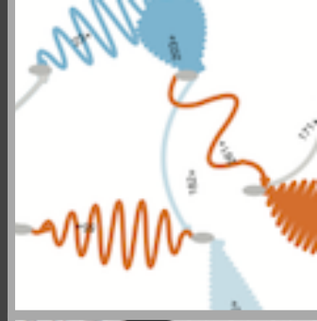
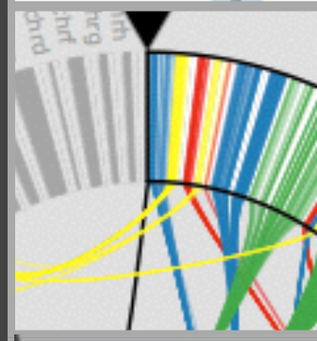
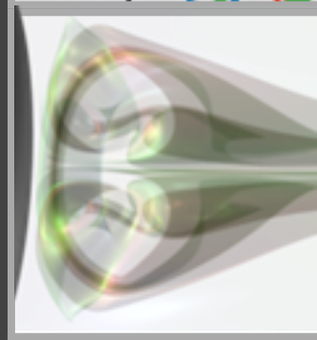
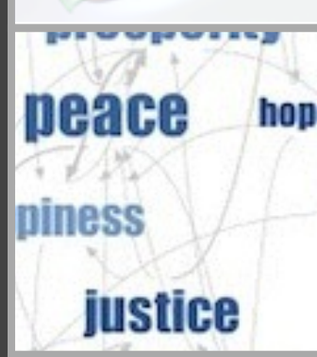
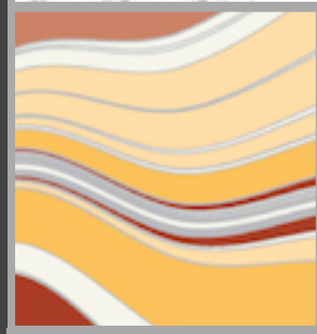
cs6630 | October 2 2014

FILTERING & AGGREGATION

Miriah Meyer
University of Utah

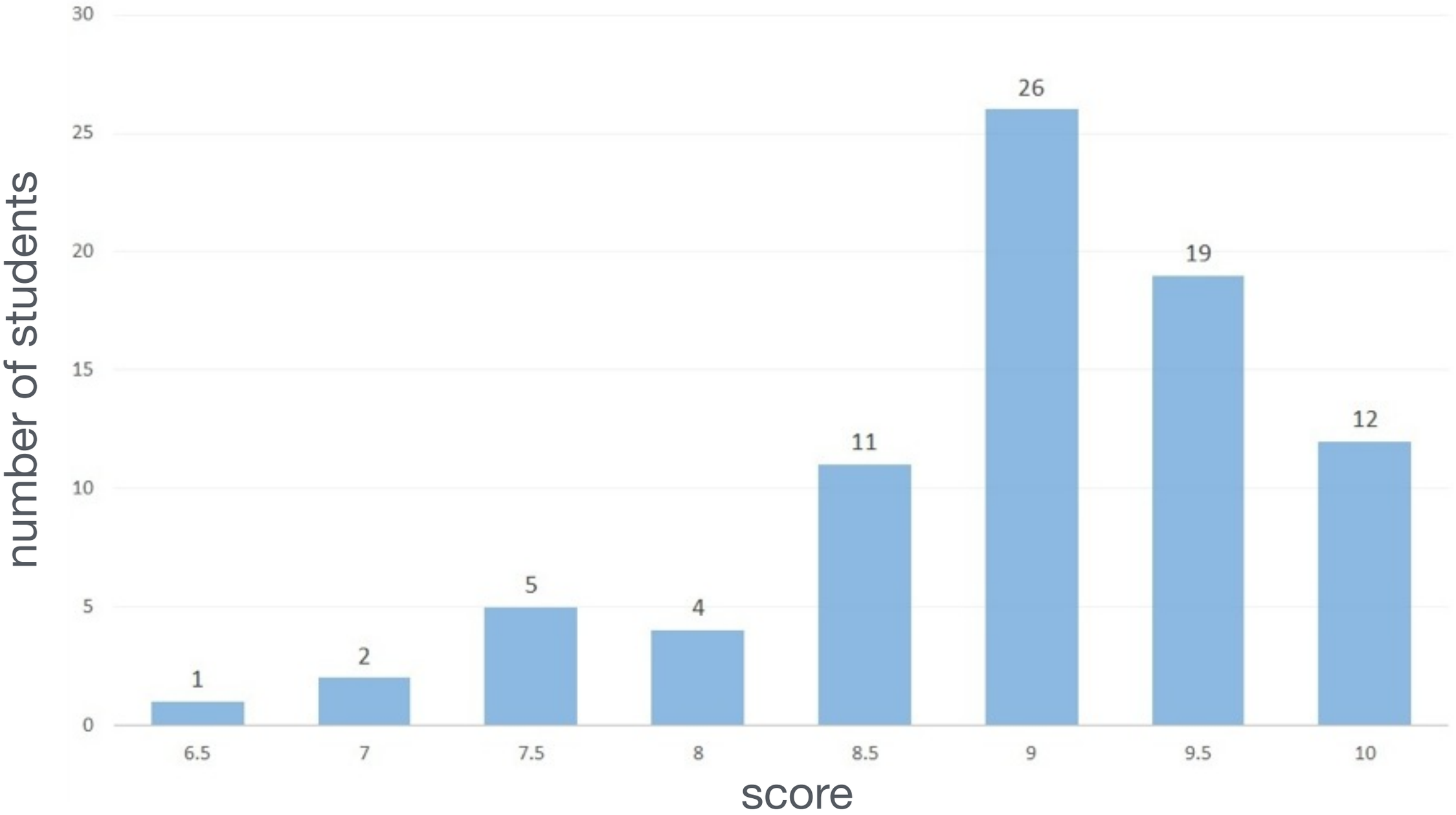








administrivia . . .

- exam on Tuesday
- data exploration grades are in

data exploration assignment



last time . . .

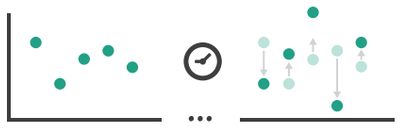
FOCUS + CONTEXT

carefully pick what to show

hint at what you are not showing

Manipulate

Change

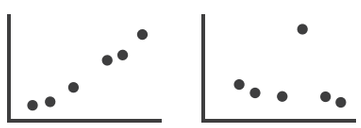


Select

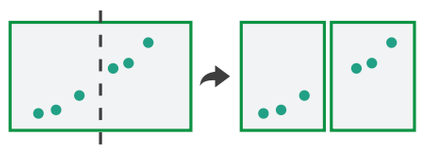
Navigate

Facet

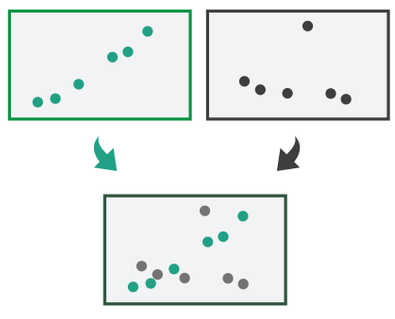
Juxtapose



Partition



Superimpose



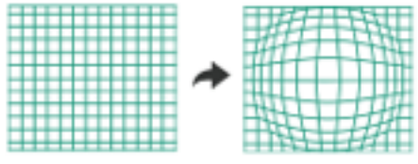
Reduce

Embed

- Elide Data

- Superimpose Layer

- Distort Geometry



what is elision?

what is elision?

focus items shown in detail, other items summarized for context

- superimpose

superimpose

focus layer limited to a local region of view, instead of stretching across the entire view

distort

distort

use geometric distortion of the contextual regions to make room for the details in the focus region(s)

distortion concerns

- **unsuitable for relative spatial judgements**
- **overhead of tracking distortion**
- **visual communication of distortion**
  - gridlines, shading
- **target acquisition problem**
  - lens displacing items away from screen location
- **mixed results compared to separate views and temporal navigation**
- **fisheye follow-up: concern with enthusiasm over distortion**
  - *what* is being shown: selective filtering
  - *how* it is being shown: distortion as one possibility

today . . .

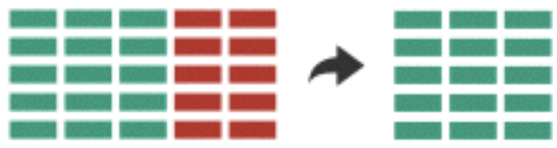
- filtering
- aggregation

Reducing Items and Attributes

Filter

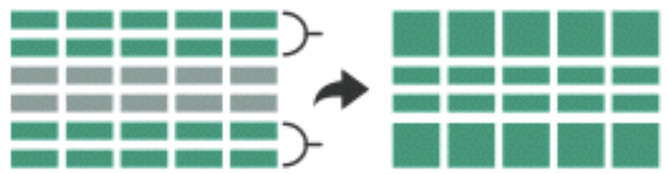
- Items

- Attributes

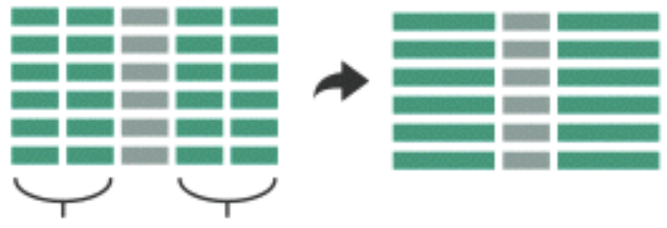


Aggregate

- Items



- Attributes



why reduce?

filter vs aggregation

filter

elements are eliminated

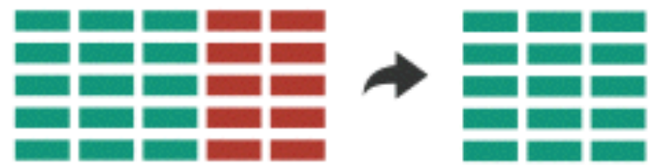
→ Items

→ Attributes

filter
elements are eliminated

- Items

- Attributes



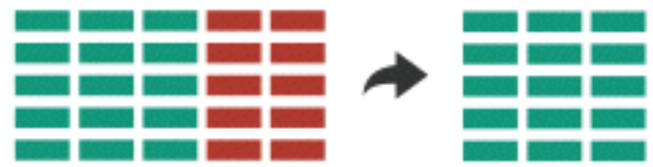
dynamic queries

filter

elements are eliminated

- Items

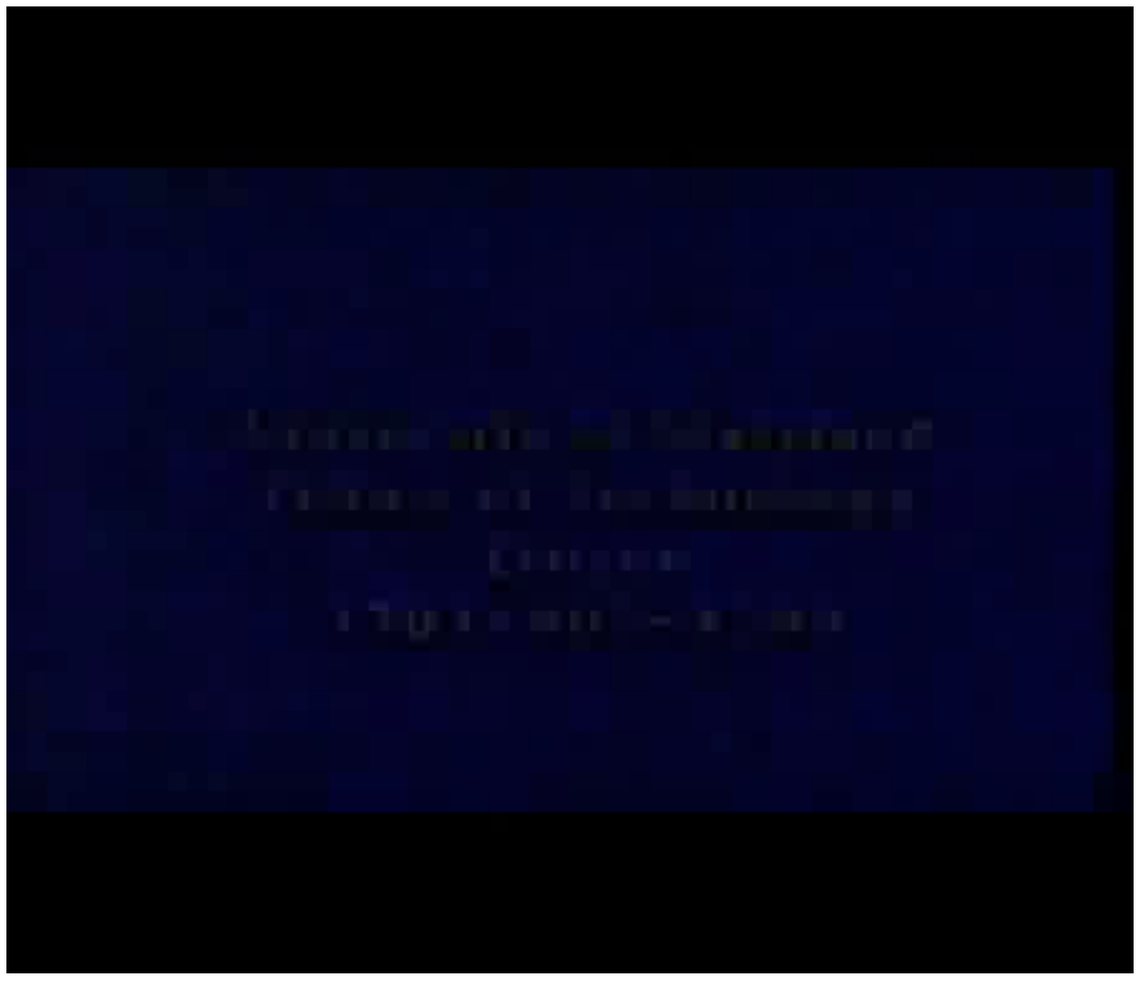
- Attributes



dynamic queries

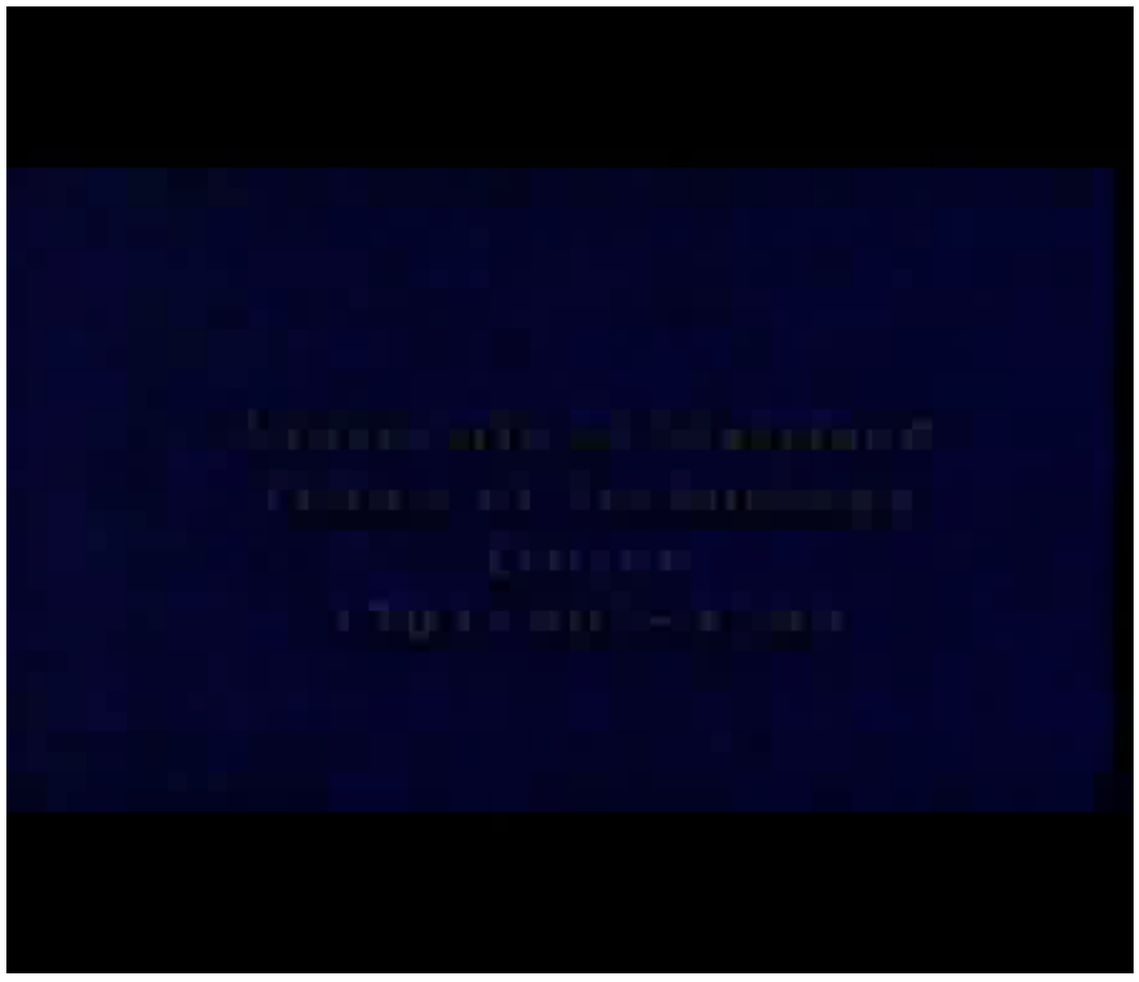
coupling between encoding and interaction so that user can immediately see the results of an action

ITEM FILTERING



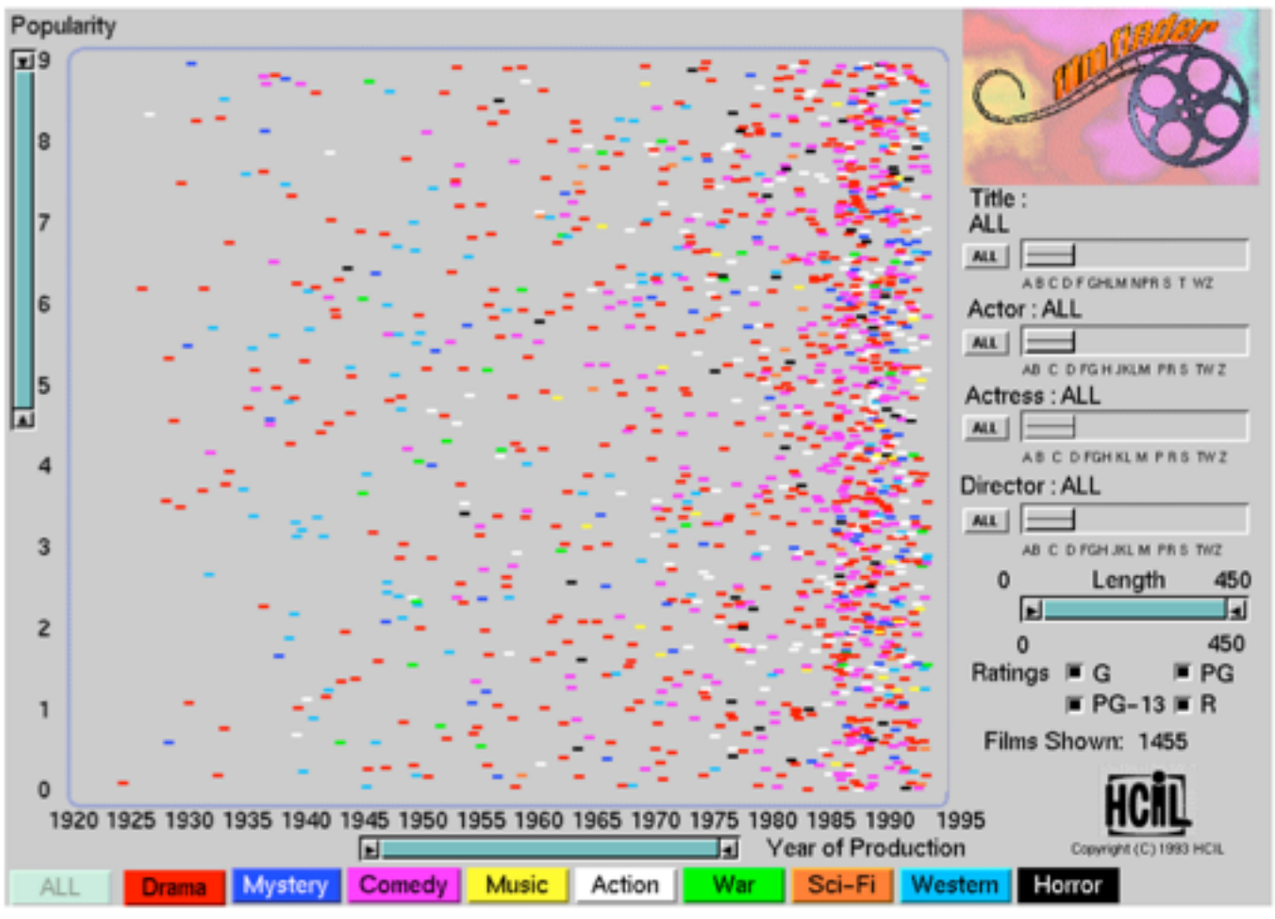
Ahlberg 1994

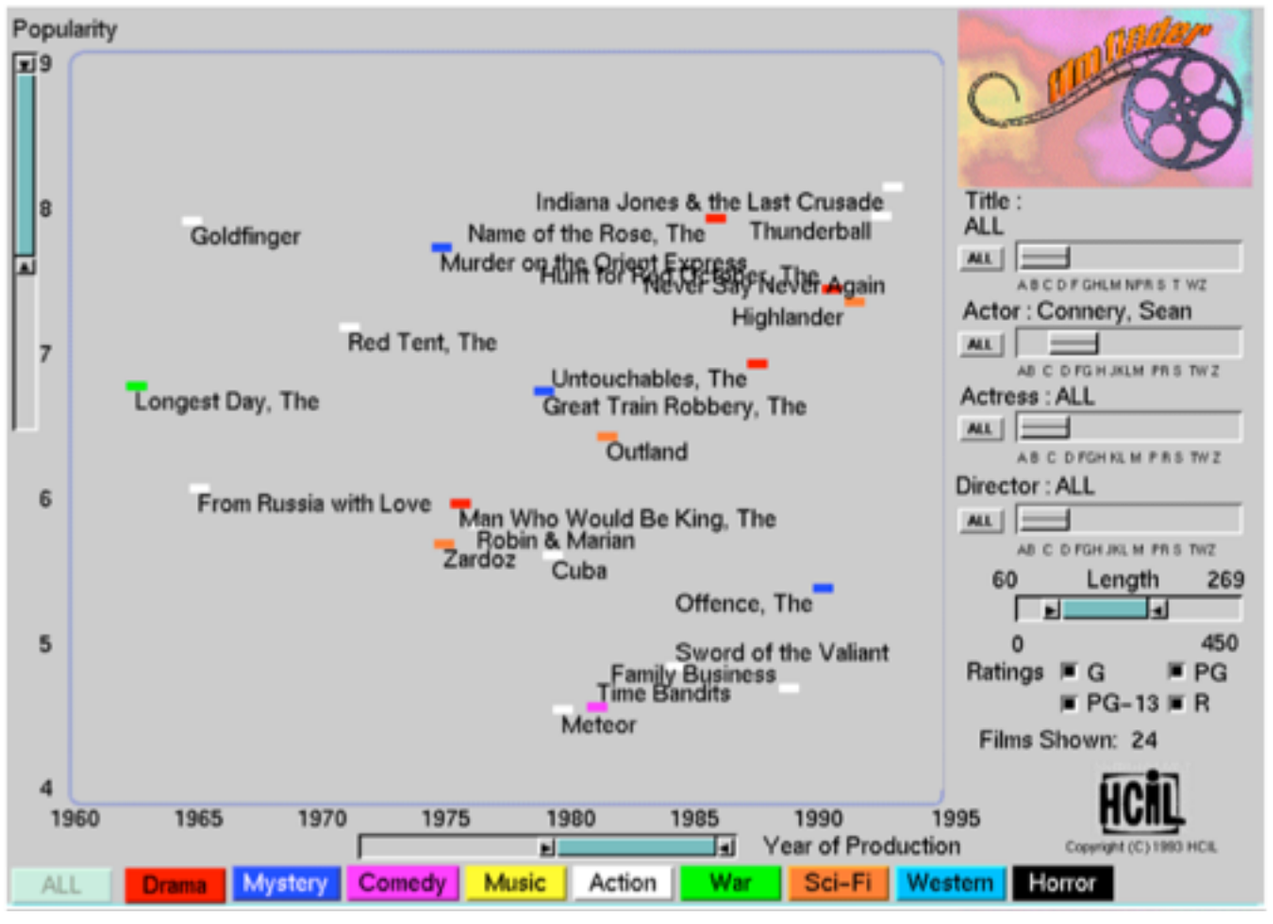
ITEM FILTERING



Ahlberg 1994

ITEM FILTERING





Ahlberg 1994

New York Health Department Restaurant Ratings Map - Interactive Map - NYTimes.com

http://www.nytimes.com/interactive/dining/new-york-health-department-restaurant-ratings-map.html?ref

nytimes interactive graphics

Google Maps Wikipedia UT hiking

UPDATED June 25, 2012

New York Health Department Restaurant Ratings Map

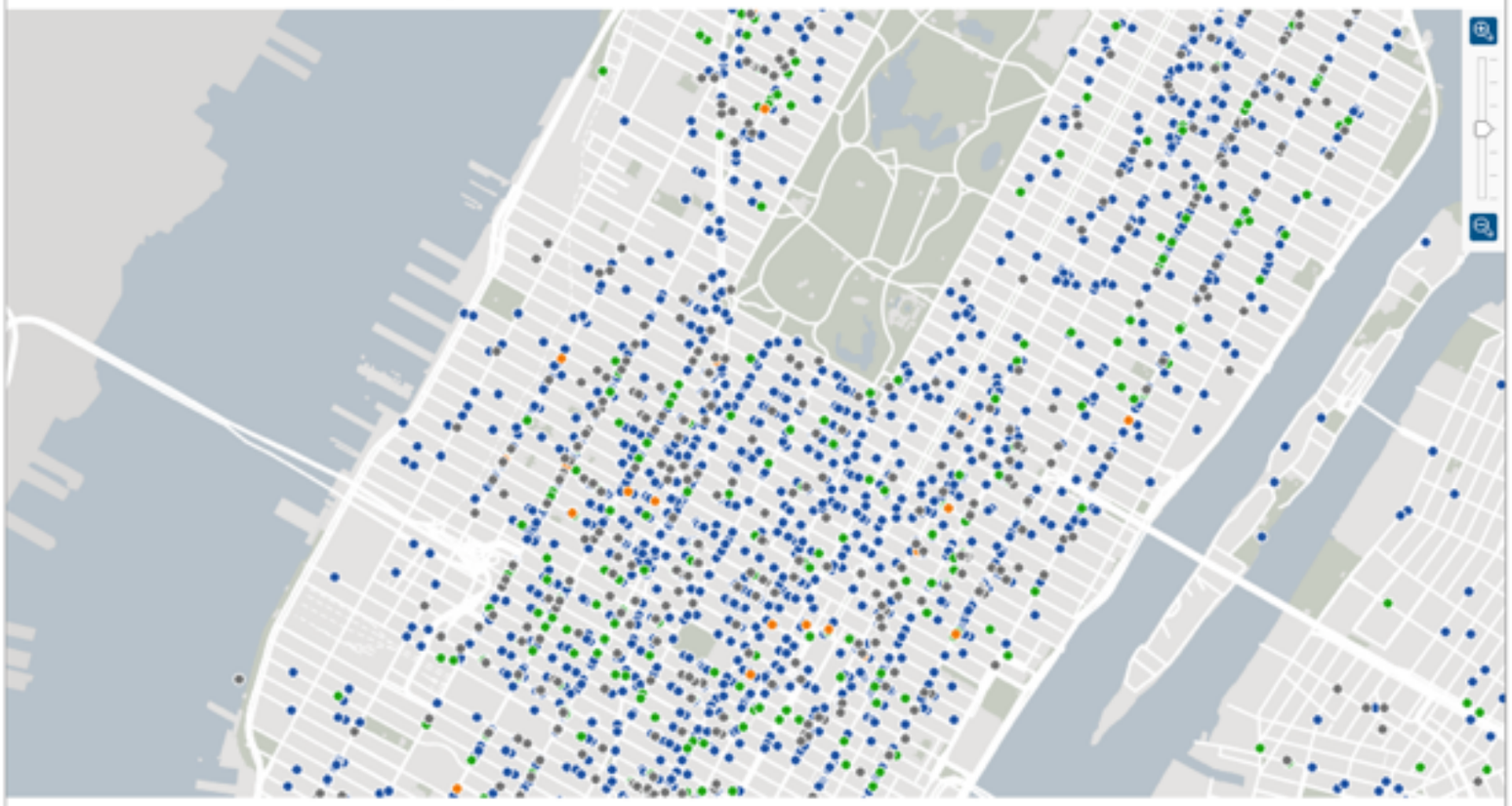
The New York City Department of Health and Mental Hygiene performs unannounced sanitary inspections of every restaurant at least once per year. Violation points result in a letter grade, which can be explored in the map below, along with violation descriptions. The information on this map will be updated every two weeks. For menus and reviews by New York Times critics, visit our restaurants guide. Related Article »

FIND A RESTAURANT | FIND A LOCATION

Name of restaurant

FILTER

All grades | All violations | All cuisines



Restaurant locations are derived from the New York City Department of Health and Mental Hygiene database. Due to the limitations of the Health Department's database, some restaurants could not be placed.

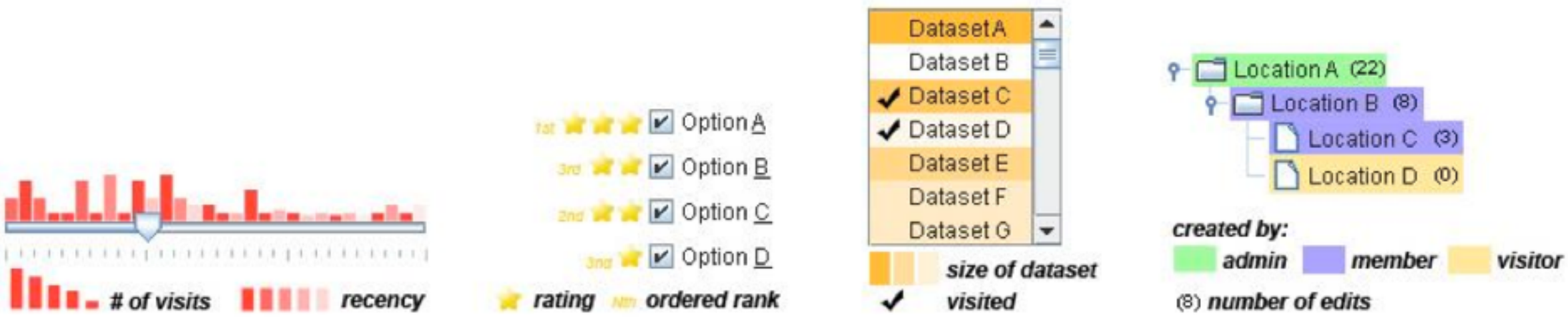
By JEREMY WHITE | Send Feedback

Source: New York City Department of Health and Mental Hygiene

scented widgets

information scent: user's (imperfect) perception of data

GOAL: lower the cost of information foraging through better cues

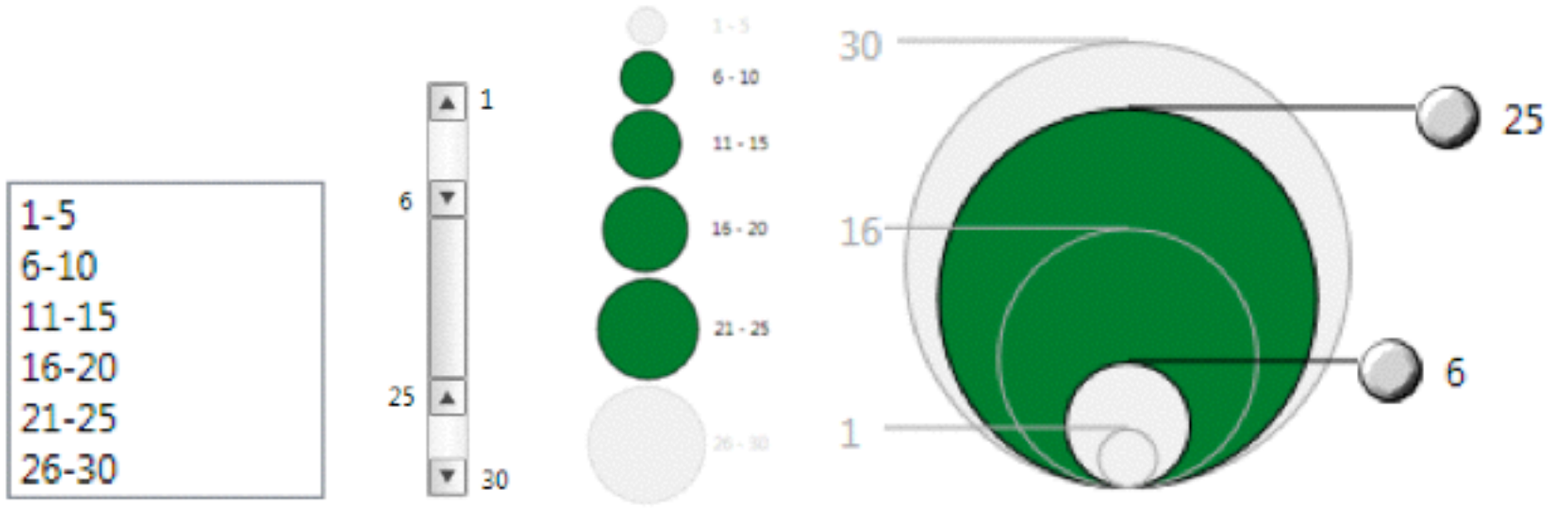


Willett 2007

interactive legends

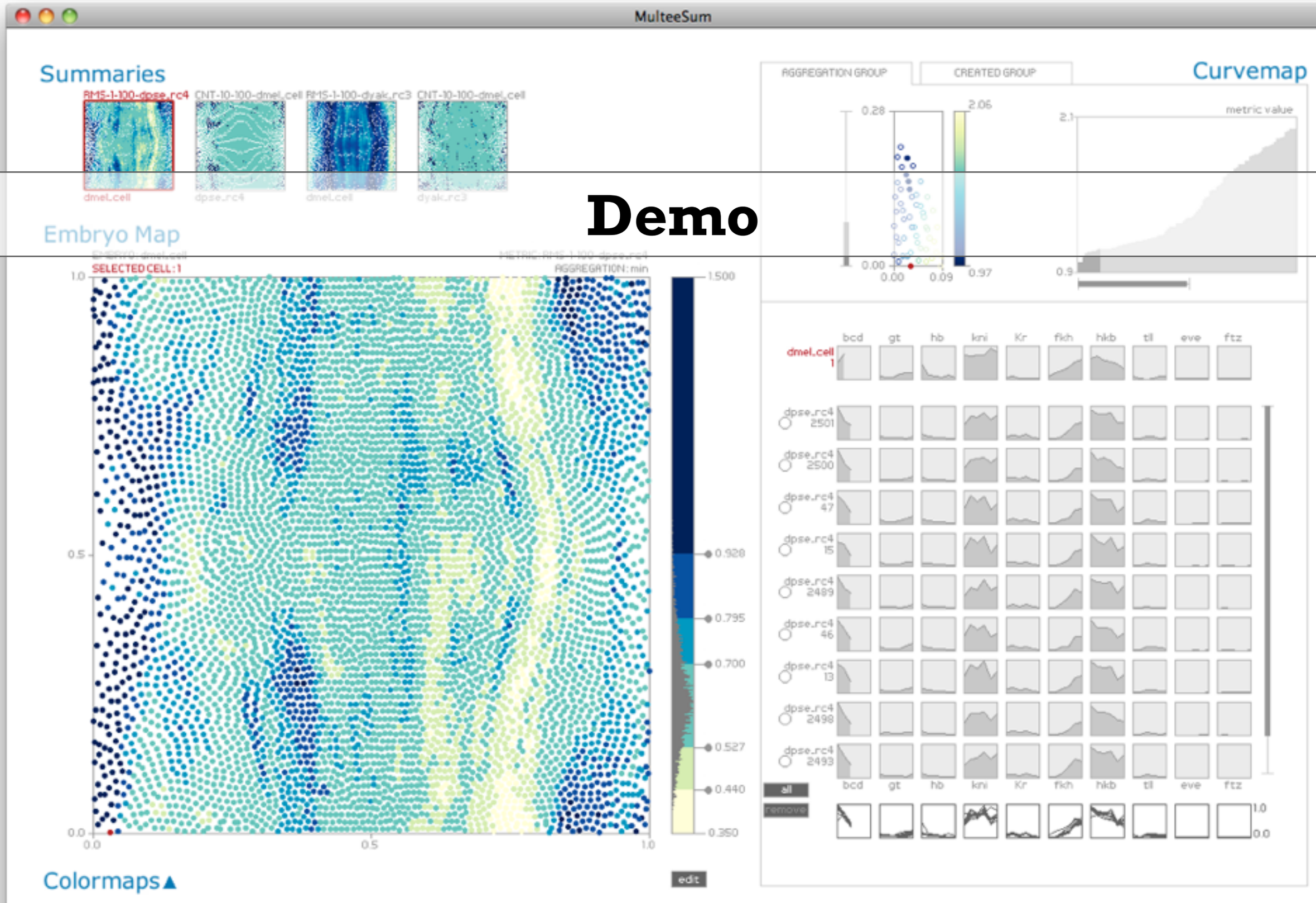
controls combining the visual representation of static legends with interaction mechanisms of widgets

define and control visual display together

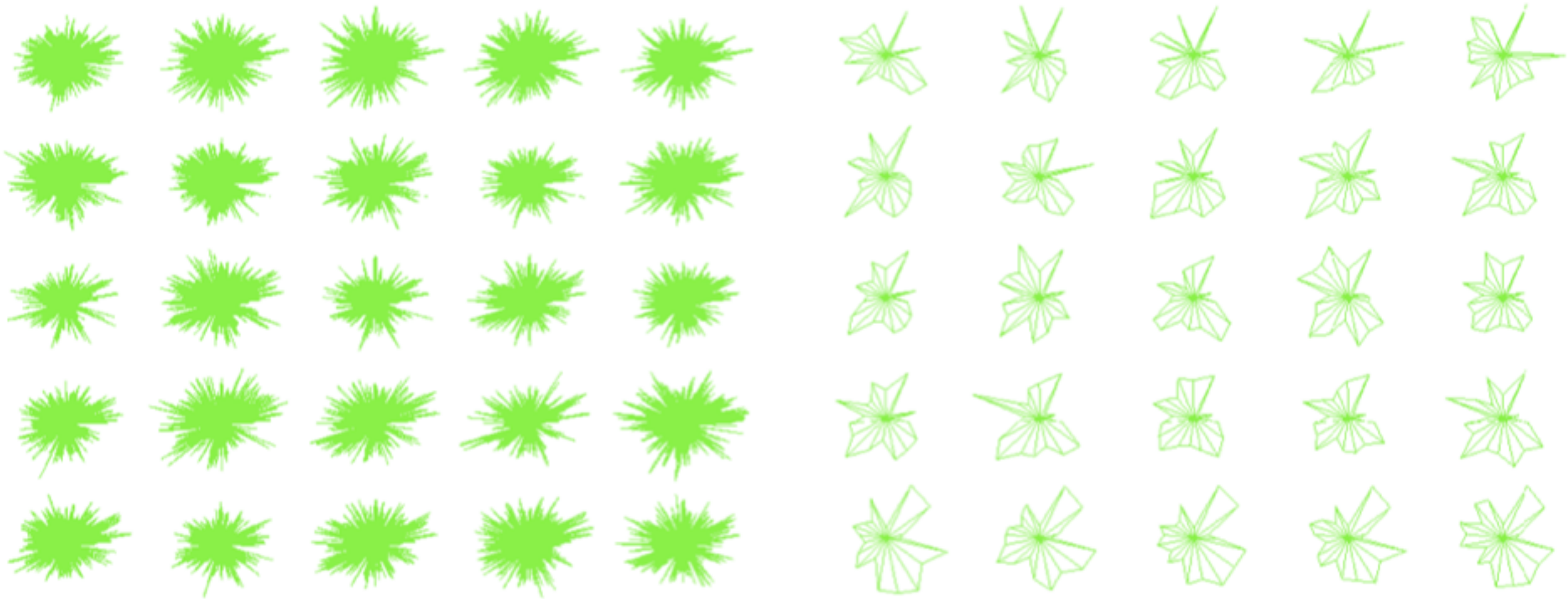


Riche 2010

Demo


MulteeSum
Summaries
RM5-1-100-dpse_rc4
CNT-10-100-dmel_cell
RM5-1-100-dyak_rc3
CNT-10-100-dmel_cell
dmel_cell
dpse_rc4
dmel_cell
dyak_rc3
Embryo Map
SELECTED CELL: 1
METRIC: RM5-1-100-dpse_rc4
AGGREGATION: min
1.0
0.5
0.0
0.0
0.5
1.0
1.500
0.920
0.795
0.700
0.527
0.440
0.350
edit
Colormaps▲
AGGREGATION GROUP
CREATED GROUP
Curvemap
0.28
2.06
0.00
0.00
0.09
0.97
metric value
2.1
0.9
bcd
gt
hb
kni
Kr
fkh
hkb
tll
eve
ftz
dmel_cell 1
dpse_rc4 2501
dpse_rc4 2500
dpse_rc4 47
dpse_rc4 15
dpse_rc4 2489
dpse_rc4 46
dpse_rc4 13
dpse_rc4 2498
dpse_rc4 2493
all
remove
1.0
0.0

ATTRIBUTE FILTERING

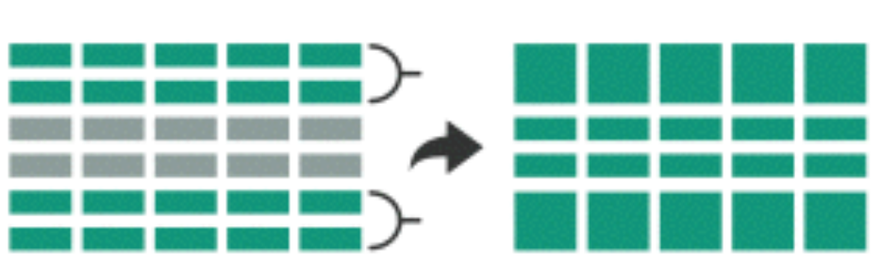


Wang 2003

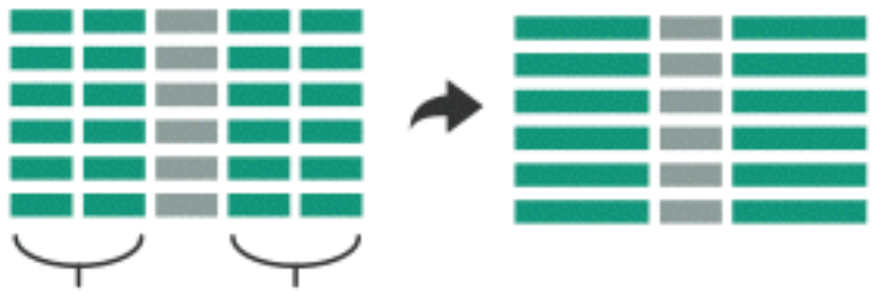
aggregate

a group of elements is represented by a new derived element that stands in for the entire group

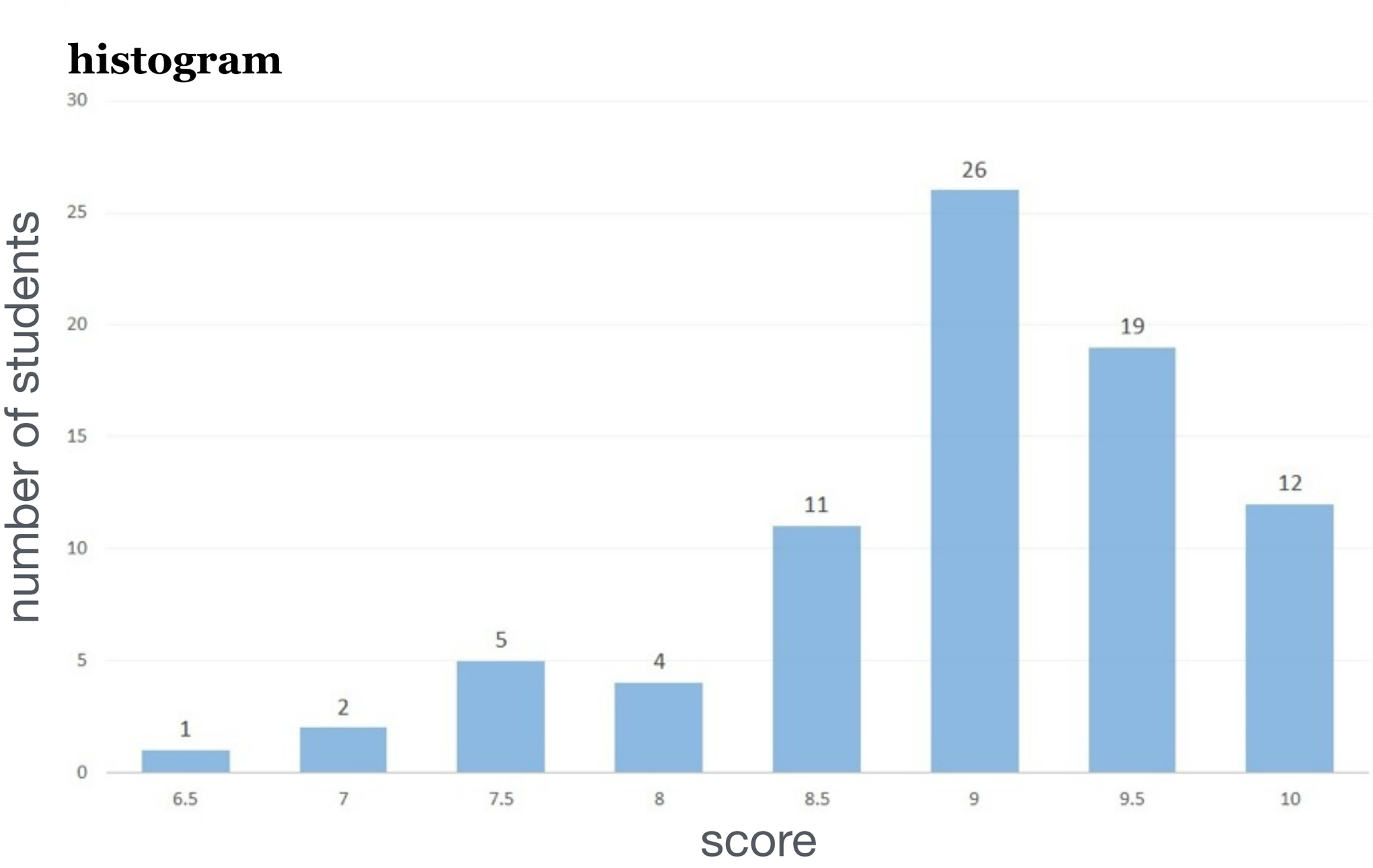
→ Items



→ Attributes

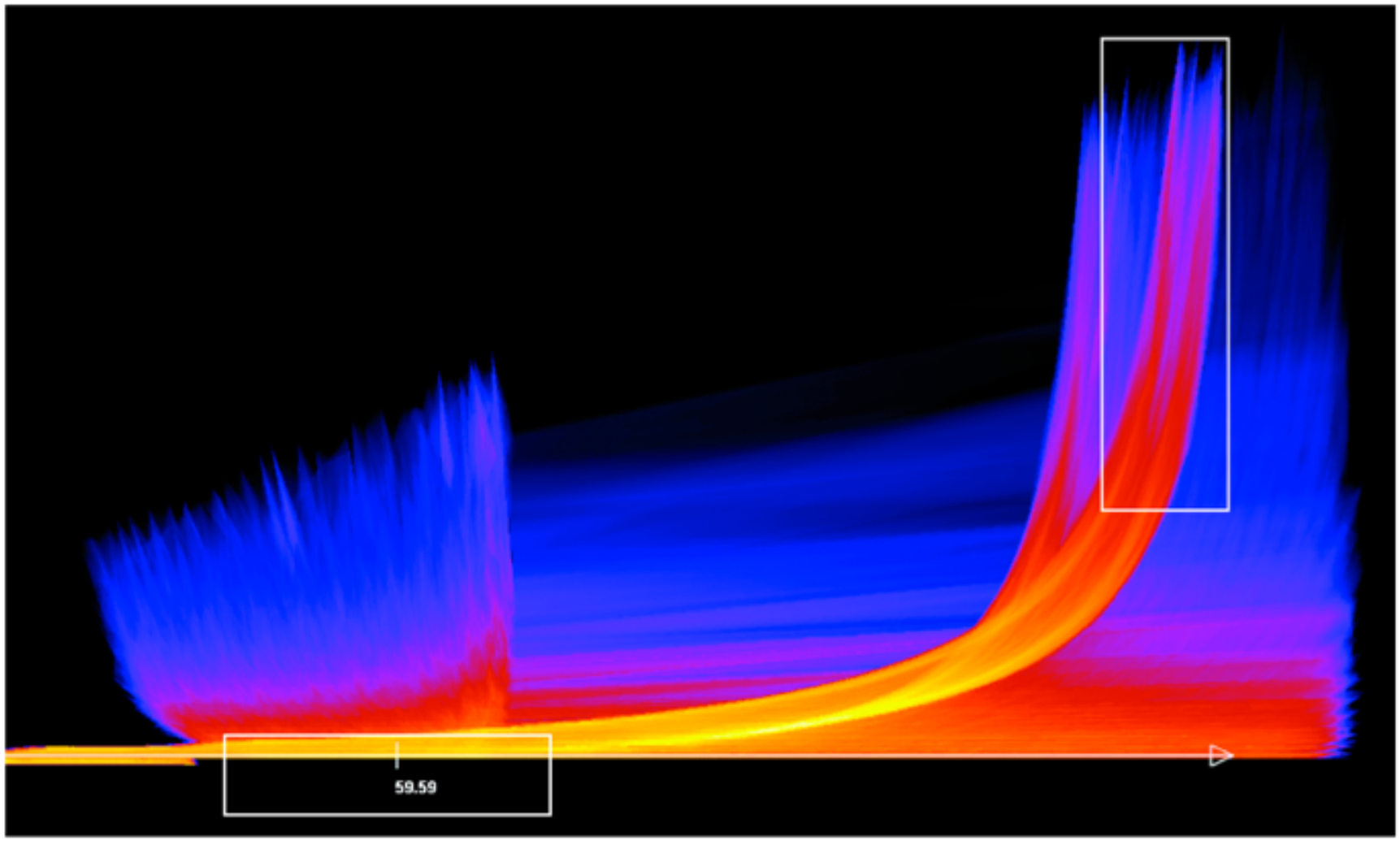


item aggregation



item aggregation

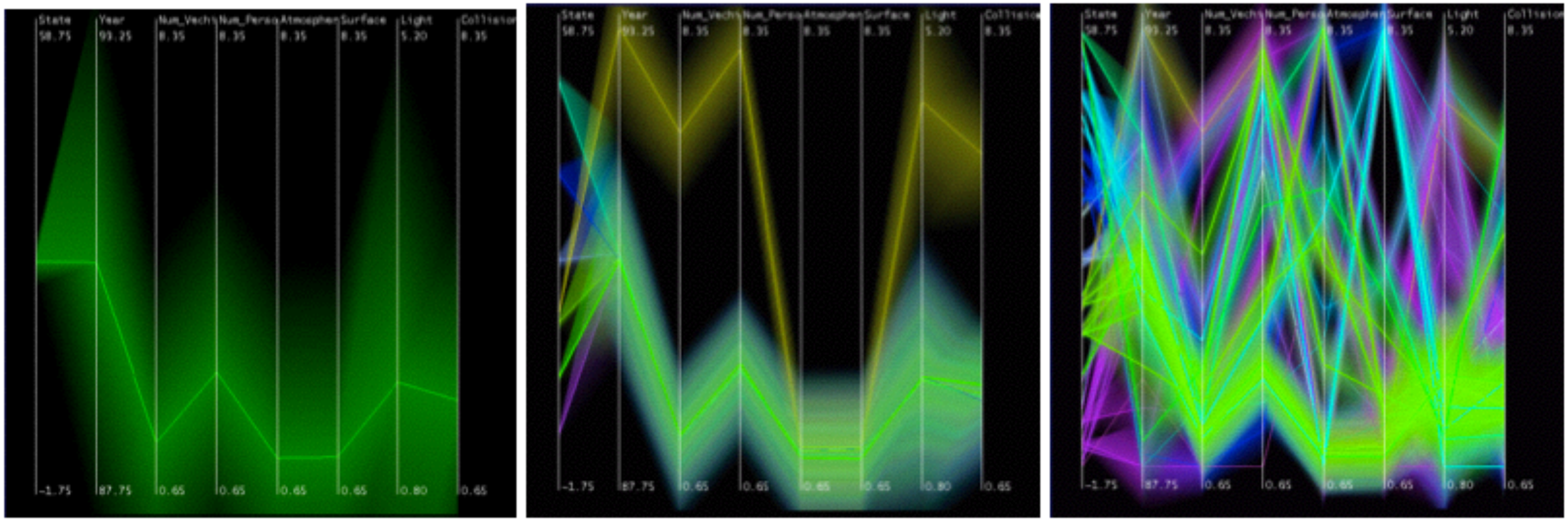
continuous scatterplot



Bachthaler 2008

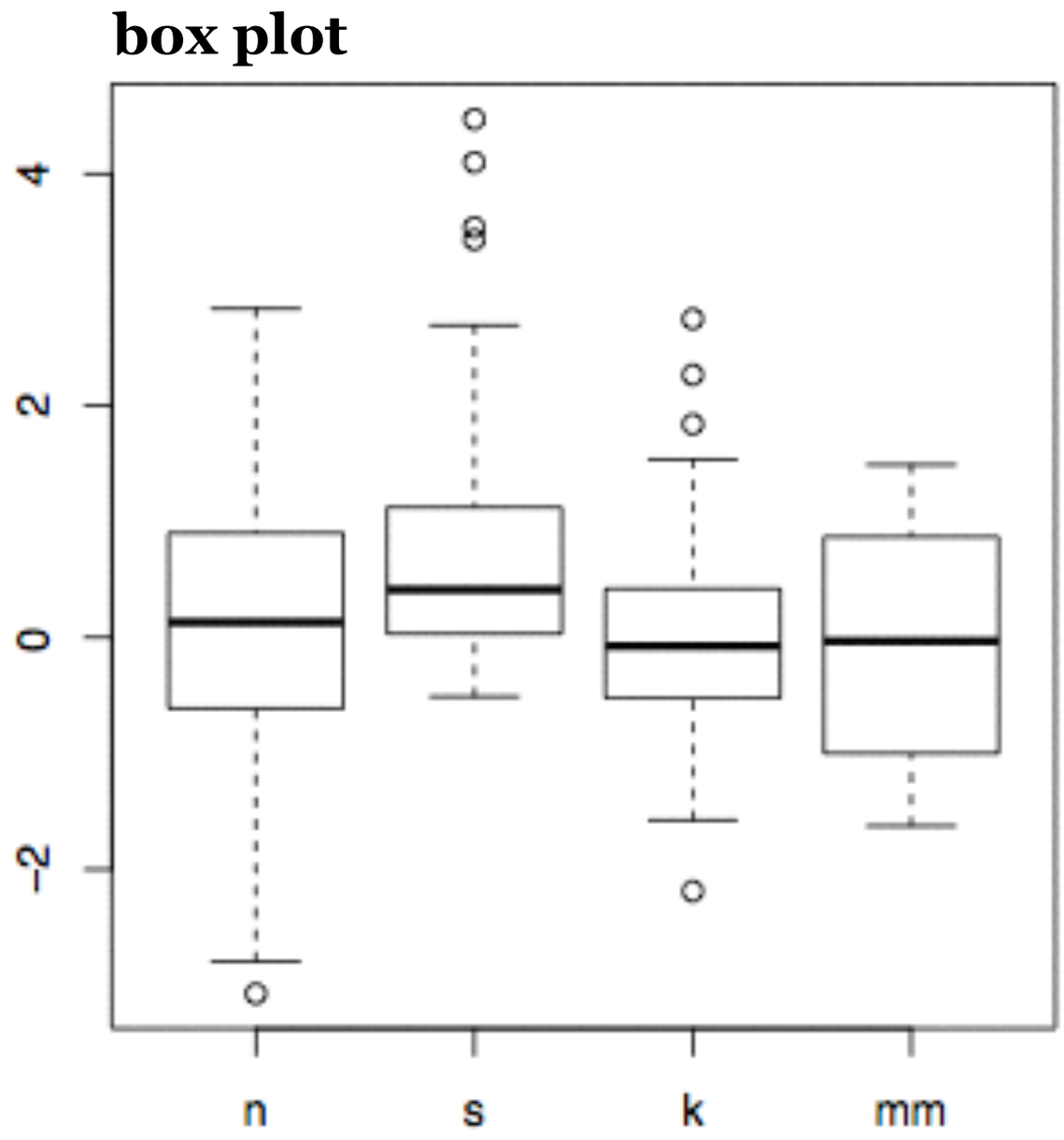
item aggregation

hierarchical parallel coordinates

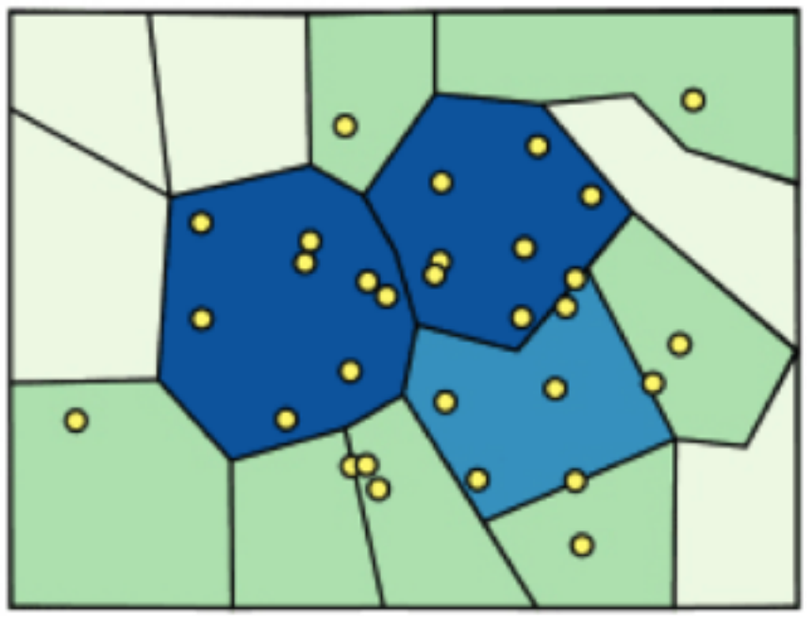


Fua 1999

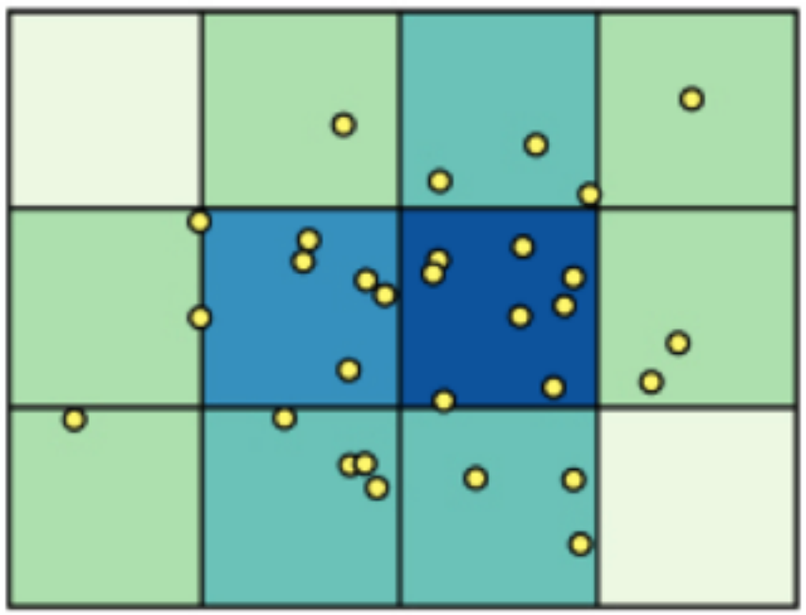
item aggregation

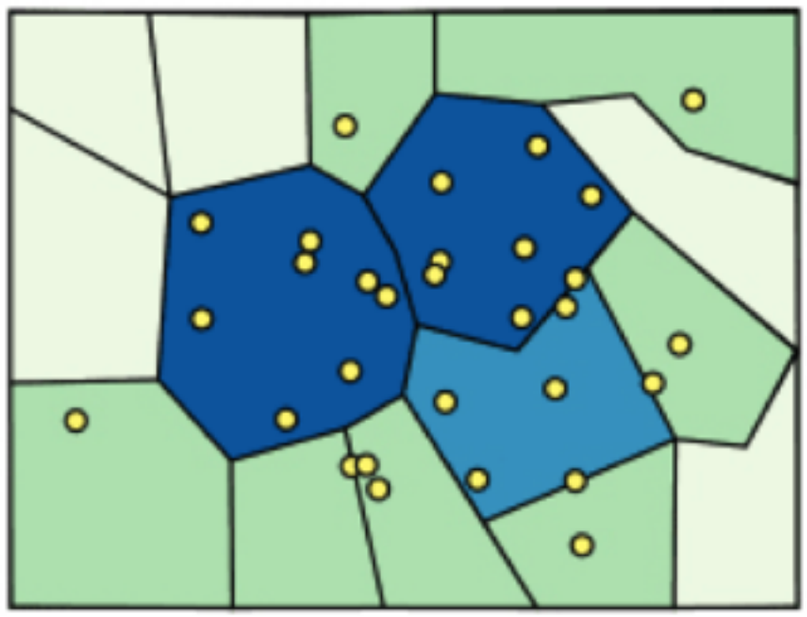


spatial aggregation

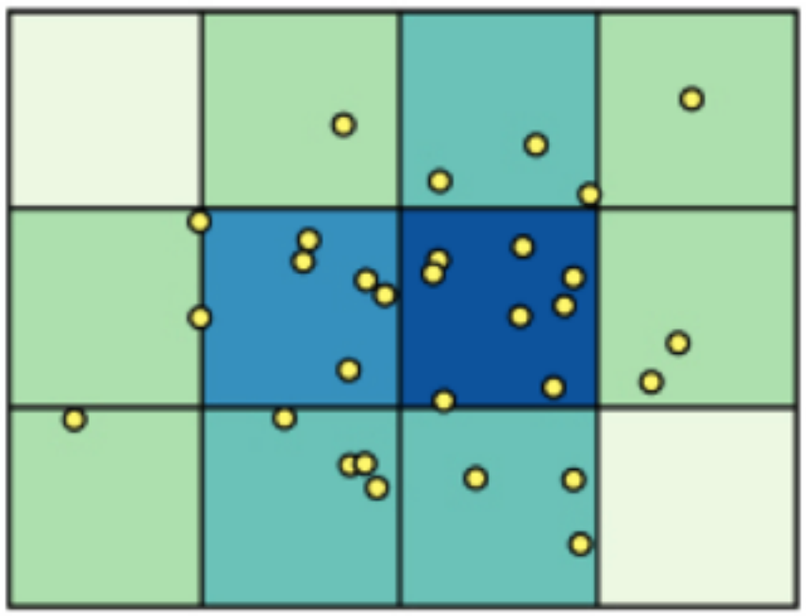


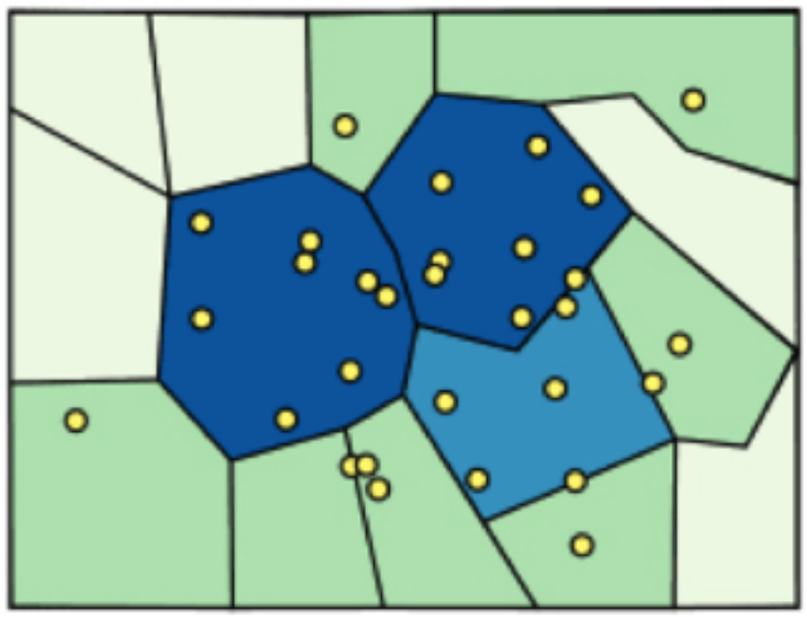
spatial aggregation



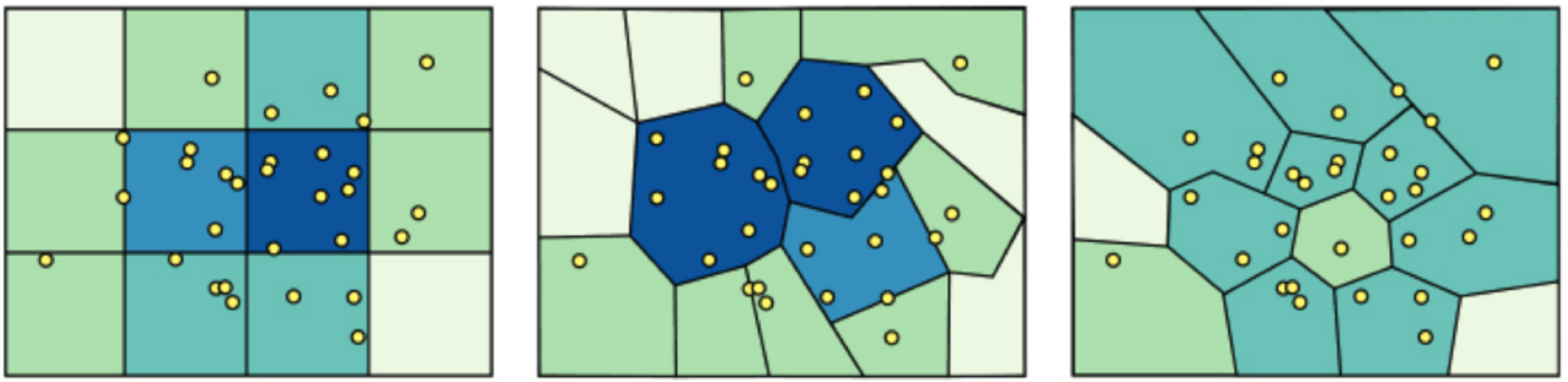


spatial aggregation





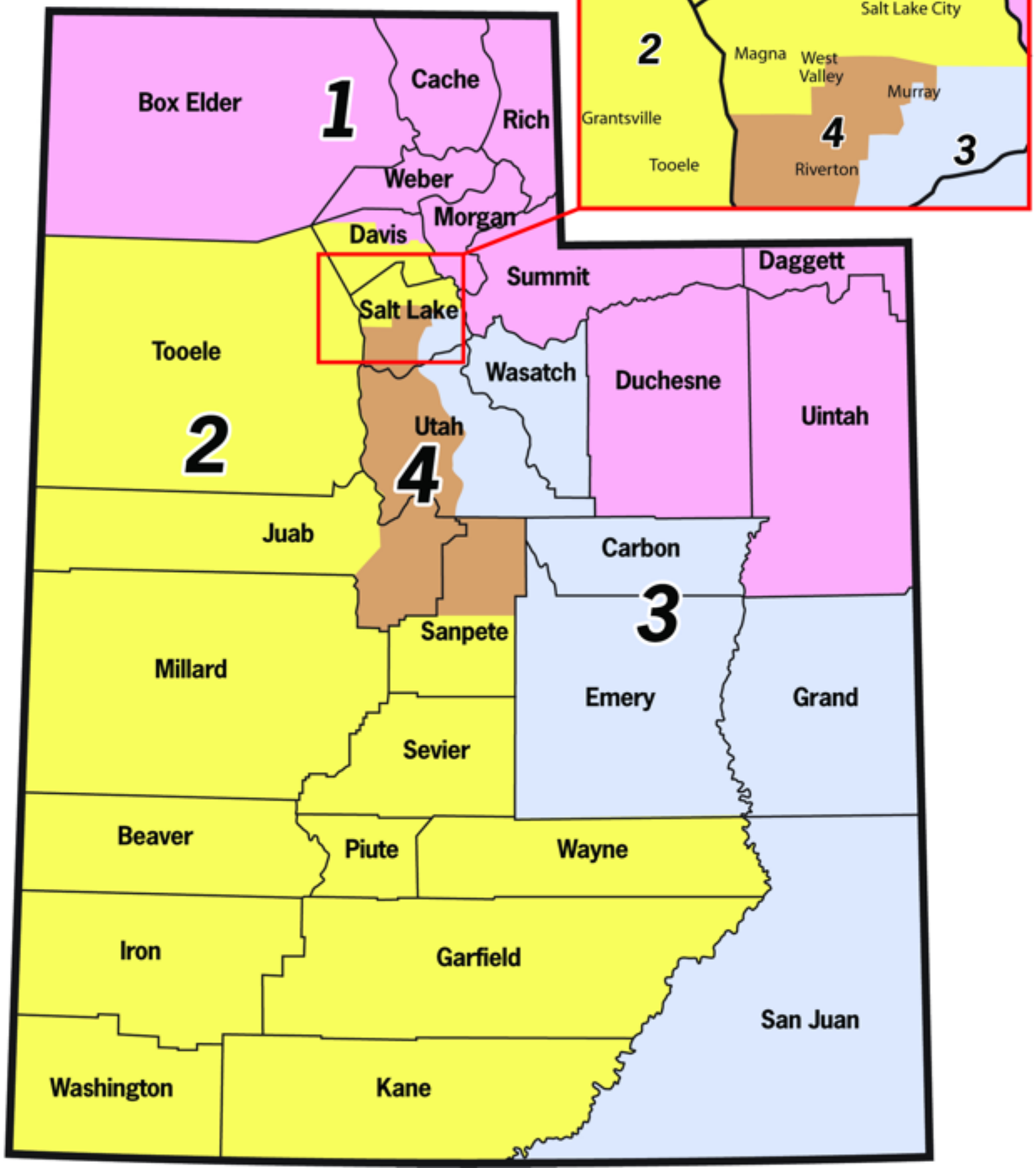
spatial aggregation



modifiable areal unit problem

in cartography, changing the boundaries of the regions used to analyze data can yield dramatically different results

Congressional Districts



attribute aggregation

1) group attributes and compute a similarity score across the set

2) dimensionality reduction, to preserve meaningful structure

attribute aggregation

1) group attributes and compute a similarity score across the set

2) dimensionality reduction, to preserve meaningful structure

MulteeSum

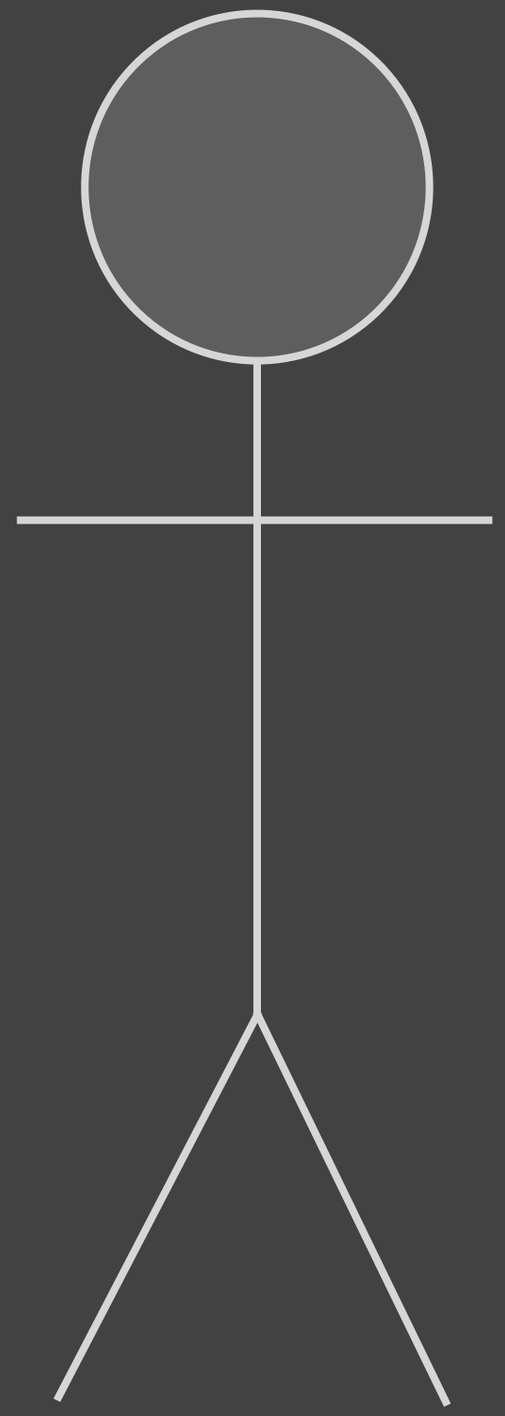
A Tool for Comparative Spatial and Temporal Gene Expression Data

Miriah Meyer,[1] Tamara Munzner,[2] Angela DePace,[3] Hanspeter Pfister[1]

[1] Harvard University

[2] University of British Columbia

[3] Harvard Medical School



TTTAGCACGAA

TTTAGCACGAA

TTTAGCACGAA

TTTAGCACGAA

hair

TTTAGCACGAA

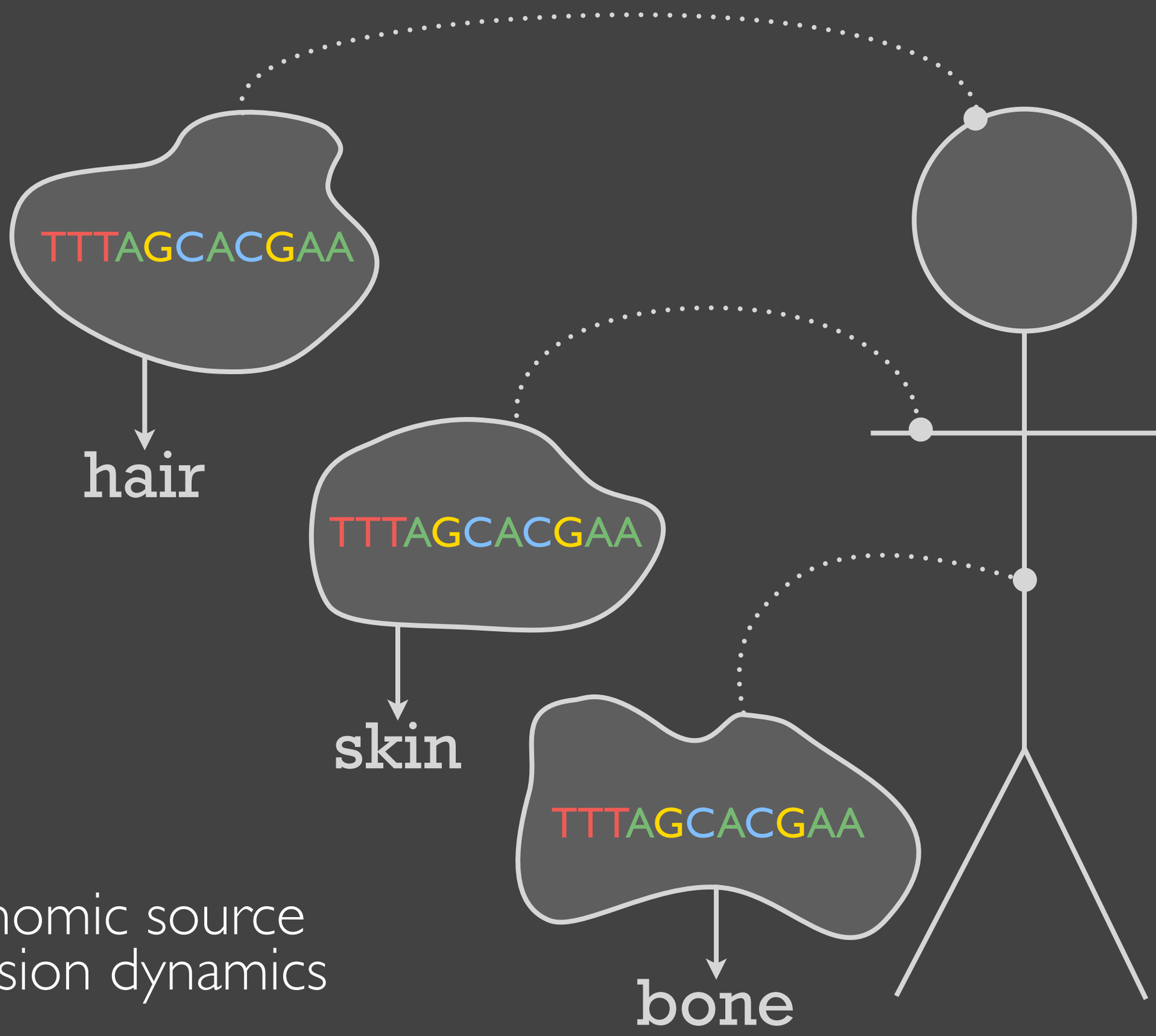
skin

TTTAGCACGAA

bone

grand challenge in biology:

understand genomic source of gene expression dynamics



collaborators: DePace lab at Harvard Medical School

model species: fruit fly

collaborators: DePace lab at Harvard Medical School

model species: fruit fly

scientific goal: link changes in the regulatory part of the genome to species variation

collaborators: DePace lab at Harvard Medical School

model species: fruit fly

scientific goal: link changes in the regulatory part of the genome to species variation

requires: characterize differences in gene expression patterns between species

collaborators: DePace lab at Harvard Medical School

model species: fruit fly

scientific goal: link changes in the regulatory part of the genome to species variation

requires: characterize differences in gene expression patterns between species

MulteeSum

process

two year collaboration

two early prototype systems

feedback from six biologists

informal interviews, emails

one day a week in biology lab

tool deployed

currently used several times a week

data & tool & tasks

summaries & groups

encodings & interaction

conclusions

gene expression is

gene expression is ...

... the measured level of how much a gene is on or off.

gene expression is ...

... the measured level of how much a gene is on or off.

... a single quantitative value.

0.2

gene expression is ...

... the measured level of how much a gene is on or off.

... a single quantitative value.

collaborators measure it ...

... for multiple time points.

gene expression is ...

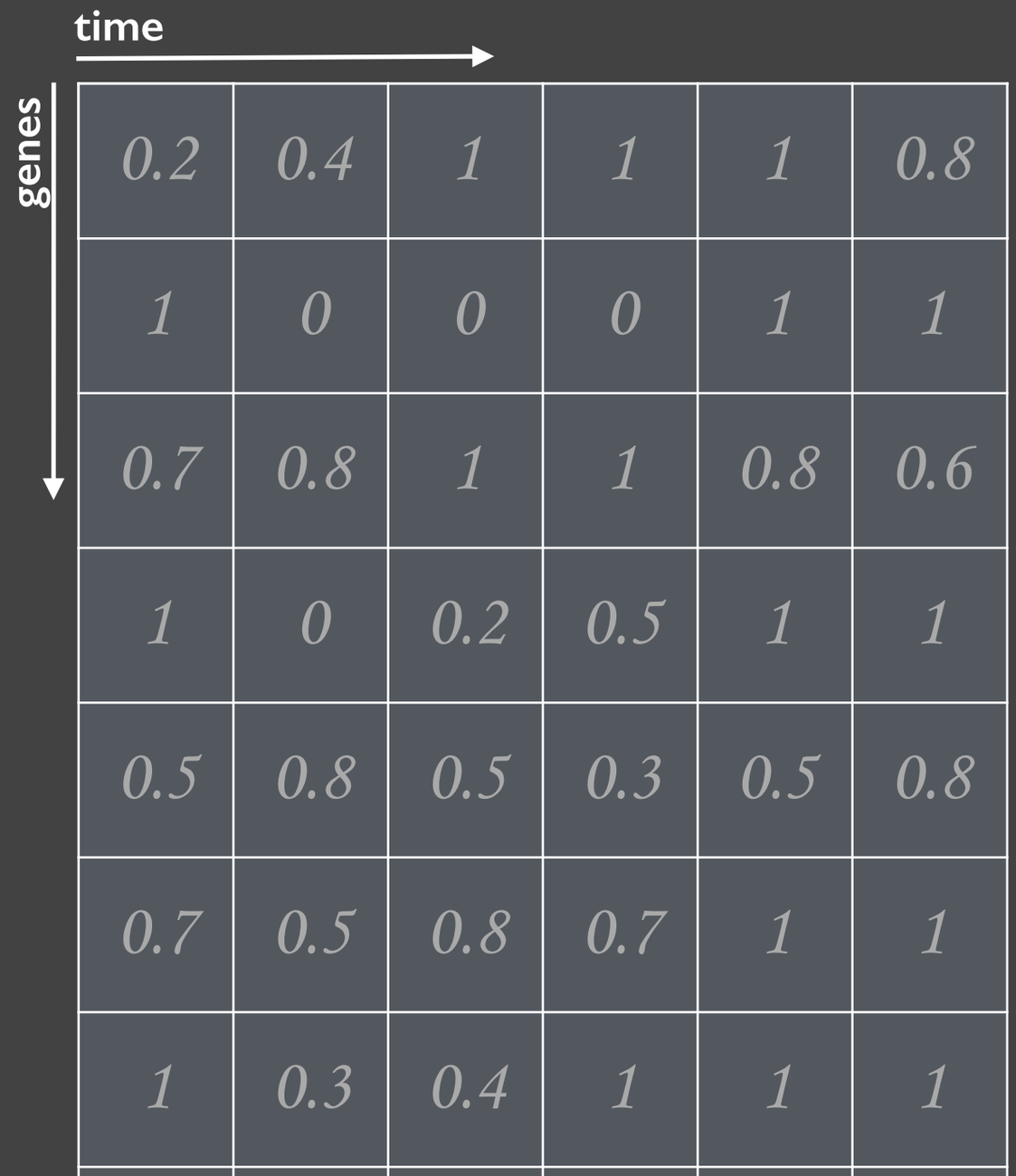
... the measured level of how much a gene is on or off.

... a single quantitative value.

collaborators measure it ...

... for multiple time points.

... for multiple genes.



gene expression is ...

... the measured level of how much a gene is on or off.

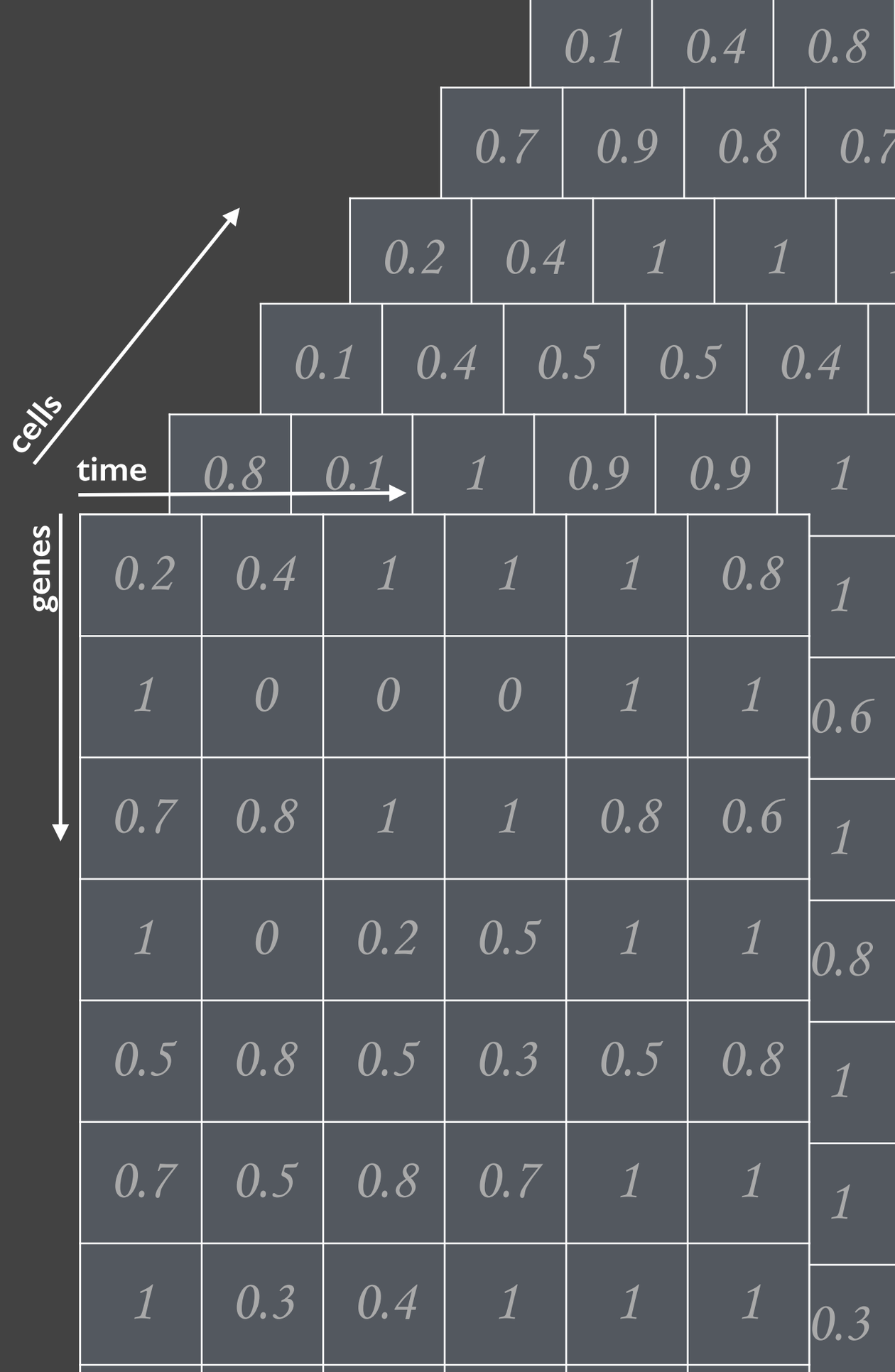
... a single quantitative value.

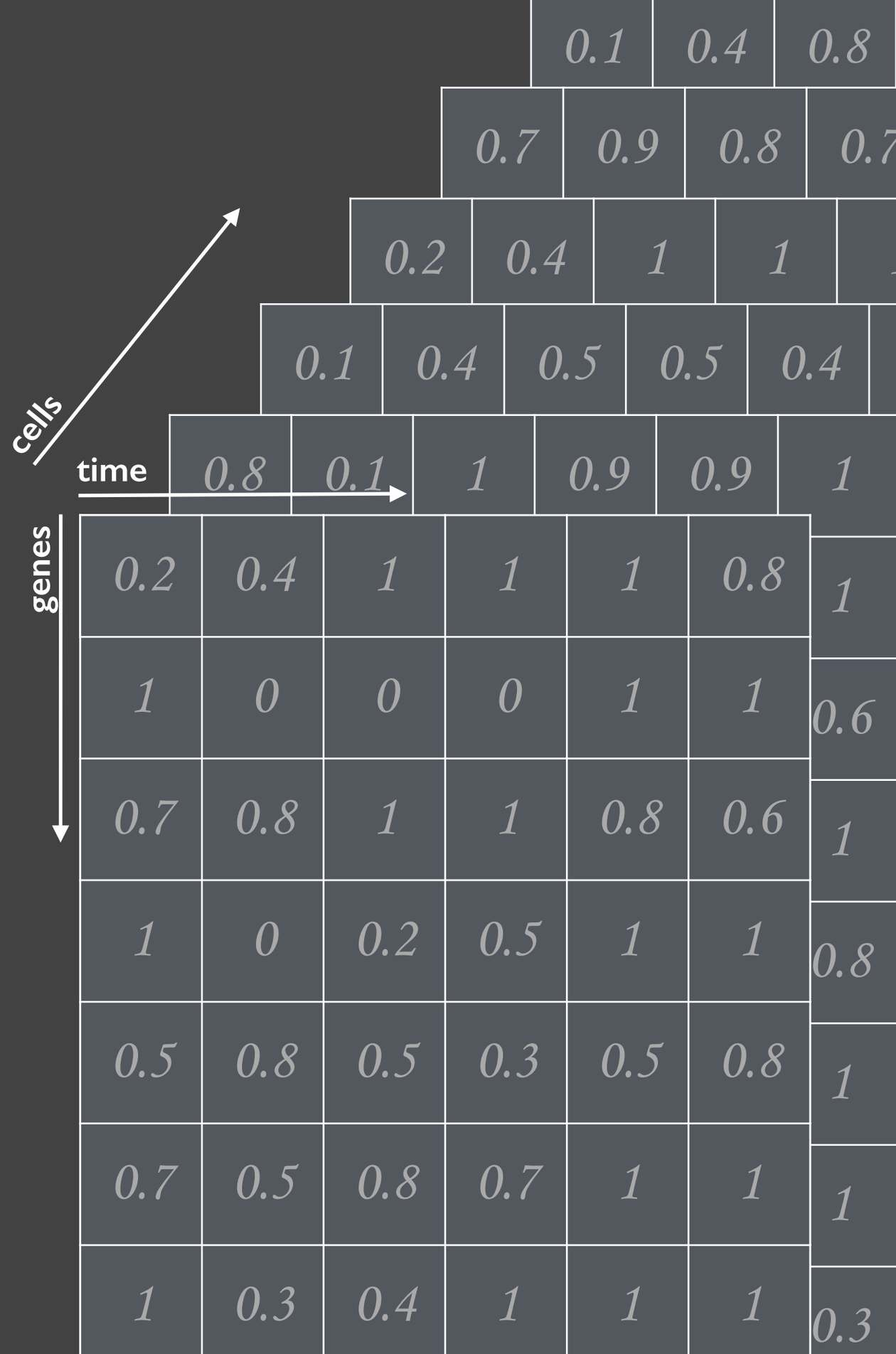
collaborators measure it ...

... for multiple time points.

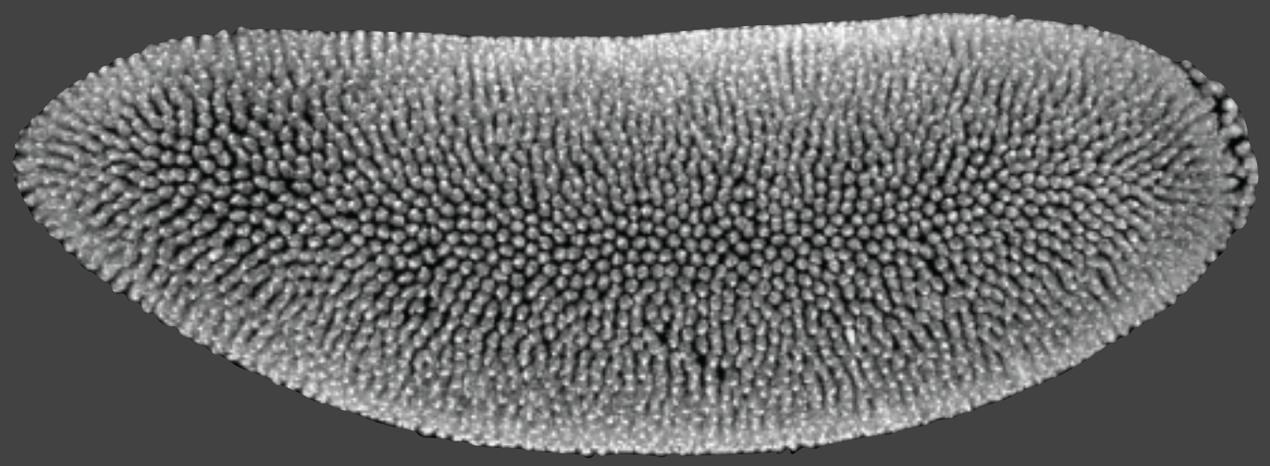
... for multiple genes.

... in many cells.




cells
time
genes
0.1 0.4 0.8
0.7 0.9 0.8 0.7
0.2 0.4 1 1
0.1 0.4 0.5 0.5 0.4
0.8 0.1 1 0.9 0.9 1
0.2 0.4 1 1 1 0.8 1
1 0 0 0 1 1 0.6
0.7 0.8 1 1 0.8 0.6 1
1 0 0.2 0.5 1 1 0.8
0.5 0.8 0.5 0.3 0.5 0.8 1
0.7 0.5 0.8 0.7 1 1 1
1 0.3 0.4 1 1 1 0.3

measurements taken for every cell in a single fruit fly embryo.



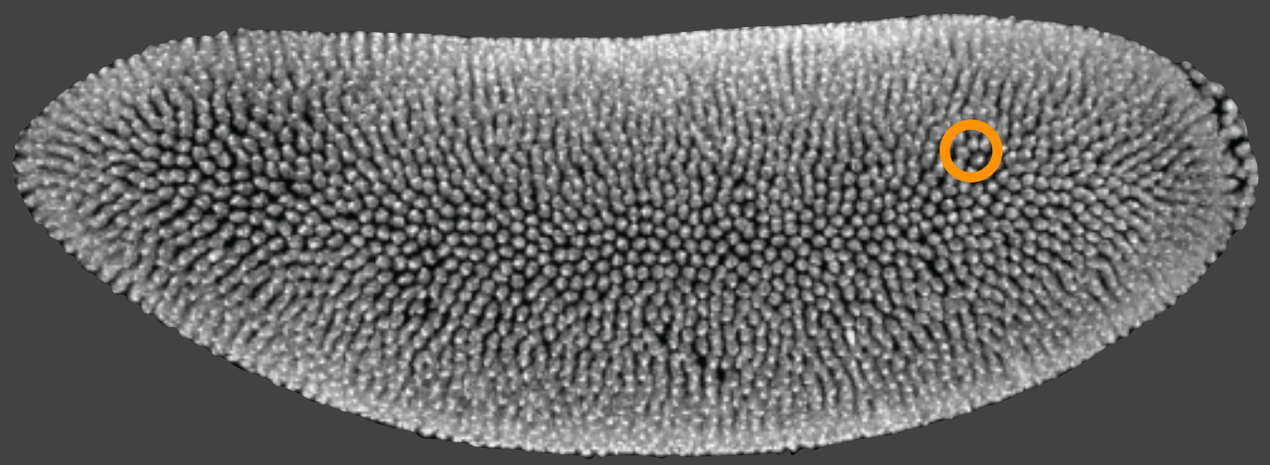
cells

time

genes

0.2	0.4	1	1	1	0.8	1
1	0	0	0	1	1	0.6
0.7	0.8	1	1	0.8	0.6	1
1	0	0.2	0.5	1	1	0.8
0.5	0.8	0.5	0.3	0.5	0.8	1
0.7	0.5	0.8	0.7	1	1	1
1	0.3	0.4	1	1	1	0.3

measurements taken for every cell in a single fruit fly embryo.



correlate gene expression with spatial location.

cells

time

genes

0.1	0.4	0.8				
0.7	0.9	0.8	0.7			
0.2	0.4	1	1			
0.1	0.4	0.5	0.5	0.4		
0.8	0.1	1	0.9	0.9	1	
0.2	0.4	1	1	1	0.8	1
1	0	0	0	1	1	0.6
0.7	0.8	1	1	0.8	0.6	1
1	0	0.2	0.5	1	1	0.8
0.5	0.8	0.5	0.3	0.5	0.8	1
0.7	0.5	0.8	0.7	1	1	1
1	0.3	0.4	1	1	1	0.3

virtual embryo

virtual embryo

several thousand cells
[5,000 ± 1,000]

virtual embryo

several thousand cells
[5,000 ± 1,000]

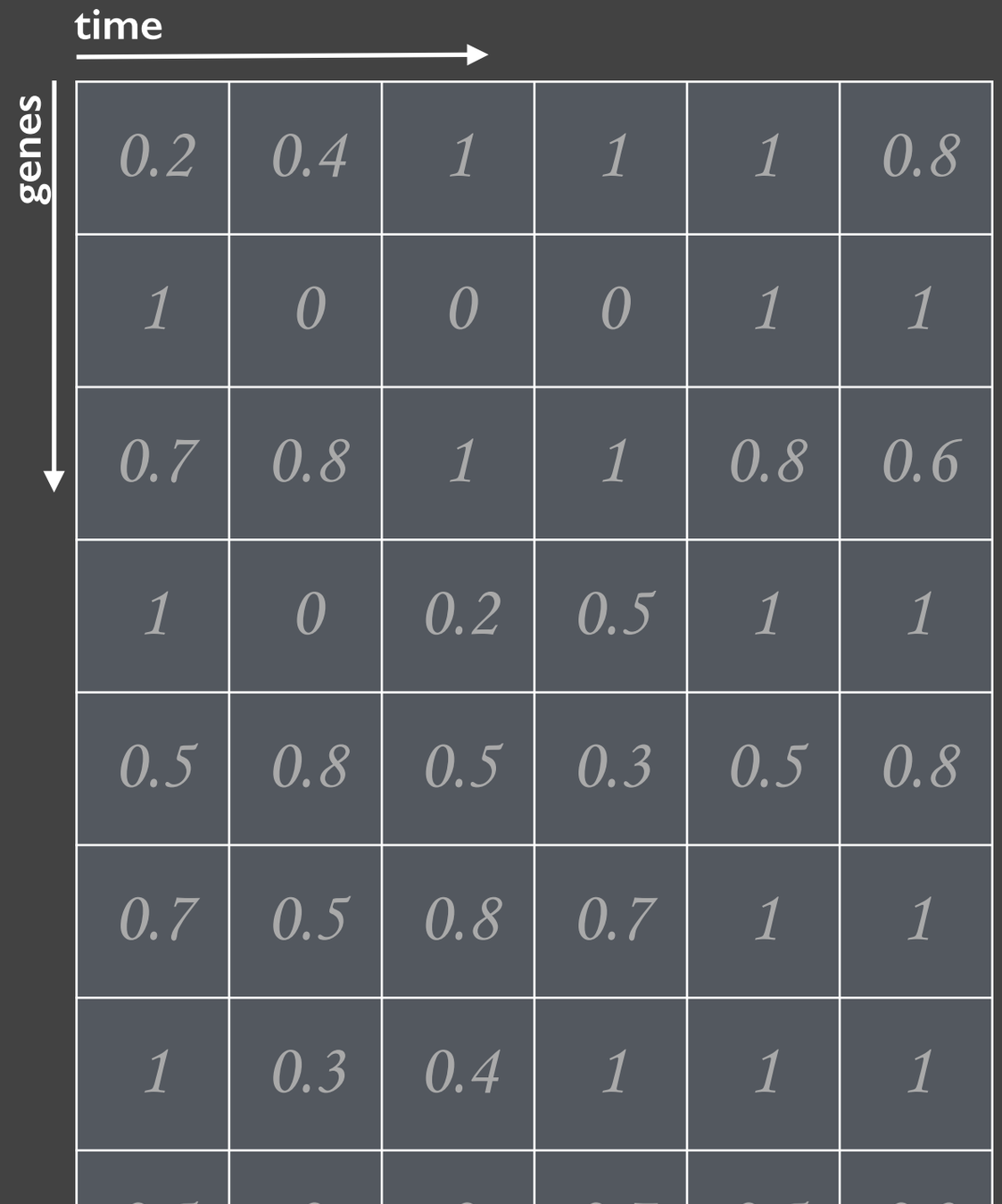
each cell has:

virtual embryo

several thousand cells
[5,000 ± 1,000]

each cell has:

expression profile
[6 time points x 50 genes]



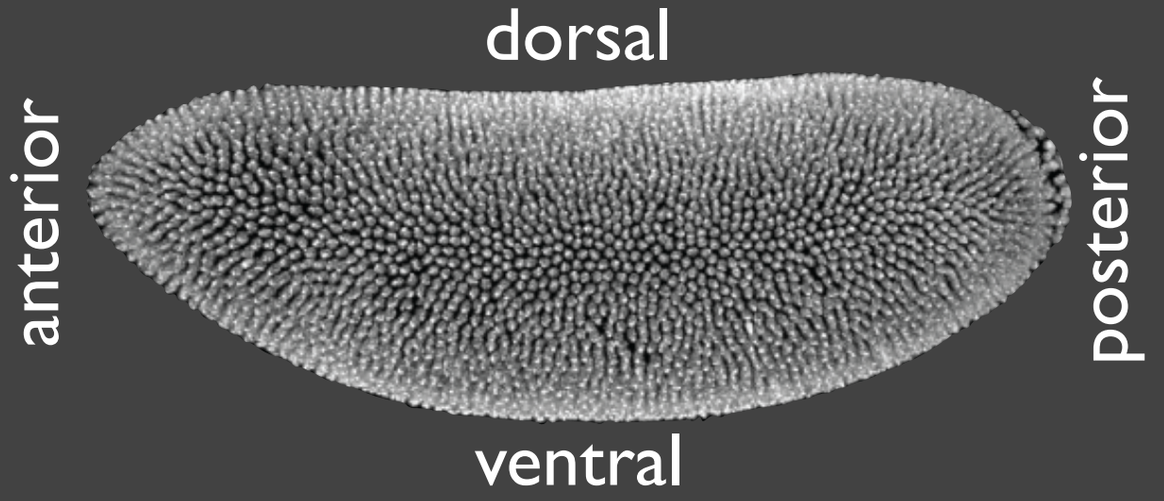
virtual embryo

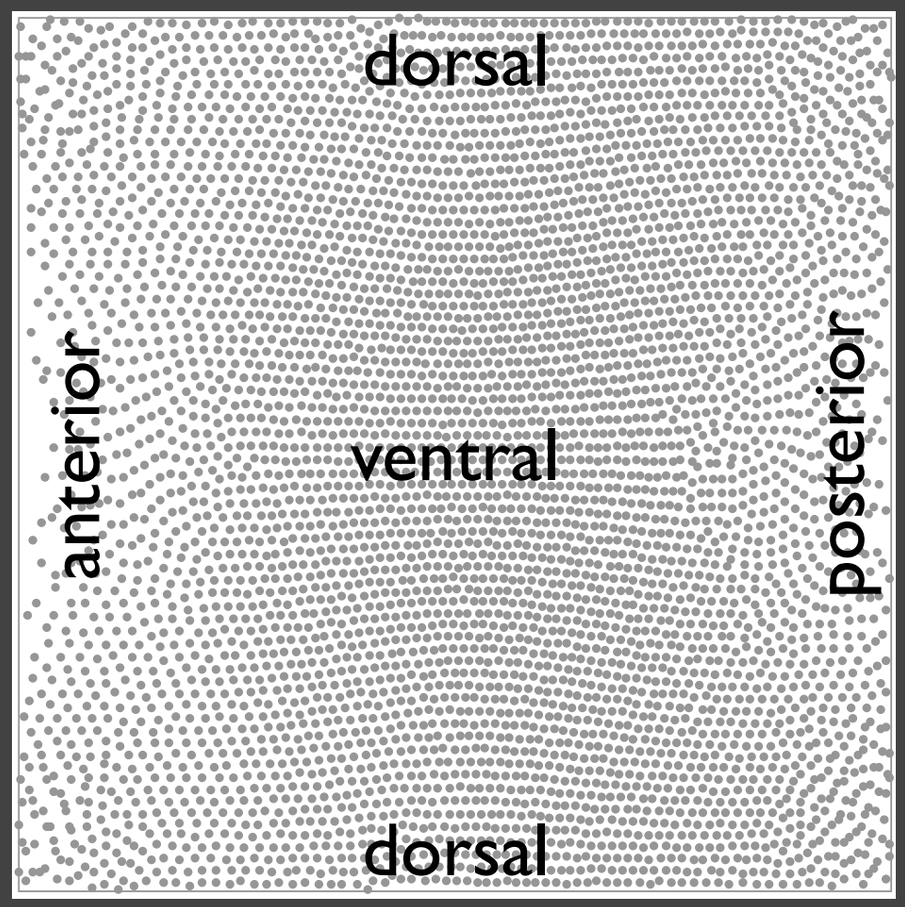
several thousand cells
[5,000 ± 1,000]

each cell has:

expression profile
[6 time points x 50 genes]

spatial position
[3D and 2D coordinates]





virtual embryo

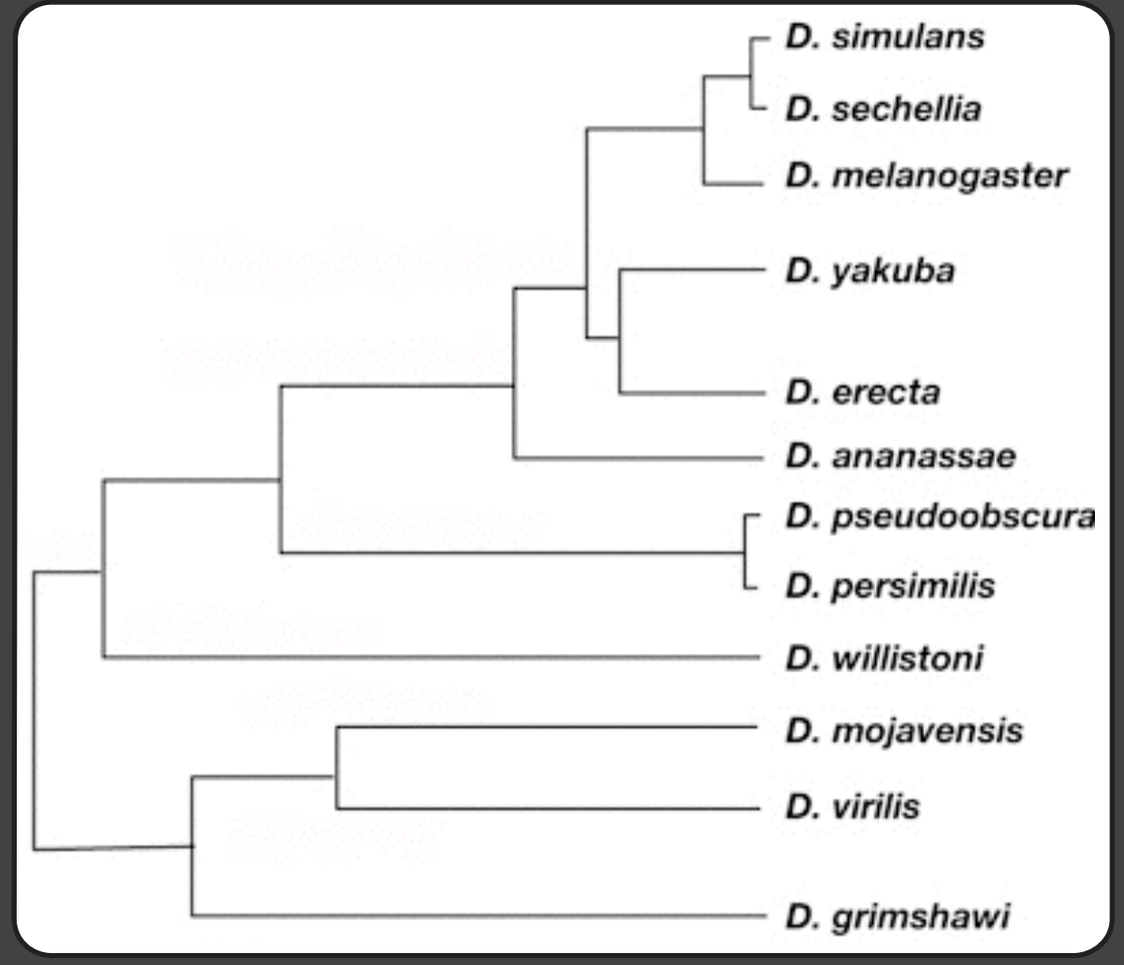
several thousand cells
[5,000 ± 1,000]

each cell has:

expression profile
[6 time points x 50 genes]

spatial position
[3D and 2D coordinates]

12 related species



virtual embryo

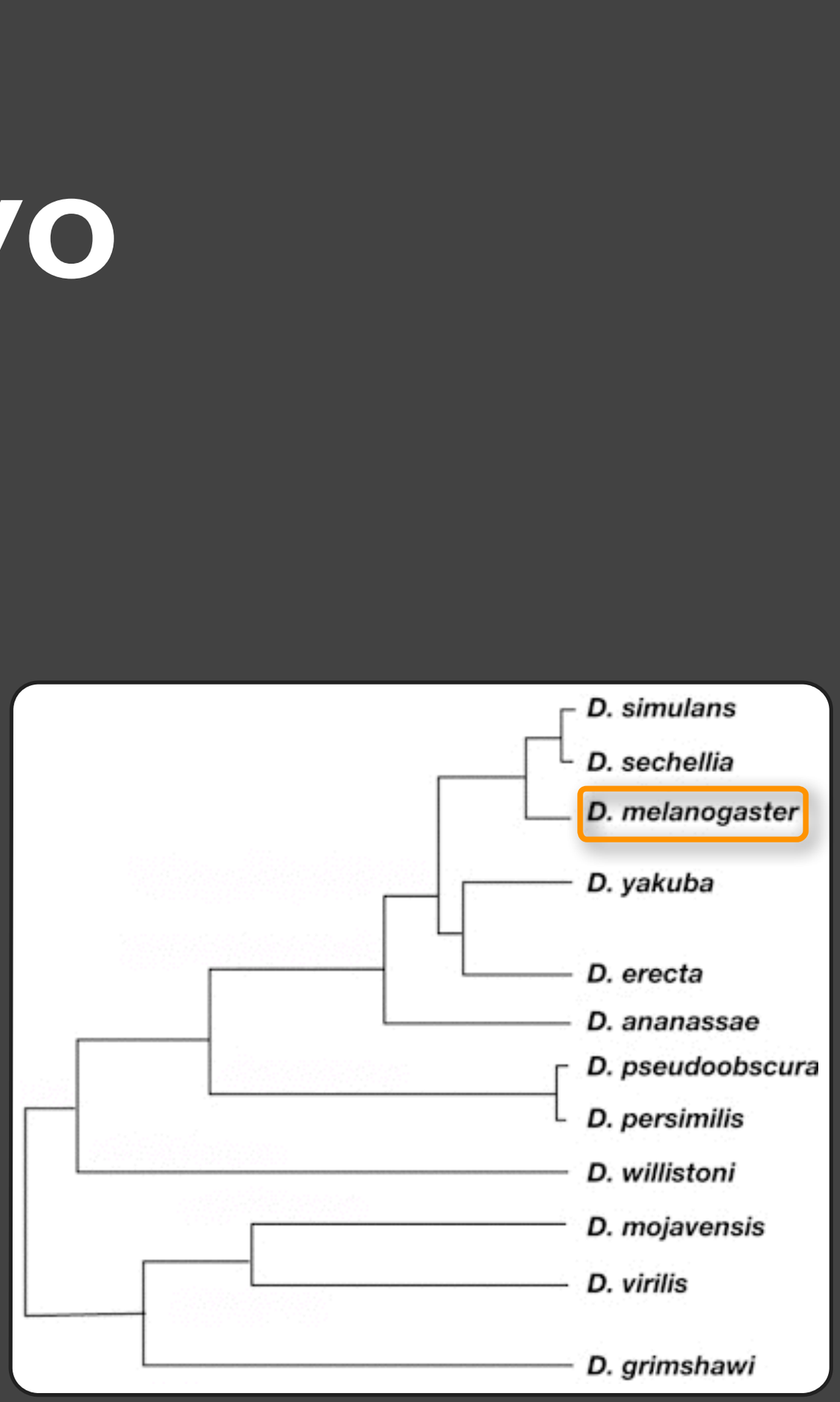
several thousand cells
[5,000 ± 1,000]

each cell has:

expression profile
[6 time points x 50 genes]

spatial position
[3D and 2D coordinates]

12 related species
one complete



virtual embryo

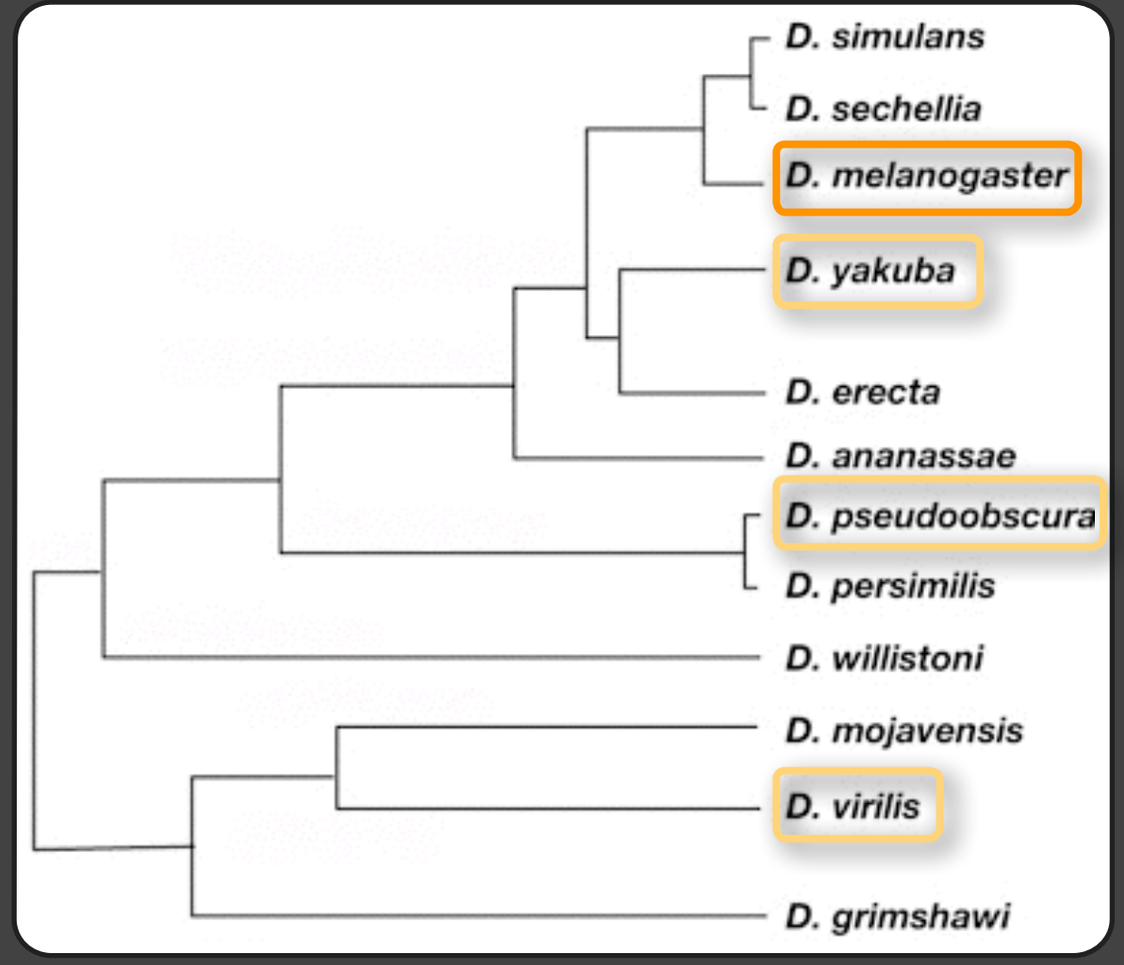
several thousand cells
[5,000 ± 1,000]

each cell has:

expression profile
[6 time points x 50 genes]

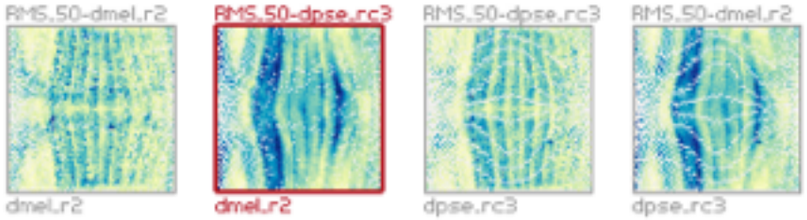
spatial position
[3D and 2D coordinates]

12 related species
one complete
three preliminary

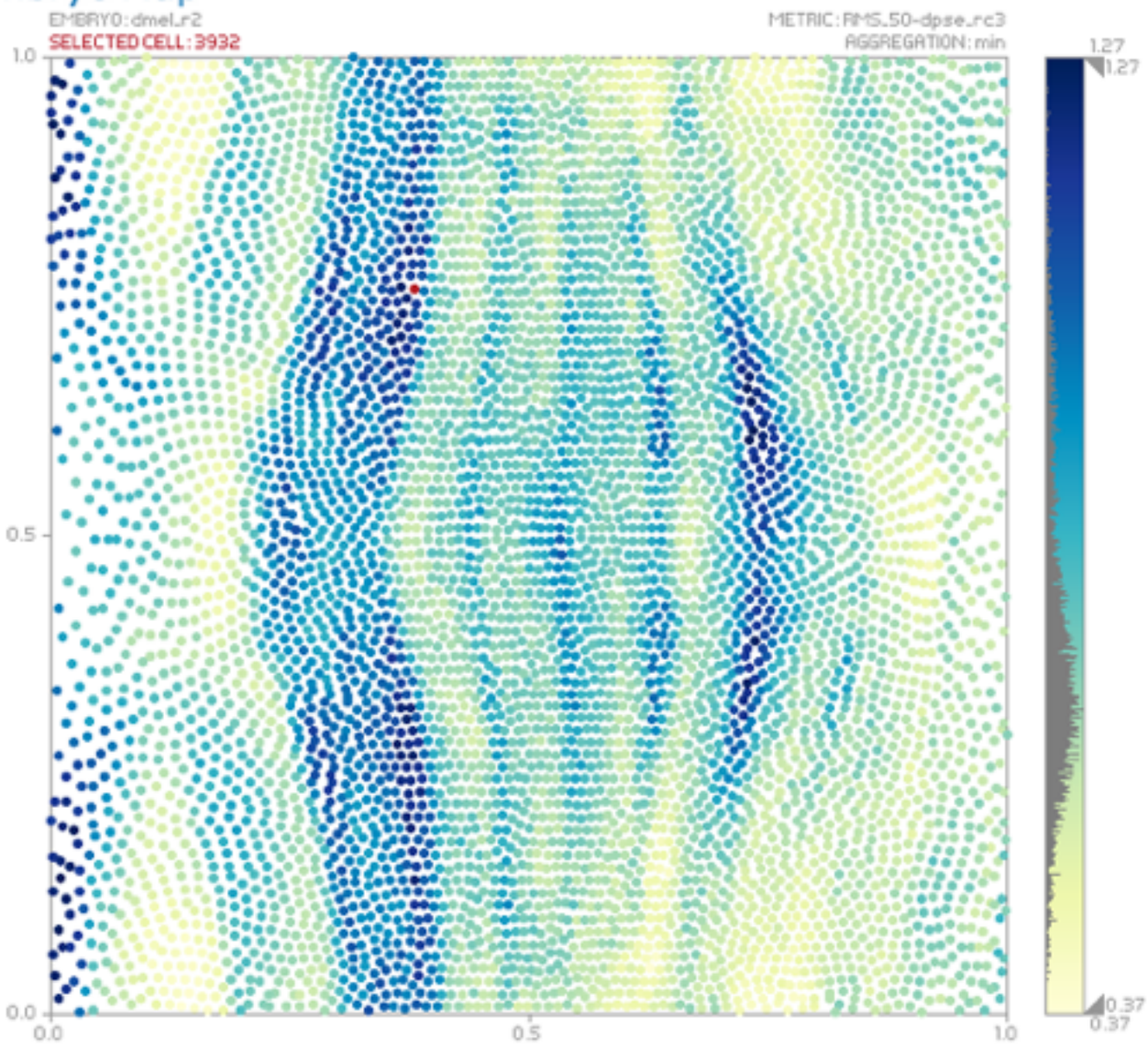


MulteeSum

Summaries



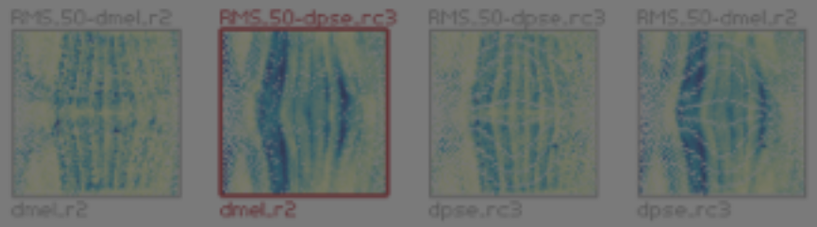
Embryo Map



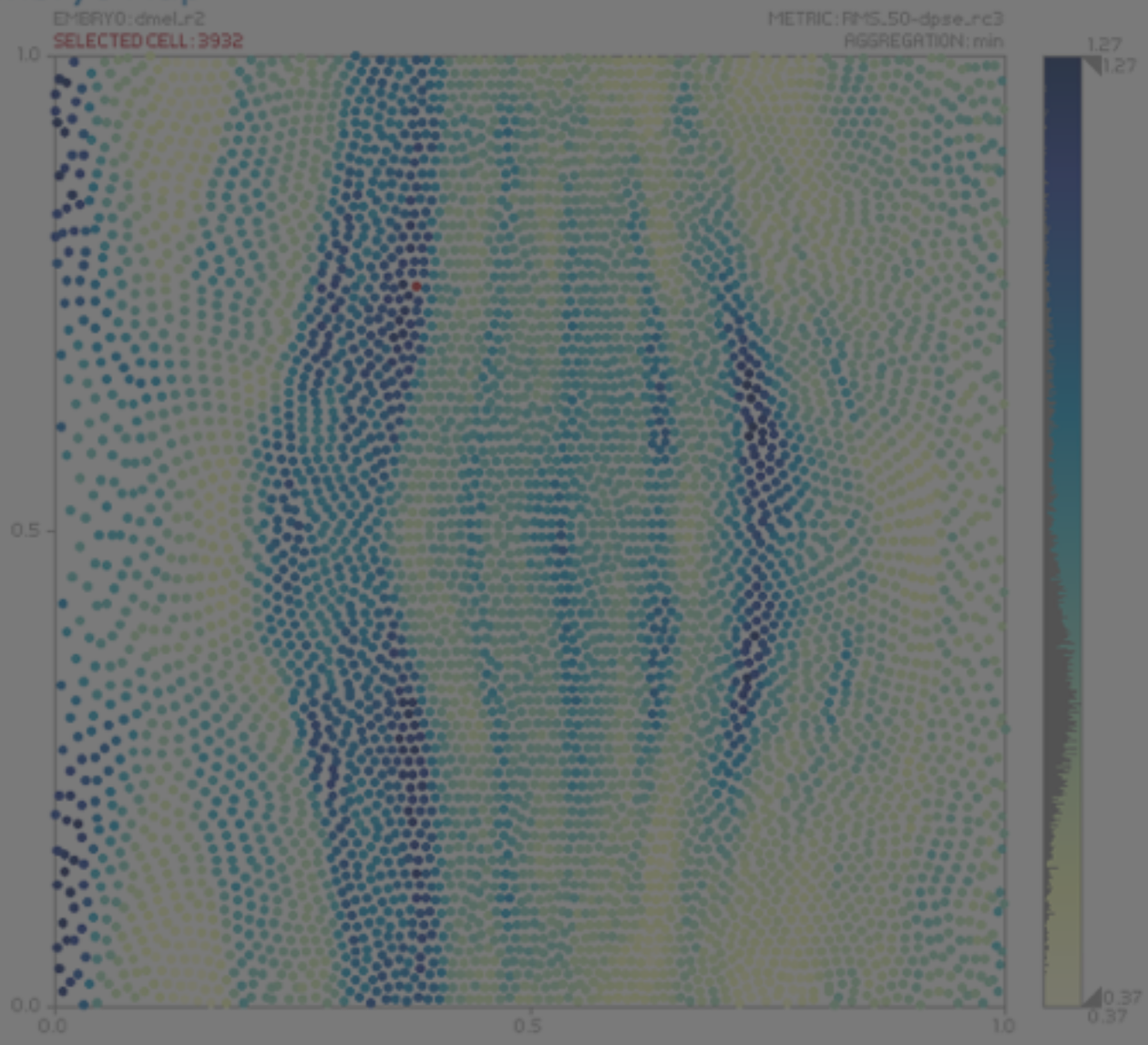
Curvemap

MulteeSum

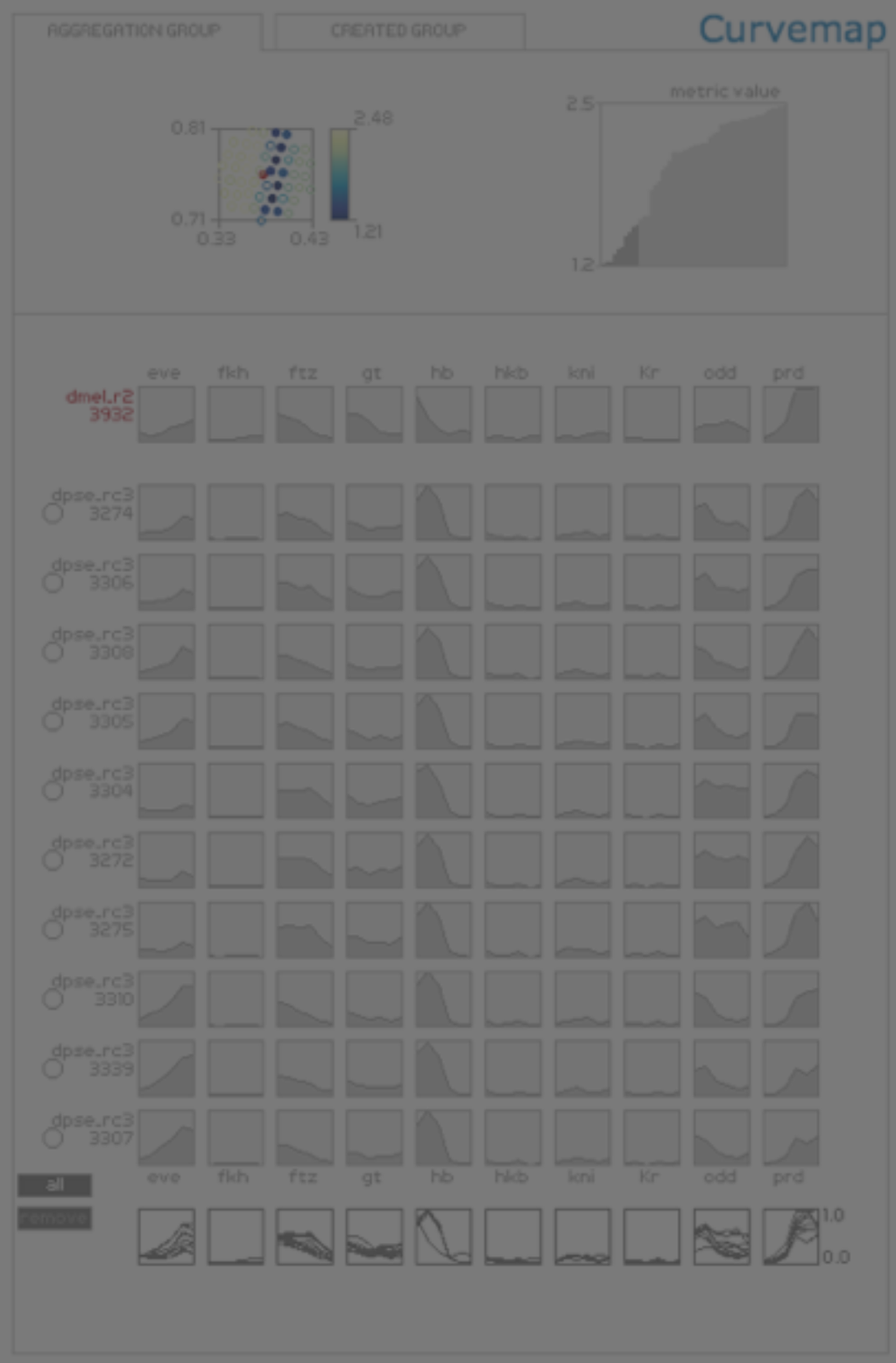
Summaries



Embryo Map



Curvemap



MulteeSum

Summaries

RMS_50-dmel_r2 — dmel_r2

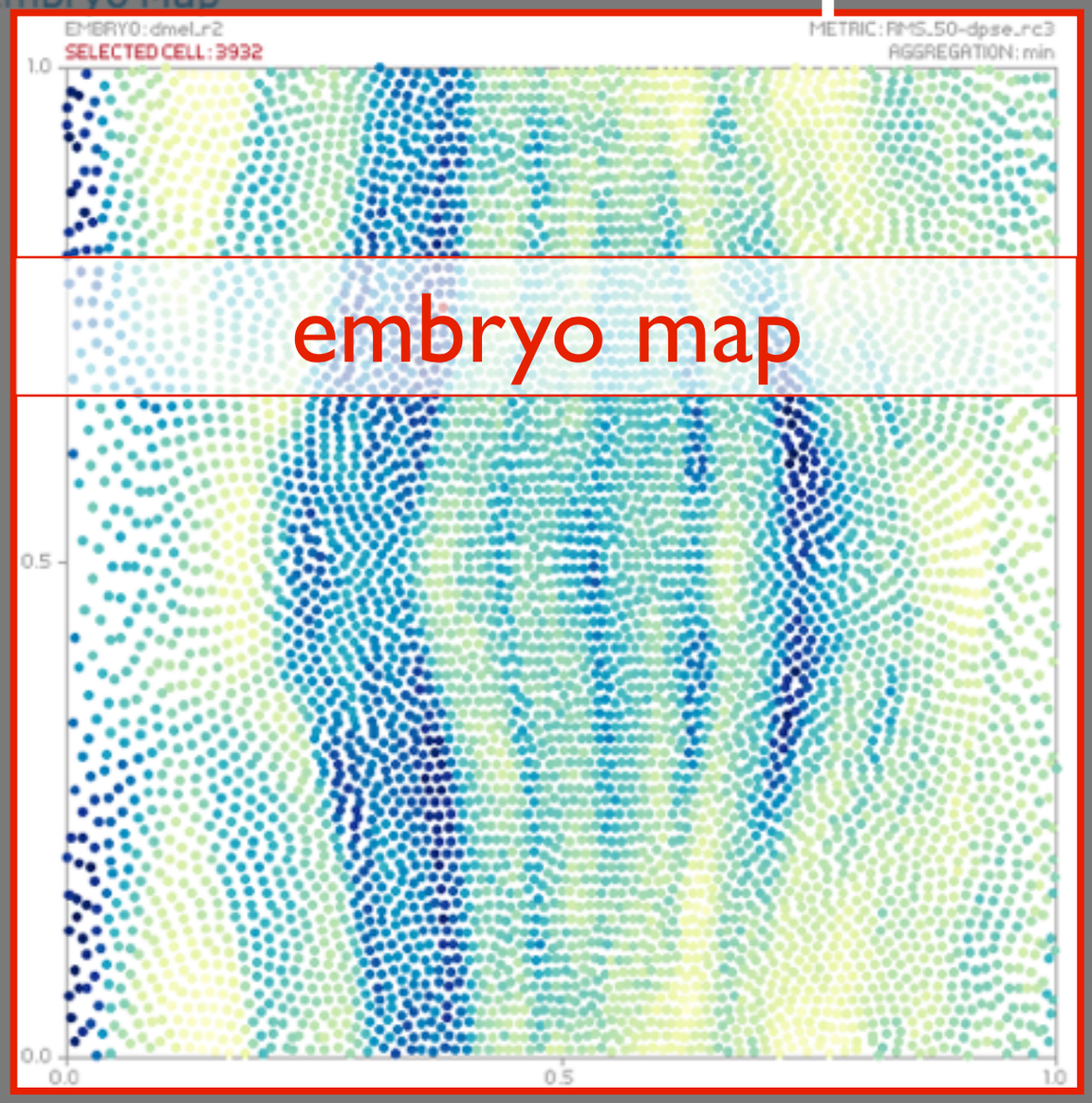
RMS_50-dpse_rc3 — dmel_r2

RMS_50-dpse_rc3 — dpse_rc3

RMS_50-dmel_r2 — dpse_rc3

Embryo Map

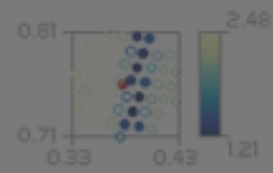
spatial

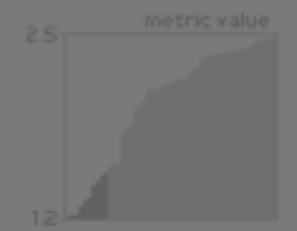


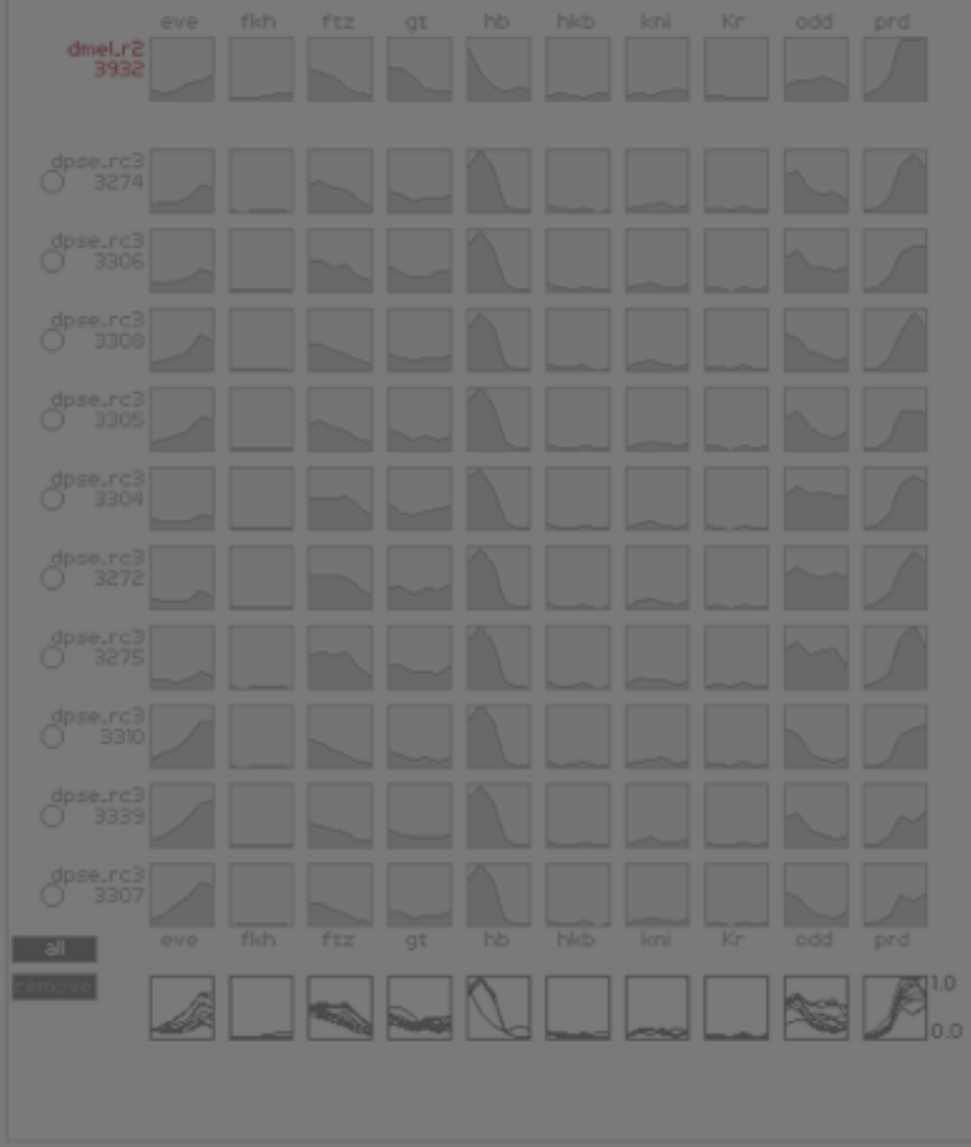
embryo map

Curvemap

AGGREGATION GROUP | CREATED GROUP







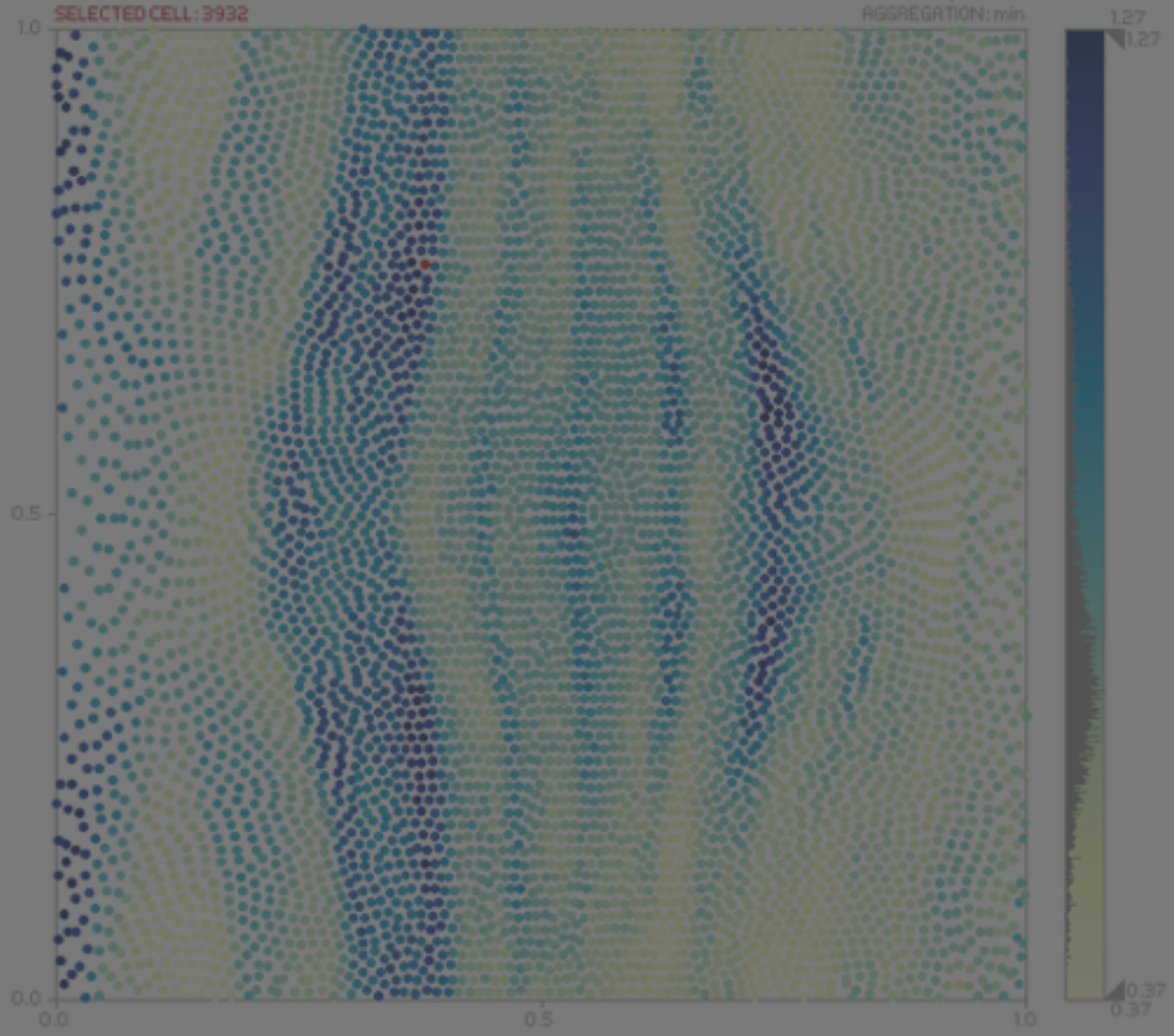
MulteeSum

Summaries

RMS.50-dmel.r2 dmel.r2 | RMS.50-dpse.rc3 dmel.r2 | RMS.50-dpse.rc3 dpse.rc3 | RMS.50-dmel.r2 dpse.rc3

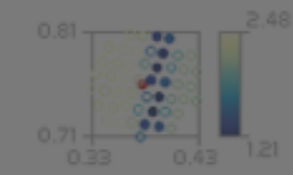
Embryo Map

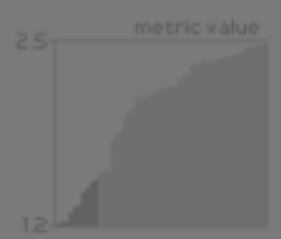
EMBRYO: dmel.r2
SELECTED CELL: 3932
METRIC: RMS.50-dpse.rc3
AGGREGATION: min



AGGREGATION GROUP | CREATED GROUP

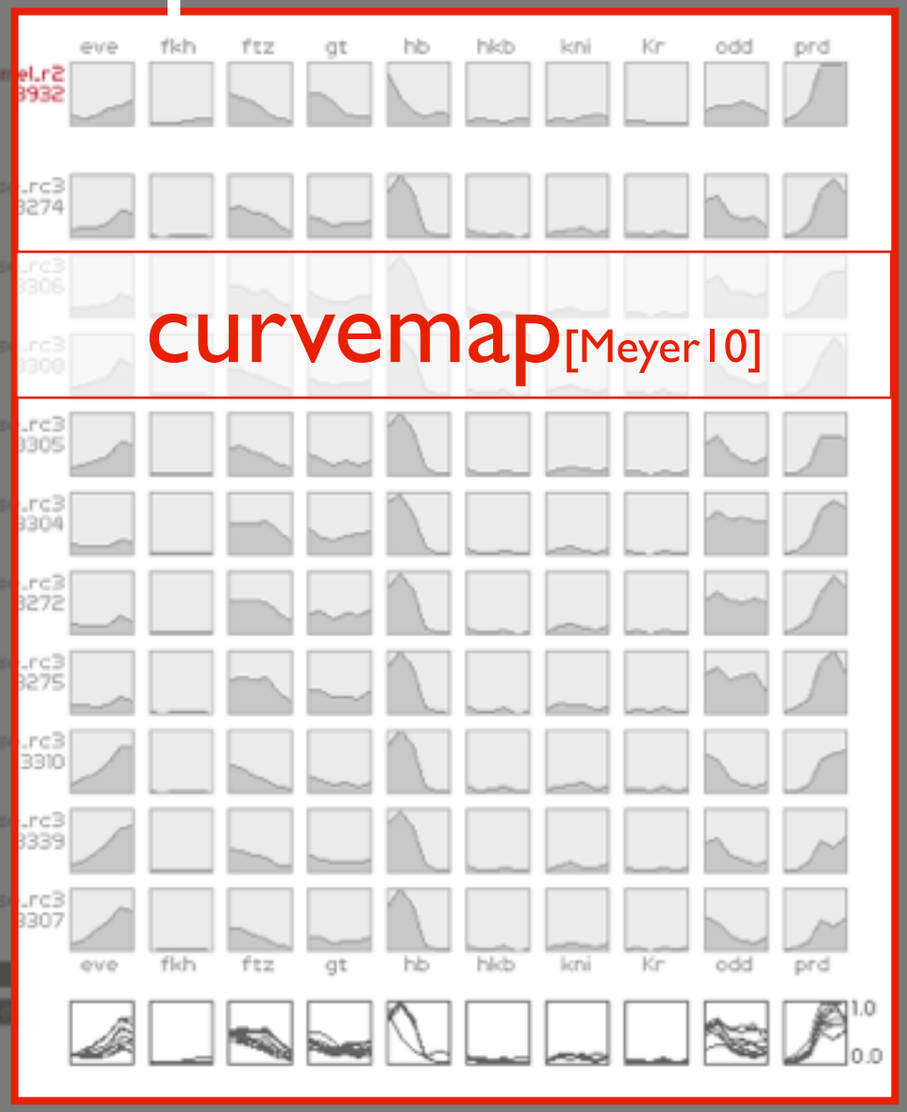
Curvemap





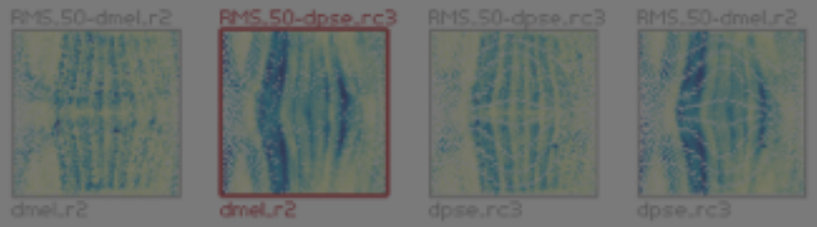
expression

curvemap[Meyer10]

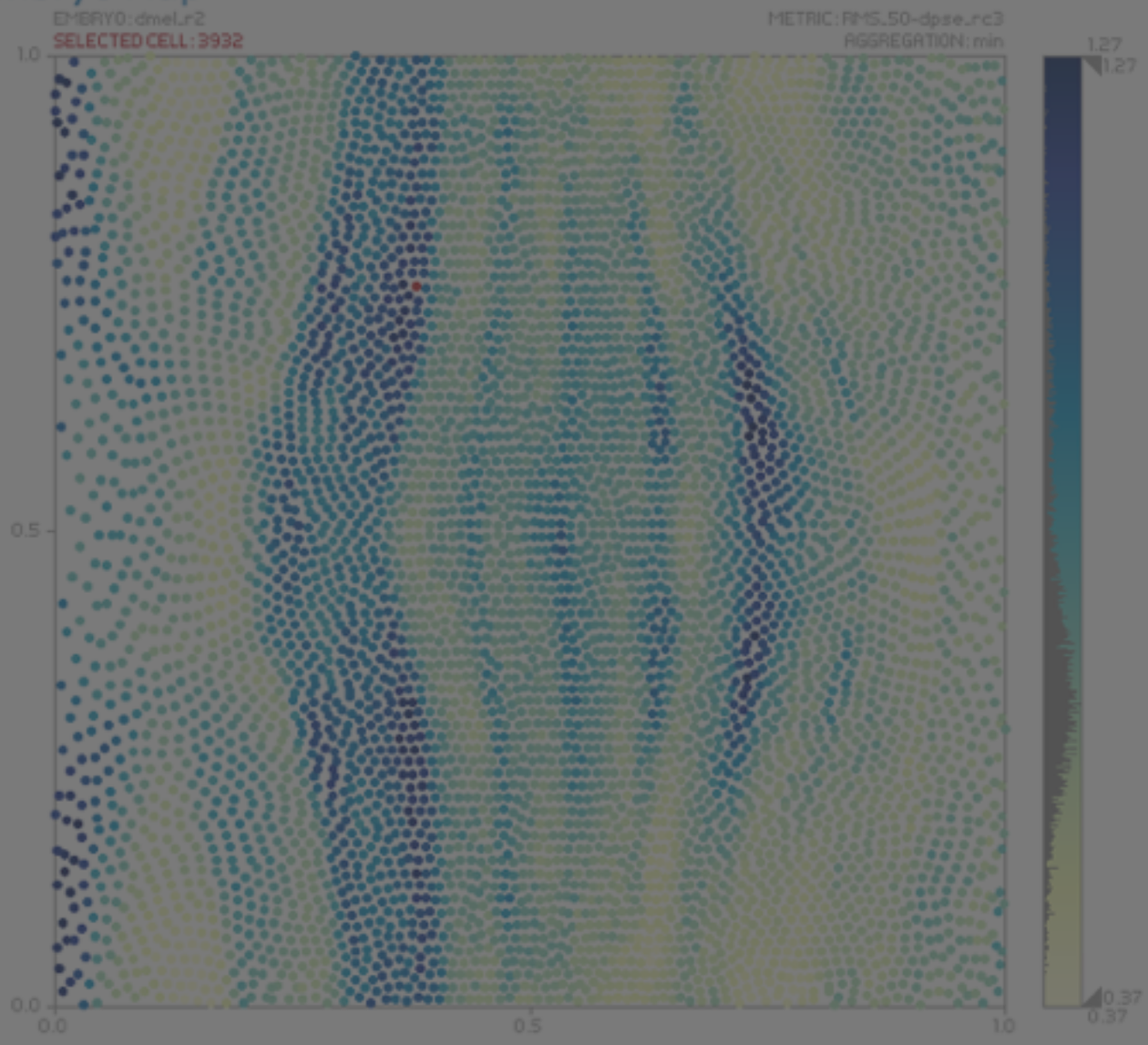


MulteeSum

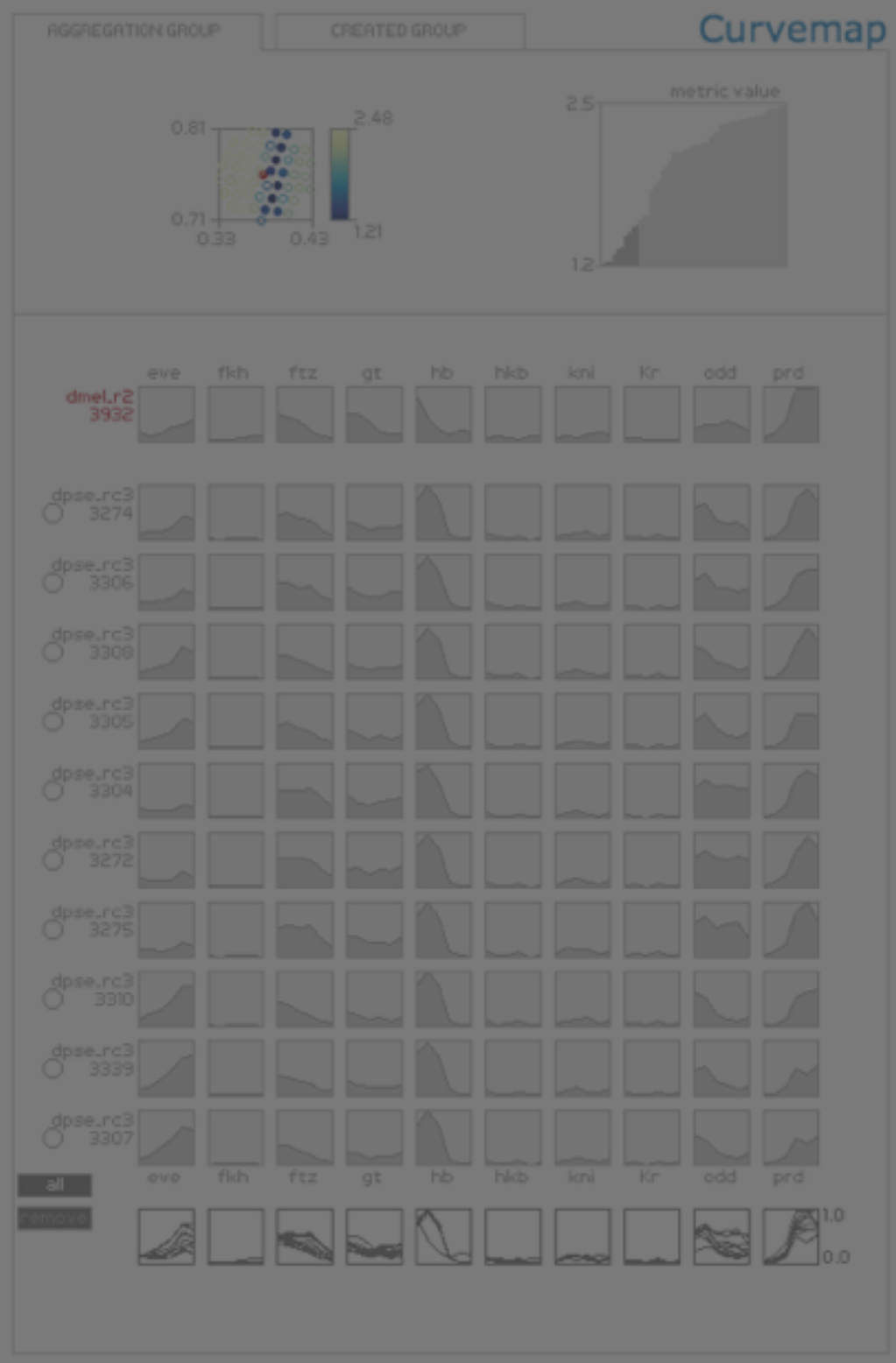
Summaries



Embryo Map

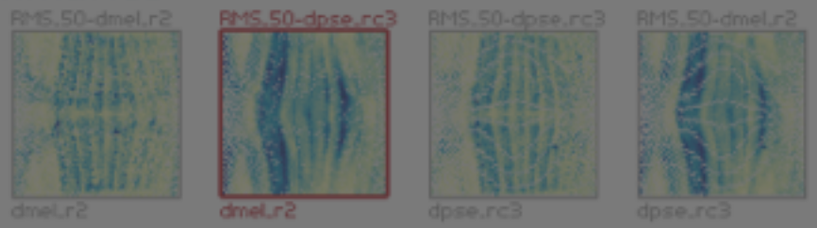


Curvemap

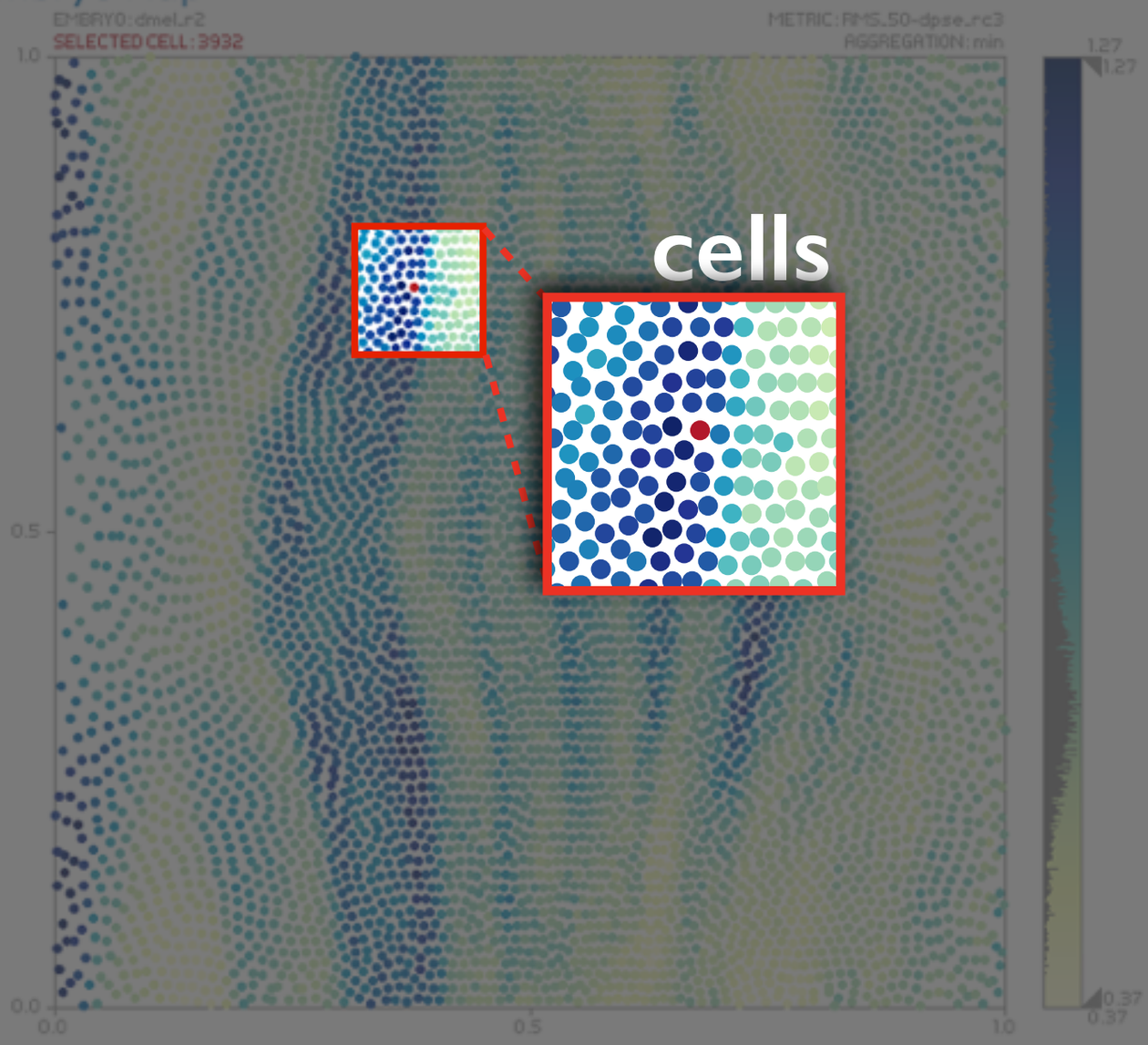


MulteeSum

Summaries

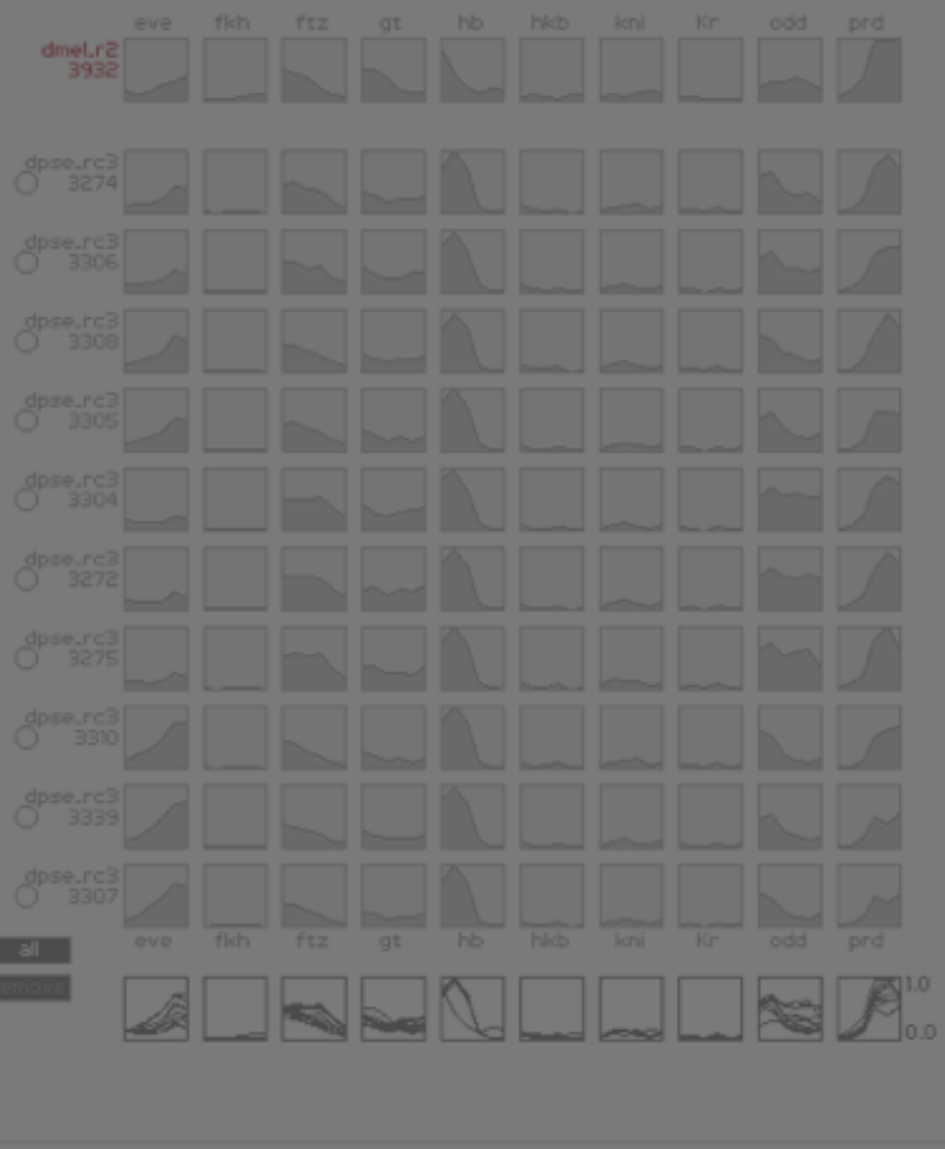


Embryo Map



AGGREGATION GROUP | CREATED GROUP

Curvemap



MulteeSum

Summaries

RMS.50-dmel.r2 — dmel.r2

RMS.50-dpse.rc3 — dmel.r2

RMS.50-dpse.rc3 — dpse.rc3

RMS.50-dmel.r2 — dpse.rc3

Embryo Map

EMBRYO: dmel.r2
SELECTED CELL: 3932
METRIC: RMS.50-dpse.rc3
AGGREGATION: min

1.0
0.5
0.0
0.0
0.5
1.0

1.27
1.27
0.37
0.37

cells

AGGREGATION GROUP | CREATED GROUP

Curvemap

0.81
0.71
0.33
0.43
2.48
1.21

metric value
2.5
1.2

expression profile

eve fkh ftz gt hb hkb kni Kr odd prd

dmel.r2 3932

dpse.rc3 3274
dpse.rc3 3306
dpse.rc3 3309
dpse.rc3 3305
dpse.rc3 3304
dpse.rc3 3272
dpse.rc3 3275
dpse.rc3 3310
dpse.rc3 3339
dpse.rc3 3307

eve fkh ftz gt hb hkb kni Kr odd prd

all
none

1.0
0.0

MulteeSum

Summaries

RMS.50-dmel.r2 — dmel.r2

RMS.50-dpse.rc3 — dmel.r2

RMS.50-dpse.rc3 — dpse.rc3

RMS.50-dmel.r2 — dpse.rc3

Embryo Map

EMBRYO: dmel.r2
SELECTED CELL: 3932
METRIC: RMS.50-dpse.rc3
AGGREGATION: min

1.0
0.5
0.0
0.0
0.5
1.0

1.27
1.27
0.37
0.37

eve

expression

time

Curvemap

AGGREGATION GROUP | CREATED GROUP

0.81
0.71
0.33
0.43
2.48
1.21

metric value
2.5
1.2

eve fkh ftz gt hb hkb kni Kr odd prd

dmel.r2 3932

dpse.rc3 3274
dpse.rc3 3306
dpse.rc3 3308
dpse.rc3 3305
dpse.rc3 3304
dpse.rc3 3272
dpse.rc3 3275
dpse.rc3 3310
dpse.rc3 3339
dpse.rc3 3307

eve fkh ftz gt hb hkb kni Kr odd prd

all
remove

1.0
0.0

MulteeSum

Summaries

RMS.50-dmel.r2 — dmel.r2

RMS.50-dpse.rc3 — dmel.r2

RMS.50-dpse.rc3 — dpse.rc3

RMS.50-dmel.r2 — dpse.rc3

Embryo Map

EMBRYO: dmel.r2
SELECTED CELL: 3932
METRIC: RMS.50-dpse.rc3
AGGREGATION: min

1.0
0.5
0.0
0.0
0.5
1.0

1.27
1.27
0.37
0.37

AGGREGATION GROUP | CREATED GROUP

Curvemap

0.81
0.71
0.33
0.43
2.48
1.21

metric value
2.5
1.2

eve fkh ftz gt hb hkb kni Kr odd prd

dmel_r2 3932

dpse_rc3 3274
dpse_rc3 3275
dpse_rc3 3310
dpse_rc3 3339
dpse_rc3 3307

all
none

1.0
0.0

MulteeSum

Summaries

RMS.50-dmel.r2 | RMS.50-dpse.rc3 | RMS.50-dpse.rc3 | RMS.50-dmel.r2

dmel.r2 | dmel.r2 | dpse.rc3 | dpse.rc3

Embryo Map

EMBRYO: dmel.r2
SELECTED CELL: 3932
METRIC: RMS.50-dpse.rc3
AGGREGATION: min

1.0 | 0.5 | 0.0 | 0.0 | 0.5 | 1.27 | 1.27

AGGREGATION GROUP | CREATED GROUP

Curvemap

0.81 | 0.71 | 0.33 | 0.43 | 2.48 | 1.21

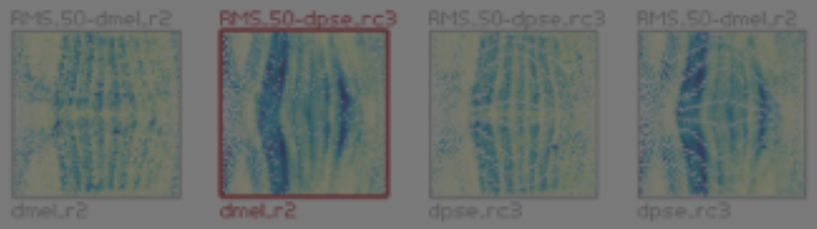
metric value
2.5 | 1.2

eve fkh ftz gt hb hkb kni Kr odd prd

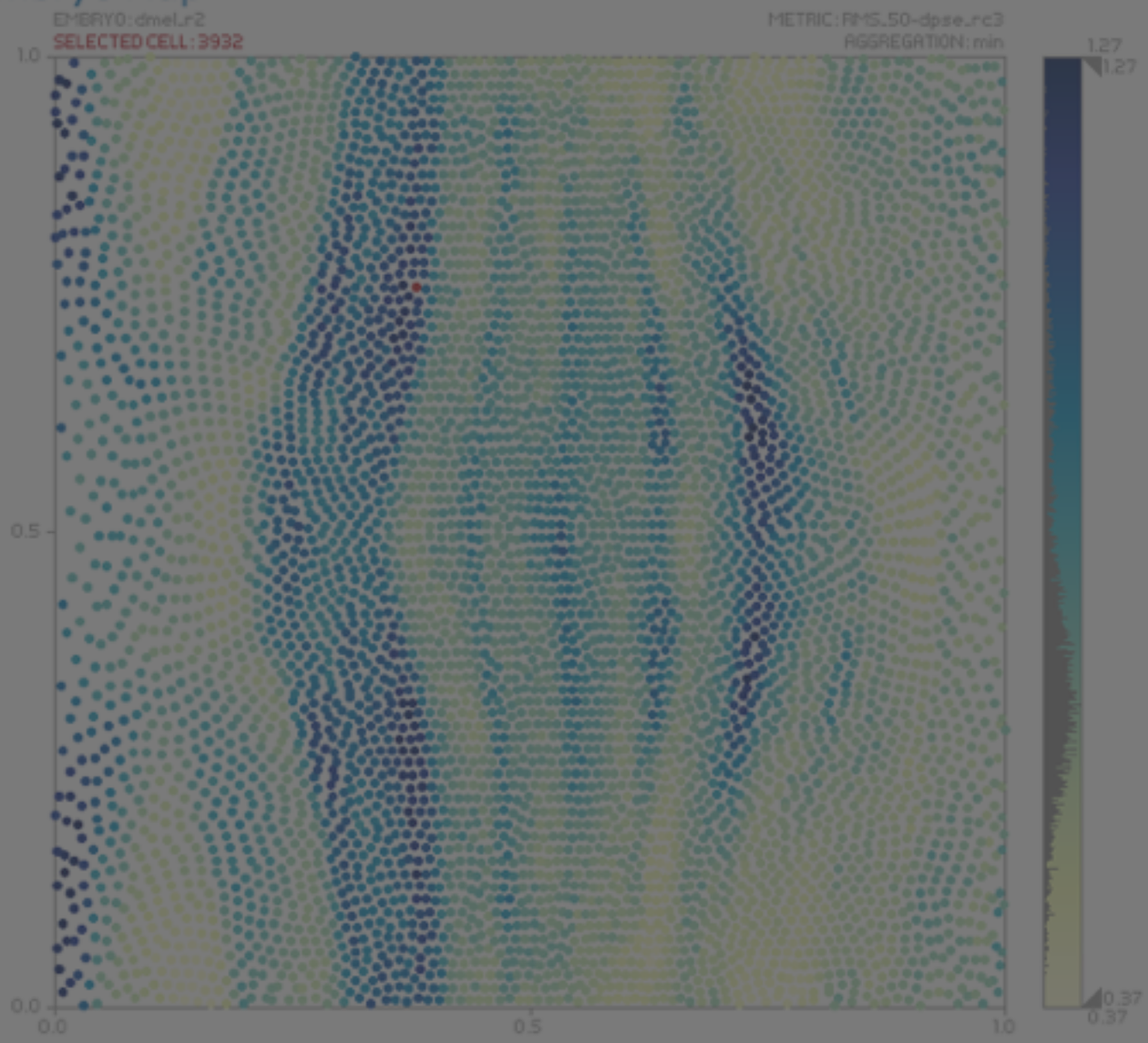
dmel_r2 3932
dpse_rc3 3274
dpse_rc3 3306
dpse_rc3 3308

MulteeSum

Summaries



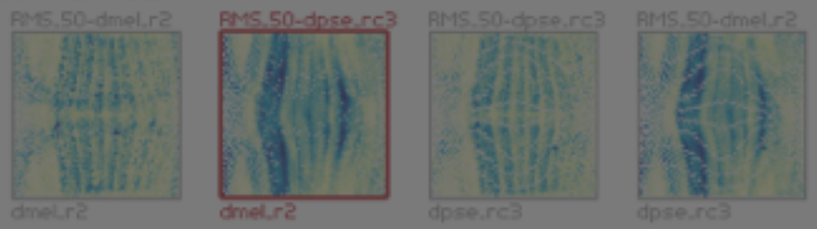
Embryo Map



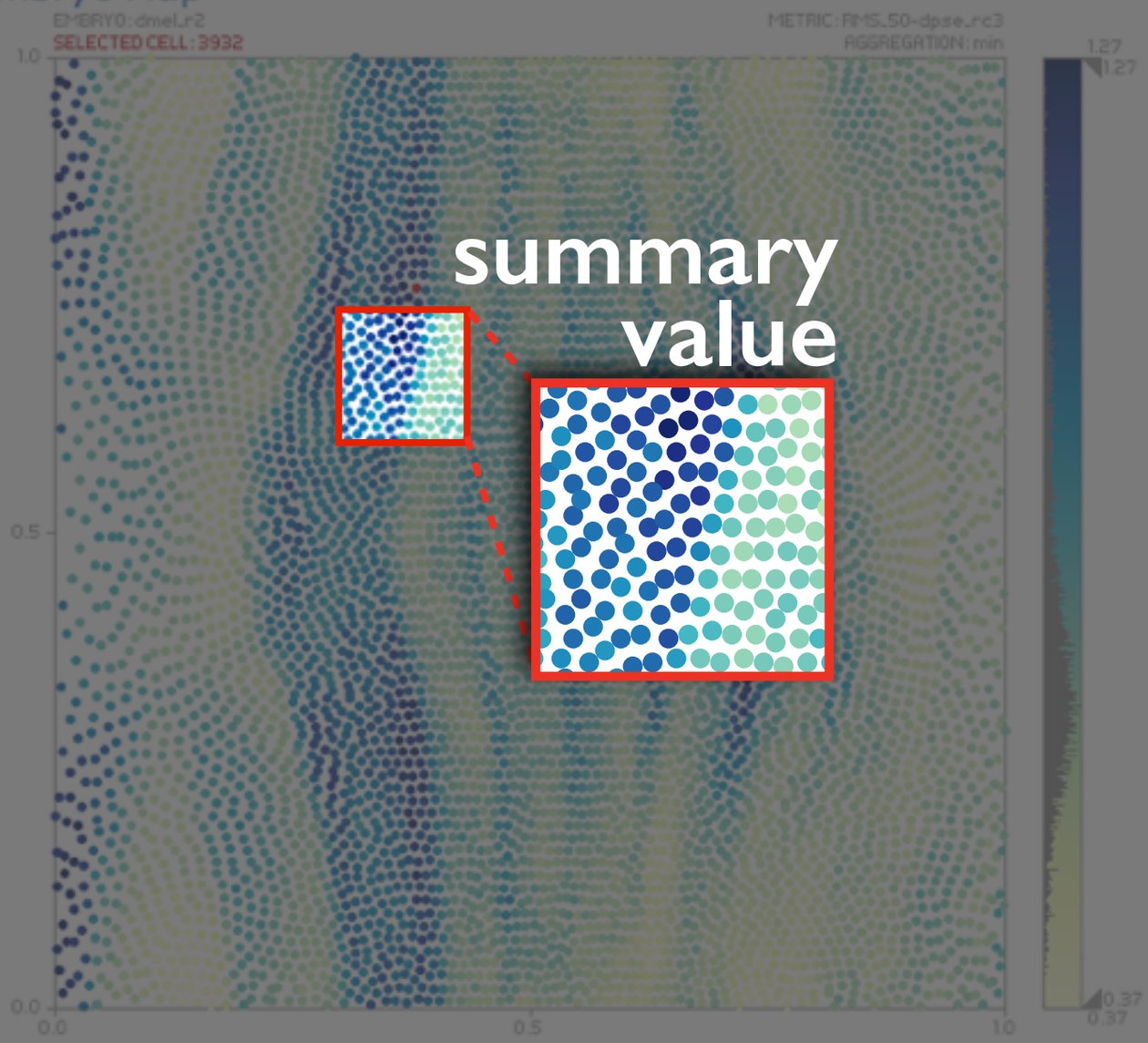
Curvemap

MulteeSum

Summaries



Embryo Map

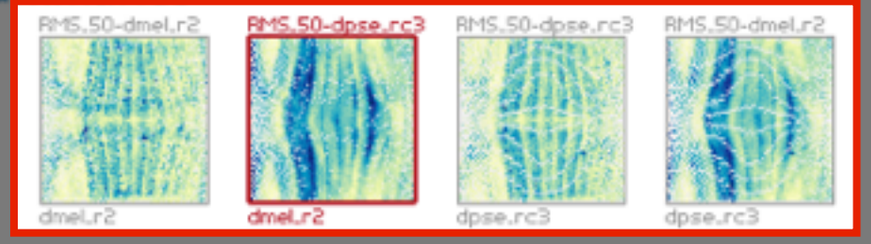


Curvemap

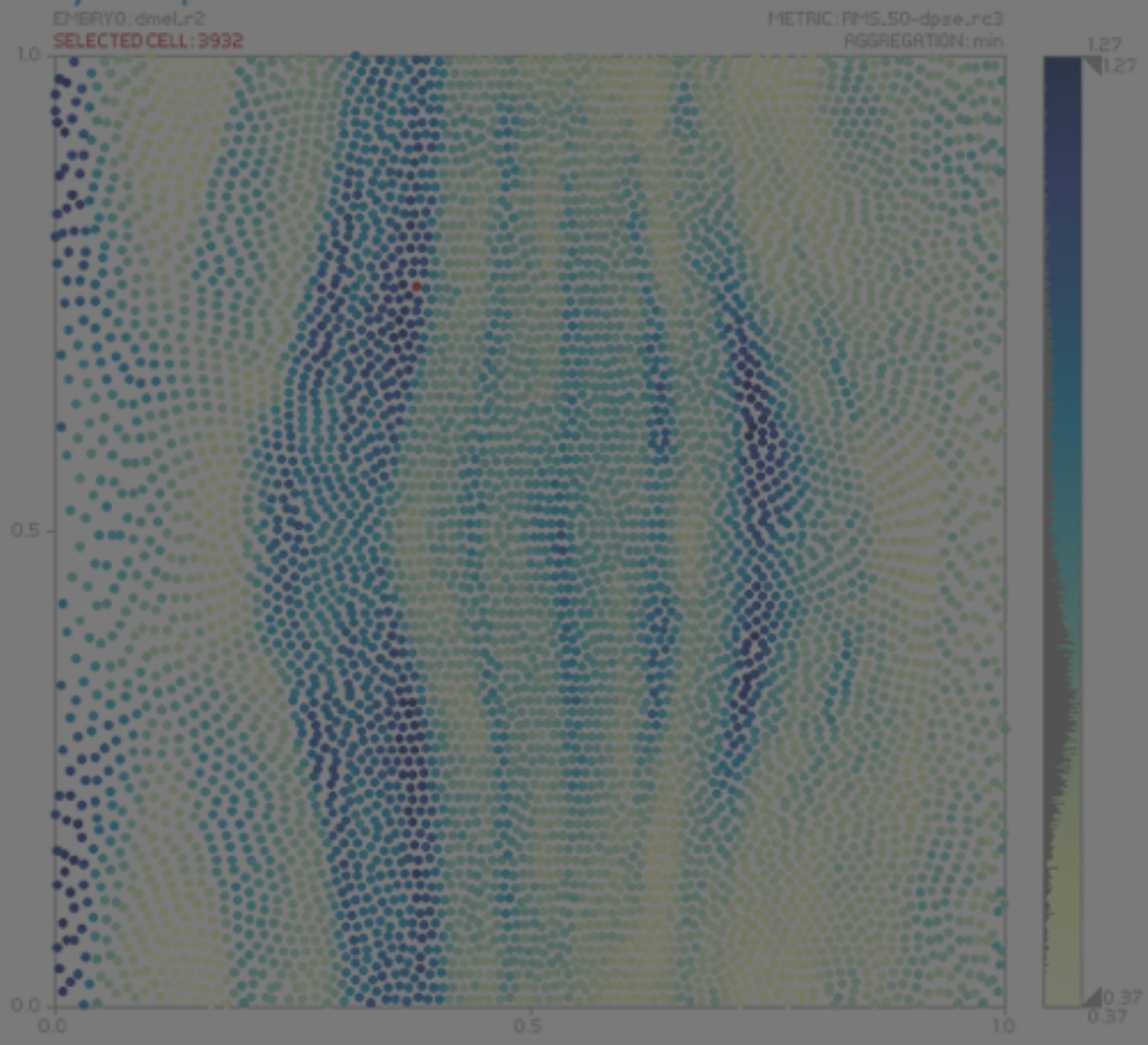
MulteeSum

summaries

Summaries



Embryo Map



Curvemap

Characterize differences in gene expression patterns between species.

Characterize differences in gene expression patterns between species.

differences related to:

spatial position

gene expression profiles

complex combination

challenging to characterize manually

support mechanisms:

summaries, groups

data & tool & tasks

summaries & groups

encodings & interaction

conclusions

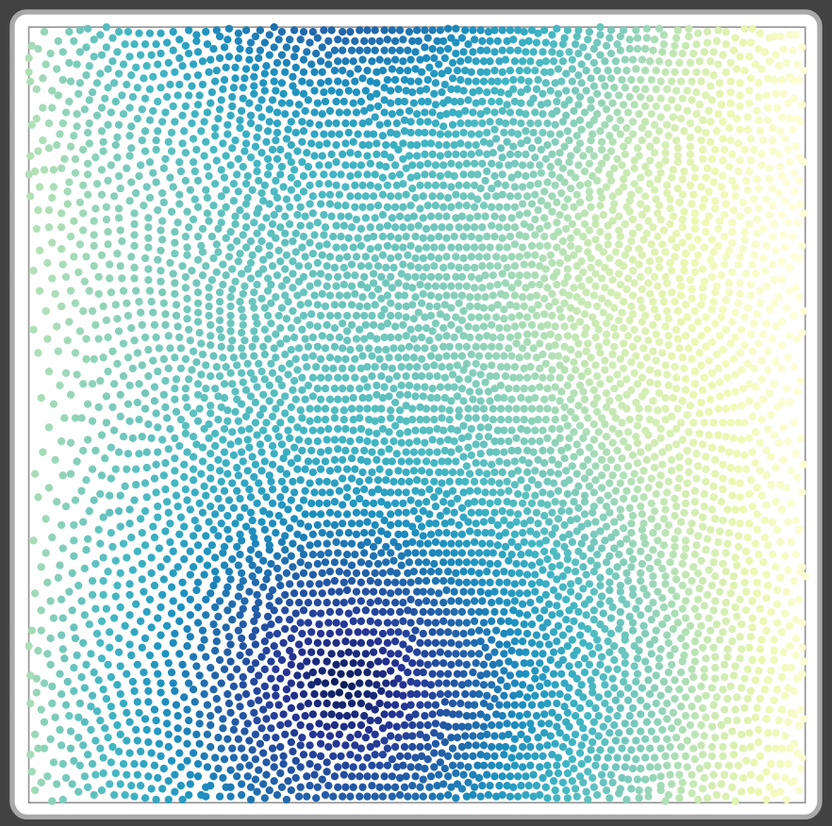
summary

a value for each cell that summarizes the underlying data

summary

a value for each cell that summarizes the underlying data

spatial

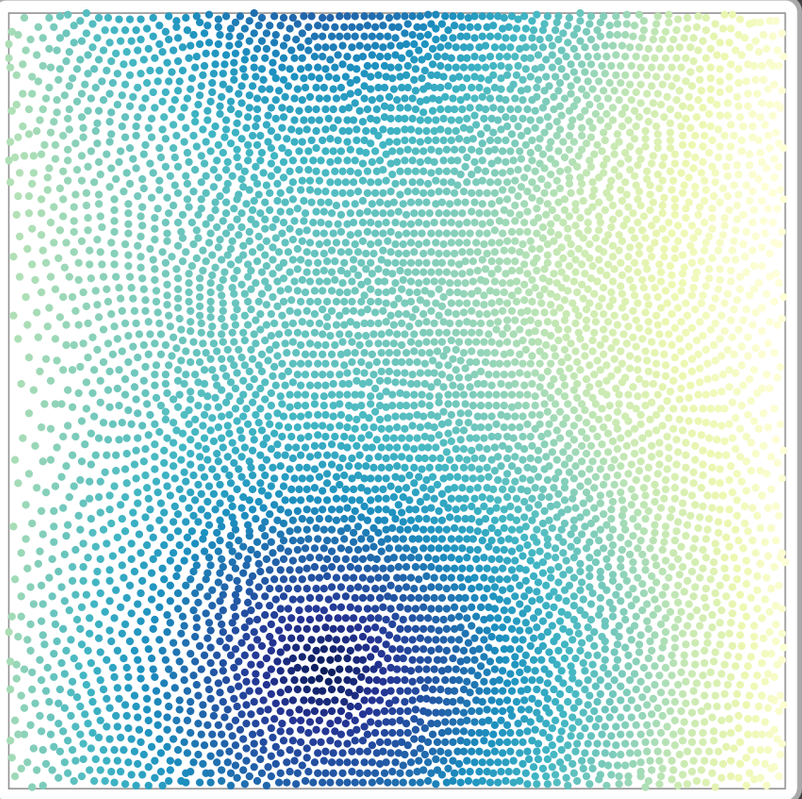


distance to cell *i*

summary

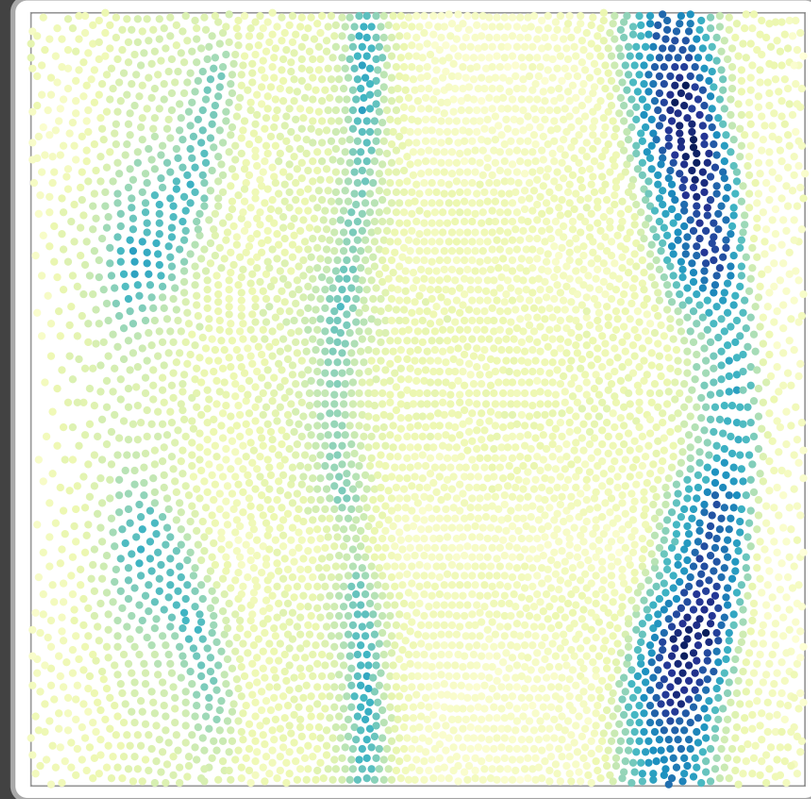
a value for each cell that summarizes the underlying data

spatial



distance to cell *i*

expression

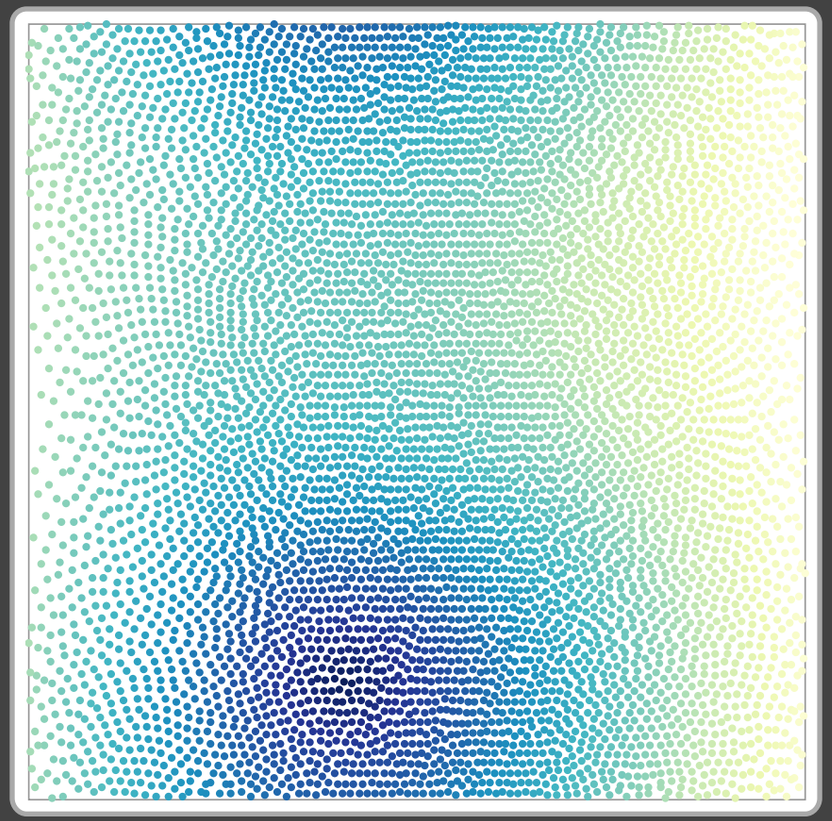


gene *j* at timepoint *k*

summary

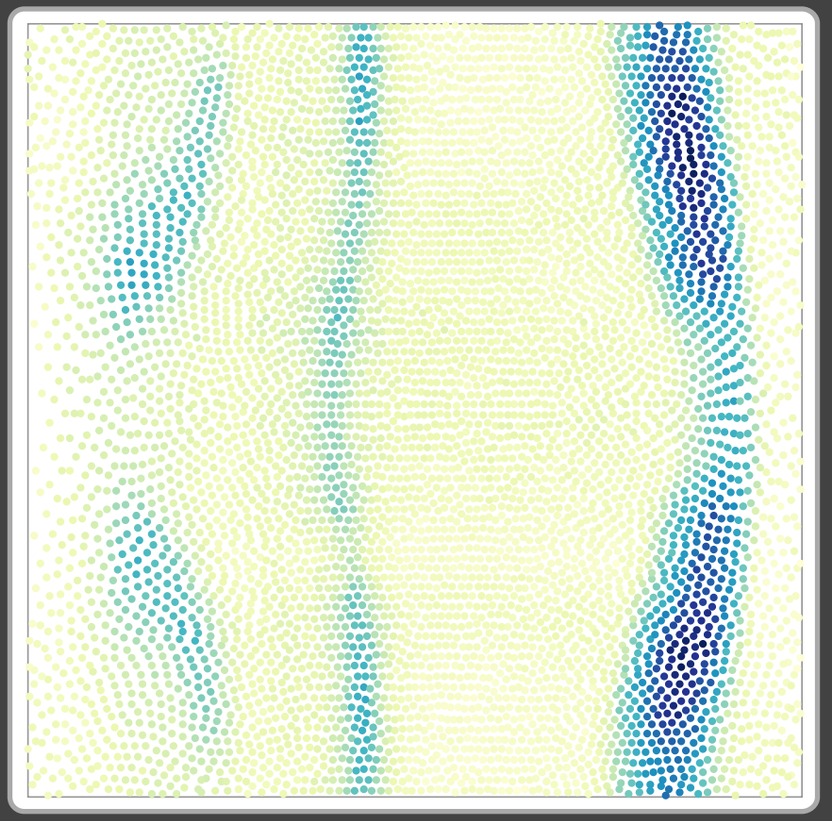
a value for each cell that summarizes the underlying data

spatial



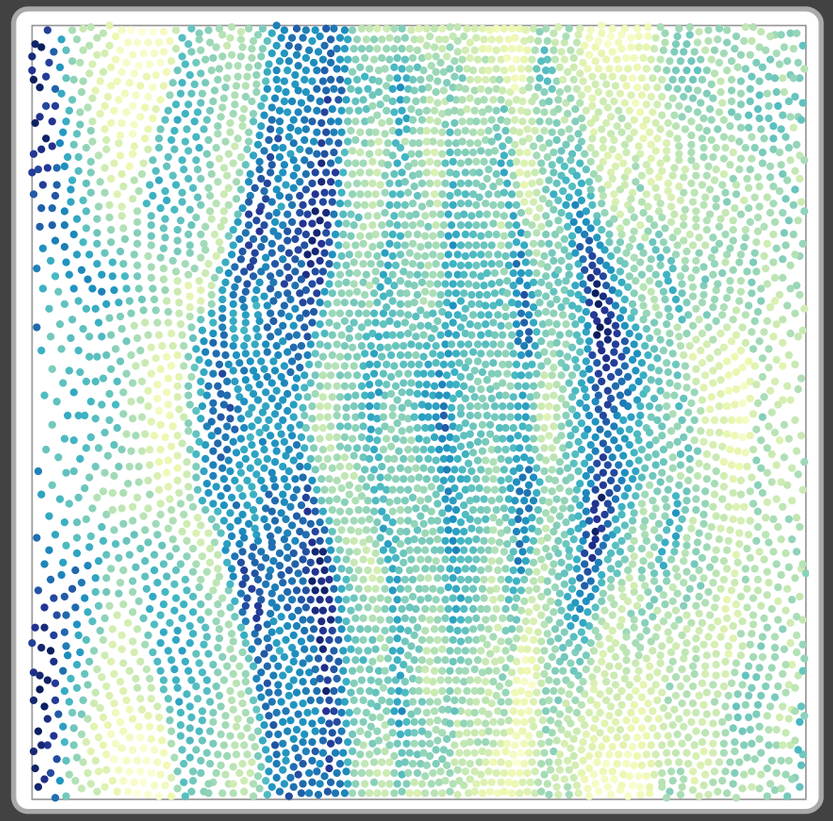
distance to cell *i*

expression



gene *j* at timepoint *k*

combination

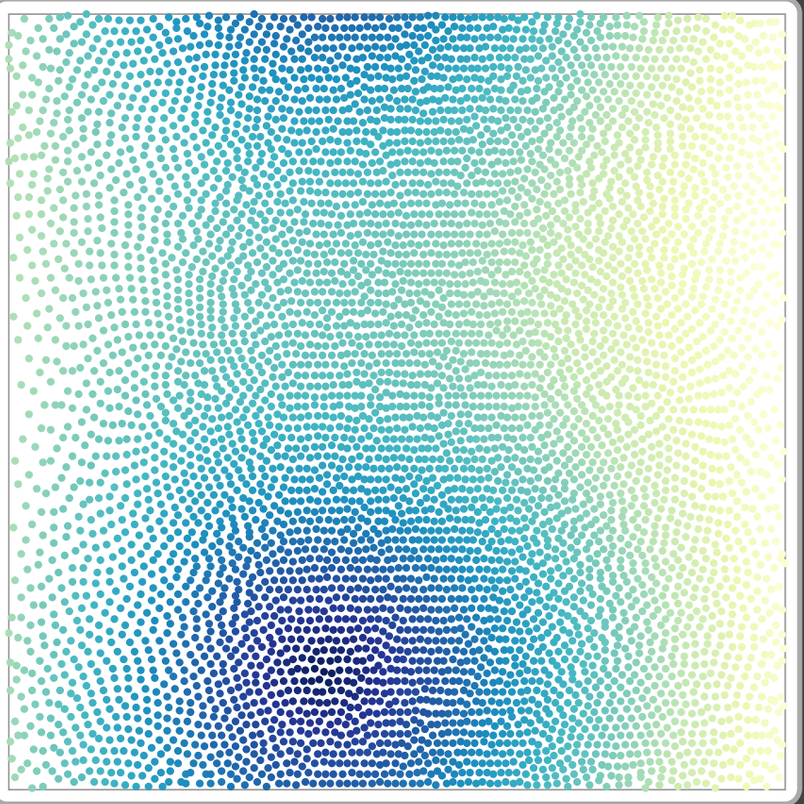


compare embryos *A* and *B*

summary

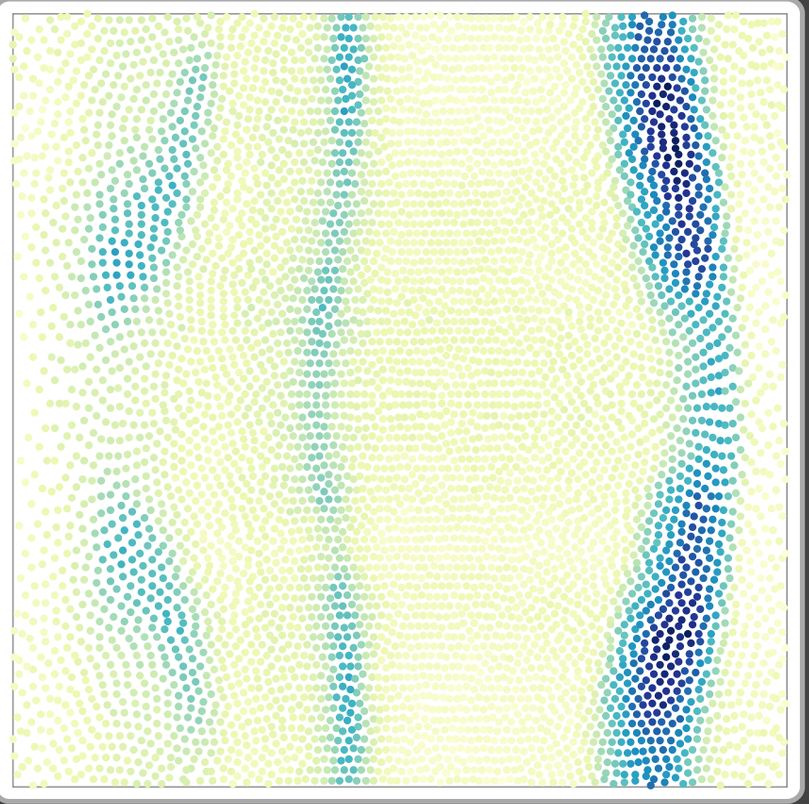
a value for each cell that summarizes the underlying data

spatial



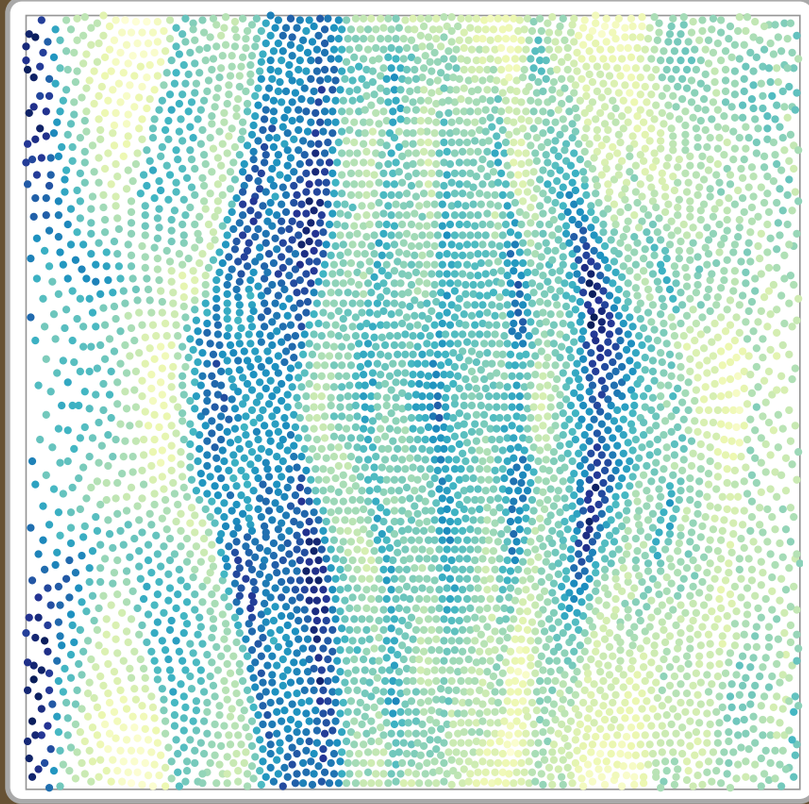
distance to cell *i*

expression



gene *j* at timepoint *k*

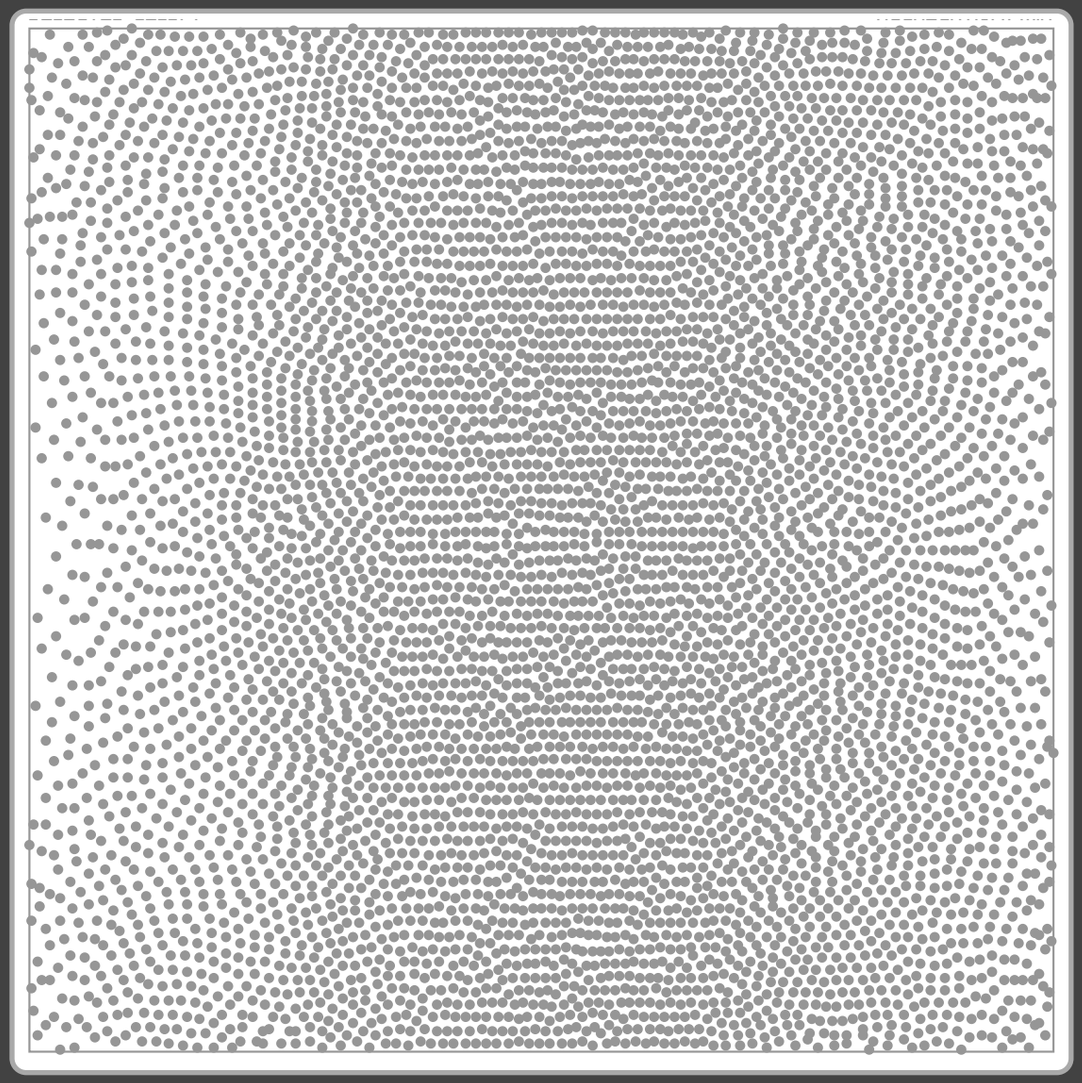
combination



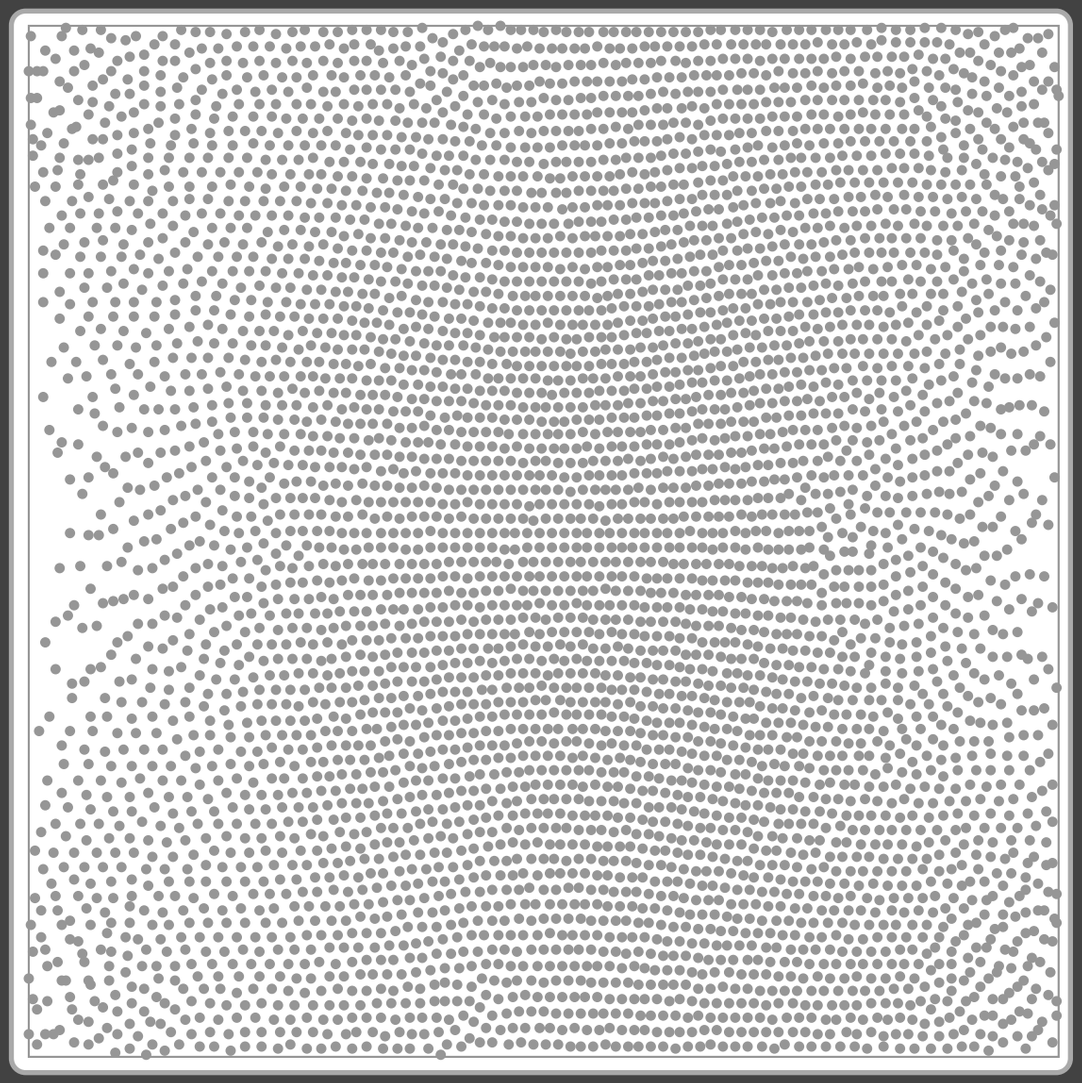
compare embryos *A* and *B*

creating a comparison summary

creating a comparison summary

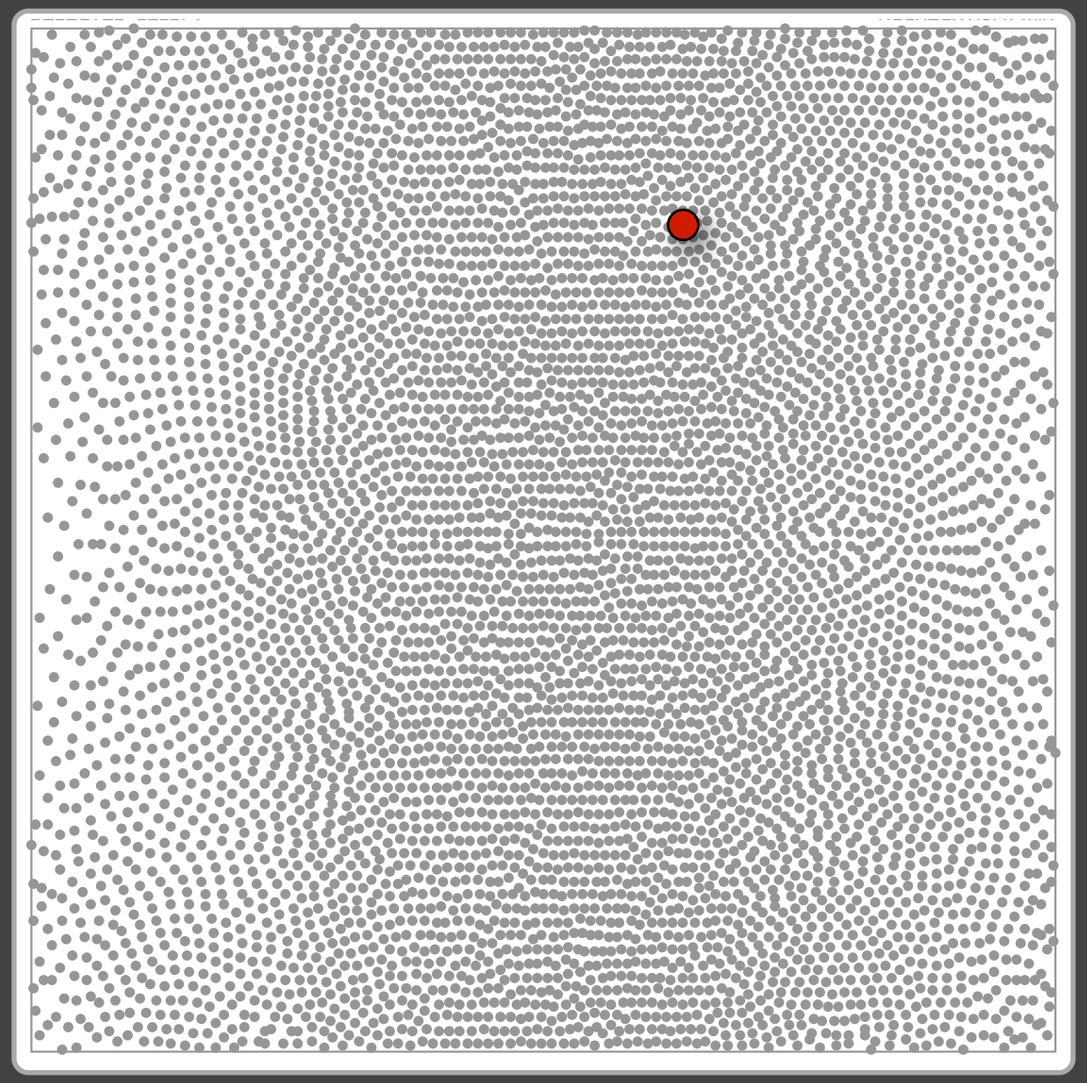


embryo A

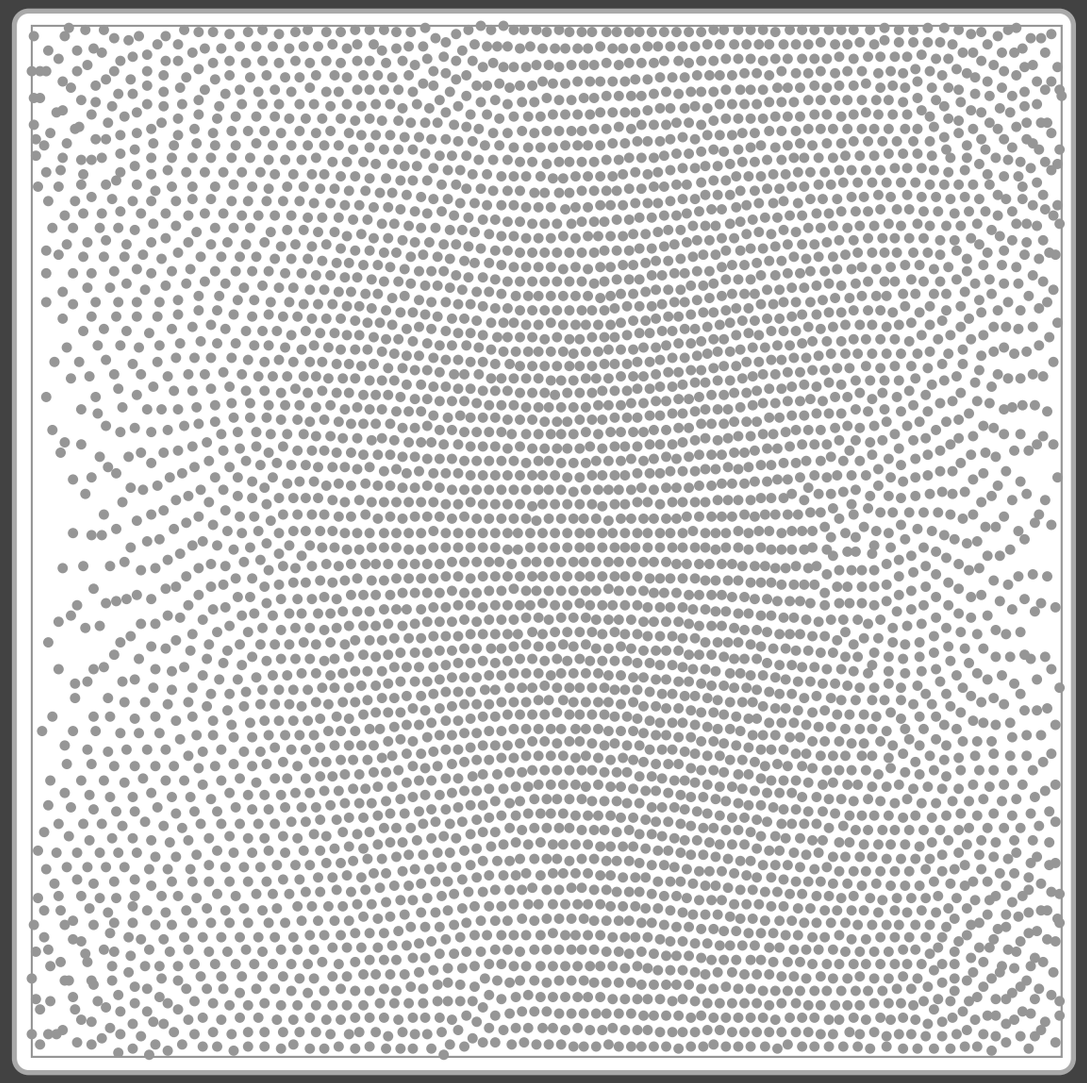


embryo B

creating a comparison summary

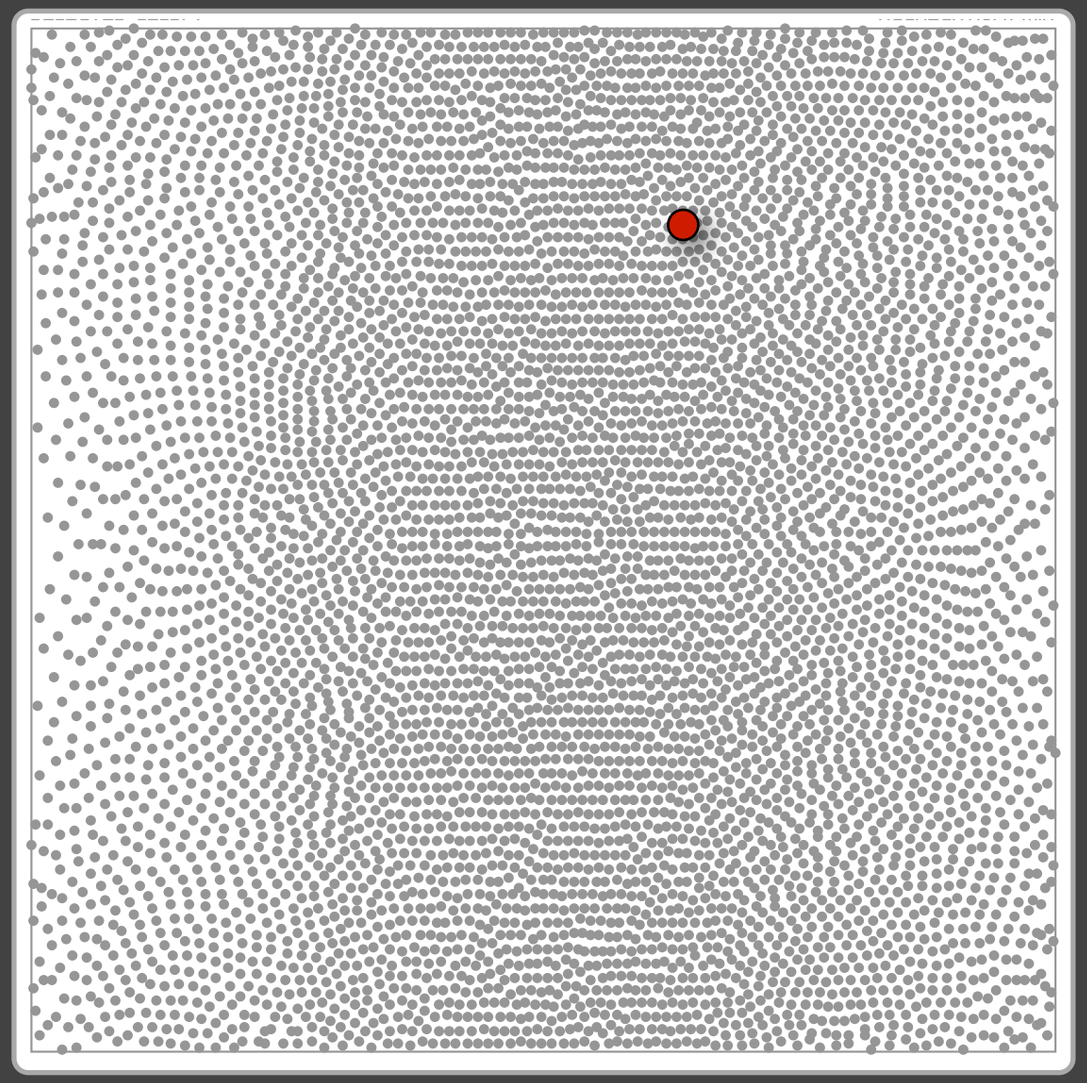


embryo A

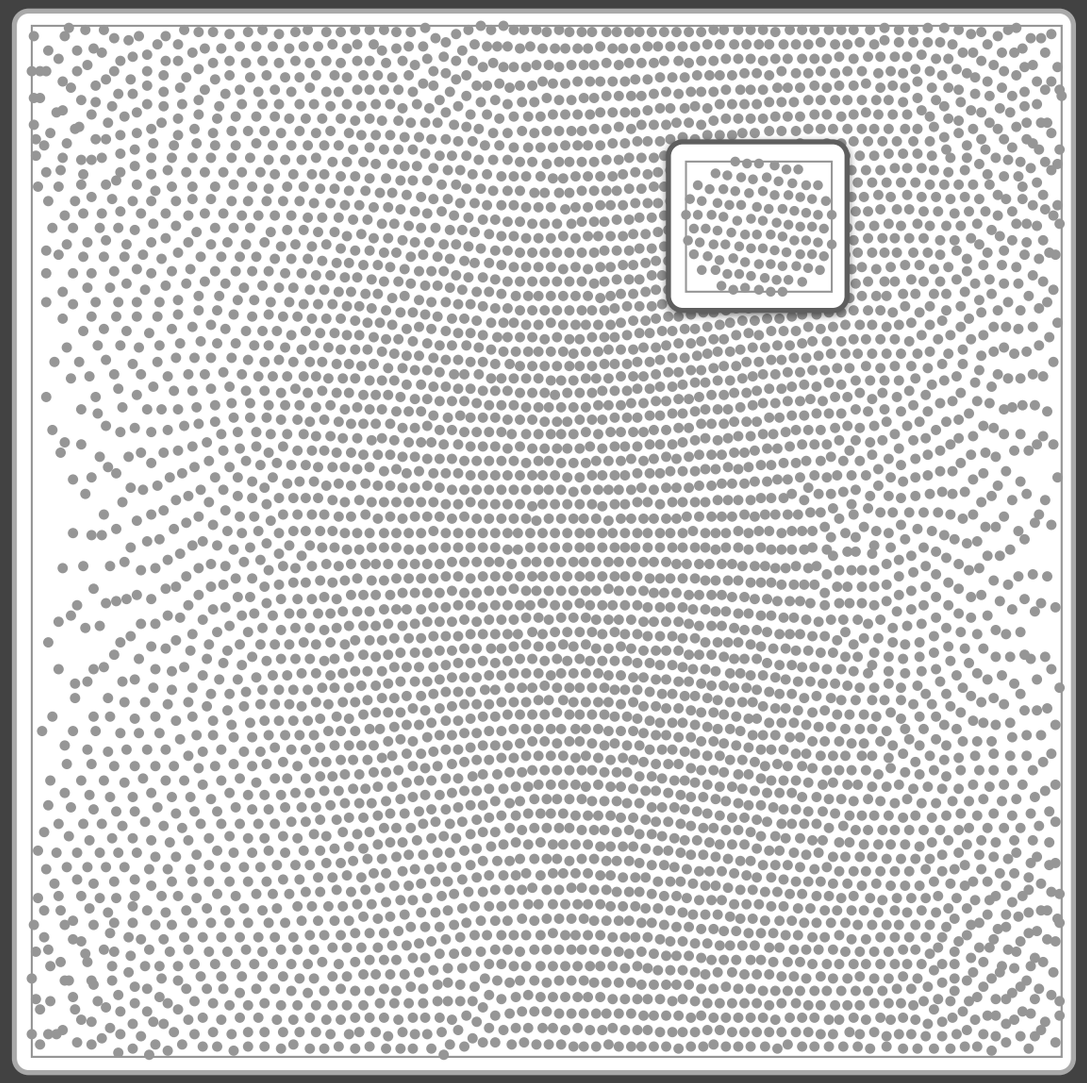


embryo B

creating a comparison summary

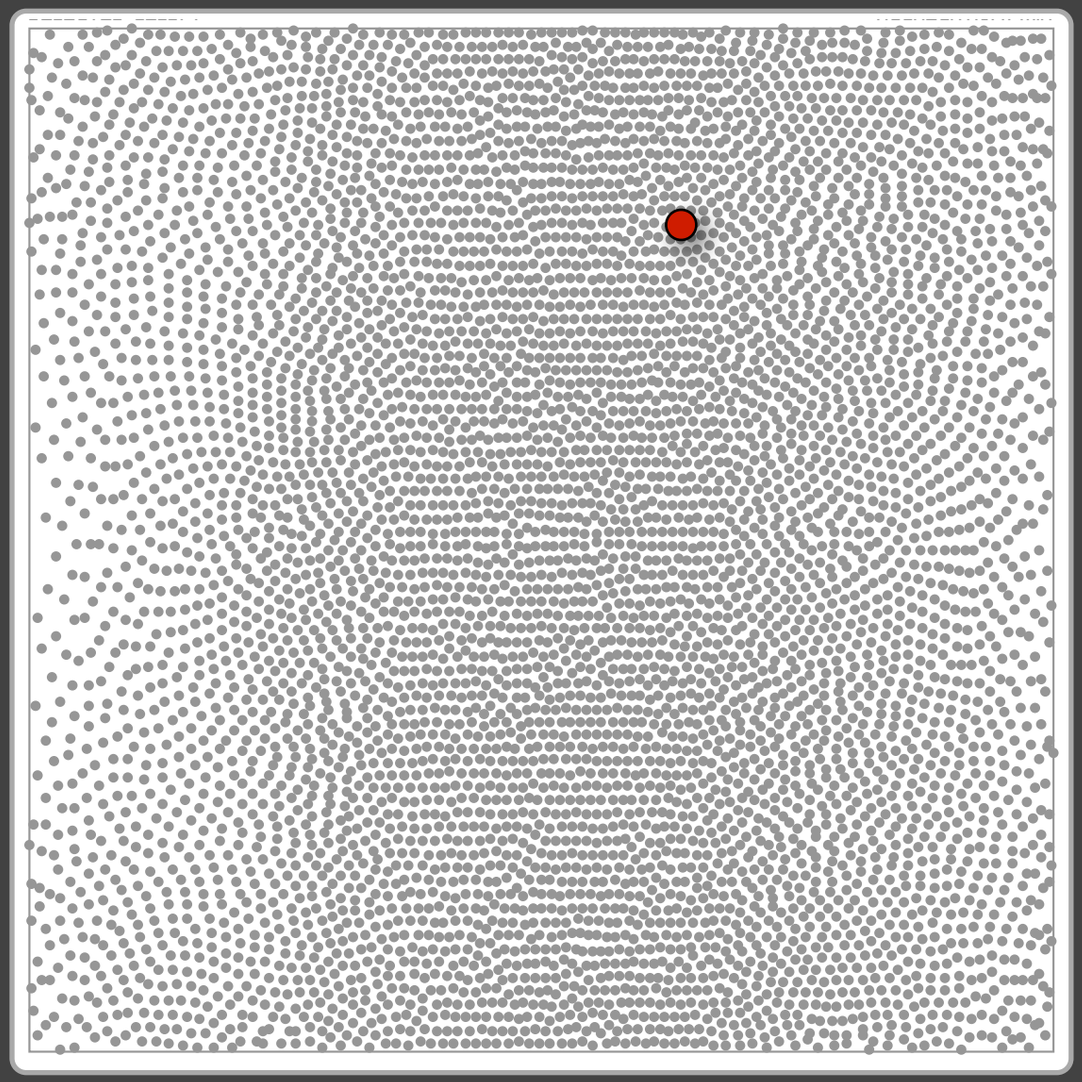


embryo A



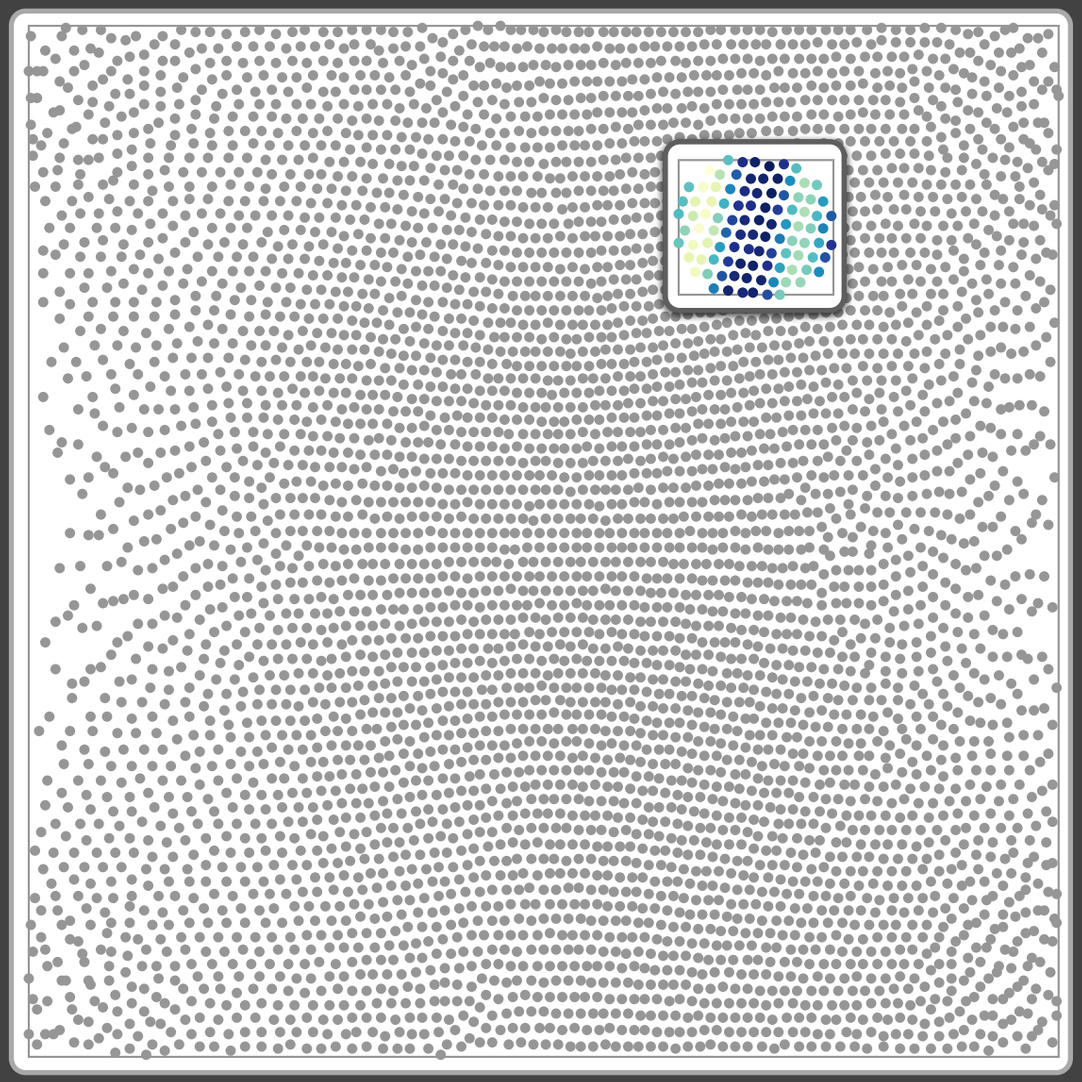
embryo B

creating a comparison summary



embryo A

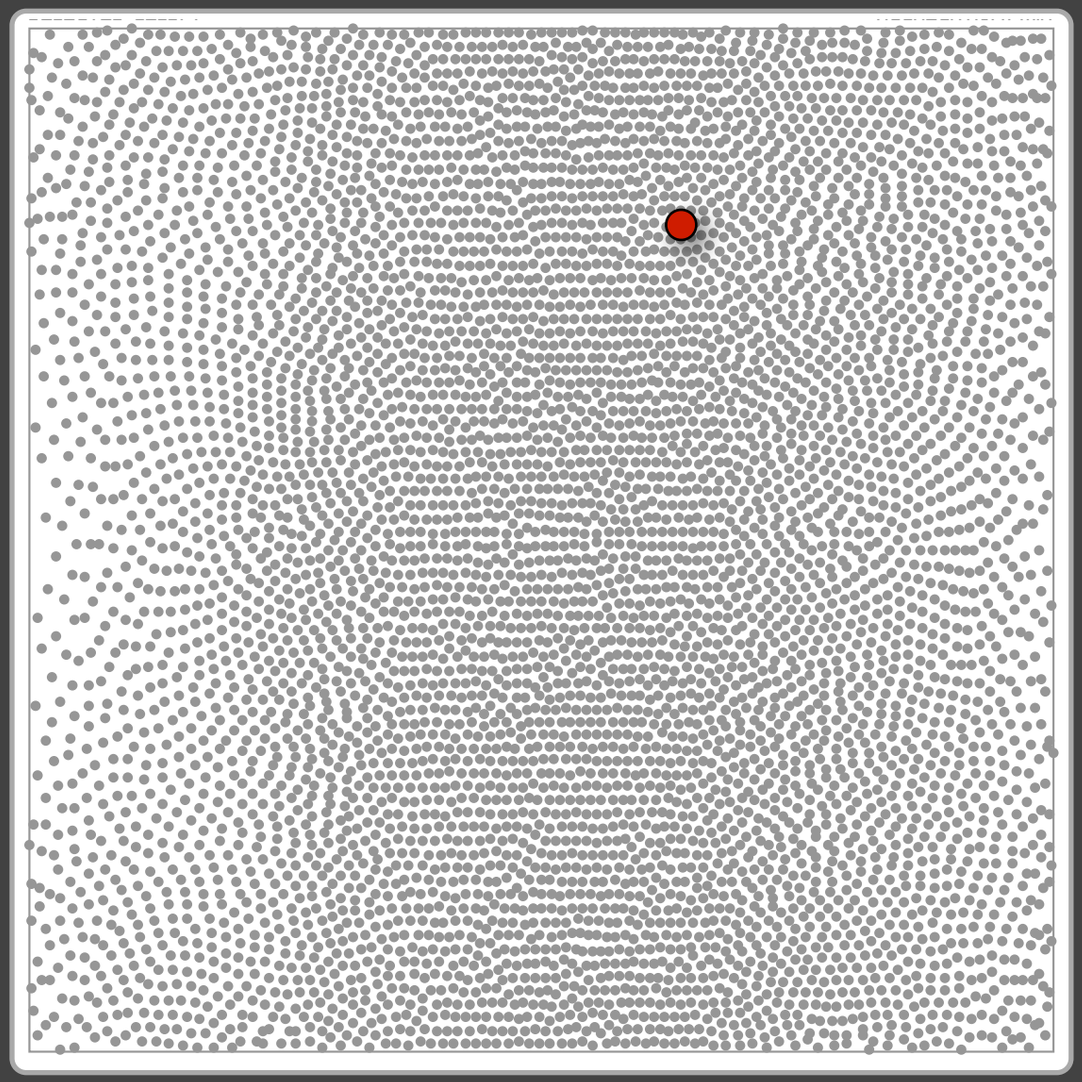
root-mean-square distance



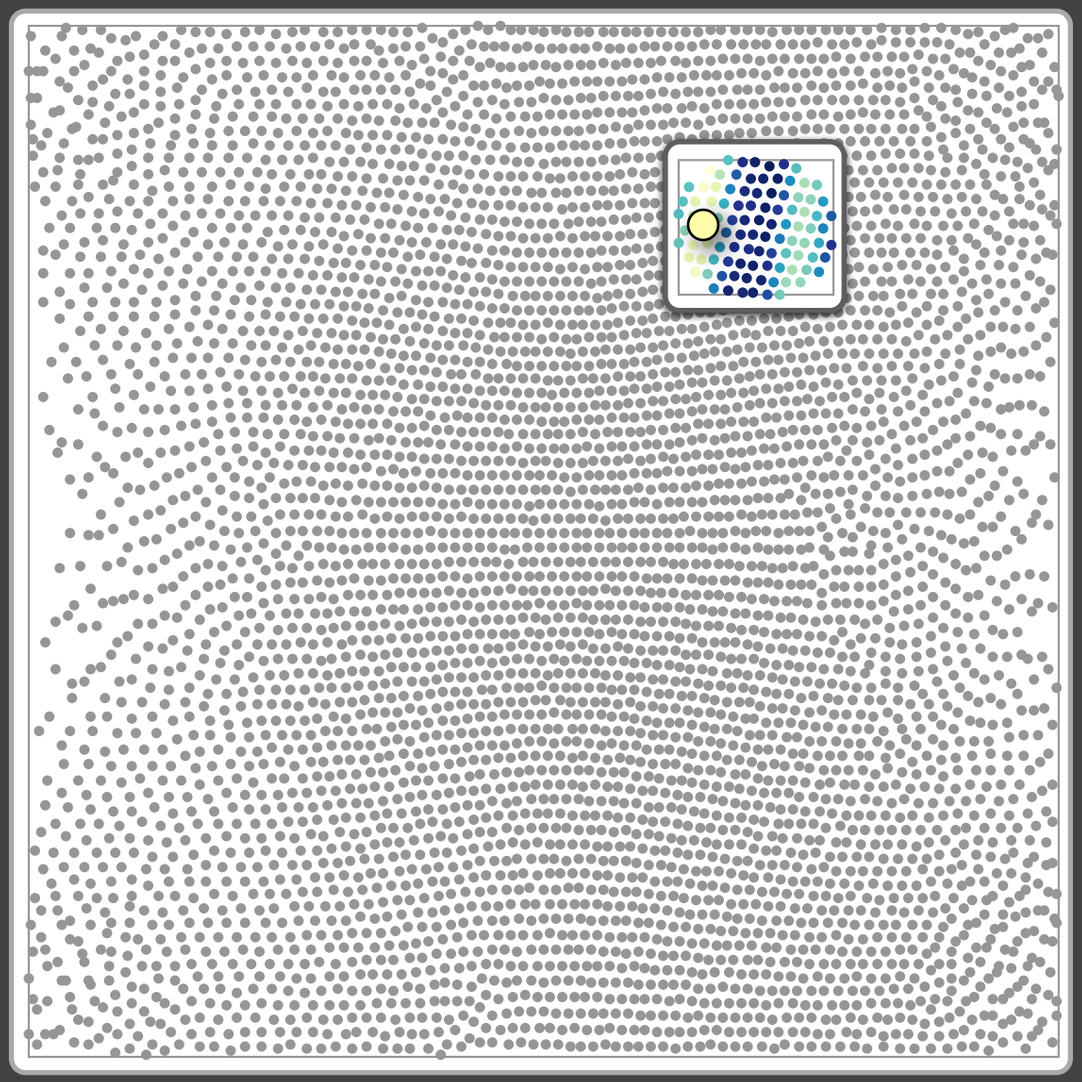
embryo B

creating a comparison summary

root-mean-square distance

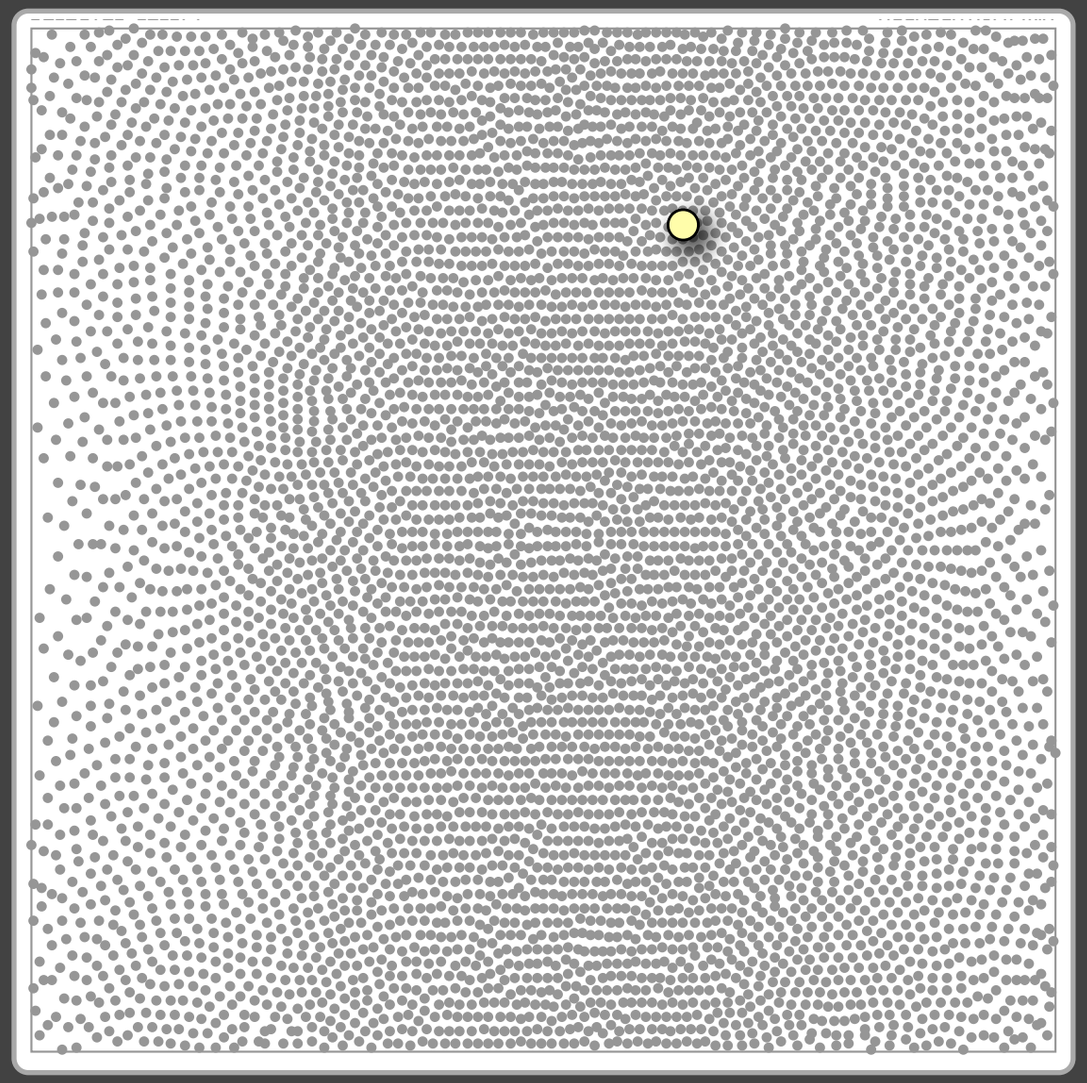


embryo A



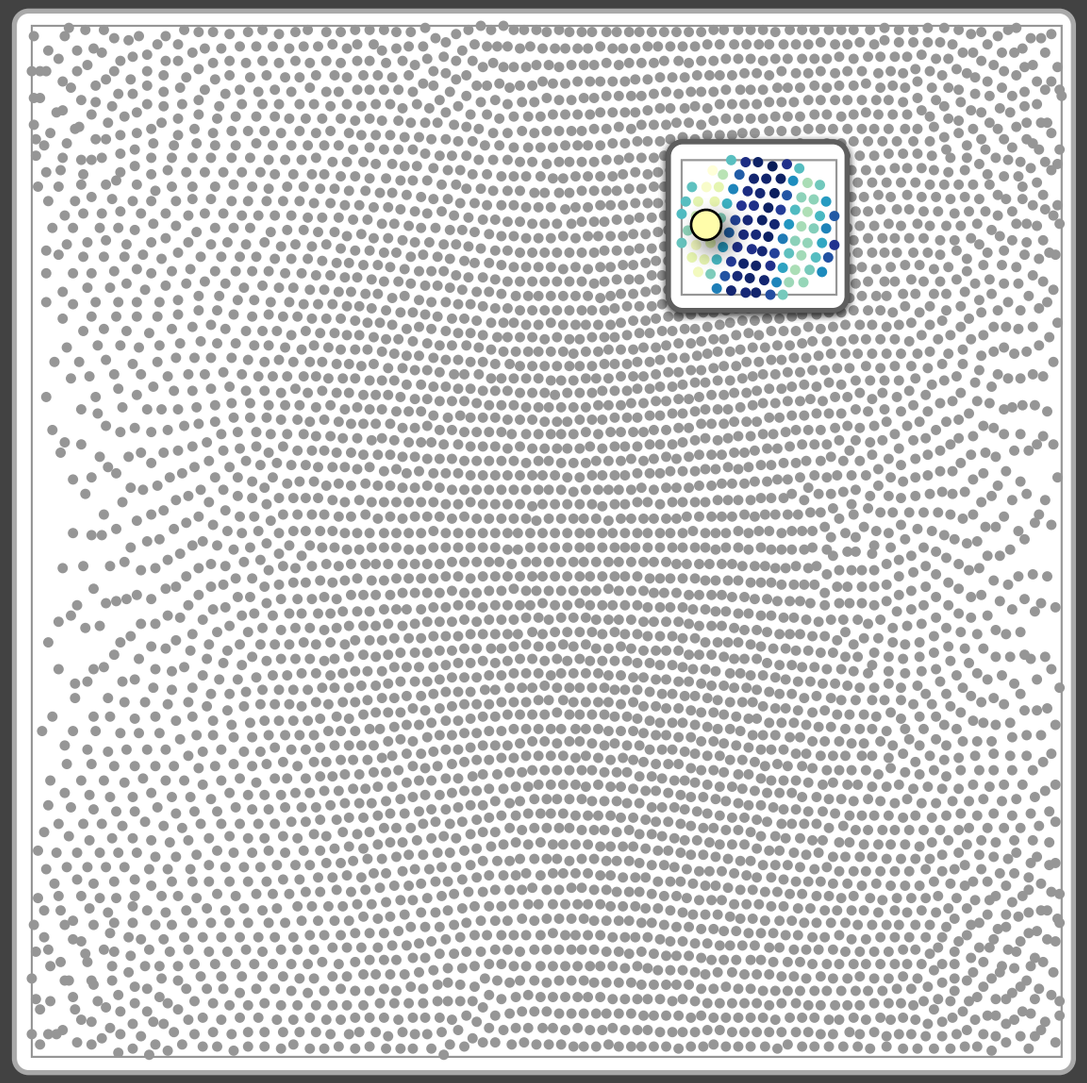
embryo B

creating a comparison summary



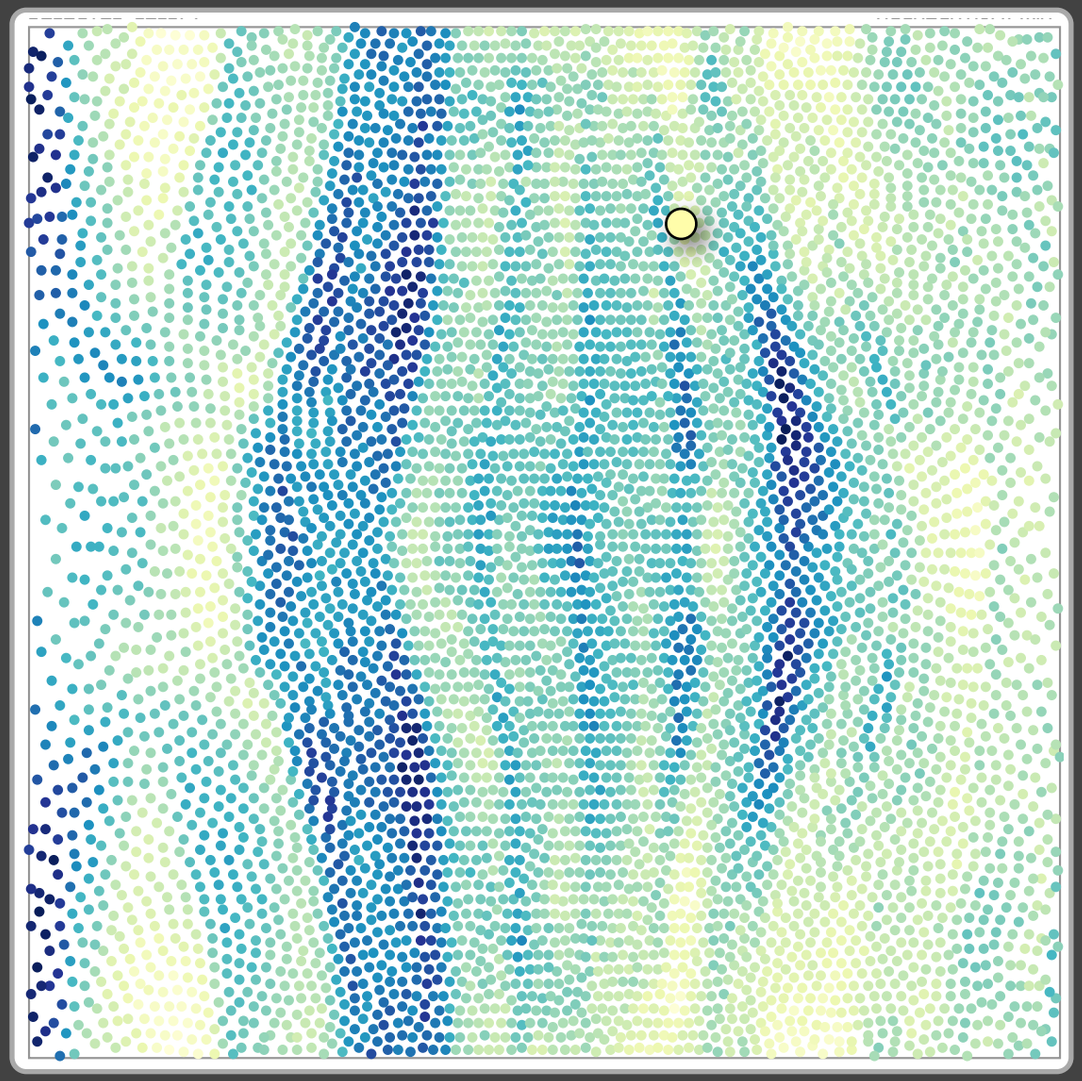
embryo A

root-mean-square distance



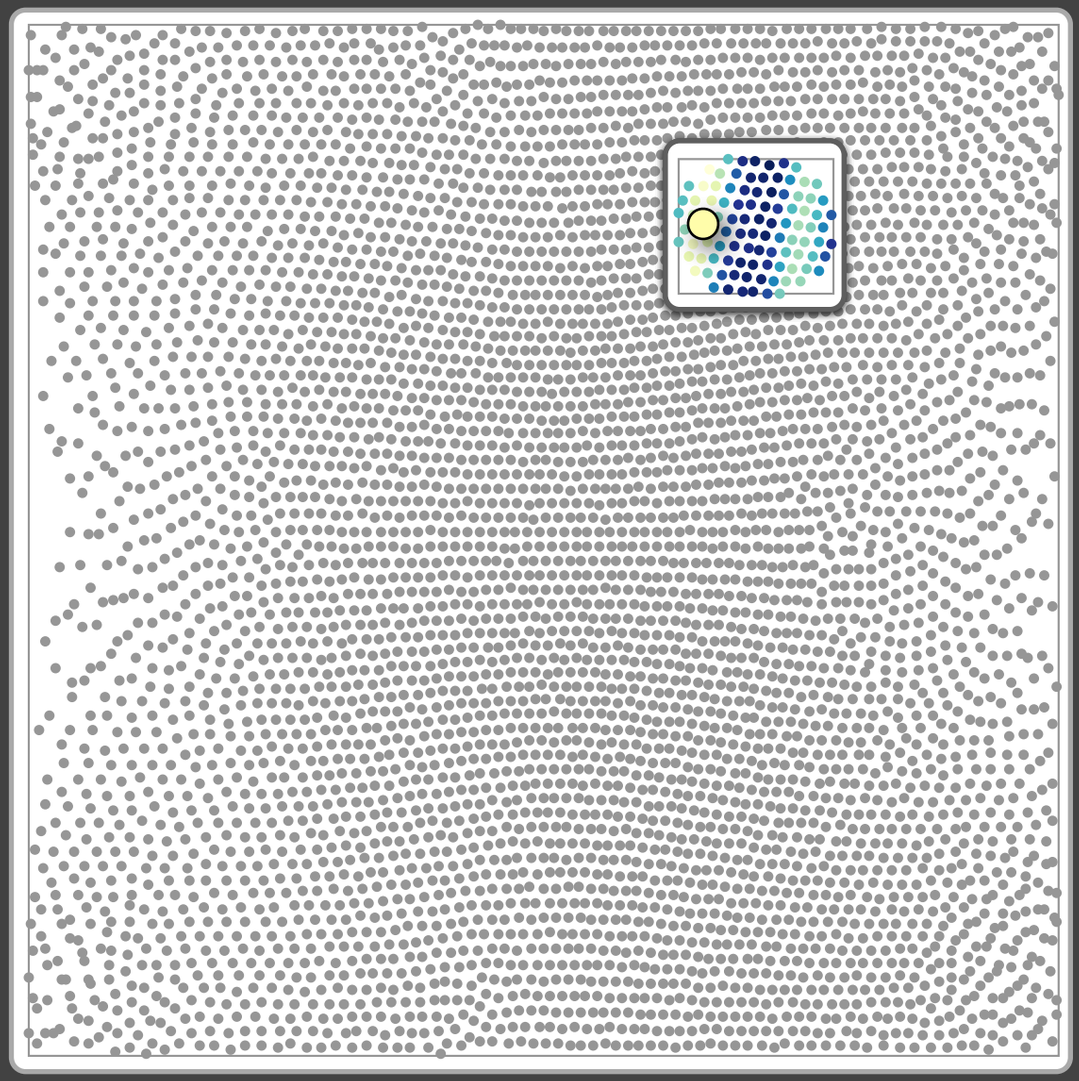
embryo B

creating a comparison summary



embryo A

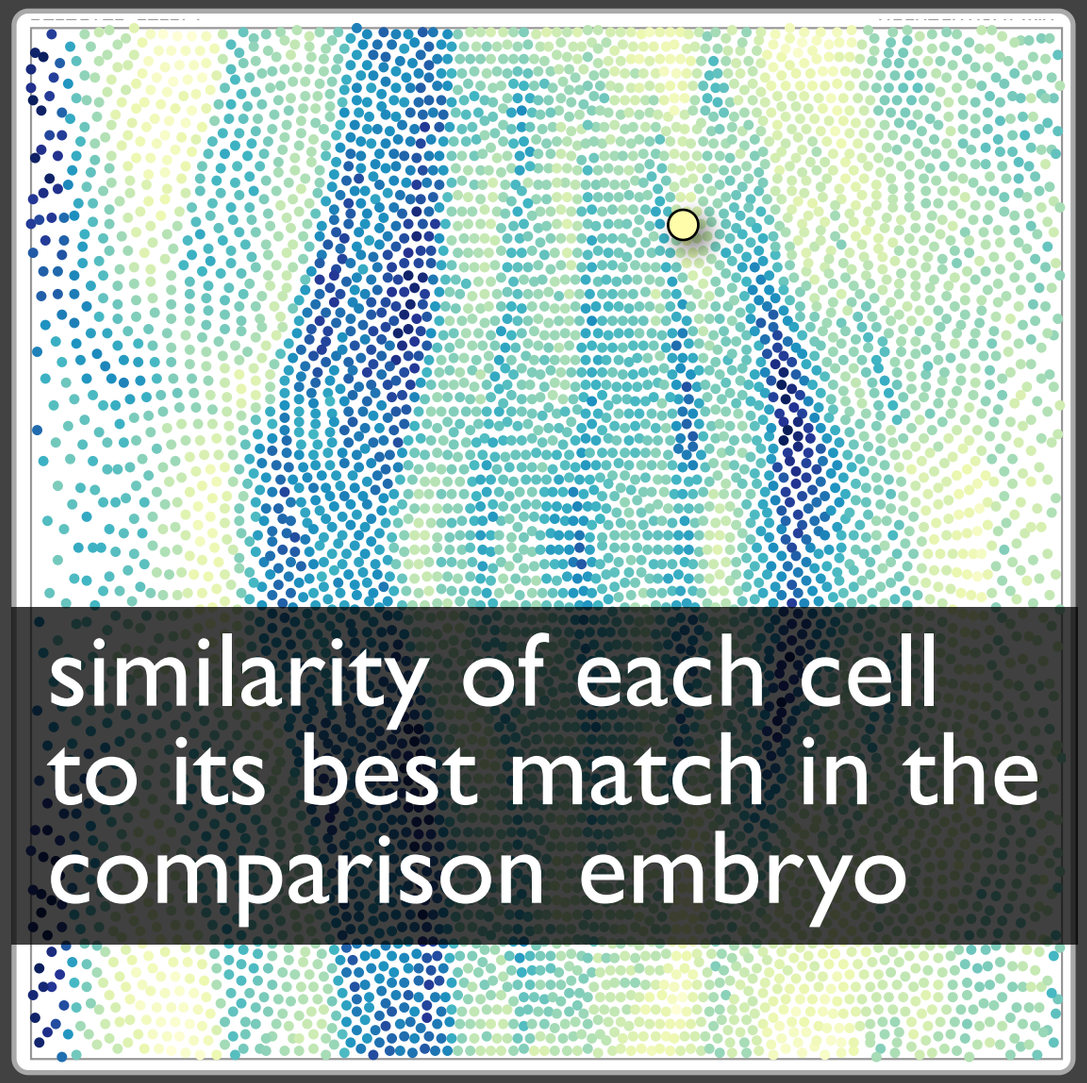
root-mean-square distance



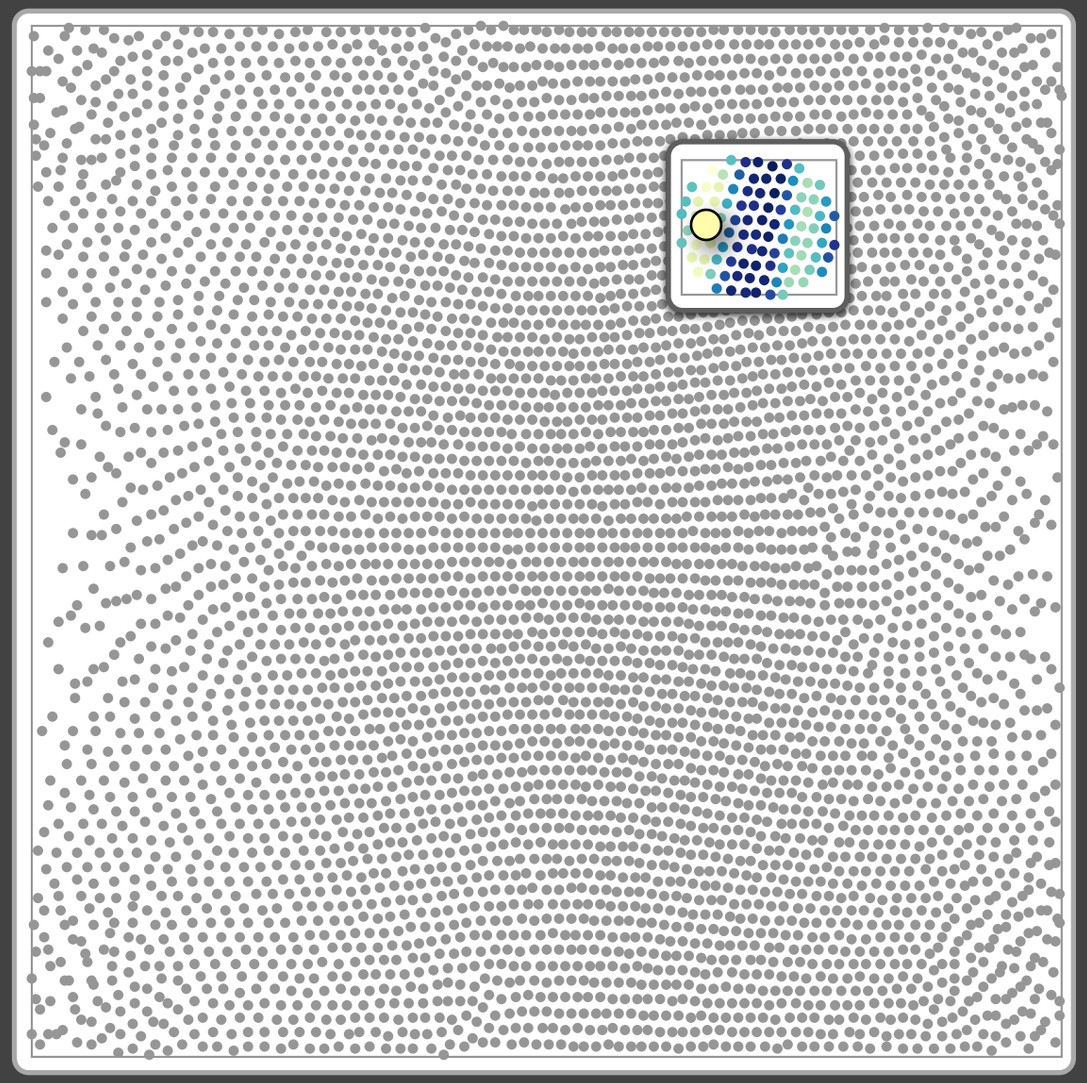
embryo B

creating a comparison summary

root-mean-square distance



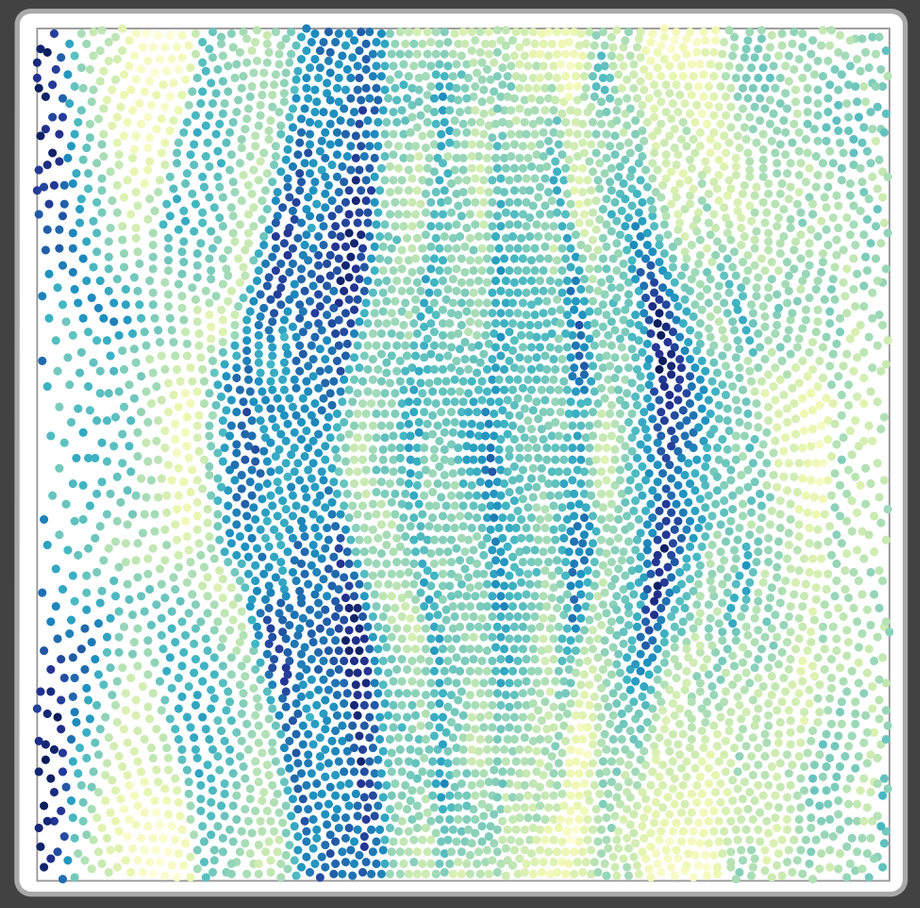
similarity of each cell to its best match in the comparison embryo



embryo A

embryo B

comparative summary components

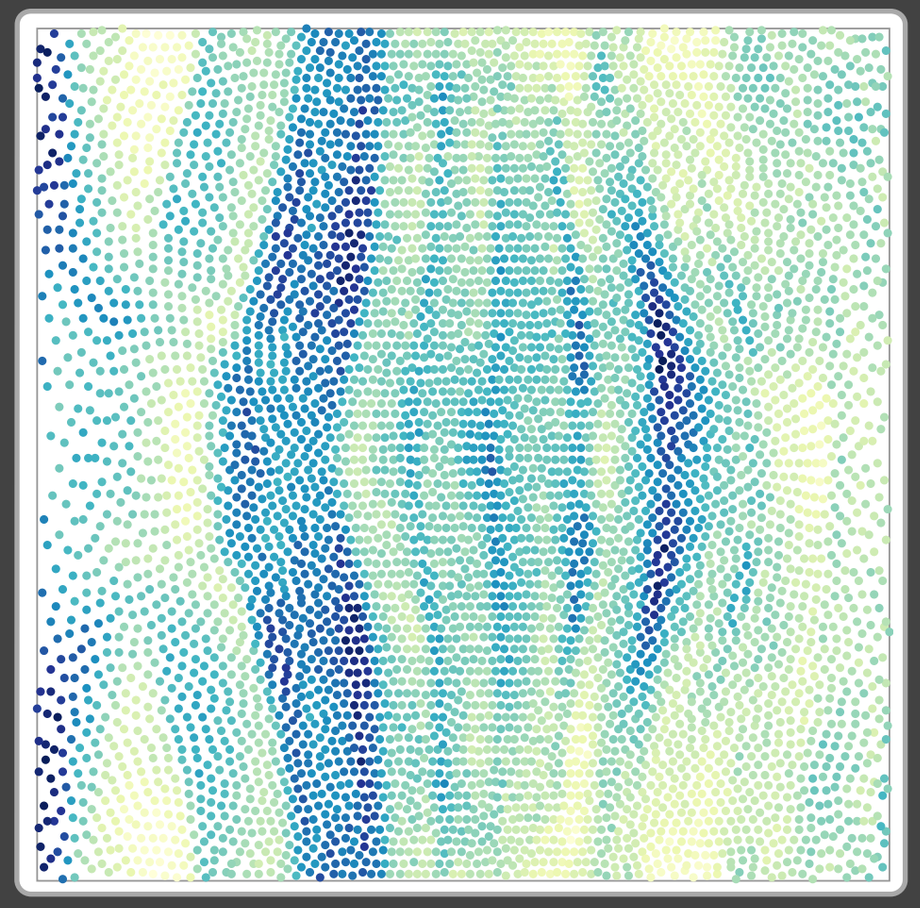


aggregation group

metric

aggregation

comparative summary components

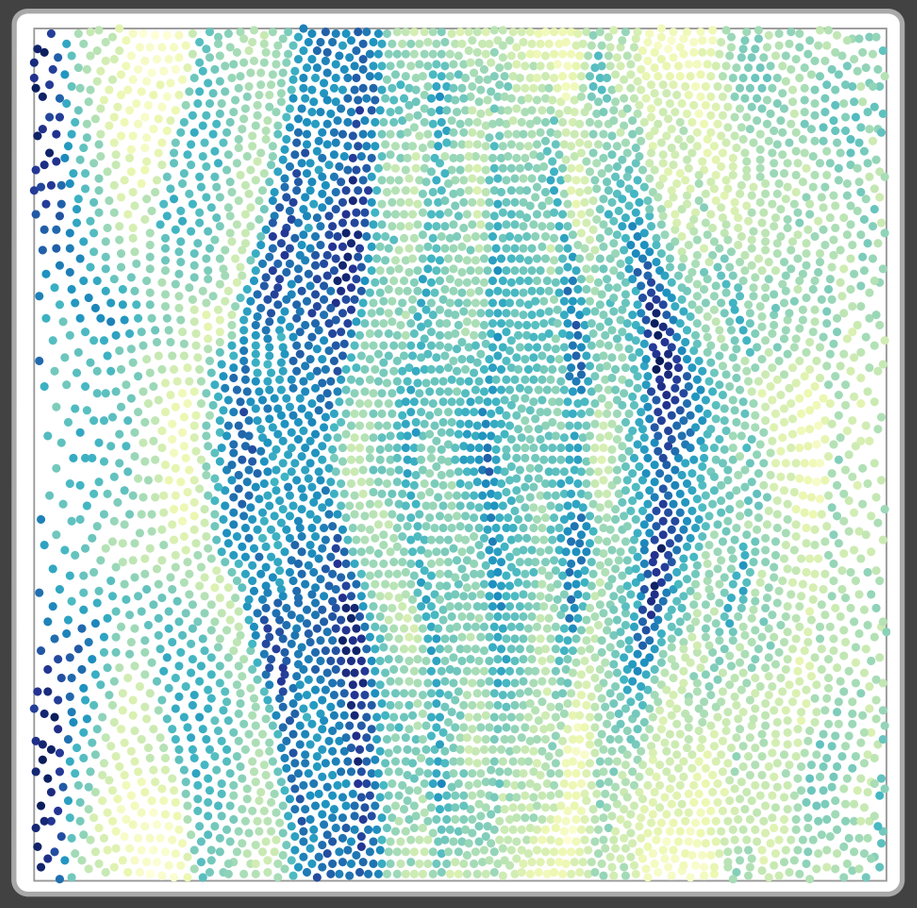


aggregation group: *100 spatially closest cells*

metric

aggregation

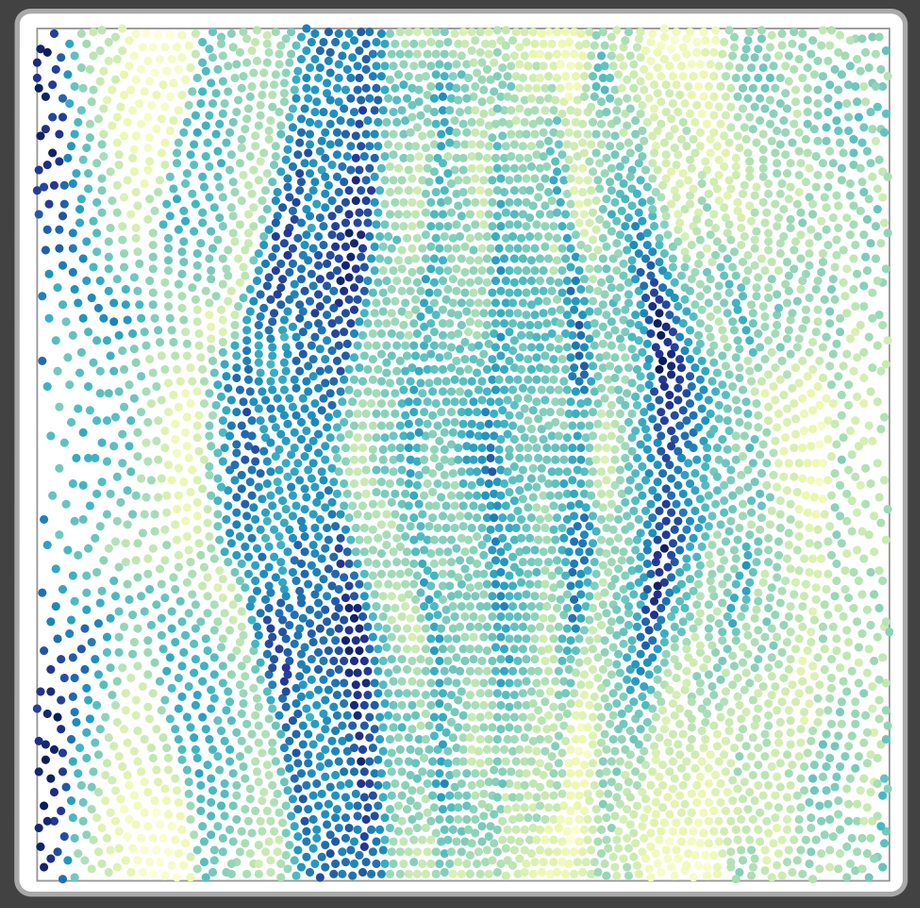
comparative summary components



aggregation group: *100 spatially closest cells*

metric: *root-mean-square distance*

aggregation



comparative summary components

aggregation group: *100 spatially closest cells*

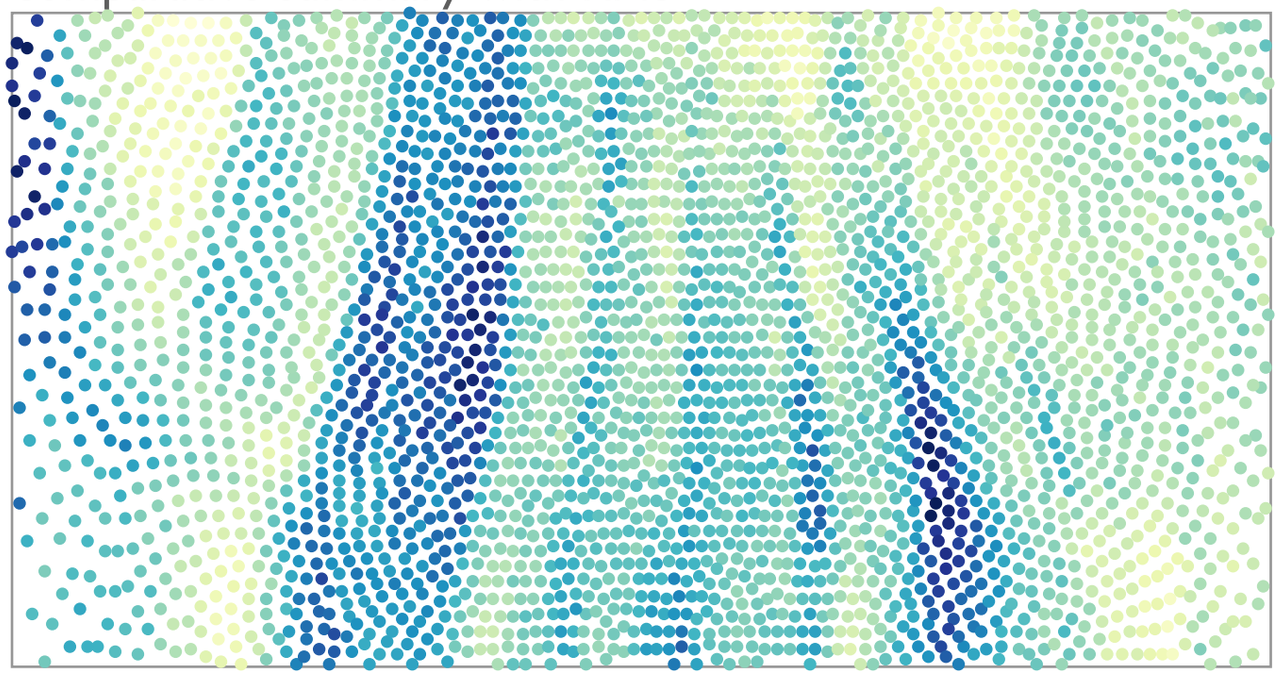
metric: *root-mean-square distance*

aggregation: *min operator*

case study one:

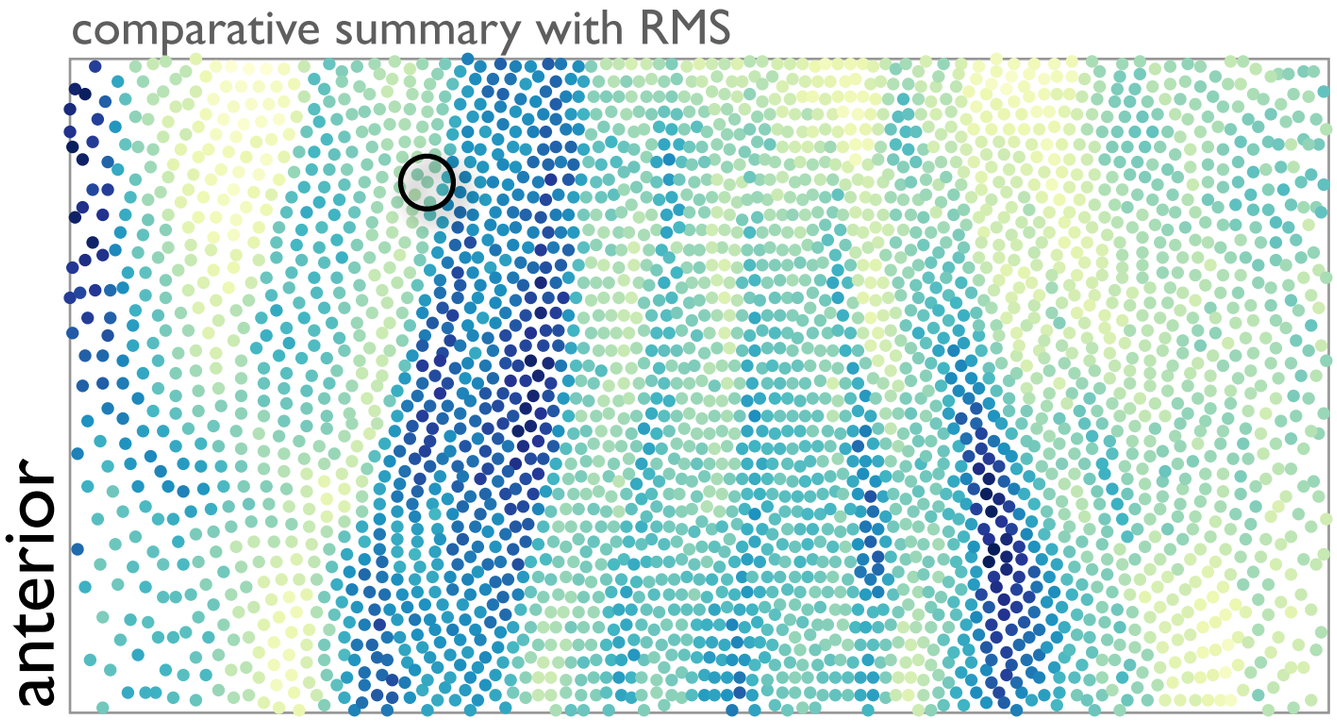
characterizing a summary

comparative summary with RMS



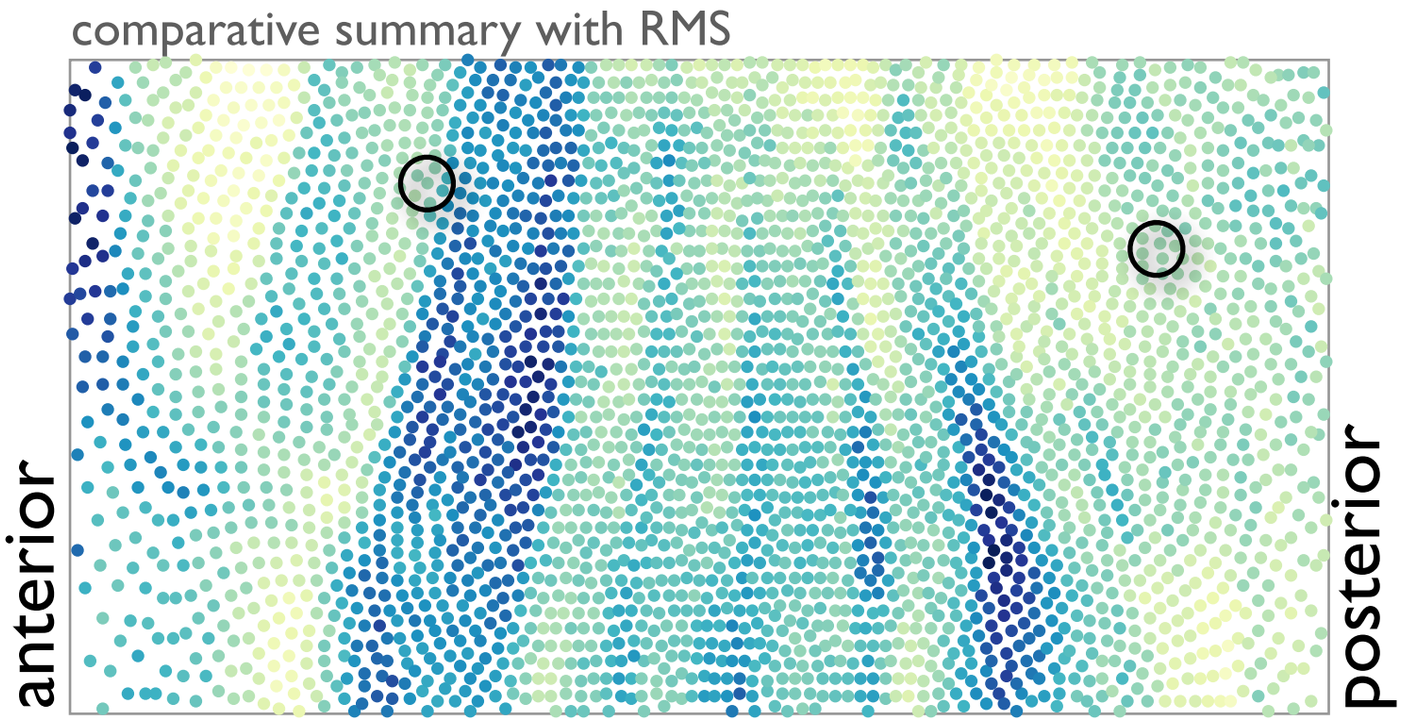
case study one:

characterizing a summary



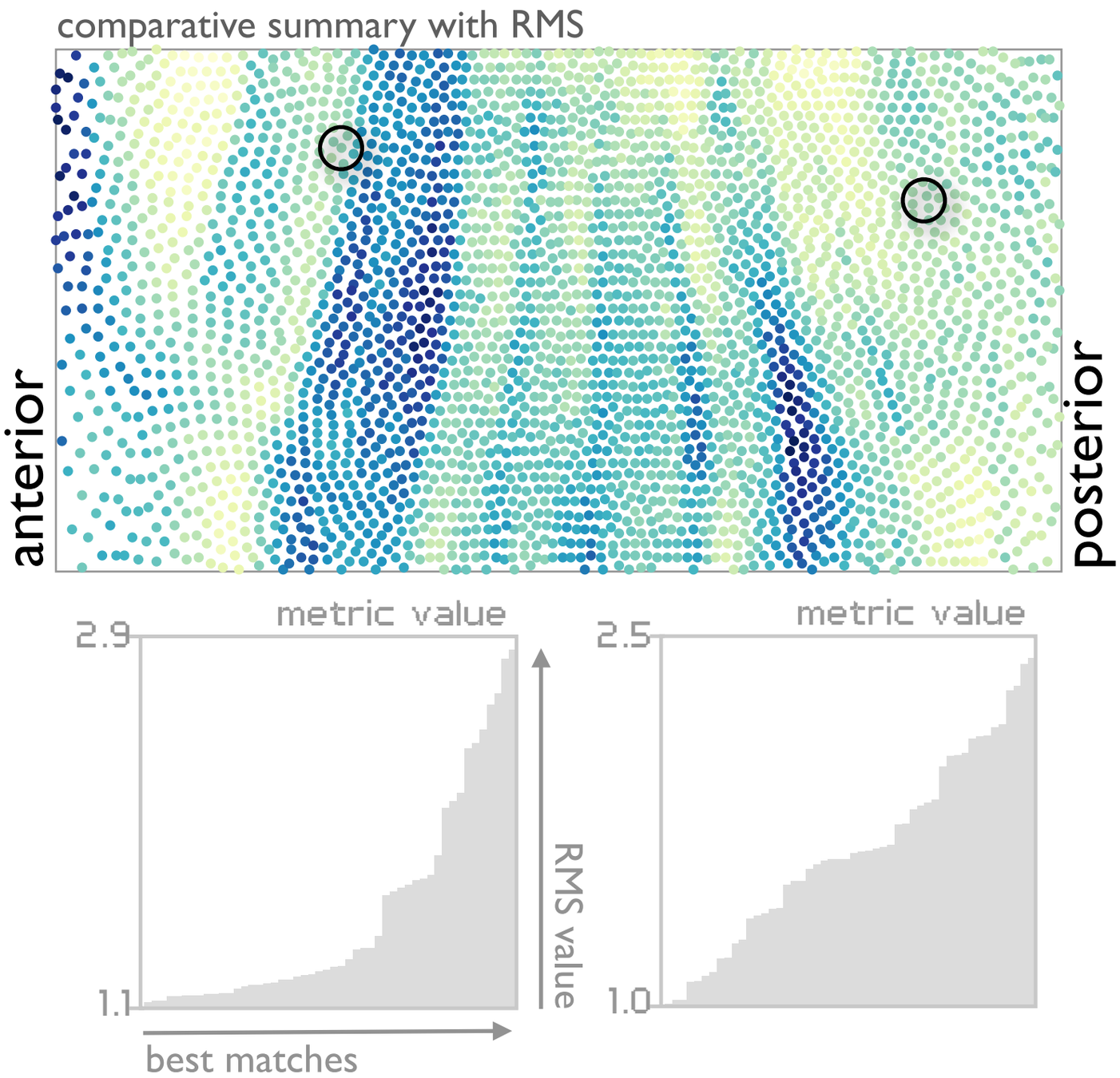
case study one:

characterizing a summary



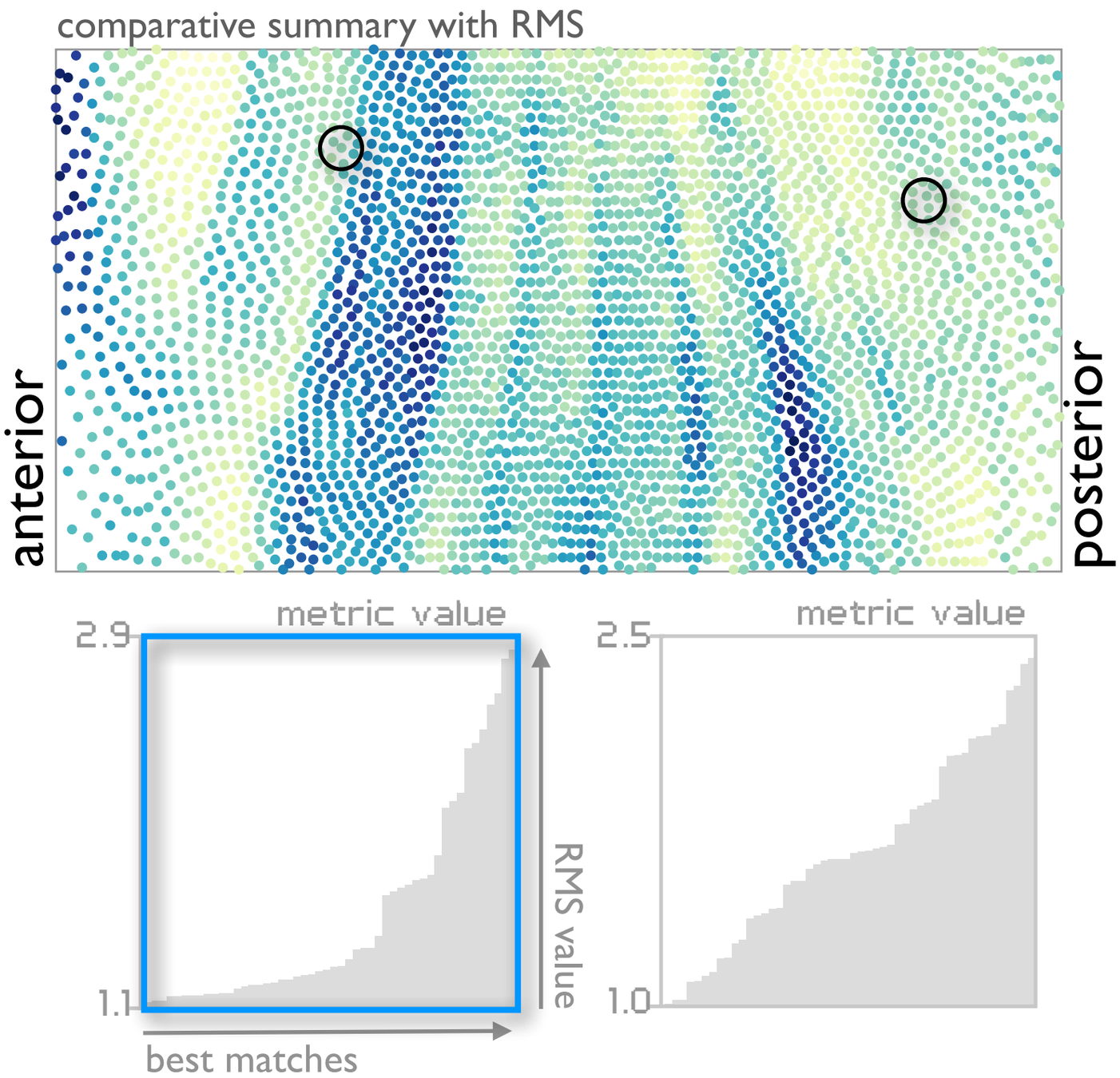
case study one:

characterizing a summary



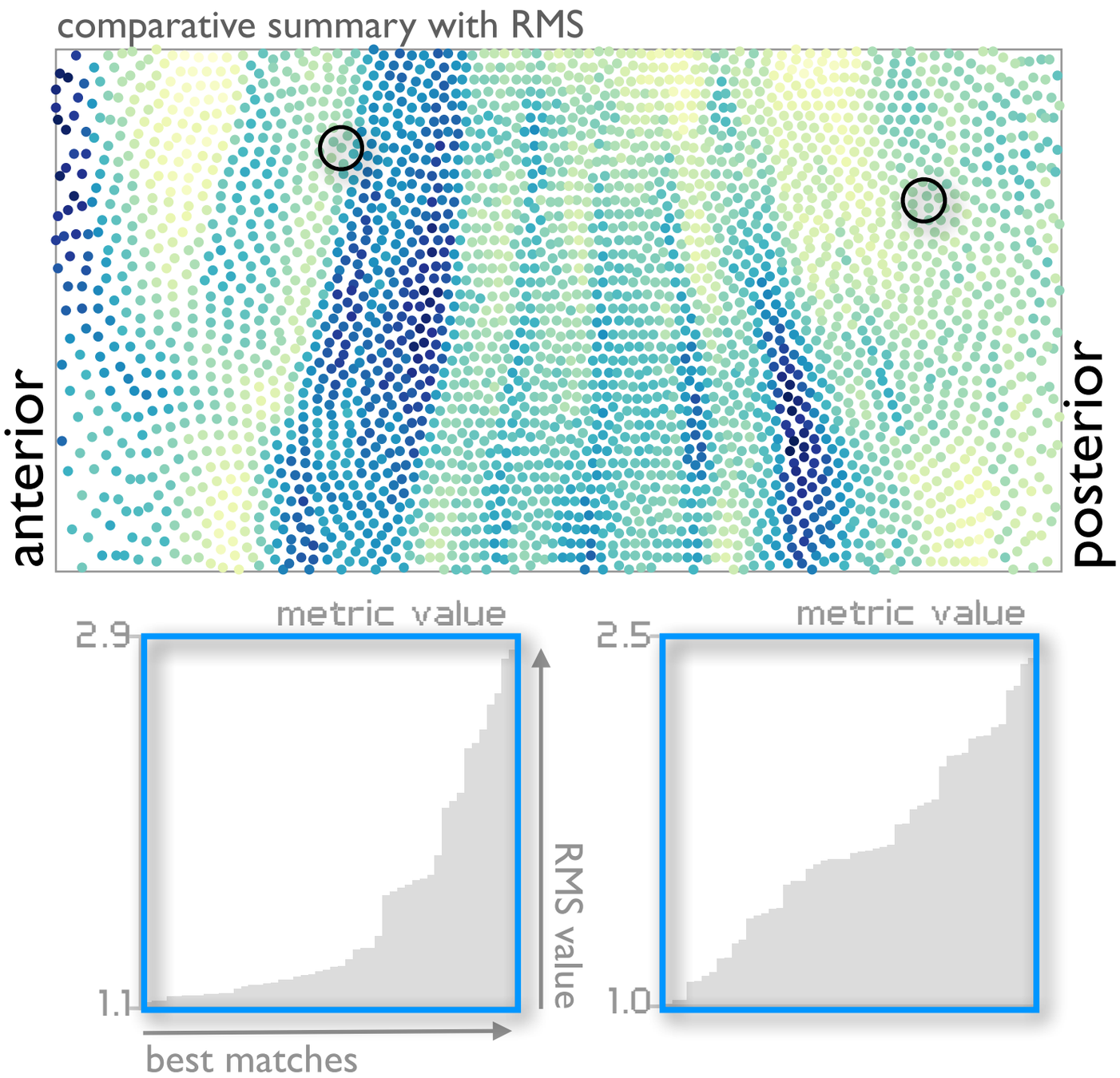
case study one:

characterizing a summary



case study one:

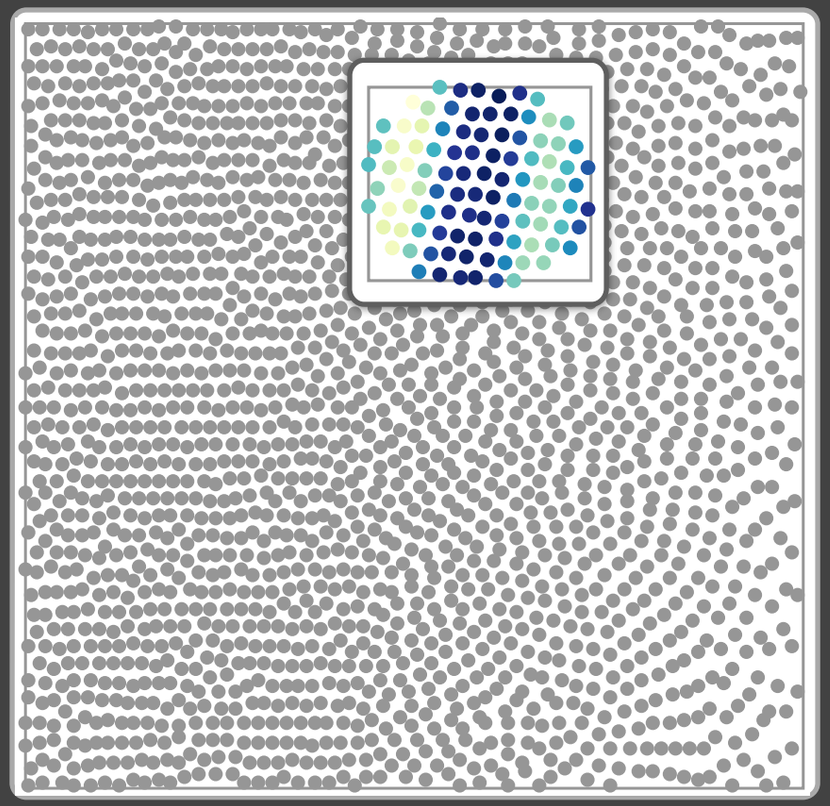
characterizing a summary



group
a set of cells

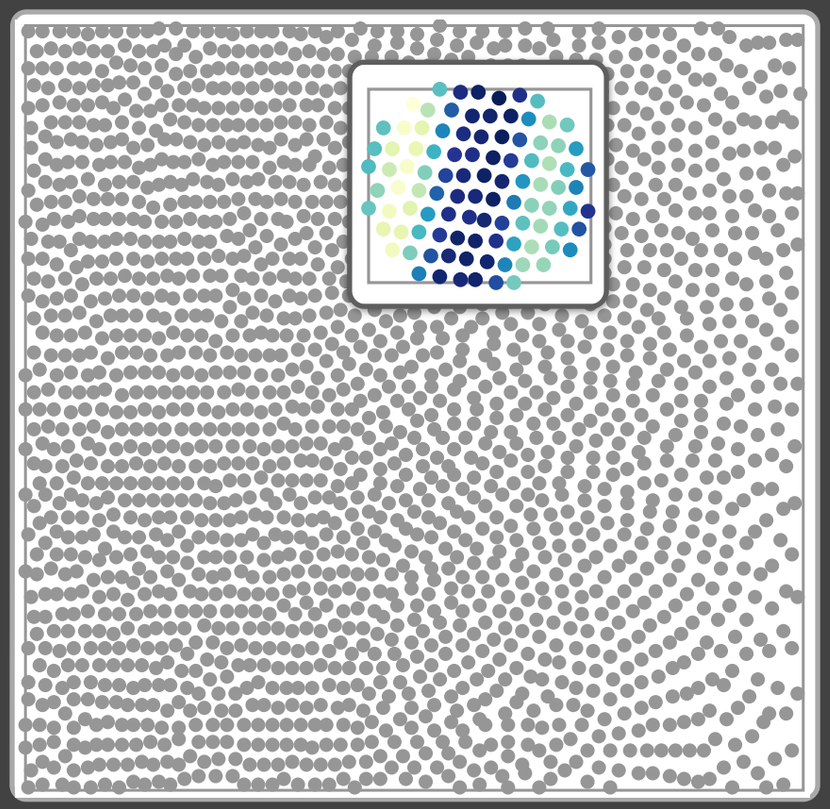
group
a set of cells

aggregation group

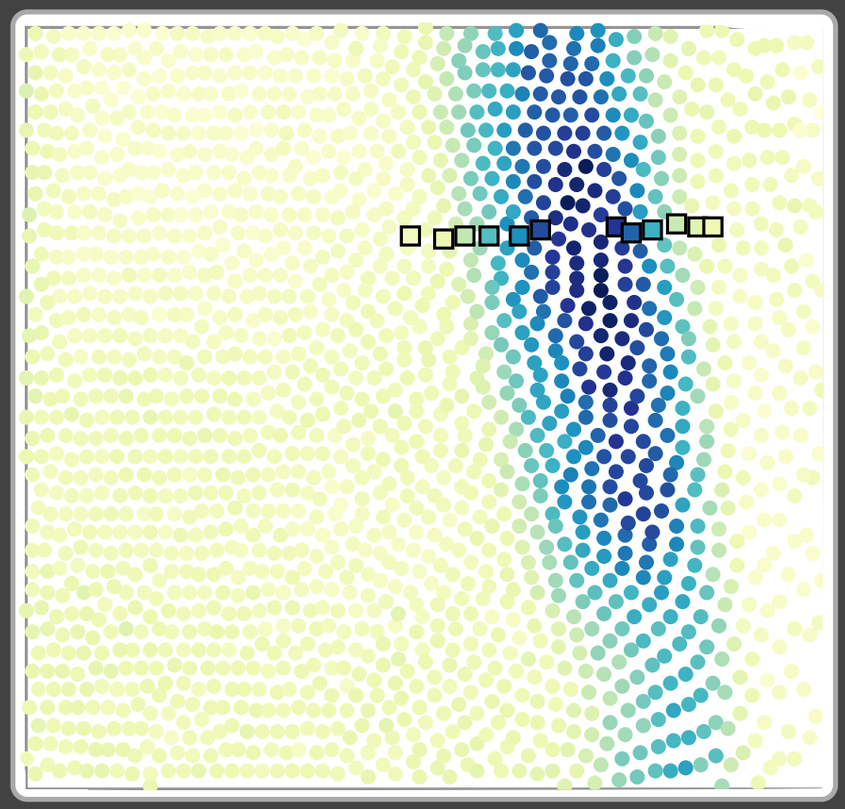


group
a set of cells

aggregation group

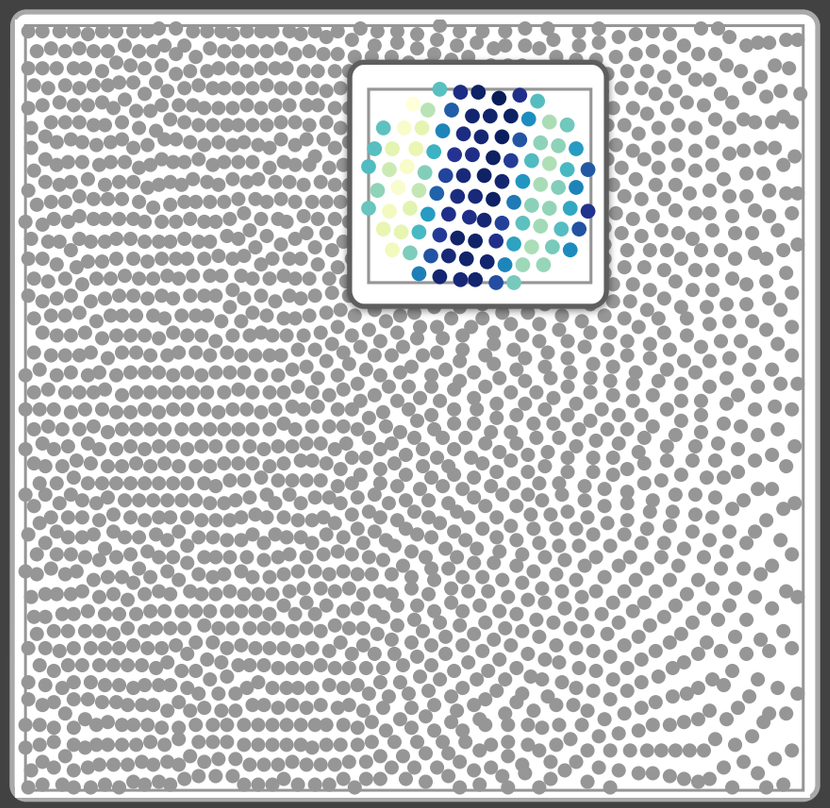


created group

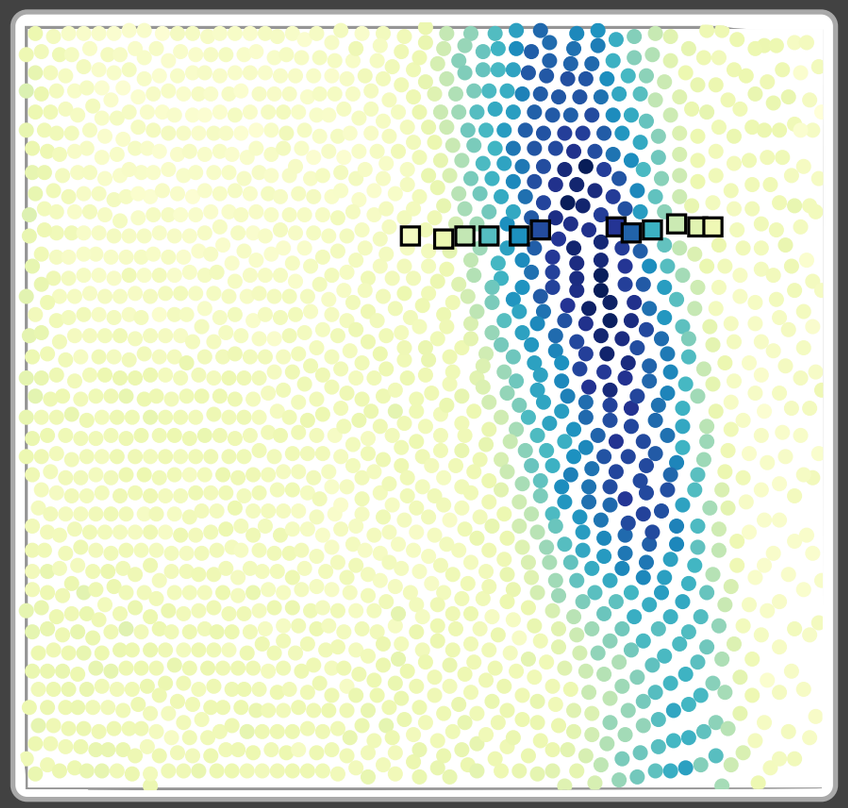


group
a set of cells

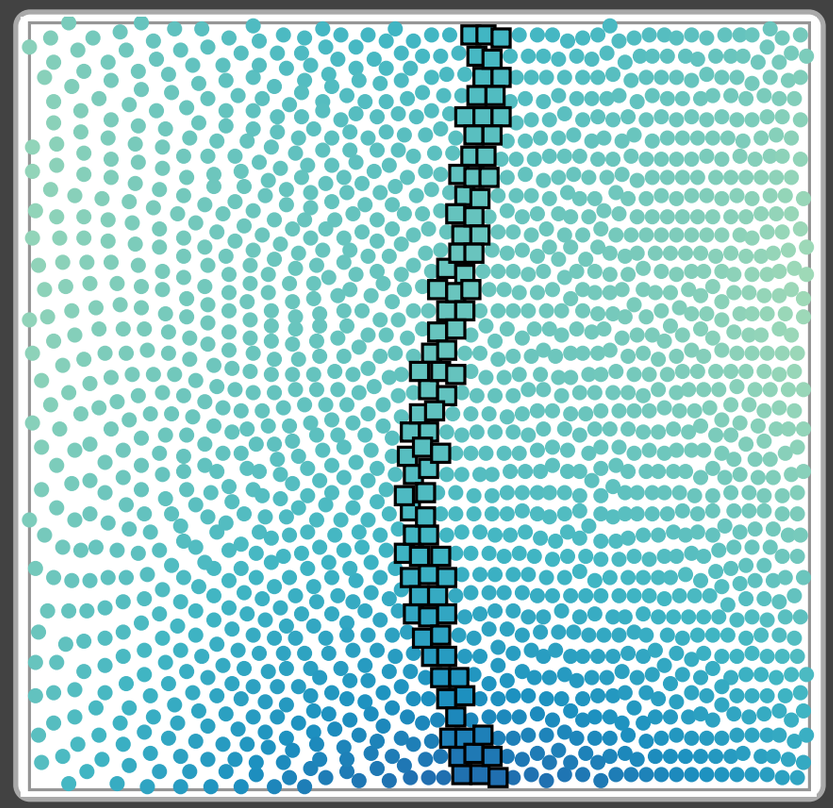
aggregation group



created group



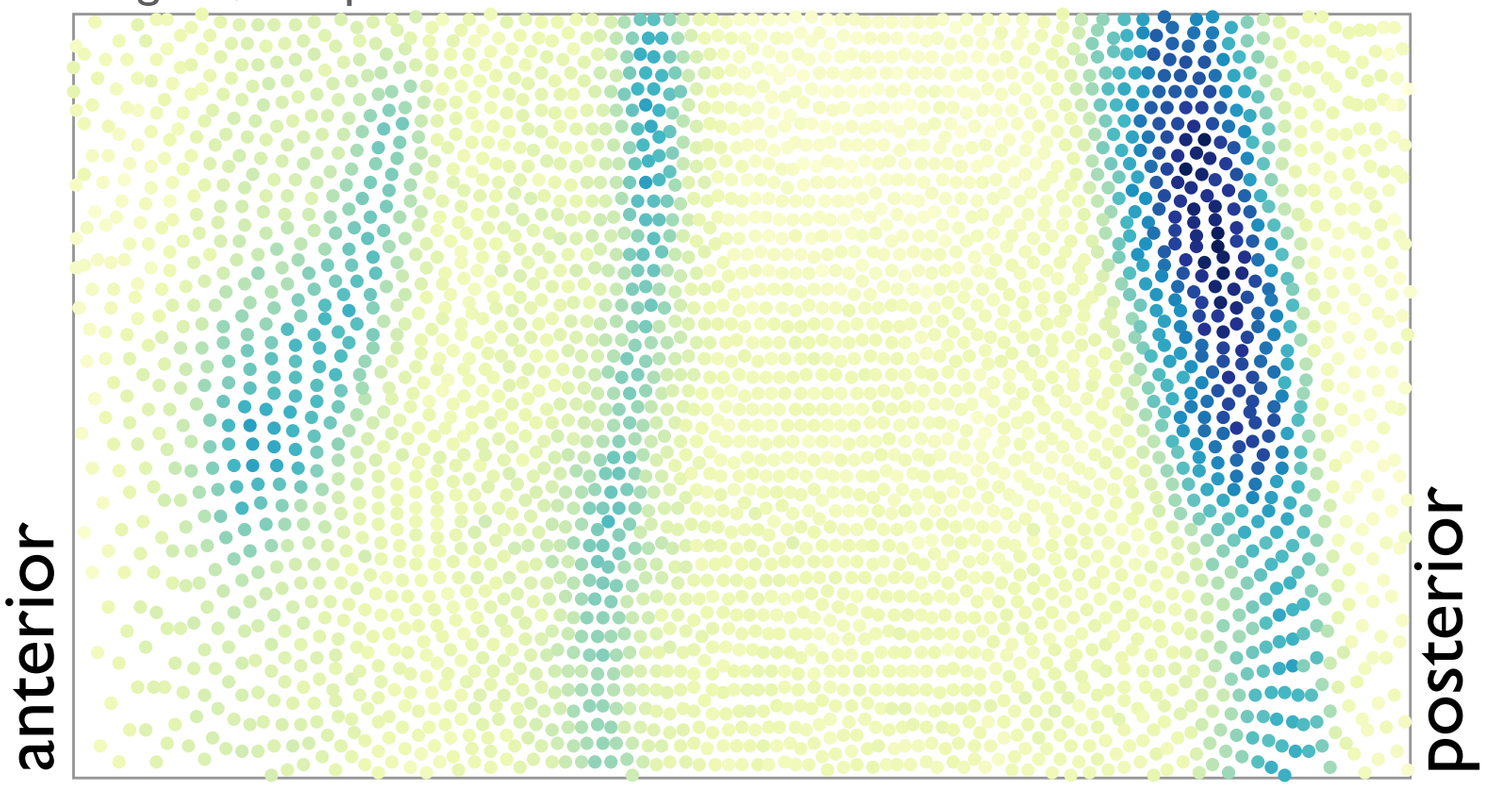
existing group



case study two:

exploring groups

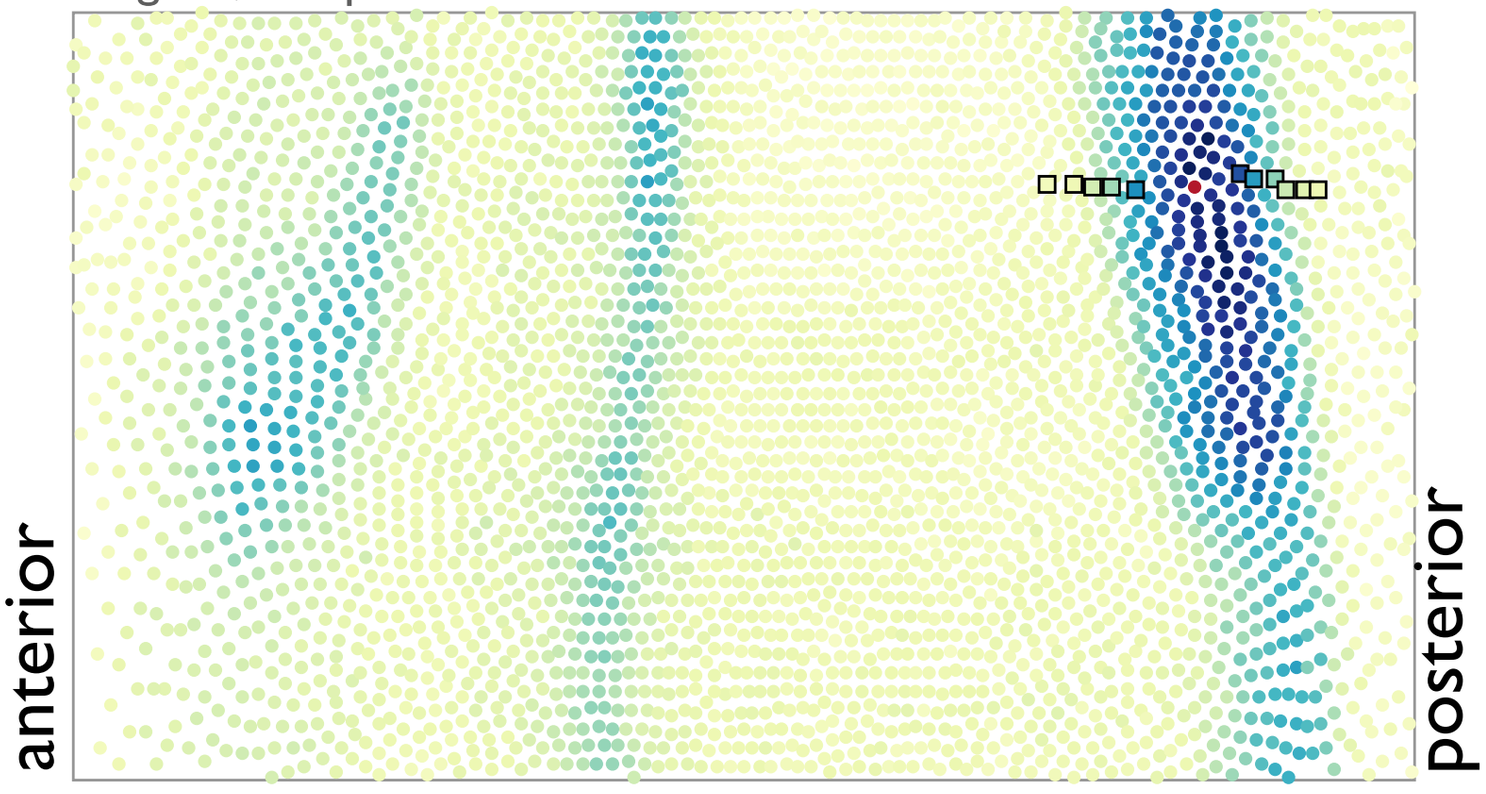
hb gene, timepoint 4



case study two:

exploring groups

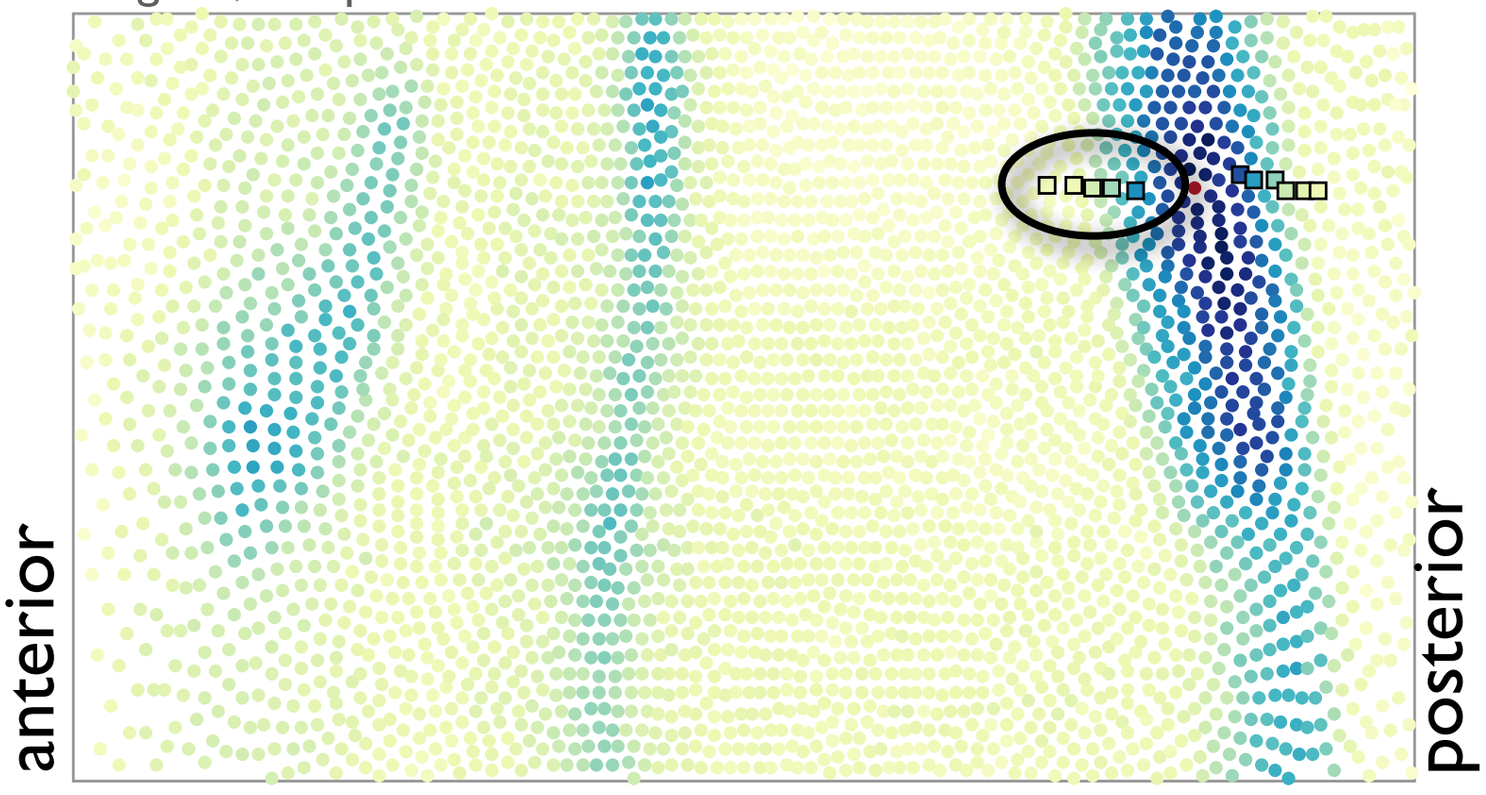
hb gene, timepoint 4



case study two:

exploring groups

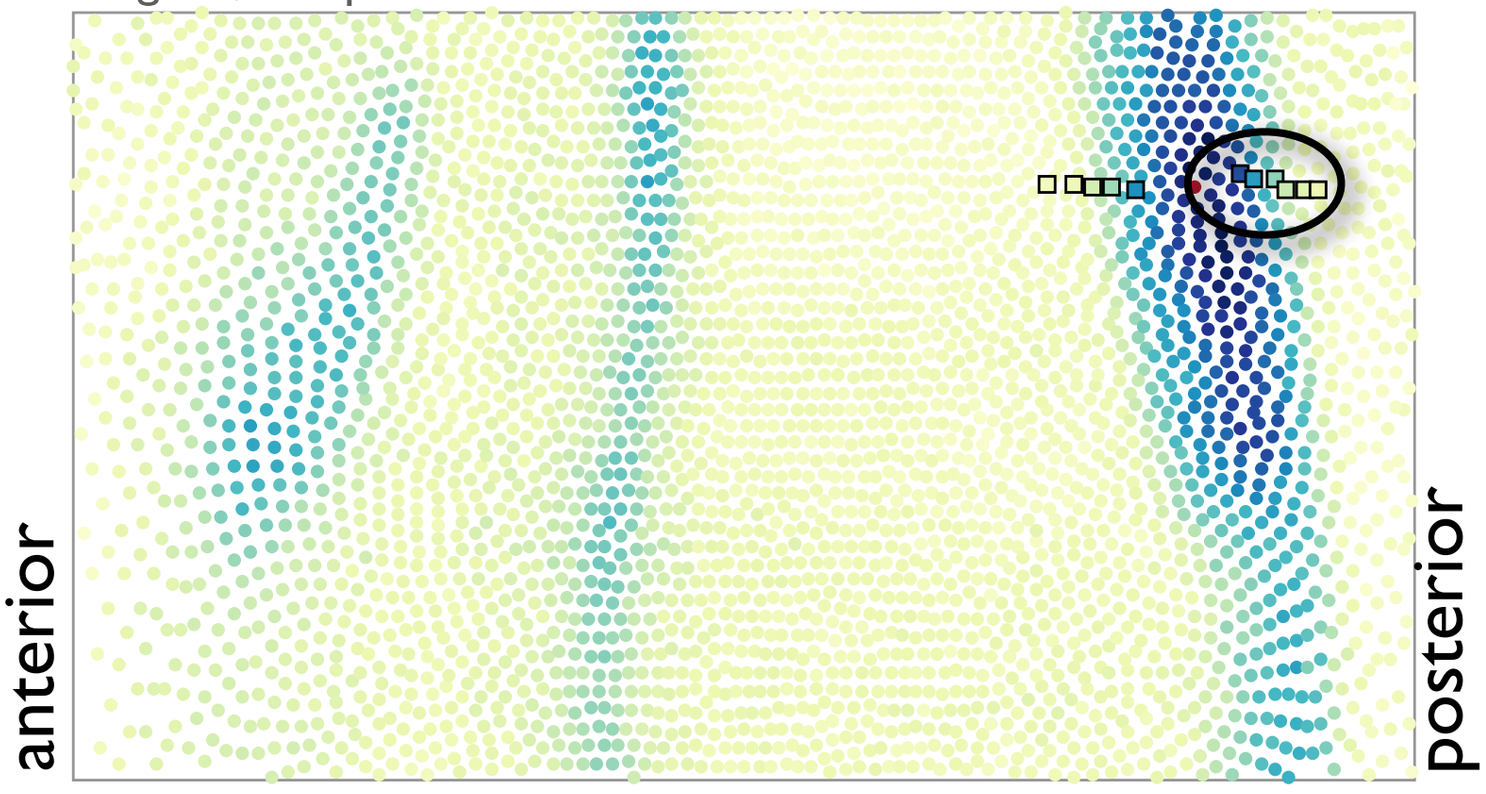
hb gene, timepoint 4



case study two:

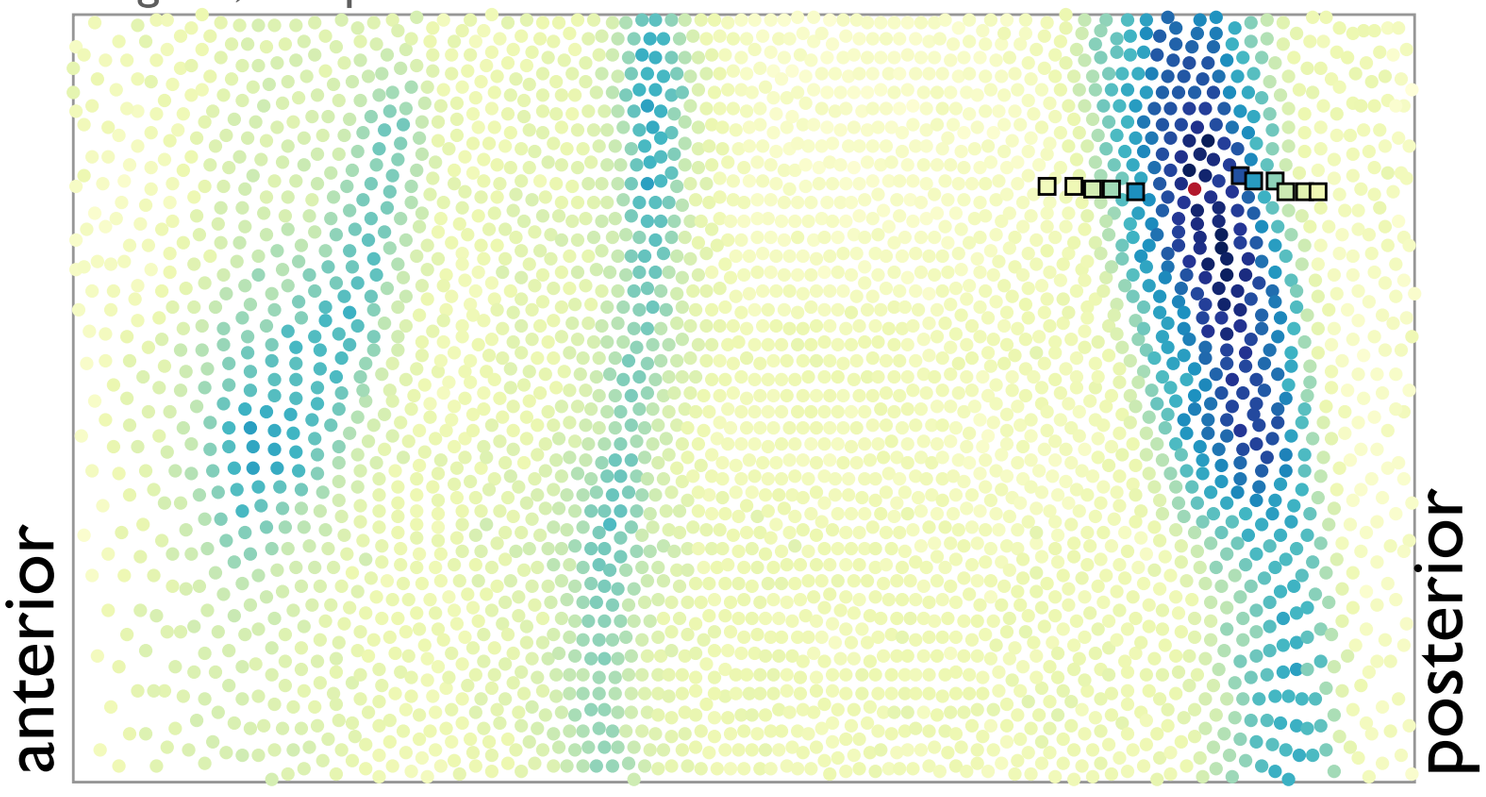
exploring groups

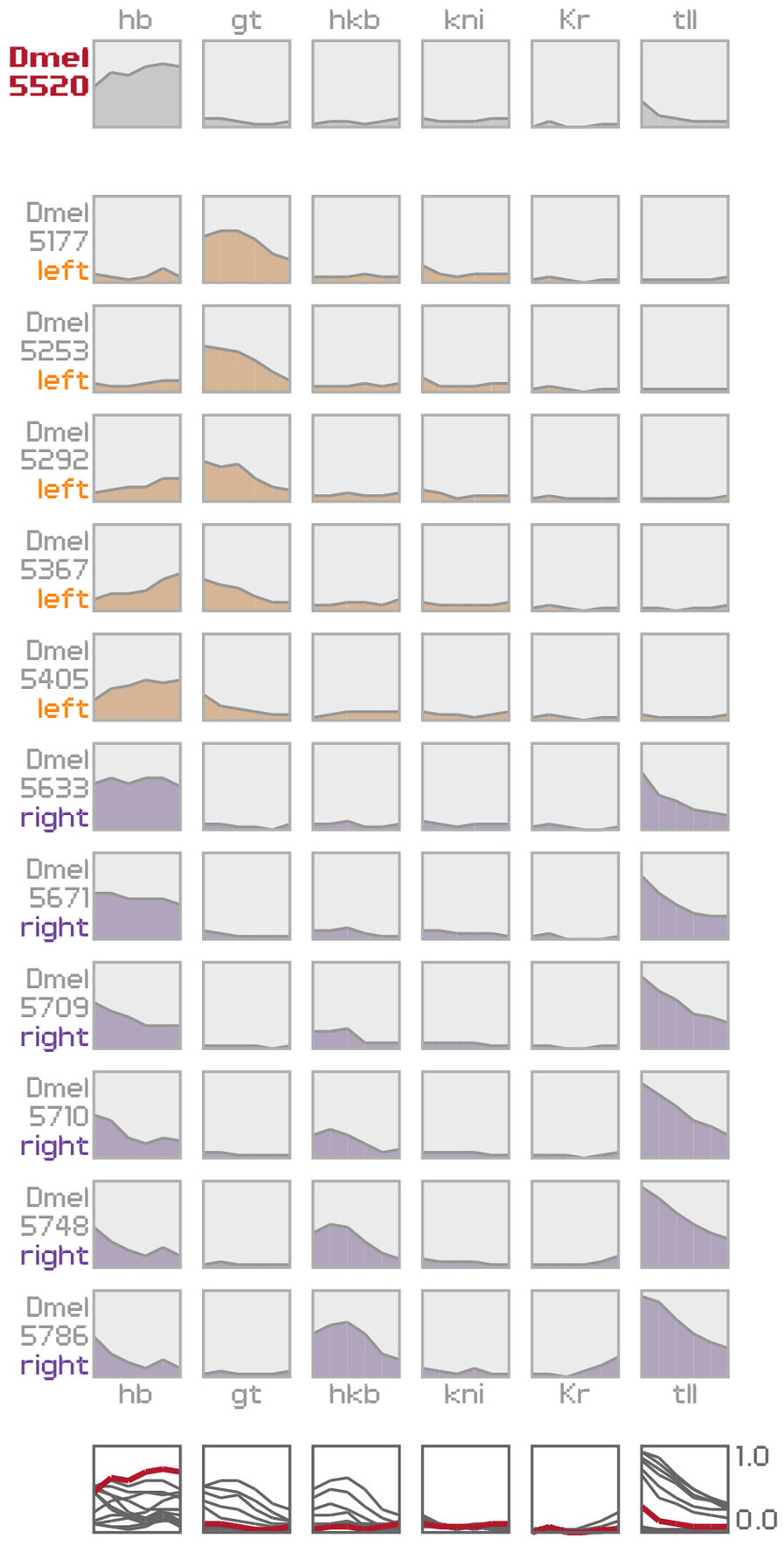
hb gene, timepoint 4



case study two:
exploring groups

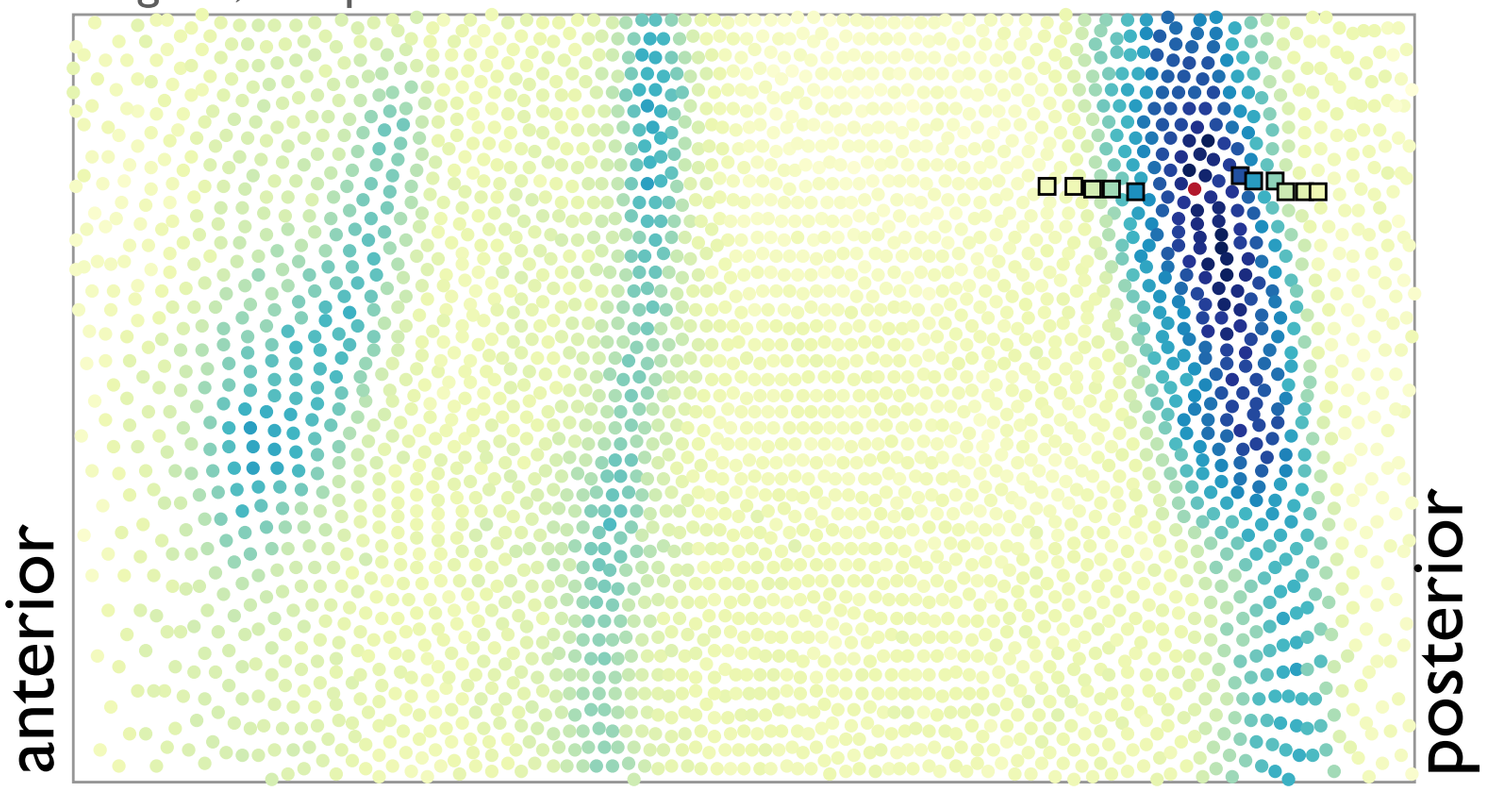
hb gene, timepoint 4

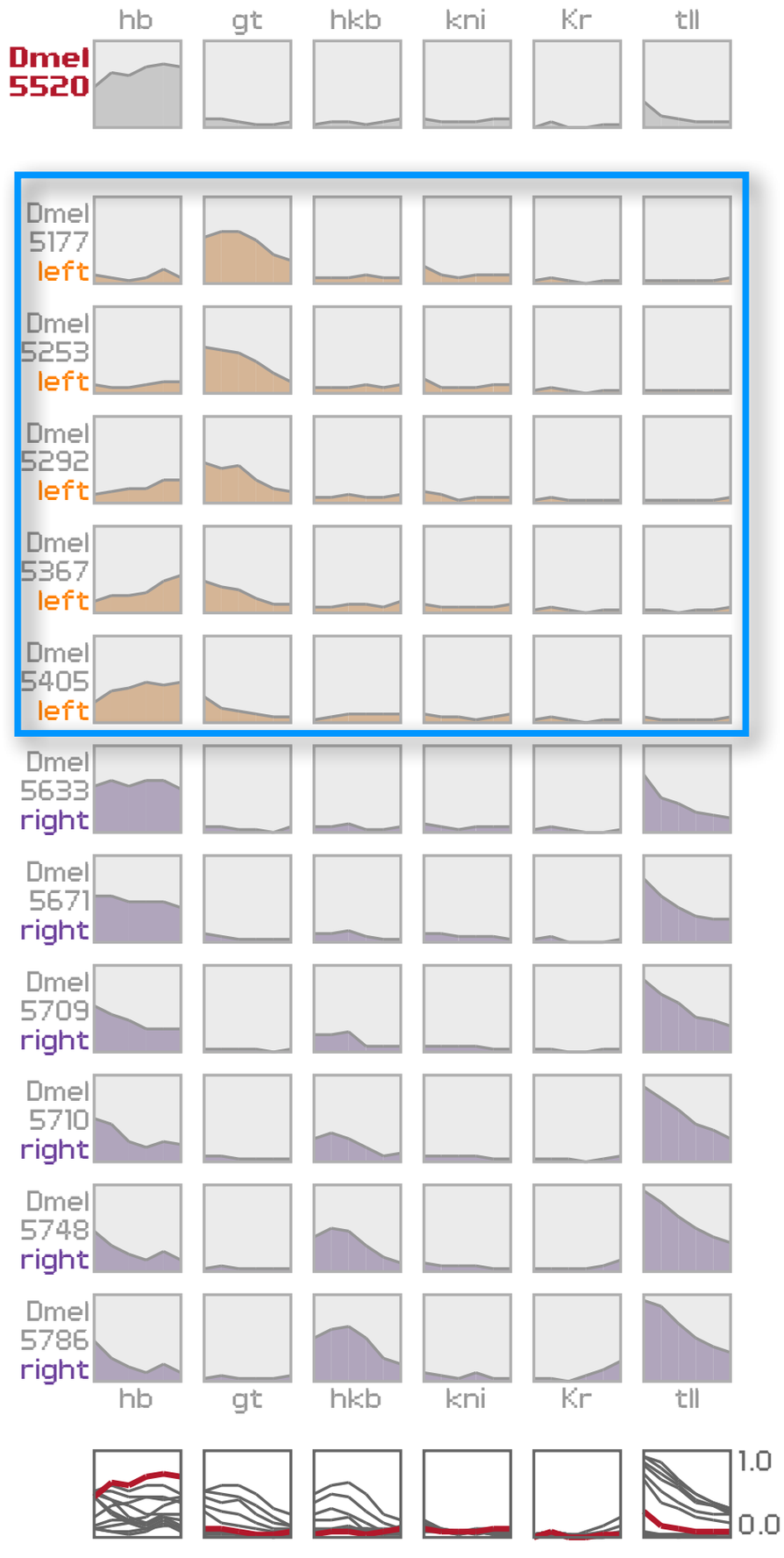




case study two:
exploring groups

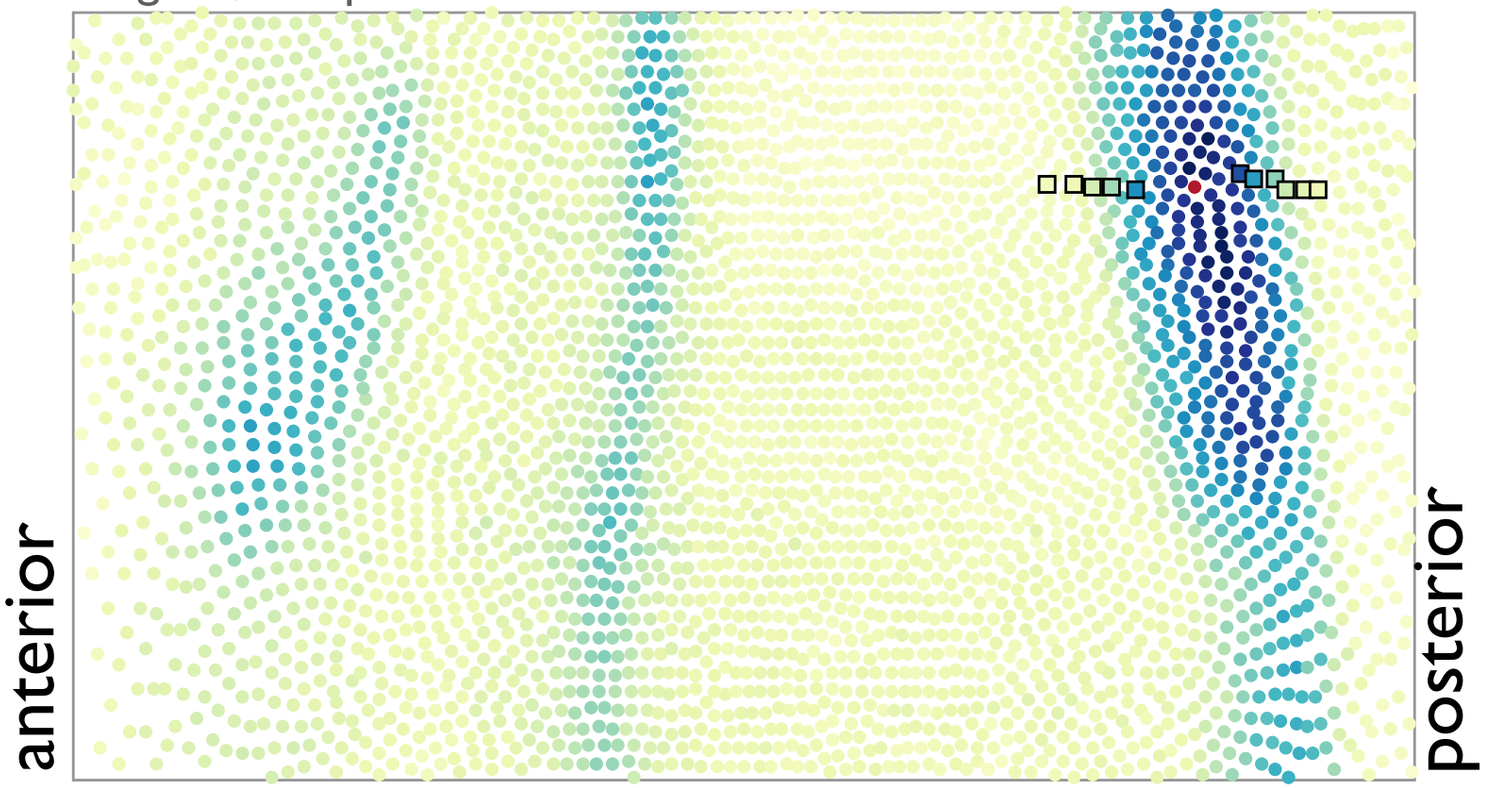
hb gene, timepoint 4

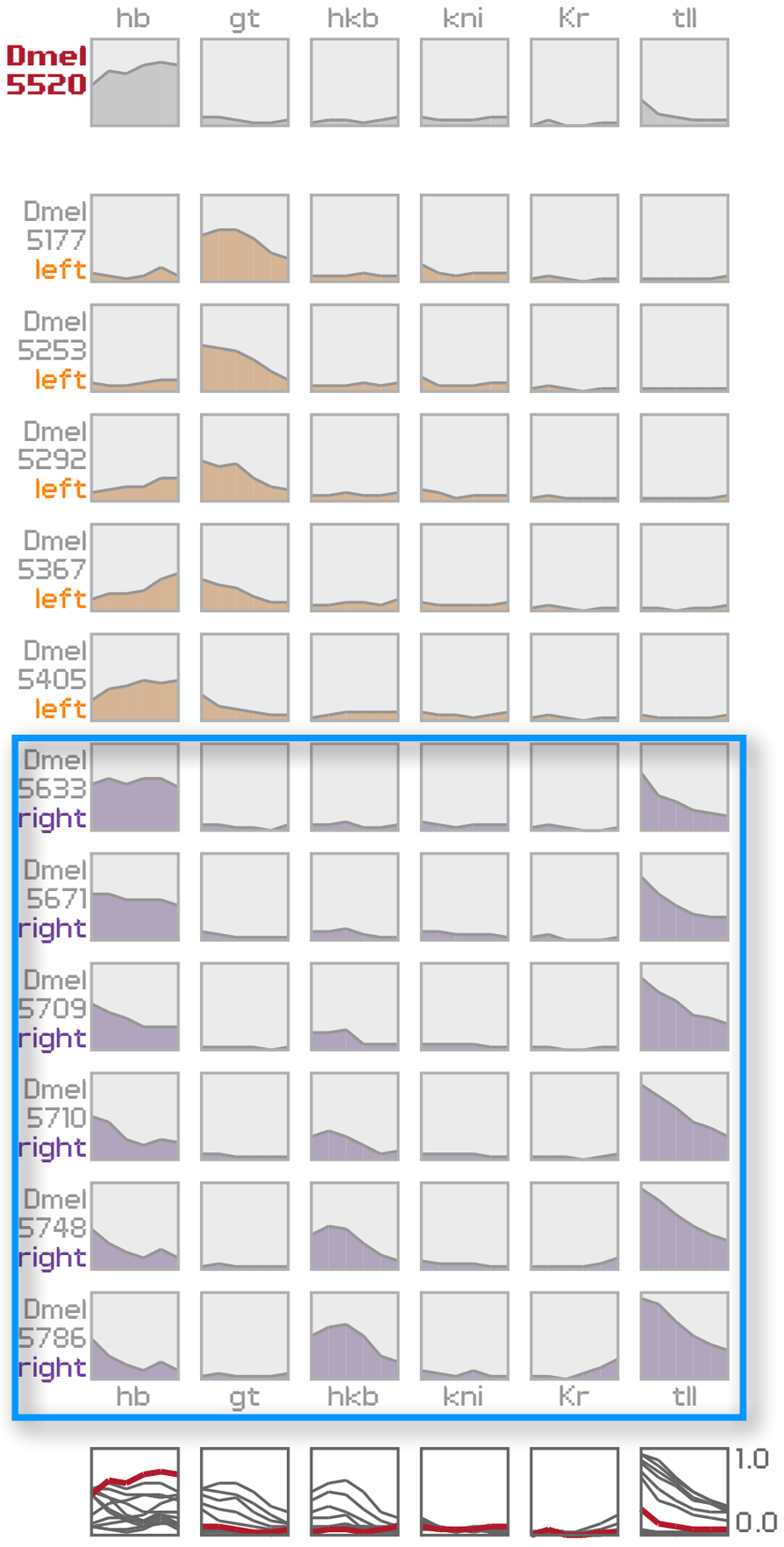




case study two:
exploring groups

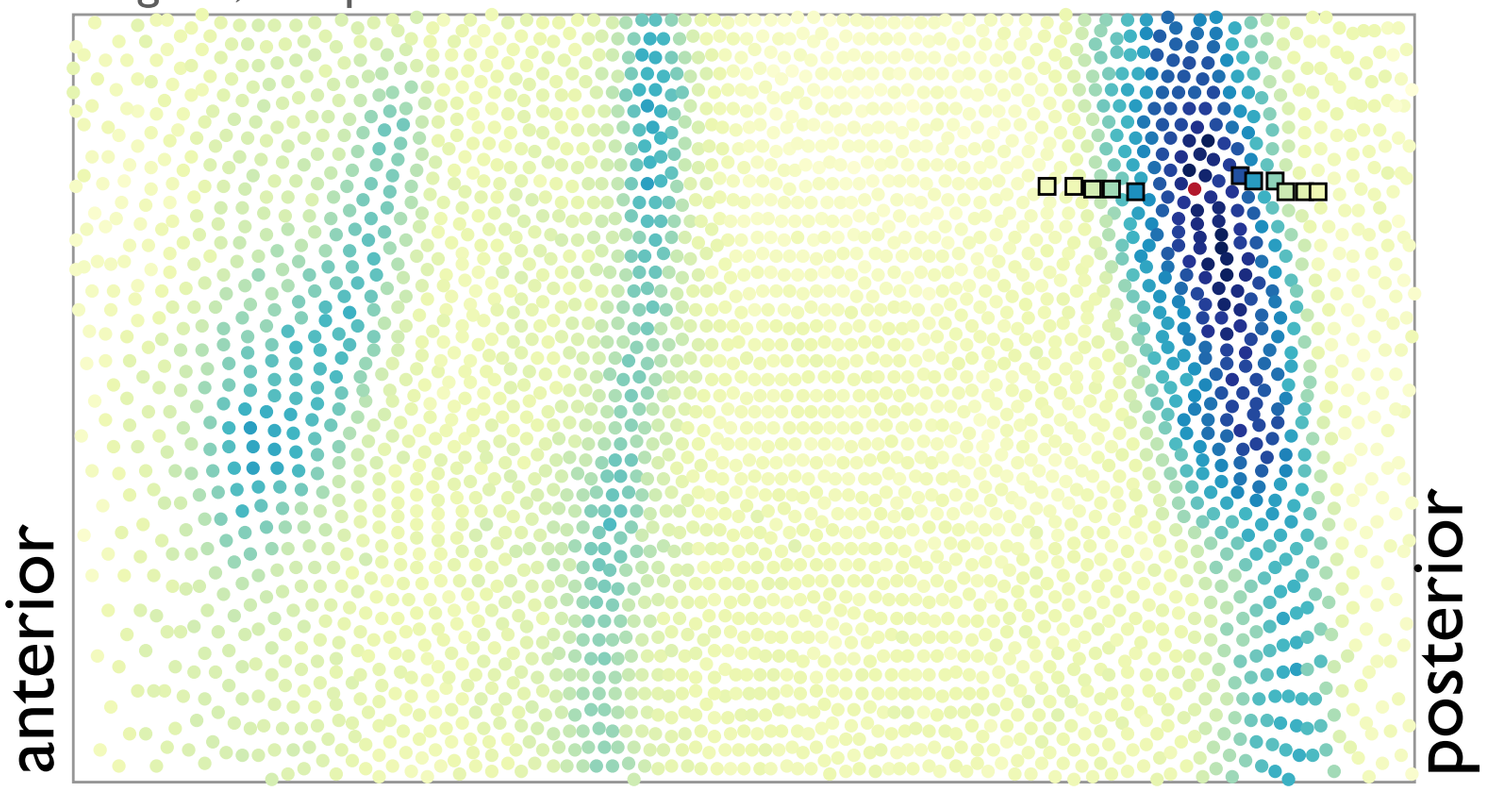
hb gene, timepoint 4

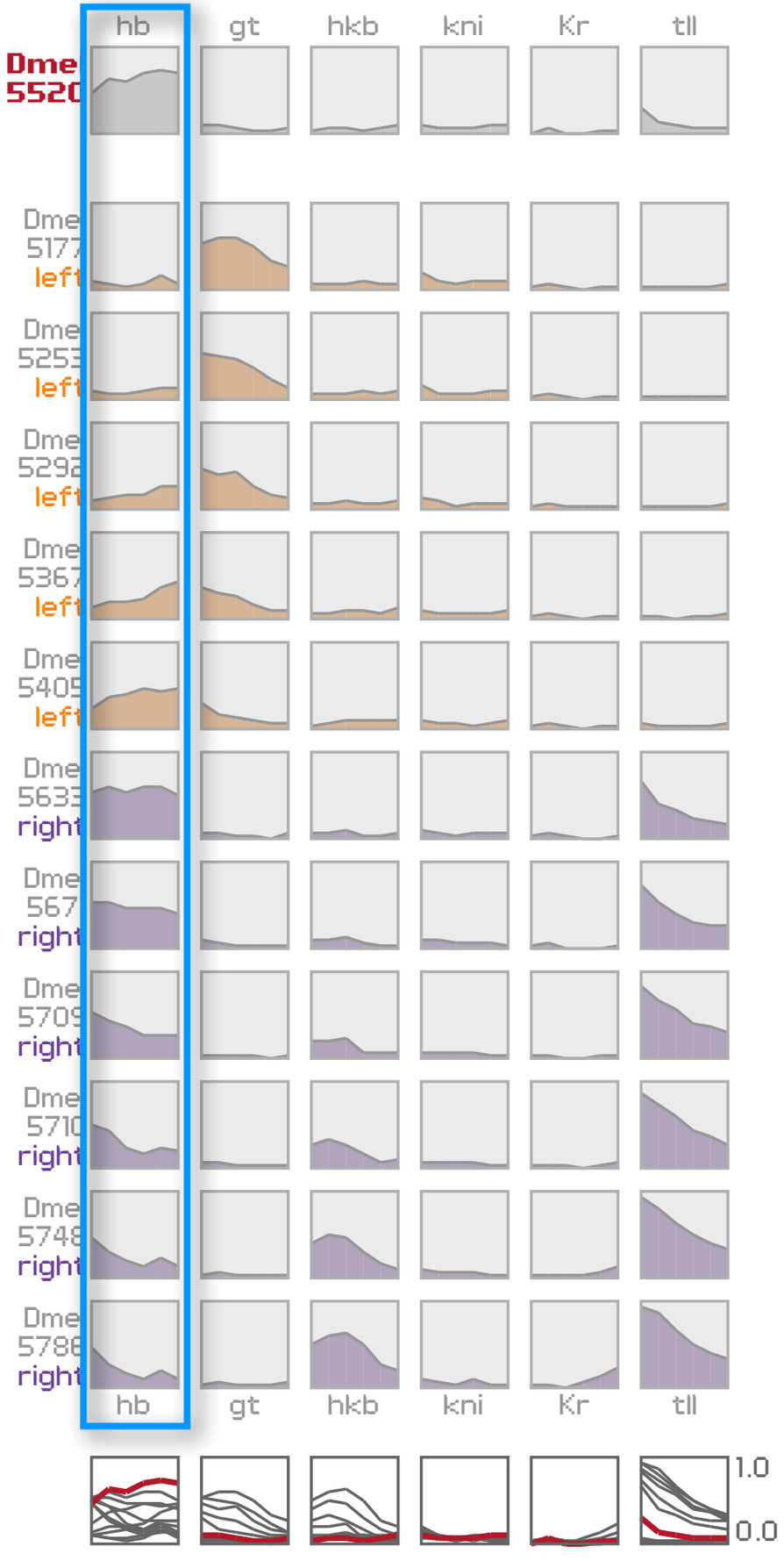




case study two:
exploring groups

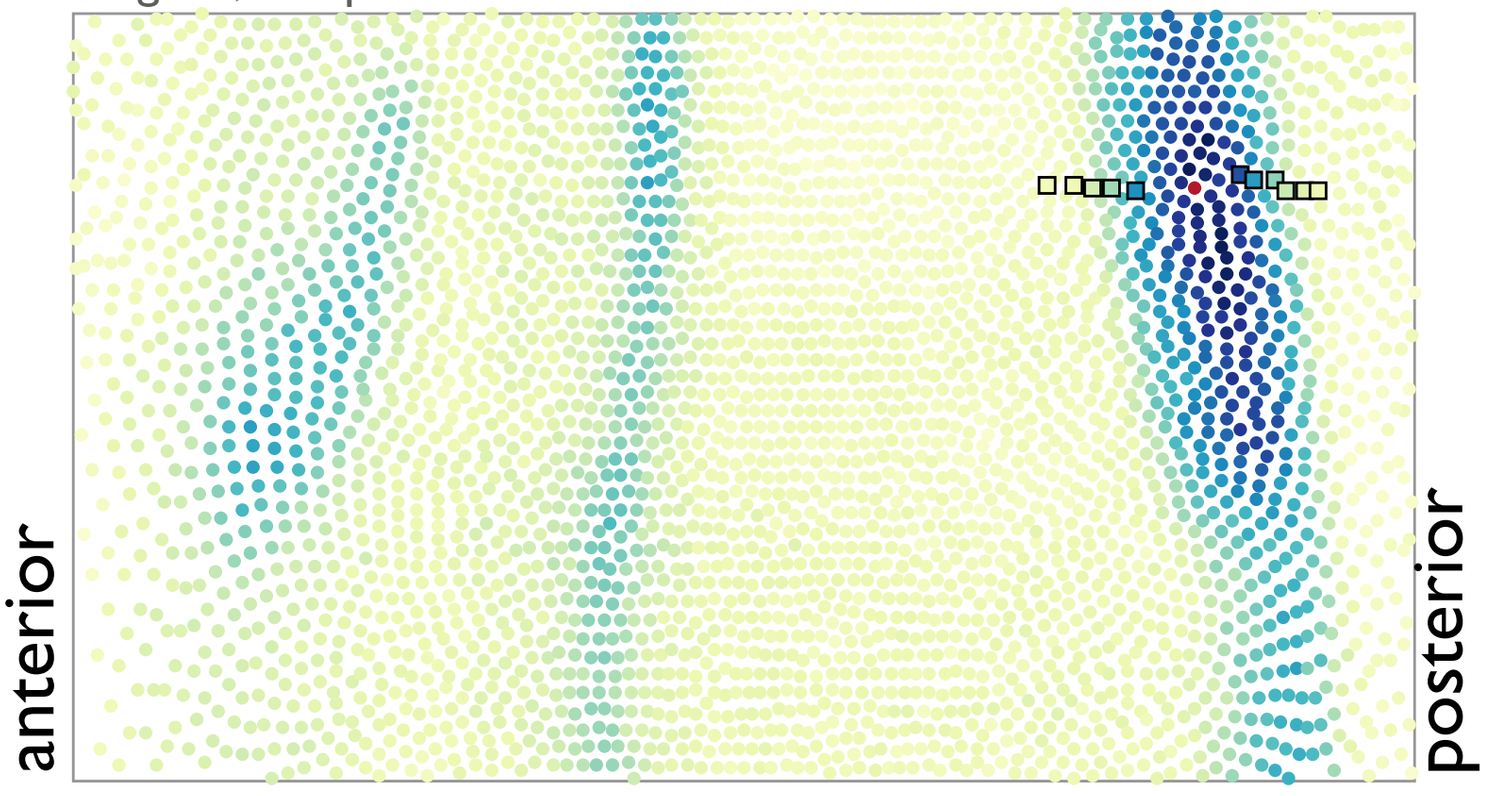
hb gene, timepoint 4

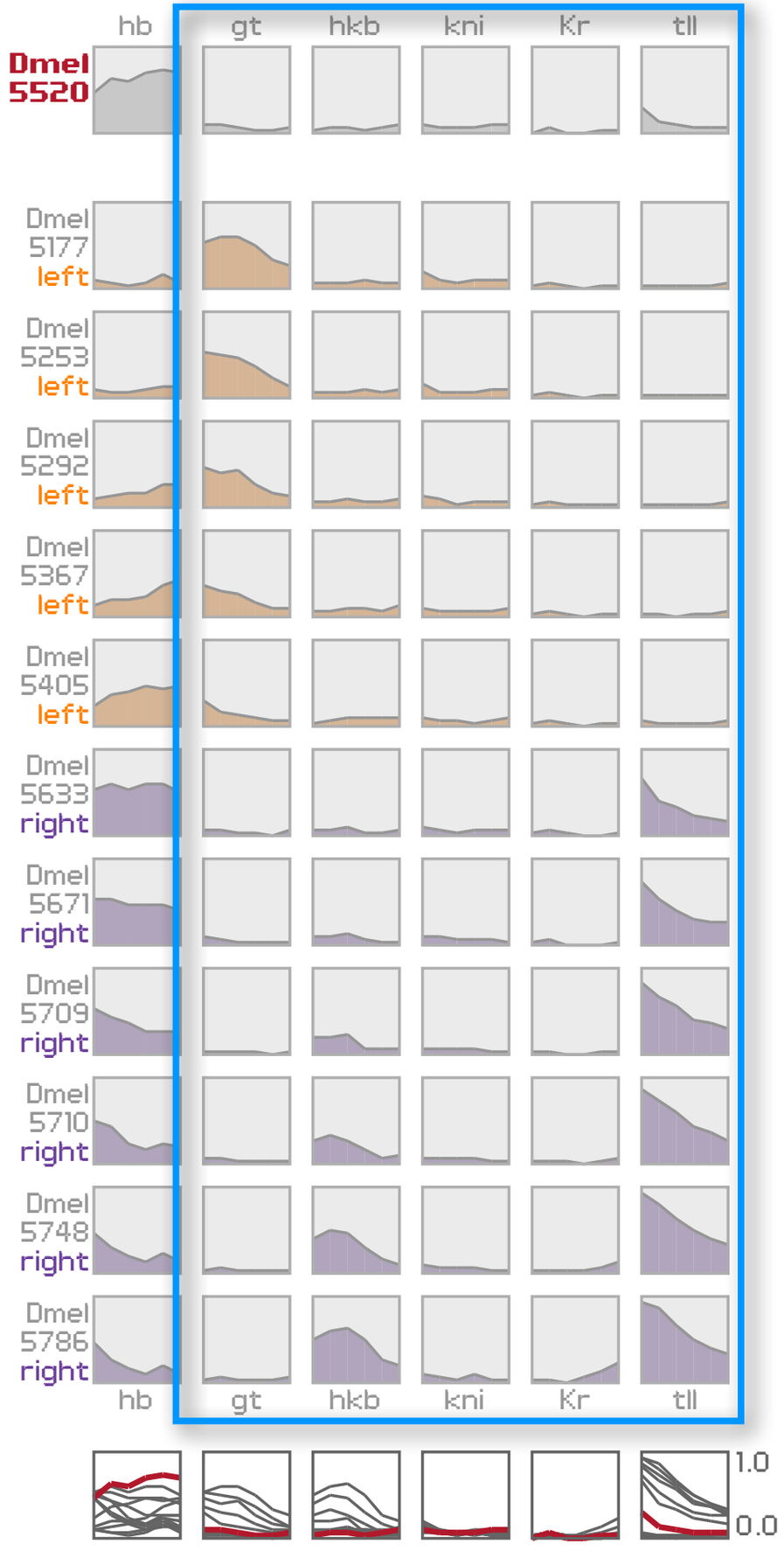




case study two:
exploring groups

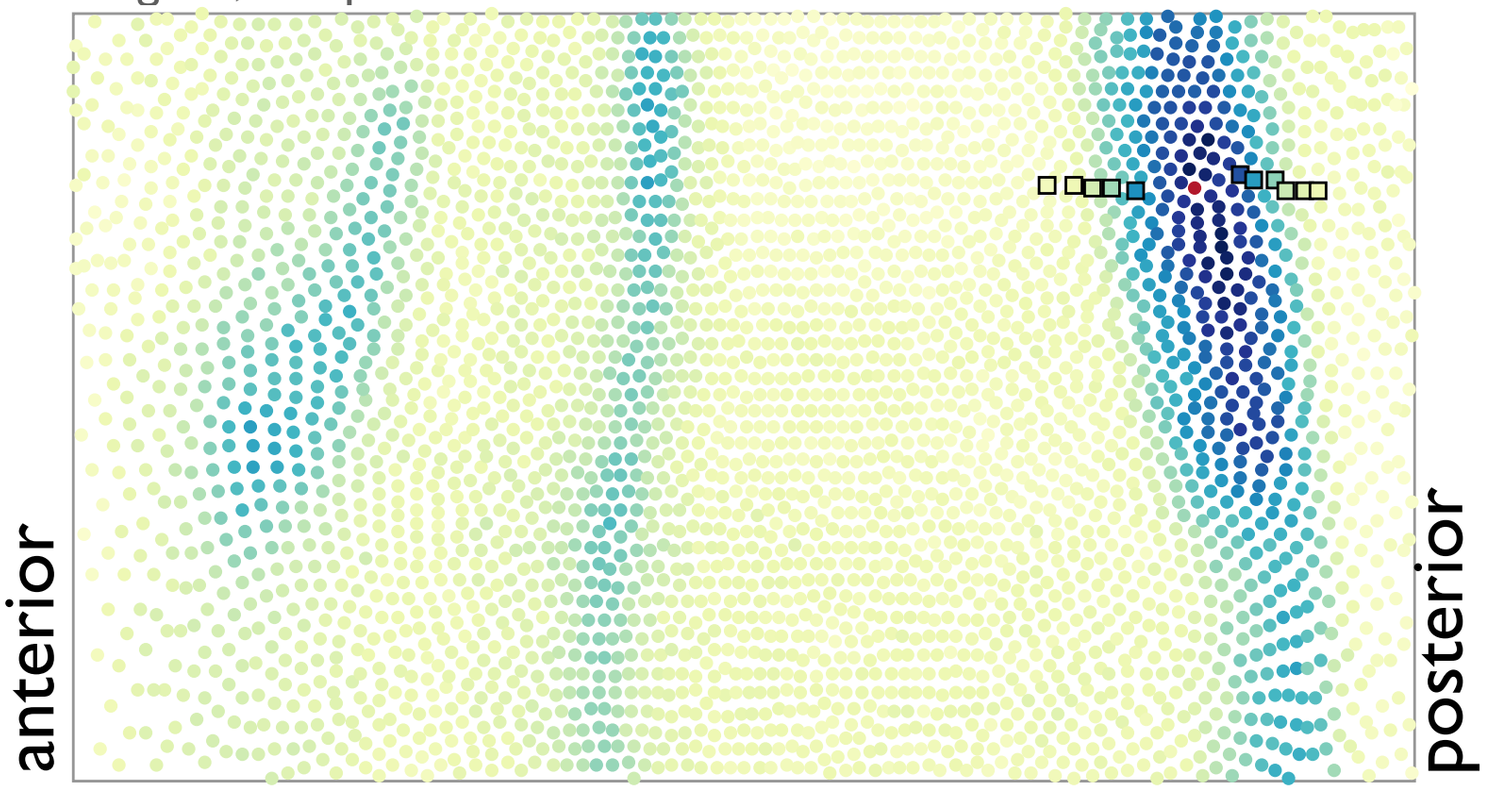
hb gene, timepoint 4

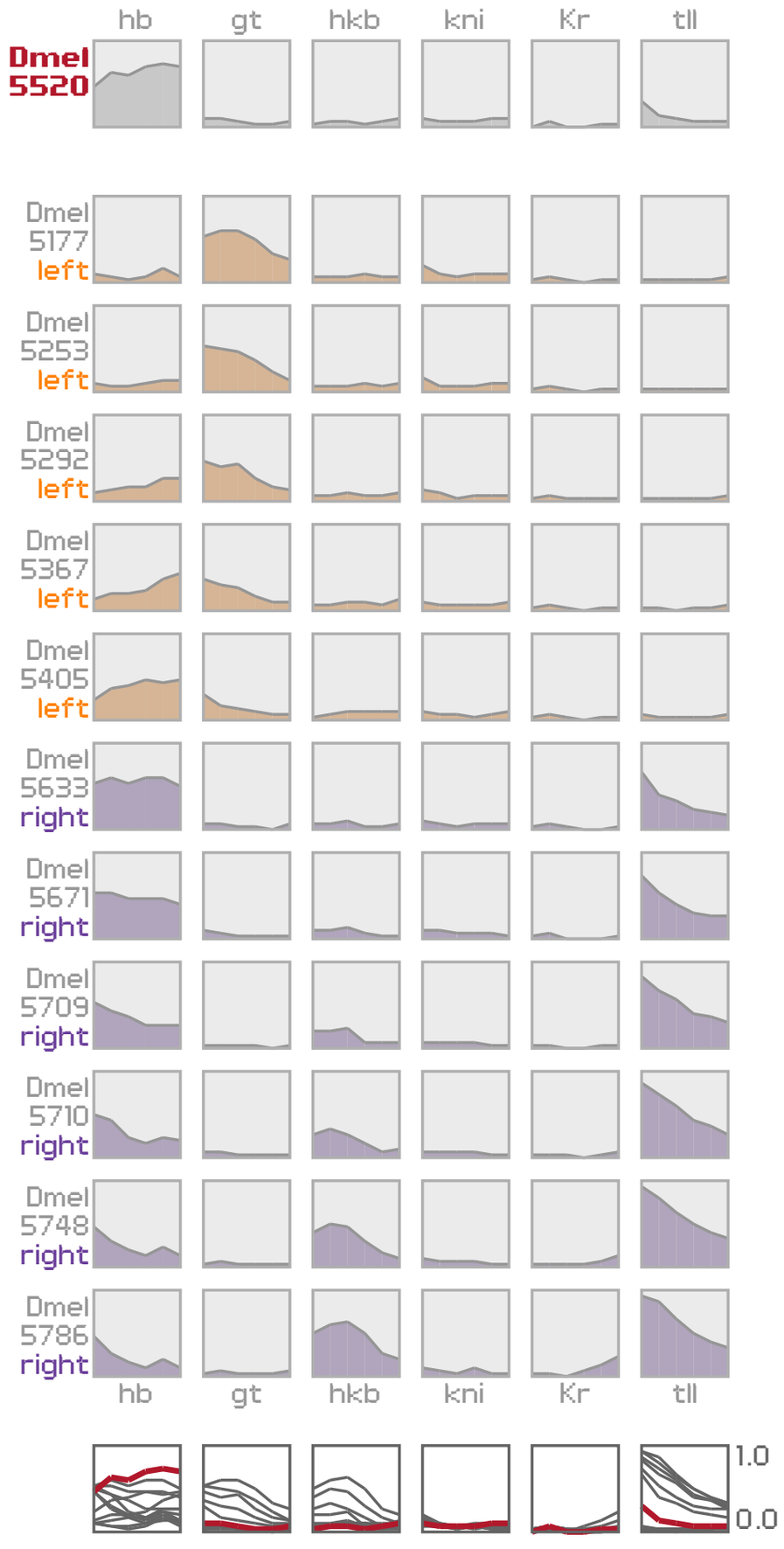




case study two:
exploring groups

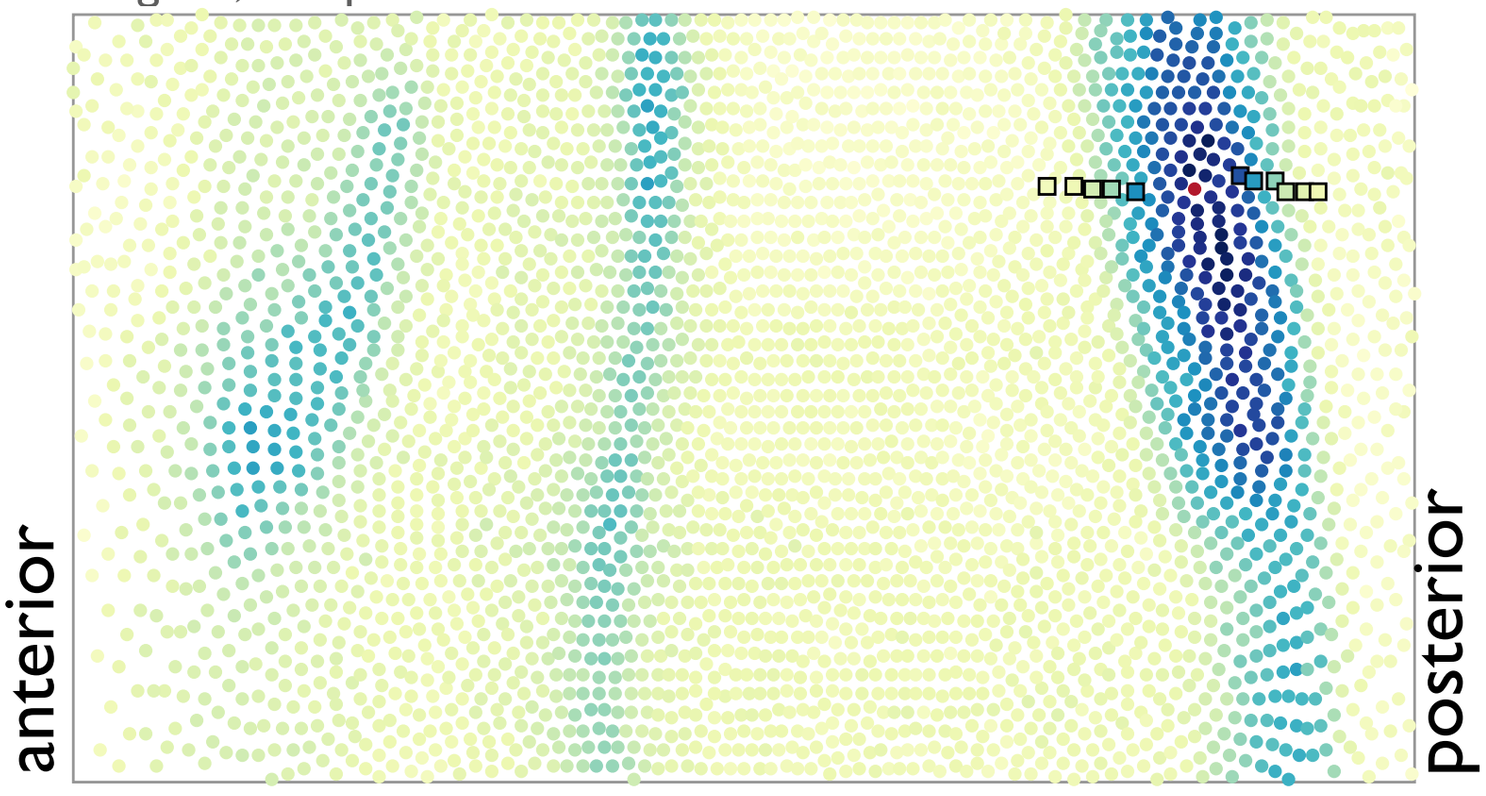
hb gene, timepoint 4

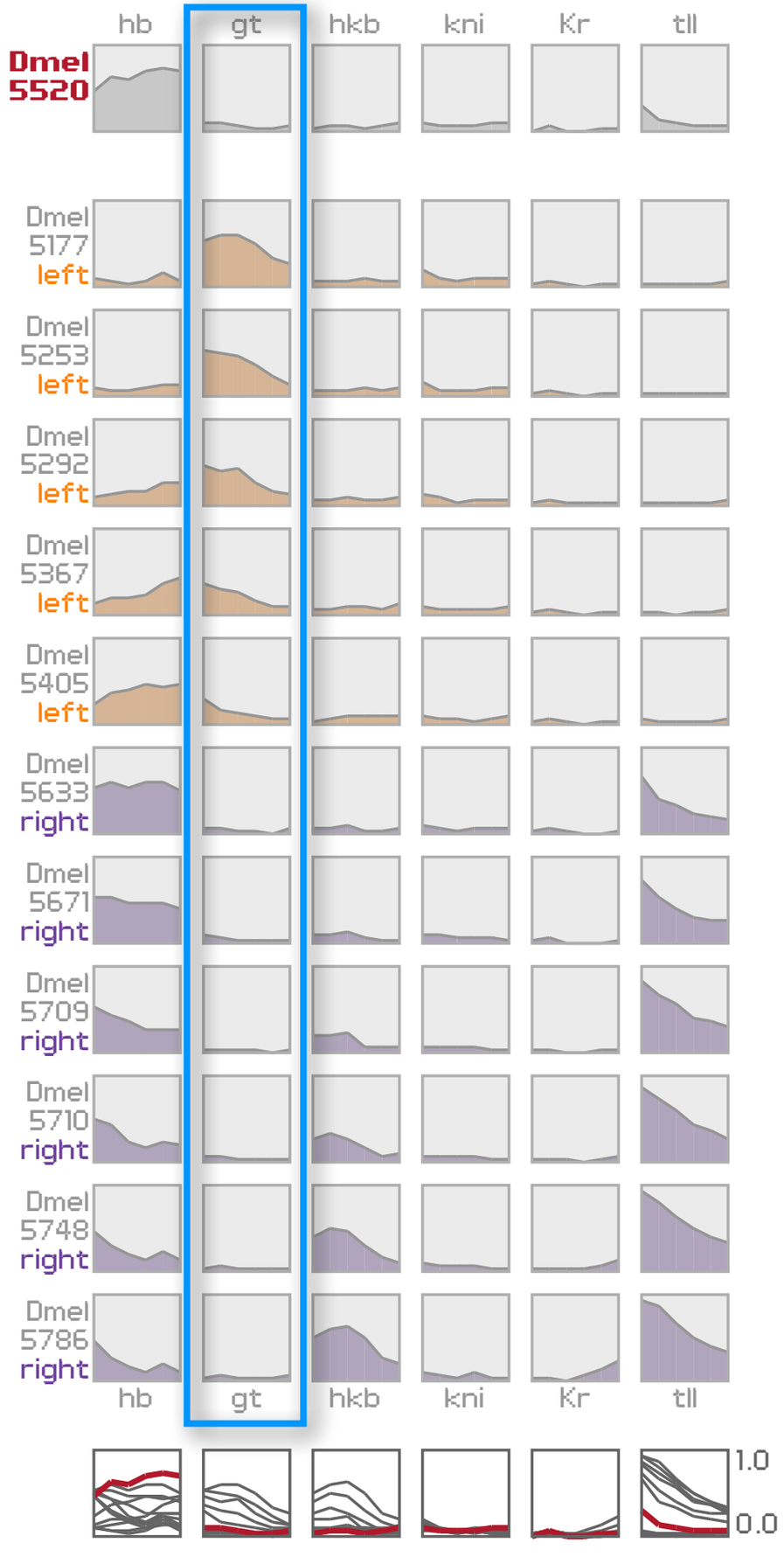




case study two:
exploring groups

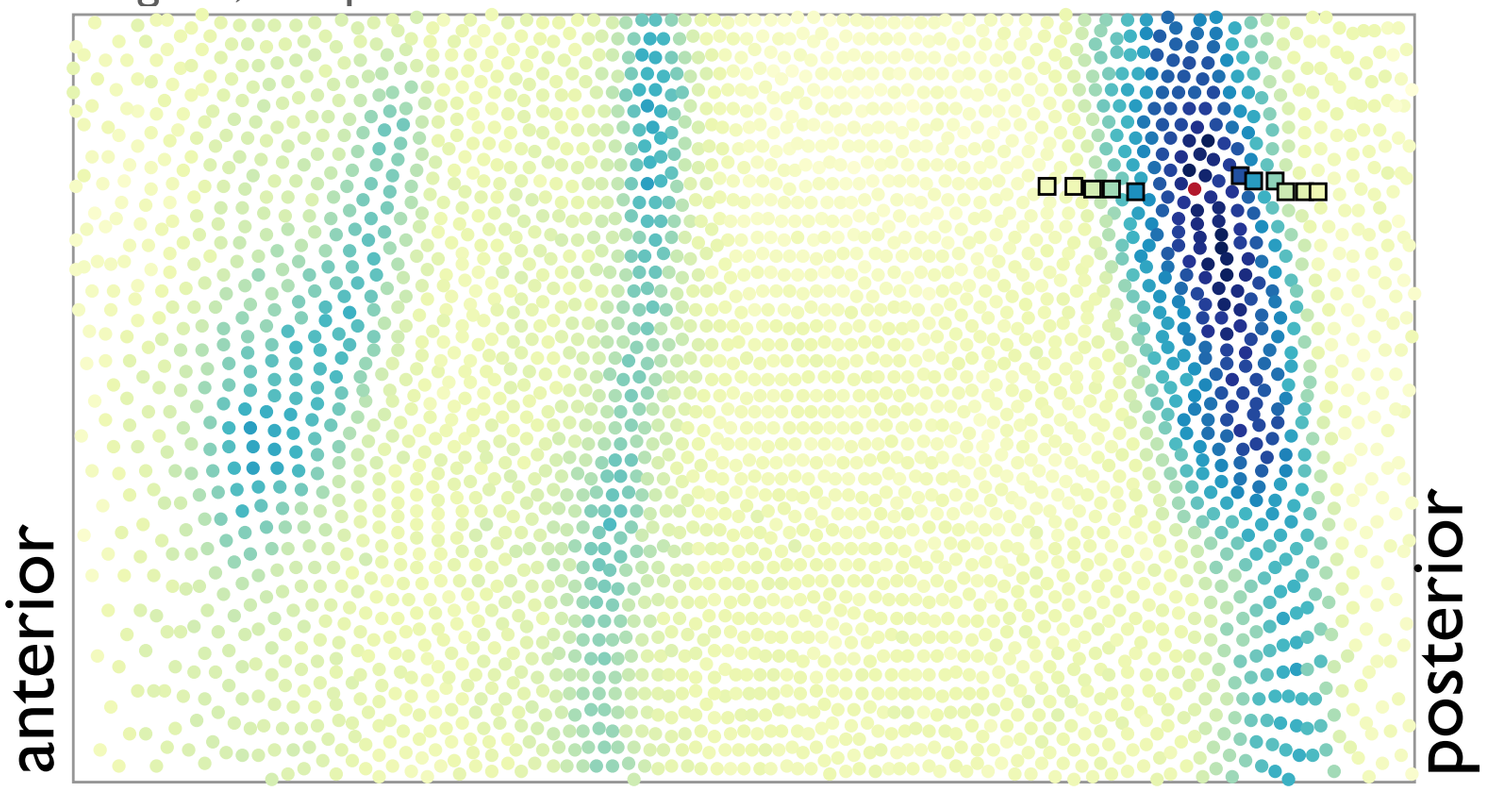
hb gene, timepoint 4

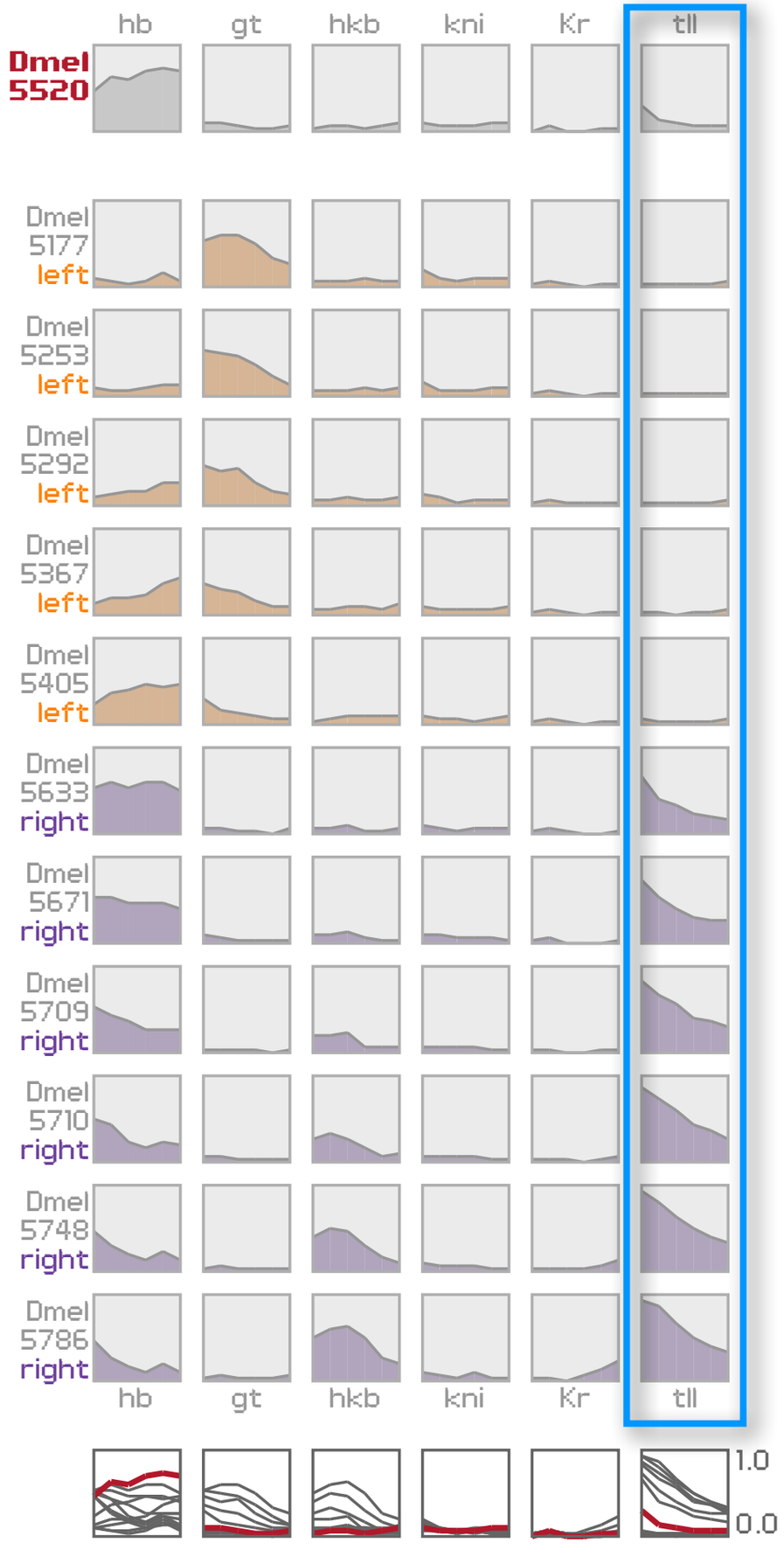




case study two:
exploring groups

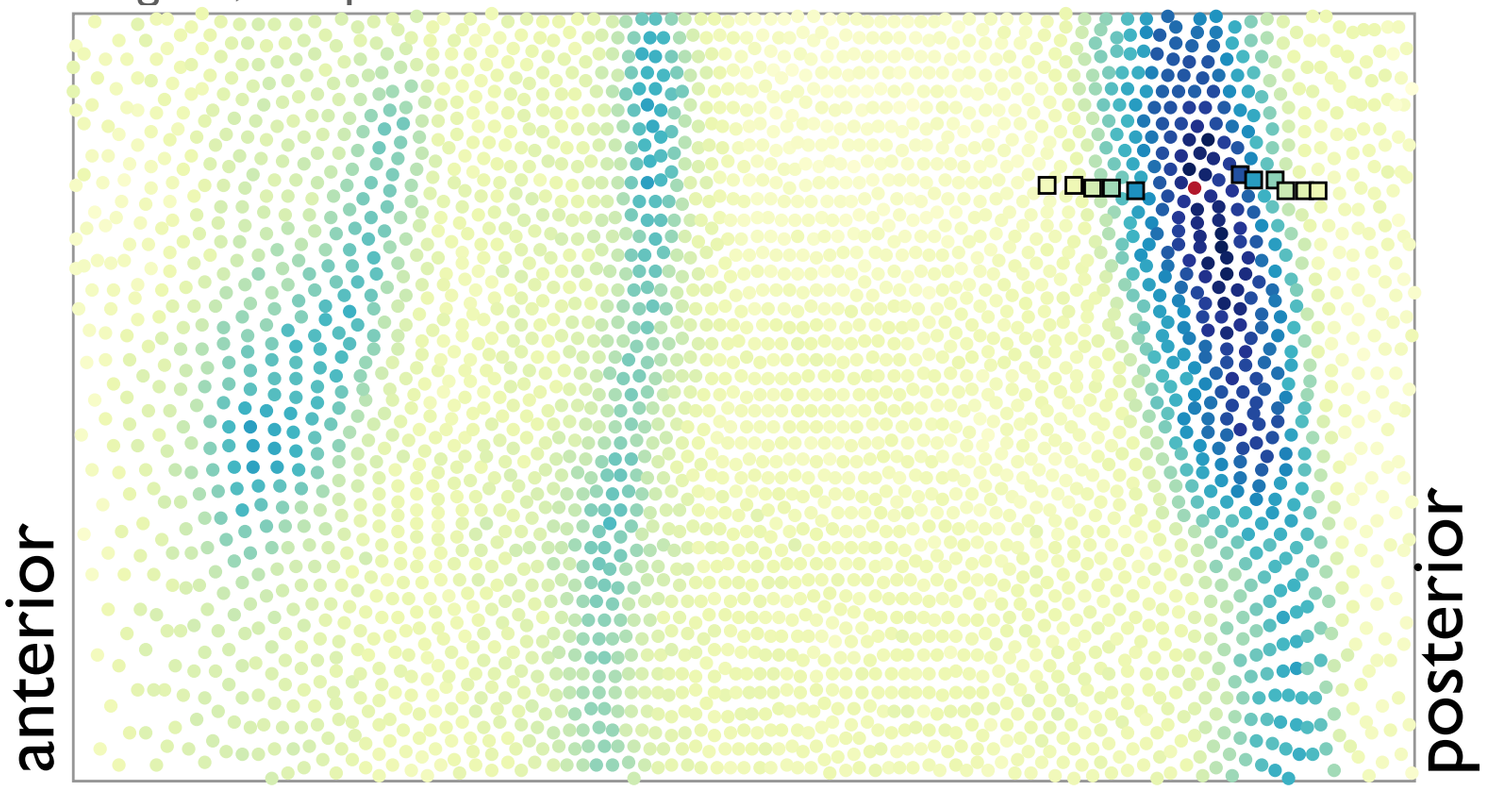
hb gene, timepoint 4

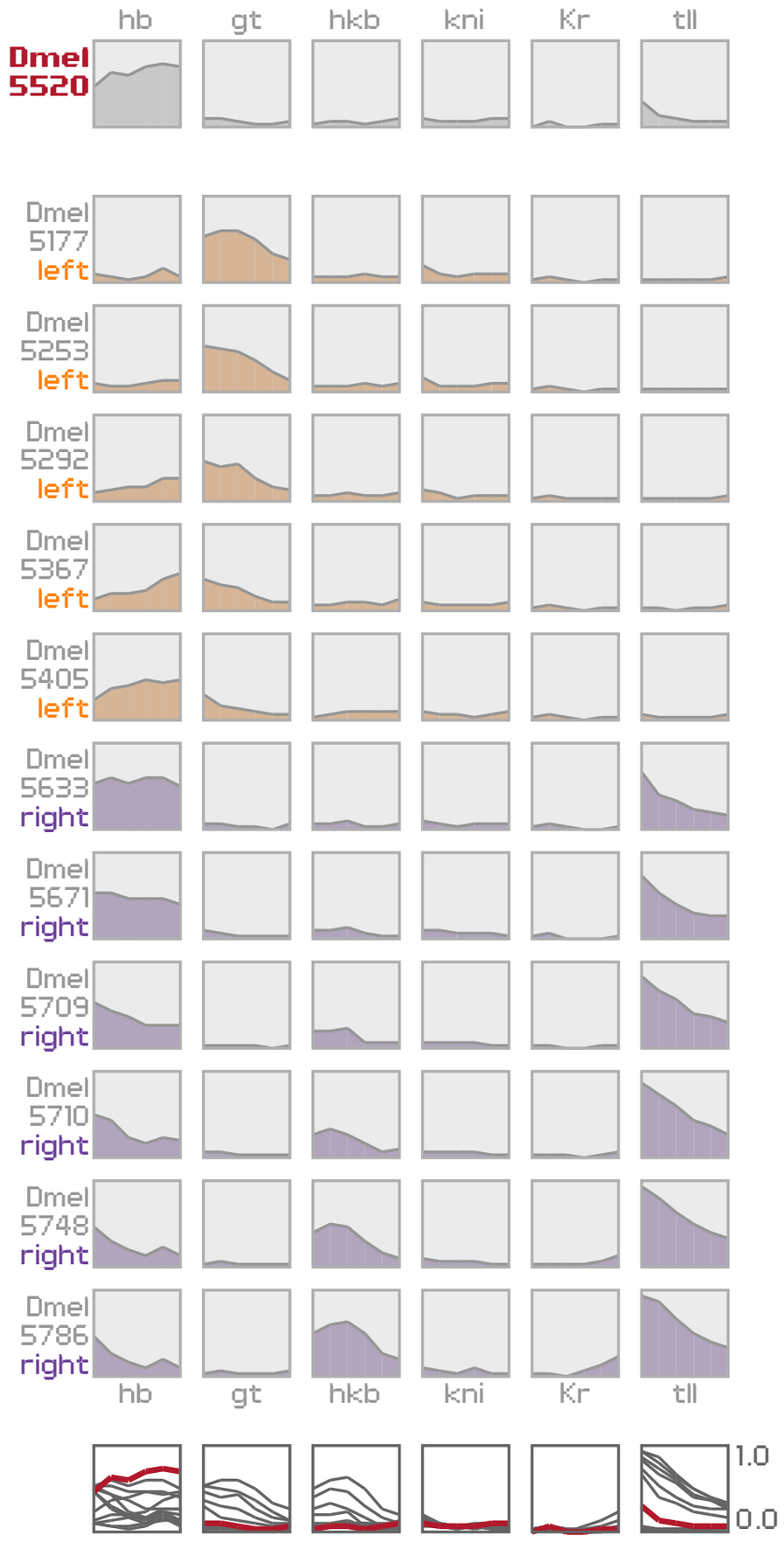




case study two:
exploring groups

hb gene, timepoint 4





data & tool & tasks

summaries & groups

encodings & interaction

conclusions

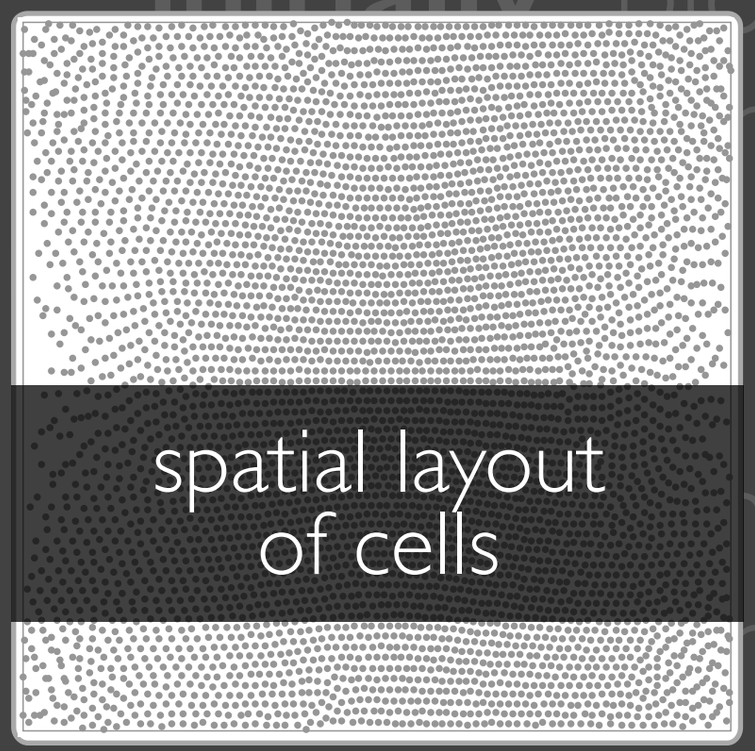
initially: biologists relied on computational methods to explore data

goal: link underlying data with computational results

visualize: triad of data

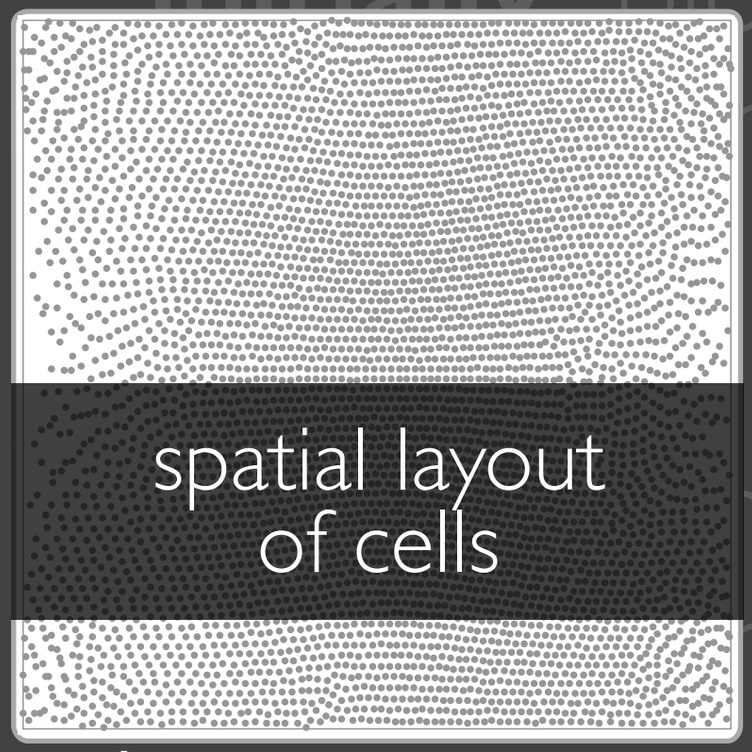
initially: biologists relied on
nal methods to explore data

nderlying data with
nal results

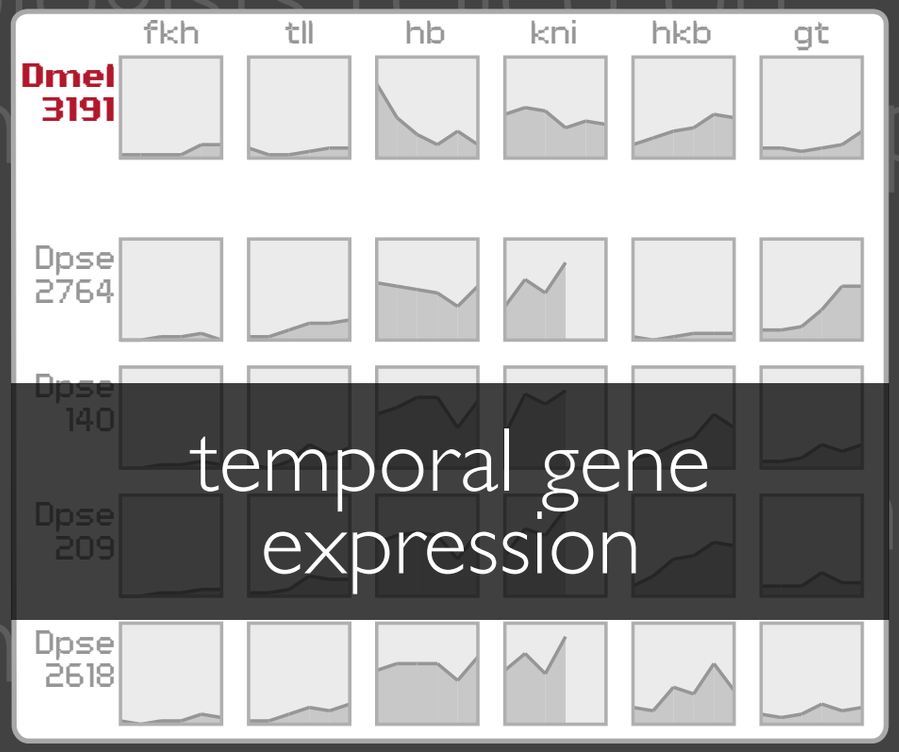


visualize: triad of data

initially: biologists relied on [illegible] plore data



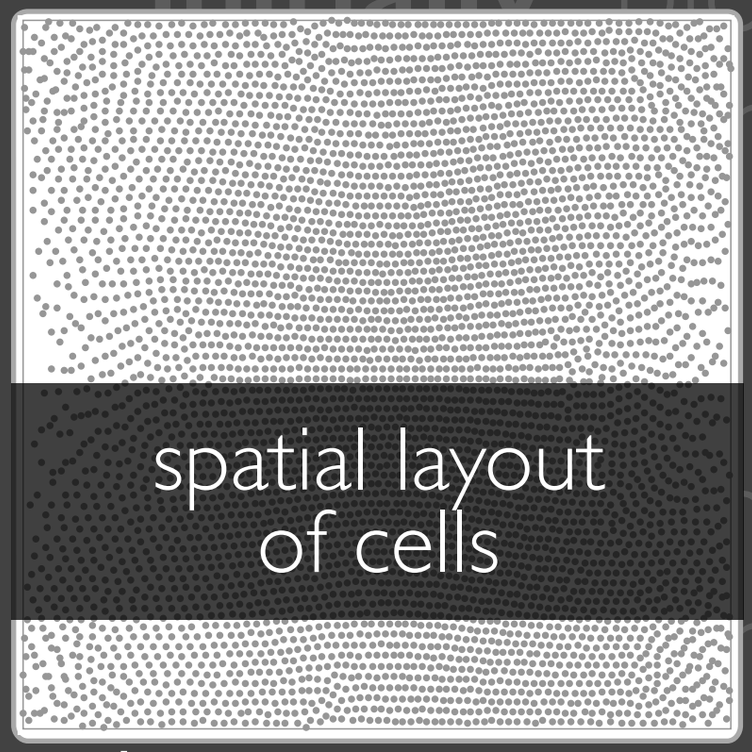
embryo map



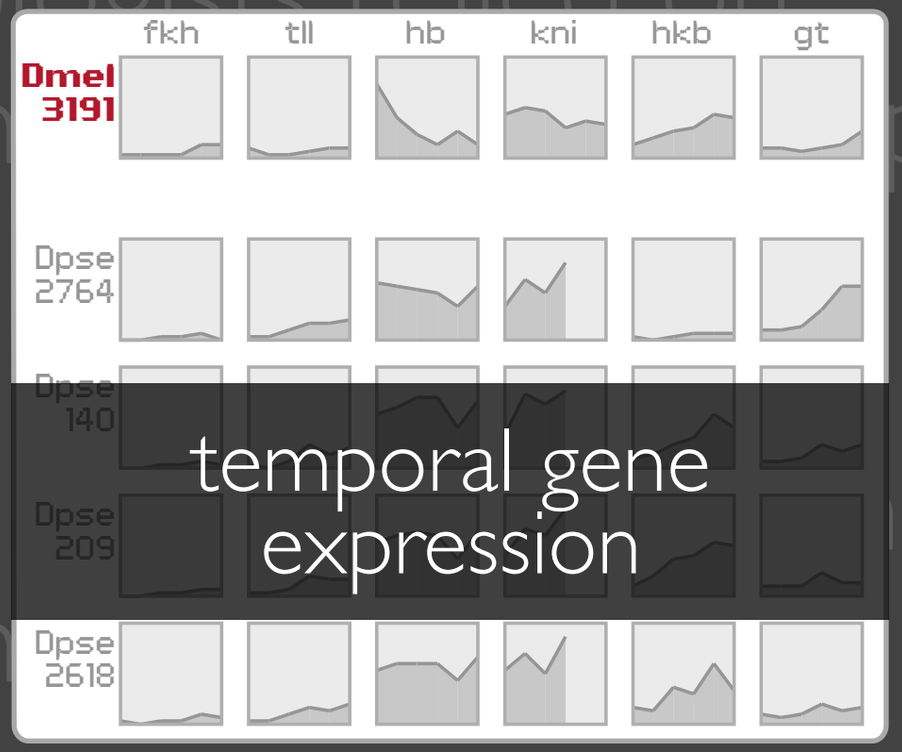
curvemap [Meyer10]

visualize: triad of data

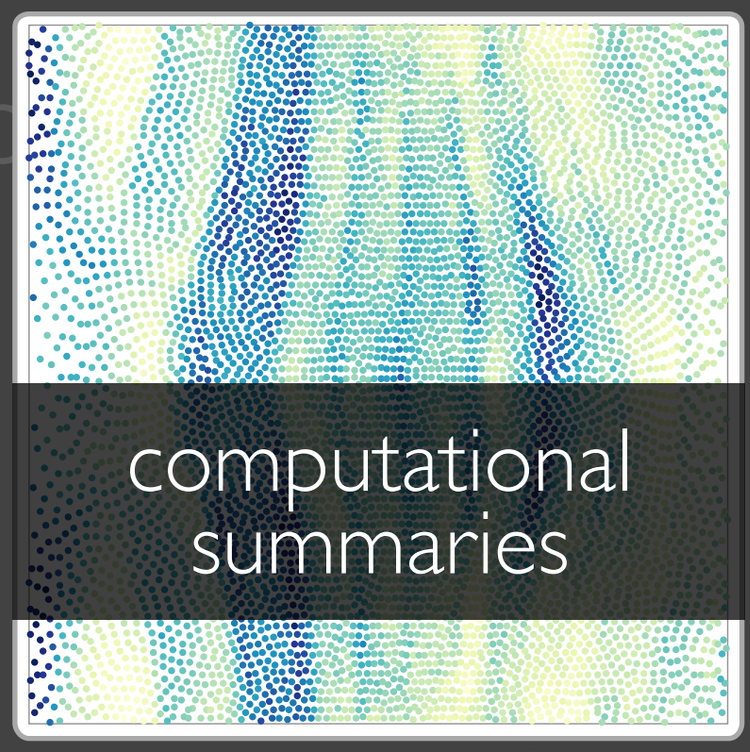
initially: biologists relied on



embryo map



curvemap [Meyer10]

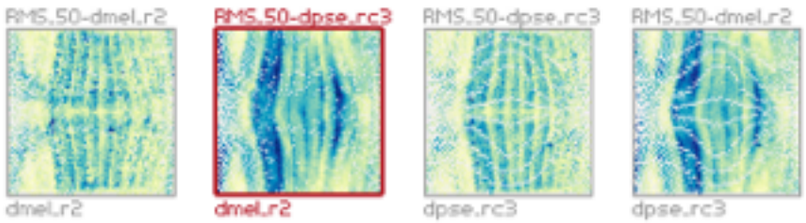


colormap

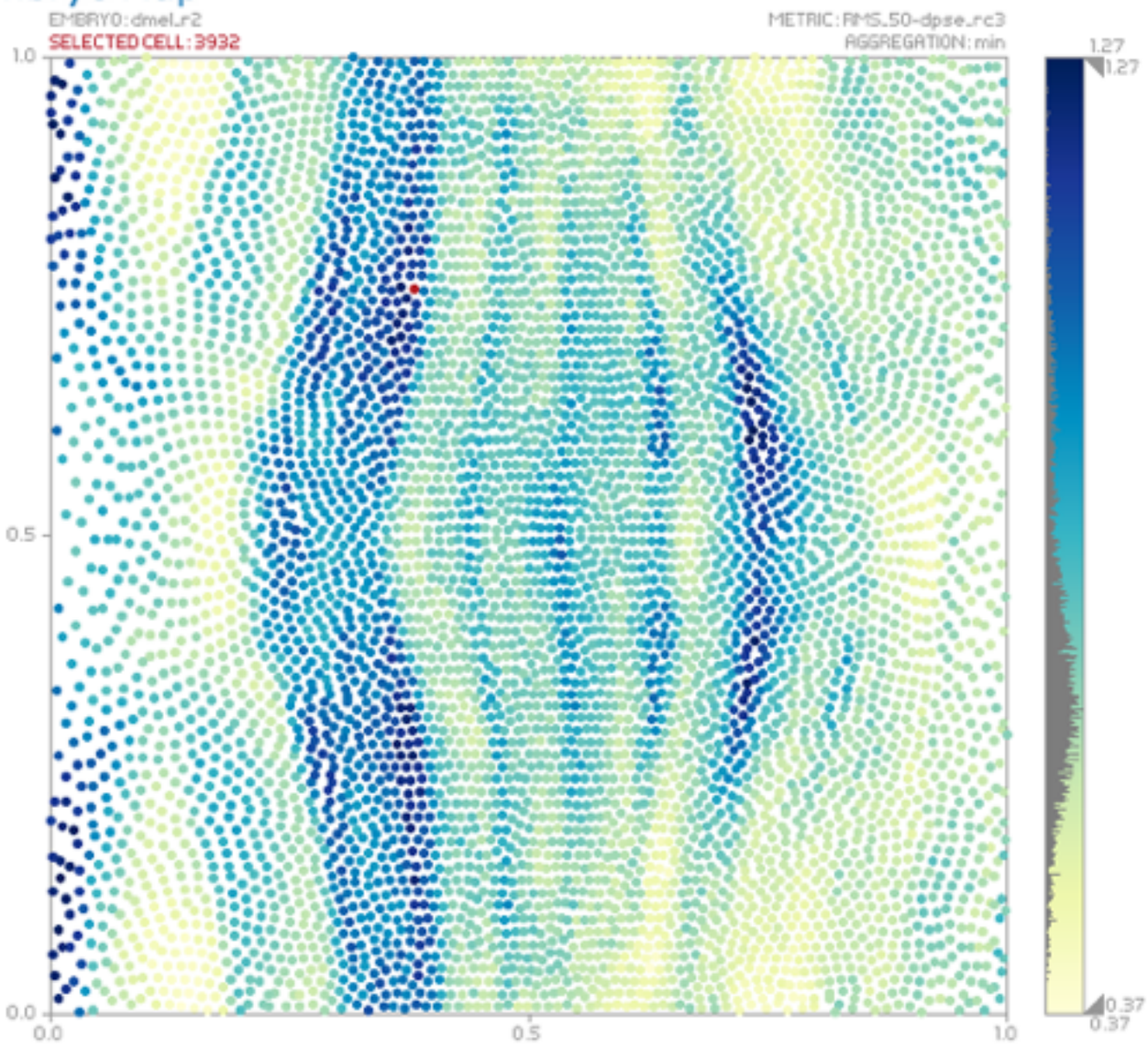
visualize: triad of data

MulteeSum

Summaries



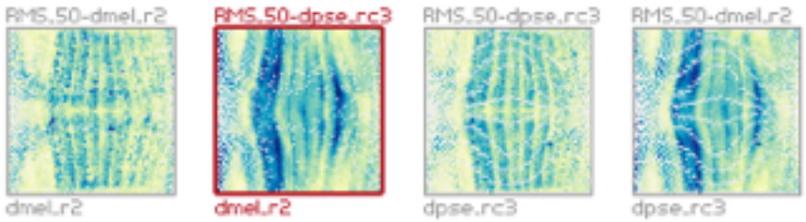
Embryo Map



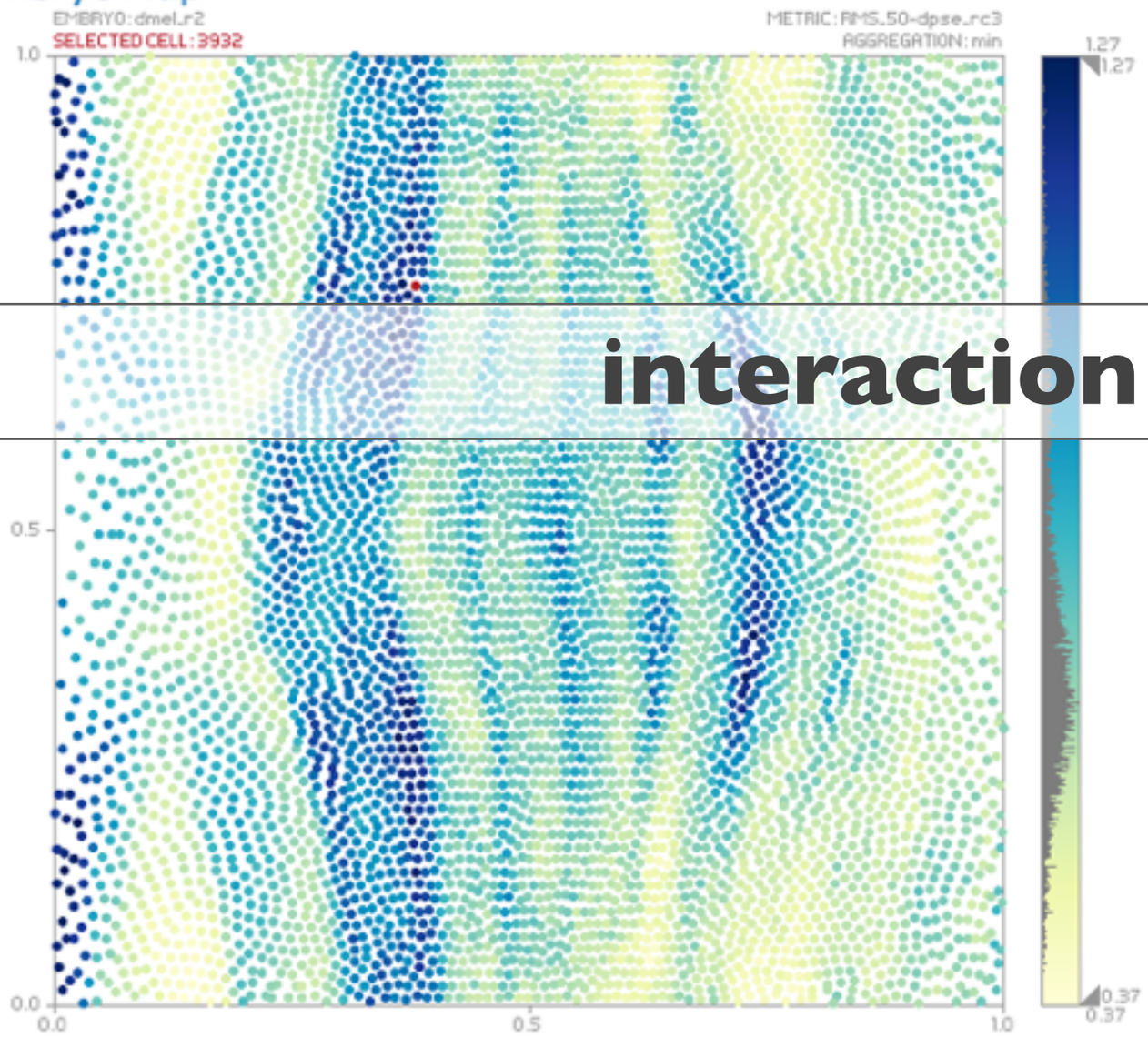
Curvemap

MulteeSum

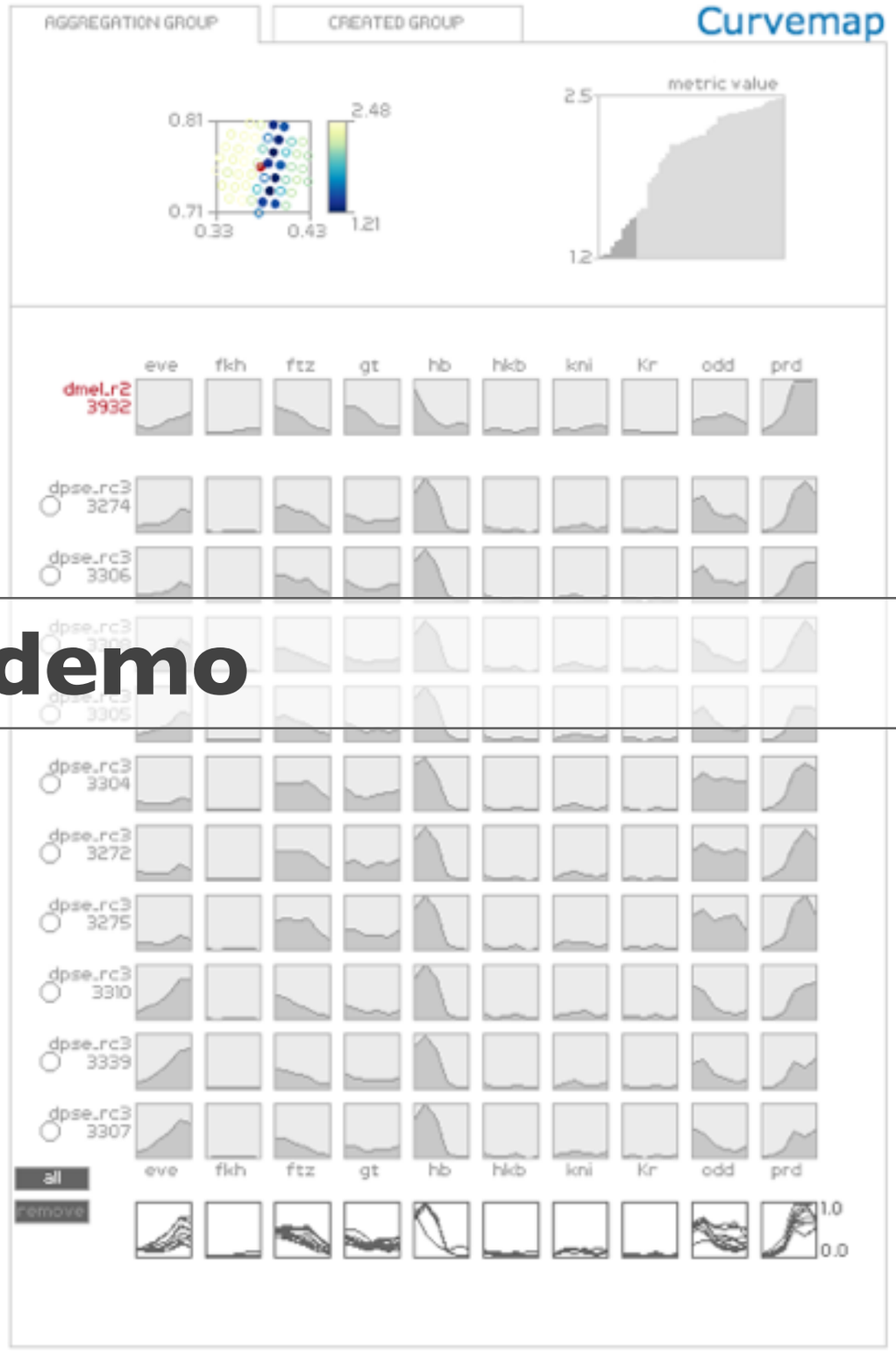
Summaries



Embryo Map



Curvemap



interaction demo

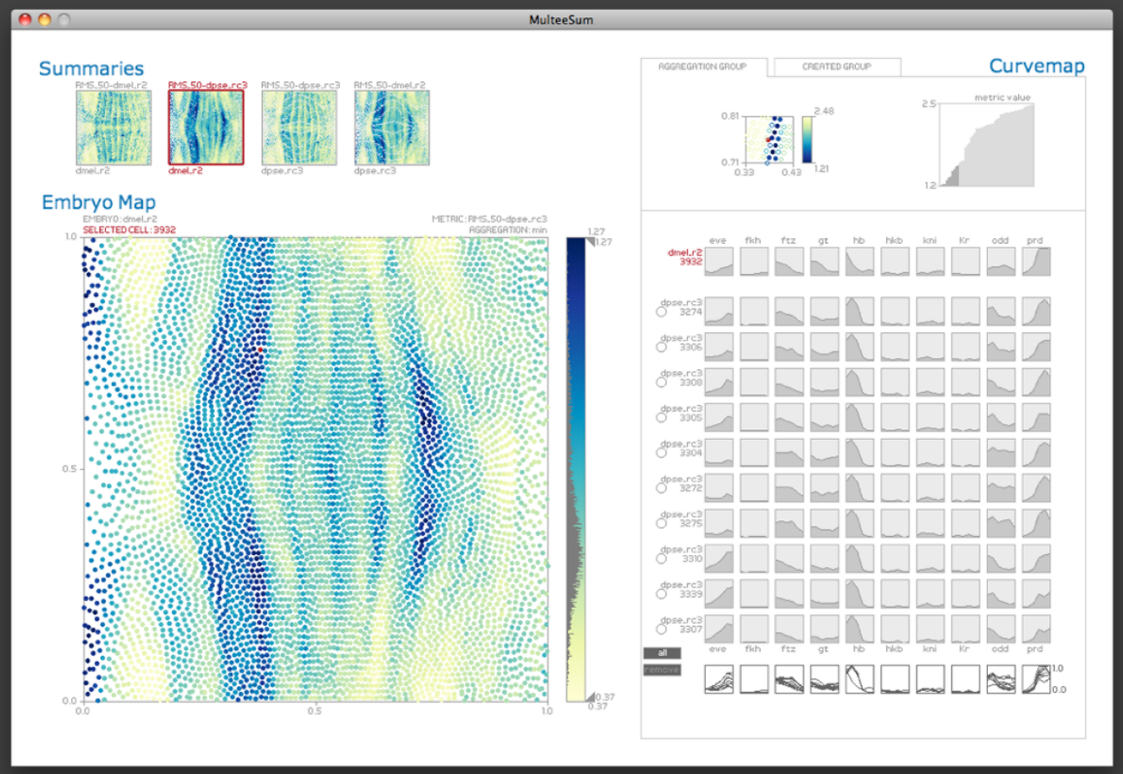
data & tool & tasks

summaries & groups

encodings & interaction

conclusions

contributions



MulteeSum

spatial and temporal gene expression data from multiple species

workflow

visualization supports upstream computation via summaries

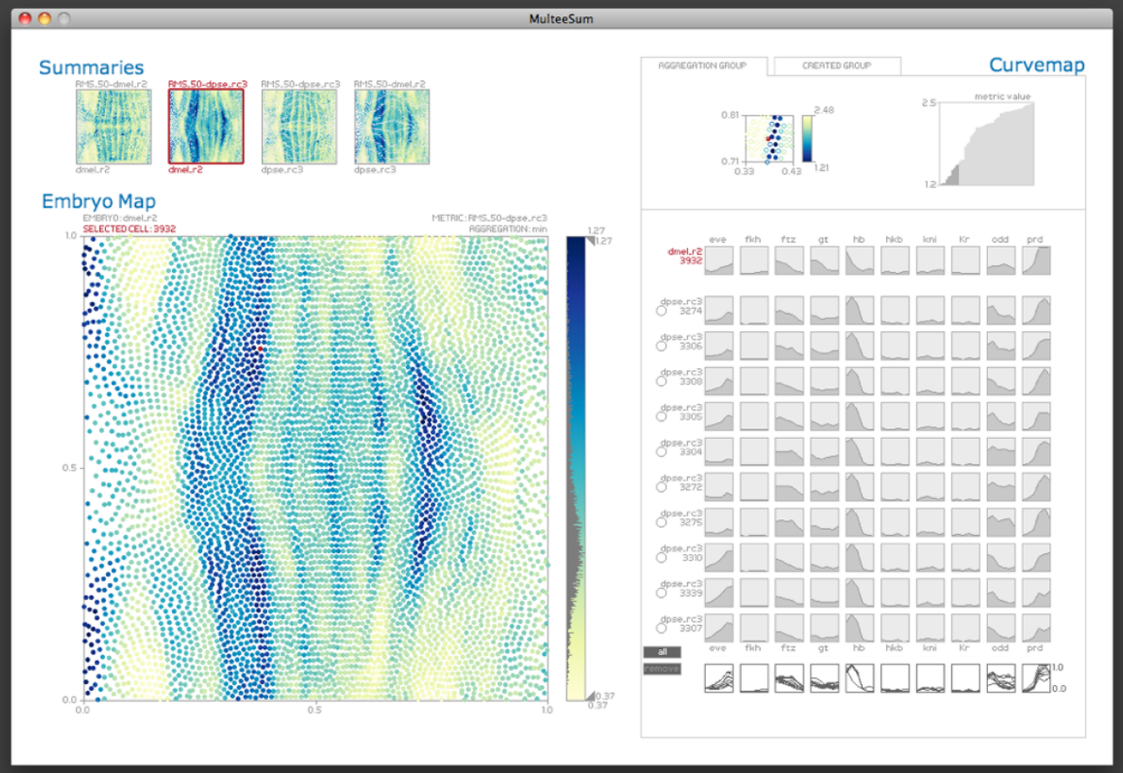
validation

case studies, deployment

contributions

MulteeSum

spatial and temporal gene
expression data from multiple species



workflow

visualization supports upstream computation via summaries

validation

case studies, deployment

Cellular resolution comparison of gene expression in Drosophila reveals coordinated shifts in the segmentation network.

DePace et. al, in preparation.

attribute aggregation

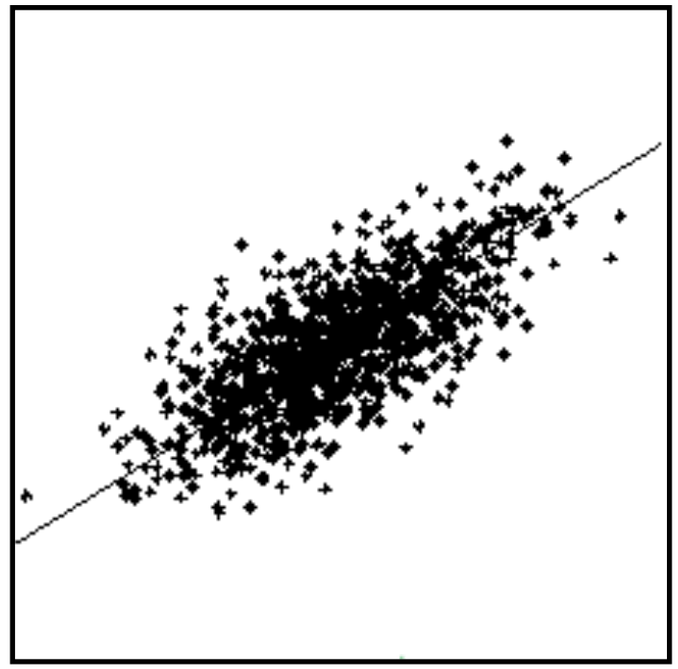
1) group attributes and compute a similarity score across the set

2) dimensionality reduction, to preserve meaningful structure

dimensionality reduction

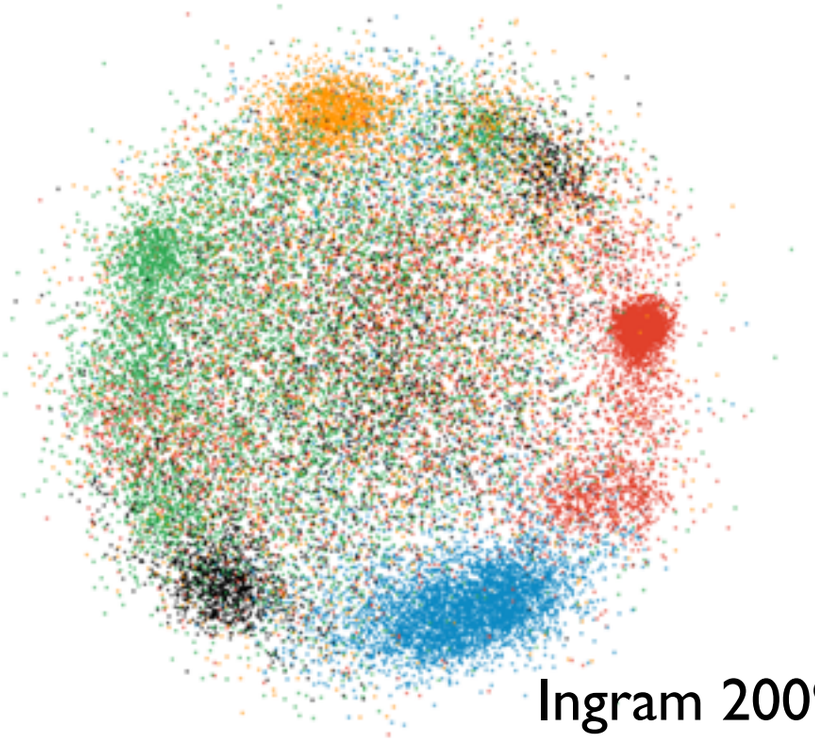
-PCA

- linear approach
- new dimensions are weighted combinations of original ones
- new dimensions created in order of maximum variance



-MDS

- nonlinear class of approaches
- maximize differences in distances from high dim space in the low dim space

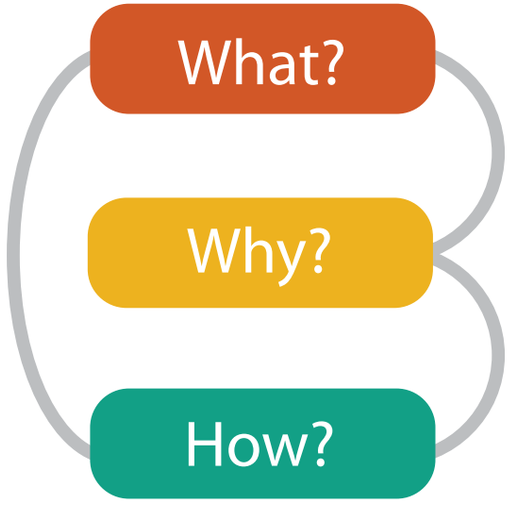


Ingram 2009

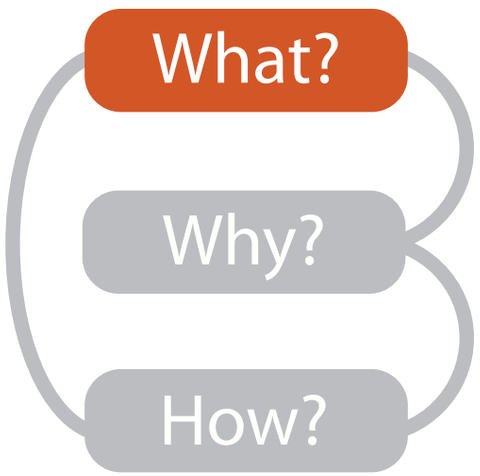
end of foundations...

Analysis: What, why, and how

- what is shown?
  - **data** abstraction
- why is the user looking at it?
  - **task** abstraction
- how is it shown?
  - **idiom**: visual encoding and interaction



- abstract vocabulary avoids domain-specific terms
  - translation process iterative, tricky
- what-why-how analysis framework as scaffold to think systematically about design space



What?

Datasets

Data Types

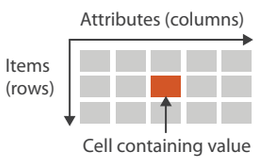
➔ Items ➔ Attributes ➔ Links ➔ Positions ➔ Grids

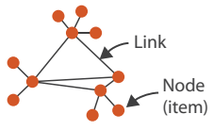
Data and Dataset Types

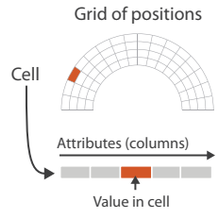
Tables	Networks & Trees	Fields	Geometry	Clusters, sets, lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

Dataset Types

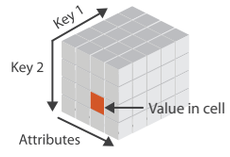
➔ Tables ➔ Networks ➔ Fields (Continuous)







➔ *Multidimensional Table*



➔ *Trees*

➔ Geometry (Spatial)

Dataset Availability

➔ Static

➔ Dynamic

Attributes

Attribute Types

➔ Categorical

➔ Ordered

➔ *Ordinal*

➔ *Quantitative*

Ordering Direction

➔ Sequential

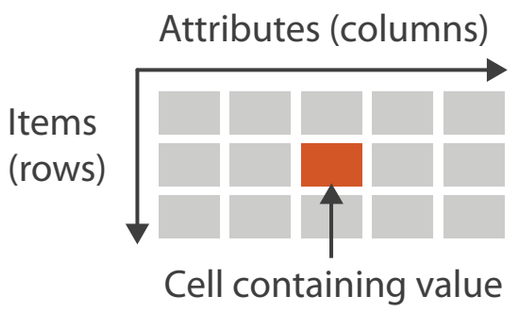
➔ Diverging

➔ Cyclic

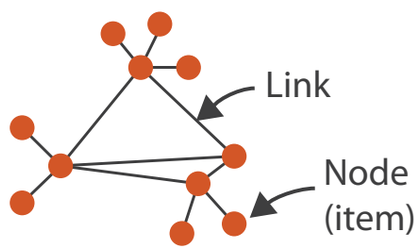
Dataset types

➔ Dataset Types

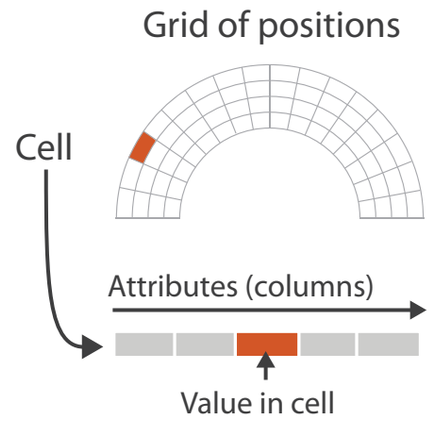
➔ Tables



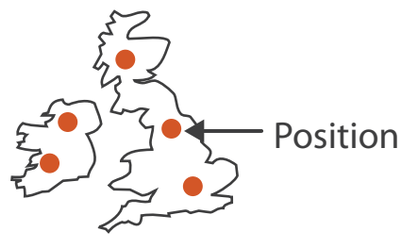
➔ Networks



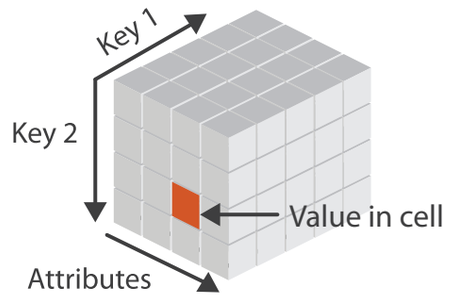
➔ Fields (Continuous)



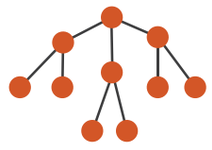
➔ Geometry (Spatial)



➔ *Multidimensional Table*



➔ *Trees*



Dataset and data types

Data and Dataset Types

Tables	Networks & Trees	Fields	Geometry	Clusters, sets, lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

Data Types

➔ Items ➔ Attributes ➔ Links ➔ Positions ➔ Grids

Dataset Availability

➔ Static ➔ Dynamic

Attribute types

- **Attribute Types**
  - Categorical

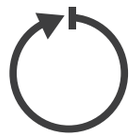
  - Ordered
    - *Ordinal*

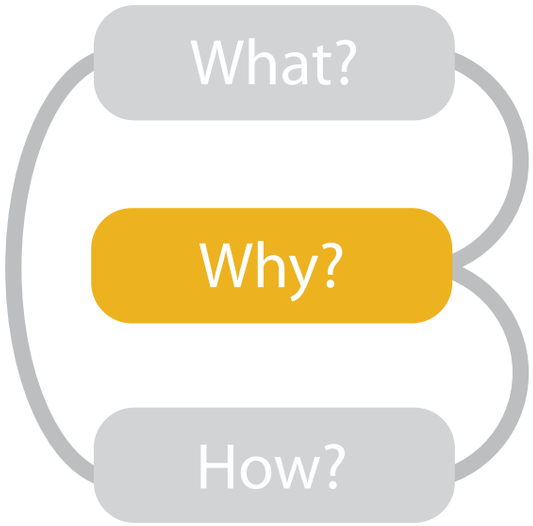
    - *Quantitative*

- **Ordering Direction**
  - Sequential

  - Diverging

  - Cyclic





Why?

Actions

Analyze

→ Consume

→ *Discover* → *Present* → *Enjoy*

→ Produce

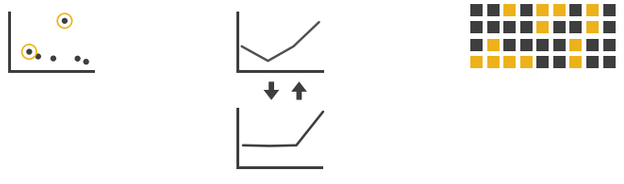
→ *Annotate* → *Record* → *Derive*

Search

	Target known	Target unknown
Location known	*Lookup*	*Browse*
Location unknown	*Locate*	*Explore*

Query

→ Identify → Compare → Summarise



Targets

All Data

→ Trends → Outliers → Features

Attributes

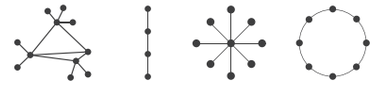
→ One → Many

→ *Distribution* → *Dependency* → *Correlation* → *Similarity*

→ *Extremes*

Network Data

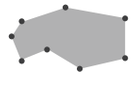
→ Topology

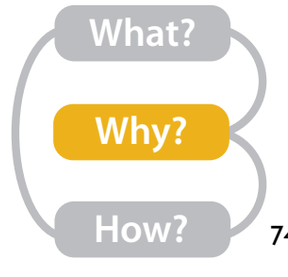


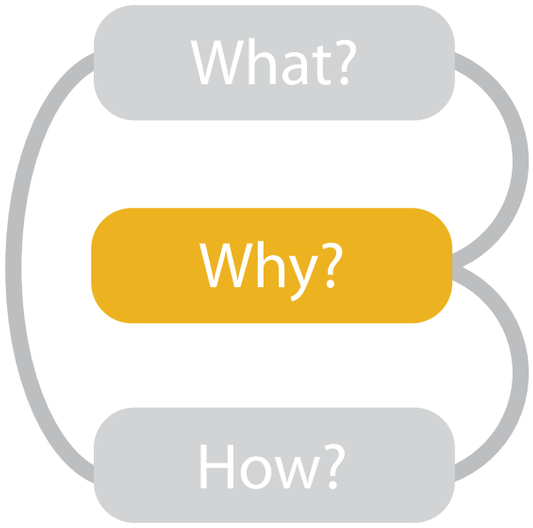
→ *Paths*

Spatial Data

→ Shape







- {action, target} pairs
  - *discover distribution*
  - *compare trends*
  - *locate outliers*
  - *browse topology*

Why?

Actions

Analyze

→ Consume

→ *Discover* → *Present* → *Enjoy*

→ Produce

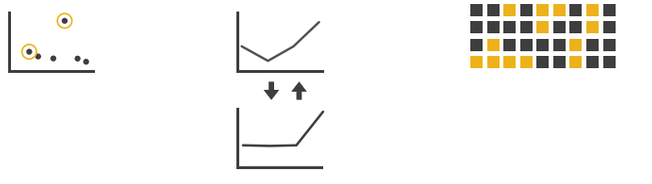
→ *Annotate* → *Record* → *Derive*

Search

	Target known	Target unknown
Location known	*Lookup*	*Browse*
Location unknown	*Locate*	*Explore*

Query

→ Identify → Compare → Summarise



Targets

All Data

→ Trends → Outliers → Features

Attributes

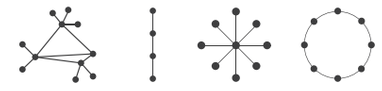
→ One → Many

→ *Distribution* → *Dependency* → *Correlation* → *Similarity*

→ *Extremes*

Network Data

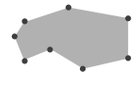
→ Topology

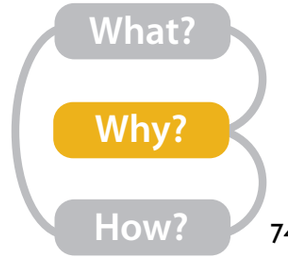


→ *Paths*

Spatial Data

→ Shape





High-level actions: Analyze

- consume
  - discover vs present
    - classic split
    - aka explore vs explain
  - enjoy
    - newcomer
    - aka casual, social
- produce
  - annotate, record
  - derive
    - crucial design choice

➔ **Analyze**

➔ Consume

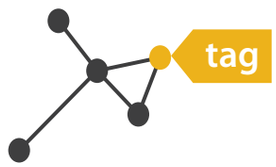
➔ *Discover*

➔ *Present*

➔ *Enjoy*

➔ Produce

➔ *Annotate*



➔ *Record*

➔ *Derive*

Actions: Mid-level search, low-level query

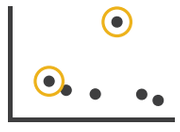
- what does user know?
  - target, location
- how much of the data matters?
  - one, some, all

Search

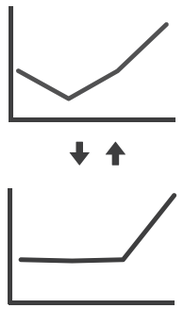
	Target known	Target unknown
Location known	*Lookup*	*Browse*
Location unknown	*Locate*	*Explore*

Query

→ Identify



→ Compare



→ Summarise

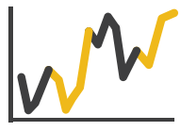
Why: Targets

ALL DATA

- Trends

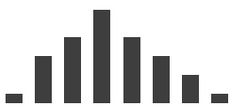
- Outliers

- Features

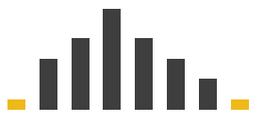


ATTRIBUTES

- One
  - *Distribution*

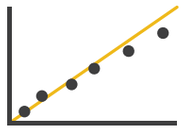


  - *Extremes*

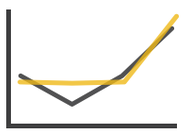


- Many
  - *Dependency*

  - *Correlation*

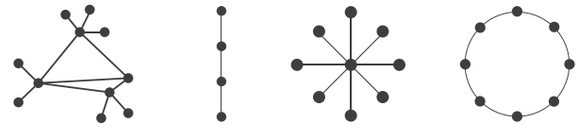


  - *Similarity*



NETWORK DATA

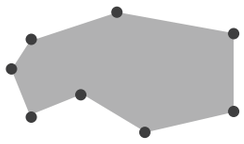
- Topology



- *Paths*

SPATIAL DATA

- Shape



How?

Encode

Arrange

- Express
- Separate
- Order
- Align
- Use

Map

from categorical and ordered attributes

- Color
  - *Hue*
  - *Saturation*
  - *Luminance*
- Size, Angle, Curvature, ...
- Shape
- Motion

Direction, Rate, Frequency, ...

Manipulate

Change

Select

Navigate

Facet

Juxtapose

Partition

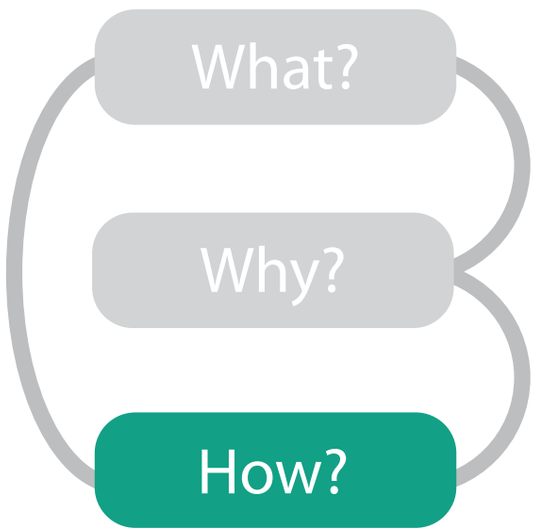
Superimpose

Reduce

Filter

Aggregate

Embed



+ perception and design guidelines...

L12: Tables

REQUIRED READING

Chapter 7

Arrange Tables

7.1 The Big Picture

Figure 7.1 shows the four visual encoding design choices for how to arrange tabular data spatially. One is to express values. The other three are to separate, order, and align regions. The spatial orientation of axes can be rectilinear, parallel, or radial. Spatial layouts may be dense, and they may be space-filling.

► A fifth arrangement choice, to use a given spatial layout, is not an option for nonspatial information; it is covered in Chapter 8.

7.2 Why Arrange?

The **arrange** design choice covers all aspects of the use of spatial channels for visual encoding. It is the most crucial visual encoding choice because the use of space dominates the user's mental model of the dataset. The three highest ranked effectiveness channels for quantitative and ordered attributes are all related to spatial position: planar position along an aligned scale, planar position against a common scale, and length. The highest ranked effectivness channel for categorical attributes, grouping items within the same region, is also about the use of space. Moreover, there are no nonspatial channels that are highly effective for all attribute types: the others are split into being suitable for either ordered or categorical attributes, but not both, because of the principle of expressiveness.

► The primacy of the spatial position channels is discussed at length in Chapter 5, as are the principles of effectiveness and expressiveness.

7.3 Arrange by Keys and Values