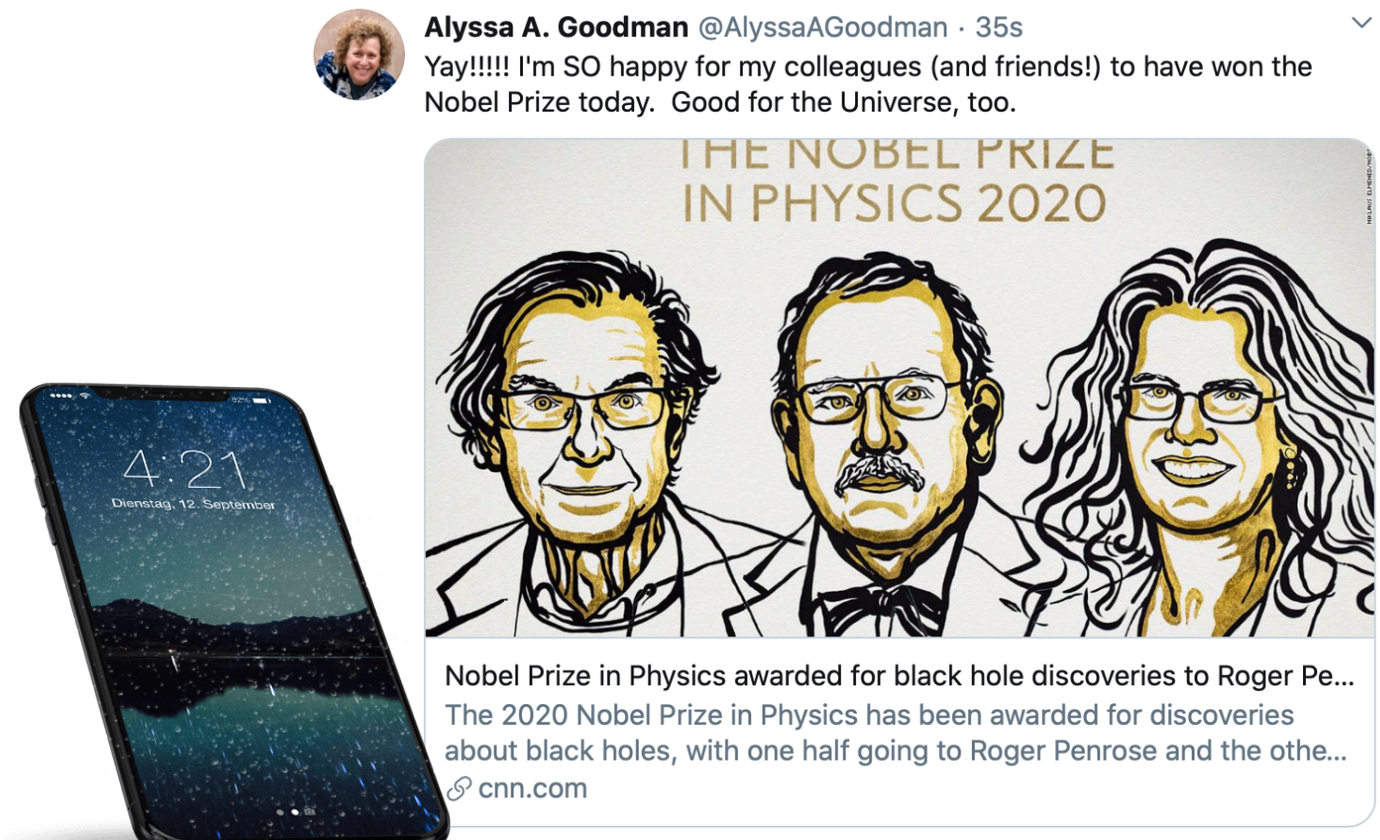
If you would like to participate in live polls today… open the camera app on your phone, scan the QR code, or type PollEV.com/prediction, to get ready.



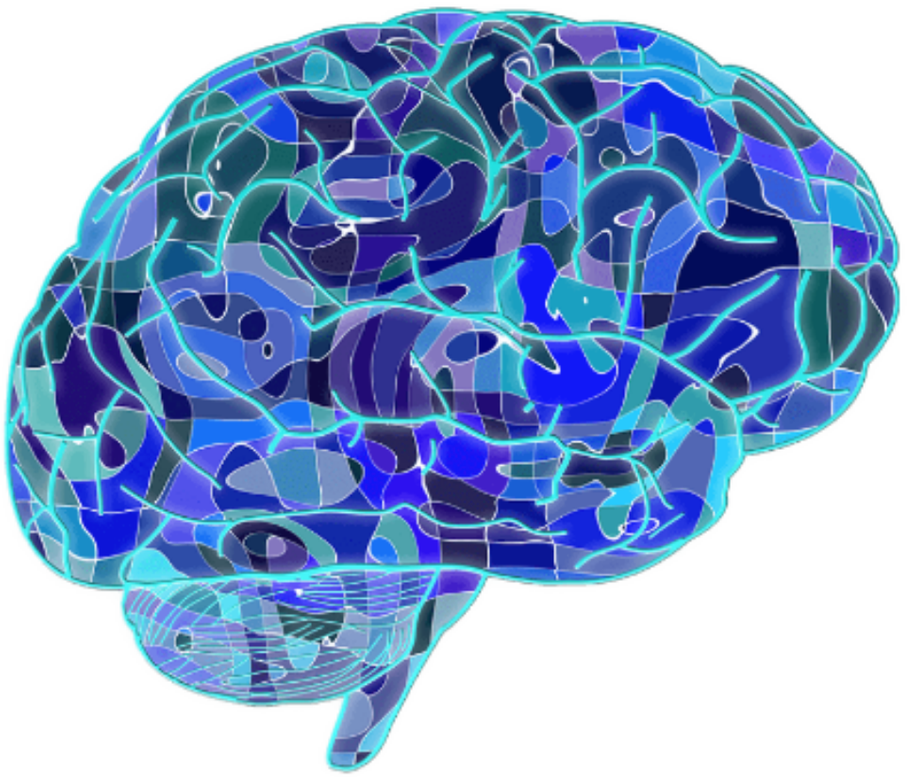
Meanwhile, a thank you to glue's sponsors…

Your Expertise (Select all that apply)

- Computer Science
- Biology
- Statistics
- Physics
- Data Visualization
- Software Development
- Instrumentation
- None of the above

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Your interests (Select all that apply)



- Computer Science
- Biology
- Statistics
- Physics
- Data Visualization
- Software Development
- Instrumentation
- None of the above

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Discovery through Data Diversity: Exploratory Data Visualization with glue

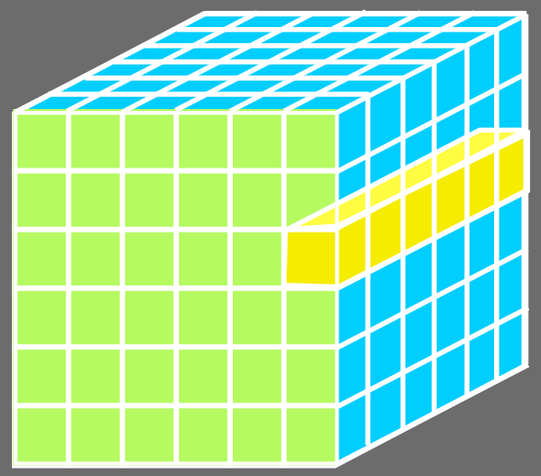
Alyssa A. Goodman, Harvard-Smithsonian Center for Astrophysics,
Radcliffe Institute for Advanced Study, and glue solutions, inc.

PRACTICALITY

PRINCIPLES

PHILOSOPHY

glue
multidimensional data exploration



DIVERSE TOOLS

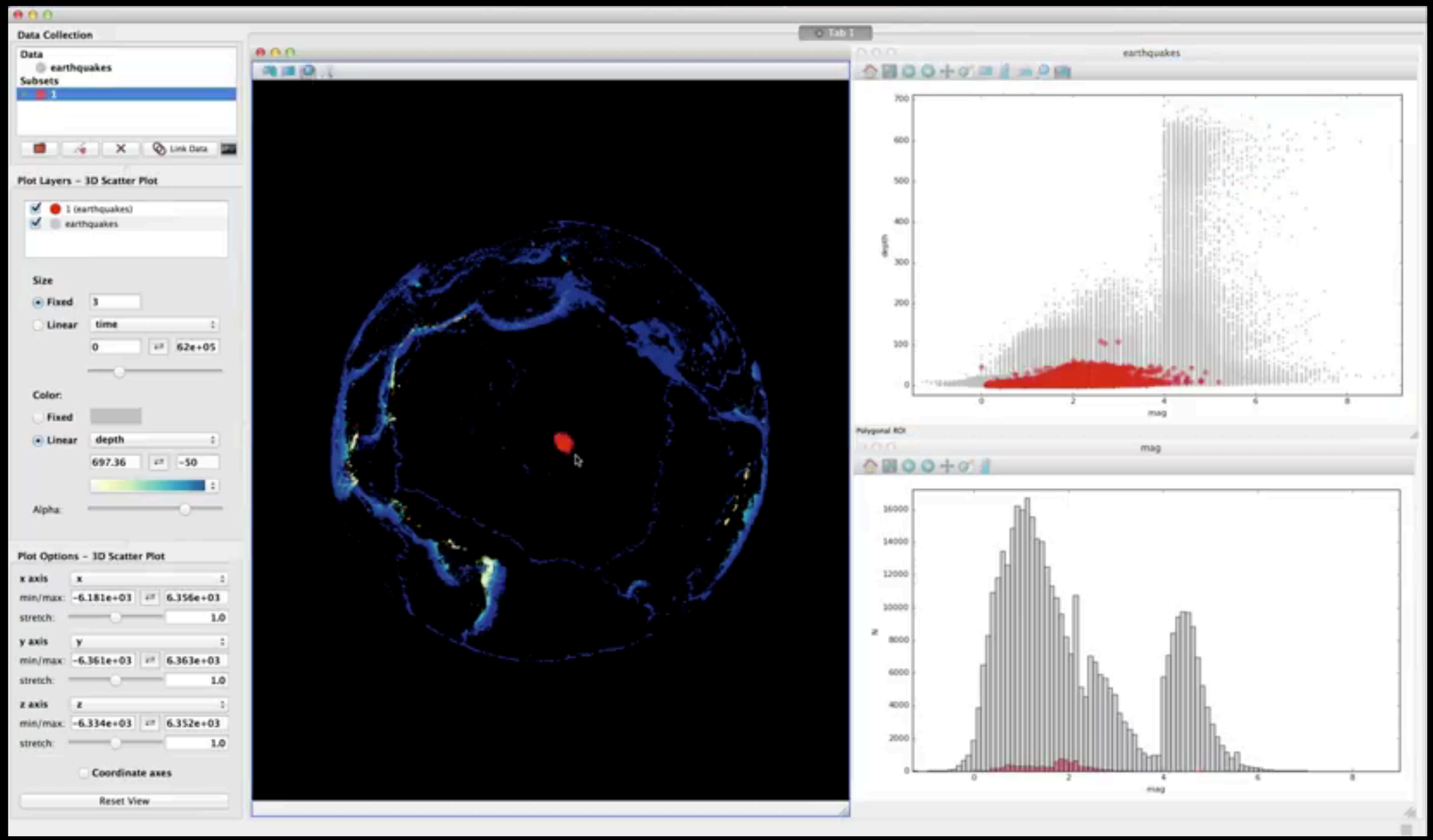
DIVERSE DATA

DIVERSE VIEWS

PRACTICALITY

Linked Views of High-dimensional Data (in Python)

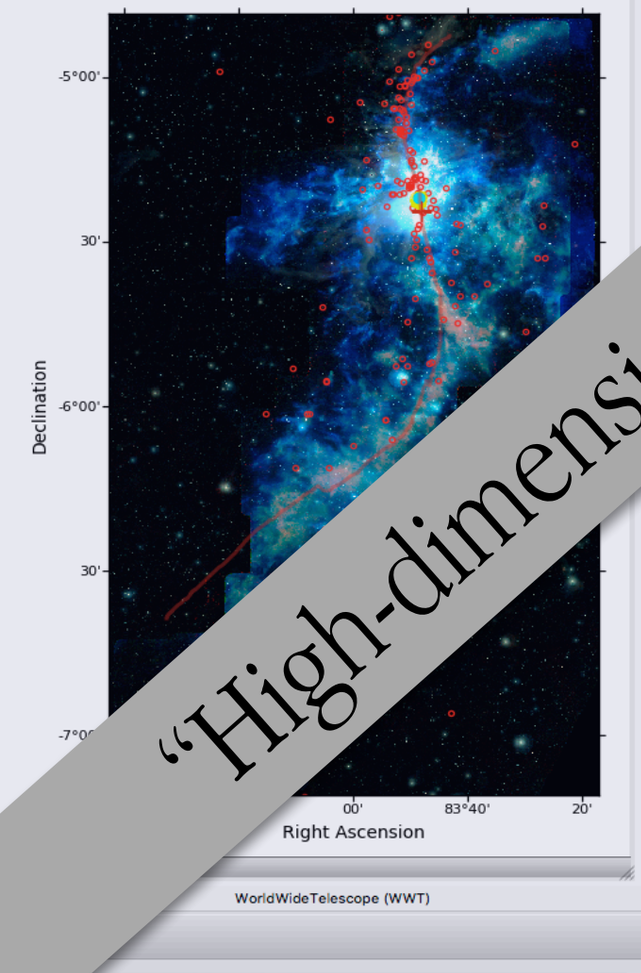
glue

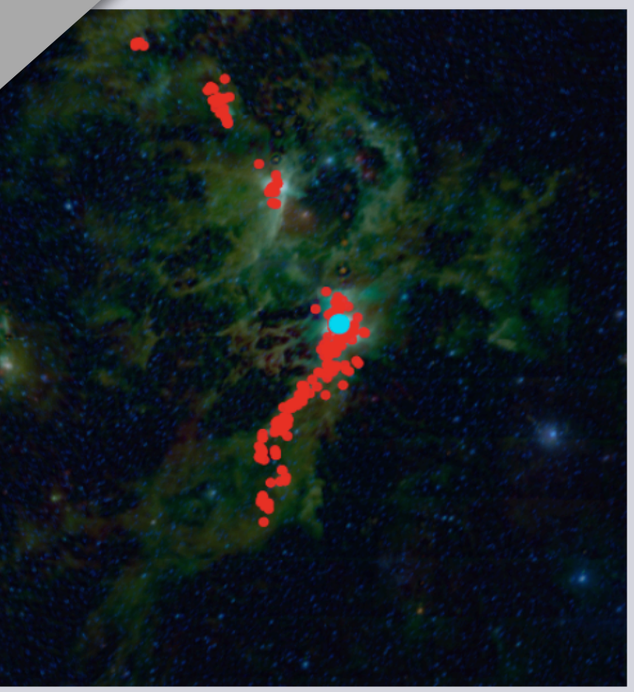


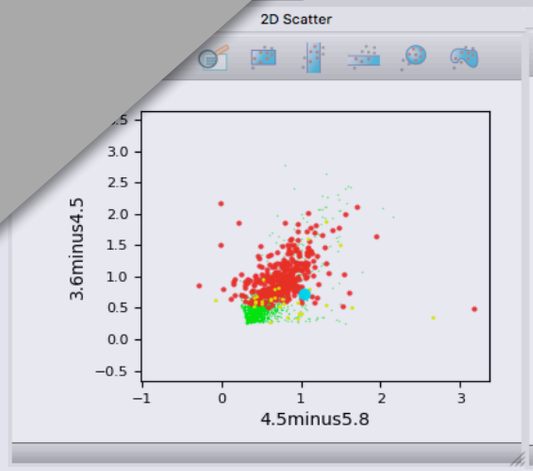
video by Tom Robitaille, lead glue developer

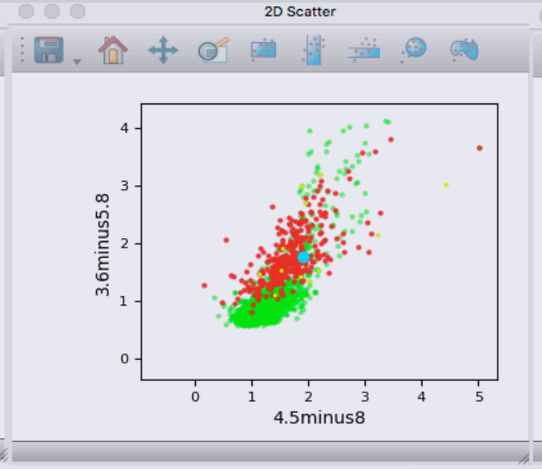
glue created by: C. Beaumont, M. Borkin, M. Breddels, T. Robitaille, C. Zucker, and A. Goodman, PI

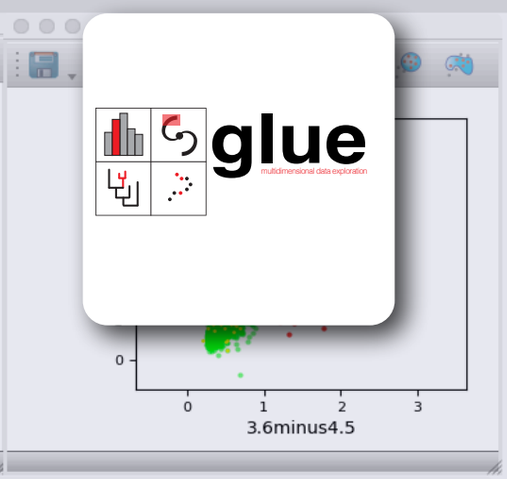
“High-dimensional Data”

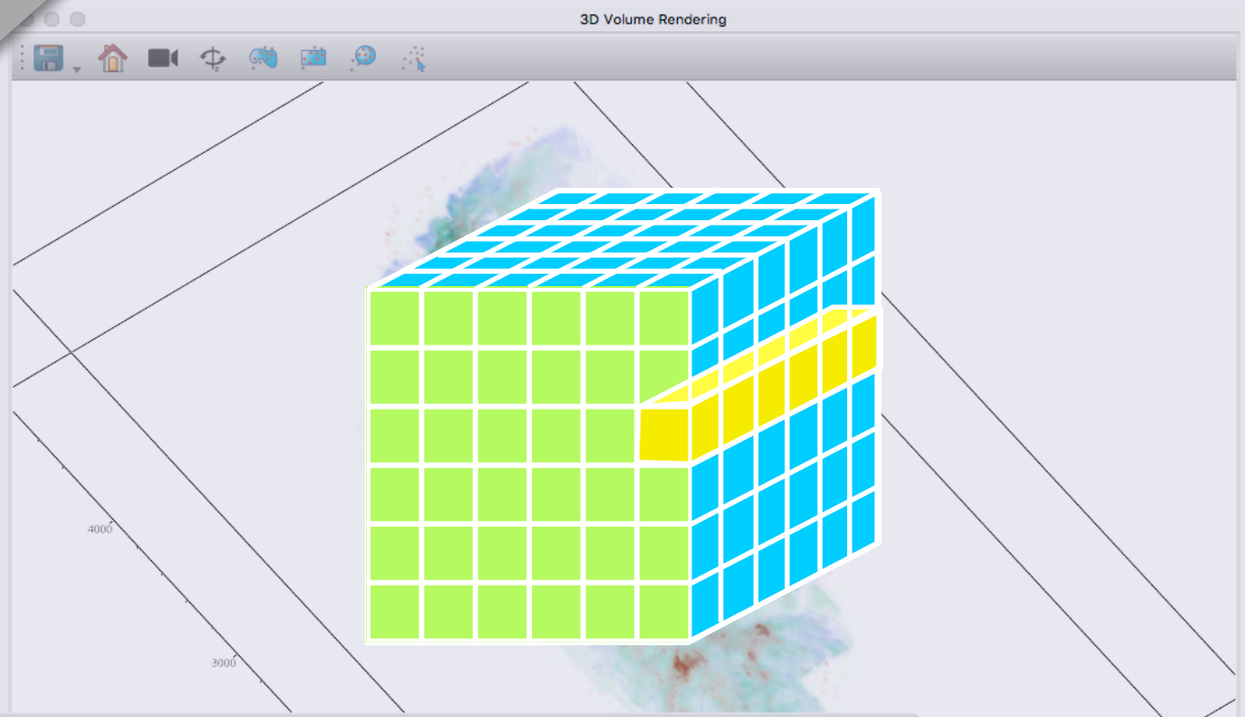


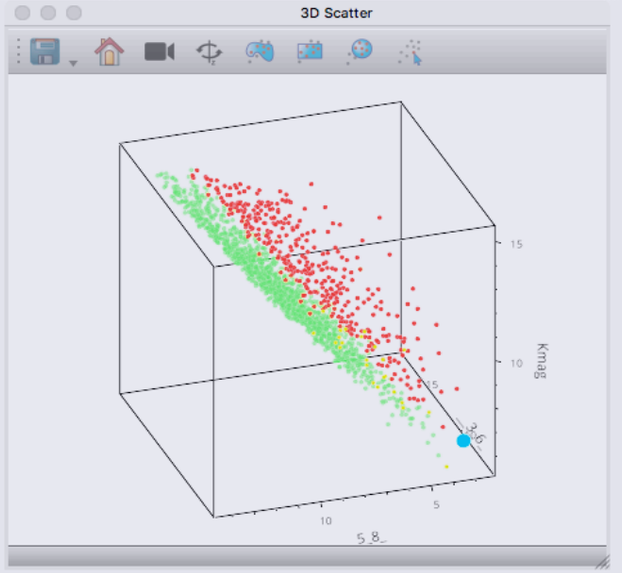


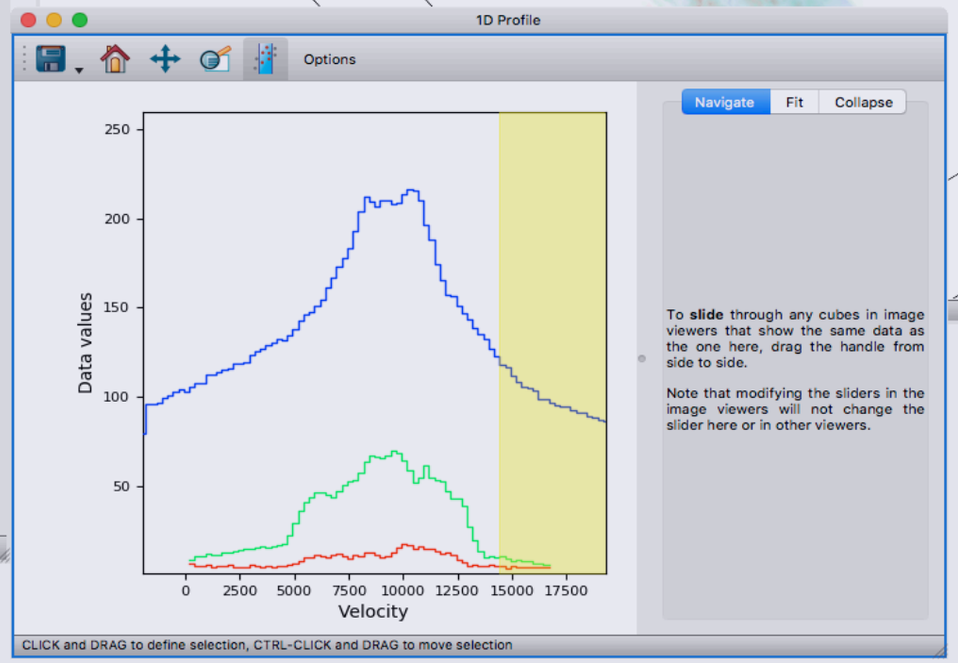


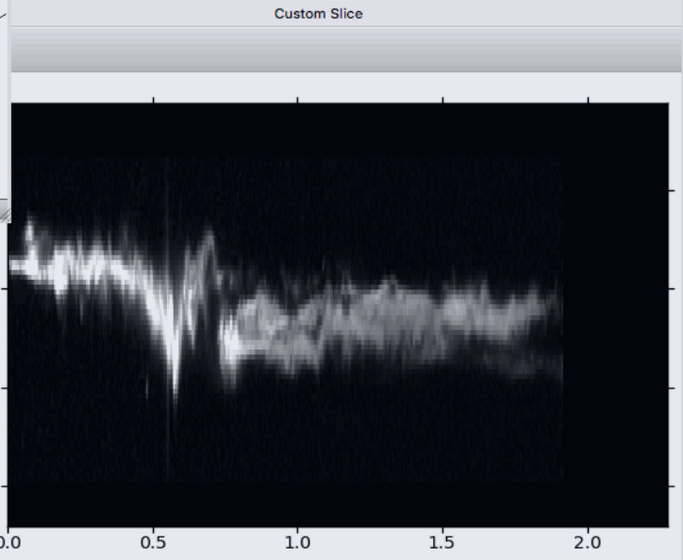




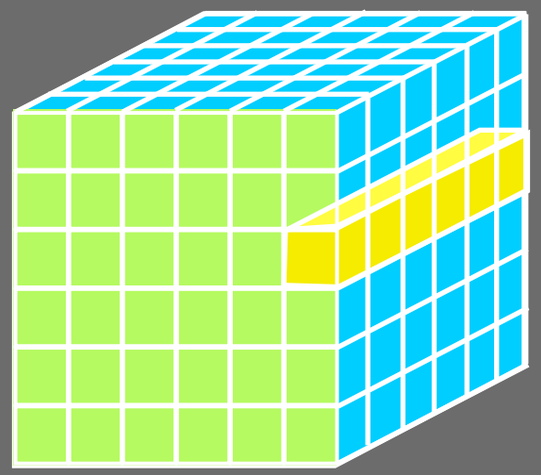






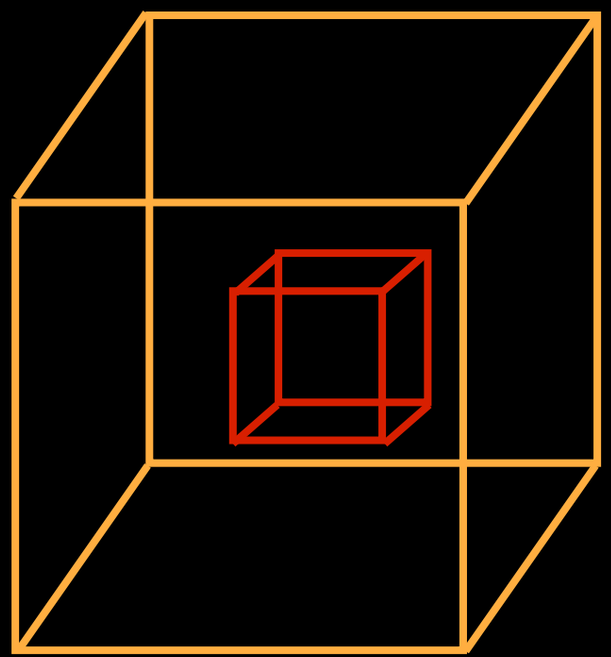


PRINCIPLES

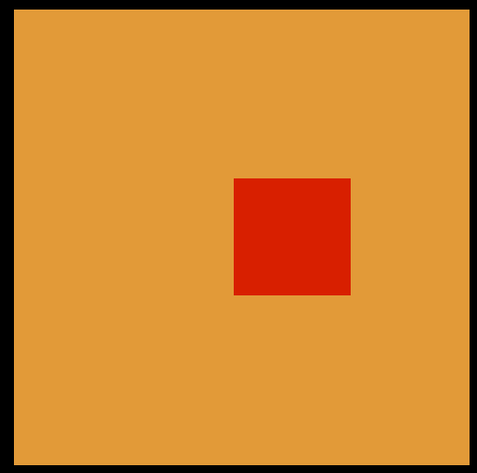


Linked Views of High-dimensional Data

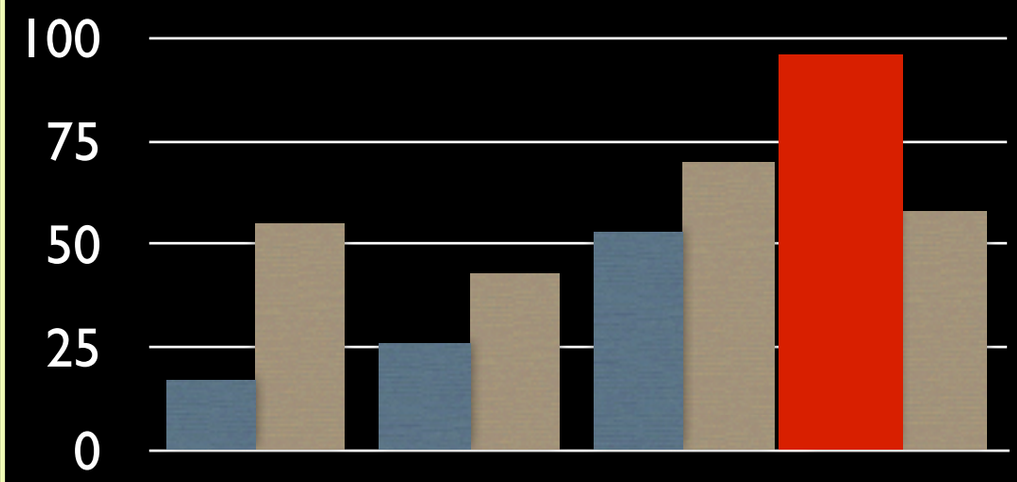
John Tukey



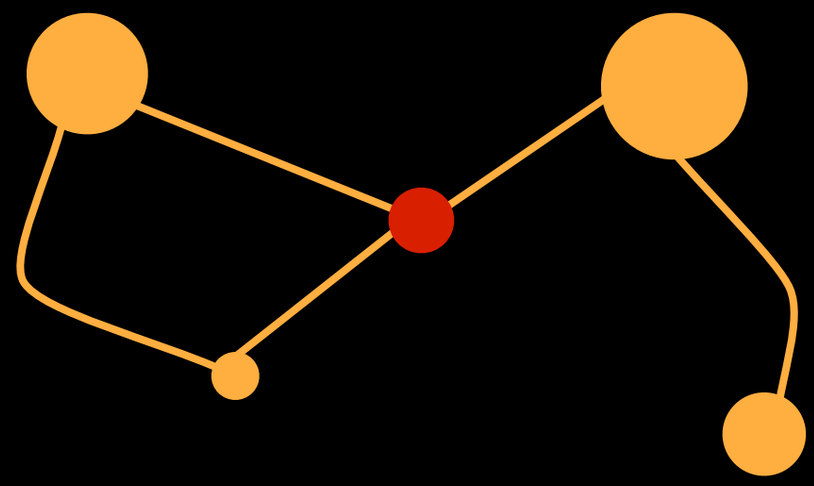
3D



2D



Data Abstraction



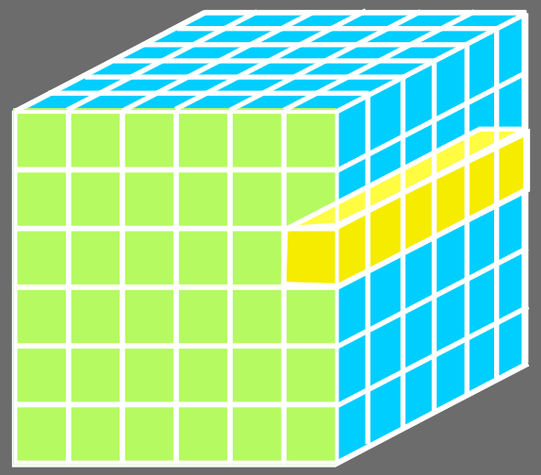
figure, by M. Borkin, reproduced from Goodman 2012, "Principles of High-Dimensional Data Visualization in Astronomy"

PRACTICALITY

glue

multidimensional data exploration

PRINCIPLES



glue

multidimensional data exploration

built-in

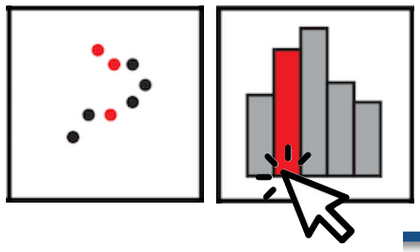
standard data loaders

Data File 1

Data File 2

Data File N

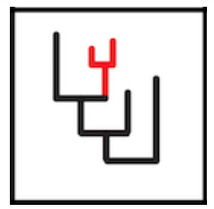
define new variables, import/export insights, interactive plots for the web, save state, all from GUI



standard 1D, 2D & 3D plots

link data files' attributes

highlight live or algorithmic selections with Boolean logic



custom plots

plug-in

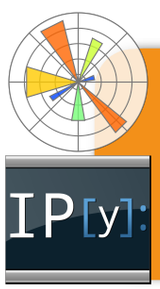
custom data loaders

custom buttons, features

?

+options

user config.py file (loaders, colors, plot types, +)



access to all matplotlib functions through built-in IPython terminal

jupyter

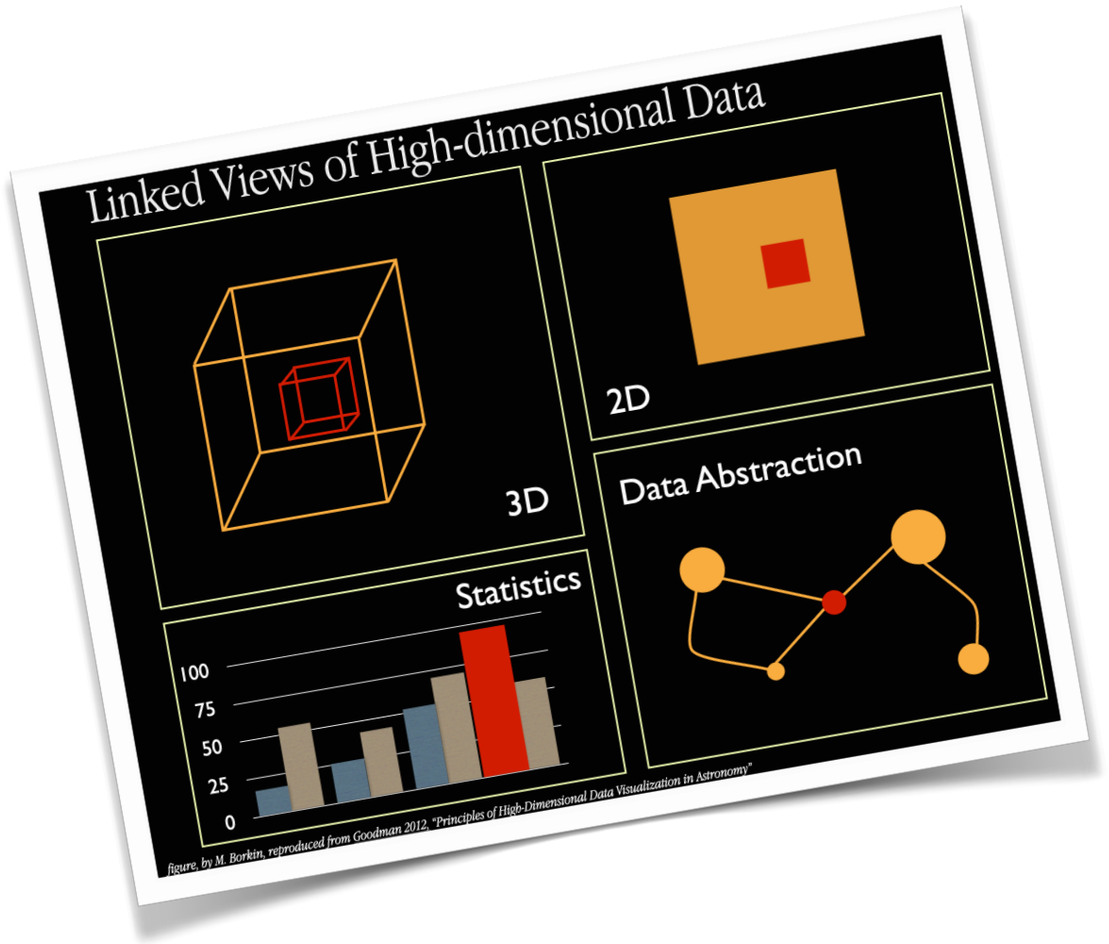
run & interact with glue from Jupyter notebook & other tools

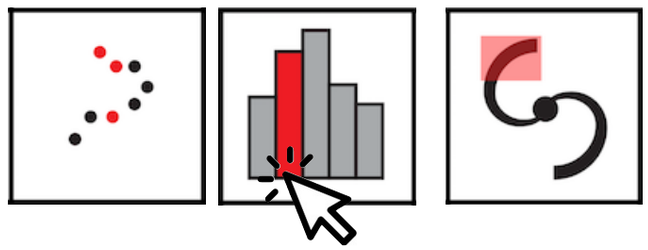
DIVERSE TOOLS DIVERSE DATA DIVERSE VIEWS

glueviz.org

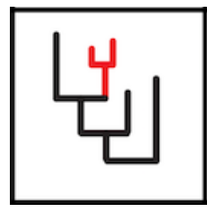
glue

multidimensional data exploration





highlight live or
algorithmic selections
with Boolean logic



DIVERSE VIEWS

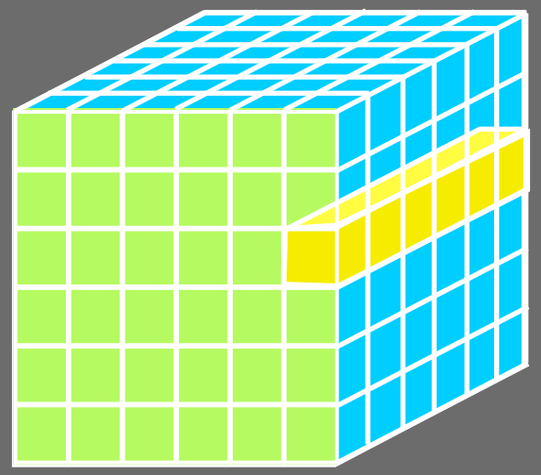
glueviz.org

PRACTICALITY

glue

multidimensional data exploration

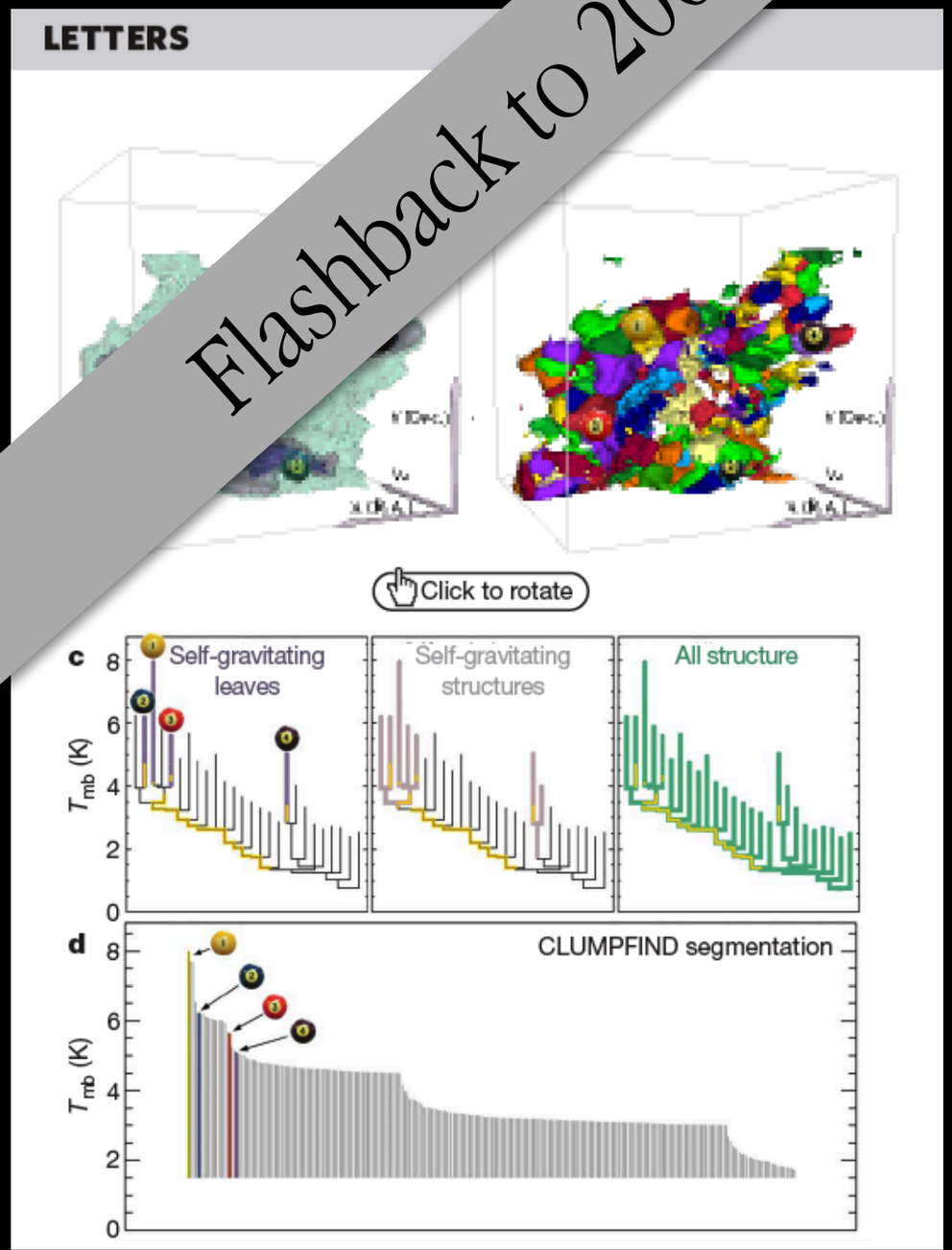
PRINCIPLES



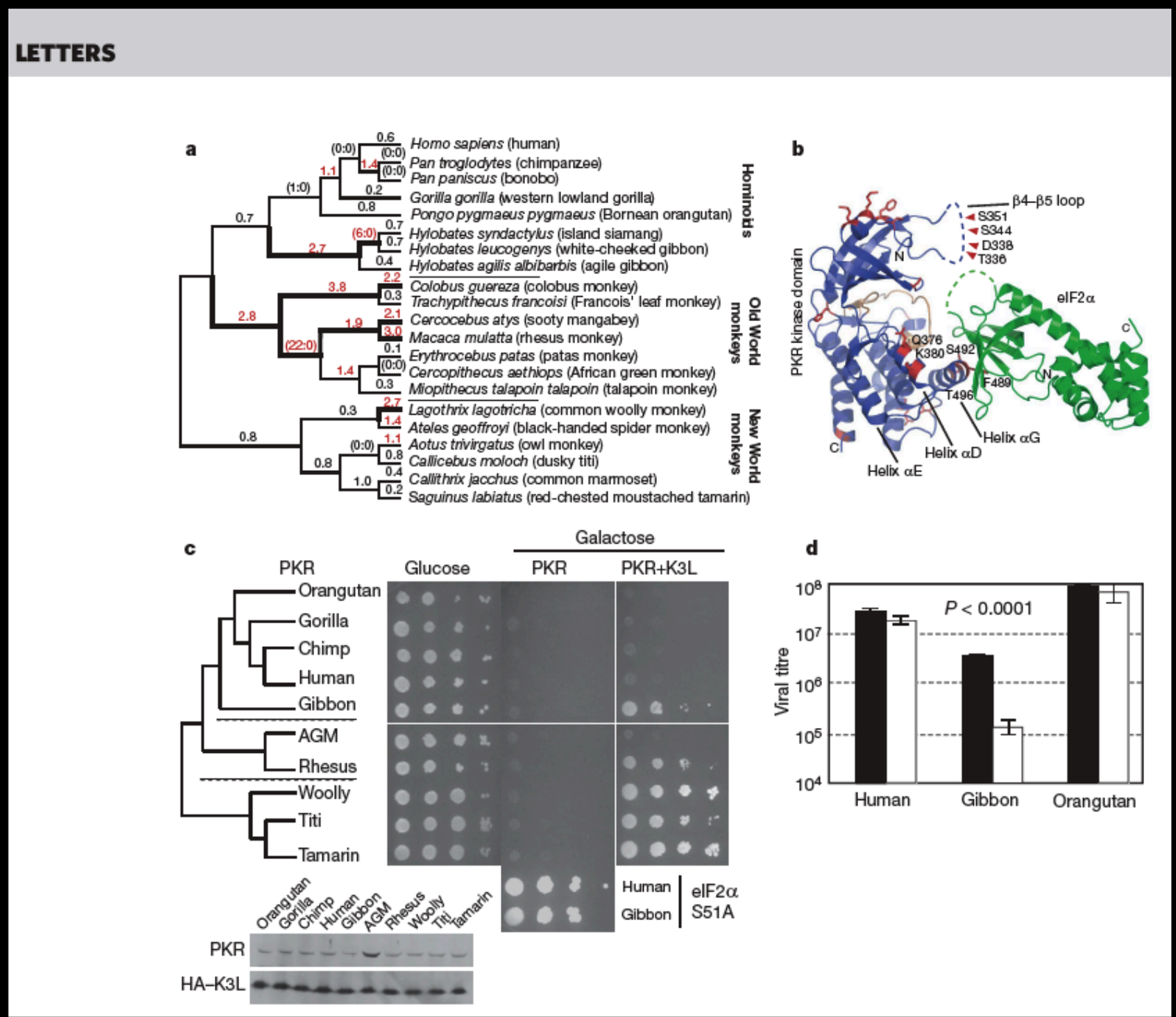
"High-dimen[illegible]" or "Multivariate" Data

(Astronomy=Biology)

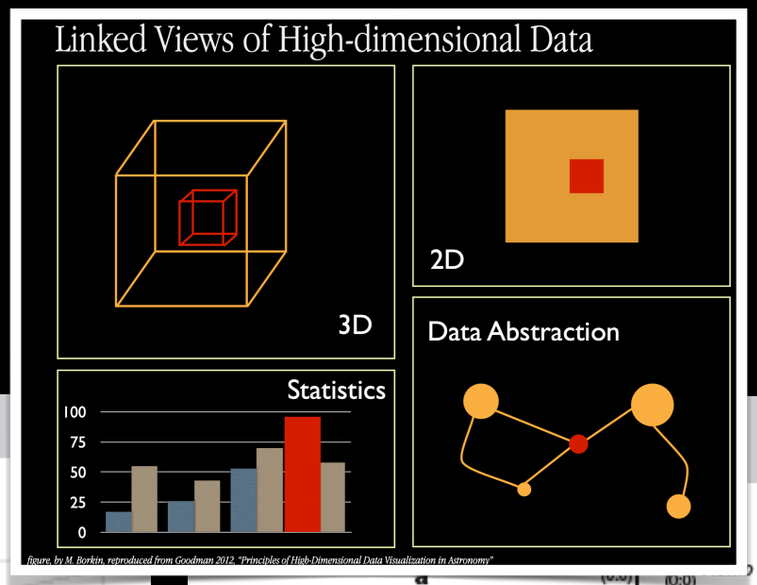
Flashback to 2009 . . .

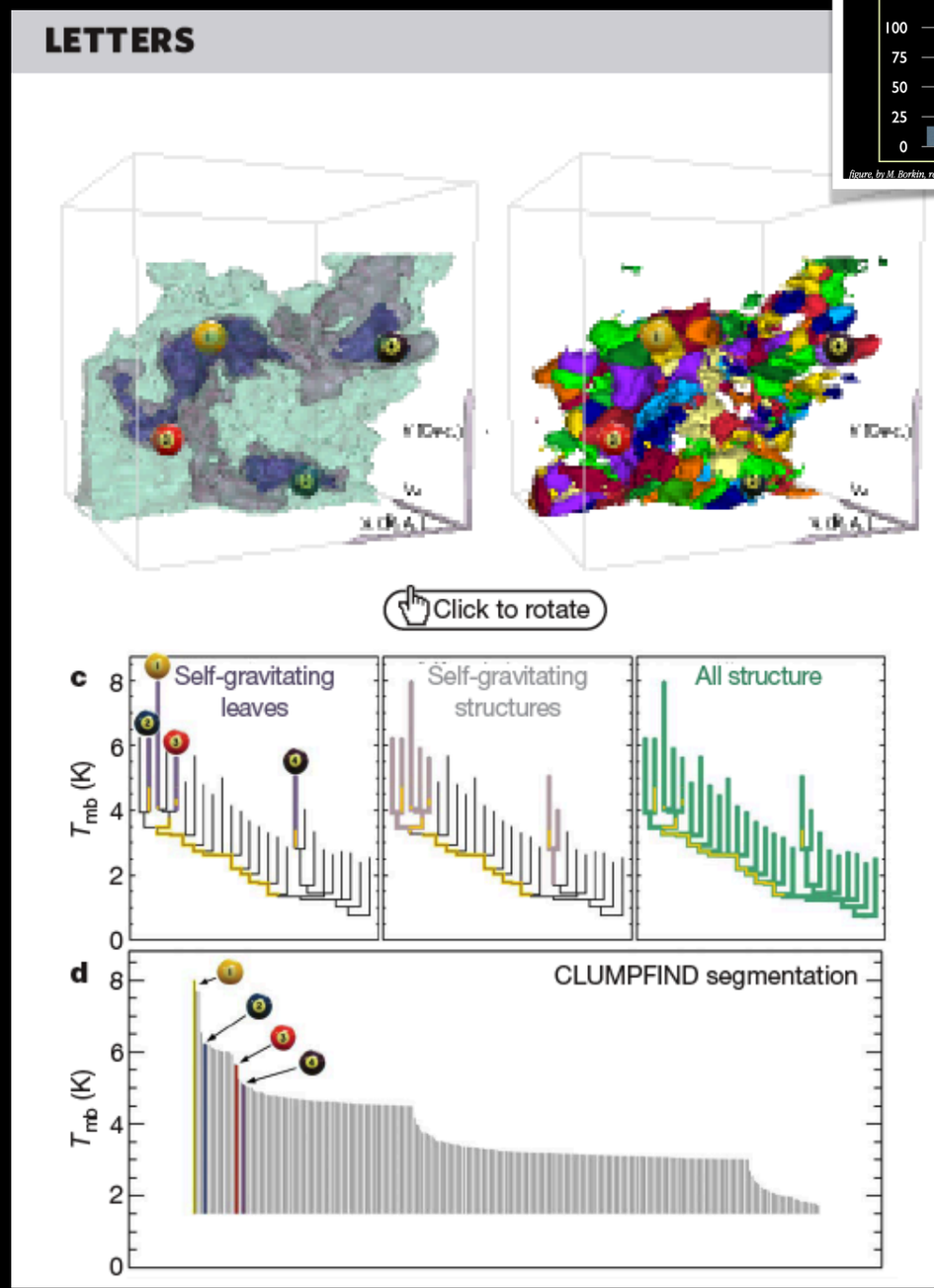


Goodman et al. *Nature*, 2009

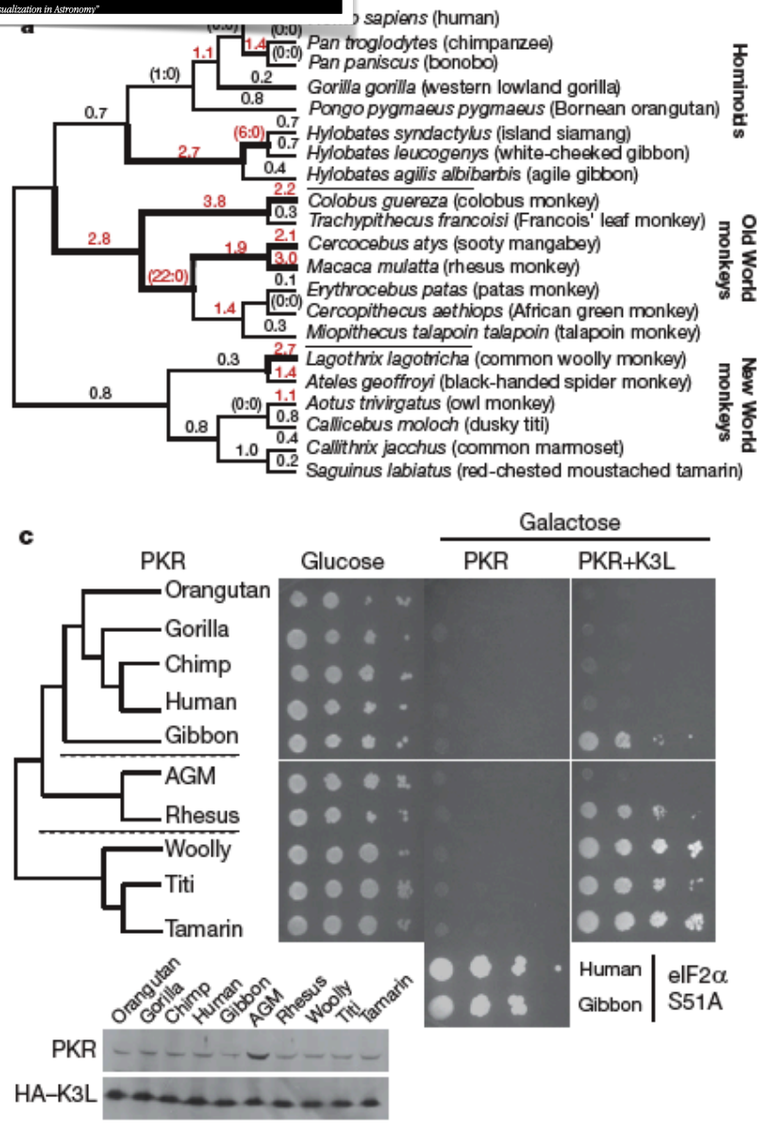


Elde et al. *Nature*, 2008





Goodman et al. *Nature*, 2009



Elde et al. *Nature*, 2008

Diverse High-D Data

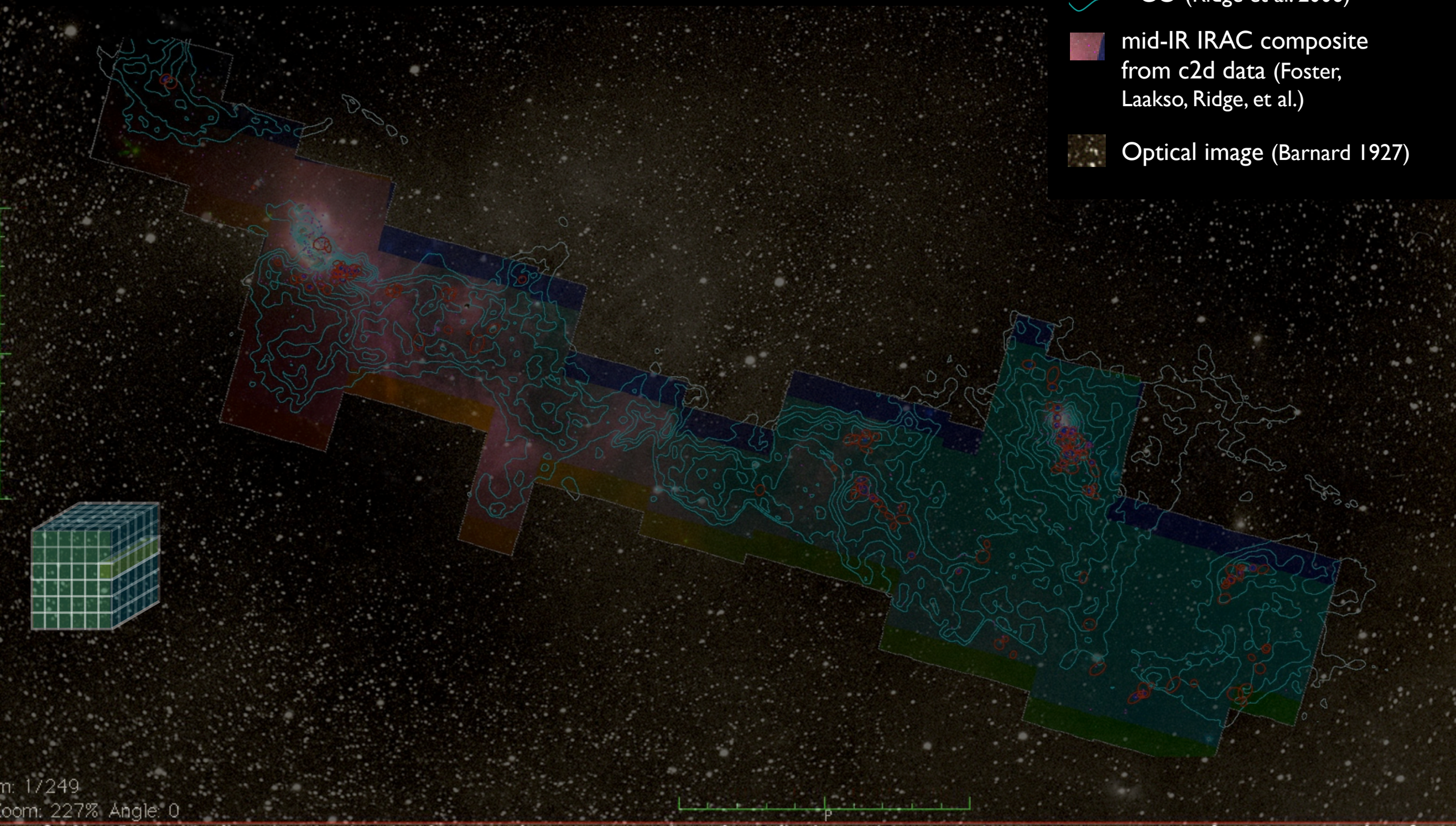
mm peak (Enoch et al. 2006)

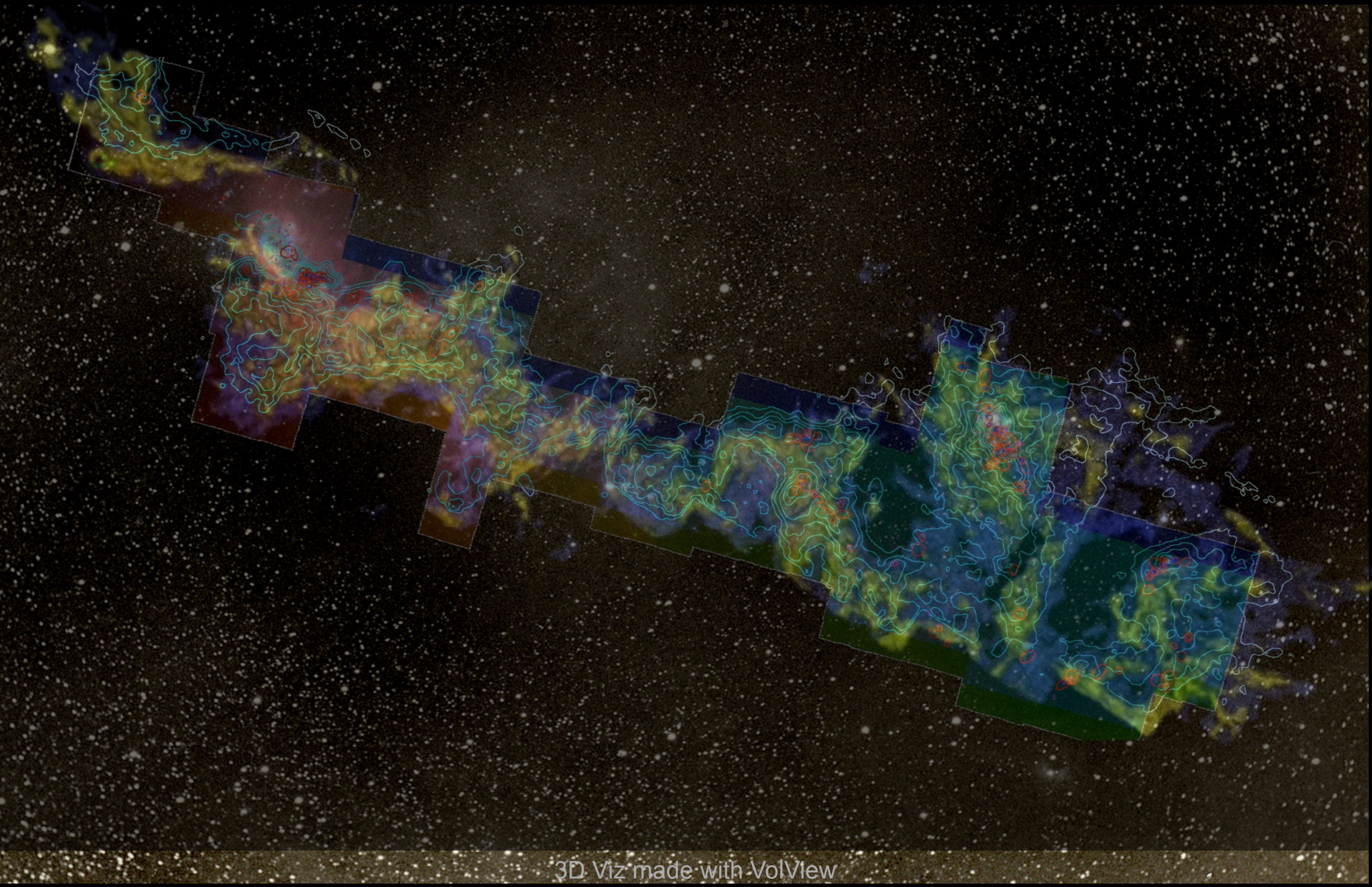
sub-mm peak (Hatchell et al. 2005, Kirk et al. 2006)

^{13}CO (Ridge et al. 2006)

mid-IR IRAC composite from c2d data (Foster, Laakso, Ridge, et al.)

Optical image (Barnard 1927)





AstronomicalMedicine@IIG

COMPLETE

TODAY

python File Edit View Canvas Data Manager Plugins Help

20% Mon Oct 22 7:15 AM Alyssa A Goodman

Glue

Open Session | Export Session | Import Data | Export Data/Subsets | Active Subset: Subset 3 | Mode: | Terminal | Preferences | Error Console

Data Collection

Data
- PerA_12coFCRAO_F_xyv
- PerA_13coFCRAO_F_xyv
- Perseus_Av_NICER
- PerA_AvTemMIPS_F_Av
- PerA_AvTemMIPS_F_T
- Perseus_fcrao_iras_2mass

Subsets
- spec_probe
- hot_highv
- Subset 3

Tab 1 | Tab 2

1D Profile

Data values

Velocity

2D Scatter

TIRAS

V13CO

2D Scatter

Dec

RA

3D Scatter

NIRAS

RA

Dec

Plot Layers -

Errors | Vectors

Plot Options - 2D Scatter

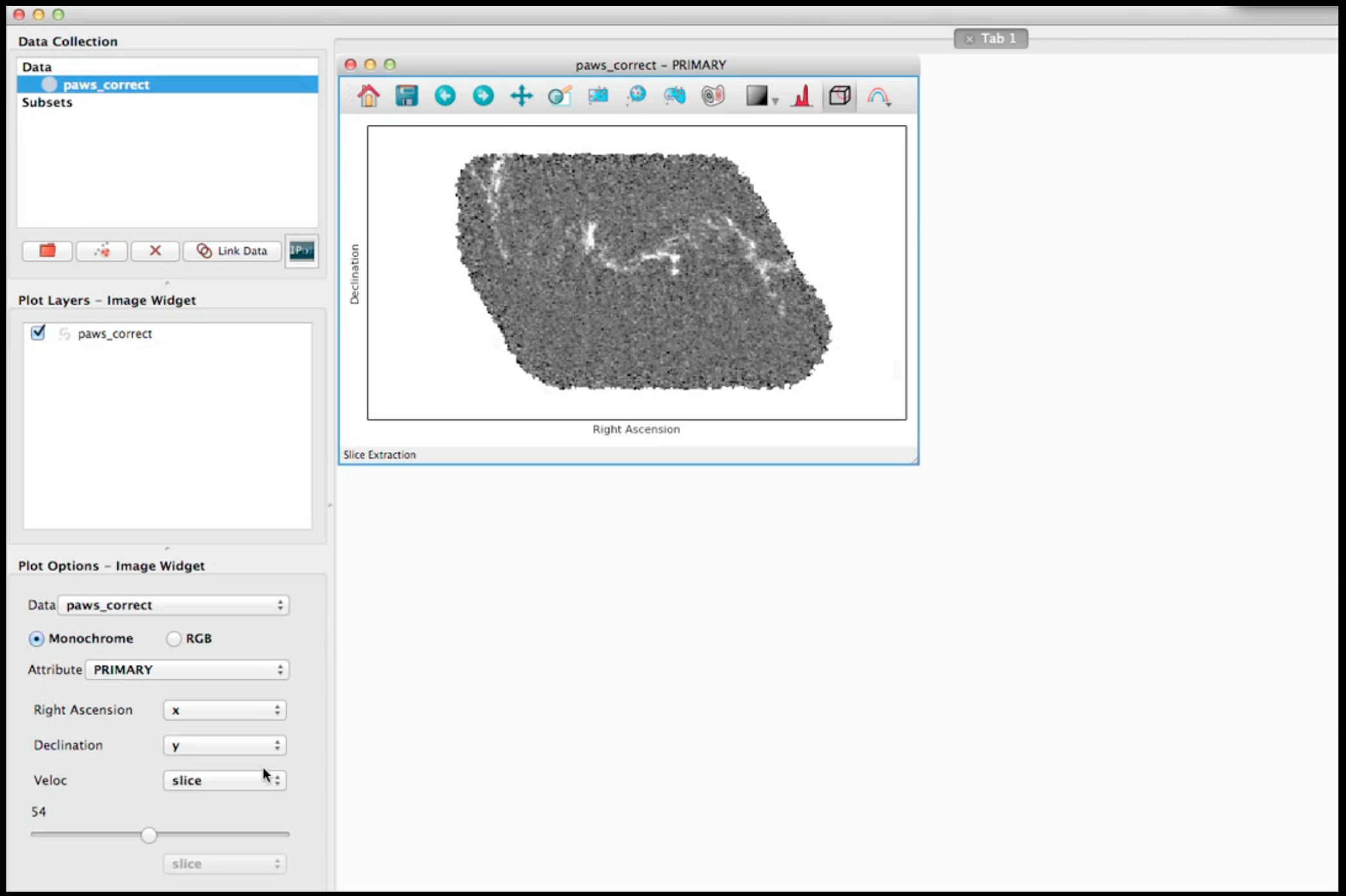
General | Limits | Axes

x axis V13CO log

y axis TIRAS log

“Dimension” doesn’t always mean x, y, or z…

glue

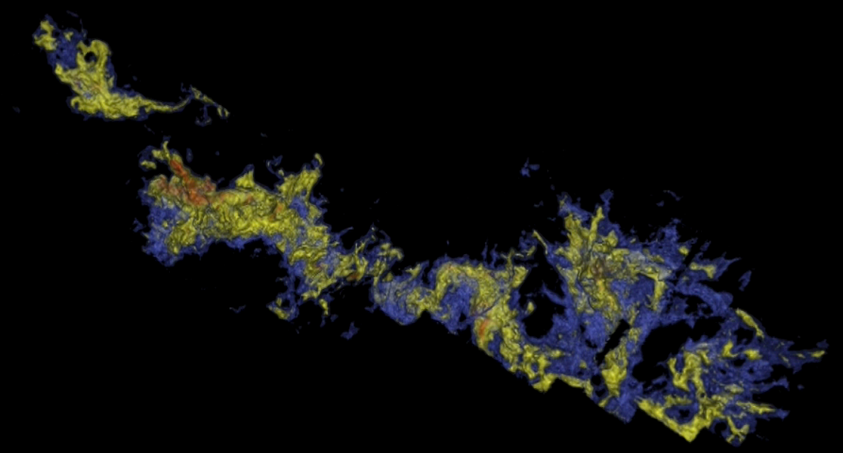


video by Chris Beaumont, glue developer

glue created by: C. Beaumont, M. Borkin, M. Breddels, T. Robitaille, C. Zucker, and A. Goodman, PI

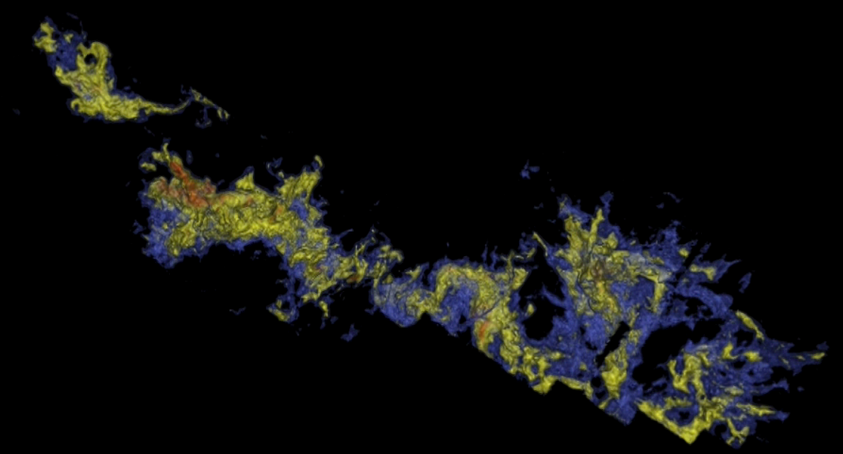
“Dimension” isn’t even always spatial…

The “3rd” dimension in this 3D plot is “velocity” coming from Doppler Spectroscopy.



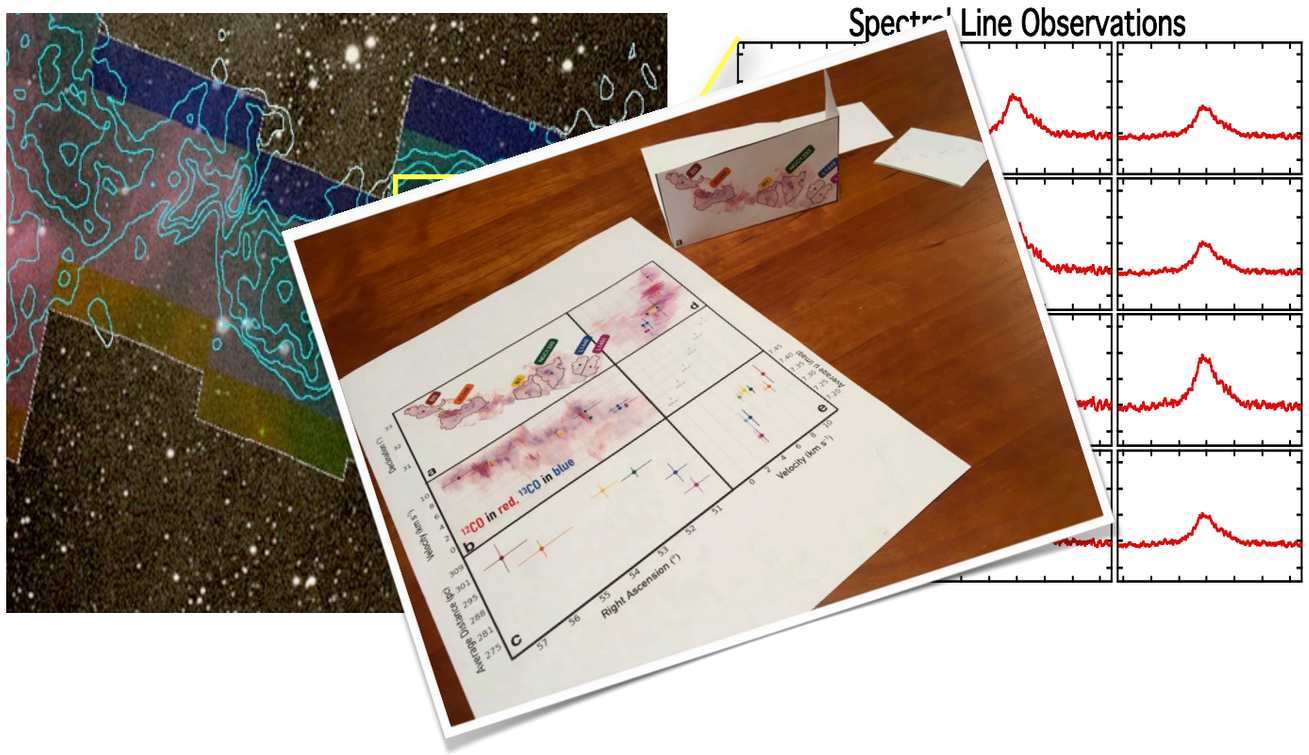
“Dimension” isn’t even always spatial…

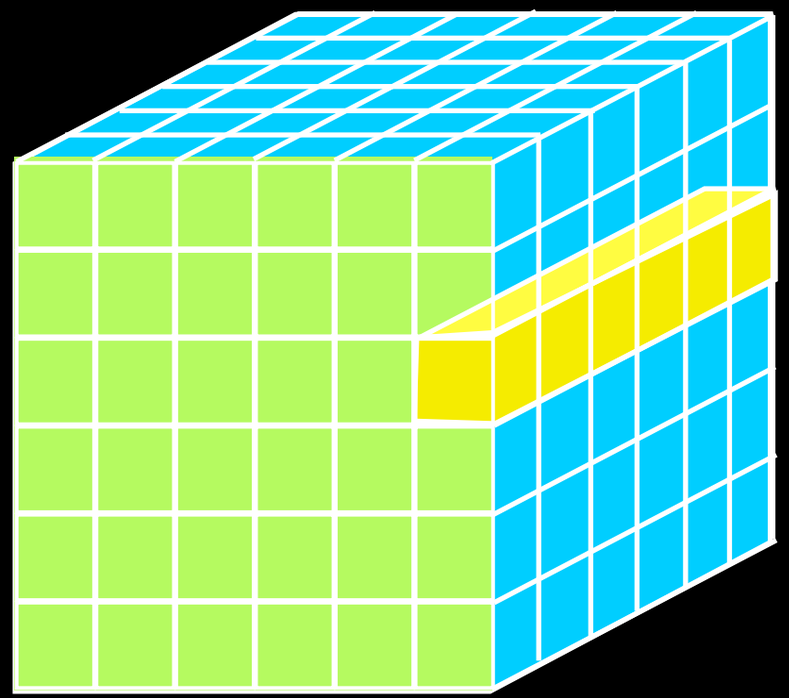
The “3rd” dimension in this 3D plot is “velocity” coming from Doppler Spectroscopy.



glue
multidimensional data exploration

“Perseus” progress preview





PRINCIPLE

"DATA, DIMENSIONS, DISPLAY"

1D: Columns = “Spectrum”,“Time Series,” “Sequence”
2D: Faces or Slices = “Images,” “Arrays”
3D: Volumes = “3D Renderings”, “2D Movies”
4D: Time Series of Volumes = “3D Movies”

DIVERSE DATA DIVERSE VIEWS

glue

multidimensional data exploration

built-in

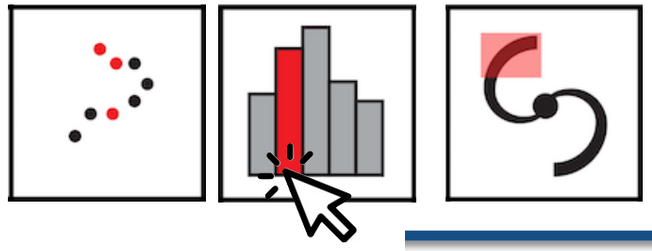
standard data loaders

Data File 1

Data File 2

Data File N

define new variables, import/export insights, interactive plots for the web, save state, all from GUI



standard 1D, 2D & 3D plots

link data files' attributes

highlight live or algorithmic selections with Boolean logic

plug-in

custom data loaders

custom buttons, features

?

custom plots

+options

user config.py file (loaders, colors, plot types, +)

access to all matplotlib functions through built-in IPython terminal

jupyter

run & interact with glue from Jupyter notebook & other tools

DIVERSE DATA

glueviz.org

python File Edit View Canvas Data Manager Plugins Help

21% Mon Oct 22 7:15 AM Alyssa A Goodman

Glue

Open Session | Export Session | Import Data | Export Data/Subsets | Link Data | Arithmetic attributes | Active Subset: Subset 3 | Mode: | Terminal | Preferences | Error Console

Data Collection

Data
- PerA_12coFCRAO_F_xyv
- PerA_13coFCRAO_F_xyv
- Perseus_Av_NICER
- PerA_AvTemMIPS_F_Av
- PerA_AvTemMIPS_F_T
- Perseus_fcrao_iras_2mass

Subsets
- spec_probe
- hot_highv
- Subset 3

Plot Layers - 2D Scatter
- Subset 3 (Perseus_fcrao_iras_2mass)
- hot_highv (Perseus_fcrao_iras_2mass)
- spec_probe (Perseus_fcrao_iras_2mass)
- Perseus_fcrao_iras_2mass

Color | Points | Line | Errors | Vectors

color: Fixed

opacity

Plot Options - 2D Scatter

General | Limits | Axes

x axis: V13CO log

y axis: TIRAS log

Tab 1 | Tab 2

2D Image

1D Profile

2D Scatter

Link Editor

Click on two datasets to set up links or click on an existing connection to edit links. Selected datasets are shown in green. When one dataset is selected, the colors show directly and indirectly linked (blue) and inaccessible (red) datasets.

PerA_AvTemMIPS_F_T
PerA_AvTemMIPS_F_Av
Perseus_Av_NICER
PerA_13coFCRAO_F_xyv
PerA_12coFCRAO_F_xyv
Perseus_fcrao_iras_2mass

Data File 1

Data File 2

Data File N

Datasets: PerA_AvTemMIPS_F_Av

Components
- Coordinate components
- Pixel Axis 0 [y]
- Pixel Axis 1 [x]
- Declination
- Right Ascension
- Main components
- PRIMARY

Glue

Datasets: PerA_12coFCRAO_F_xyv

Components
- Coordinate components
- Pixel Axis 0 [z]
- Pixel Axis 1 [y]
- Pixel Axis 2 [x]
- Velocity
- Declination
- Right Ascension
- Main components
- PRIMARY

Current Links

Function	Component 1	Component 2
identity	Right Ascension	Right Ascension
identity	Declination	Declination

Unglue

Advanced linking | Cancel | OK

V13CO

3D Scatter

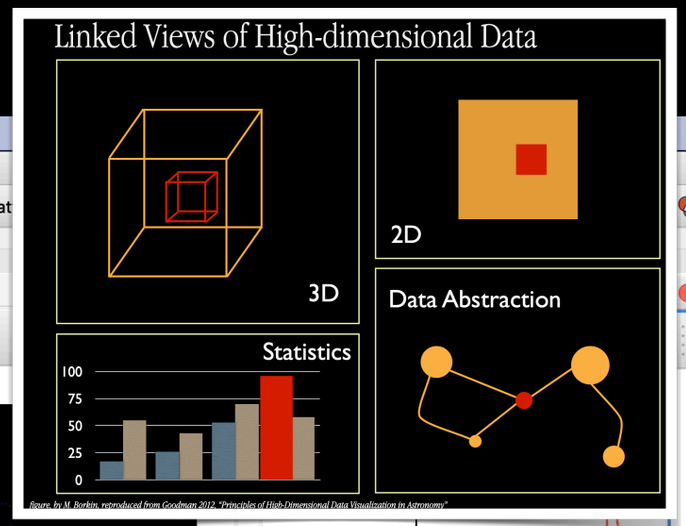
NIRAS

RA

Pixel Axis 0 [z]

Pixel Axis 1 [y]

Pixel Axis 2 [x]



python File Edit View Canvas Data Manager Plugins Help

20% Mon Oct 22 7:15 AM Alyssa A Goodman

Open Session | Export Session | Import Data | Export Data/Subsets | Link Data | Arithmetic at | Terminal | Preferences | Error Console

Data Collection

Data
- PerA_12coFCRAO_F_xyv
- PerA_13coFCRAO_F_xyv
- Perseus_Av_NICER
- PerA_AvTemMIPS_F_Av
- PerA_AvTemMIPS_F_T
- Perseus_fcrao_iras_2mass

Subsets
- spec_probe
- hot_highv
- Subset 3

Plot Layers - 2D Scatter
- Subset 3 (Perseus_fcrao_iras_2mass)
- hot_highv (Perseus_fcrao_iras_2mass)
- spec_probe (Perseus_fcrao_iras_2mass)
- Perseus_fcrao_iras_2mass

Color | Points | Line | Errors | Vectors

color: Fixed

opacity

Plot Options - 2D Scatter

General | Limits | Axes

x axis: V13CO log

y axis: TIRAS log

2D Image

2D Scatter

Dec

RA

Data values

Velocity

2D Scatter

TIRAS

V13CO

3D Scatter

NIRAS

RA

Pixel Axis 0 [z]

Pixel Axis 1

1000

DIVERSE TOOLS DIVERSE DATA DIVERSE VIEWS

glue

multidimensional data exploration

built-in

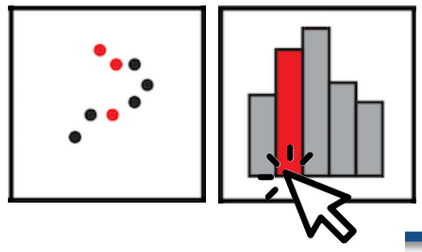
standard data loaders

Data File 1

Data File 2

Data File N

define new variables, import/export insights, interactive plots for the web, save state, all from GUI



standard 1D, 2D & 3D plots

link data files' attributes

highlight live or algorithmic selections with Boolean logic

plug-in

custom data loaders

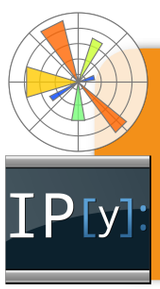
custom buttons, features

?

custom plots

+options

user config.py file (loaders, colors, plot types, +)



access to all matplotlib functions through built-in IPython terminal

jupyter

run & interact with glue from Jupyter notebook & other tools

DIVERSE TOOLS DIVERSE DATA DIVERSE VIEWS

glueviz.org

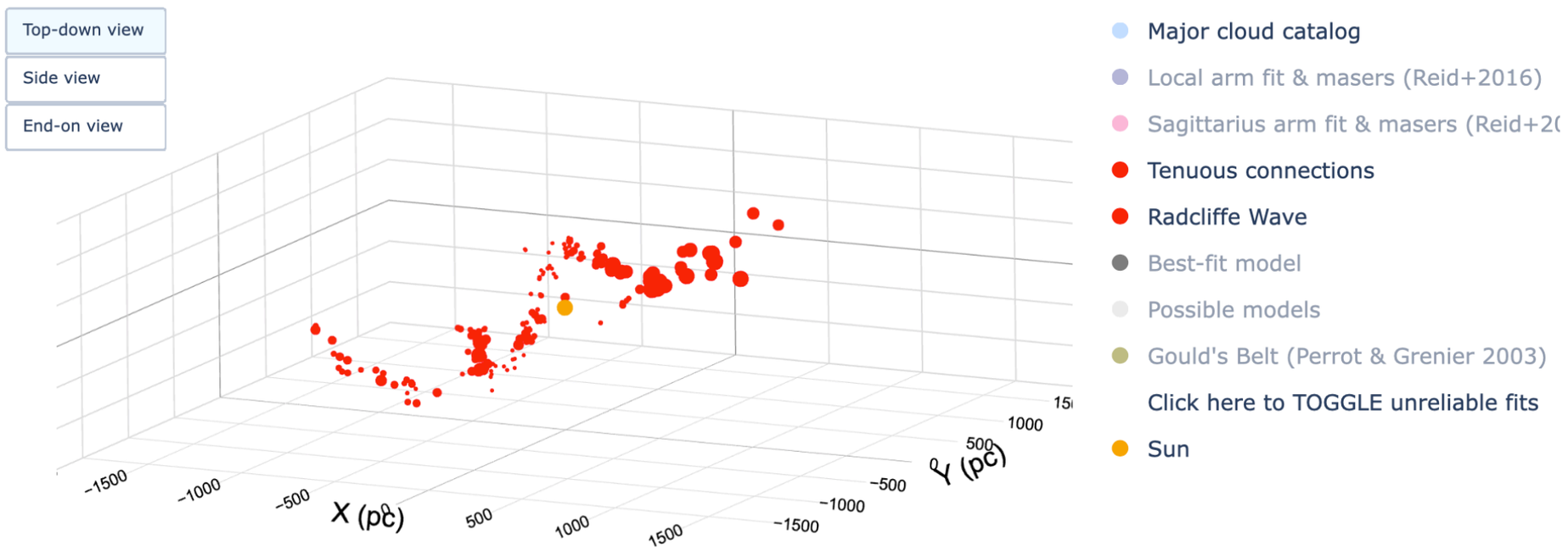
[live demo of glue + WorldWide Telescope plug-in, using Radcliffe Wave data]

VISUALS

On this page: INTERACTIVES, FIGURES, VIDEOS -- scroll down to see it all.

INTERACTIVES

Explore the RadWave in 3D



DIVERSE TOOLS

"YOUR" WORLD

SINGLE-CELL SEQUENCING & CELL VIEW

Cold Spring Harbor Laboratory
bioRxiv
THE PREPRINT SERVER FOR BIOLOGY

HOME | ABO

Search

bioRxiv is receiving many new papers on coronavirus SARS-CoV-2. A reminder: these are preliminary reports that have not bee
practice/health-related behavior, or be reported in news media as established information.

New Results

Comment on this paper

Cellular taxonomy and spatial organization of the ventral posterior hypothalamus reveals neuroanatomical parcellation of the mammillary bodies

Laura E. Mickelsen, William F. Flynn, Kristen Springer, Lydia Wilson, Eric J. Beltrami, Mohan Bolisetty, Paul Robson, Alexander C. Jackson

doi: https://doi.org/10.1101/2020.05.14.096818

This article is a preprint and has not been certified by peer review [what does this mean?].

Abstract | Full Text | Info/History | Metrics | Preview PDF

ABSTRACT

The ventral posterior hypothalamus (VPH) is an anatomically complex brain region implicated in arousal, reproduction, energy balance and memory processing. However, neuronal cell type diversity within the VPH is poorly understood, an impediment to deconstructing the roles of distinct VPH circuits in physiology and behavior. To address this question, we employed a droplet-based single cell RNA sequencing (scRNA-seq) approach to systematically classify molecularly distinct cell types in the mouse VPH. Analysis of >16,000 single cells revealed 20 neuronal and 18 non-neuronal cell populations, defined by suites of discriminatory markers. We validated differentially expressed genes in a selection of neuronal populations through fluorescence *in situ* hybridization (FISH). Focusing on the mammillary bodies (MB), we discovered transcriptionally-distinct clusters that exhibit a surprising degree of segregation within neuroanatomical subdivisions of the MB, while genetically-defined MB cell types project topographically to the anterior thalamus. This single cell transcriptomic atlas of cell types in the VPH provides a detailed resource for interrogating the circuit-level mechanisms underlying the diverse functions of VPH circuits in health and disease.

accepted for publication at eLife (Mickelsen, Flynn, …Robson, Jackson) 'Cellular taxonomy and spatial organization of the murine ventral posterior hypothalamus' doi: 10.7554/eLife.58901

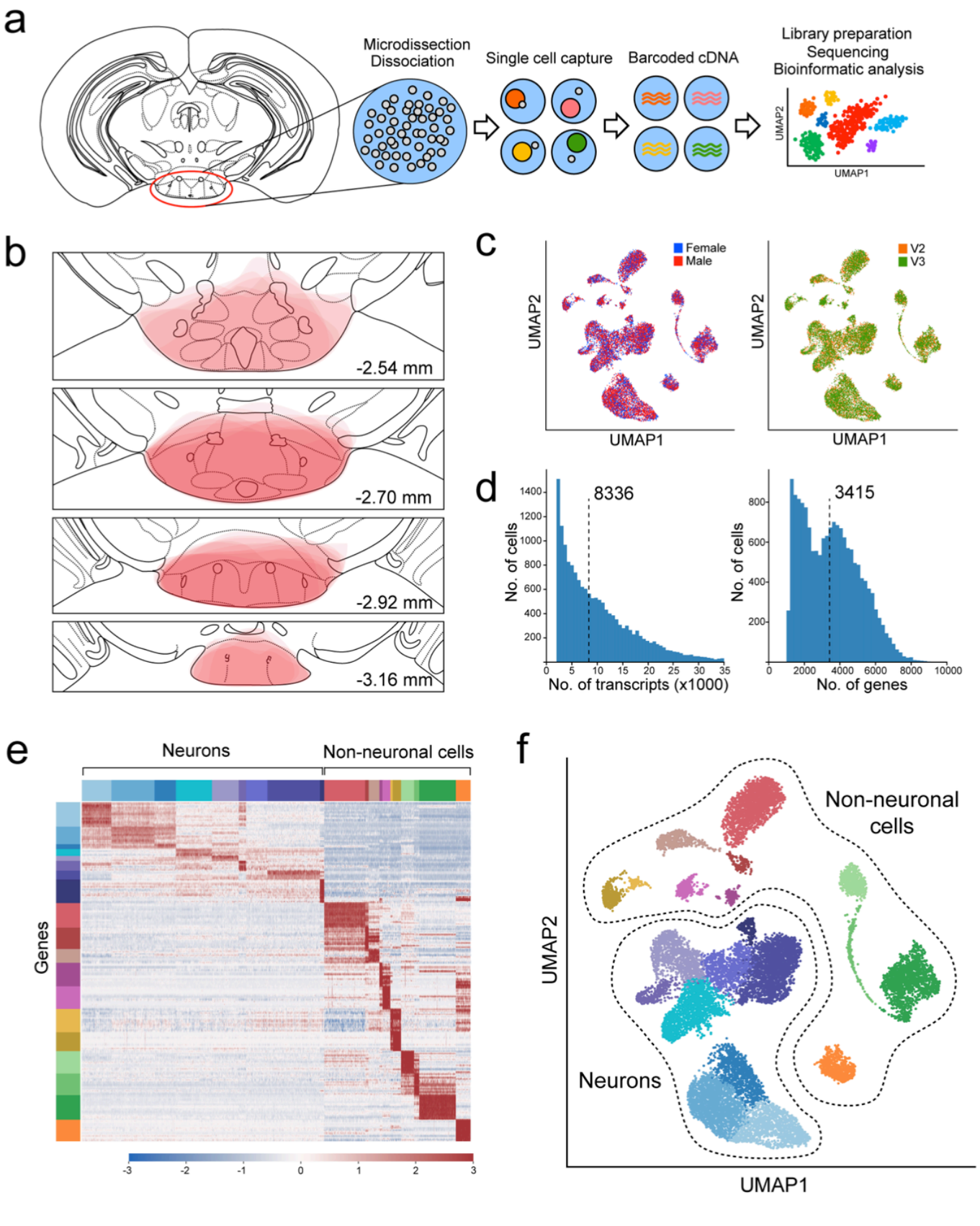


Figure 2:

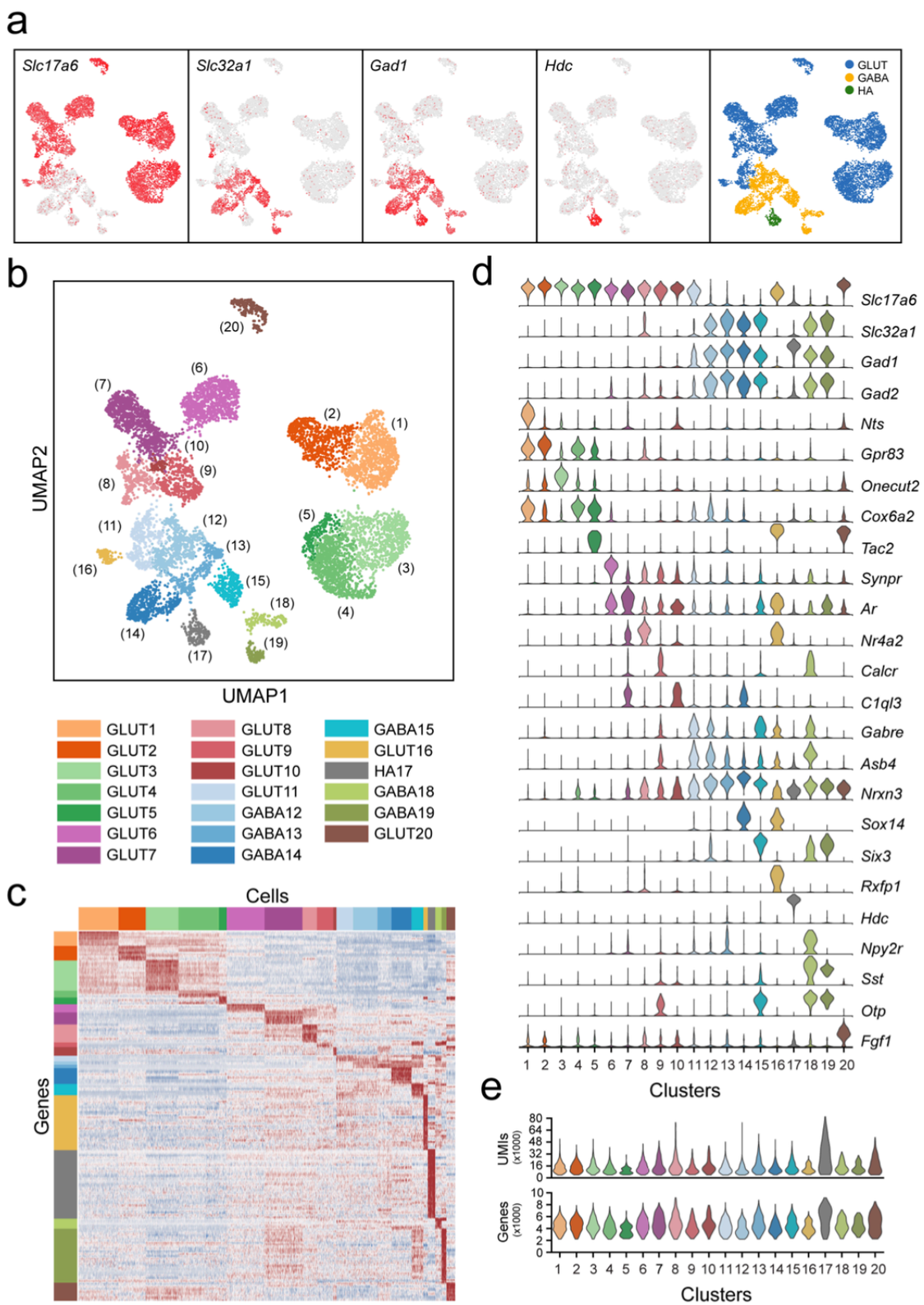
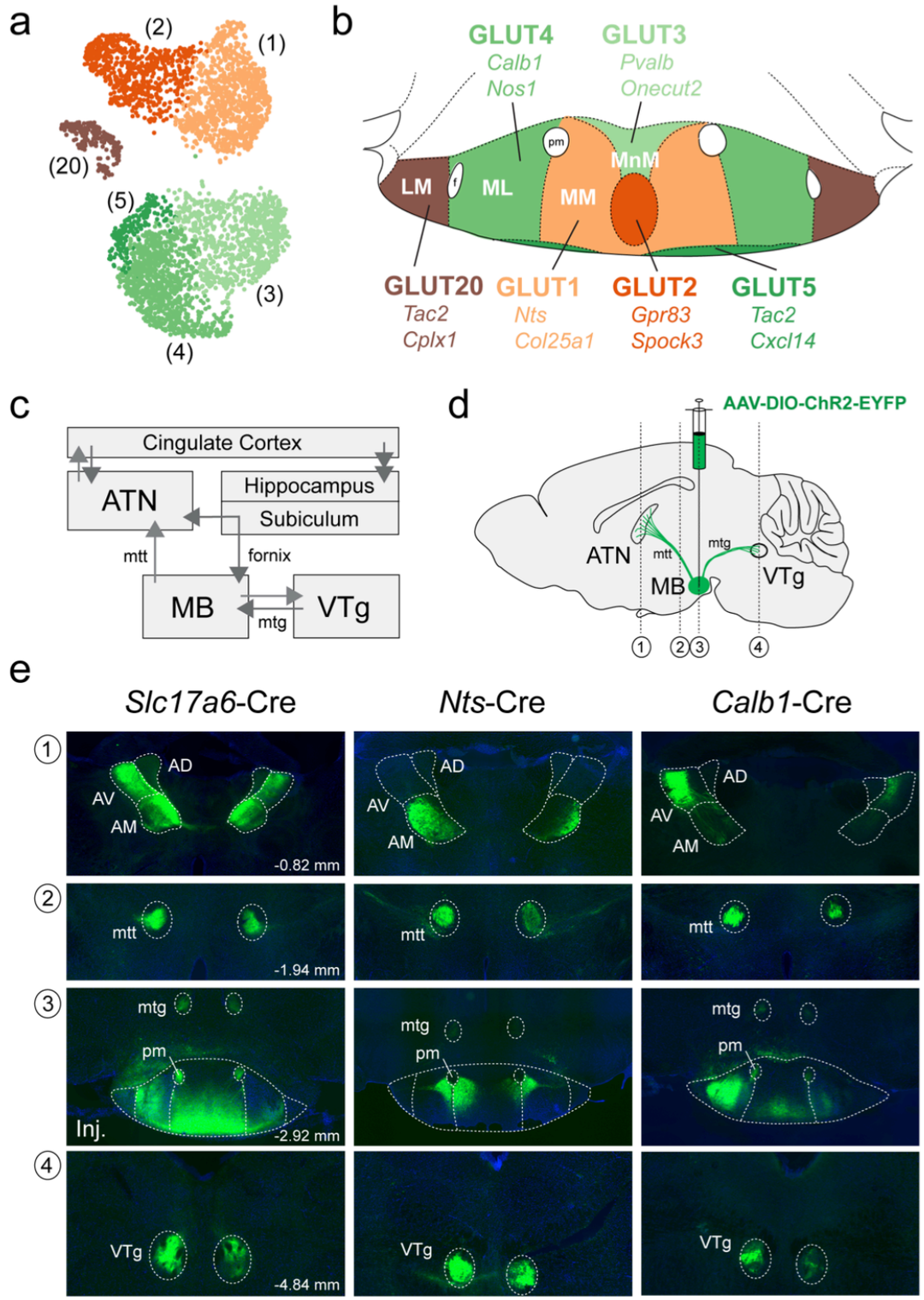


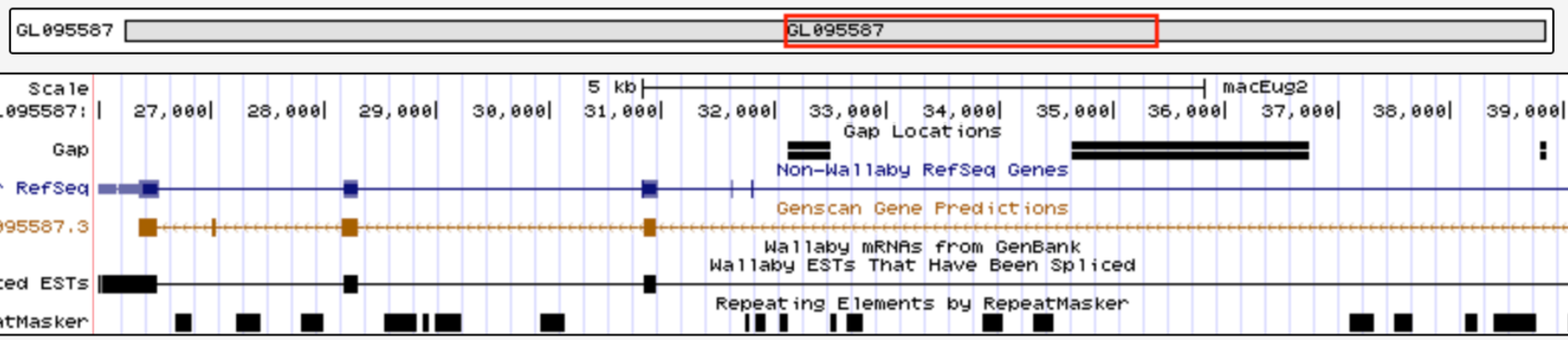
Figure 8:



UCSC Genome Browser on Wallaby Sep. 2009 (TWGS Meug_1.1/macEug2) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

GL095587:25,998-40,635 14,638 bp. enter position or search terms go



move start < 2.0 >

Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position. Press "?" for keyboard shortcuts.

move end < 2.0 >

track search | default tracks | default order | hide all | add custom tracks | track hubs | configure | multi-region | reverse | resize | refresh

collapse all

Use drop-down controls below and press refresh to alter tracks displayed.
Tracks with lots of items will automatically be displayed in more compact modes.

expand all

Mapping and Sequencing

refresh

Base Position	Assembly	Gap	GC Percent	INSDC	Restr Enzymes
dense	hide	dense	hide	hide	hide

Short Match: hide

Genes and Gene Predictions

refresh

Other RefSeq	AUGUSTUS	Genscan Genes	TransMap V5...
dense	hide	pack	hide

mRNA and EST

refresh

Wallaby mRNAs	Spliced ESTs	Wallaby ESTs
pack	dense	hide

Comparative Genomics

refresh

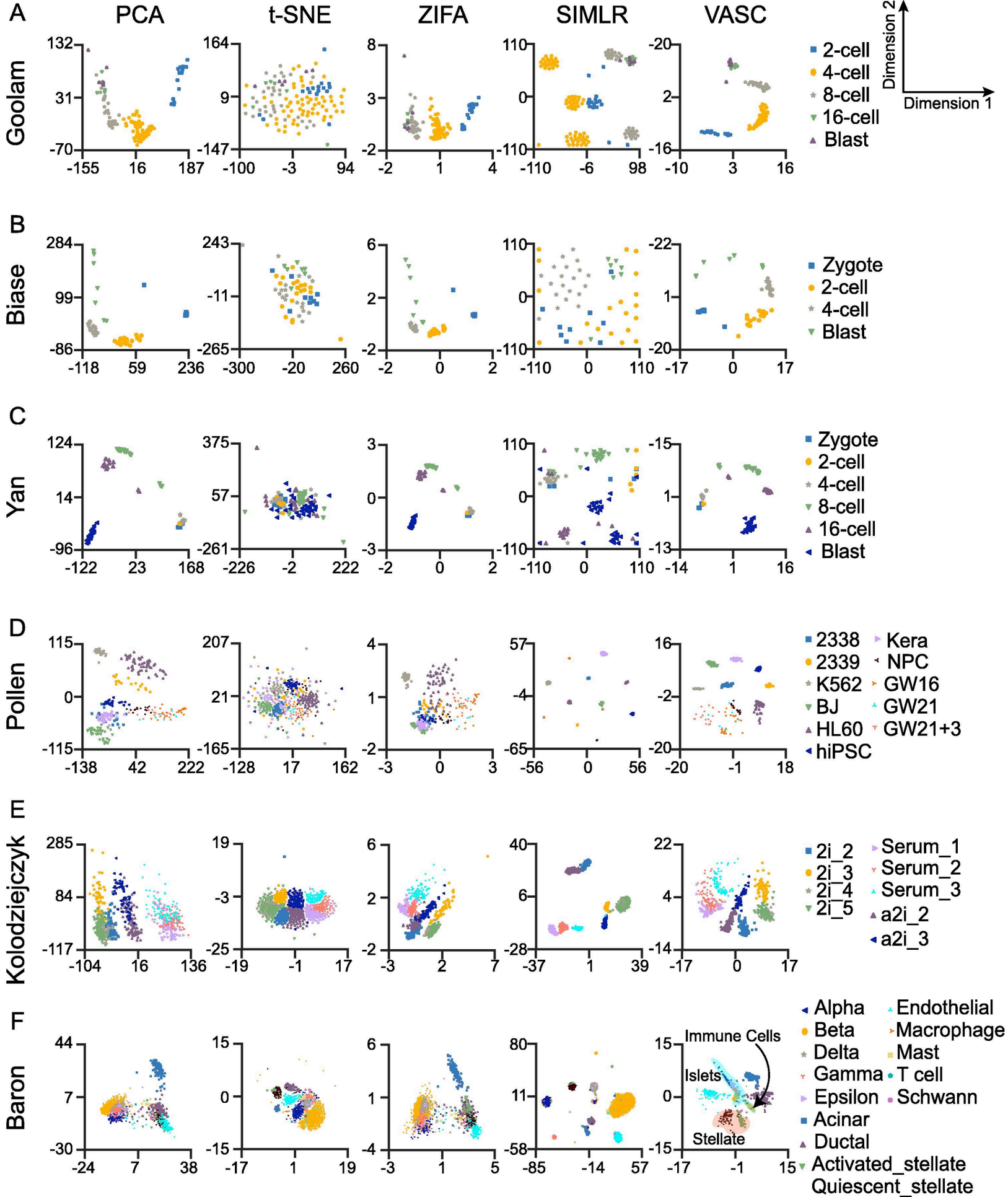
Mouse Chain/Net: hide

Variation and Repeats

refresh

RepeatMasker	Interrupted Rpts	Microsatellite	Simple Repeats	WM + SDust
dense	hide	hide	hide	hide

refresh



Genomics, Proteomics & Bioinformatics
Volume 16, Issue 5, October 2018, Pages 320-331

Method

VASC: Dimension Reduction and Visualization of Single-cell RNA-seq Data by Deep Variational Autoencoder

Dongfang Wang [a], Jin Gu [b] ✉

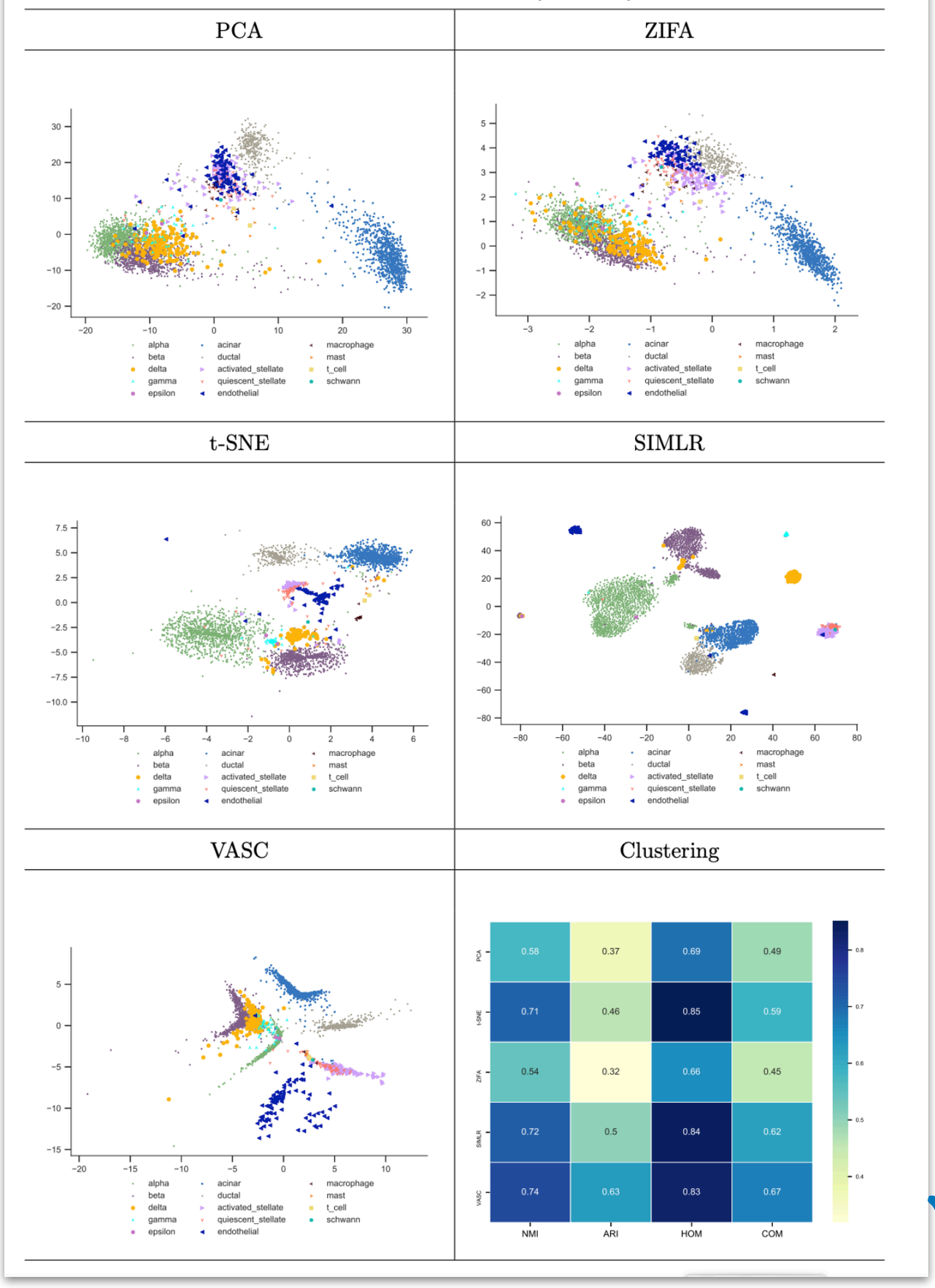
MOE Key Laboratory of Bioinformatics, BNRIST Bioinformatics Division & Center for Synthetic and Systems Biology, Department of Automation, Tsinghua University, Beijing 100084, China

Received 23 March 2018, Revised 9 July 2018, Accepted 8 August 2018, Available online 18 December 2018.

Figure 2. **Visualization of scRNA-seq datasets using different methods**

Each data point represents a cell. Different cell types are indicated in different colors and shapes. All datasets were run by PCA, t-SNE, ZIFA, SIMLR, and VASC respectively. Cell type information was retrieved from original studies. Shown in the figures are clustering output from the Goolam [22] (**A**), Biase [18] (**B**), Yan [30] (**C**), Pollen [27] (**D**), Kolodziejczyk [24] (**E**), and Baron_human-1 [17] (**F**) datasets. Visualization of other datasets is provided in the Section 4 of File S1. PCA, principal components analysis; t-SNE, t-distributed stochastic neighbor embedding; ZIFA, zero-inflated factor analysis; SIMLR, single-cell interpretation via multiple kernel learning.

The "Baron" dataset contains a large number of cells from human and mouse pancreas. Totally, there are 4 human donors with 1937, 1724, 3605, and 1303 cells, and 2 mice with 822 and 1064 cells, respectively



HOSTED BY
ELSEVIER
Genomics, Proteomics & Bioinformatics
Volume 16, Issue 5, October 2018, Pages 320-331

Method

VASC: Dimension Reduction and Visualization of Single-cell RNA-seq Data by Deep Variational Autoencoder

Dongfang Wang [a], Jin Gu [b] ✉

MOE Key Laboratory of Bioinformatics, BNRIST Bioinformatics Division & Center for Synthetic and Systems Biology, Department of Automation, Tsinghua University, Beijing 100084, China

Received 23 March 2018, Revised 9 July 2018, Accepted 8 August 2018, Available online 18 December 2018.

Figure 2. **Visualization of scRNA-seq datasets using different methods**

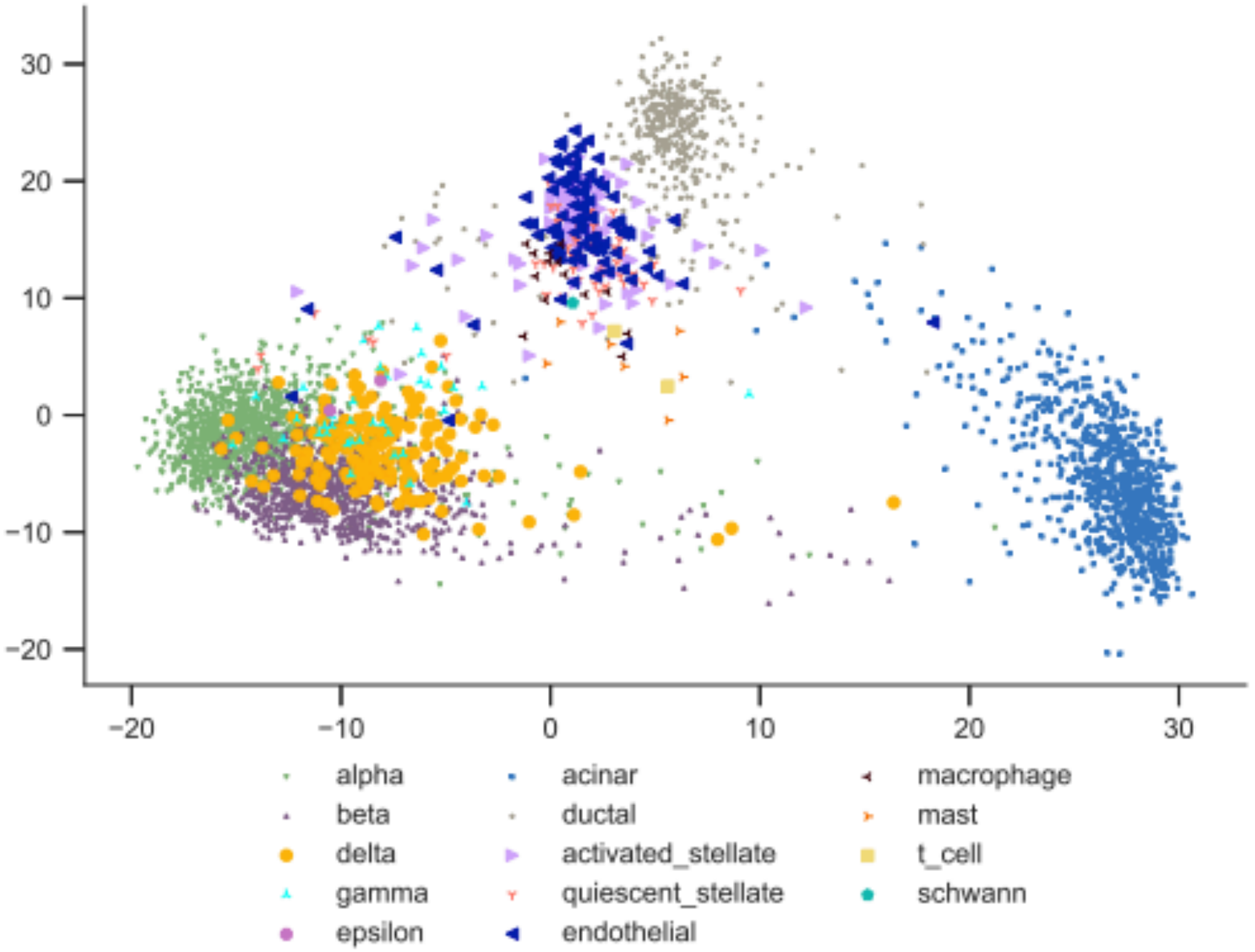
Each data point represents a cell. Different cell types are indicated in different colors and shapes. All datasets were run by PCA, t-SNE, ZIFA, SIMLR, and VASC respectively. Cell type information was retrieved from original studies. Shown in the figures are clustering output from the Goolam [22] (**A**), Biase [18] (**B**), Yan [30] (**C**), Pollen [27] (**D**), Kolodziejczyk [24] (**E**), and Baron_human-1 [17] (**F**) datasets. Visualization of other datasets is provided in the Section 4 of File S. PCA, principal components analysis; t-SNE, t-distributed stochastic neighbor embedding; ZIFA, zero-inflated factor analysis; SIMLR, single-cell interpretation via multiple kernel learning.

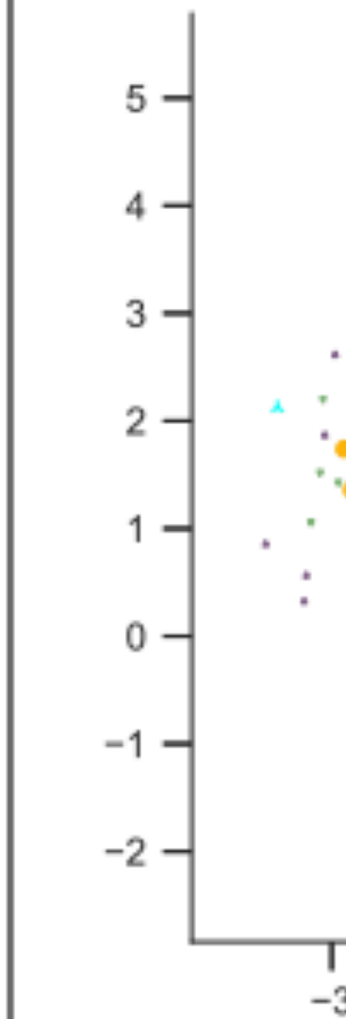
The “Baron” dataset contains a large number of cells from human and mouse pancreas. Totally, there are 4 human donors with 1937, 1724, 3605, and 1303 cells, and 2 mice with 822 and 1064 cells, respectively

PCA

Tiny symbols show types of cells, color helps differentiate, as some symbol shapes are repeated.

(Symbol size not meaningful?)





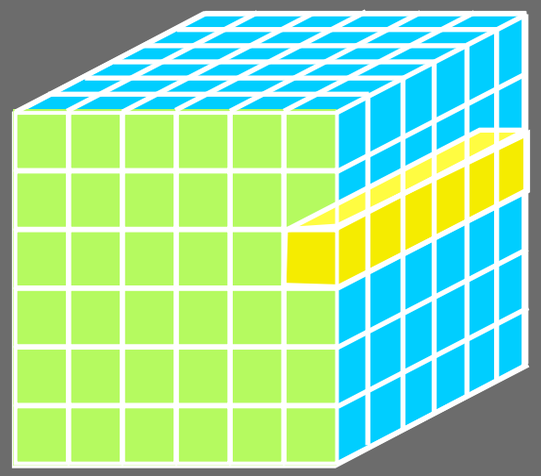
10QViz.org

PRACTICALITY

PRINCIPLES

PHILOSOPHY

glue
multidimensional data exploration



DIVERSE TOOLS

DIVERSE DATA

DIVERSE VIEWS

My world & your

philosophy

The Jackson Laboratory

Search th

RESEARCH & FACULTY

EDUCATION & LEARNING

JAX MICE & SERVICES

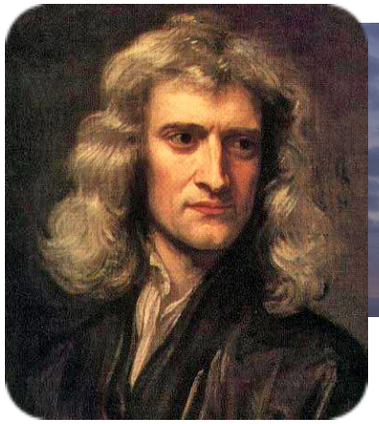
PERSONALIZED MEDICINE

MICE IN SPACE UPDATE

Researchers show that JAX Mig stayed mighty during their time microgravity.

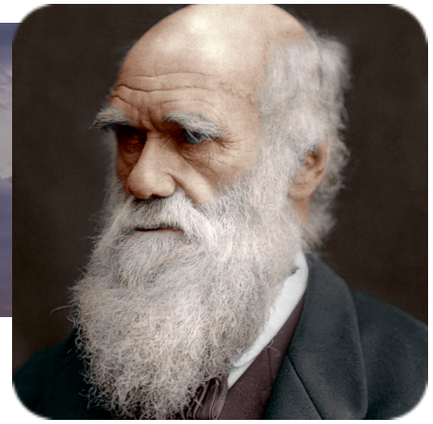
LEARN MORE +

DIVERSE TOOLS DIVERSE DATA DIVERSE VIEWS



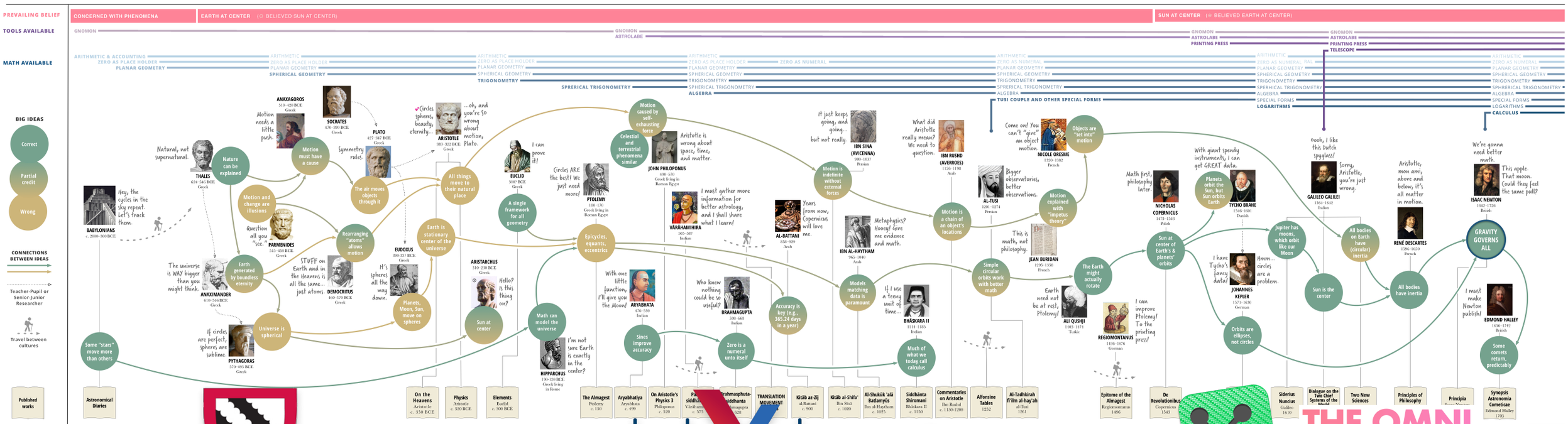
PATH TO

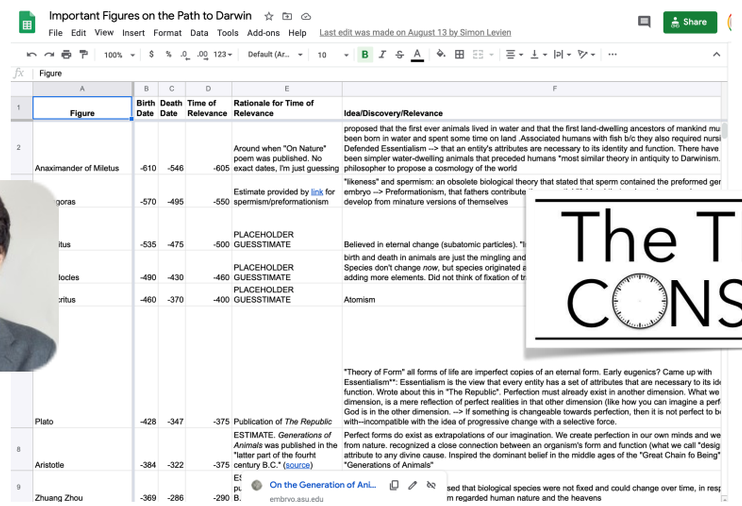
A project to track the evolution of science

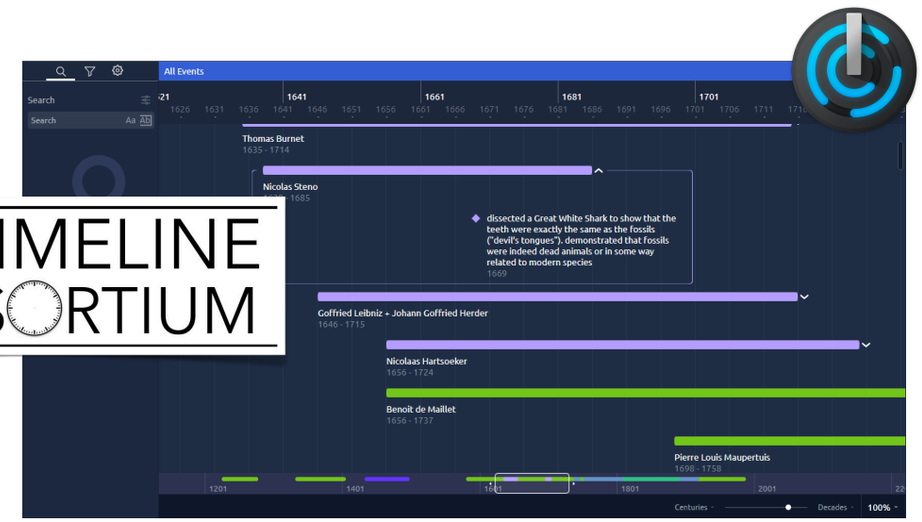


The Path to Newton

PredictionX







LabXchange™
Science and learning–connected.

The TIMELINE CONSORTIUM

THE OMNI SHOW

PredictionX

predictionx.org

PredictionX: The Past & Present of the Future

Essentials

Predictive Systems Framework

Phenomena →Predictions

Understanding Uncertainty

Study Design

Timelines

Why predict?

Omens, Oracles & Prophecies

- *Mesopotamian Haruspicy*
- *Egyptian Priests*
- *Yoruba Ifa*
- *Tarot*
- *Roman Augury*
- *Casting Lots*
- **The Diviner's Guide**
- *Chinese Oracle Bones*
- *Greek Astronomy*
- *Oracle of Delphi*
- *Turkish Tasseography*
- *Astrology*
- *Aztec Rituals*
- *Maya Spacetime*
- *Comets of Doom*

cross-cultural conversations

The Rise of Theory

- *Ancient Mesopotamia, Egypt, Greece & Rome*
- *Islamic Science*

The Path to Newton

- *Indian Mathematics*
- *European Renaissance*

The Royal Society

Lost without Longitude (Navigation)

- *Help, I'm Lost!*
- *Tools of the Navigator*

Modern Simulation

John Snow & Cholera

- *Cholera Map*

Health

- *Epidemiology*
- *Personal Genomics*
- *Population Genetics*

Wealth

- *Climate & Wealth*
- *Behavioral Economics*

The Future of the Future

- *AI, Derek's Day*
- *Philosophy*
- *Uncertainty*

Earth

- *Climate & Energy*
- *Climate Policy*
- *Tent Tarot*
- *Earthquakes*

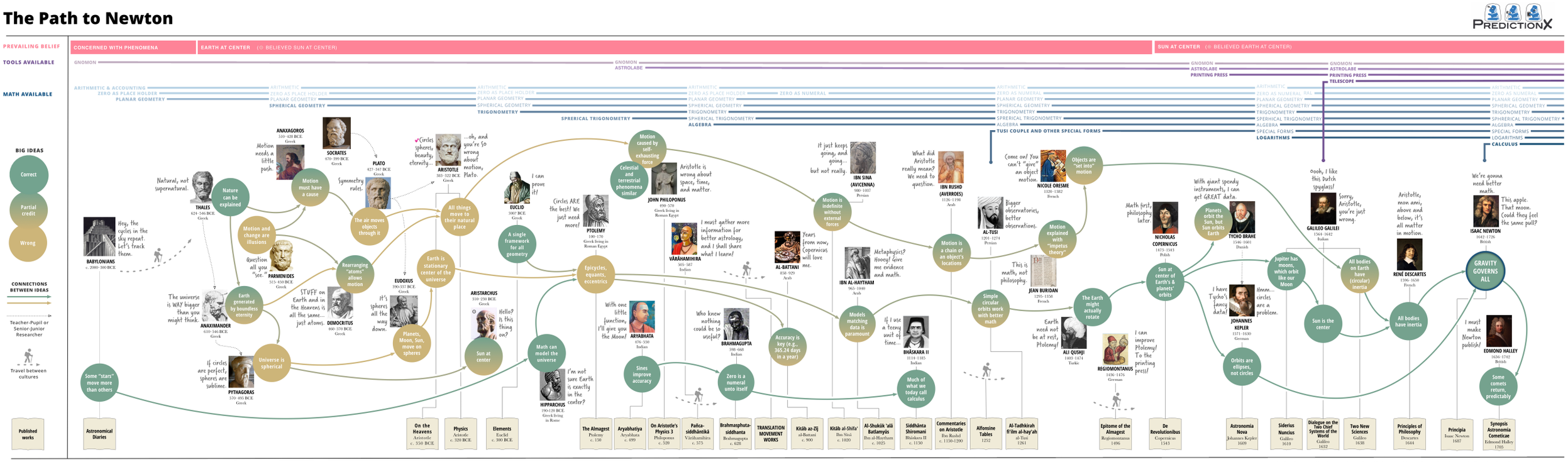
Space

- *Futures of our Universe*
- *SETI*

Coming Soon · Interactive Resource · ▶ video(s) · edX on edX as mini-course, otherwise find material via predictionx.org

THE "PADUA" RAINBOW

Phenomenon	Observation*	Data	Rule	Theory	Explanation	Prediction



or, Experiment

Mendel

Phenomenon | Observation | Data | Rule | Prediction

Darwin

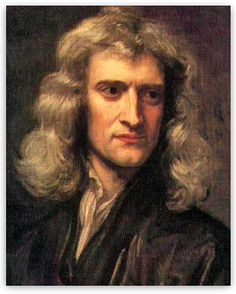
Phenomenon | Observation | Data | Theory | Explanation | Prediction

BIOLOGY

PHYSICS

Kepler

Phenomenon | Observation | Data | Rule | Prediction



Newton

Phenomenon | Observation | Data | Rule | Theory | Explanation | Prediction

Mendel *1865*

Phenomenon	Obser	Data	Rule	Prediction

Darwin *18*

Ph	Observation	Data	Theory	Explanation	Prediction

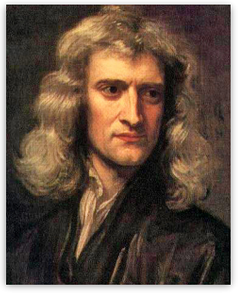
"No Predictive Biology" —Ed Liu, 1 hour ago

NO FULLY PREDICTIVE GENERAL THEORY

LLY PREDICTIVE GENERAL THOERY

Kepler *1609*

Phenomenon	Observation	Data	Rule	Prediction



Newton *1687*

Phenomenon	Observation	Data	Rule	Theory	Explanation	Prediction

THE FUTURE OF THE FUTURE

20th century

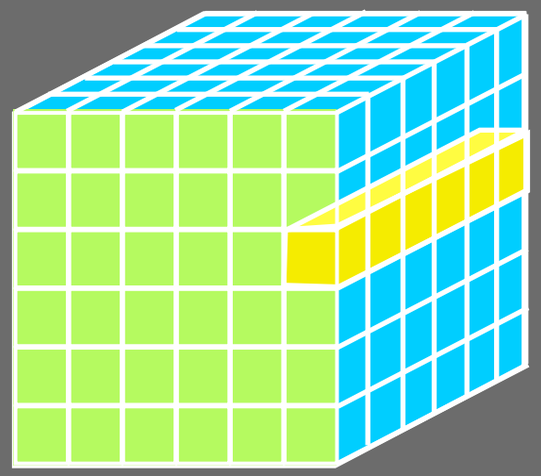
21st century?

PRACTICALITY

PRINCIPLES

PHILOSOPHY

glue
multidimensional data exploration



DIVERSE TOOLS

DIVERSE DATA

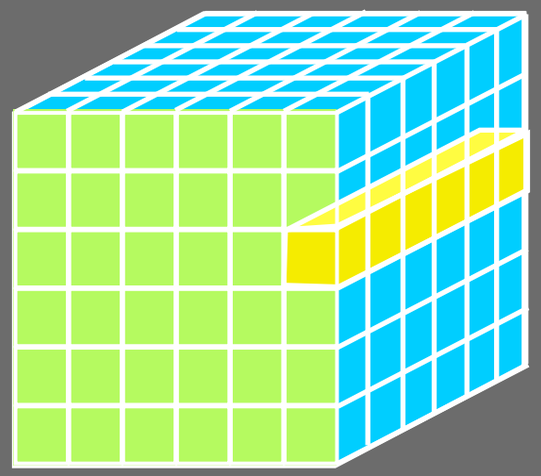
DIVERSE VIEWS

THE FUTURE OF PUBLISHING

PRACTICALITY

glue
multidimensional data exploration

PRINCIPLES

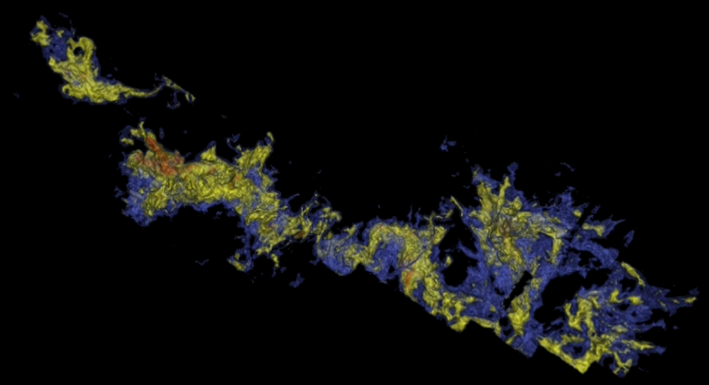


PHILOSOPHY

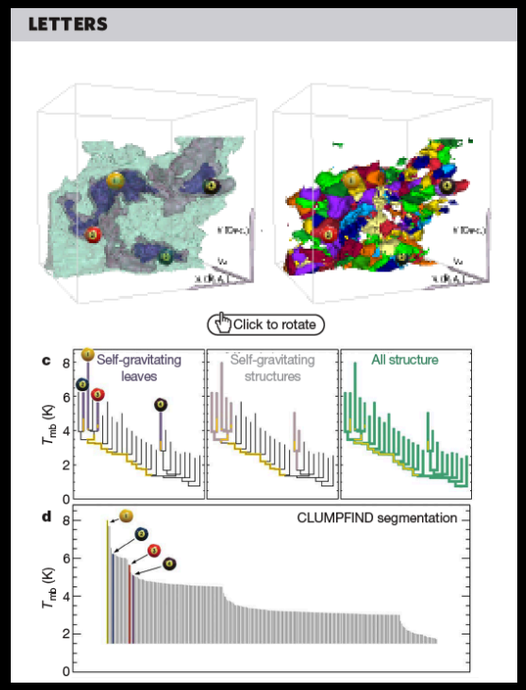
DIVERSE TOOLS | DIVERSE DATA | DIVERSE VIEWS

"Perseus" progress

2008



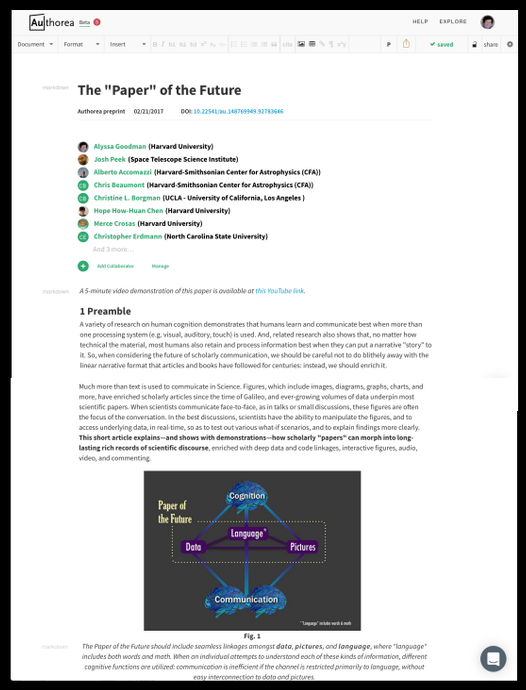
2009



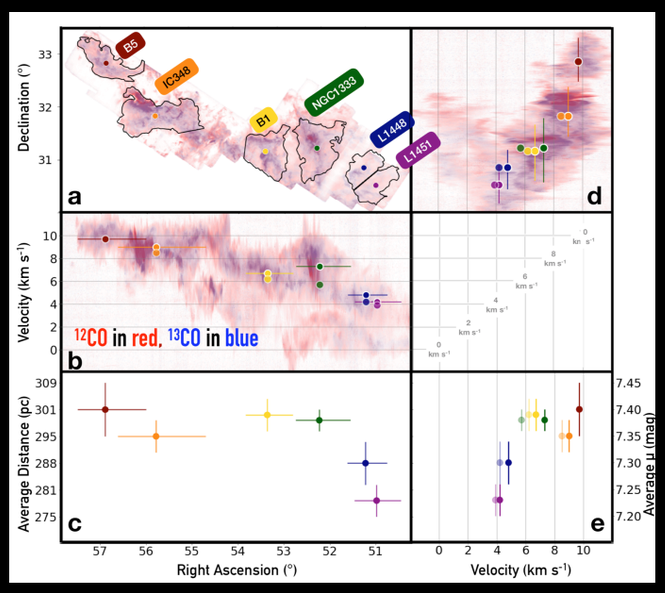
2012

glue
multidimensional data exploration

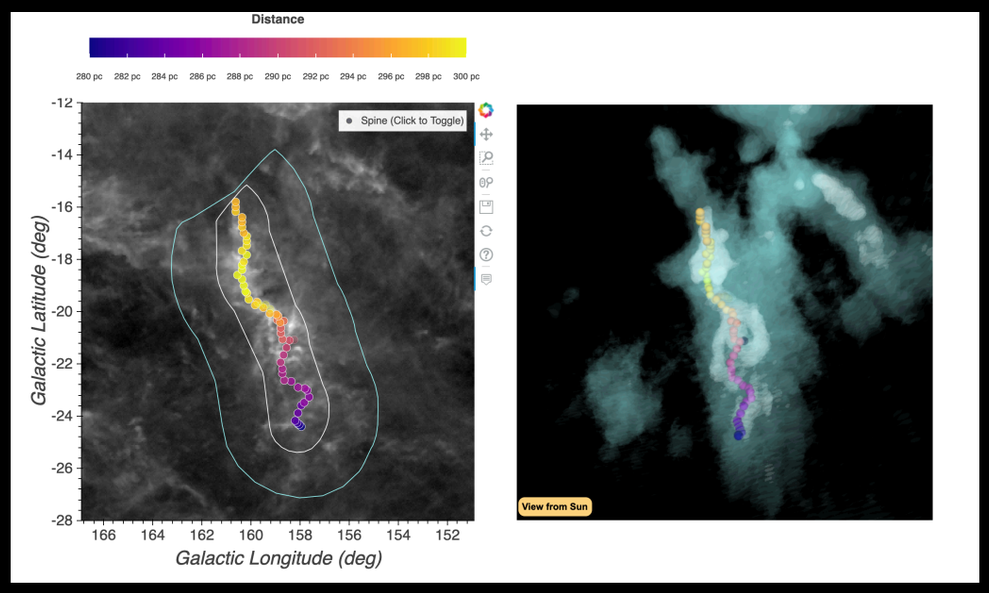
2015



2018



2020

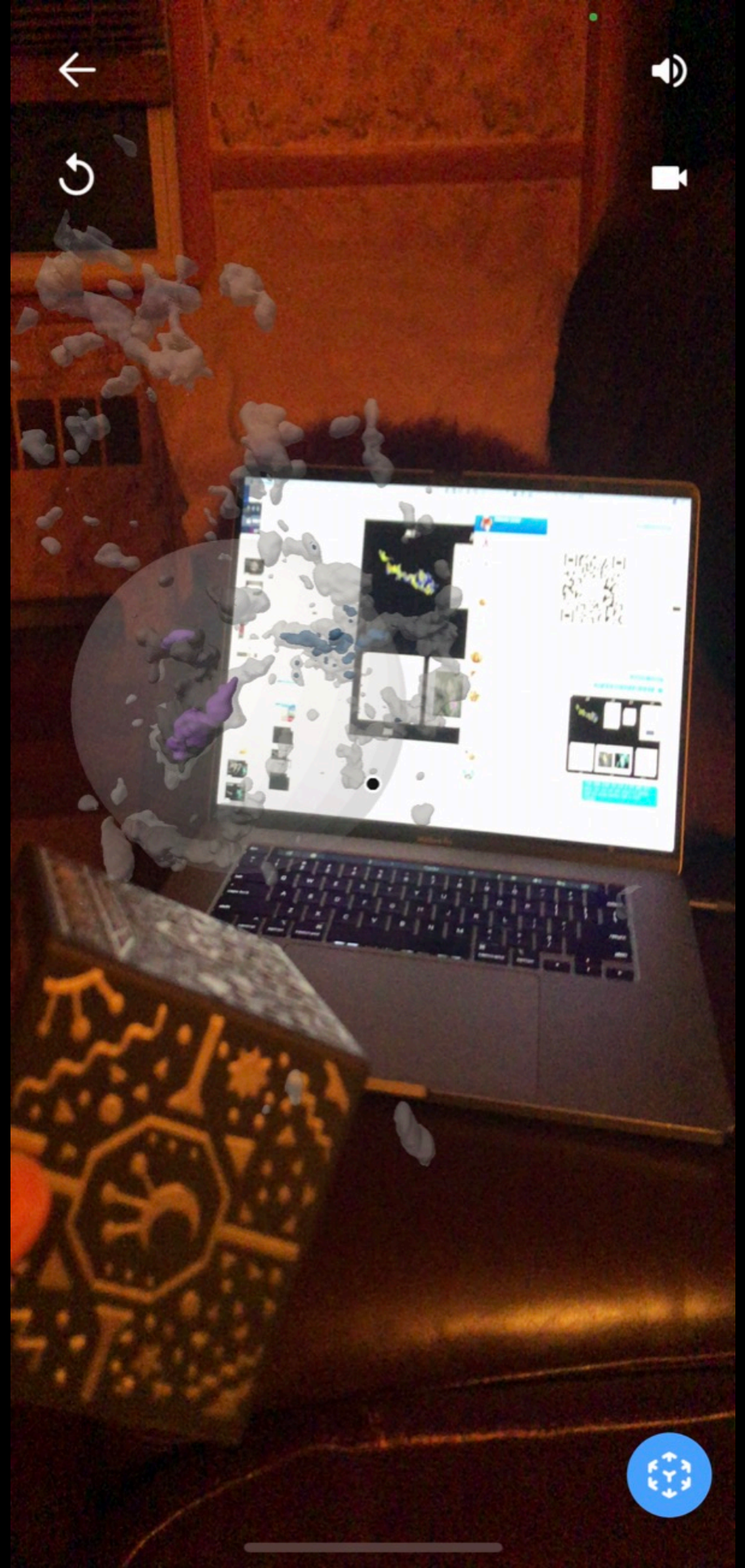


2020+

VBB
PZL

THE VERY NEAR FUTURE

VBB
PZL



PERSEUS IN ACTUAL 3D

Start the presentation to see live content. Still no live content? Install the app or get help at **PollEv.com/app**

10QViz.org

MENU

TEN QUESTIONS TO ASK WHEN CREATING A VISUALIZATION

The 10 Questions

1. **Who** | Who is your audience? How expert will they be about the subject and/or display conventions?
2. **Explore-Explain** | Is your goal to explore, document, or explain your data or ideas, or a combination of these?
3. **Categories** |Do you want to show or explore pre-existing, known, human-interpretable, categories?
4. **Patterns** |Do you want to identify new, previously unknown or undefined patterns?
5. **Predictions & Uncertainty** | Are you making a comparison between data and/or predictions? Is representing uncertainty a concern?
6. **Dimensions** | What is the intrinsic number of dimensions (not necessarily spatial) in your data, and how many do you want to show at once?
7. **Abstraction & Accuracy** | Do you need to show all the data, or is summary or abstraction OK?
8. **Context & Scale** | Can you, and do you want to, put the data into a standard frame of reference, coordinate system, or show scale(s)?
9. **Metadata** | Do you need to display or link to non-quantitative metadata? (including captions, labels, etc.)
10. **Display Modes** | What display modes might be used in experiencing your display?

Join the 10QViz Conversation!

*To learn more about this site, please visit the **About** page.*

To read an in-process manuscript giving the scholarship behind the recommendations on this site, see Coltekin & Goodman 2018.

glueviz.org

glue
solutions
inc.

gluesolutions.io

predictionx.org

10QViz.org