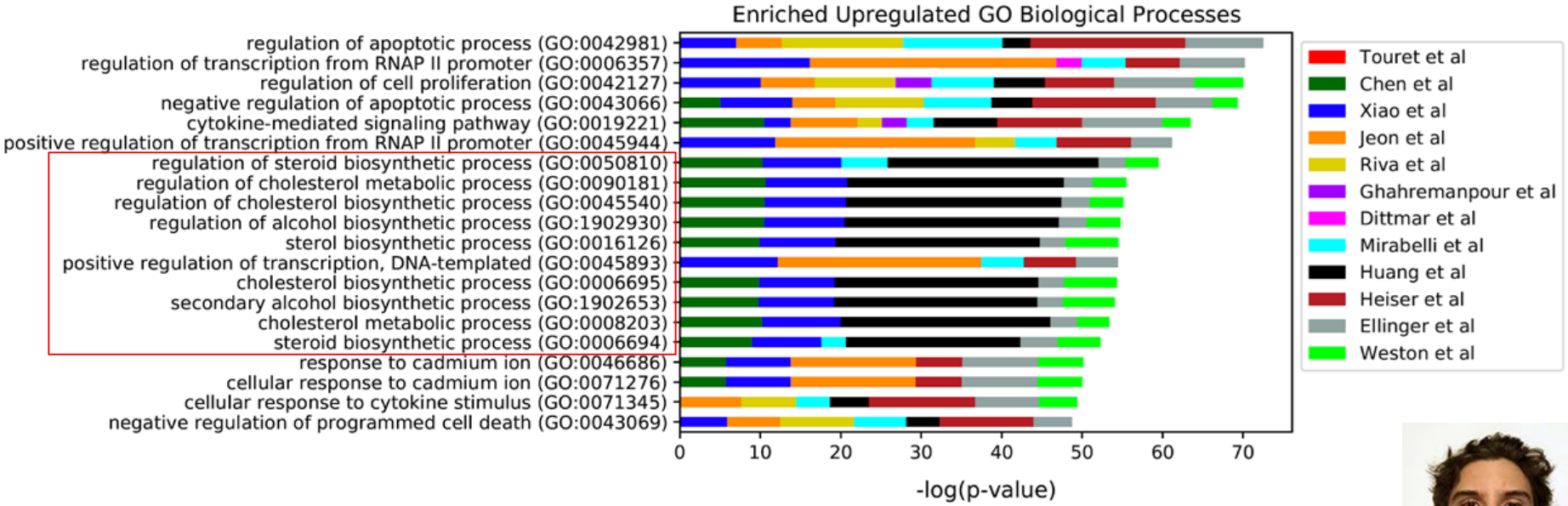
Search in description, metadata or drugs: e.g. Remdesivir

Description	Drugs	DrugEnrichr link
15 SARS-CoV-2 inhibitors in Vero E6 cells from Jan et al. 2021	15 drugs	
24 FDA-approved drugs inhibiting SARS-CoV-2 from Xiao et al.	24 drugs	
25 compounds against SARS-CoV-2 via AlphaLISA RBD-ACE2 assay from Hanson et	25 drugs	

https://maayanlab.cloud/covid19/

Patterns 2020 Sep 11;1(6):100090.

Drugs from Published In-Vitro Screens are Hitting the Same Region in Expression Space



https://appyters.maayanlab.cloud/#/Drugmonizome_Consensus_Terms

appyters

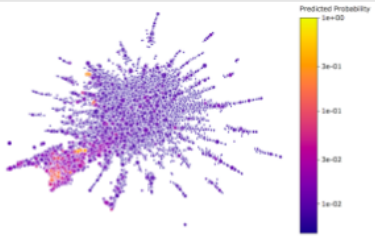
A catalog of appyter notebooks

What is an Appyter? | Creating Appyters | Publishing Appyters | About

Search appyters...

Drugmonizome

RNA-seq scRNA-seq Enrichr Machine Learning Harmonizome L1000 Compare Sets microRNAs Kinome Aging



Drugmonizome-ML

Views: 377 Starts: 357
Runs: 169 Retrievals: 131

Machine learning pipeline to predict novel drug indications from small molecule attributes

v0.0.10 Apache-2.0 Machine Learning Drugmonizome

Select

Drugmonizome
Extract, Transform, Load

Drugmonizome ETL: DrugRepurposingHub

Views: 47 Starts: 26
Runs: 9 Retrievals: 9

An appyter to process drug-target and drug-mechanism of action associations from the Drug Repurposing Hub.

v0.0.2 MIT Drug Repurposing Hub Drugmonizome ETL Script

Select

Drugmonizome Consensus Terms

Views: 45 Starts: 49
Runs: 29 Retrievals: 82

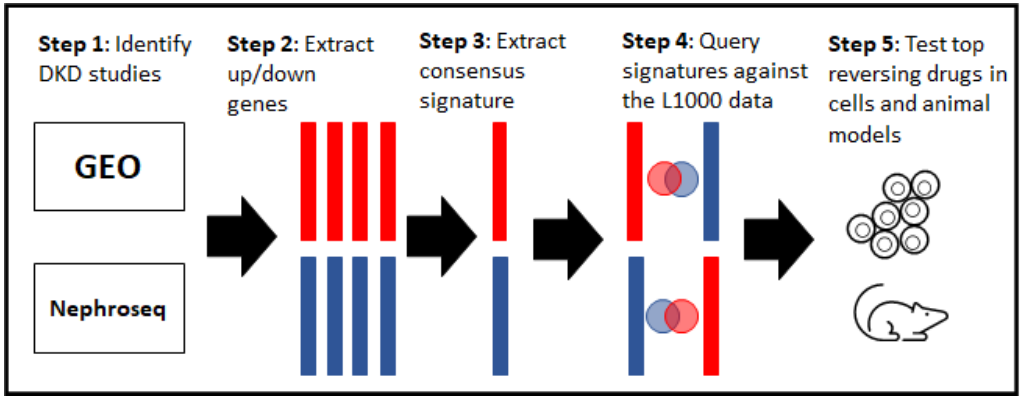
An appyter that queries an input collection of drug sets against libraries in Drugmonizome and returns the top enriched consensus terms

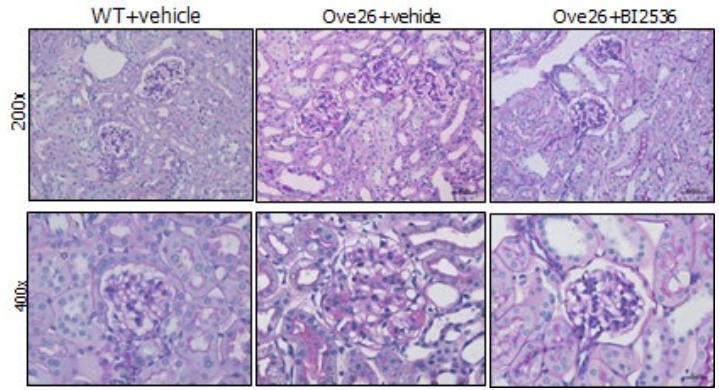
v0.0.3 MIT Drugmonizome Drug Set Enrichment

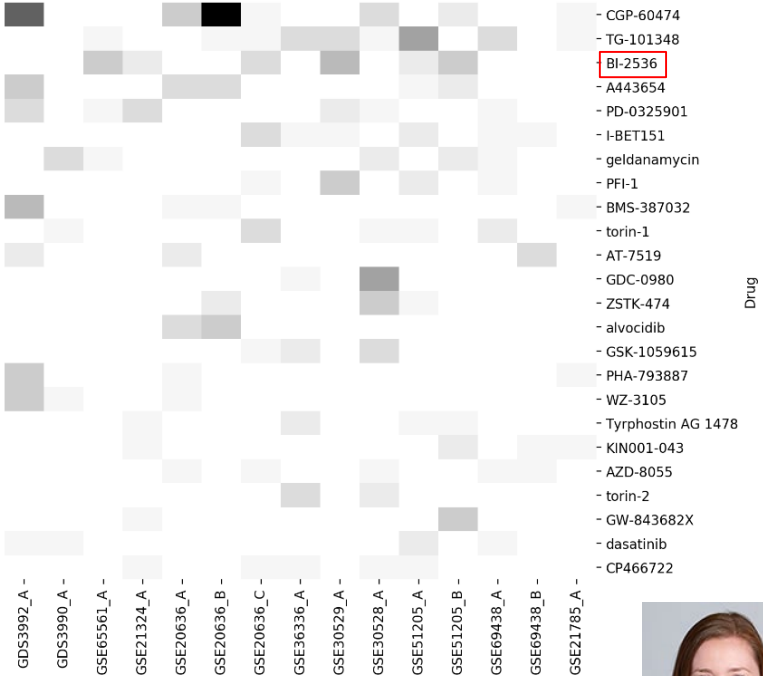
Select

https://appyters.maayanlab.cloud/#/?tags=Drugmonizome

PLK1 Inhibitor as a Potential Drug to Treat Diabetic Kidney Disease







https://appyters.maayanlab.cloud/#/L1000FWD_Consensus_Drugs

Diabetes. 2020 Oct 16;db200580.

appyters

A catalog of appyter notebooks

What is an Appyter? | Creating Appyters | Publishing Appyters | About

consensus

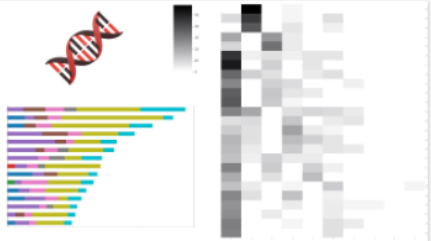
RNA-seq scRNA-seq Enrichr Machine Learning Harmonizome Drugmonizome L1000 Compare Sets microRNAs Kinome Aging

Drugmonizome Consensus Terms

Views: 66 Starts: 66
Runs: 75 Retrievals: 92

An appyter that queries an input collection of drug sets against libraries in Drugmonizome and returns the top enriched consensus terms

v0.0.3 CC-BY-NC-SA-4.0 Drugmonizome Drug Set Enrichment



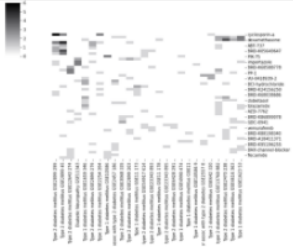
Enrichr Consensus Terms

Views: 161 Starts: 192
Runs: 191 Retrievals: 200

This appyter returns consensus Enrichr terms using a set of gene sets

v0.2.6 CC-BY-NC-SA-4.0 Enrichr Gene Set Enrichment

Select



L1000FWD Consensus Drugs

Views: 141 Starts: 183
Runs: 194 Retrievals: 118

Returns Consensus mimicker and reverser drugs from a set of up and down genesets

v0.1.3 CC-BY-NC-SA-4.0 L1000FWD L1000

Select

https://appyters.maayanlab.cloud/#/?q=consensus

https://lincsproject.org/LINCS/tools

Please cite this article in press as: Clarke et al., Appyters: turning Jupyter Notebooks into data-driven web apps, Patterns (2021), https://doi.org/10.1016/j.patter.2021.100213

Patterns

CellPress
OPEN ACCESS

Article

Appyters: turning Jupyter Notebooks into data-driven web apps

Daniel J.B. Clarke,[1] Minji Jeon,[1] Daniel J. Stein,[1] Nicole Moiseyev,[1] Eryk Kropiwnicki,[1] Charles Dai,[1] Zhuorui Xie,[1] Megan L. Wojciechowicz,[1] Skylar Litz,[1] Jason Hom,[1] John Erol Evangelista,[1] Lucas Goldman,[1] Serena Zhang,[1] Christine Yoon,[1] Tahmid Ahamed,[1] Samantha Bhuiyan,[1] Minxuan Cheng,[1] Julie Karam,[1] Kathleen M. Jagodnik,[1] Ingrid Shu,[1] Alexander Lachmann,[1] Sam Ayling,[2] Sherry L. Jenkins,[1] and Avi Ma'ayan[1,3,*]

[1]Department of Pharmacological Sciences, Mount Sinai Center for Bioinformatics, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, Box 1603, New York, NY 10029, USA

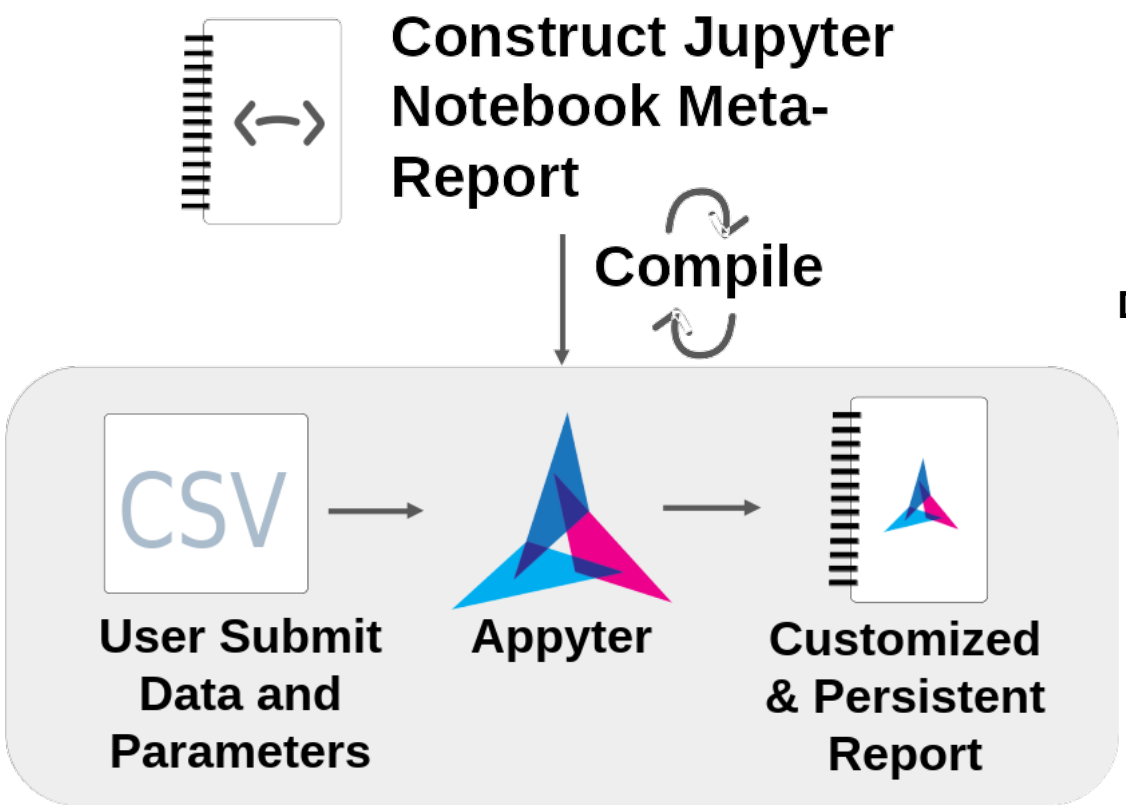
[2]Pencil Worx Design, 345 West 88th Street, New York, NY 10024, USA

[3]Lead contact

*Correspondence: avi.maayan@mssm.edu

https://doi.org/10.1016/j.patter.2021.100213

So What is an Appyter and How You Can Create One?

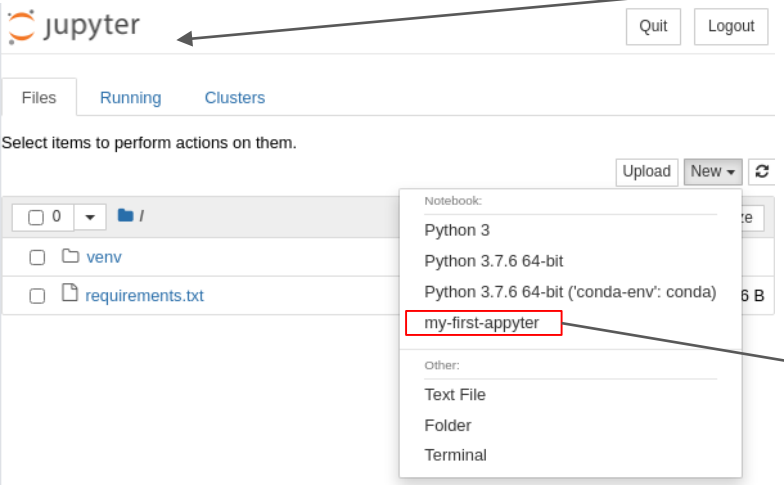


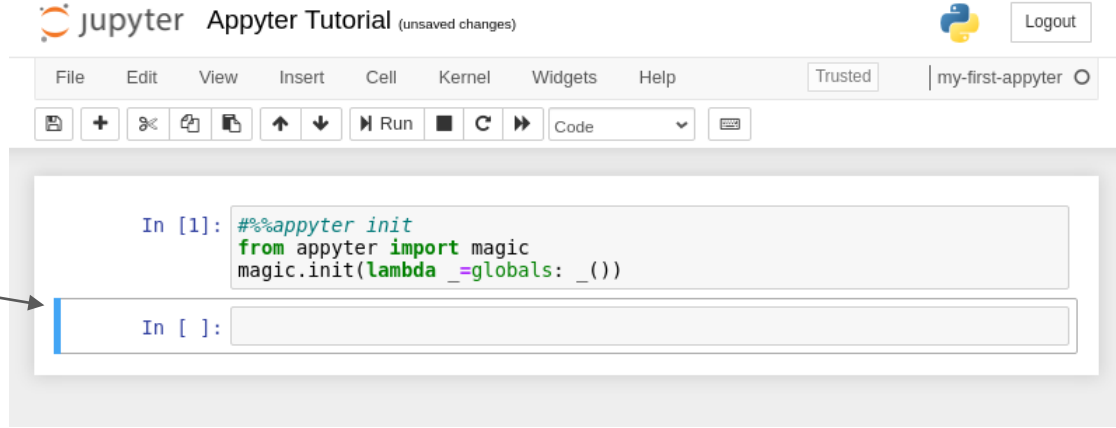
Daniel Clarke

Clarke et al., Appyters: turning Jupyter Notebooks into data-driven web apps, Patterns (2021), https://doi.org/10.1016/j.patter.2021.100213

Developing an Appyter: Start a jupyter-notebook

```
(venv) my-first-appyter % pip install appyter
...
(venv) my-first-appyter % jupyter notebook .
...
[I 16:46:10.066 NotebookApp] The Jupyter Notebook is running at:
[I 16:46:10.067 NotebookApp] http://localhost:8888/?token=797ac2fae005b3a1754cbeb43c9cc8277964a7ed9e3e1564
```

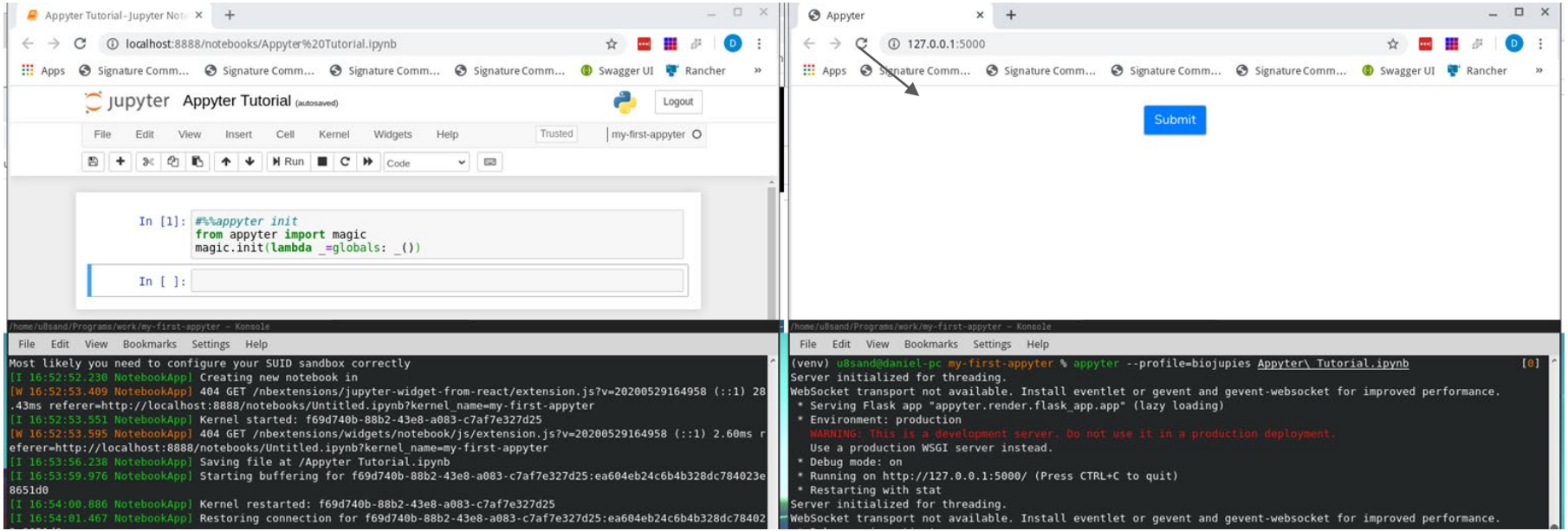




Clarke et al., Appyters: turning Jupyter Notebooks into data-driven web apps, Patterns (2021), https://doi.org/10.1016/j.patter.2021.100213

Developing an Appyter: Start the Appyter Server

```
(venv) my-first-appyter % appyter --profile=biojupies Appyter\ Tutorial.ipynb
...
* Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
```



Clarke et al., Appyters: turning Jupyter Notebooks into data-driven web apps, Patterns (2021), https://doi.org/10.1016/j.patter.2021.100213

Use Appyter Magic for jinja2 Meta-Programming

Minor changes can turn an existing jupyter notebook into a full-blown web application allowing others to use your data processing pipeline.

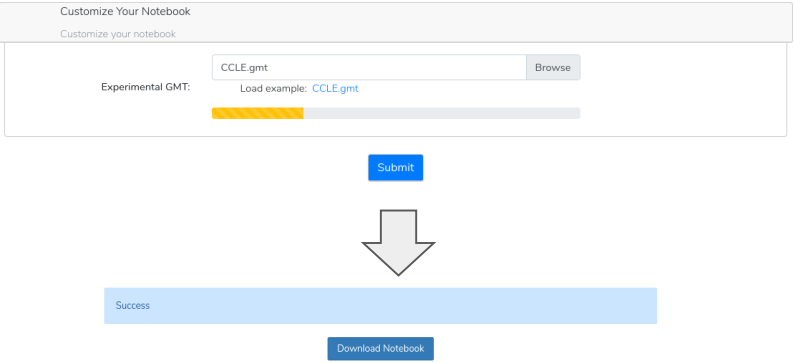
In [1]:
```
#%%appyter init
from appyter import magic
magic.init(lambda _=globals: _())
```

My Appyter

In [2]:
```
%%appyter code_eval
from maayanlab_bioinformatics.parse import gmt_read_pd
data = gmt_read_pd(open({{
    FileField(
        name='file',
        label='Experimental GMT',
        default='CCLE.gmt',
        examples={'CCLE.gmt': 'https://amp.pharm.mssm.edu/Enrichr/geneSetLibrary?mode=text&libraryName=Cancer_Cell_
    )
}}, 'r'))
data.describe()
```

```
from maayanlab_bioinformatics.parse import gmt_read_pd
data = gmt_read_pd(open('CCLE.gmt', 'r'))
data.describe()
```

	PECAPJ49 UPPER AERODIGESTIVE TRACT	VMCUB1 URINARY TRACT	NCIH596 LUNG	SNU410 PANCREAS	TE9 OESOPHAGUS	697 HAEMATOPOIETIC AND LYMPHOID TISSUE	HOS BONE	SCC25 UPPER AERODIGESTIVE TRACT	NCIH1734 LUNG	5637 URINARY TRACT	...	HS683 CENTRAL NERVOUS SYSTEM
count	46.0	119.0	55.0	31.0	125.0	199.000000	52.0	101.0	79.0	72.0	...	40.0
mean	1.0	1.0	1.0	1.0	1.0	1.190955	1.0	1.0	1.0	1.0	...	1.0



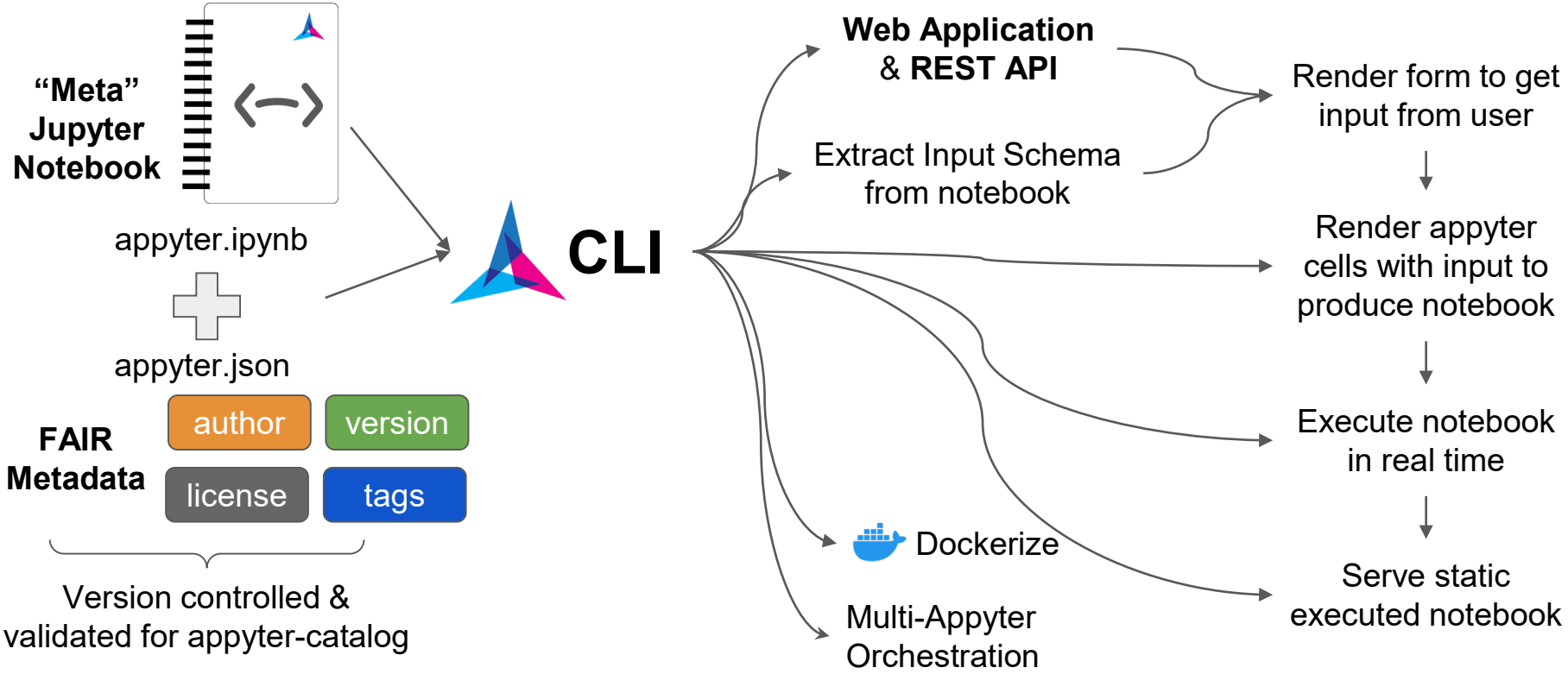
My Appyter

In [1]:
```
from maayanlab_bioinformatics.parse import gmt_read_pd
data = gmt_read_pd(open('CCLE.gmt', 'r'))
data.describe()
```

	PECAPJ49 UPPER AERODIGESTIVE TRACT	VMCUB1 URINARY TRACT	NCIH596 LUNG	SNU410 PANCREAS	TE9 OESOPHAGUS	697 HAEMATOPOIETIC AND LYMPHOID TISSUE	HOS BONE	SCC25 UPPER AERODIGESTIVE TRACT	NCIH1734 LUNG	5637 URINARY TRACT	...	HS683 CENTRAL NERVOUS SYSTEM	HS934 SKI
count	46.0	119.0	55.0	31.0	125.0	199.000000	52.0	101.0	79.0	72.0	...	40.0	125.00000
mean	1.0	1.0	1.0	1.0	1.0	1.190955	1.0	1.0	1.0	1.0	...	1.0	1.03200

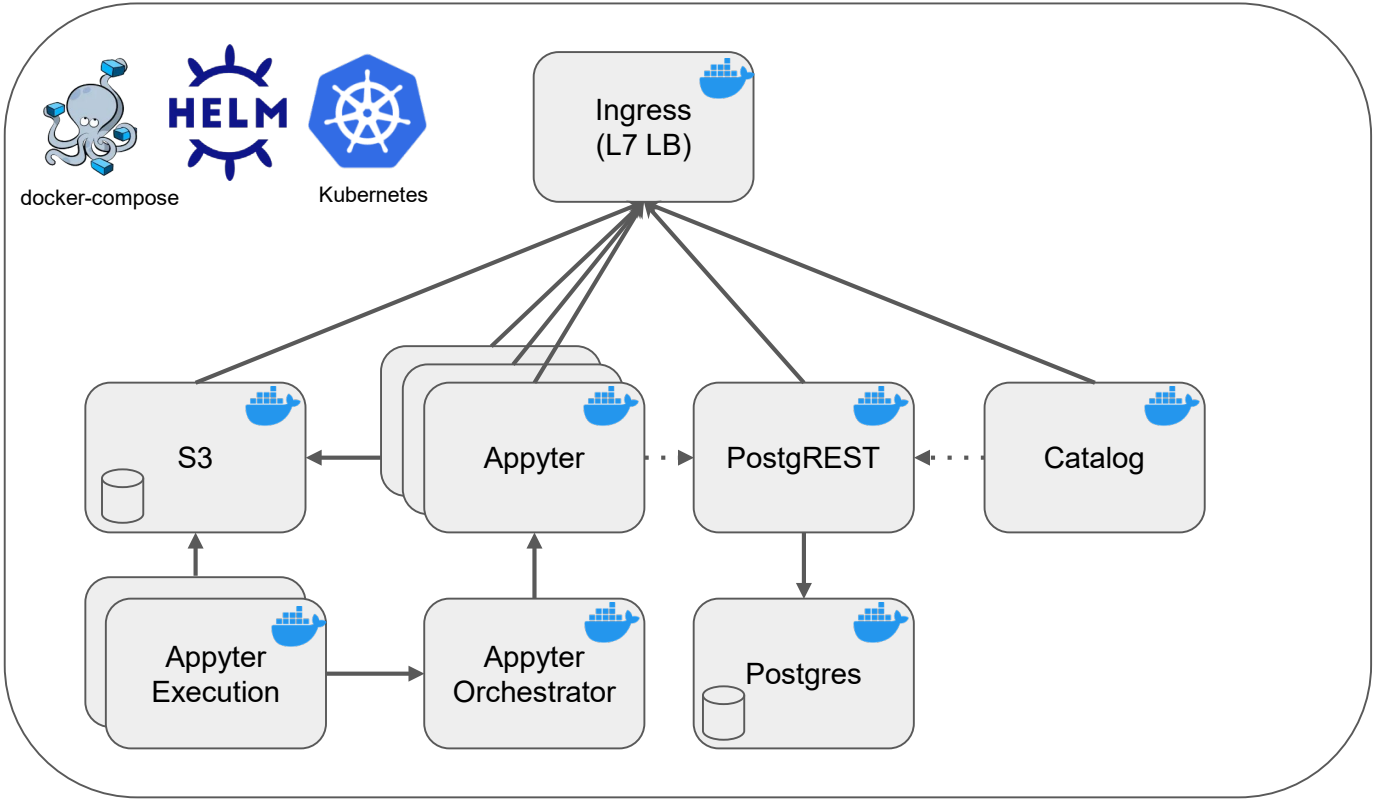
Clarke et al., Appyters: turning Jupyter Notebooks into data-driven web apps, Patterns (2021), https://doi.org/10.1016/j.patter.2021.100213

Single Source of Truth: Jupyter Notebook (ipynb)



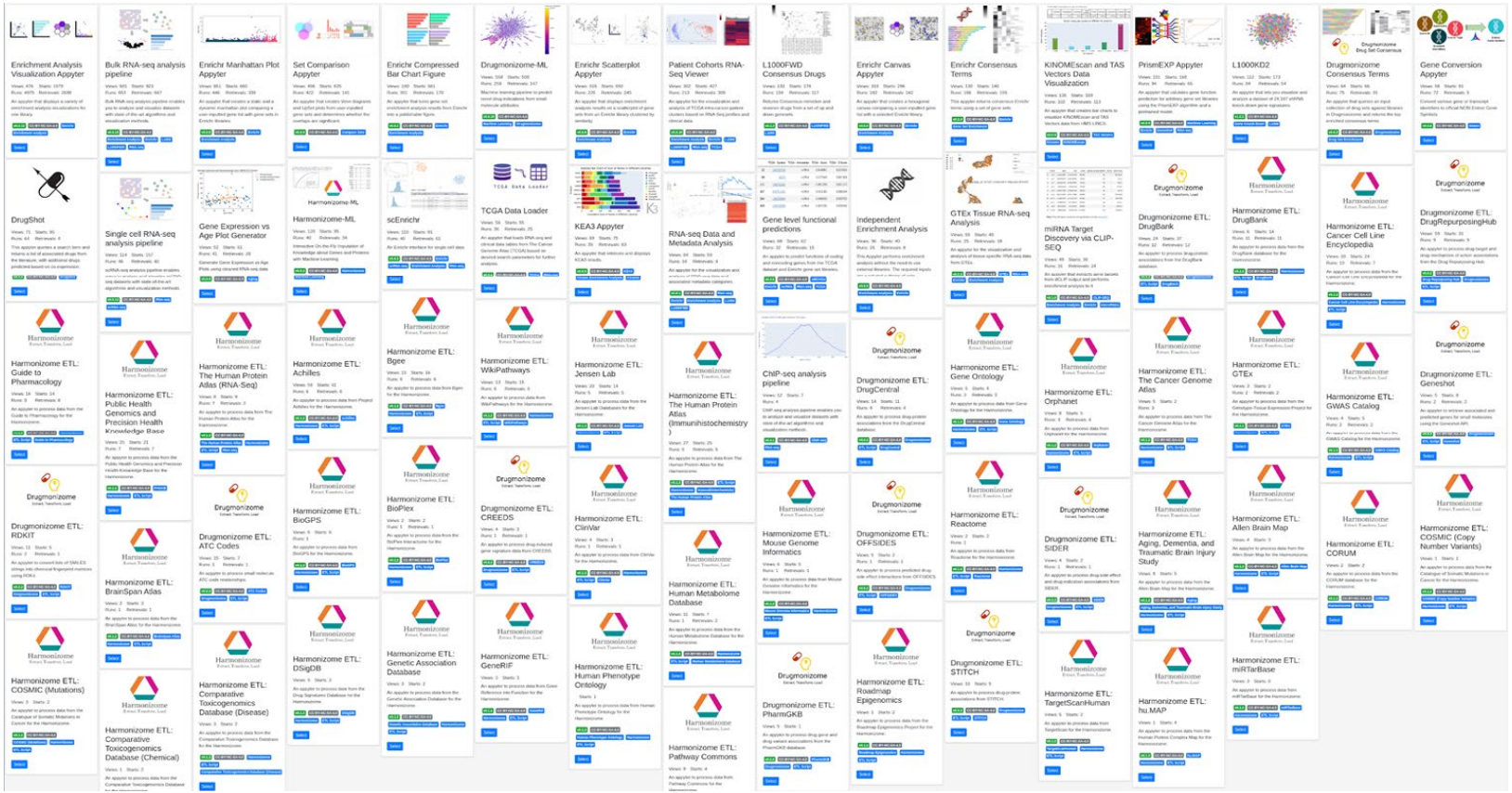
Clarke et al., Appyters: turning Jupyter Notebooks into data-driven web apps, Patterns (2021), https://doi.org/10.1016/j.patter.2021.100213

Platform Agnostic Multi-Appyter Orchestration



Clarke et al., Appyters: turning Jupyter Notebooks into data-driven web apps,
Patterns (2021), https://doi.org/10.1016/j.patter.2021.100213

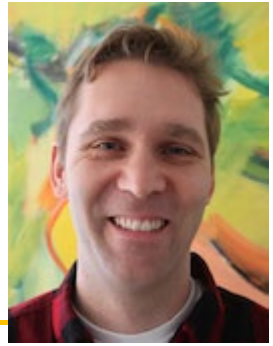
The Appyters Catalog Currently Contains 77 Appyters



Clarke et al., Appyters: turning Jupyter Notebooks into data-driven web apps, Patterns (2021), https://doi.org/10.1016/j.patter.2021.100213

https://appyters.maayanlab.cloud/

ARCHS^4 Resource



nature communications

nature › nature communications › articles › article

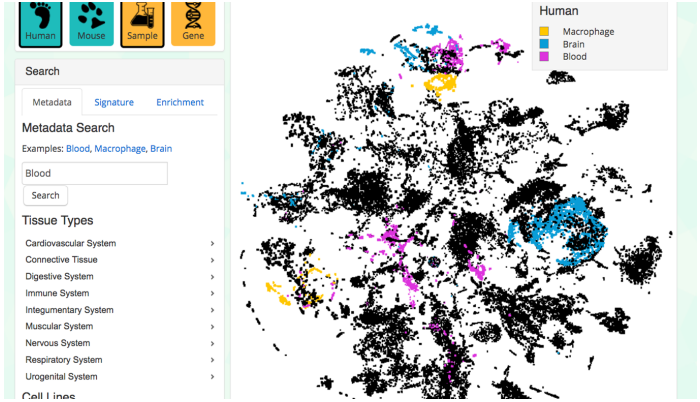
Article | Open Access | Published: 10 April 2018

Massive mining of publicly available RNA-seq data from human and mouse

Alexander Lachmann, Denis Torre, Alexandra B. Keenan, Kathleen M. Jagodnik, Hoyjin J. Lee, Lily Wang, Moshe C. Silverstein & Avi Ma'ayan

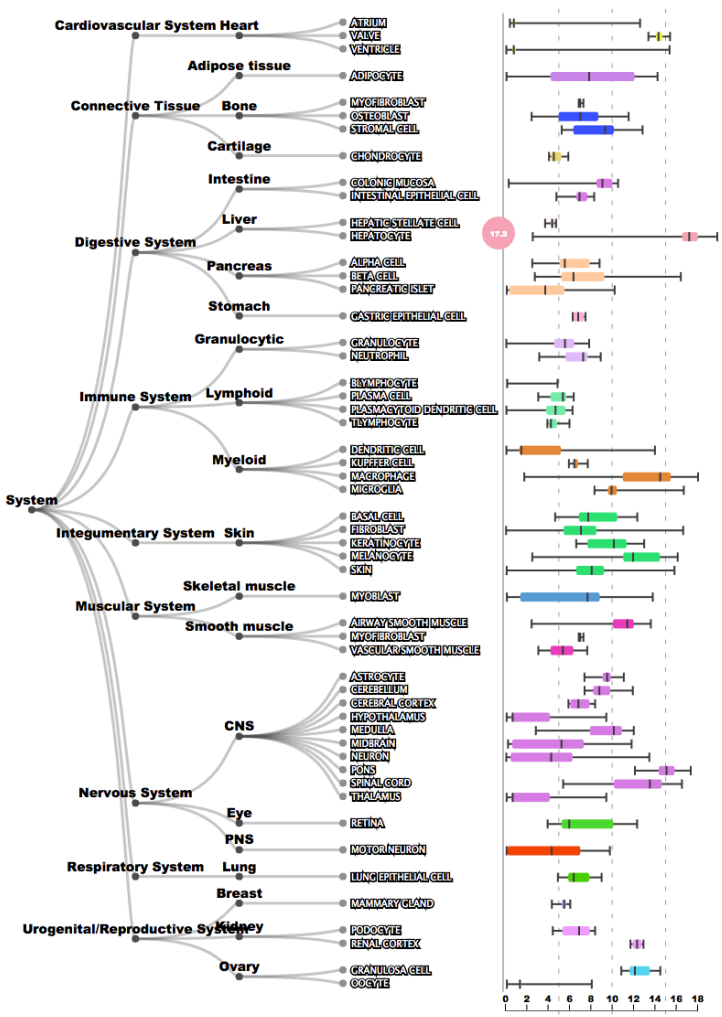
Nature Communications **9**, Article number: 1366 (2018) | Cite this article

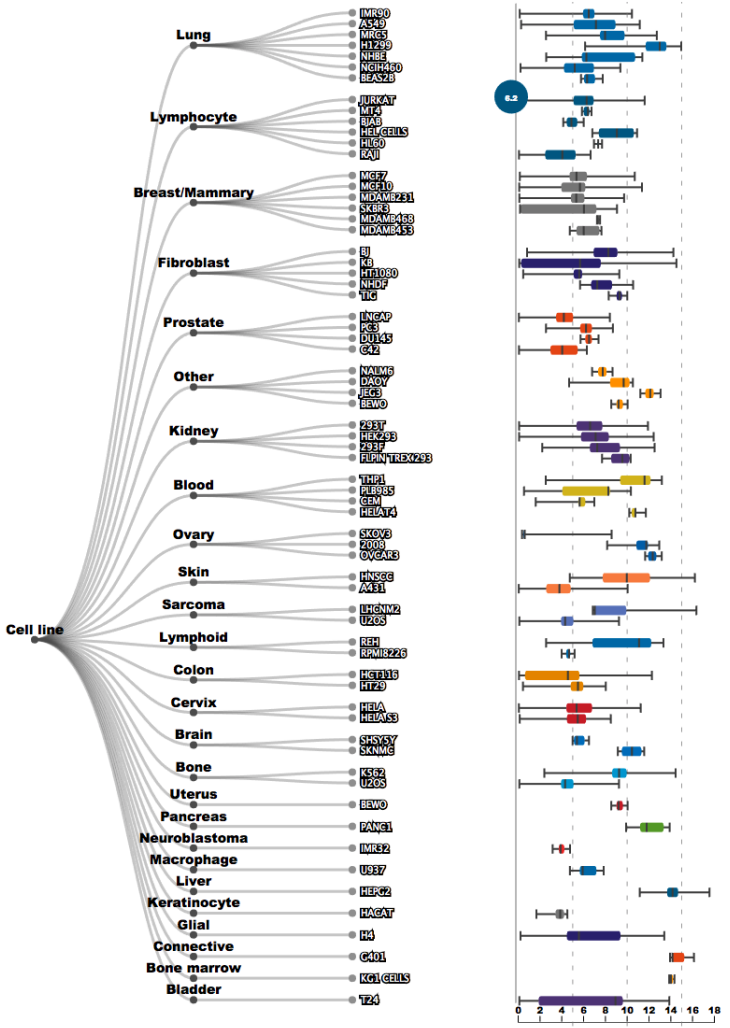
26k Accesses | **112** Citations | **69** Altmetric | Metrics



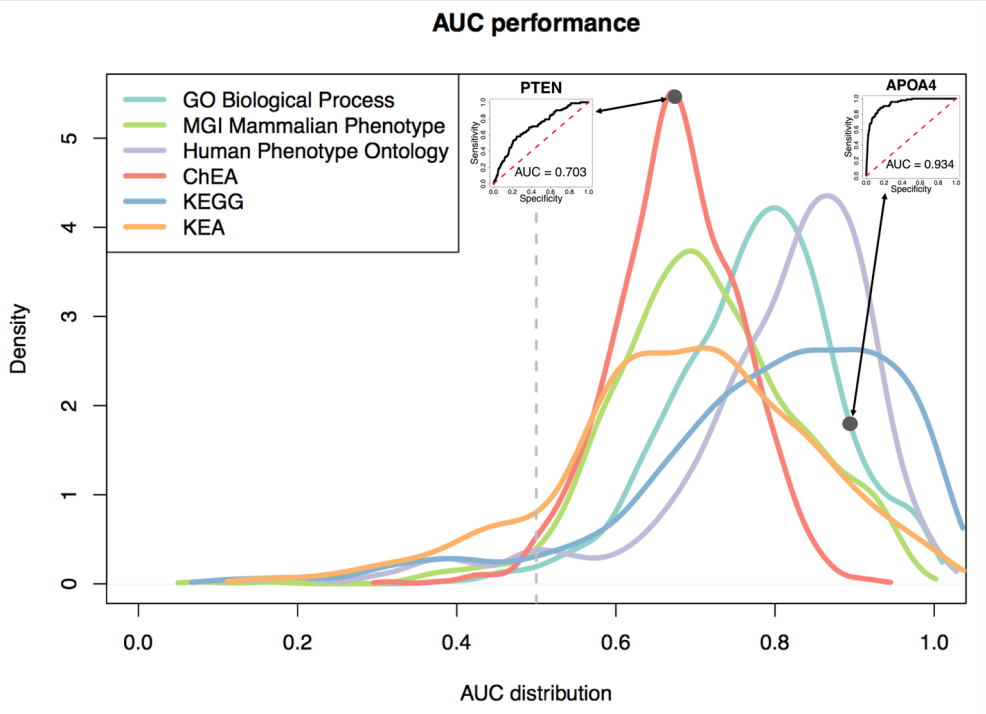
- Samples can be searched by:
  - Meta data, text annotation from SRA
  - Data driven sample query (highly expressed and low expressed genes)
  - Sample query through functional enrichment
  - Manual selection

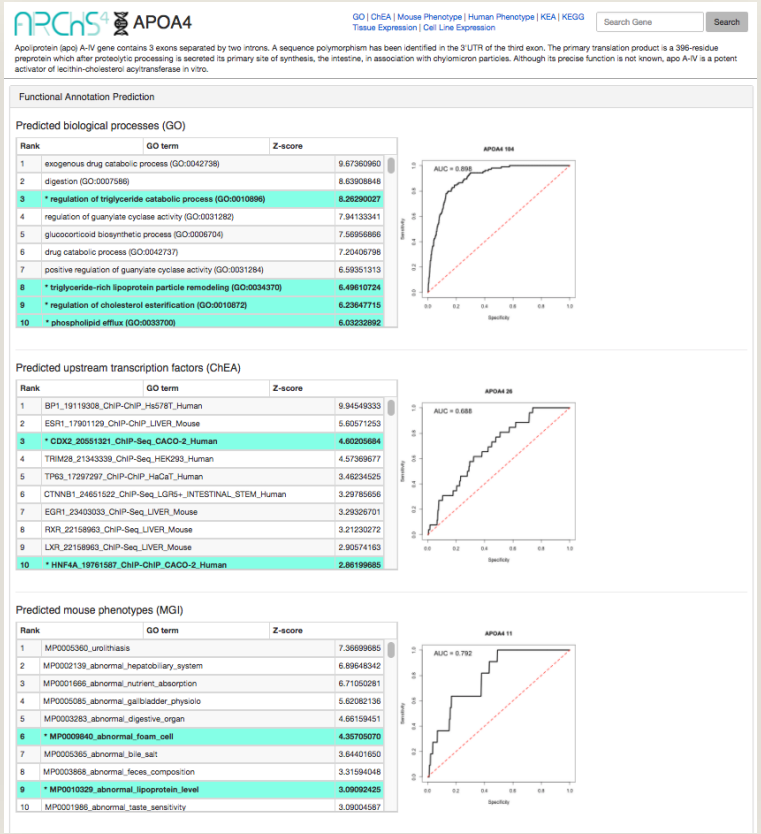
Gene Expression Atlas for Tissues and Cell Lines



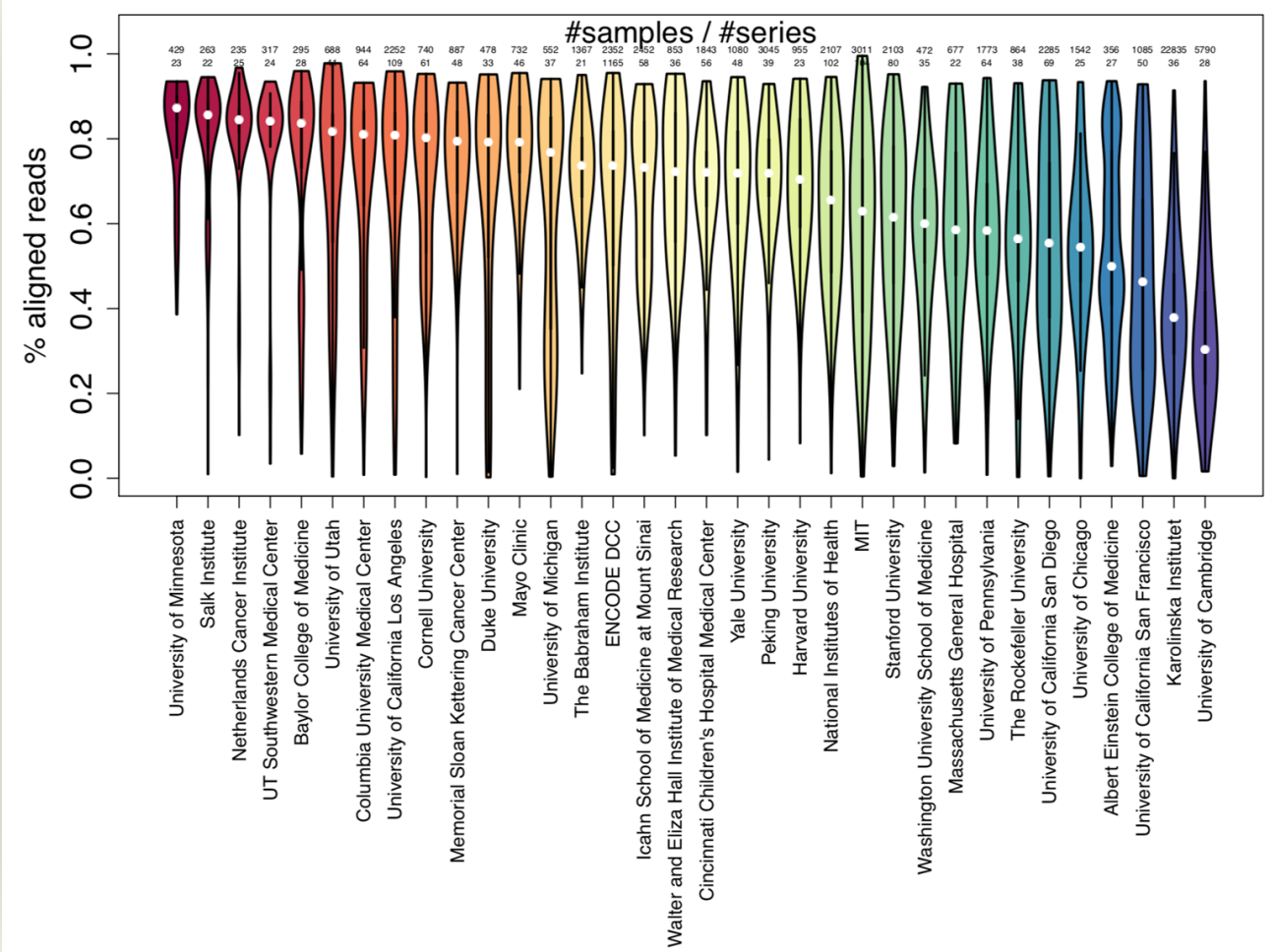


Prediction of Gene Function using ARCHS4 Data



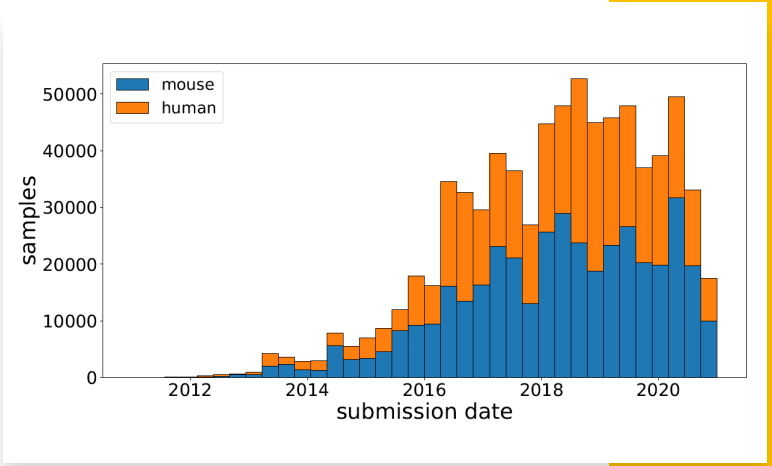


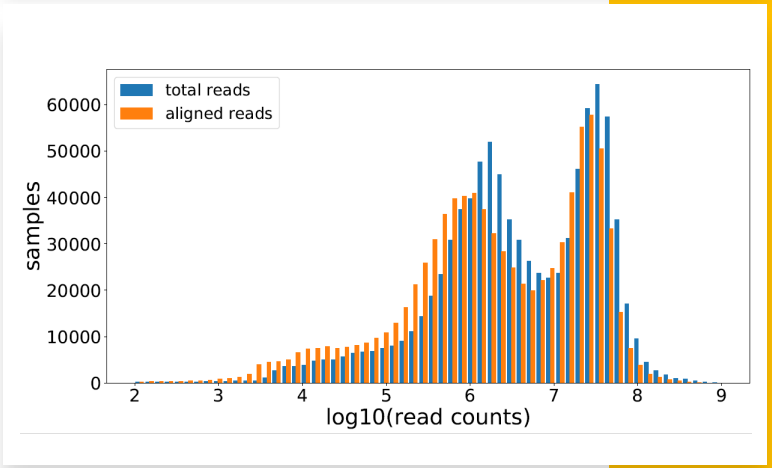
Prediction of Gene Function using ARCHS4 Data



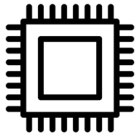
Mining Massive Gene Expression Repositories

- ARCHS4 was used to align 16 trillion reads (16*10^12) to human and mouse reference genomes
- > petabyte of data (1 million GB)
- 900,809 samples from 23,739 experiments

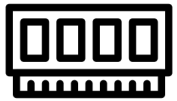




Cost challenges

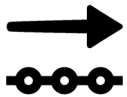


- CPU
  - Hardware purchases are high upfront and require maintenance
  - CPU rentals are high if used continuously

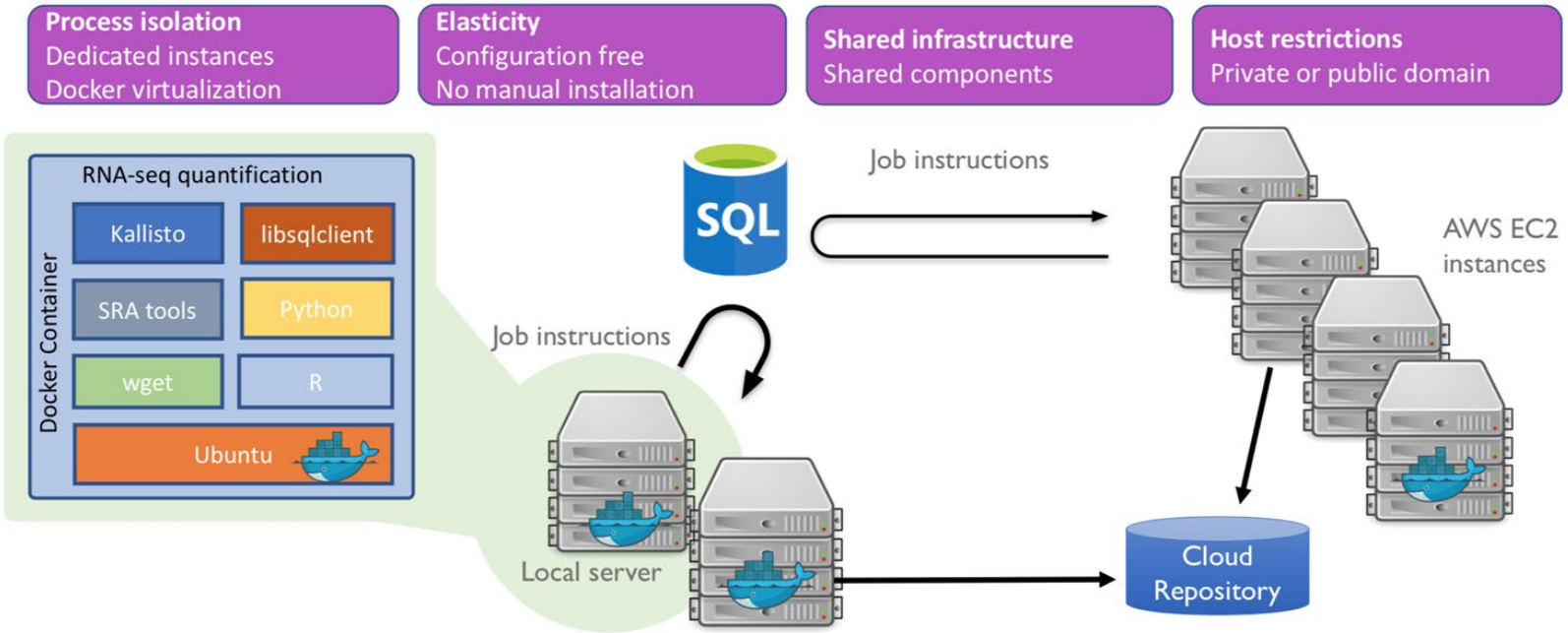


- Memory
  - Aligning reads to a reference genome requires efficient in memory index structures
  - Depending on alignment algorithm this can vary from 4GB (kallisto/De Bruijn Graph) to 16GB (STAR/suffix tree)

- Storage
  - Raw read files need to be stored for processing
  - Required storage varies significantly between samples

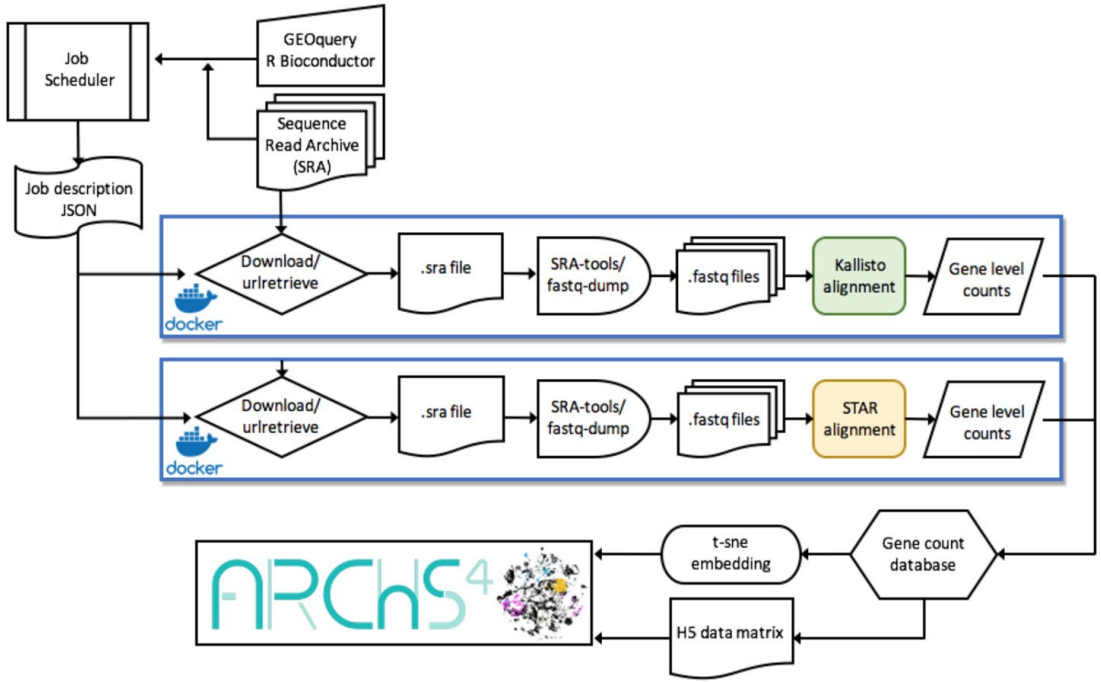


- Network bandwidth
  - Fast networking is required to retrieve raw read data
  - Slow bandwidth will increase overall processing time



Hybrid RNA-seq Pipeline Design

- Processing pipelines are dockerized for easy deployment to resource pool
- Control server hosting job commands serves job descriptions upon request of worker nodes
- Job descriptions are JSON objects
- Results are stored at an online location accessible via URL



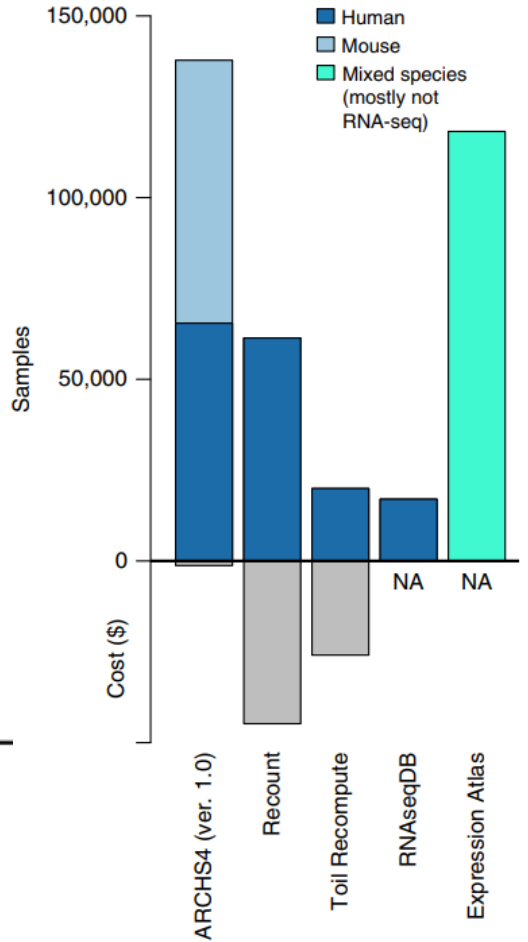
Hybrid RNA-seq Pipeline Design

- Parallelization through deployment of dockerized workflows
- Instances are precisely chosen for required resources
  - memory
  - bandwidth
  - storage

Cost comparison of RNA-seq resequencing projects

- The average cost of processing an RNA-seq sample is below $0.01
- Minimal resource allocation and low-cost high network bandwidth of cloud resources is a key advantage of the ARCHS4 pipeline
- Choice of efficient alignment algorithm (kallisto) results in low memory footprint

RNA-seq resource	ARCHS4	Recount	Toil Recompute
Human samples	84,863	61,350	19,931
Mouse samples	103,083	0	0
Total samples	187,946	61,350	19,931
Cost per sample	< $0.01	$0.73	$1.30



Mount Sinai Center for Bioinformatics

Ma'ayan Laboratory of Computational Systems Biology

2021 Summer Research Training Program in Biomedical Big Data Science

Icahn School of Medicine at Mount Sinai — Center for Bioinformatics

Research intensive 10-week training program for undergraduate and master's students

Program Dates: June 7 - August 13, 2021

Students selected for summer session 2021 of our **research training program in the Ma'ayan Laboratory at the Icahn School of Medicine at Mount Sinai** will conduct faculty-mentored independent research projects in the following areas:

- Data Harmonization
- Machine Learning
- Cloud Computing
- Dynamic Data Visualization

Application Deadline:
February 1, 2021 at 5 PM Eastern Time

Who Should Apply:
Students majoring in Computer Science, Informatics, Mathematics, Statistics, Physics, Engineering, Chemistry/Chemical Sciences or Biological Sciences and have an interest in Biomedical Big Data Science.

Contact:
Sherry Jenkins, MS
Program Manager
E-mail: sherry.jenkins@mssm.edu

Faculty Mentor and Principal Investigator:
Avi Ma'ayan PhD, Professor and Director
Mount Sinai Center Bioinformatics
Icahn School of Medicine at Mount Sinai
New York, New York

Trainee Salary:
$8,000 salary for the 10-week training period

Eligibility:
To be considered for this program, applicants must be:

- U.S. citizen or U.S. permanent resident
- Undergraduate or master's student in good academic standing

APPLICATION DETAILS: http://labs.icahn.mssm.edu/maayanlab/summer-research-program/

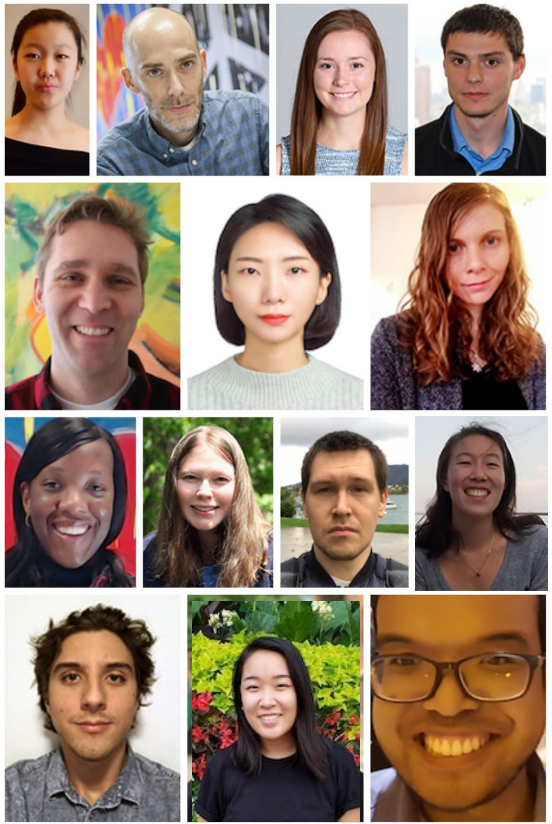
Summary

- We systematically convert publicly available omics datasets into an abstract format centered on genes and drugs.
- The resources we developed have made a big impact on the research community.
- Appyters can enable the rapid development of bioinformatics applications not just for the Mount Sinai Center for Bioinformatics but also for others at Mount Sinai.
- Machine Learning to impute knowledge about gene and drug functions.
- SignatureCommons is a template system to quickly bring up a data portal to serve data and metadata.

Acknowledgements

Ma'ayan Lab

Sherry Jenkins, MS - Project Manager
Daniel Clarke, MS - Data Science Analyst
Alexander Lachmann, PhD - Assistant Professor
Kathleen Jagodnik, PhD - Postdoctoral Fellow
Minji Jeon, PhD - Postdoctoral Fellow
Megan Wojciechowicz, MS - PhD Student
John Erol Evangelista, MS - Bioinformatician
Sherry Xie, BS - Bioinformatician
Eryk Kropiwnicki, MS - Bioinformatician
Maxim Kuleshov, MS - Bioinformatician
Ingrid Shu, BS - Bioinformatician
Allison Bailey, MPH - Associate Researcher



BD2K-LINCS
DATA COORDINATION AND INTEGRATION CENTER

COMMON FUND DATA ECOSYSTEM

IDG

Ma'ayan Lab

Icahn School of Medicine at Mount Sinai

Center for Bioinformatics

LYME MIND