Alexander Lex
@alexander_lex
http://alexander-lex.net

THE UNIVERSITY OF UTAH

Visualization in Oncological Data Science Now and in the Future: From Cancer Cell Microscopy to Reproducible Visual Analysis.

visualization design lab

SCI
www.sci.utah.edu

visualization

The purpose of computing is insight,

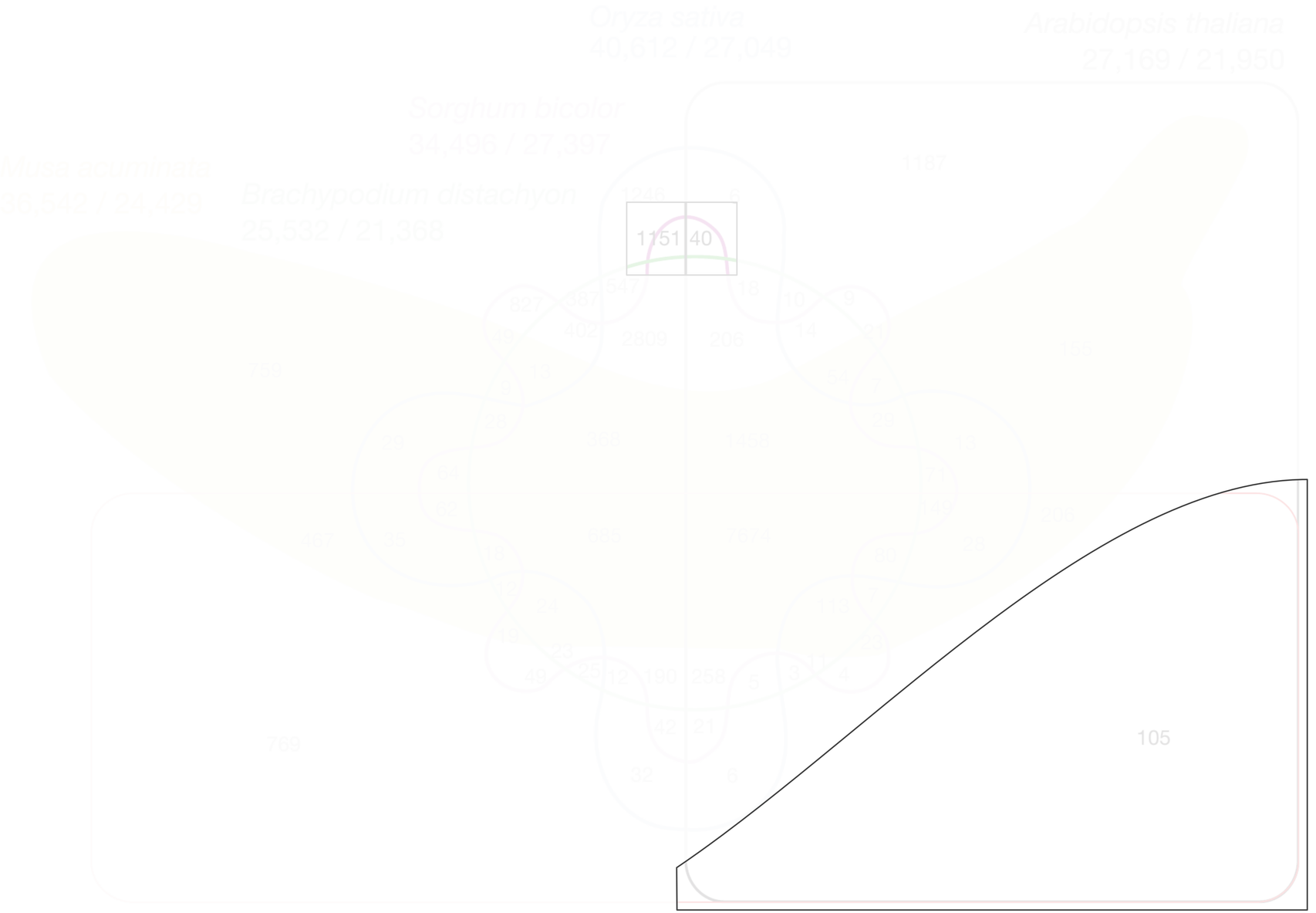
not numbers.

pictures

[Card, Mackinlay, Shneiderman]

[Richard Wesley Hamming]

Banana	*M. acuminata*
Date	*P. dactylifera*
Cress	*Arabidopsis thaliana*
Rice	*Oryza sativa*
Sorghum	*Sorghum bicolor*
Brome	*Brachypodium distachyon*



[D'Hont et al., Nature, 2012]

GOOD DATA VISUALIZATION

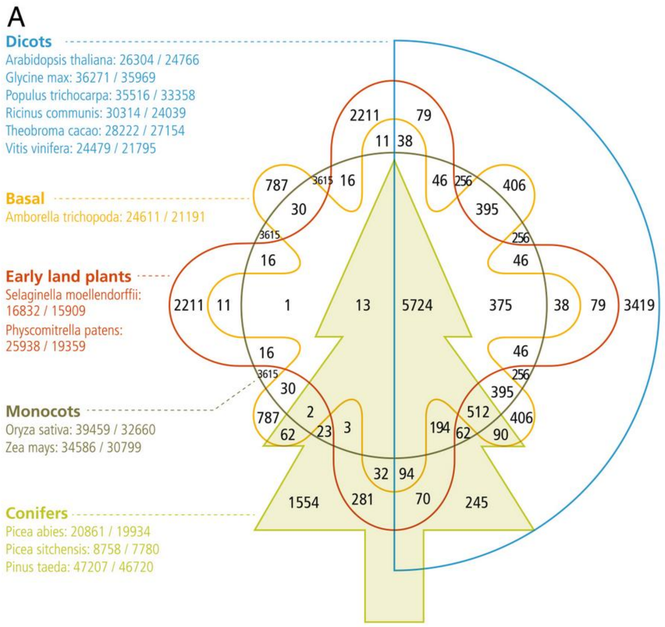
... makes data accessible

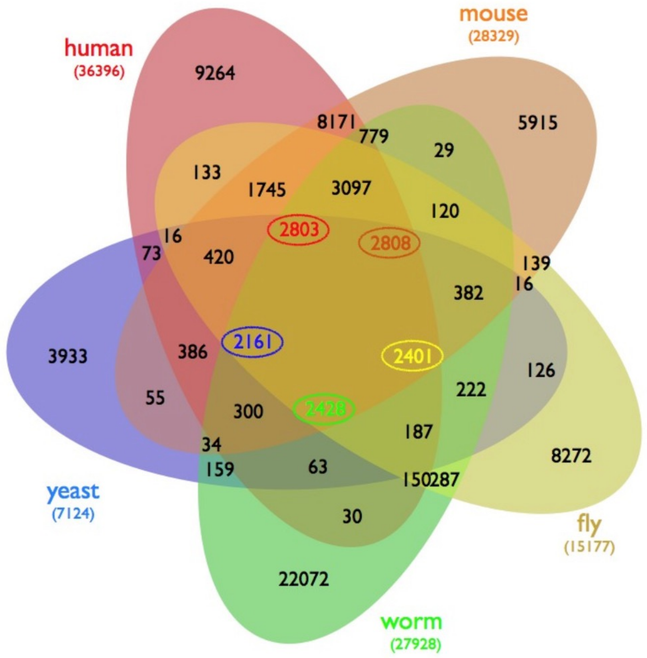
... combines strengths of humans and computers

... enables insight

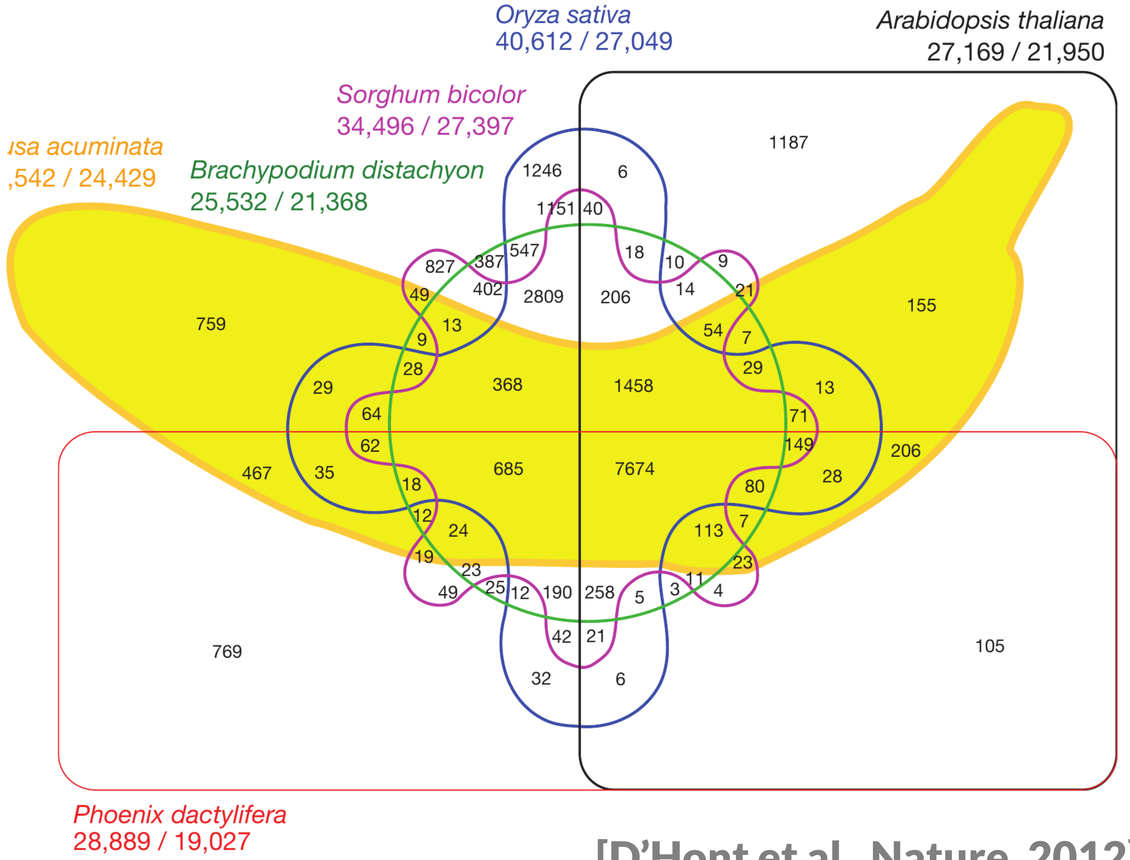
... communicates

[Neale et al., BMC Genome Biology, 2014]

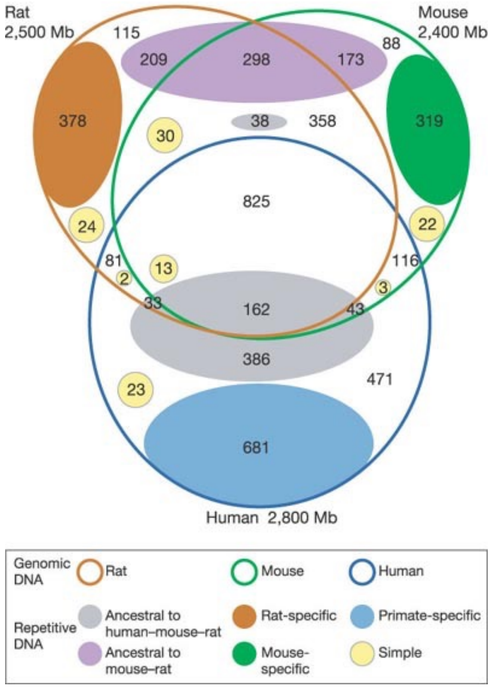




[Wiles et al., BMC Systems Biology]

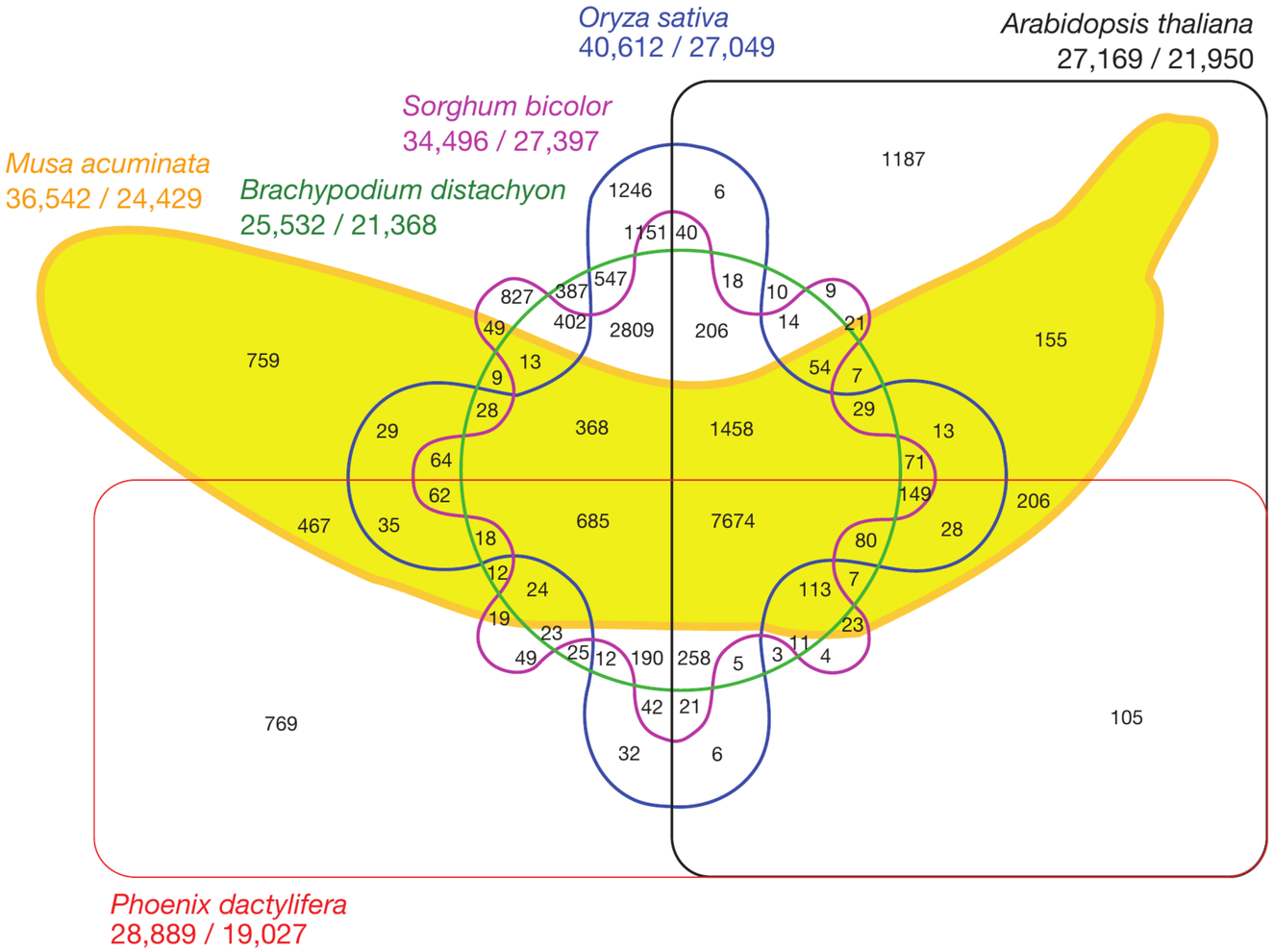


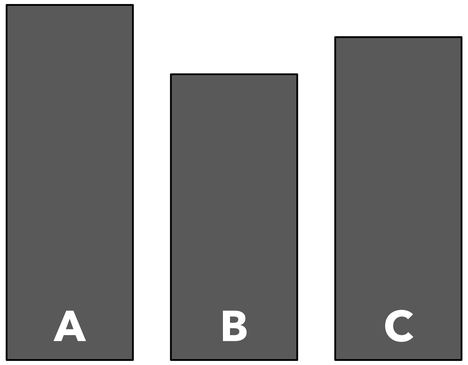
[D'Hont et al., Nature, 2012]

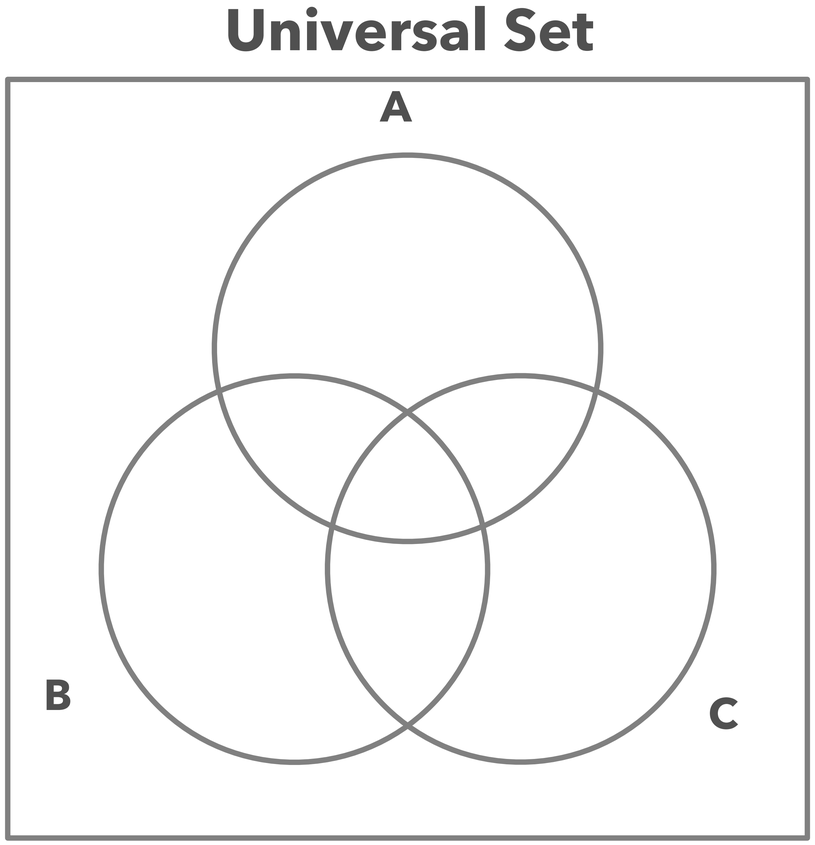


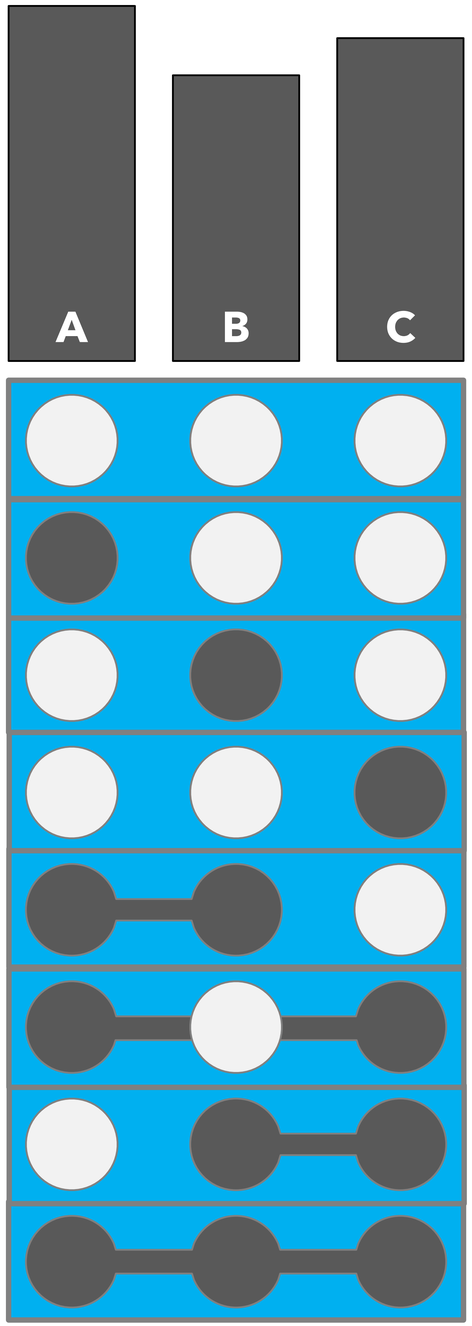
[Gibbs et al., Nature, 2004]

SO CAN WE DO BETTER?

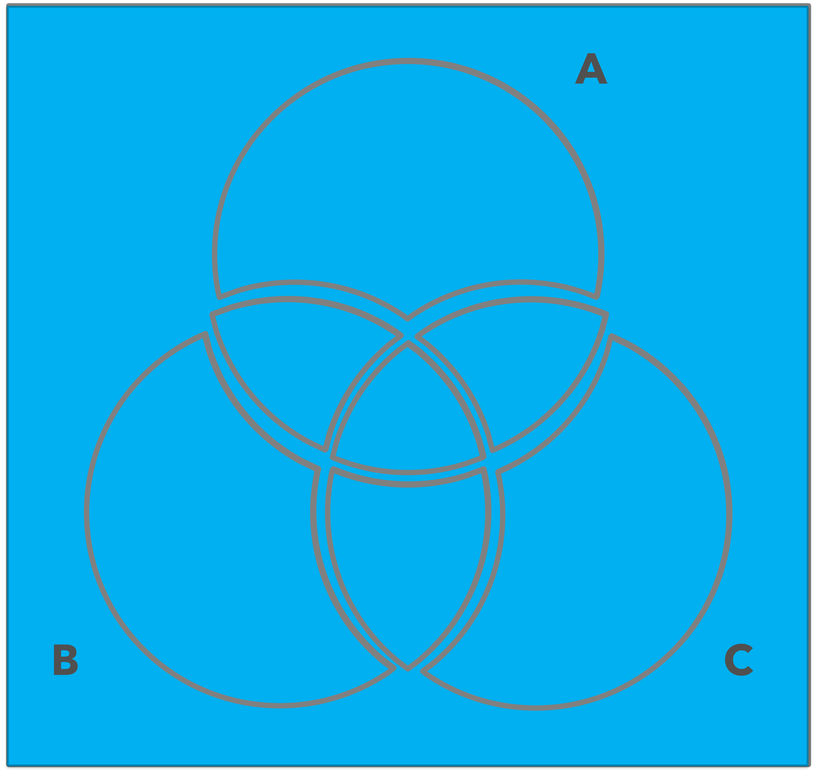


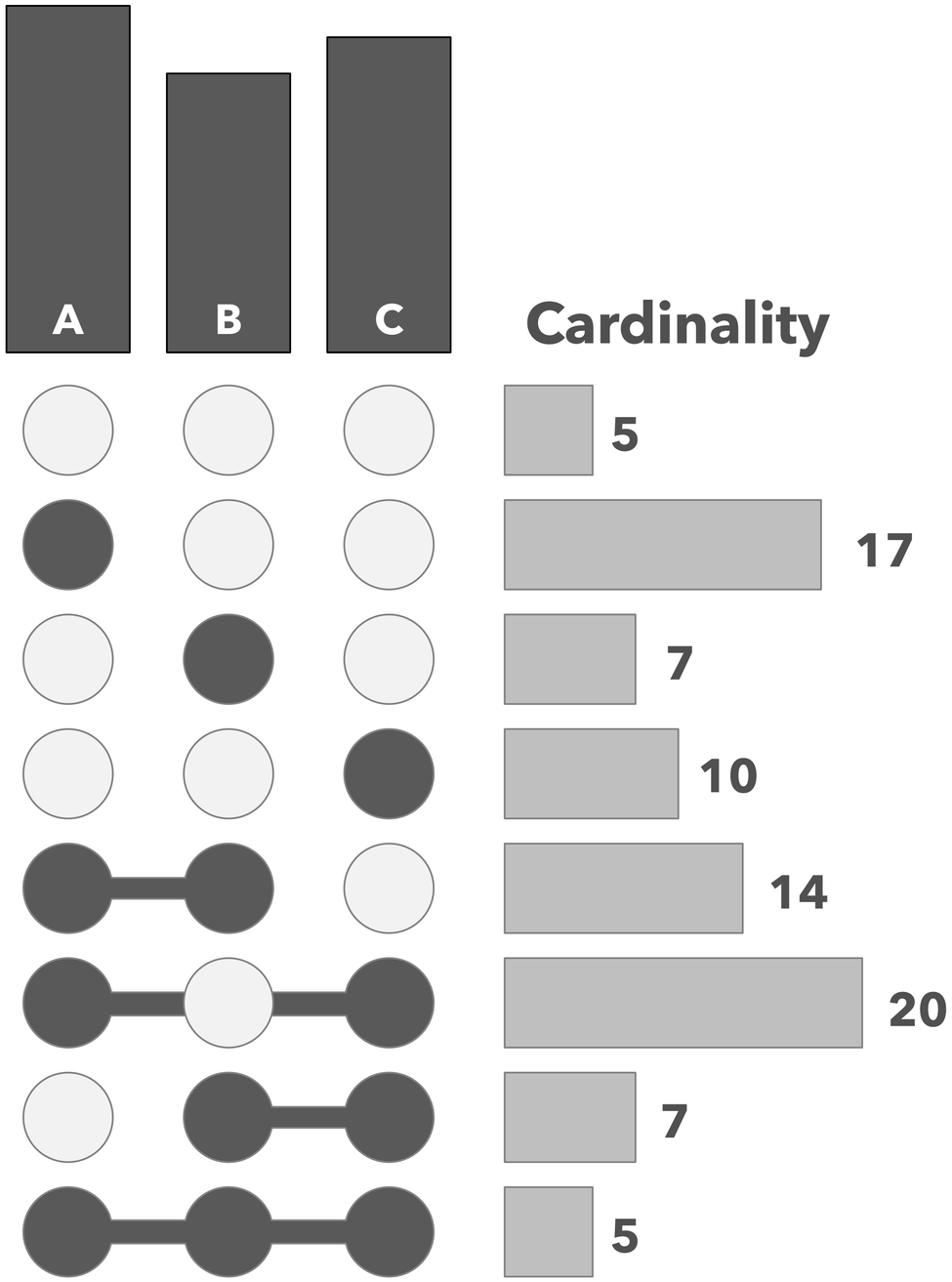

A
B
C

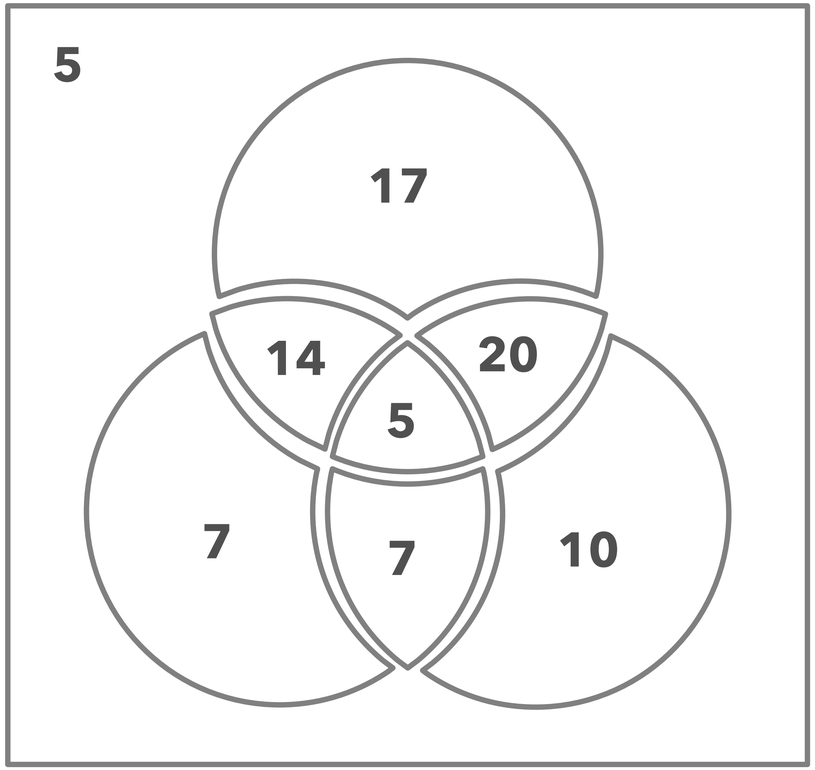

Universal Set
A
B
C



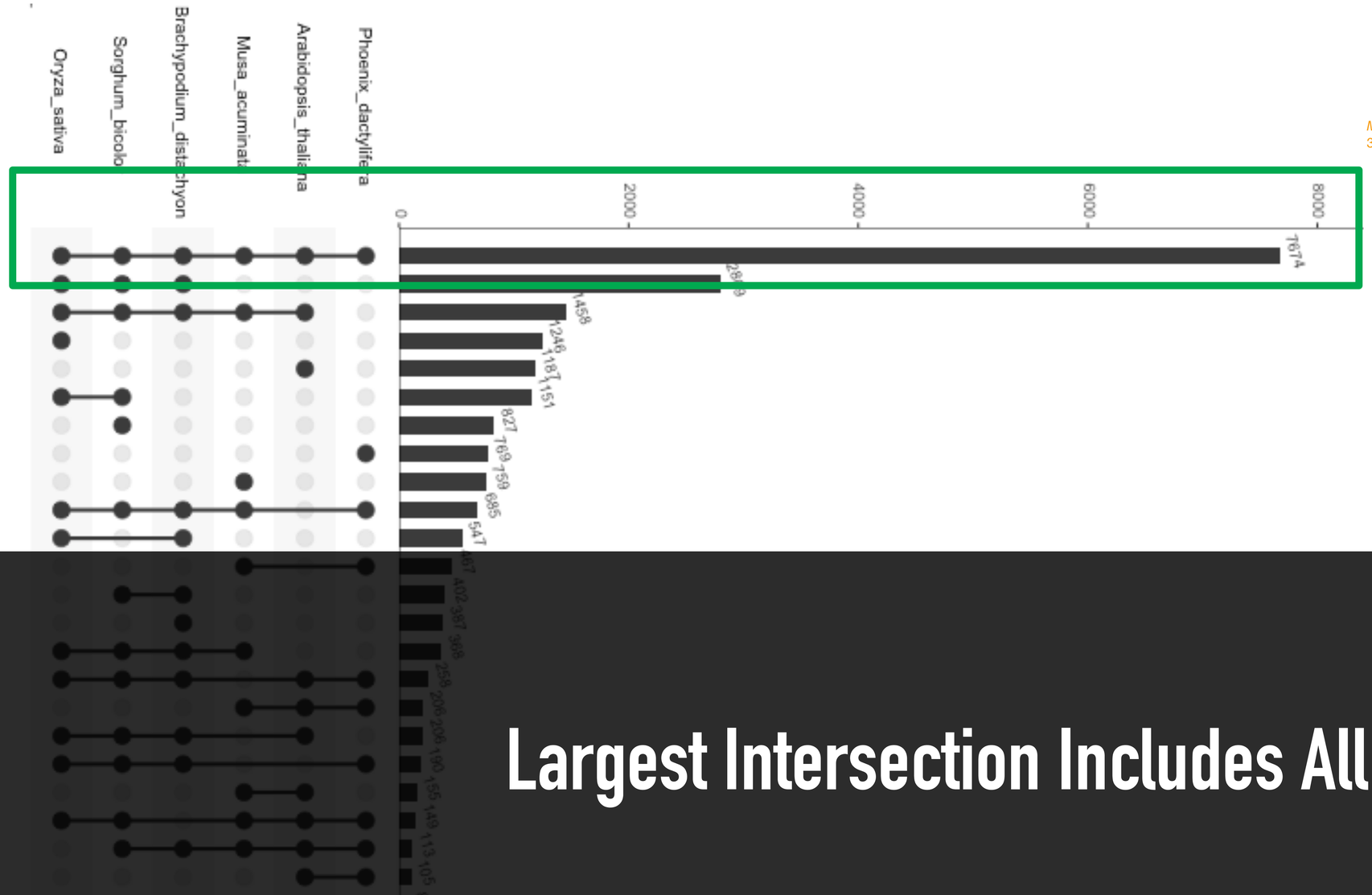
Universal Set

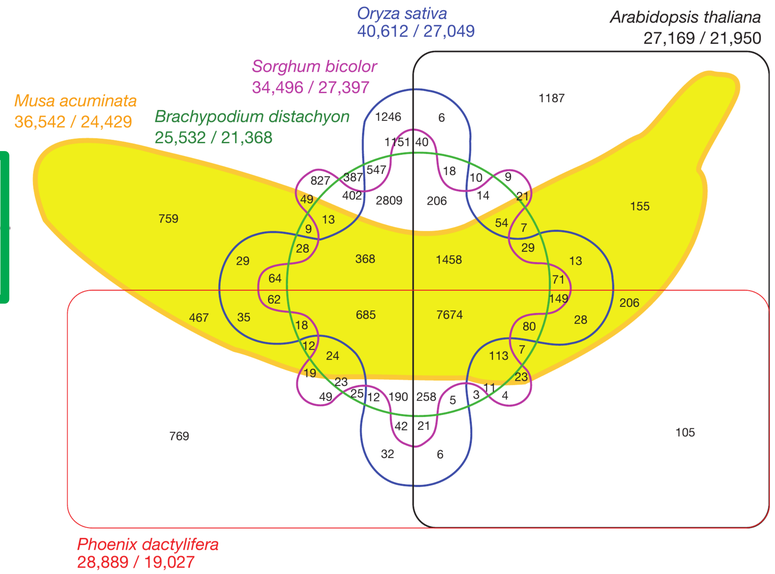



A
B
C
Cardinality
5
17
7
10
14
20
7
5


5
17
14
20
5
7
7
10

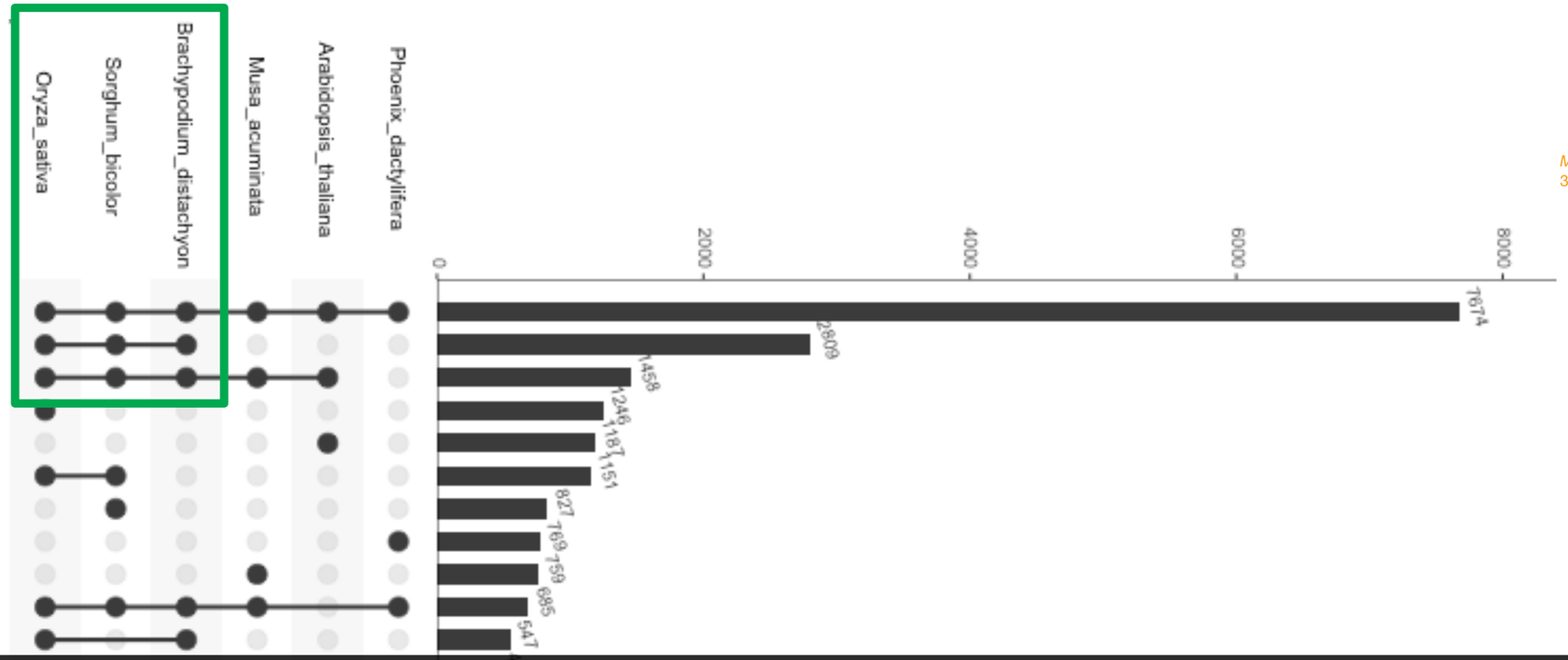
THE BANANA CHART REDESIGNED: UPSET

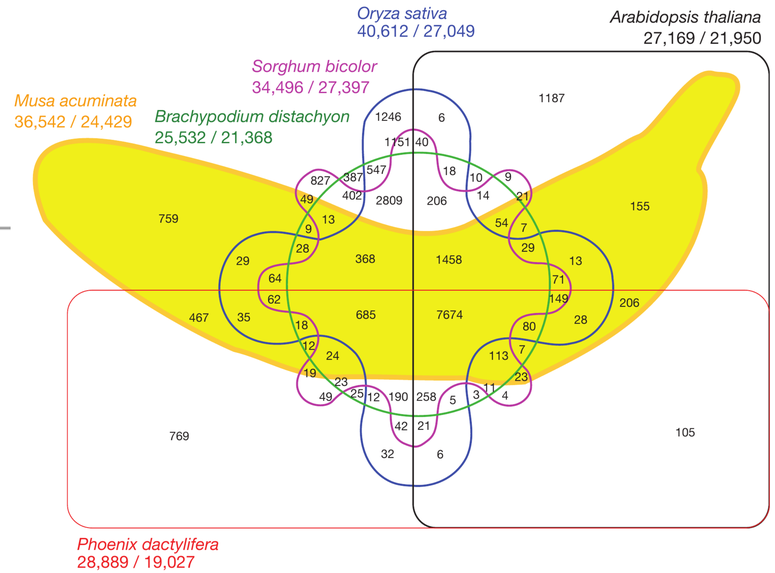




Largest Intersection Includes All Sets

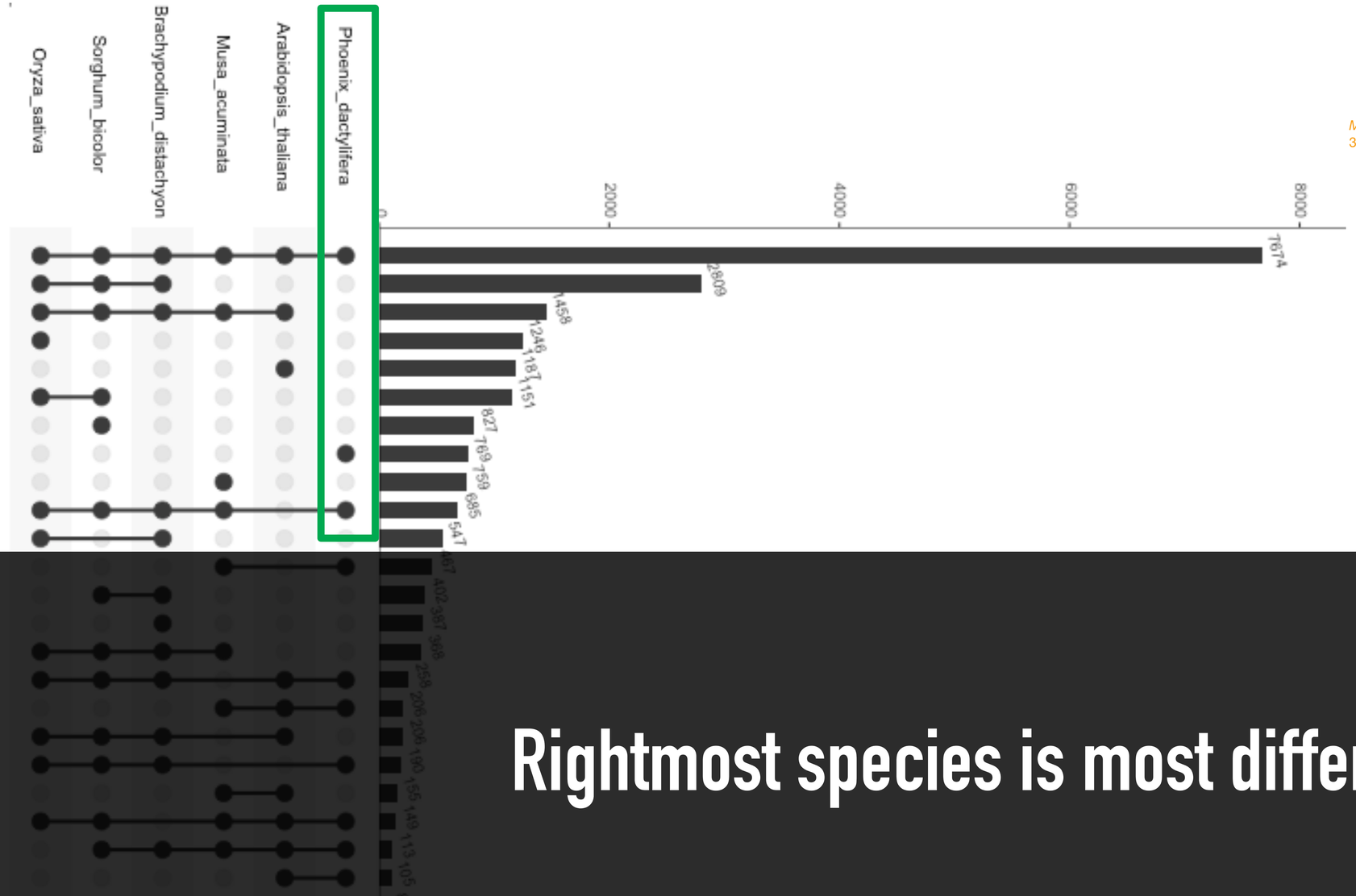
THE BANANA CHART REDESIGNED: UPSET

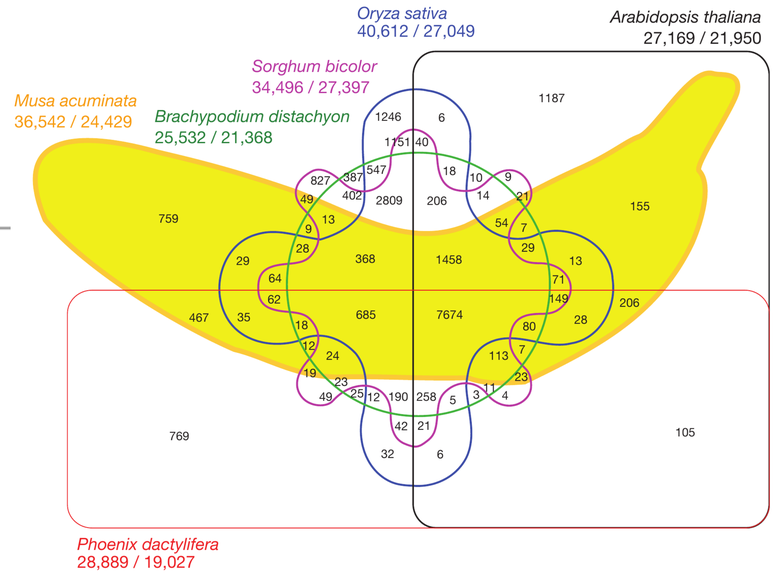




Three Leftmost Species Are Most Similar

THE BANANA CHART REDESIGNED: UPSET





Rightmost species is most different

Proper Visualization

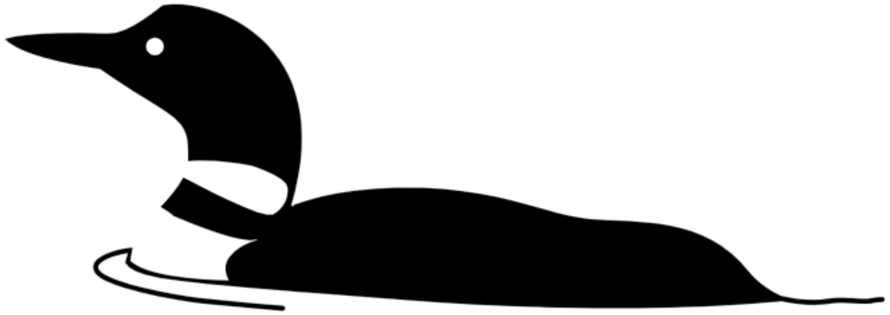
Makes a Difference

Loon

Using Exemplars to Visualize Large Scale Microscopy Data

Devin Lange, Eddie Polanco, Robert Judson-Torres, Thomas Zangle, Alexander Lex

http://loon.sci.utah.edu/



THE UNIVERSITY OF UTAH

visualization design lab

Zangle Lab
Department of Chemical Engineering
THE UNIVERSITY OF UTAH

UNIVERSITY OF UTAH
HUNTSMAN CANCER INSTITUTE

HEALTH
UNIVERSITY OF UTAH

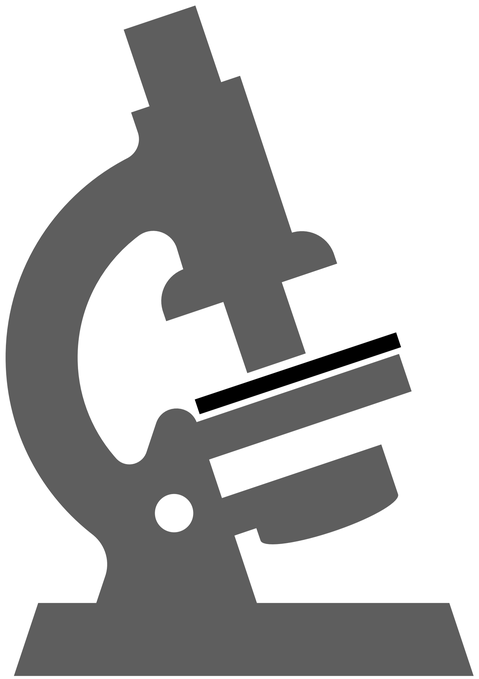
MOTIVATION: IMPROVE CANCER TREATMENT

Challenges:

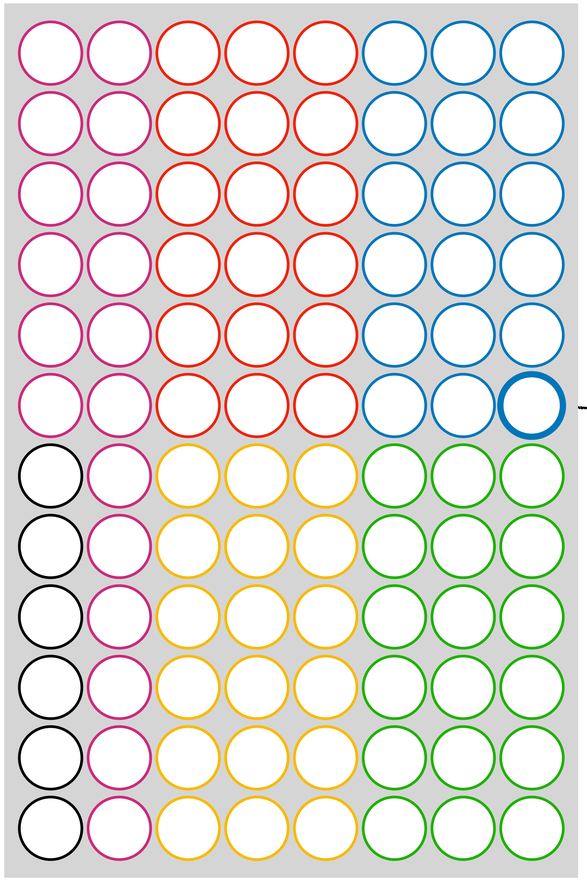
- **Cancer is heterogeneous**
- **Many treatment options exist**

Goal:

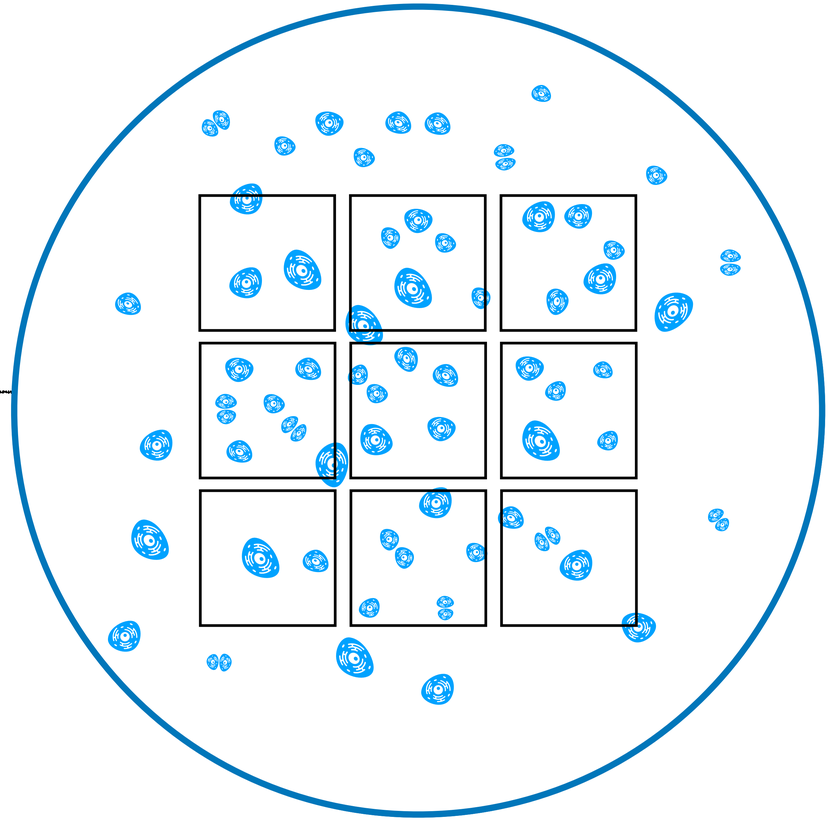
- **Identify effective treatment option for individuals through time-series microscopy**



Quantitative Phase Microscope

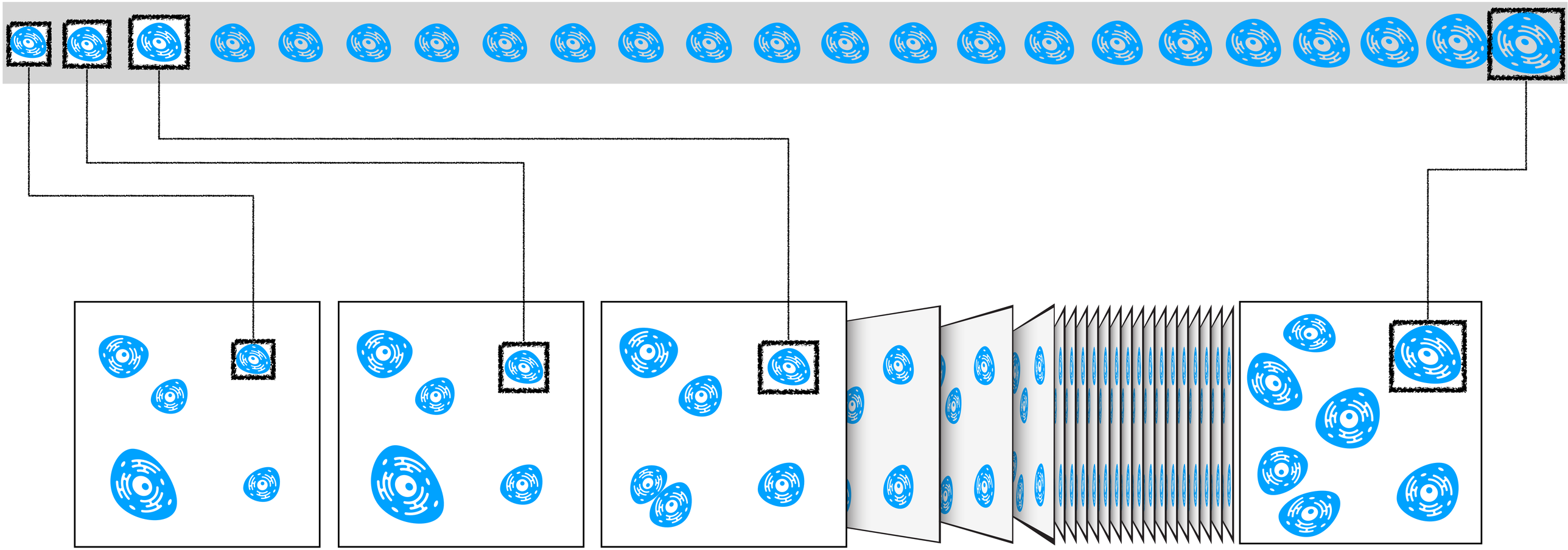


96-Well Plate

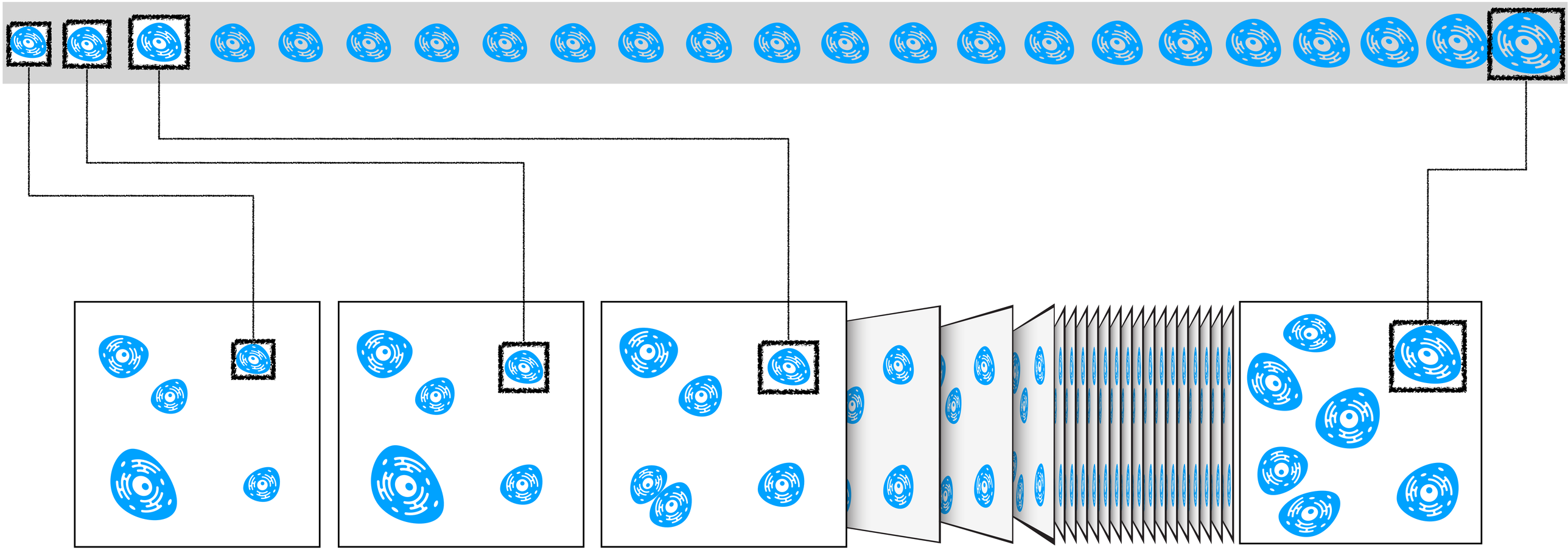


Imaging a Single Well

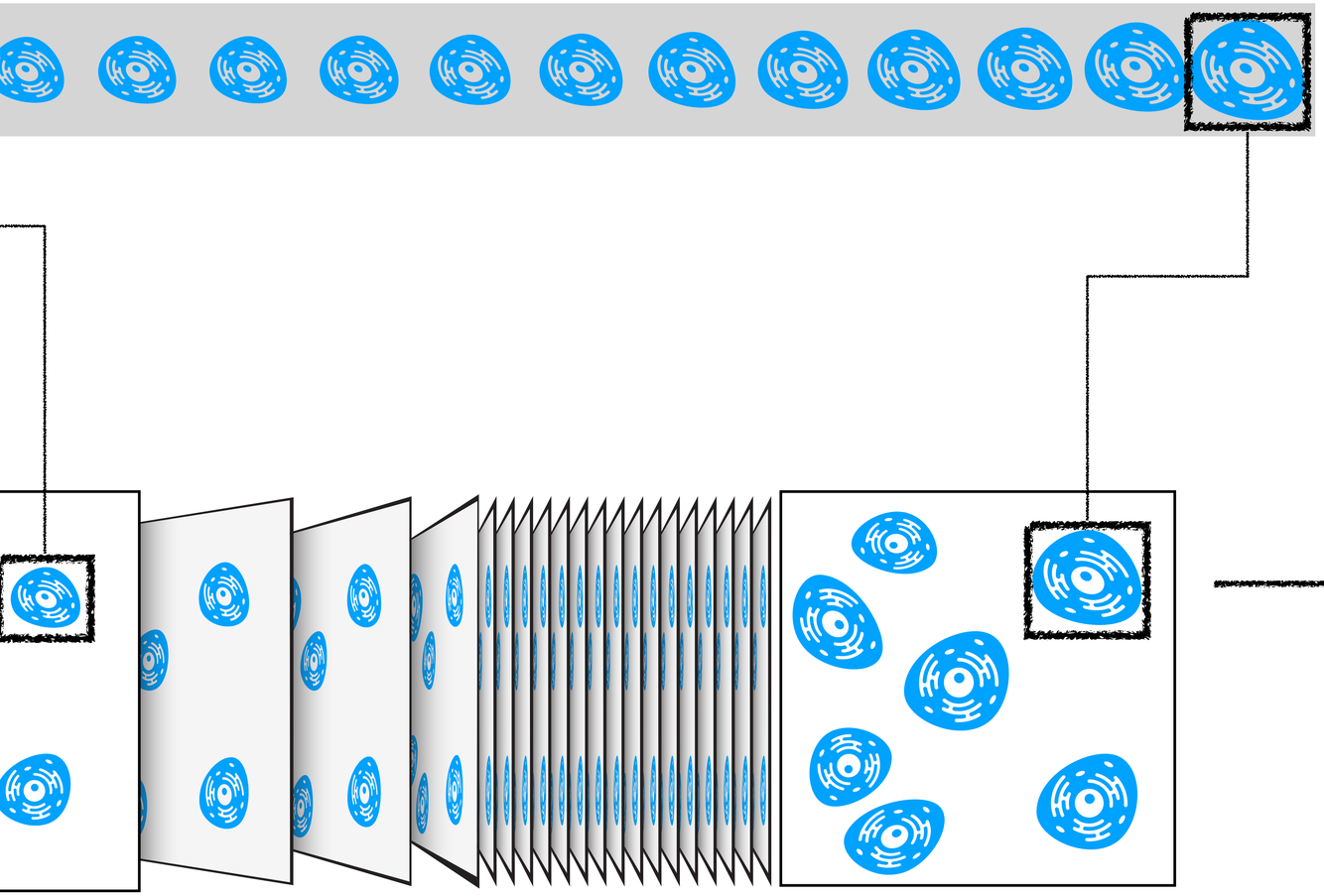
Tracking Cell Over Time



Tracking Cell Over Time



Cell and Track Attributes

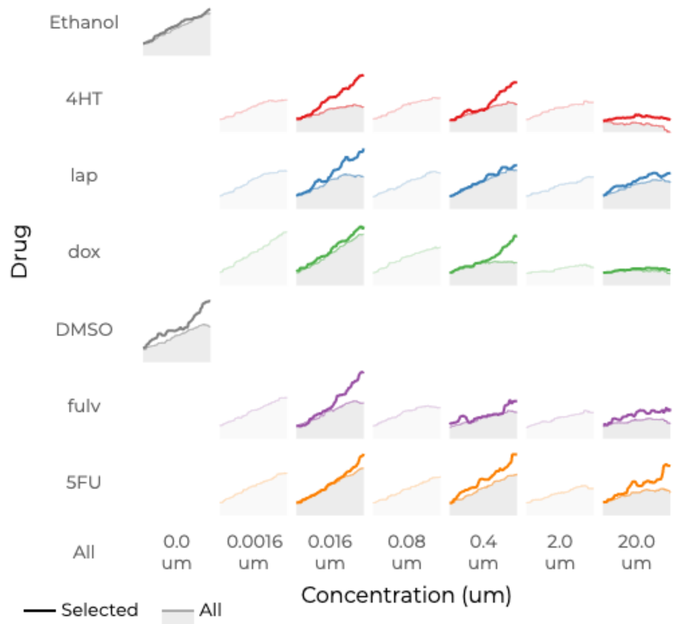


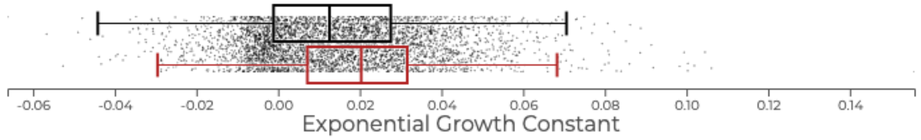
Average Mass
Growth Rate
Track Length

Current Mass
Shape Factor
Density

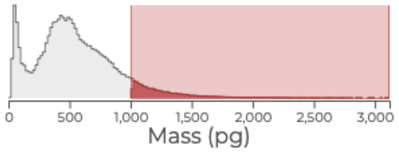
Facet | Compare

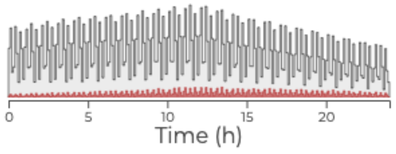
Axes: Averaged: Mass_norm over Frame ID

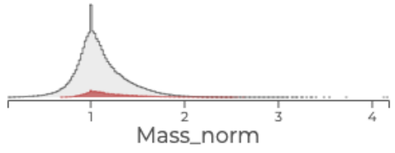




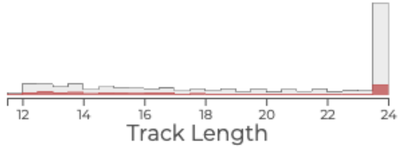
Cell-Level Attributes

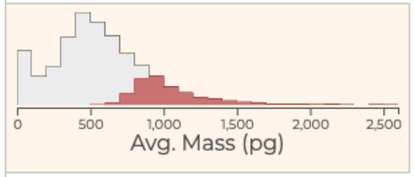


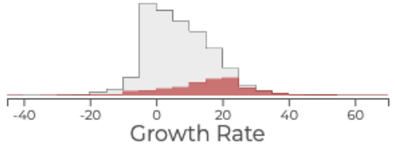




Track-Level Attributes





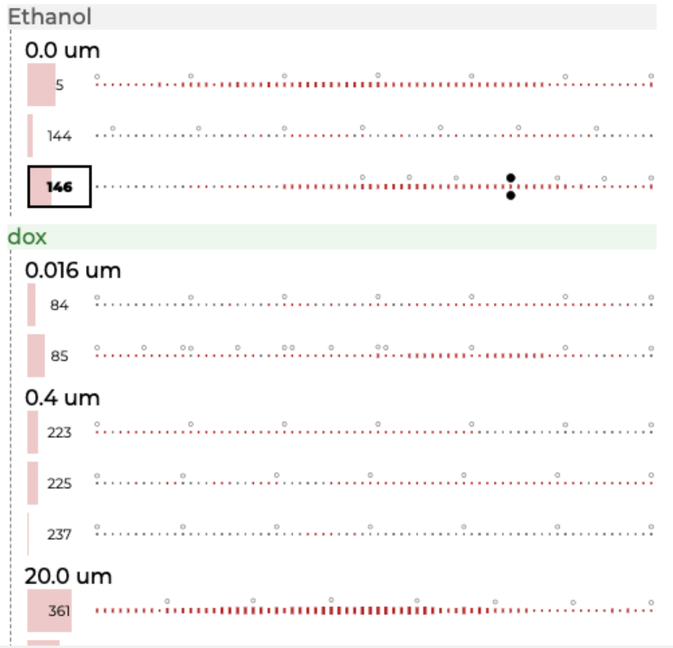


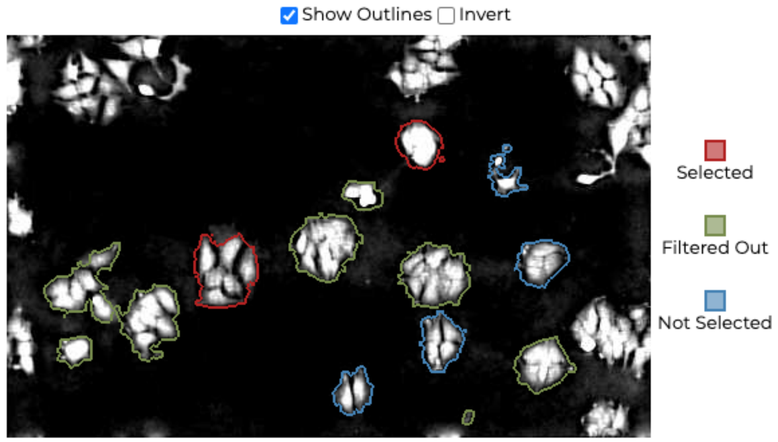
Group by: Drug | Concentration (um) +

and: Drug | Concentration (um) −

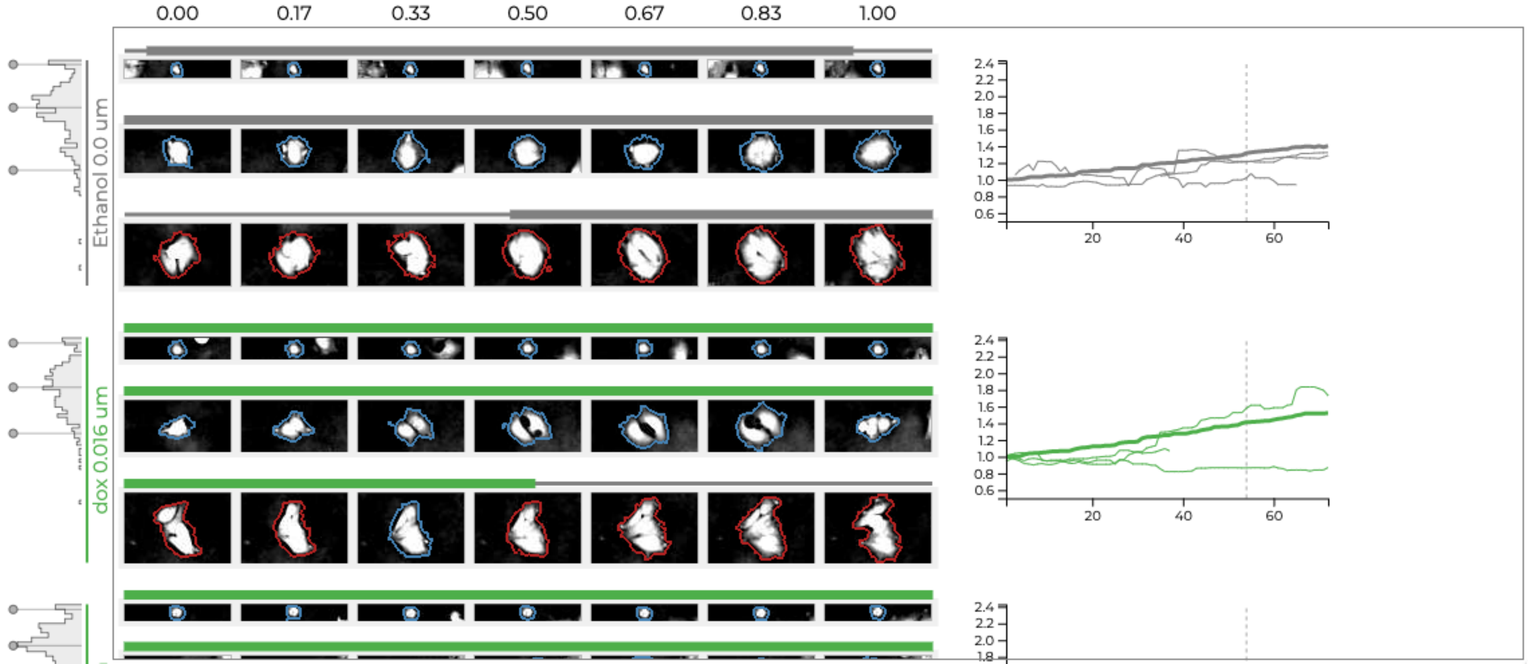
Legend

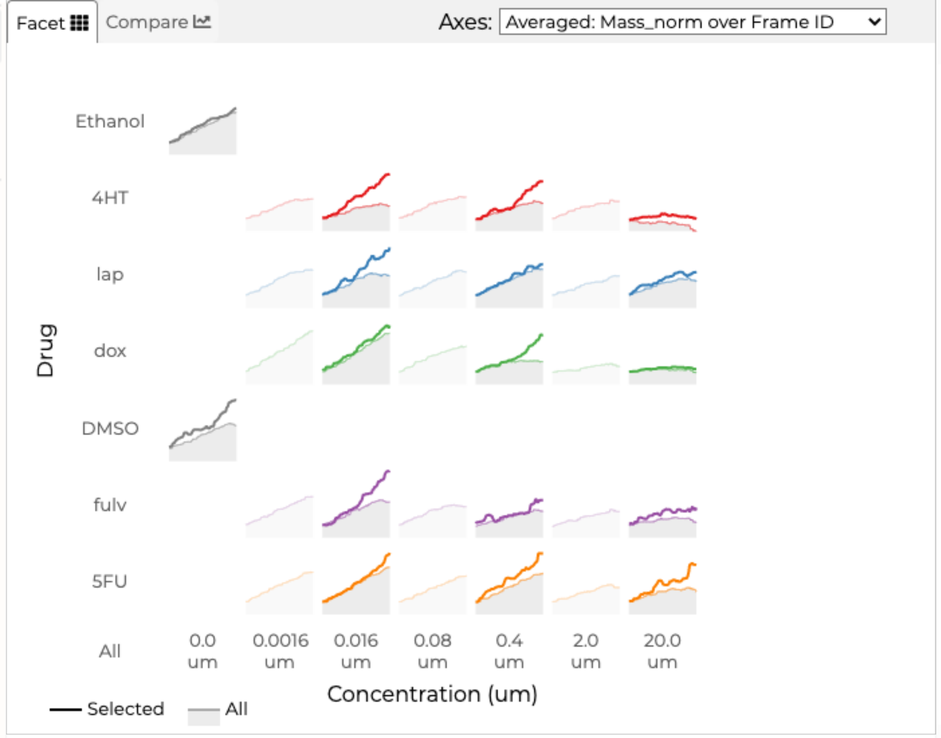
Condition: **Ethanol 0.0 um** Location: **146** Frame: **54**



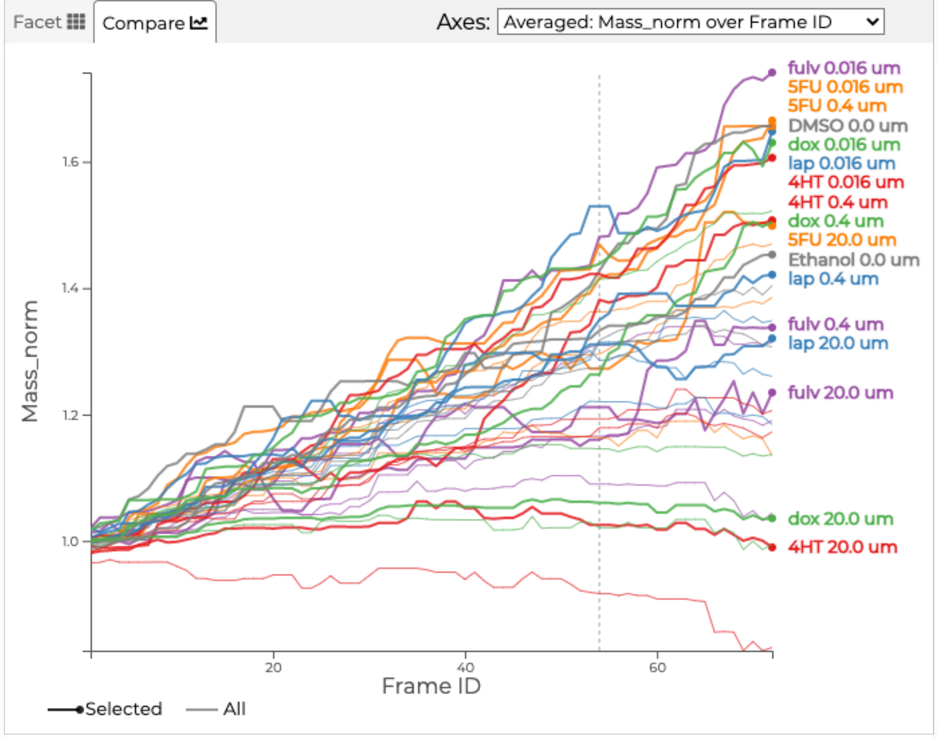


Exemplars of Avg. Mass (pg) - Sampled at: 5th percentile, Median, 95th percentile

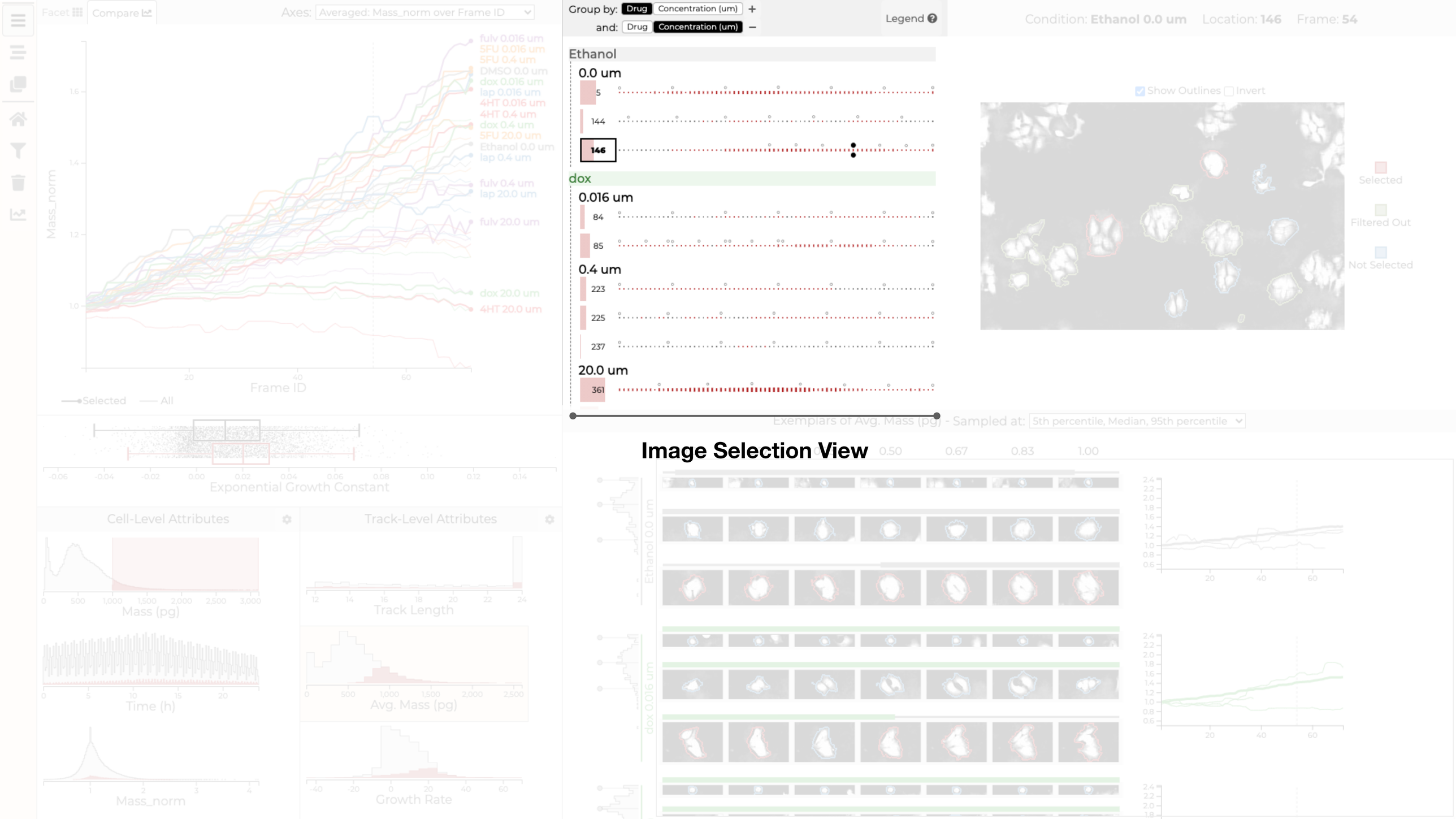




Condition Selector



Growth Curves


Facet
Compare
Axes: Averaged: Mass_norm over Frame ID
fulv 0.016 um
5FU 0.016 um
5FU 0.4 um
DMSO 0.0 um
dox 0.016 um
lap 0.016 um
4HT 0.016 um
4HT 0.4 um
dox 0.4 um
5FU 20.0 um
Ethanol 0.0 um
lap 0.4 um
fulv 0.4 um
lap 20.0 um
fulv 20.0 um
dox 20.0 um
4HT 20.0 um
Mass_norm
Frame ID
Selected
All
Exponential Growth Constant
Cell-Level Attributes
Track-Level Attributes
Mass (pg)
Track Length
Time (h)
Avg. Mass (pg)
Mass_norm
Growth Rate
Group by: Drug Concentration (um)
and: Drug Concentration (um)
Legend
Ethanol
0.0 um
5
144
146
dox
0.016 um
84
85
0.4 um
223
225
237
20.0 um
361
Condition: Ethanol 0.0 um
Location: 146
Frame: 54
Show Outlines
Invert
Selected
Filtered Out
Not Selected
Exemplars of Avg. Mass (pg) - Sampled at: 5th percentile, Median, 95th percentile
Image Selection View
0.50
0.67
0.83
1.00
Ethanol 0.0 um
dox 0.016 um

Facet | Compare | Axes: Averaged: Mass_norm over Frame ID

Group by: Drug | Concentration (um) + and: Drug | Concentration (um) −

Legend

Condition: **Ethanol 0.0 um** Location: **146** Frame: **54**

Ethanol
- 0.0 um: 5, 144, 146
dox
- 0.016 um: 84, 85
- 0.4 um: 223, 225, 237
- 20.0 um: 361

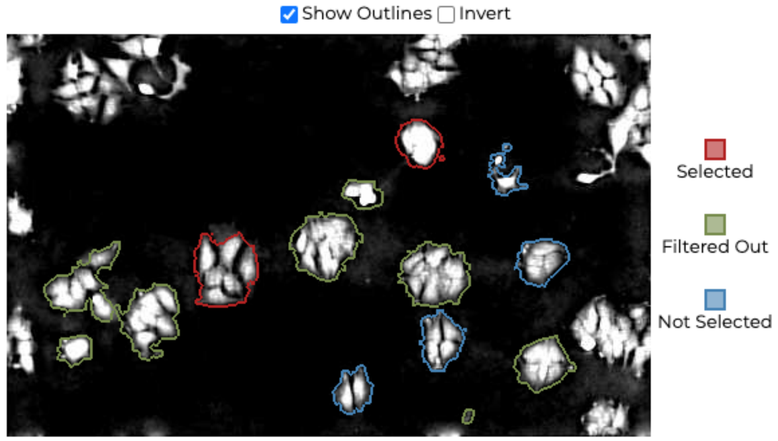
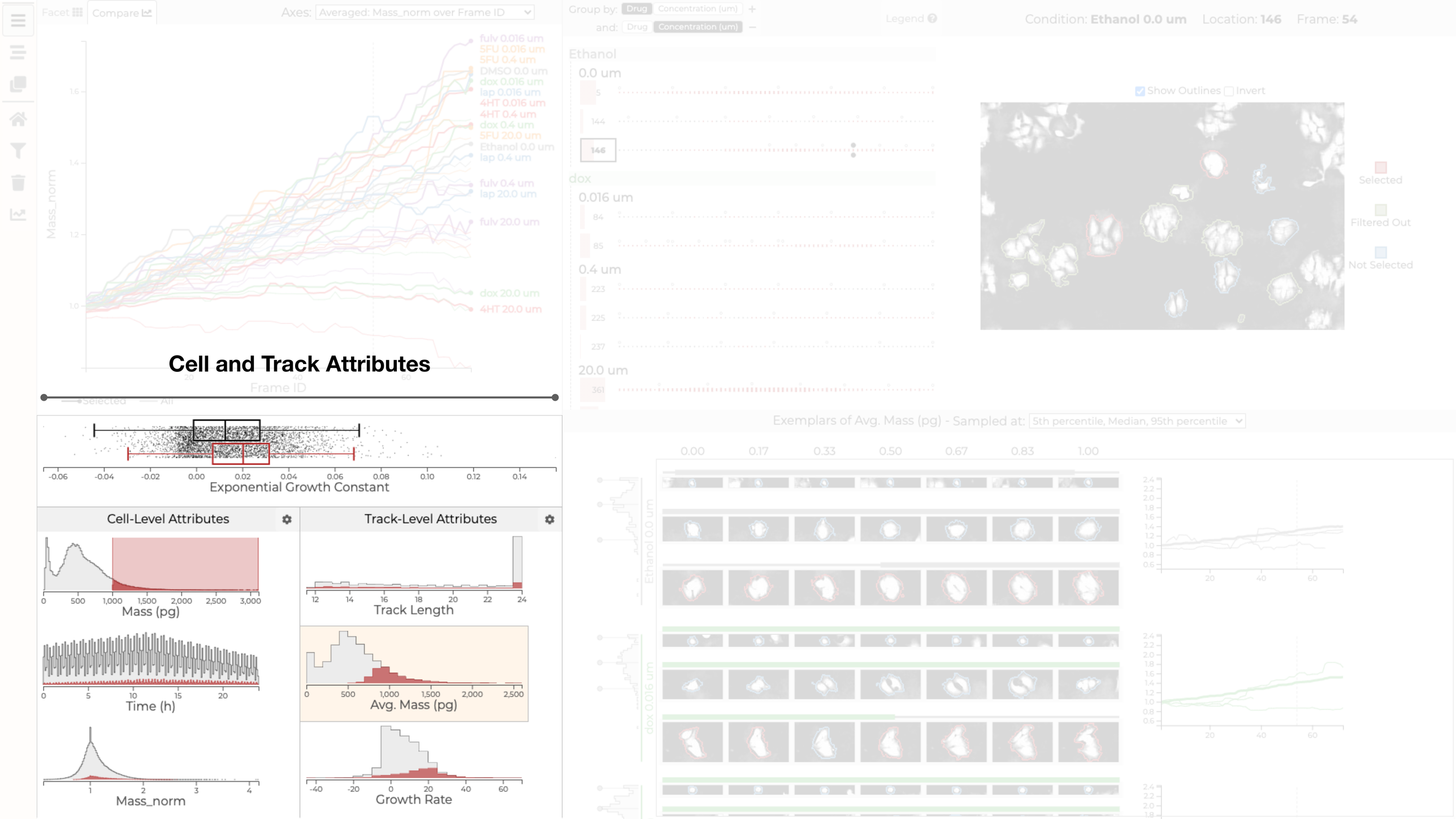


Image View

Exemplars of Avg. Mass (pg) - Sampled at: 5th percentile, Median, 95th percentile


Facet
Compare
Axes: Averaged: Mass_norm over Frame ID
fulv 0.016 um
5FU 0.016 um
5FU 0.4 um
DMSO 0.0 um
dox 0.016 um
lap 0.016 um
4HT 0.016 um
4HT 0.4 um
dox 0.4 um
5FU 20.0 um
Ethanol 0.0 um
lap 0.4 um
fulv 0.4 um
lap 20.0 um
fulv 20.0 um
dox 20.0 um
4HT 20.0 um
Mass_norm
Frame ID
Cell and Track Attributes
Selected
All
Exponential Growth Constant
Cell-Level Attributes
Track-Level Attributes
Mass (pg)
Track Length
Time (h)
Avg. Mass (pg)
Mass_norm
Growth Rate
Group by: Drug Concentration (um)
and: Drug Concentration (um)
Legend
Condition: Ethanol 0.0 um
Location: 146
Frame: 54
Ethanol
0.0 um
dox
0.016 um
0.4 um
20.0 um
Show Outlines
Invert
Selected
Filtered Out
Not Selected
Exemplars of Avg. Mass (pg) - Sampled at: 5th percentile, Median, 95th percentile
Ethanol 0.0 um
dox 0.016 um

Facet | Compare

Axes: Averaged: Mass_norm over Frame ID

Cell-Level Attributes

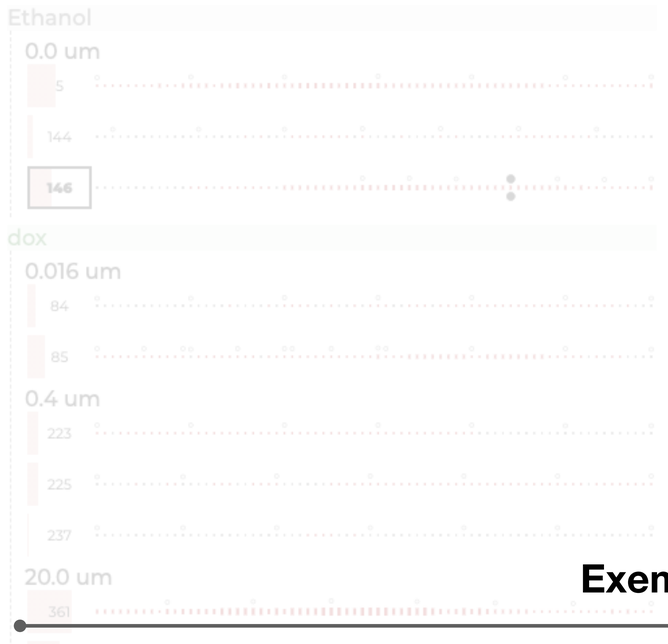
Track-Level Attributes

Group by: Drug | Concentration (um) +

and: Drug | Concentration (um) −

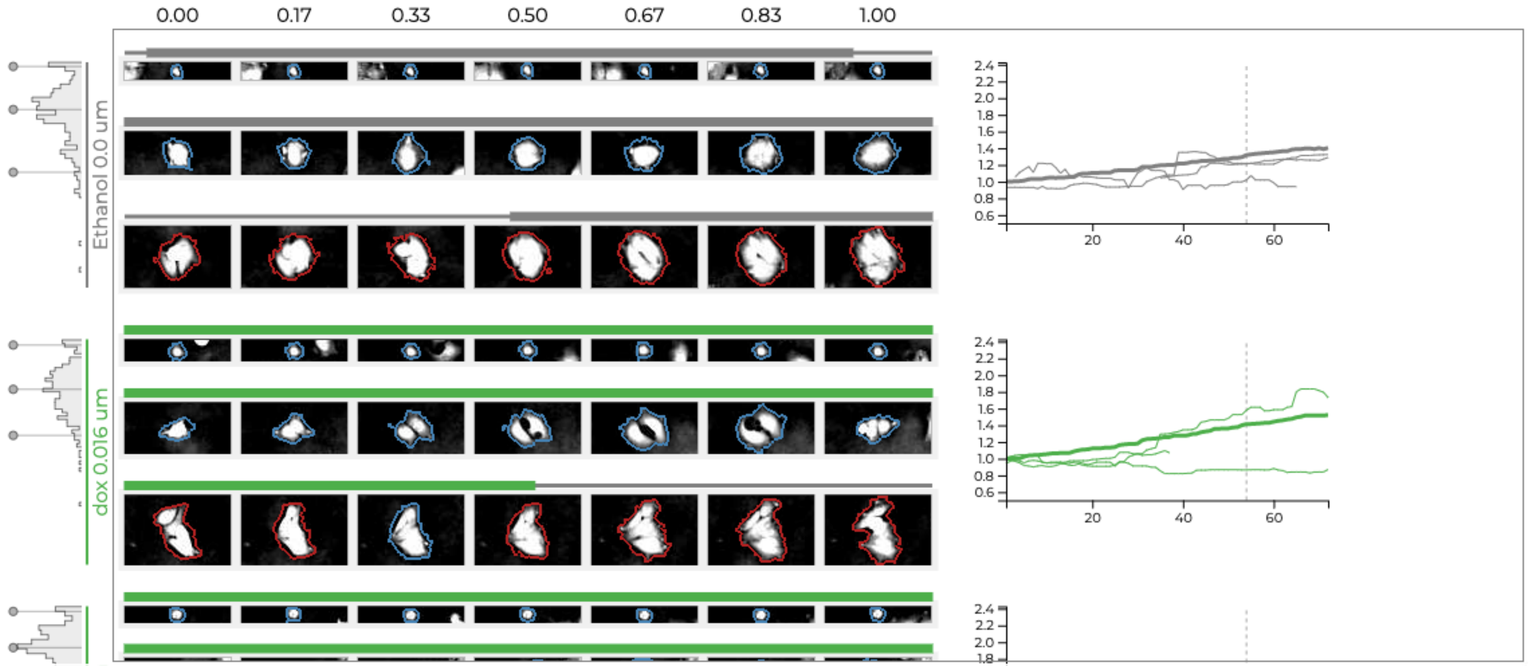
Legend

Condition: **Ethanol 0.0 um** Location: **146** Frame: **54**



Exemplar Tracks View

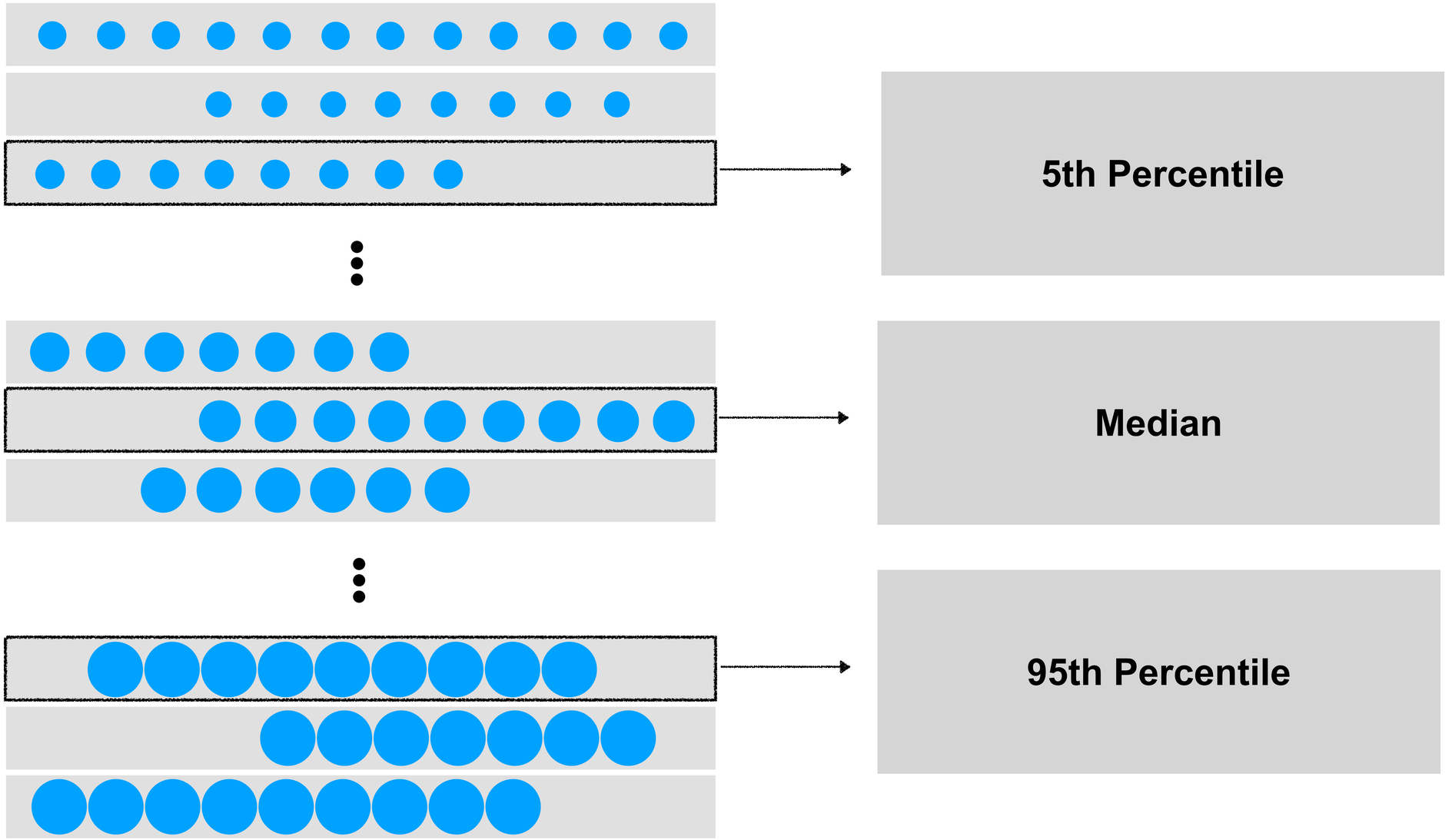
Exemplars of Avg. Mass (pg) - Sampled at: 5th percentile, Median, 95th percentile



WHY EXEMPLARS?

Want to see good representative examples.

Tricky to do given the scale of the data.

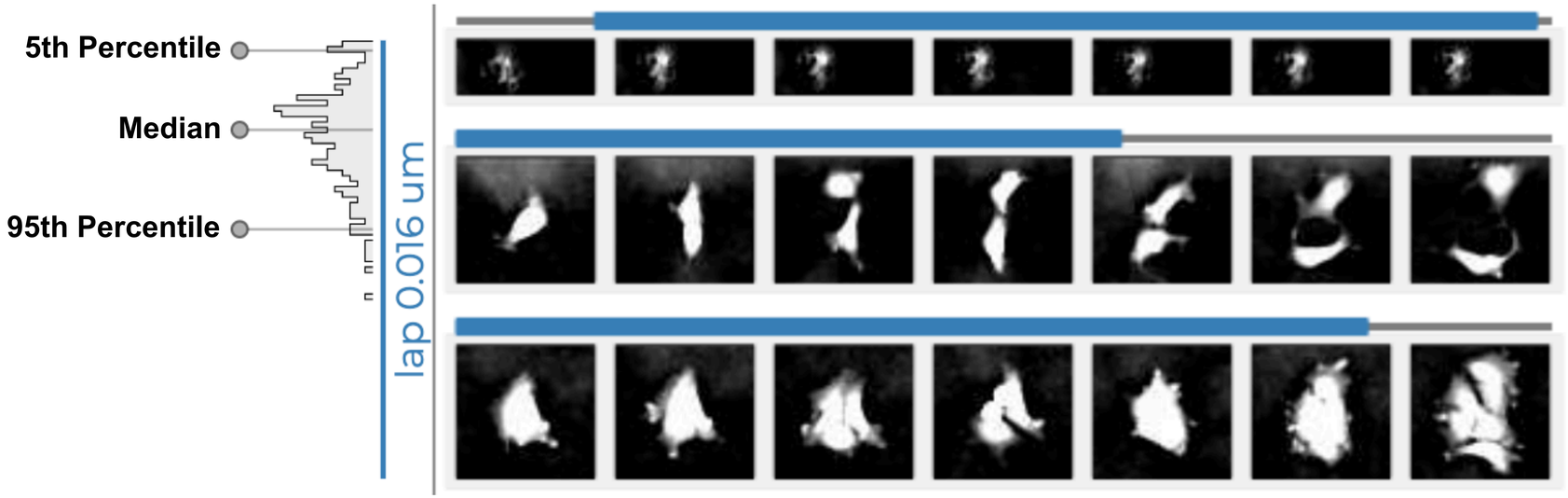


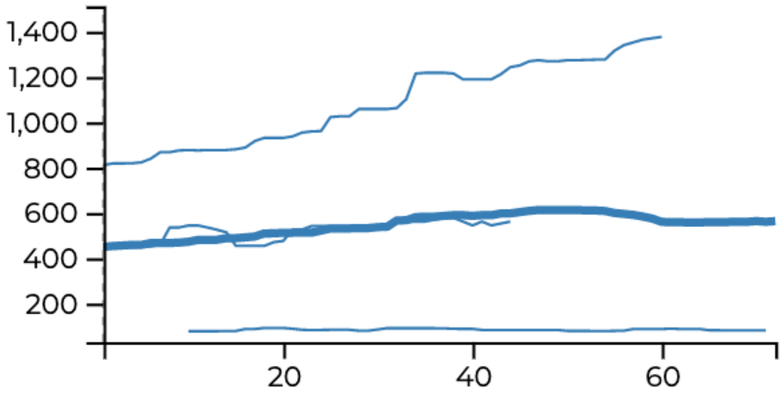
Sort by

Mass
Shape Factor
Growth Rate

Sample Tracks

5th, Median, 95th
Min, Median, Max
25th, Median, 75th





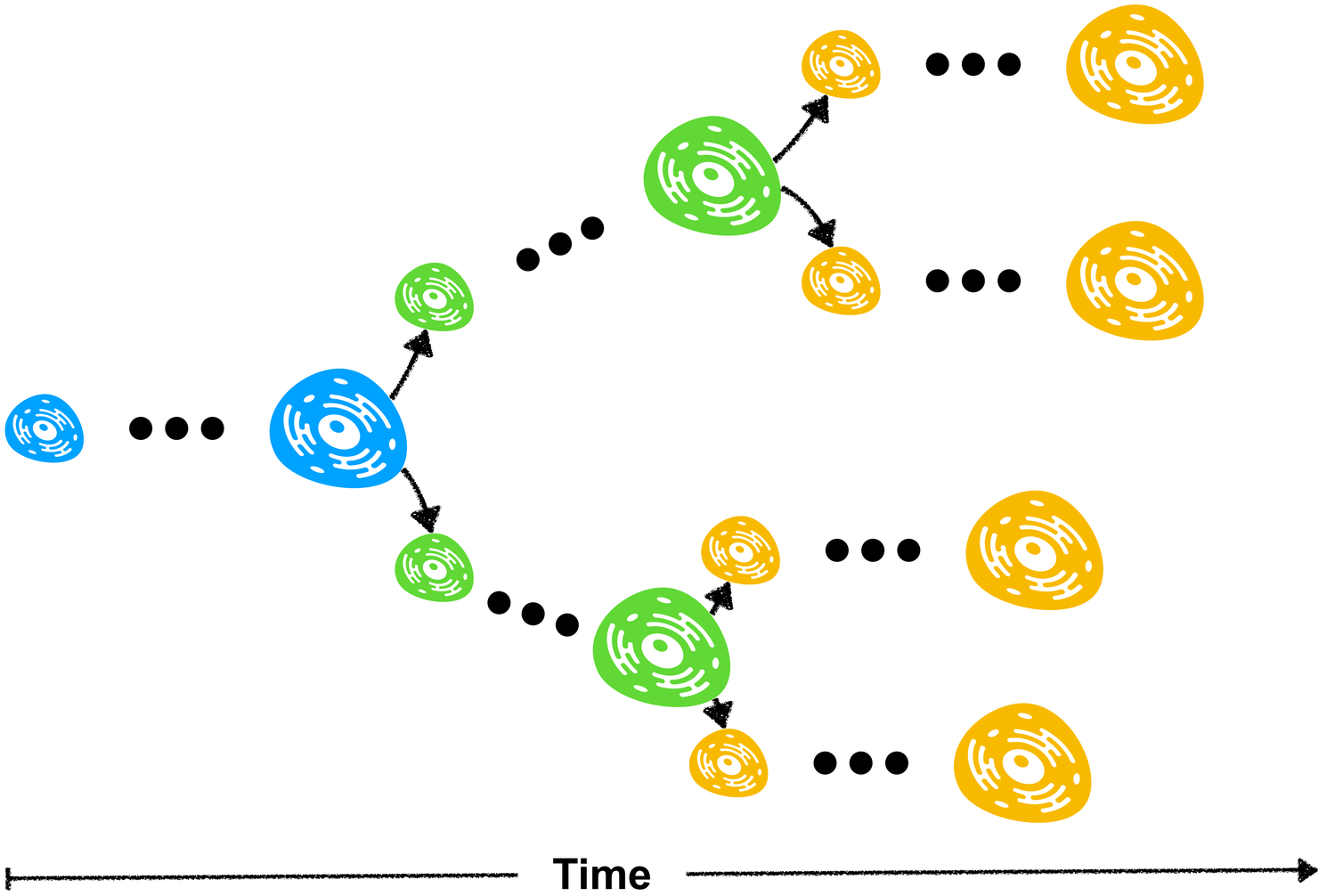
Final Display

When to use exemplars?

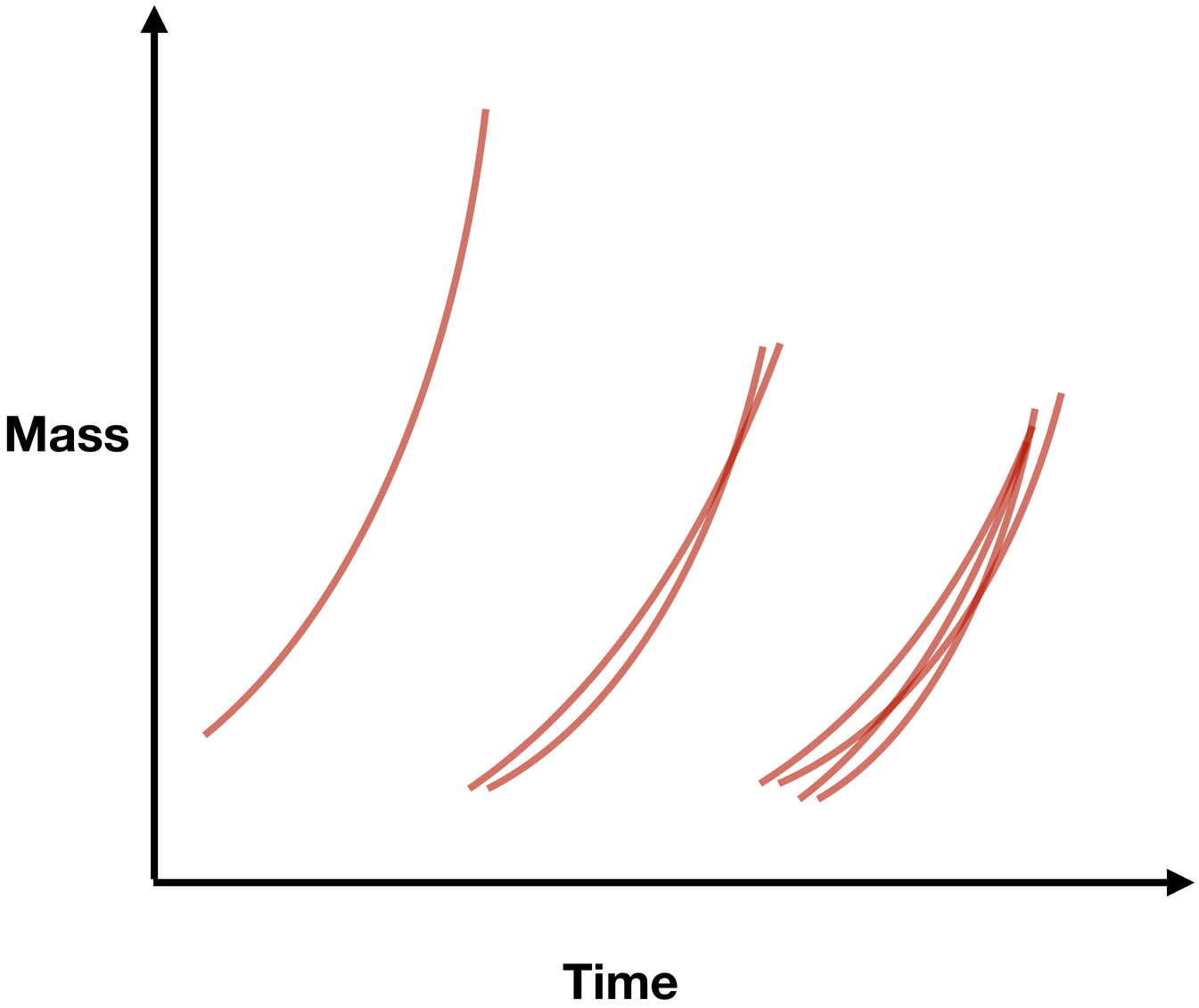
Rich contextual data that cannot be summarized

Dataset is sufficiently large

Can we analyze cells across generations?



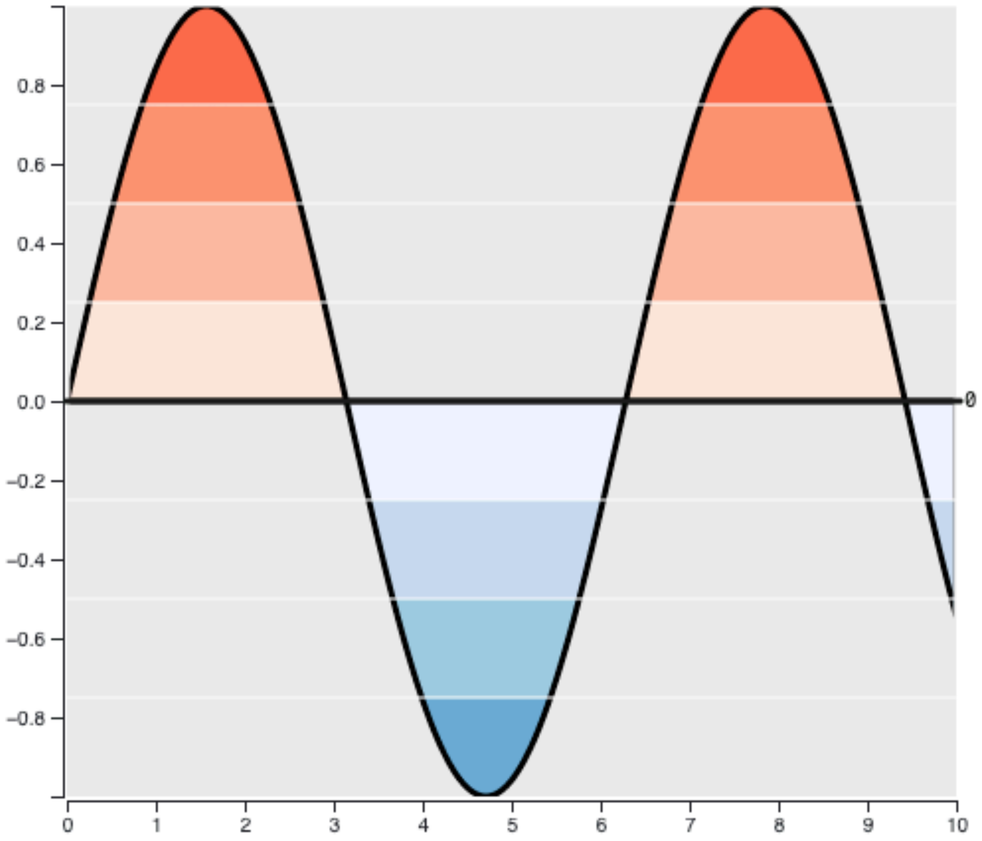
Existing Techniques



Time-Series Charts

Horizon Chart

Explanation Chart



https://www.horizon-chart-explanation.devinlange.com/

Applying to Lineages

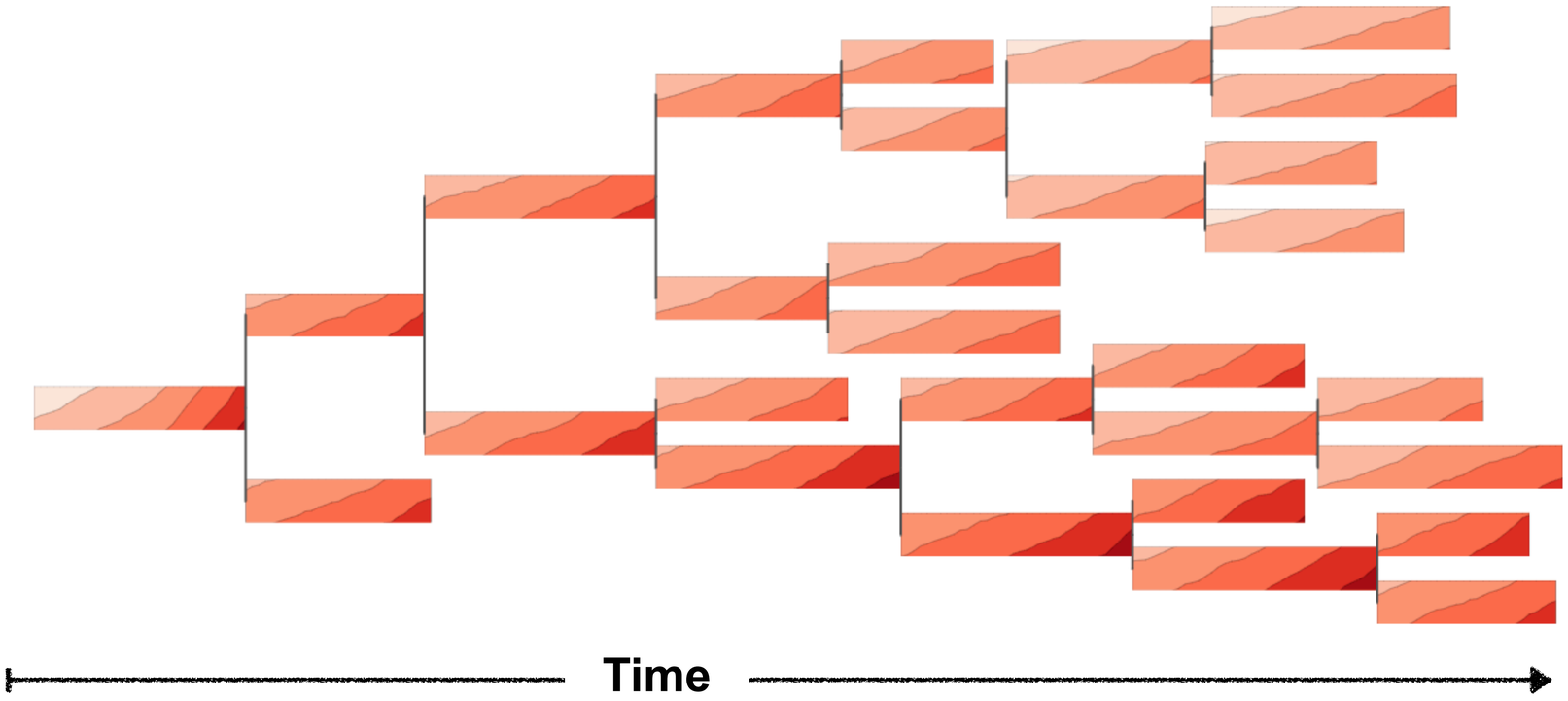
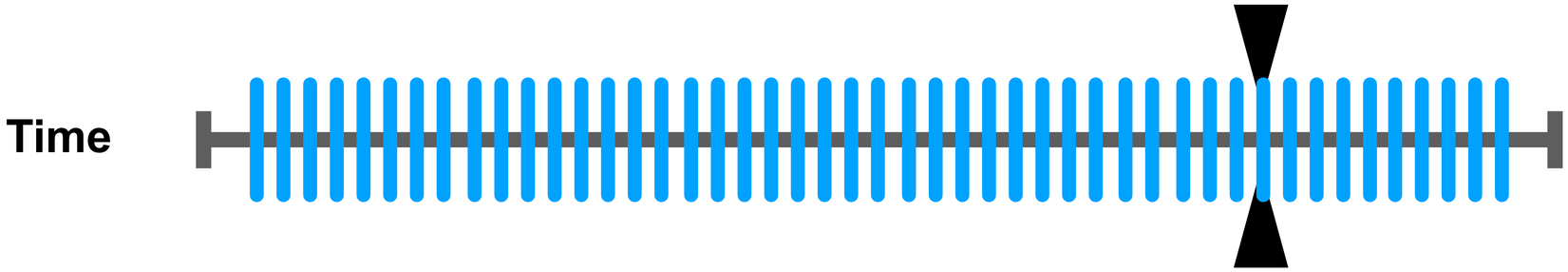


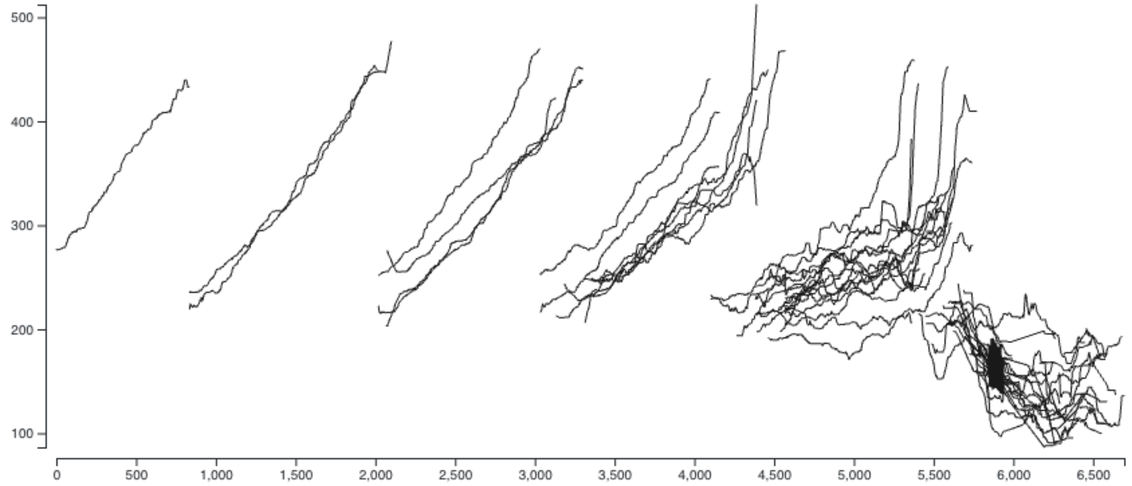
Image Scrubber



AggregateLineChart

Show mass for cell tracks

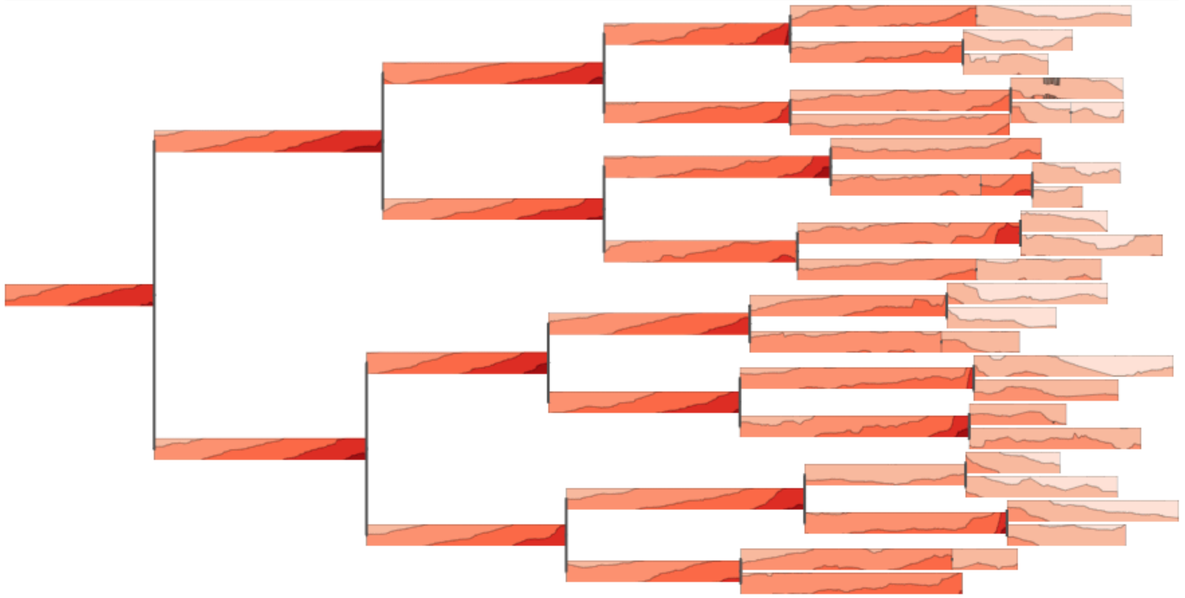
Smooth:



LooneageView

Attribute

mass



SWITCHING GEARS:

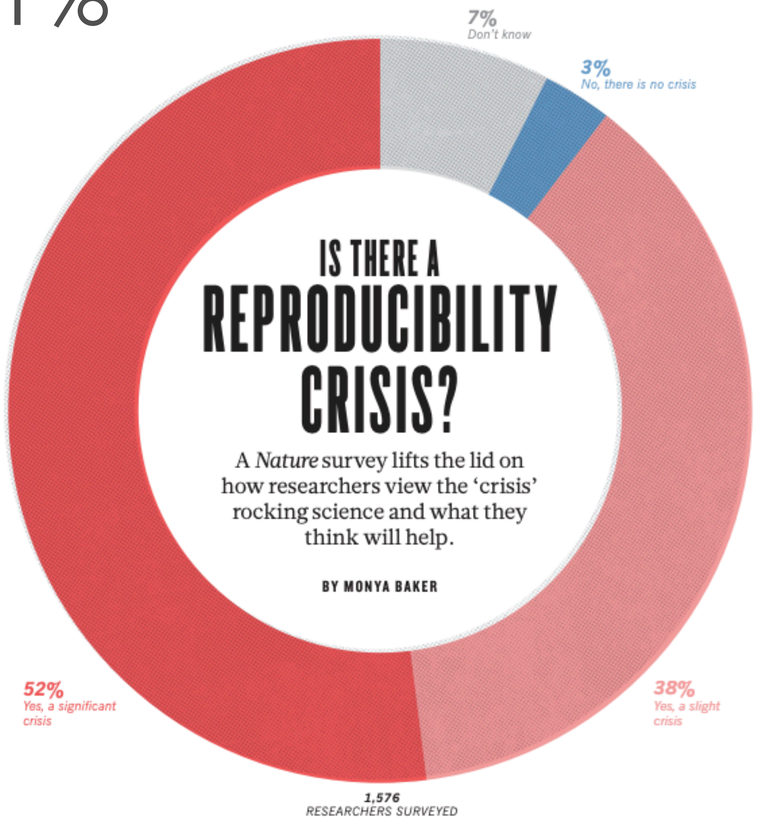
Reproducible Visual Analysis.

REPRODUCIBILITY CRISIS IN SCIENCE

Reproducibility Rates

Psychology: 40%

Cancer Biology: 11%



[Baker, Penny, 2016]
[Open Science Collaboration, 2016]
[Begley, Ellis, 2012]

WHY IS THERE A REPRODUCIBILITY CRISIS?

Perverse incentives (publish or perish)

Bias for “flashy” results

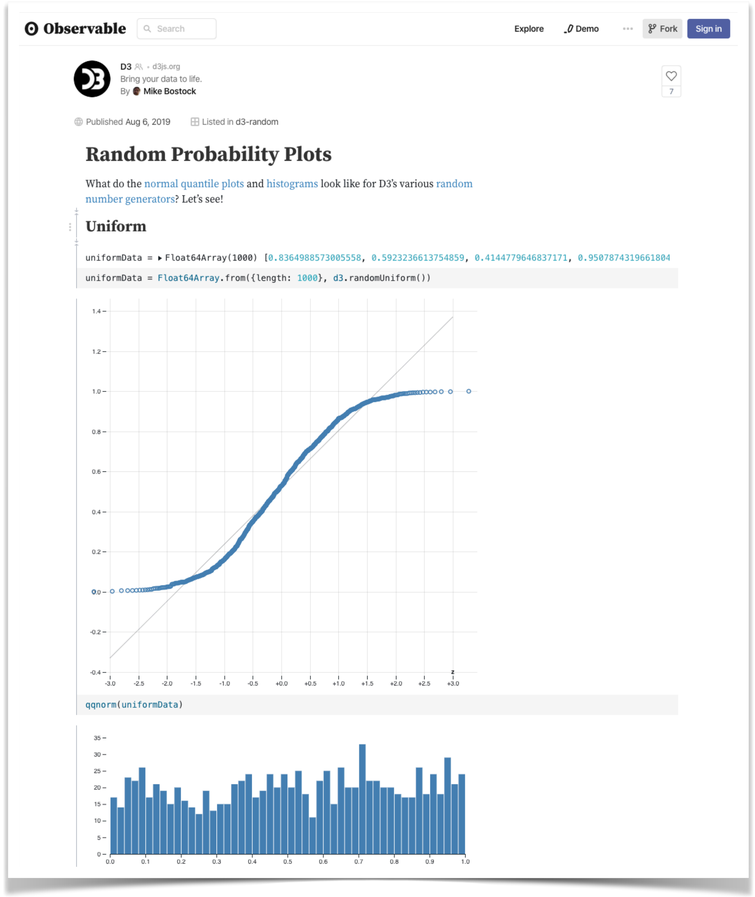
Problems in data analysis

Lack of data sharing

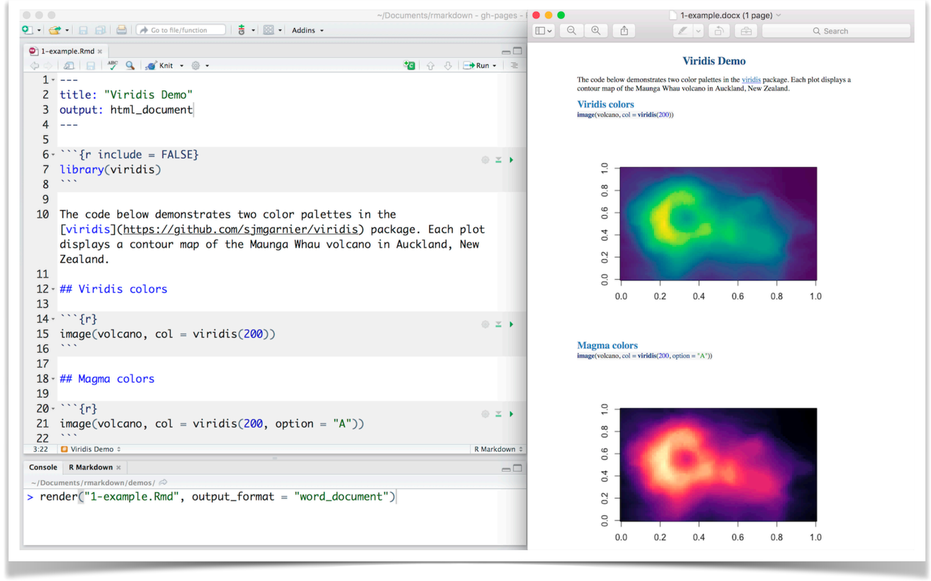
Lack of sharing the analysis process

…

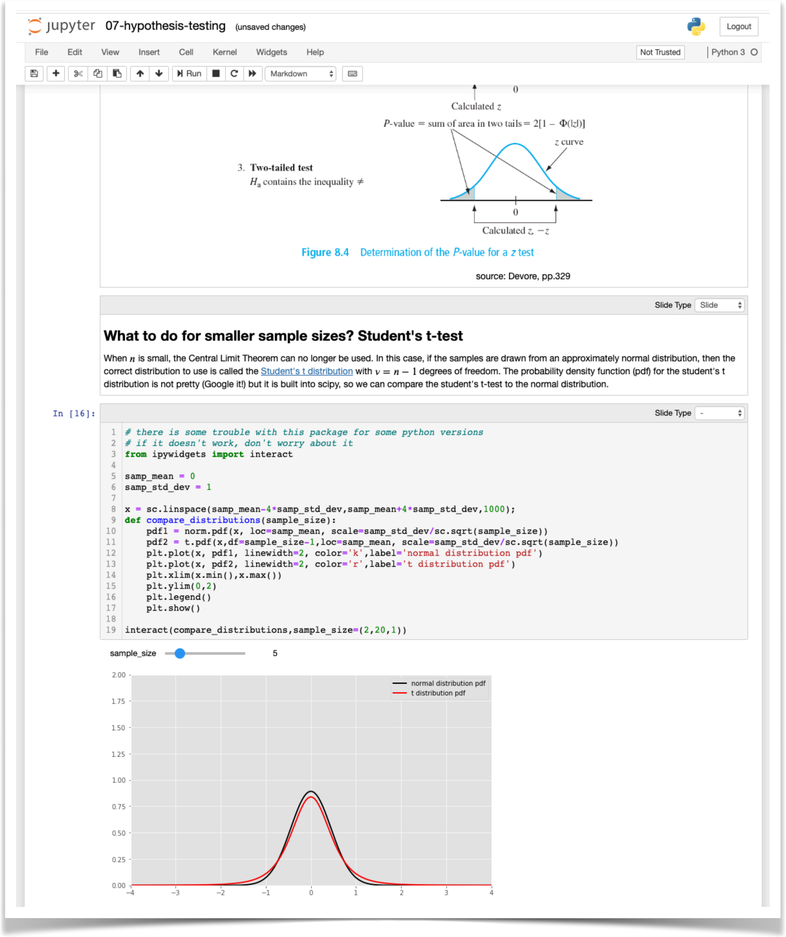
REPRODUCIBLE ANALYSIS IN THE WILD



Observable



R Markdown

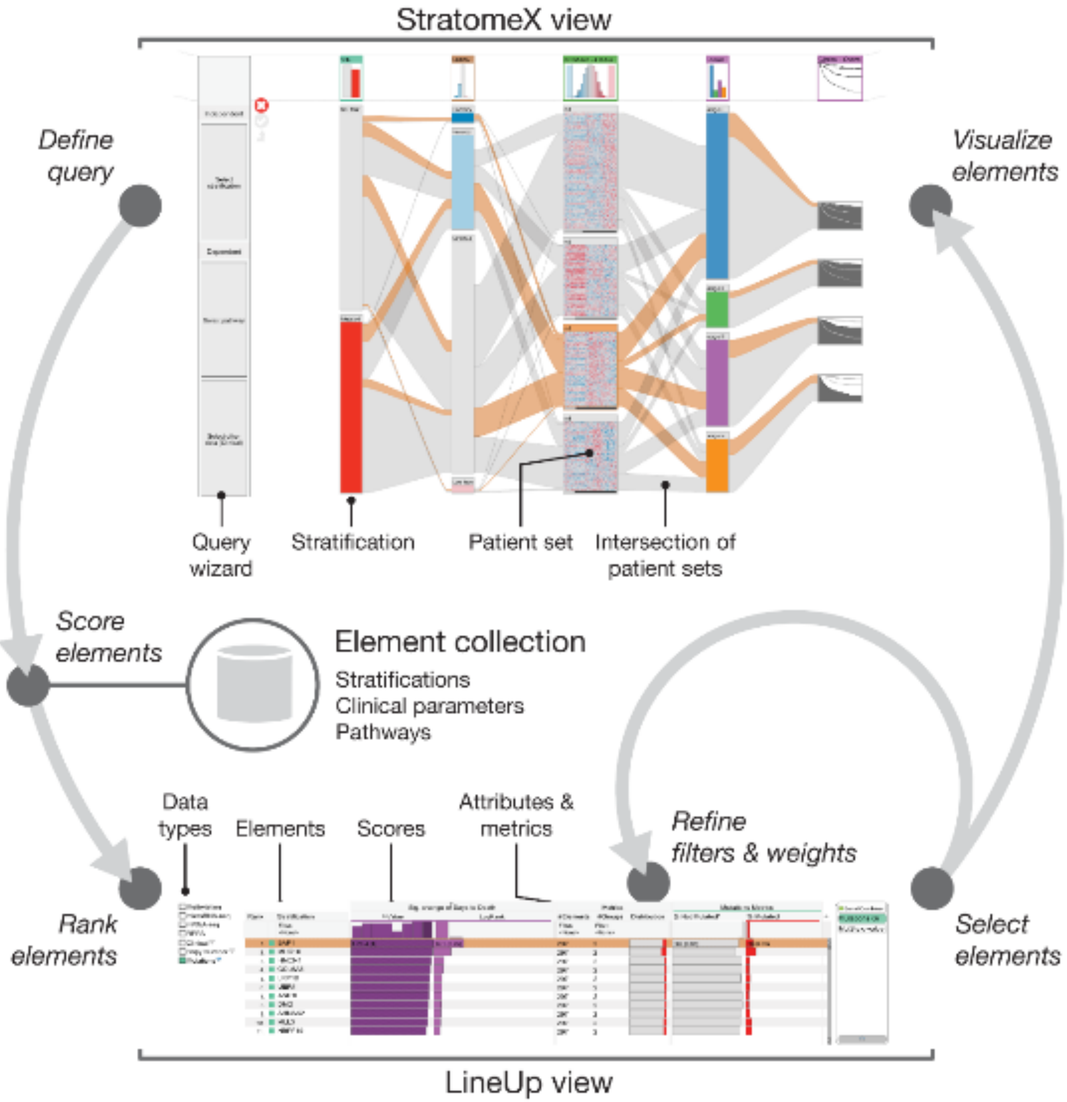


Jupyter Notebooks

THERE IS NO STRAIGHTFORWARD WAY TO DO

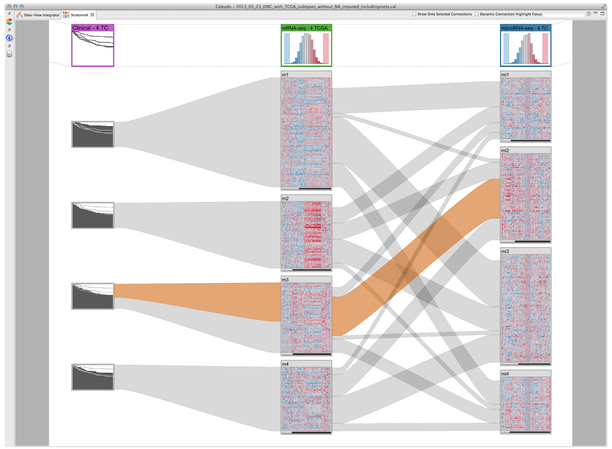
REPRODUCIBLE DATA VISUALIZATION

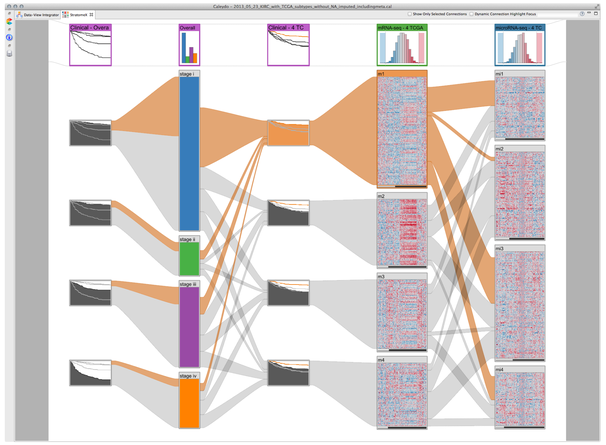
A MANUAL ATTEMPT AT LITERATE VISUALIZATION

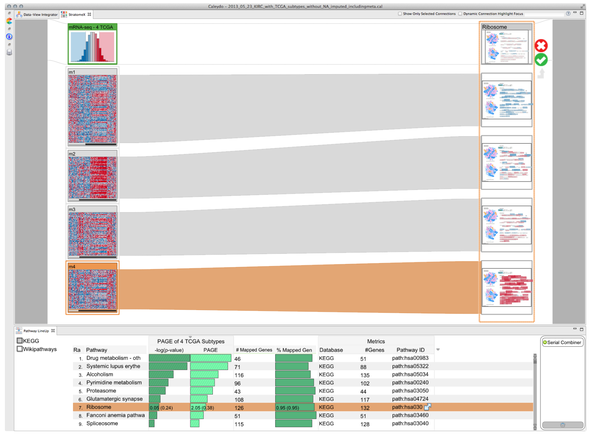


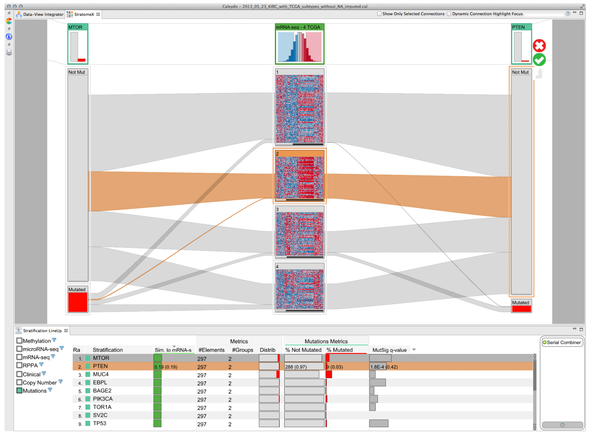
Guided Visual Exploration of Genomic Stratifications in Cancer

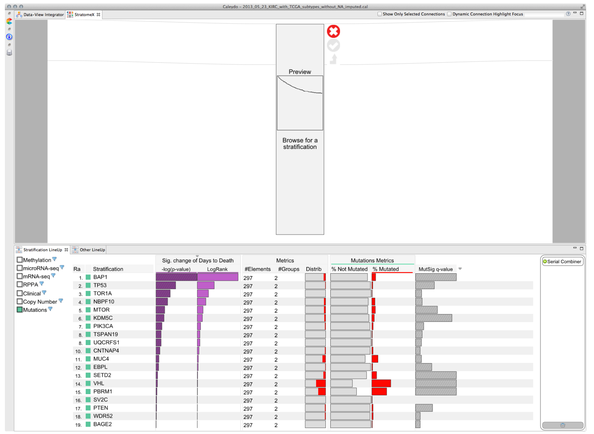
Nature Methods 11, 9 (2014), 884–885

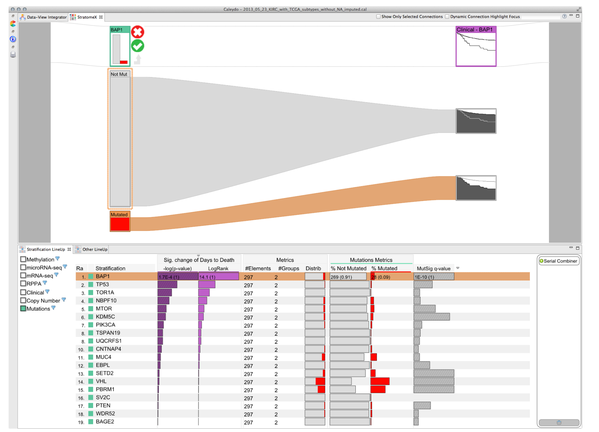


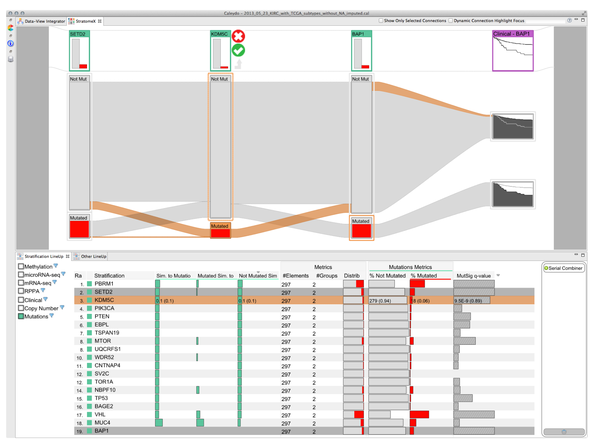


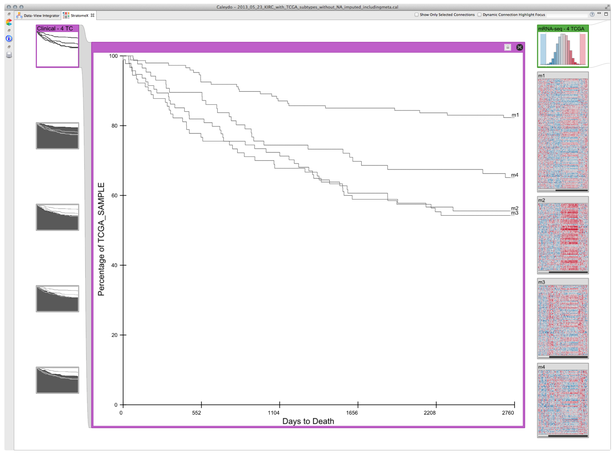










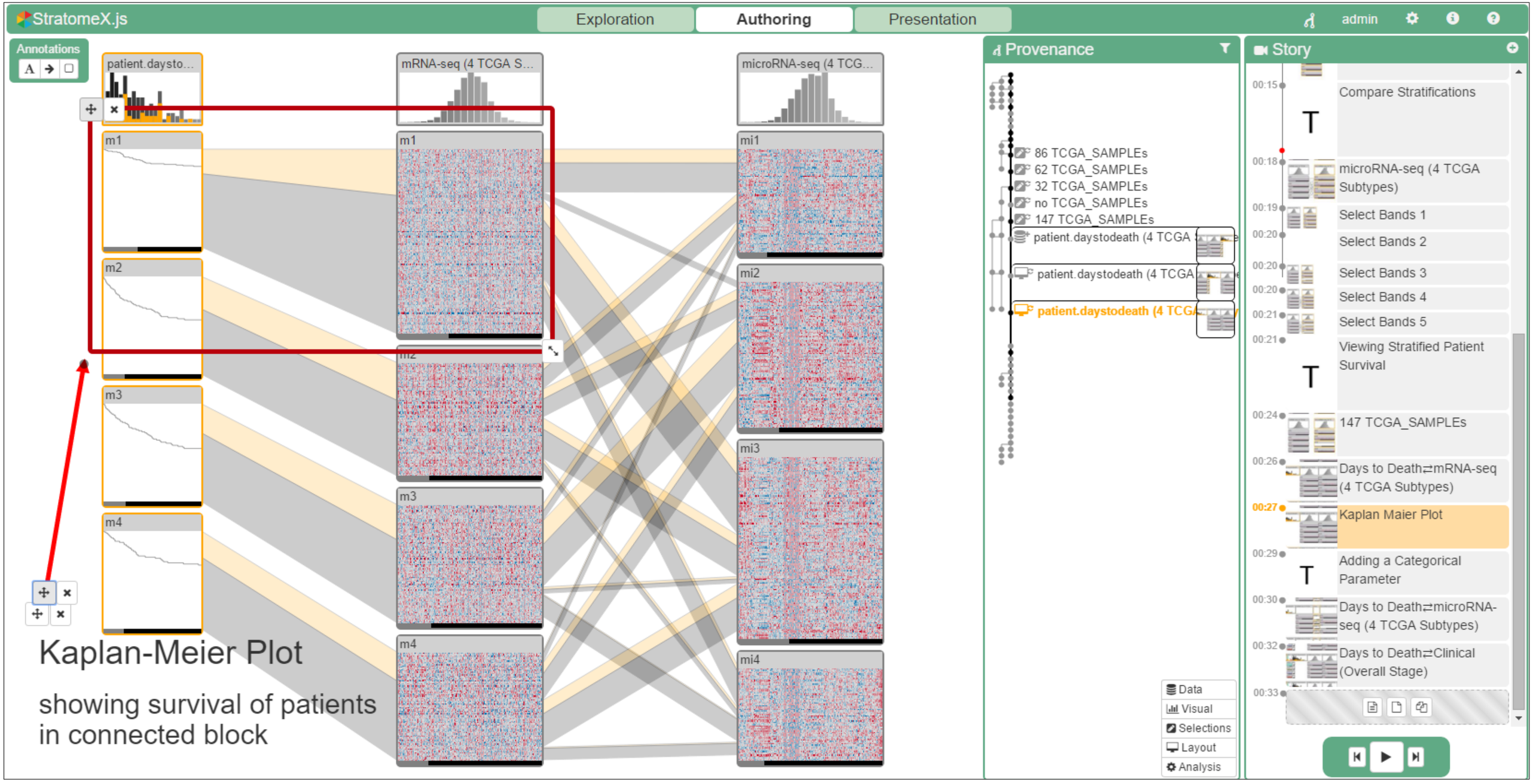


CASE STUDY

Streit et al., Guided Visual Exploration of Genomic Stratifications in Cancer, Nature Methods, 2014

IDEA: USE ANALYSIS SESSION PROVENANCE AS BASIS FOR A DATA STORY

EuroVis 2016. Samuel Gratzl, Alexander Lex,
Nils Gehlenborg, Nicola Cosgrove, Marc Streit



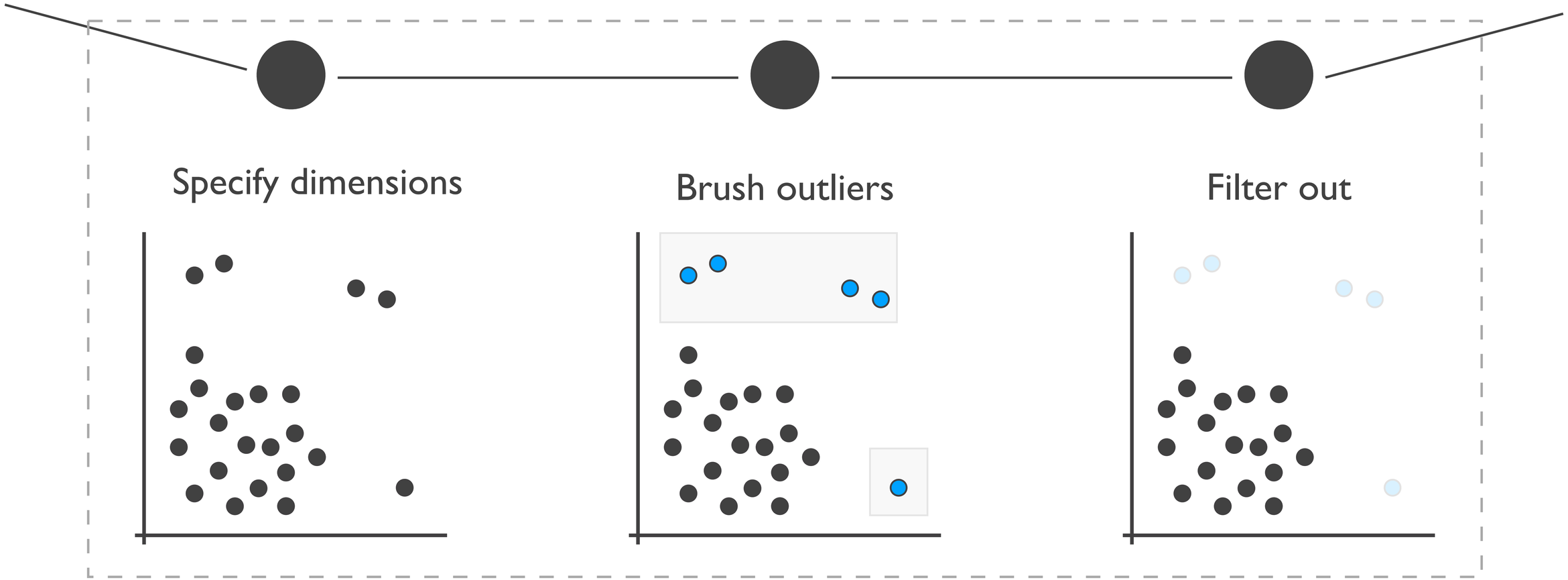
http://vistories.org/

WORKFLOWS

Can we “instrument” user actions?

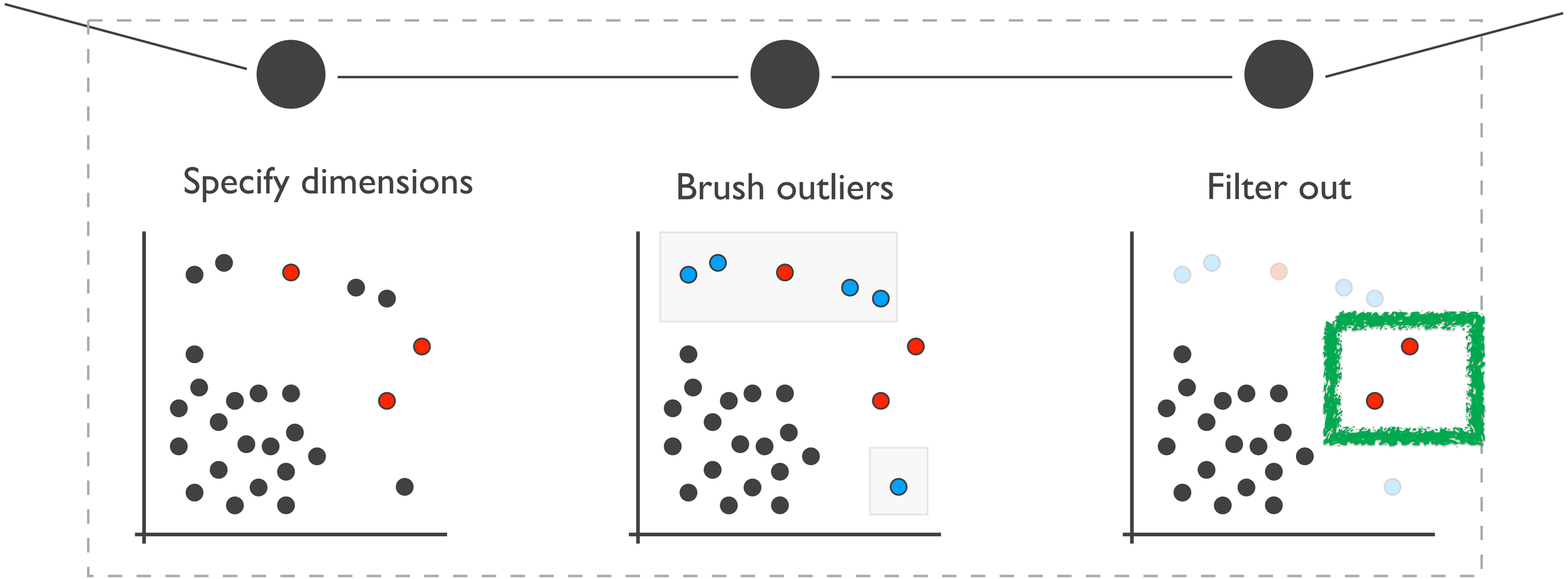
EuroVis 2022. Kiran Gadhave, Zach Cutler, Alexander Lex

IDEA: CAPTURING WORKFLOWS

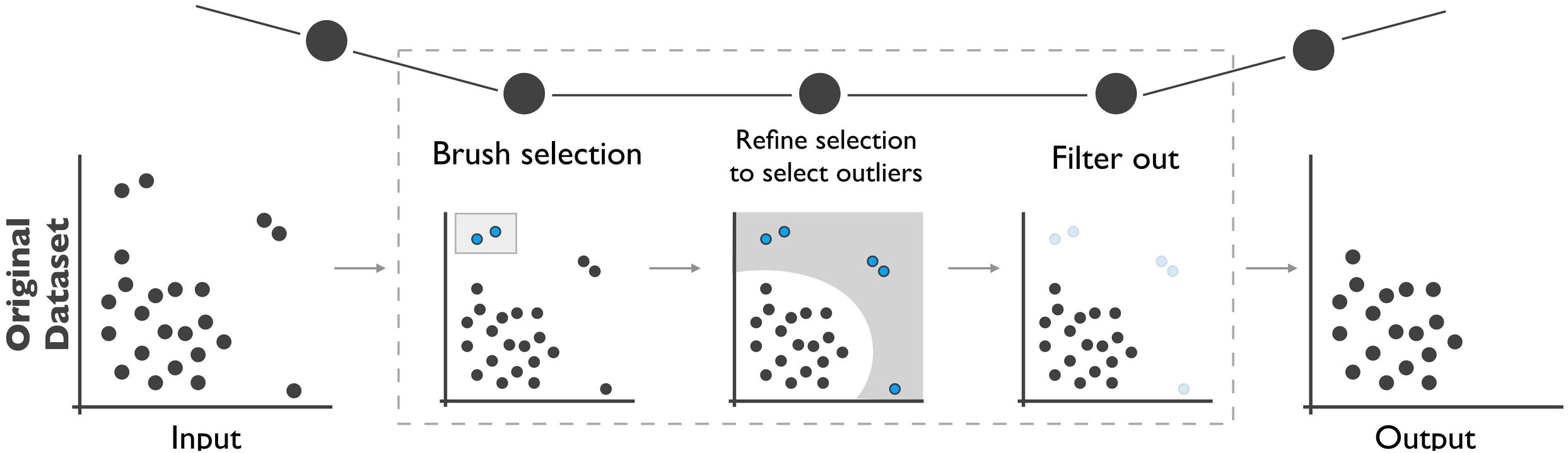


"Filter Outliers" Workflow

USING WORKFLOW ON UPDATED DATA

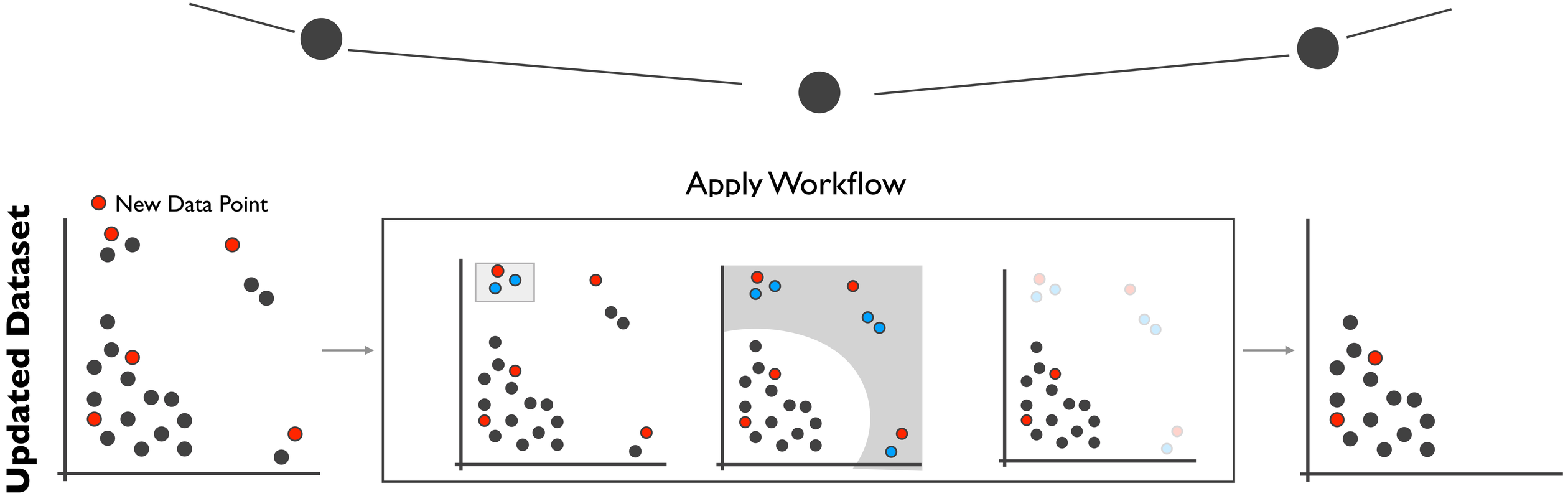


CAPTURING SEMANTICS OF WORKFLOWS

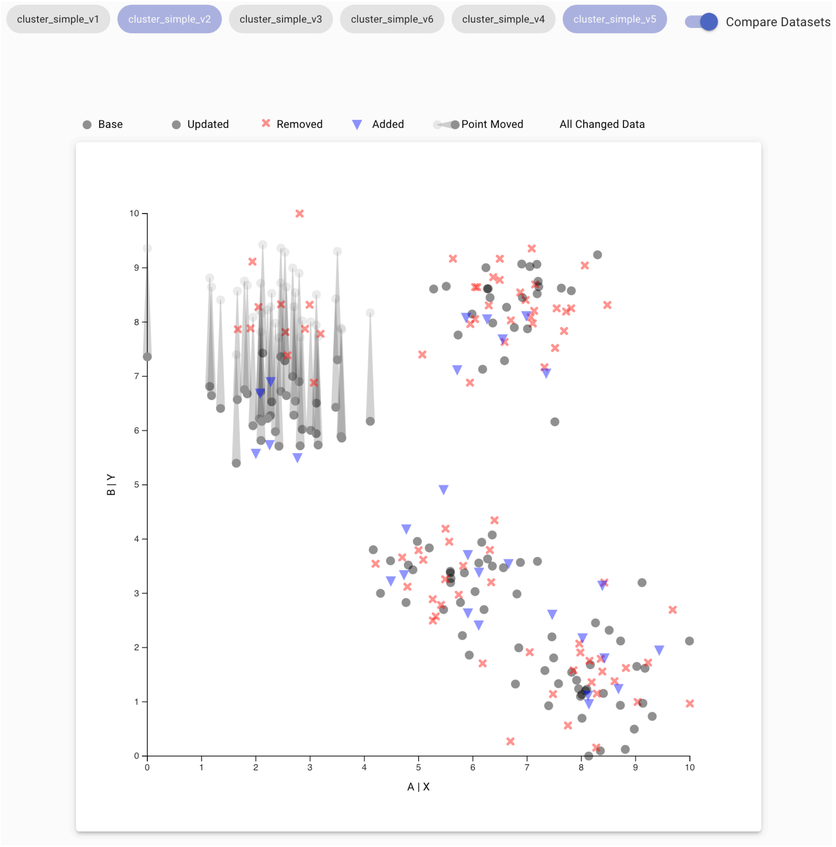


Robust “Filter Outliers” Workflow

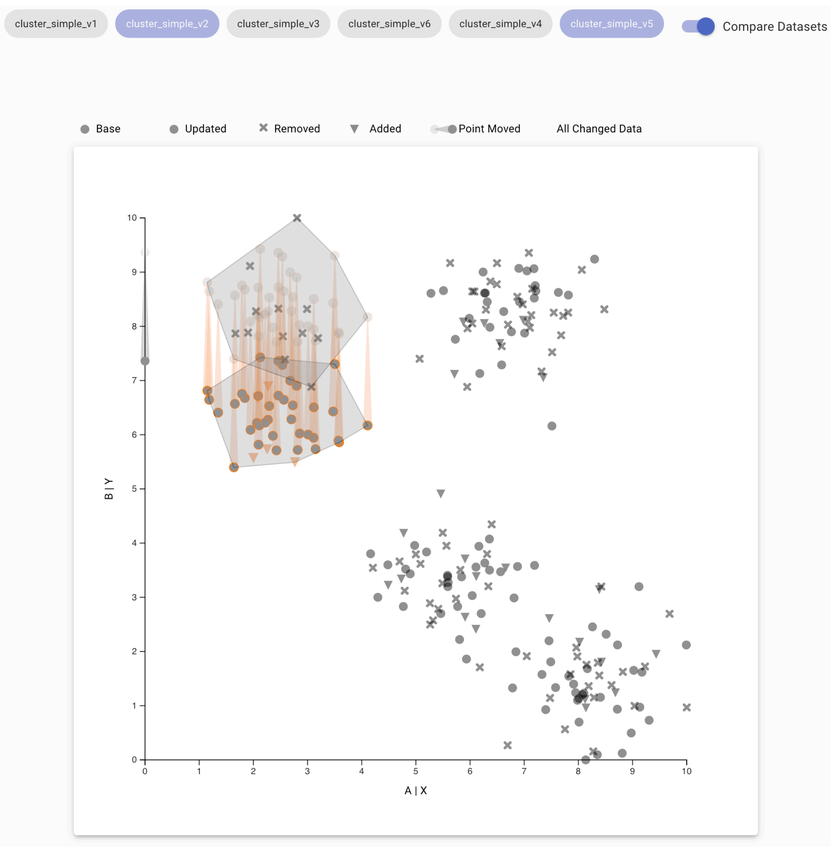
REUSING WORKFLOWS ON UPDATED DATA



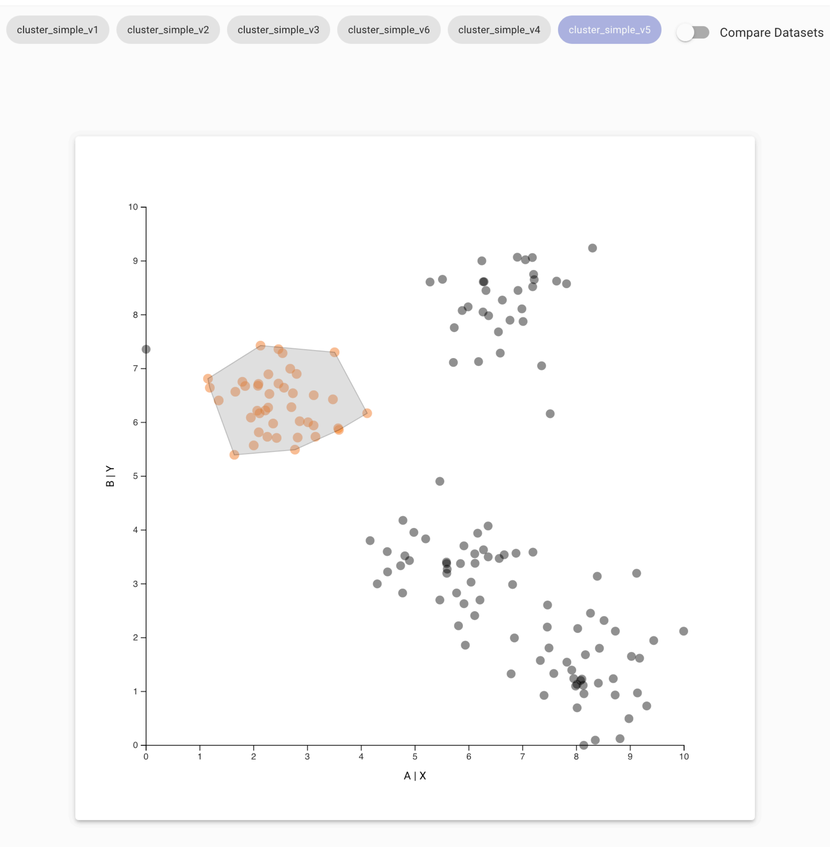
REUSING SELECTIONS ON UPDATED DATASETS



Changed Dataset



Tracking A Selected Cluster

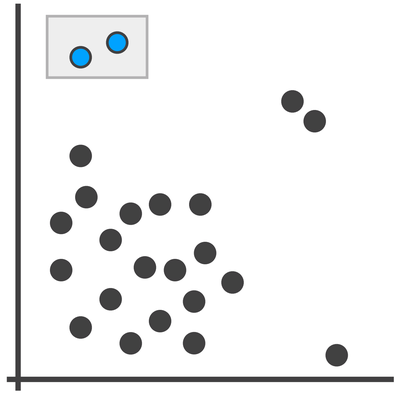


Selected Cluster on Changed Dataset

BRIDGING BETWEEN TOOLS

PROS AND CONS

Interactive Visual Analysis



Intuitive and Fast

Uses Human Perceptual Capabilities

Need to redo the analysis
when datasets update

Computational Analysis

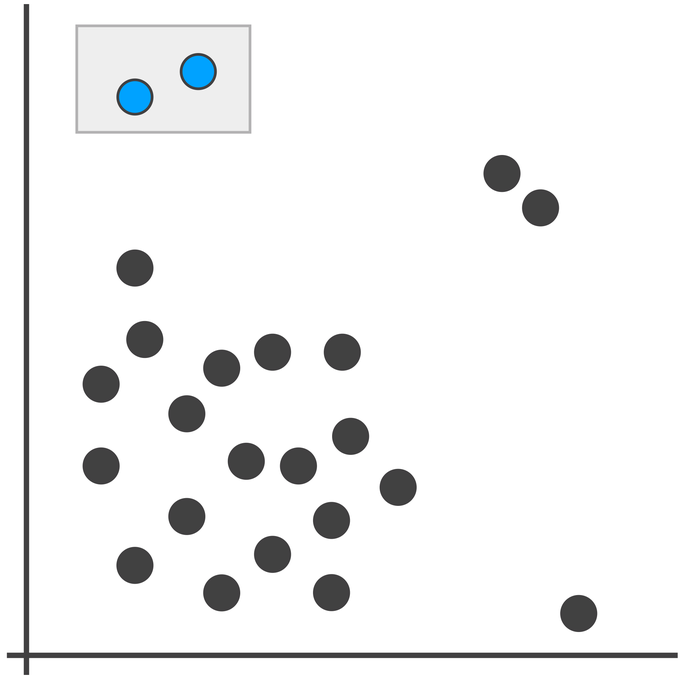
Flexible, powerful, reusable.

Time consuming to write

Have to be able to program

Difficult to see what's in the data

Interactive Visual Analysis



Workflow
Database

Computational Analysis

Library that tracks and re-executes actions

USING WORKFLOW IN A COMPUTATIONAL NOTEBOOK

```
# Installing the reapply-workflows adds a module called backend
# This module exposes the Reapply class which initializes the library

from backend import Reapply
```

```
# Here we load the reapply_workflows library.
r = Reapply()

# We add a workflow from our workflow database.
workflow = r.load_workflow("workflow16178086816 20")
```

```
# Print the workflow name
print("Workflow: ", workflow.name, "\n")

# Description of the workflow and the operations in it
workflow.describe
```

```
Workflow:  Deleting Cluster

| Root
 +--| Add Plot
     +--| Added brush to: X-Y
         +--| Cluster Selection
             +--| Filter: Out
```

```
# Prints the reapply results for all interactions, along with review status.
```

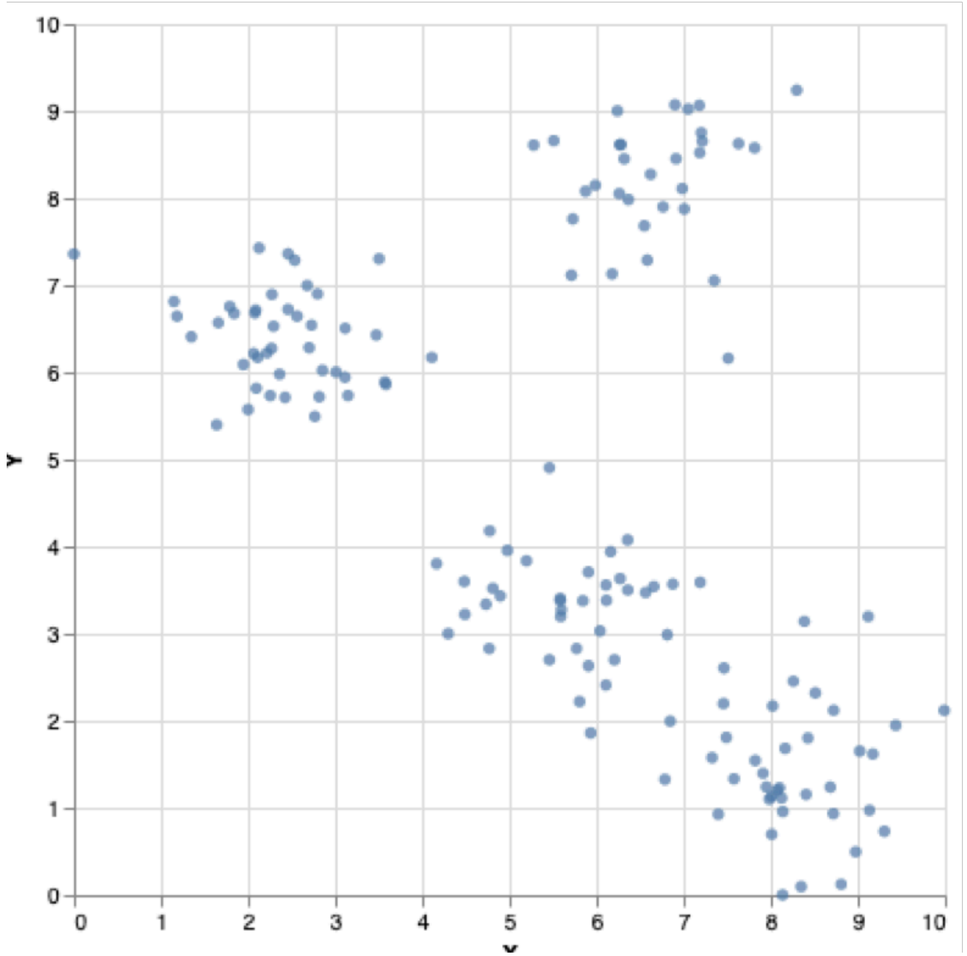
```
# Apply the workflow to target dataset.
# apply function requires the target dataset
# and the label column as arguments.
res = workflow.apply(target, "Label")

# Results is an array of datasets for each interaction
# we grab the final one.
result_dataset = res.results[-1]['data']
result_dataset
```

```
This workflow has not been reviewed for all interactions.
Please go to following url: https://reapply-workflows.gith
```

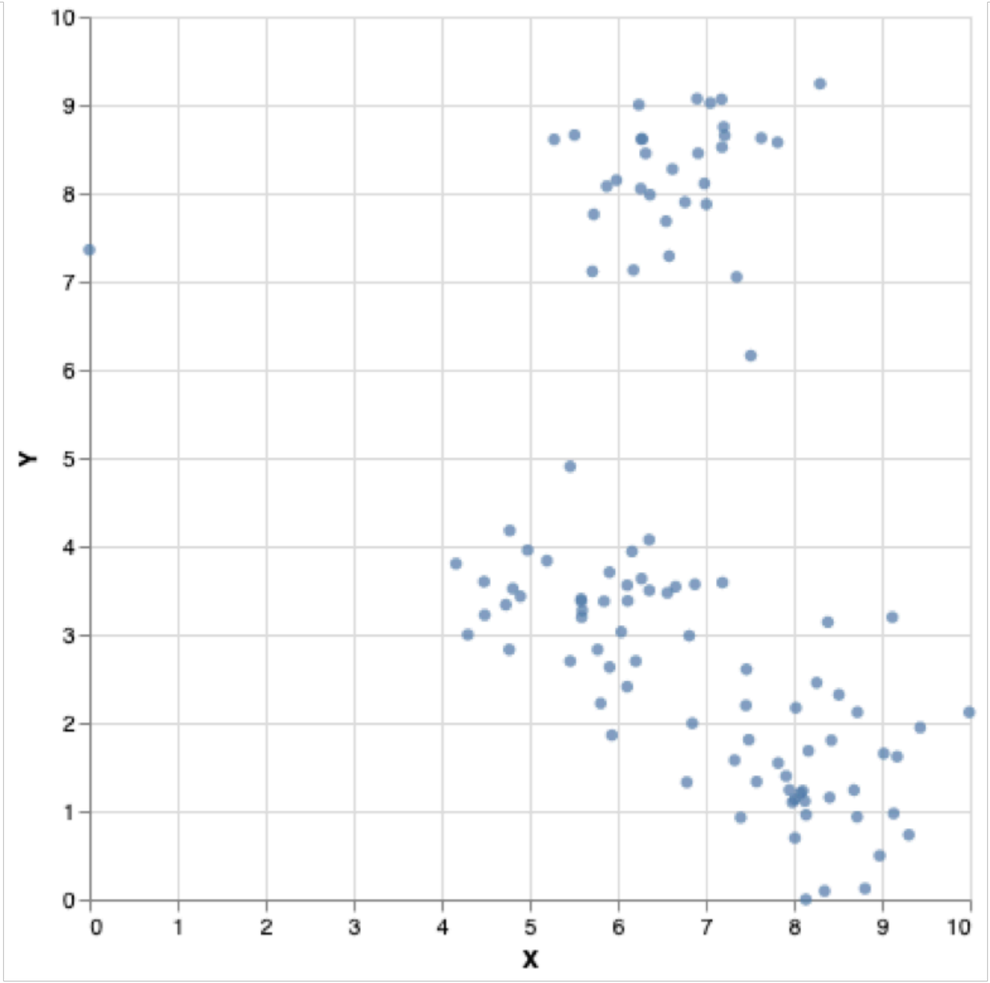
	Label	X	Y
3	P52	6.58351	7.28796
5	P171	4.77421	4.17980
8	P199	8.34966	0.09550
9	P183	8.42670	1.80299
10	P61	4.29760	2.99981
...	...	...	...
141	P138	7.35179	7.05215
142	P46	6.62171	8.27311

BEFORE AND AFTER

150 rows × 3 columns



108 rows × 3 columns



TOWARDS LITERATE & REUSABLE VISUALIZATION

With tracking, we can make interactive visualization reproducible.

Makes it possible to reuse interactive analysis processes on updated datasets.

And seamlessly integrate computational and interactive workflows

Alexander Lex
@alexander_lex
http://alexander-lex.net

Thanks to: **Devin Lange, Tom Zangle, Robert Judson-Torres, Kiran Gadhave, Zach Cutler**, Carolina Nobre, Marc Streit, Jochen Görtler, Oliver Deussen, Miriah Meyer, Jeff Phillips, Samuel Gratzl, Holger Stitz, Nils Gehlenborg, Hendrik Strobelt, Romain Vuillemot, Hanspeter Pfister, and many others!

visualization
design lab