

# **GEN3VA: aggregation and analysis of gene expression signatures from related studies**

**Avi Ma'ayan, PhD**

**Professor, Department of Pharmacological Sciences**

**Director, Mount Sinai Center for Bioinformatics**

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

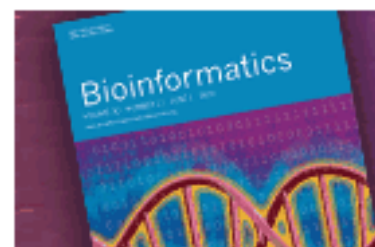


**BD2K-LINCS  
DATA COORDINATION AND  
INTEGRATION CENTER**

# Bioinformatics

[ABOUT THIS JOURNAL](#) [CONTACT THIS JOURNAL](#) [SUBSCRIPTIONS](#)[CURRENT ISSUE](#) [ARCHIVE](#) [SEARCH](#)

[Oxford Journals](#) > [Science & Mathematics](#) > [Bioinformatics](#) > [Volume 31, Issue 18](#) > Pp. 3060-3062.



Sign up for e-alerts from *Bioinformatics*  
to receive new content as soon as it publishes online.



## GEO2Enrichr: browser extension and server app to extract gene sets from GEO and analyze them for biological functions



[« Previous | Next Article »](#)  
[Table of Contents](#)

Search this journal:



[Advanced »](#)

### Current Issue

[November 1, 2016 32 \(21\)](#)



[Gregory W. Gundersen<sup>1,2,3</sup>](#), [Matthew R. Jones<sup>1,2,3</sup>](#), [Andrew D. Rouillard<sup>1,2,3</sup>](#), [Yan Kou<sup>1,2,3</sup>](#),  
[Caroline D. Monteiro<sup>4</sup>](#), [Axel S. Feldmann<sup>1,2,3</sup>](#), [Kevin S. Hu<sup>1,2,3</sup>](#) and [Avi Ma'ayan<sup>1,2,3,\\*</sup>](#)

[+ Author Affiliations](#)

[✉](#) \*To whom correspondence should be addressed.

### This Article

Bioinformatics (2015) 31 (18):  
3060-3062.  
doi:  
10.1093/bioinformatics/btv297  
First published online: May 13,  
2015

[Abstract](#) **Free**

[» Full Text \(HTML\)](#) **Free**

[Full Text \(PDF\)](#) **Free**

Received January 12, 2015.

Revision received May 1, 2015.

Accepted May 4, 2015.



# GEO2Enrichr

offered by [maayanlab.net/g2e](http://maayanlab.net/g2e)

★★★★★ (12)

[Productivity](#)

870 users

ADDED TO CHROME



OVERVIEW

REVIEWS

RELATED

G+1

4

DS5077[ACCM] Search Clear Show All Advanced Search

DataSet Record DS5077: Expression Profiles Data Analysis Tools Sample Subsets

Transmembrane protein 88 depletion effect on stem cell cardiac differentiation in vitro

Analysis of RUES2 stem cells first identified for transmembrane protein TMEM88, and then induced to differentiate into cardiac cells. RUES2 cells examined at day 5 of differentiation. Results provide insight in

*Homo sapiens*

GPL10558: Illumina Hum

Palpant NJ, Rabon L, Rab

Sep;140(18):3799-808.

ries: GSE43805

transformed count



Please verify that your data is correct.

Accession num. DS5077

Title Transmembrane protein 88 depletion effect on stem cell cardiac differentiation in vitro

Organism Homo sapiens

Control samples GSM1071455, GSM1071454

Treatment or condition samples GSM1071457, GSM1071456

Please select differential expression analysis options.

Differential expression method Characteristic direction

Cutoff 200

Transform and normalize if necessary\* No

Please fill out these optional annotations.

Cell type or tissue RUES2 stem cells

Perturbation Transmembrane protein 88 depletion

Manipulated gene TMEM88

Relevant disease ...

Please apply metadata tags.

Metadata Tags rues2 stemcells

Extract gene lists Open results tab

\*See the [website](#) for details.

GEO2Enrichr is being developed by the [Ma'ayan Lab](#).



Compatible with your device

## Ma'ayan Lab

GEO2Enrichr helps you quickly and easily extract gene sets from the Gene Expression Omnibus and analyze these lists for common biological functions.

[Report Abuse](#)

## Additional Information

Version: 4.8

Updated: October 31, 2016

Size: 153KiB

Language: English



# **DATASET** CURATED **BROWSER**

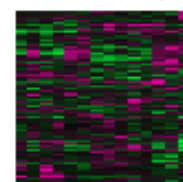


Search for

## DataSet Record GDS4892:

<b>Title:</b>	Age effect on skeletal muscle precursor cells		
<b>Summary:</b>	Analysis of FACS-sorted CD45-Ter119-Sca-1-CD29+Cxcr4+ skeletal muscle precursor (SMP) cells from young, middle-aged, and old C57BL/6 males. Results provide insight into the molecular mechanisms underlying age-related skeletal muscle and stem cell dysfunction.		
<b>Organism:</b>	<i>Mus musculus</i>		
<b>Platform:</b>	GPL1261: [Mouse430_2] Affymetrix Mouse Genome 430 2.0 Array		
<b>Citation:</b>	Sinha M, Jang YC, Oh J, Khong D et al. Restoring systemic GDF11 levels reverses age-related dysfunction in mouse skeletal muscle. <i>Science</i> 2014 May 9;344(6184):649-52. PMID: 24797481		
<b>Reference Series:</b>	GSE50821	<b>Sample count:</b>	14
<b>Value type:</b>	count	<b>Series published:</b>	2014/05/08

### Cluster Analysis




### Download

- 
- 
- 
- 
- 

 **Cite GEO Dataset**

## Data Analysis Tools

Find genes 

Compare 2 sets of samples

Cluster heatmaps

Experiment design and value distribution

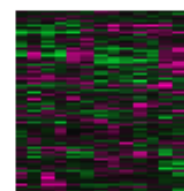
Find gene name or symbol:

Find genes that are up/down for this condition(s): ☒ age



<b>Title:</b>	Age effect on skeletal muscle precursor cells		
<b>Summary:</b>	Analysis of FACS-sorted CD45-Ter119-Sca-1-CD29+Cxcr4+ skeletal muscle precursor (SMP) cells from young, middle-aged, and old C57BL/6 males. Results provide insight into the molecular mechanisms underlying age-related skeletal muscle and stem cell dysfunction.		
<b>Organism:</b>	<i>Mus musculus</i>		
<b>Platform:</b>	GPL1261: [Mouse430_2] Affymetrix Mouse Genome 430 2.0 Array		
<b>Citation:</b>	Sinha M, Jang YC, Oh J, Khong D et al. Restoring systemic GDF11 levels reverses age-related dysfunction in mouse skeletal muscle. <i>Science</i> 2014 May 9;344(6184):649-52. PMID: 24797481		
<b>Reference Series:</b>	GSE50821	<b>Sample count:</b>	14
<b>Value type:</b>	count	<b>Series published:</b>	2014/05/08

#### Cluster Analysis



#### Download

- DataSet full SOFT file
- DataSet SOFT file
- Series family SOFT file
- Series family MINiML file
- Annotation SOFT file

Cite GEO Dataset [RIS \(.ris\)](#) [BibTeX \(.bib\)](#) [EndNote \(.enw\)](#)

### Data Analysis Tools

Find genes

Compare 2 sets of samples [?](#)

Cluster heatmaps

Experiment design and value distribution

**Step 1:** Select test and significance level

Two-tailed t-test (A vs B) ▼ Significance level: 0.100 ▼

**Step 2:** Select which Samples to put in Group A and Group B

**Step 3:** Query Group A vs. B

**Step 4:** Extract knowledge with GEO2Enrichr



Click on accessions to select samples individually, click on colored blocks and then on blinking arrows to select groups of samples.

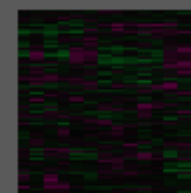
Samples, Group A	Factors	Samples, Group B
	<u>age</u>	
GSM1230351	young (2 months)	GSM1230351
GSM1230352		GSM1230352
GSM1230353		GSM1230353
GSM1230354		GSM1230354
GSM1230355		GSM1230355
GSM1230356	middle aged (12 months)	GSM1230356
GSM1230357		GSM1230357
GSM1230358		GSM1230358
GSM1230359		GSM1230359
GSM1230360	aged (24 months)	GSM1230360
GSM1230361		GSM1230361
GSM1230362		GSM1230362
GSM1230363		GSM1230363
GSM1230364		GSM1230364

Ok

Reset

Cancel

## Cluster Analysis



## Download

DataSet full SOFT file

DataSet SOFT file

Series family SOFT file

Series family MINiML file

Annotation SOFT file

30 2.0 Array

emic GDF11 levels reverses age-related dysfunction in  
) :649-52. PMID: 24797481

ount: 14

blished: 2014/05/08

## Data Analysis Tools

ct test and significance level

test (A vs B) ▼ Significance level: 0.100 ▼

Step 2: Select which Samples to put in Group A and Group B

Step 3: Query Group A vs. B

Step 4: Extract knowledge with GEO2Enrichr



middle-aged, and old C57BL/6 males. Results provide insight into the molecular mechanisms underlying age-related skeletal muscle and stem cell dysfunction.



ad

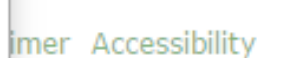
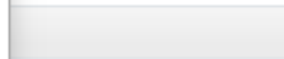
OFT file

file

OFT file

INiML file

FT file



timer Accessibility

Organism:

Platform:

Citation:

Reference Series

Value type:

 Cite GEO Dataset

Find genes

Compare 2 sets of

Cluster heatmaps

Experiment design



## GEO2Enrichr



*Please verify that your data is correct.*

Accession num.	GDS4892
Title	Age effect on skeletal muscle precursor cells
Organism	Mus musculus
Control samples	GSM1230351, GSM1230352, GSM1230353, GSM1230354, GSM1230355
Treatment or condition samples	GSM1230360, GSM1230361, GSM1230362, GSM1230363, GSM1230364

*Please select differential expression analysis options.*

Differential expression method	Characteristic direction
Cutoff	500
Transform and normalize if necessary*	No

*Please fill out these optional annotations.*

Cell type or tissue	skeletal muscle
Perturbation	age
Manipulated gene	...
Relevant disease	aging

*Please apply metadata tags.*

Metadata Tags AGING\_BD2K\_AHM\_DEMO

Extract gene lists

Check for duplicate signatures

\*See the [website](#) for details.

GEO2Enrichr is being developed by the [Ma'ayan Lab](#).

## Results

*This is a permanent results page for your analyzed data. Please save the link in your browser's address bar.*

### GEN3VA Reports for these tags

AGING\_BD2K\_AHM\_DEMO

#### SOFT file

Title / Description	Age effect on skeletal muscle precursor cells
Accession	<a href="#">GDS4892</a>
Summary	Analysis of FACS-sorted CD45-Ter119-Sca-1-CD29+Cxcr4+ skeletal muscle precursor (SMP) cells from young, middle-aged, and old C57BL/6 males. Results provide insight into the molecular mechanisms underlying age-related skeletal muscle and stem cell dysfunction.
Organism	Mus musculus
Platform	GPL1261
Normalized	False
Parsed SOFT file	<div></div>

#### Metadata

Differential expression method	<a href="#">Characteristic Direction</a>
Cutoff	500
Cell	skeletal muscle



# AGING\_BD2K\_LINCS\_DCIC\_COURSERA report

Categorized using metadata field **Cell\_type**

[All reports](#)

## Gene signatures

Signatures (244) and Custom Report Builder

Show  entries

Search:

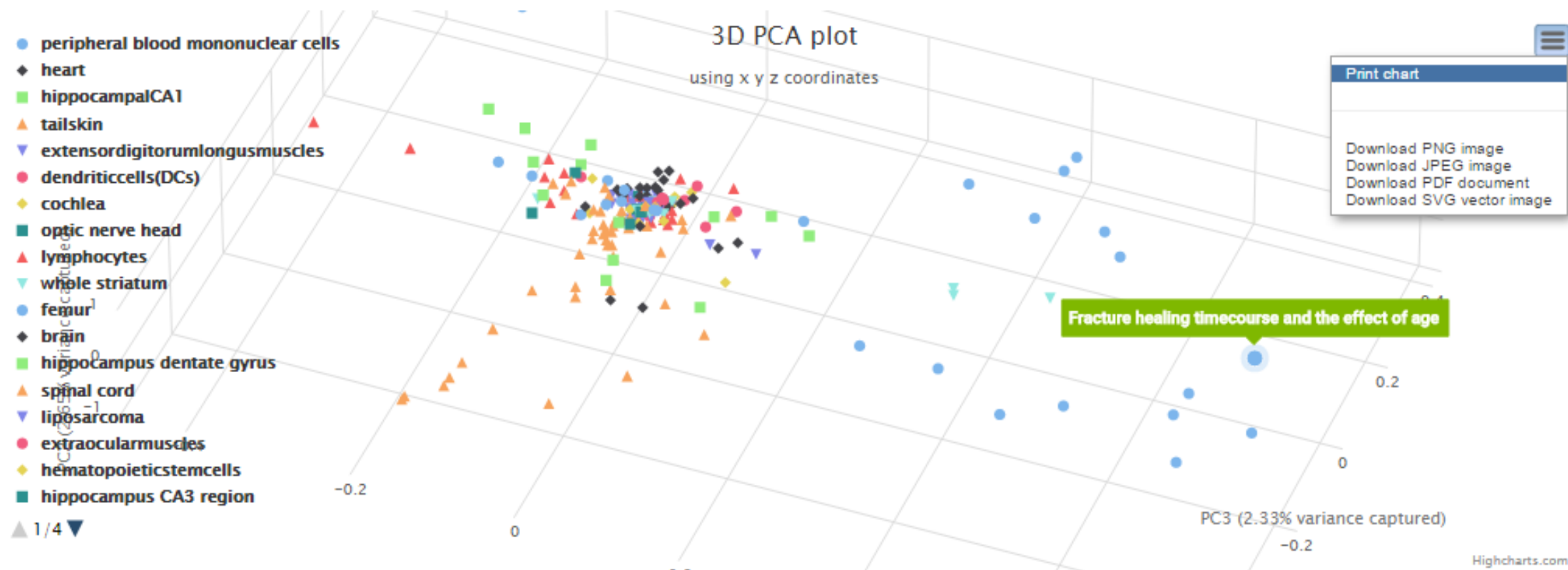
	↓↑ Title	↓↑ Organism	↓↑ Platform	↓↑
1	<a href="#">Age effect on lipopolysaccharide-induced neuroinflammation and sickness behavior</a>	Mus musculus	Mus musculus	
2	<a href="#">Age effect on lipopolysaccharide-induced neuroinflammation and sickness behavior</a>	Mus musculus	Mus musculus	
3	<a href="#">Age effect on extraocular muscles</a>	Rattus norvegicus	Rattus norvegicus	
4	<a href="#">Age effect on extraocular muscles</a>	Rattus norvegicus	Rattus norvegicus	
5	<a href="#">Age effect on extraocular muscles</a>	Rattus norvegicus	Rattus norvegicus	

Showing 1 to 5 of 244 entries

[Previous](#)
[1](#)
[2](#)
[3](#)
[4](#)
[5](#)
[...](#)
[49](#)
[Next](#)

# PCA

Interactive 3D principal component analysis of gene signatures. You can rotate the visualization and mouse over the data points.



# Genes

Hierarchical clustering of genes based on weights from the differential expression method.

Show color legend



Row Order

Alphabetically

Cluster

Rank by Sum

Rank by Variance

Column Order

Alphabetically

Cluster

Rank by Sum

Rank by Variance

Row

Search

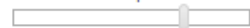
Opacity Slider



Row Group Size



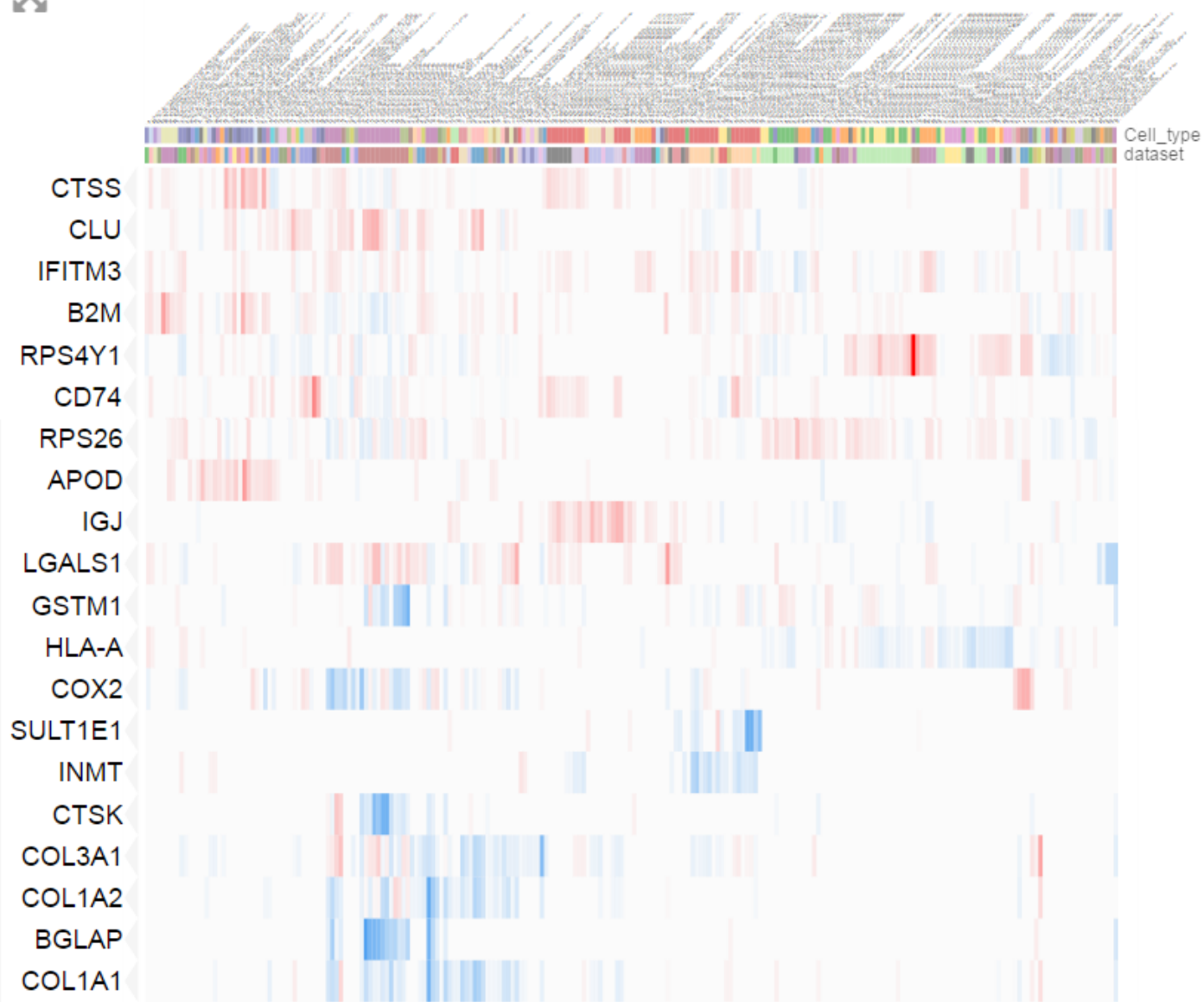
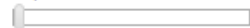
Column Group Size



Top rows sum: 20



Top rows variance: all



# Enrichr

Enrichr is a web tool that performs gene set enrichment analysis. To generate the hierarchical clusterings below, GEN3VA enriched each gene signature using Enrichr and then clustered the terms. Use the **select** button to change Enrichr's background library.

Hierarchical clustering of enriched terms from ENCODE\_TF\_ChIP-seq\_2015

Show color legend



Row Order

- Alphabetically
- Cluster
- Rank by Sum
- Rank by Variance

Column Order

- Alphabetically
- Cluster
- Rank by Sum
- Rank by Variance

Row

Search

Opacity Slider



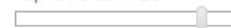
Row Group Size



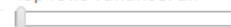
Column Group Size



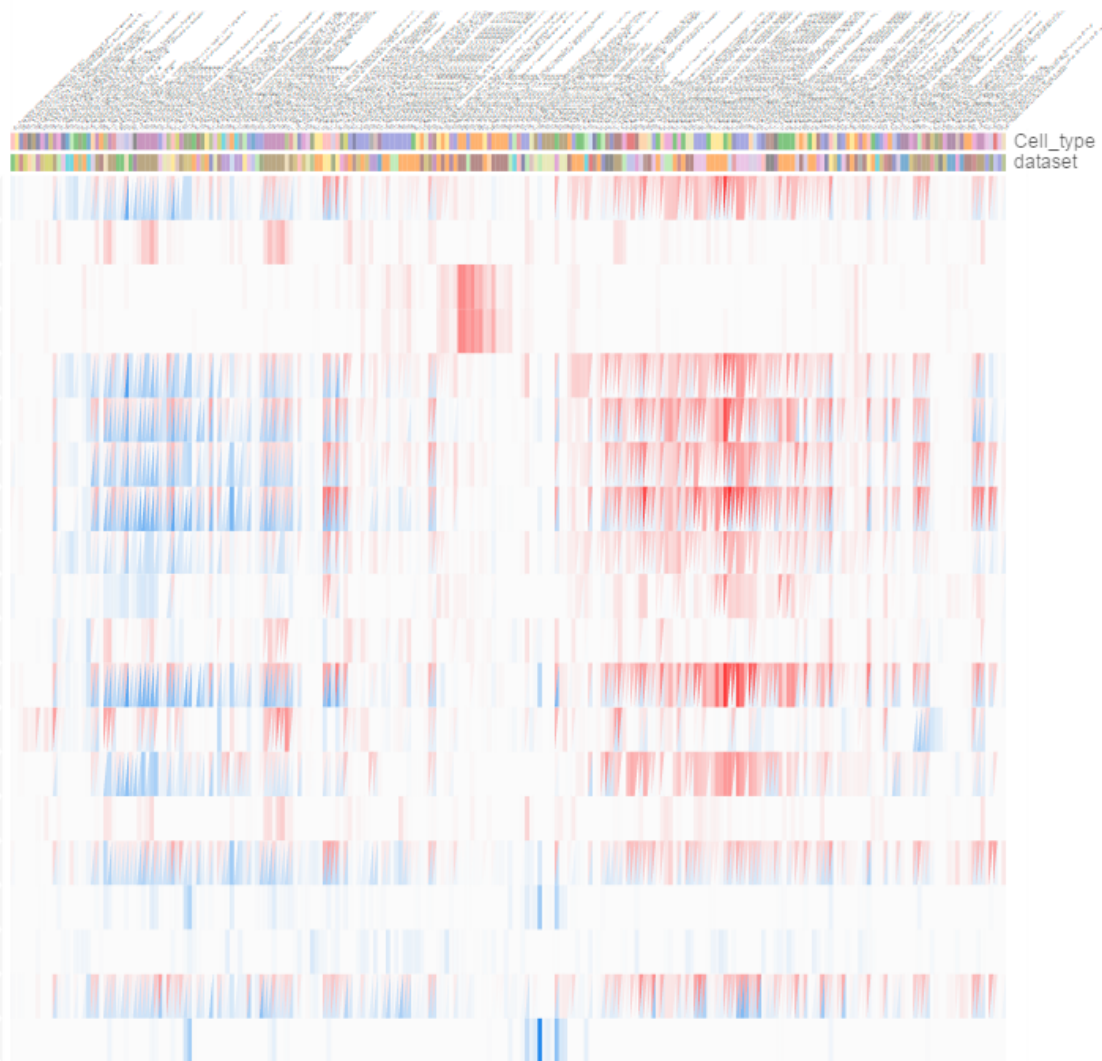
Top rows sum: 20



Top rows variance: all



RELA GM1..  
FOSL2 SK-..  
STAT1 K56..  
STAT2 K56..  
RELA GM1..  
NELFE K56..  
RELA GM1..  
TAF1 MCF-..  
RELA GM1..  
IRF4 GM12..  
TEAD4 SK-..  
KAT2A HeL..  
TCF12 myo..  
REST GM1..  
JUN HeLa-..  
POLR2A H..  
E2F4 GM12..  
UBTF MEL..  
POLR2A liv..  
E2F4 HeLa-..



# L1000CDS2

L1000CDS2 is a web tool that queries the LINCS L1000 dataset to identify small molecules that can reverse or mimic the observed input expression pattern. To generate the hierarchical clustering below, GEN3VA enriched each gene signature using L1000CDS2 to generate a list of perturbations that either mimic (red) and reverse (blue) expression.

Show color legend



Row Order

- Alphabetically
- Cluster
- Rank by Sum
- Rank by Variance

Column Order

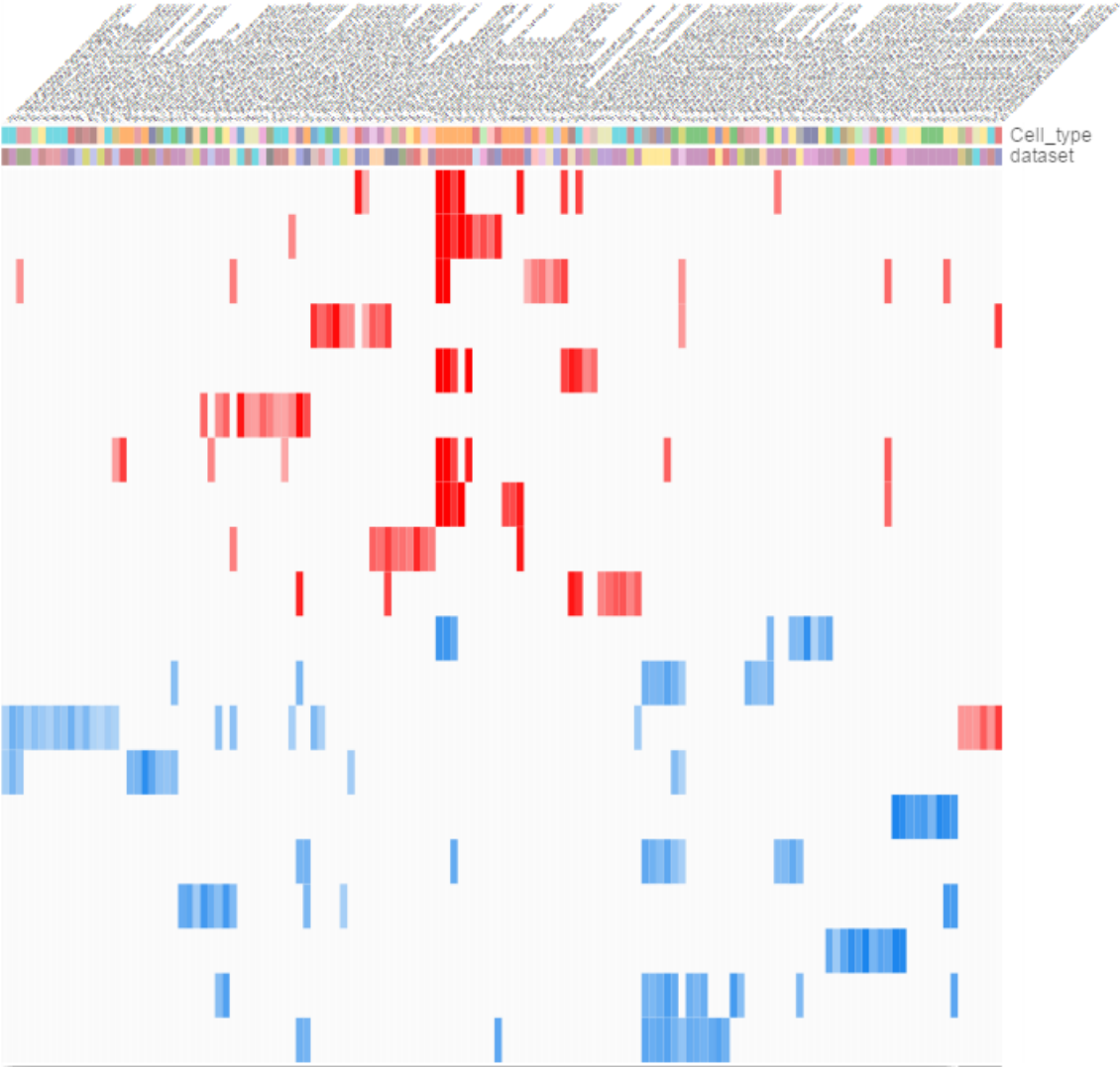
- Alphabetically
- Cluster
- Rank by Sum
- Rank by Variance

Row  
Search

- Opacity Slider
- Row Group Size
- Column Group Size
- Top rows sum: 20
- Top rows variance: all

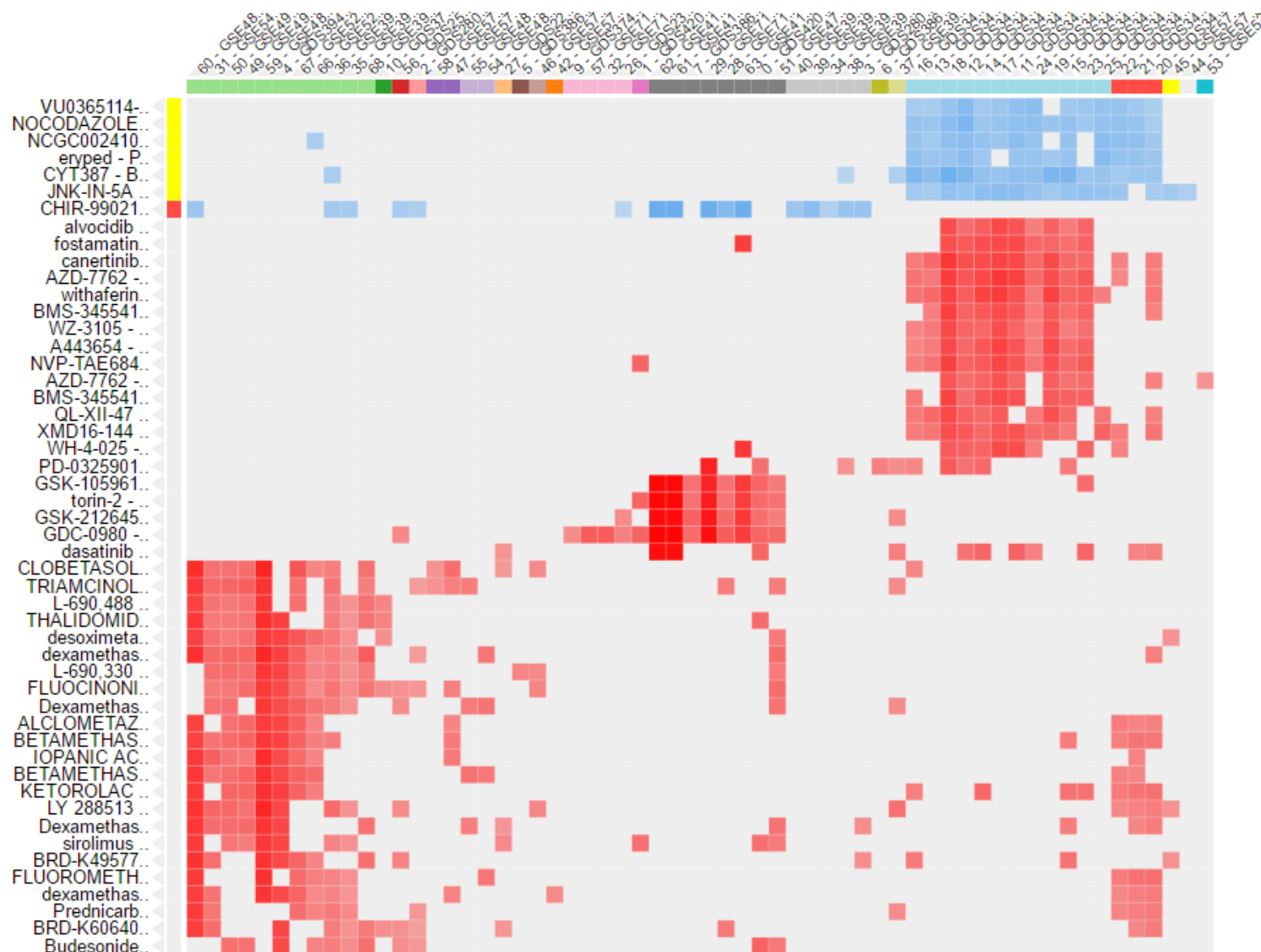


NVP-TAE68..  
AMSACRIN..  
wortmannin..  
torin-2 - MC..  
WZ-4-145 - ..  
mitoxantron..  
saracatinib ..  
blebbistatin ..  
dasatinib - ..  
BMS-34554..  
BMS-34554..  
BRD-K1087..  
radicicol - B..  
geldanamyc..  
LDN-19318..  
4-(aminome..  
celastrol - A..  
CGP-60474..  
6-diazo-5-o..  
6-diazo-5-o..





# Search for Dexamethasone Mimickers

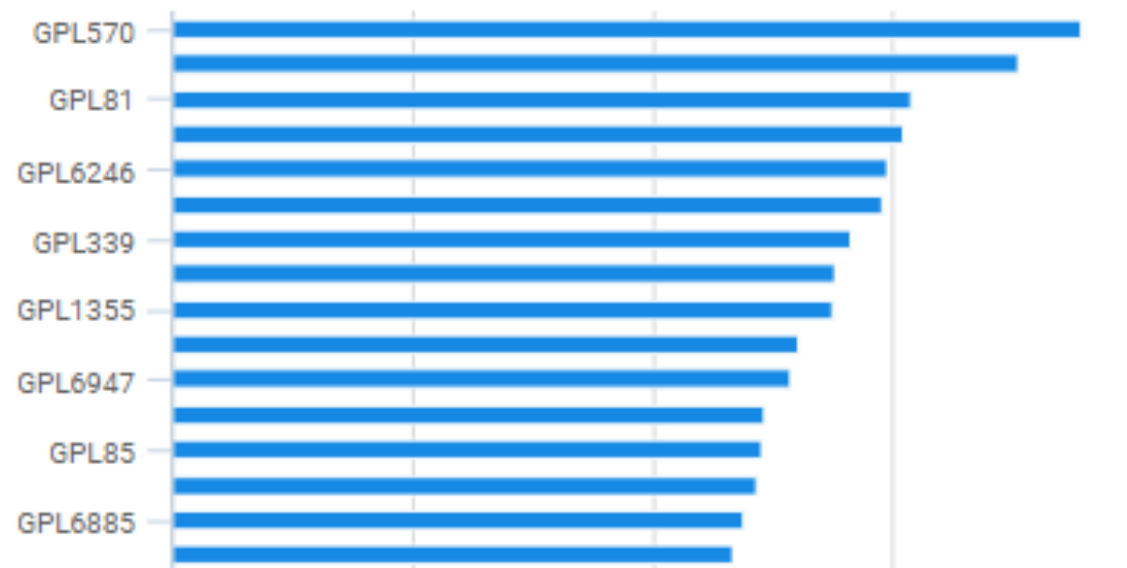


<http://amp.pharm.mssm.edu/gen3va/report/approved/Dexamethasone>

# Statistics

## Gene signatures by platform

Log10 scale (1 rendered as 0.175)



## Counts

Gene signatures	22066
Gene lists	65356
Tags	301
Reports	190
Platforms	70

# BMC Bioinformatics

[HOME](#)[ABOUT](#)[ARTICLES](#)[SUBMISSION GUIDELINES](#)[SOFTWARE](#)[OPEN ACCESS](#)

## GEN3VA: aggregation and analysis of gene expression signatures from related studies

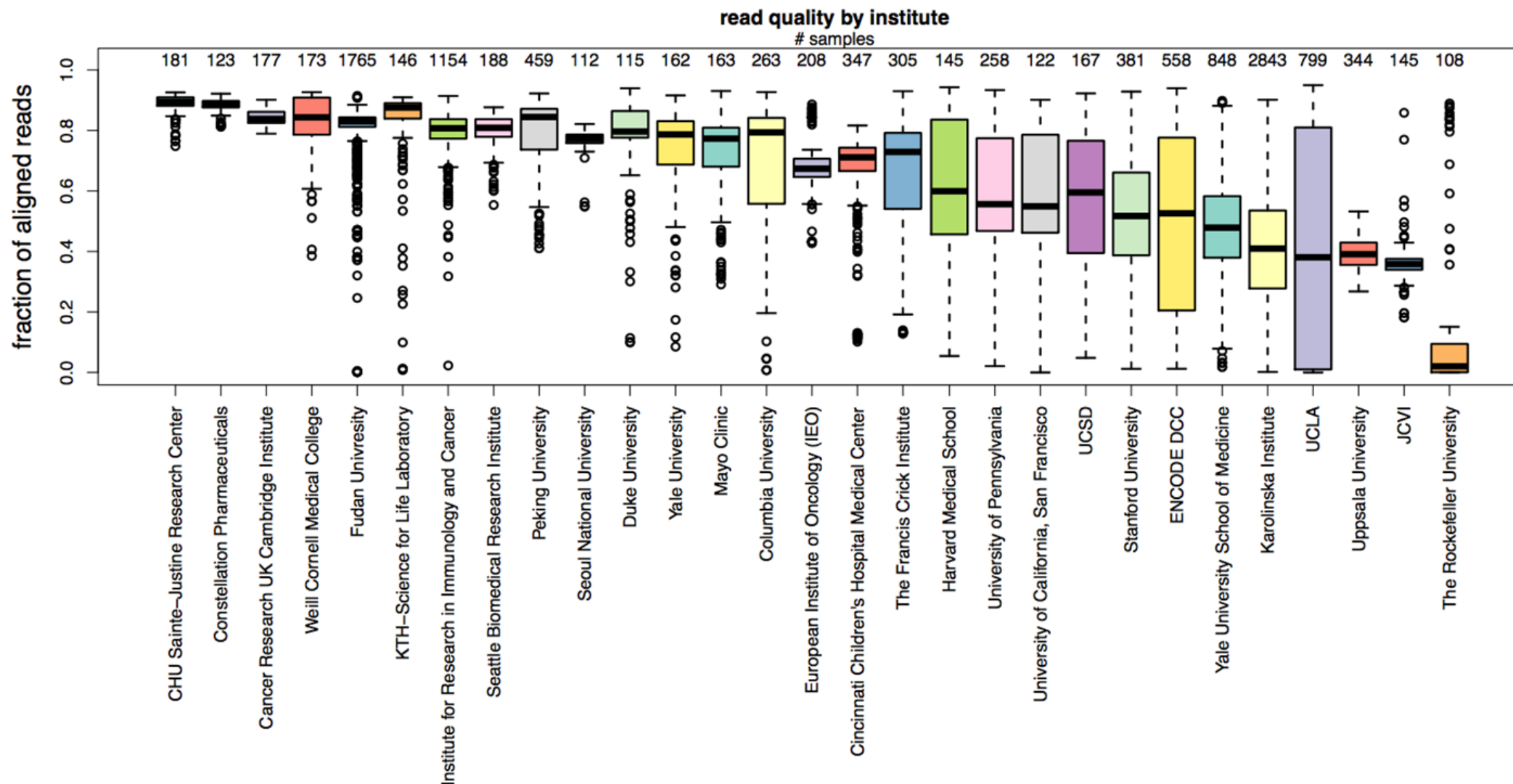
Gregory W. Gundersen, Kathleen M. Jagodnik, Holly Woodland, Nicholas F. Fernandez, Kevin Sani, Anders B. Dohlman, Peter Man-Un Ung, Caroline D. Monteiro, Avner Schlessinger and Avi Ma'ayan 

*BMC Bioinformatics* BMC series – open, inclusive and trusted 2016 17:461 |

DOI: 10.1186/s12859-016-1321-1 | © The Author(s). 2016

Received: 15 July 2016 | Accepted: 4 November 2016 | Published: 15 November 2016

# Fraction of Align Reads per Institution



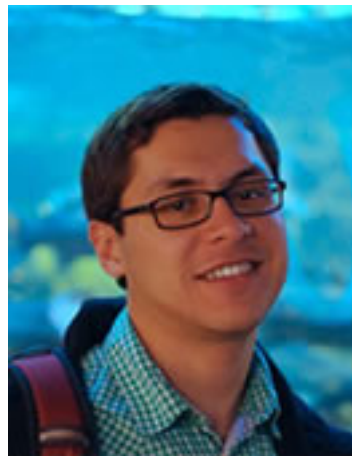
Using the Commons Cloud Credits Model Alex Lachmann Aligned All  
~180,000 Mammalian RNA-seq Samples from GEO and SRA

**GEO2Enrichr  
GEN3VA**



Gregory Gundersen

**Clustergrammer**



Nicolas Fernandez

**Crowdsourcing**



Zichen Wang

**Enrichr API**



**RNA-seq Upgrade**



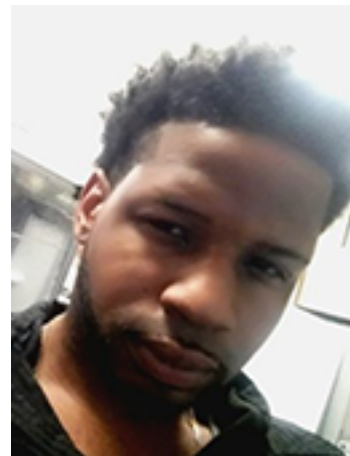
Denis Torre

**RNA-seq Upgrade**



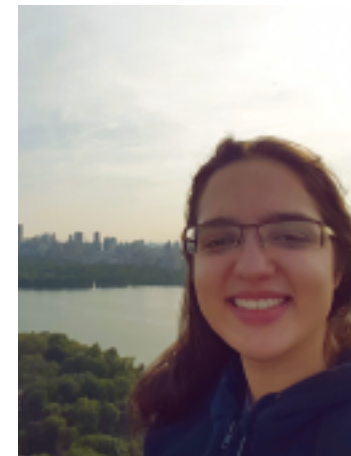
Alexander Lachmann

**Signature Extraction**



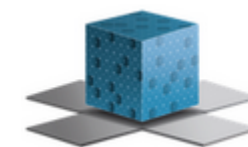
Troy Goff

**Signature Extraction**

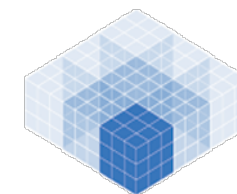


Caroline Monteiro





BD2K-LINCS  
DATA COORDINATION AND  
INTEGRATION CENTER



NIH LINCS  
PROGRAM



The NIH Common Fund



NIH Big Data to  
Knowledge (BD2K)