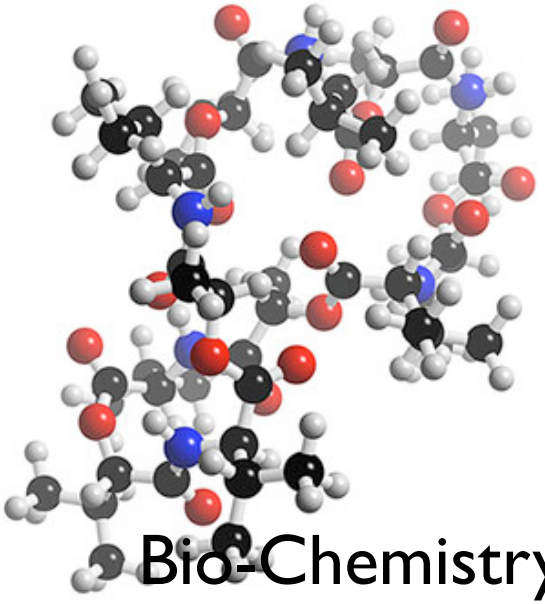
Visual Computing in Biology

Hanspeter Pfister, Harvard University
pfister@seas.harvard.edu

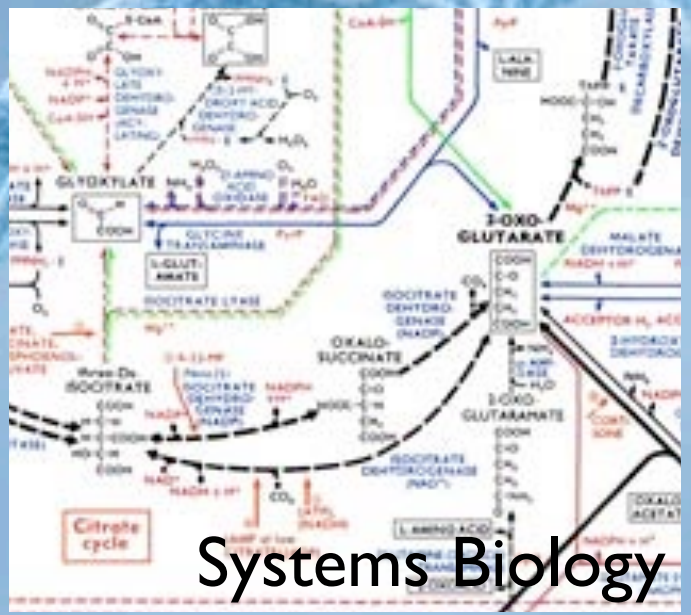
Data Explosion in Biology

Genomics

Microscopy

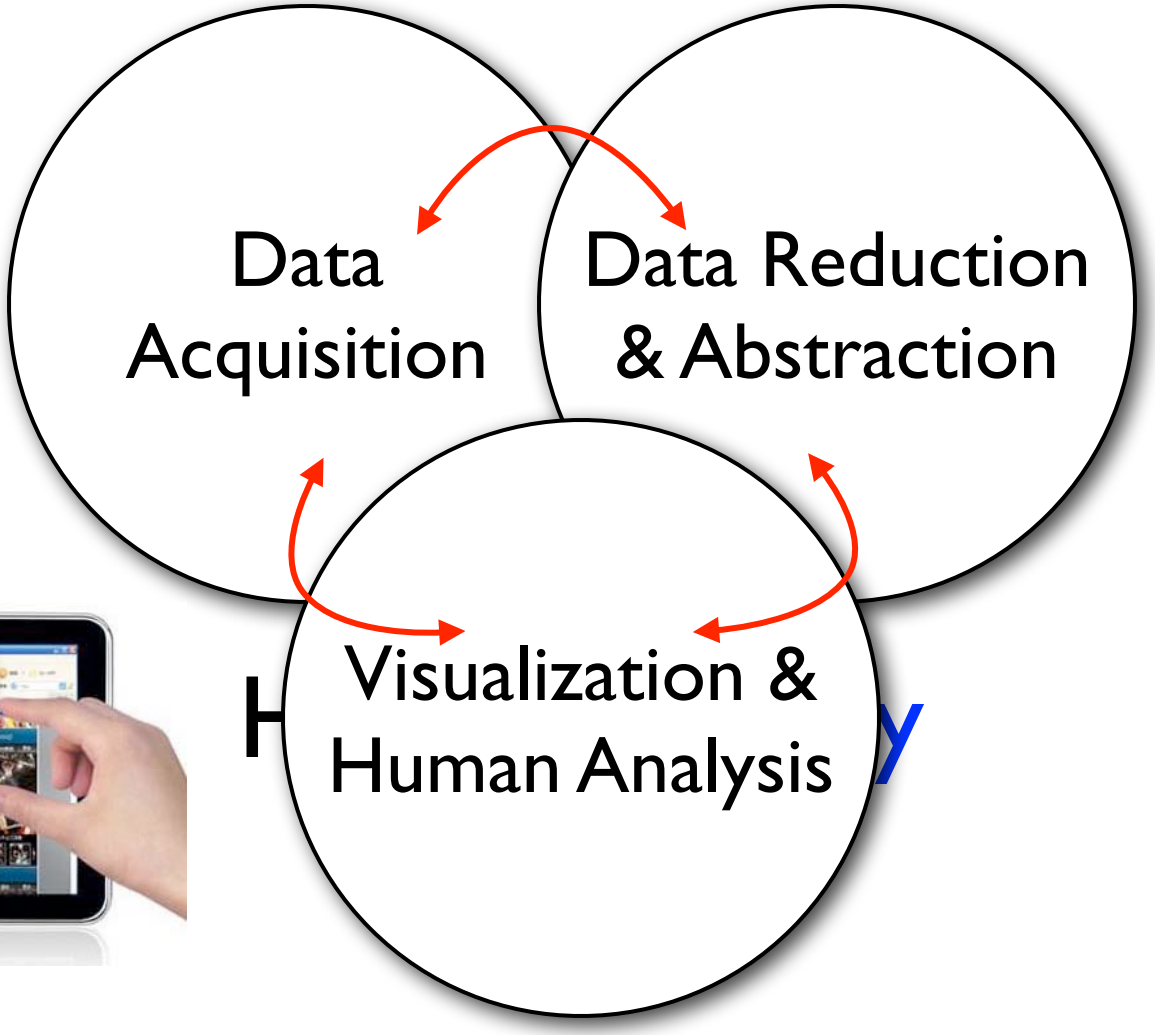


Bio-Chemistry



Systems Biology

Challenges


Data
Acquisition
Data Reduction
& Abstraction
Visualization &
Human Analysis

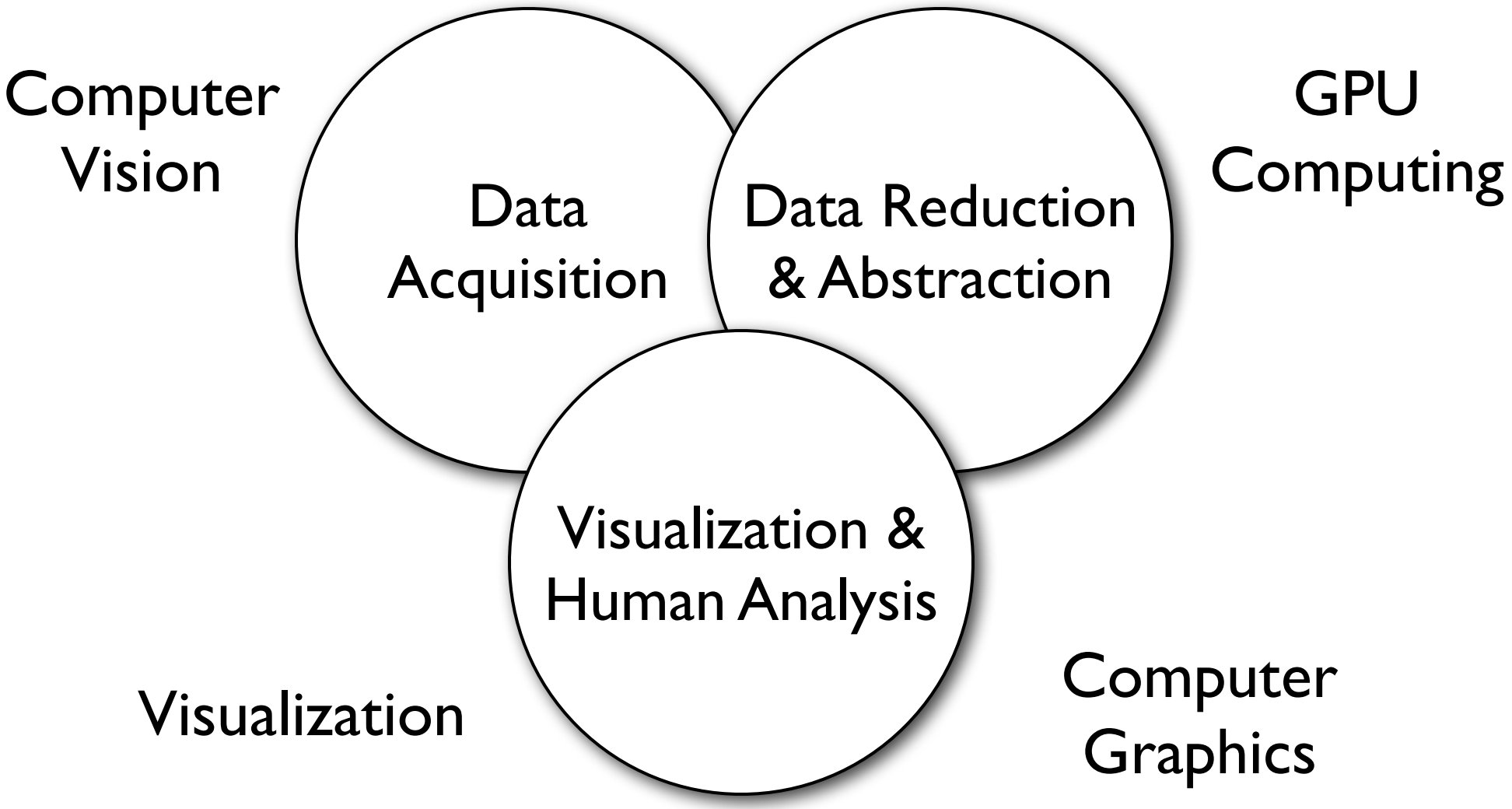
Visual Computing

Data Acquisition

Data Reduction & Abstraction

Visualization & Human Analysis

Visual Computing

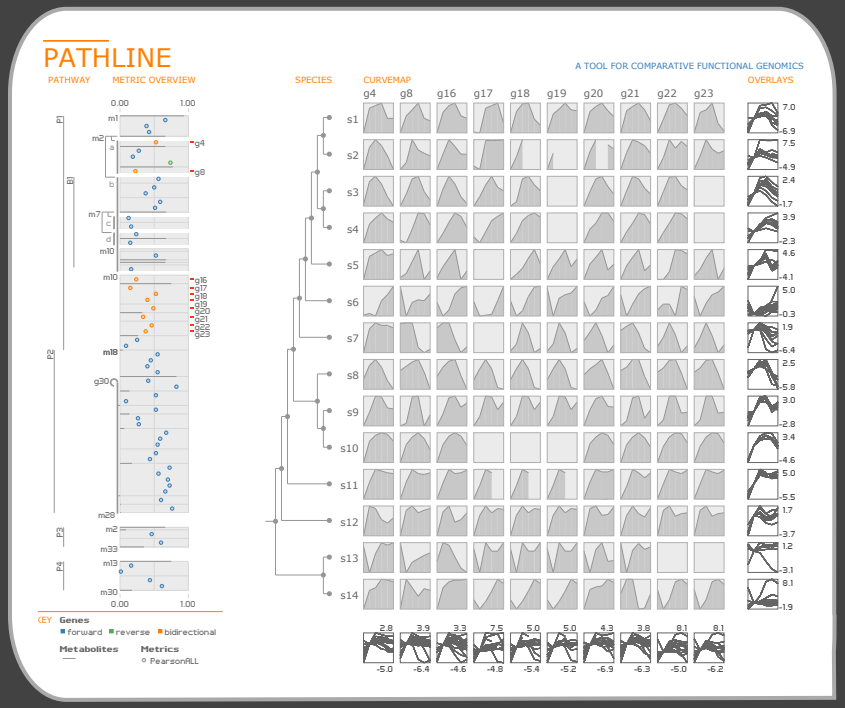


Visual Computing Group

Dr. Nicolas Bonneel
Dr. Verena Kaynig
Dr. Bjoern Andres
Moritz Baecher
Michelle Borkin
Kevin Dale
Amanda Peters
Mike Roberts
Kalyan Sunkavalli
Amelio Vazquez

Genomics and Systems Biology

Miriah Meyer
University of Utah



Pathline

A Tool for Comparative Functional Genomics Data

Miriah Meyer, Bang Wong, Mark Styczynski, Tamara Munzner, Hanspeter Pfister

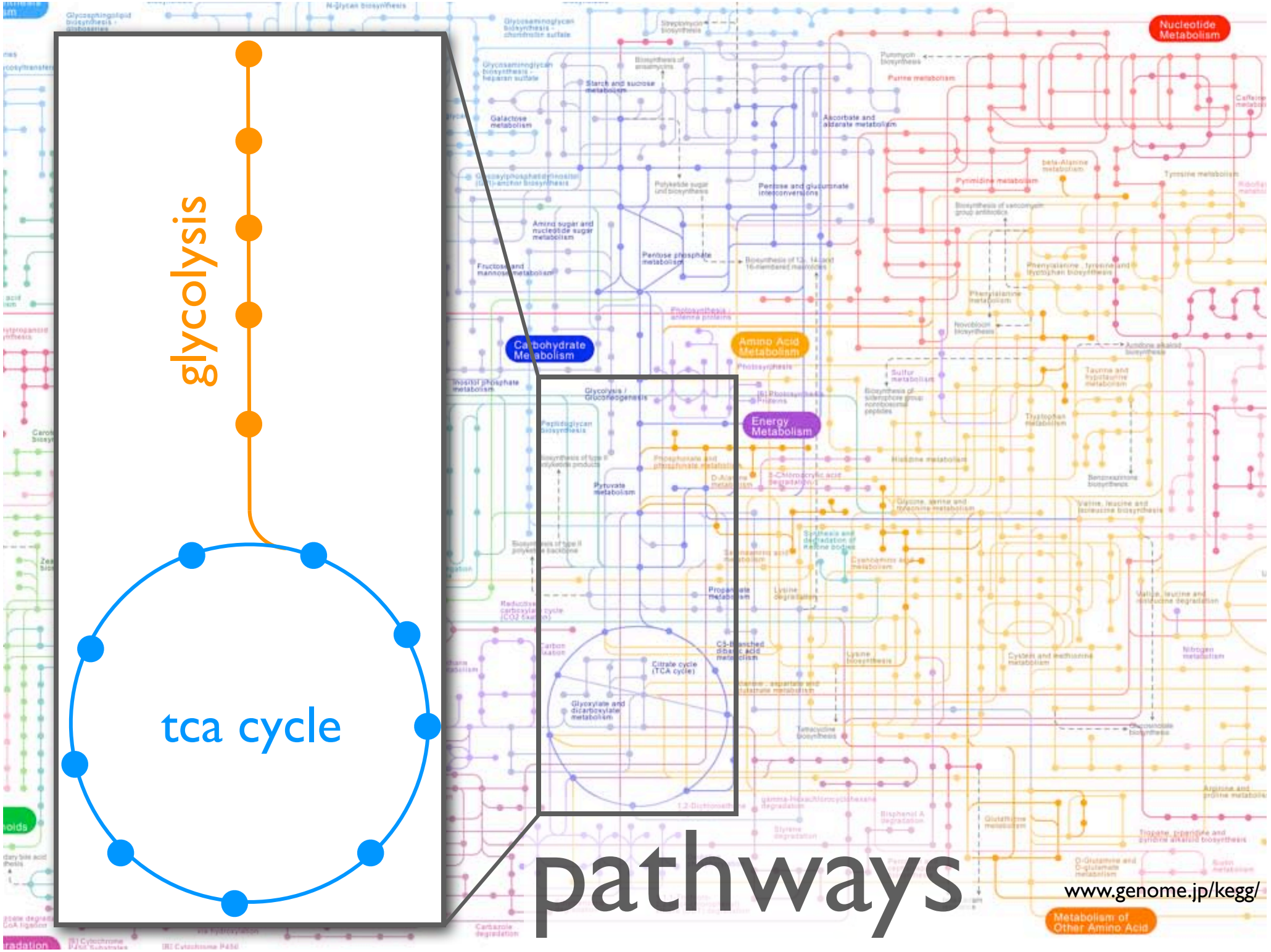
Pathline: A Tool for Comparative Functional Genomics
M. Meyer et al., IEEE/Eurographics EuroVis 2010.

functional genomics

how do genes work together to perform different functions in a cell?

functional genomics data

gene expression
molecular pathways


glycolysis
tca cycle
pathways
www.genome.jp/kegg/
Nucleotide Metabolism
Carbohydrate Metabolism
Amino Acid Metabolism
Energy Metabolism
Metabolism of Other Amino Acid

functional genomics

how do genes work together to perform different functions in a cell?

comparative functional genomics

how do the gene interactions vary across different species?

collaborators: Regev Lab at the Broad Institute

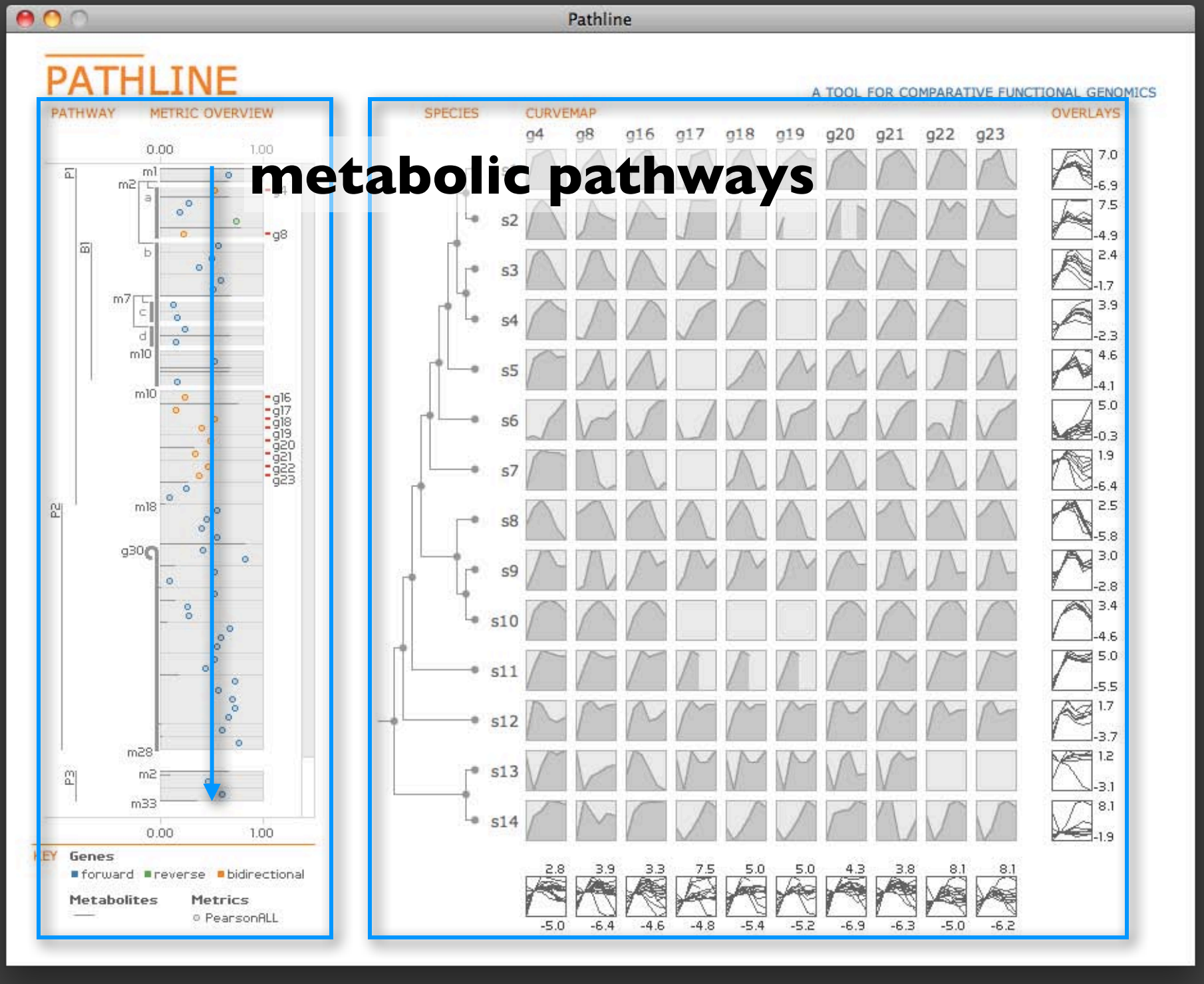
biology: metabolism in yeast

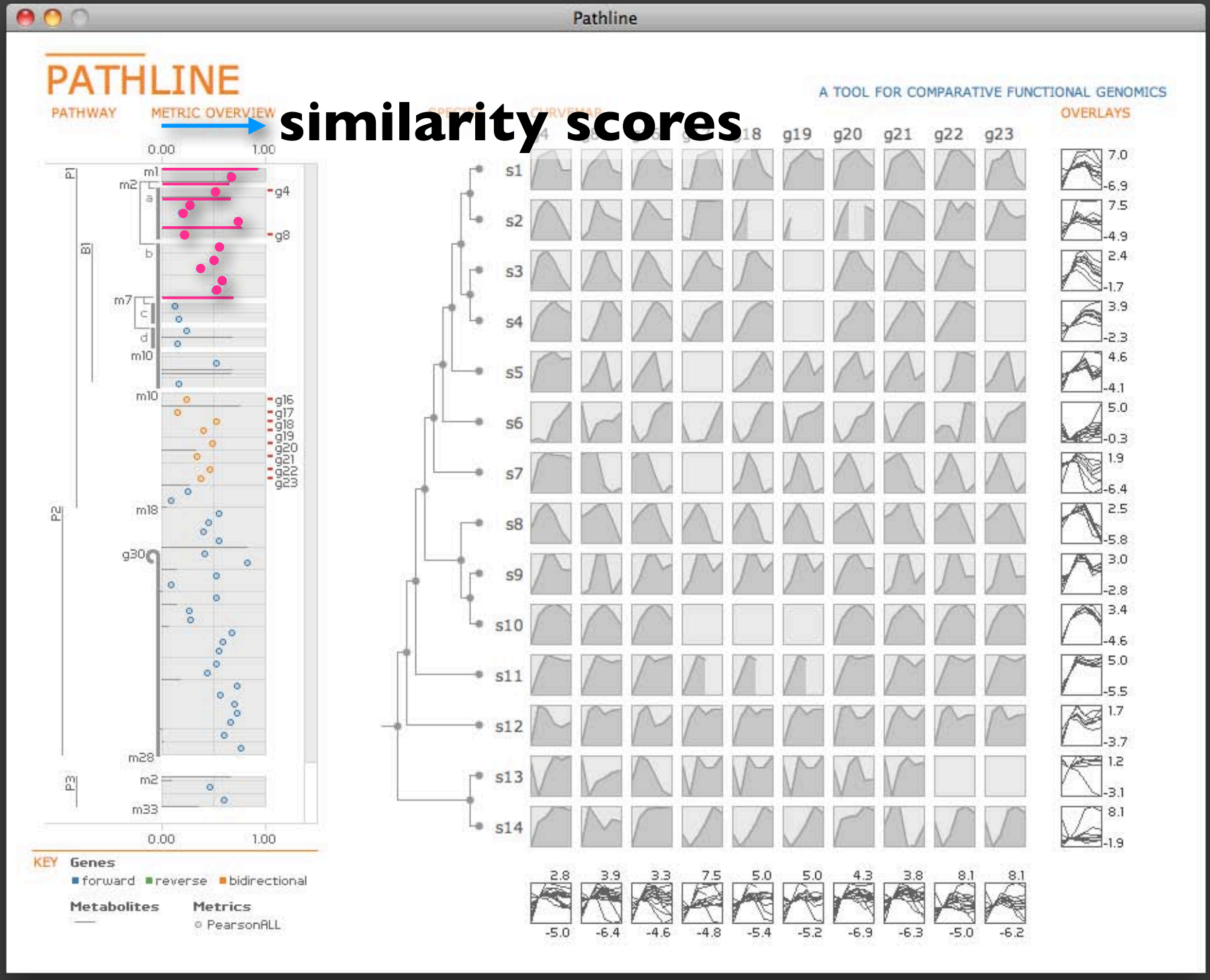
data: multiple genes
multiple time points
multiple related species
multiple pathways

problem: *existing tools can only look at a subset of this data*

comparative functional genomics

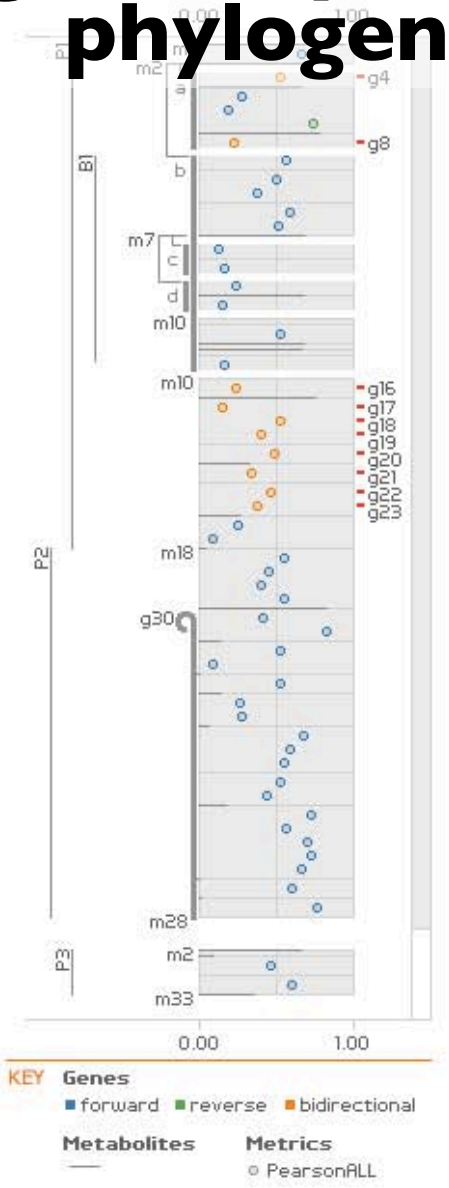
how do the gene interactions vary across different species?


Pathline
PATHLINE
A TOOL FOR COMPARATIVE FUNCTIONAL GENOMICS
PATHWAY
METRIC OVERVIEW
SPECIES
CURVEMAP
OVERLAYS
metabolic pathways
g4 g8 g16 g17 g18 g19 g20 g21 g22 g23
s2 s3 s4 s5 s6 s7 s8 s9 s10 s11 s12 s13 s14
m1 m2 m7 m10 m10 m18 g30 m28 m2 m33
g8 g16 g17 g18 g19 g20 g21 g22 g23
P1 P2 P3 B1
0.00 1.00
KEY Genes
forward reverse bidirectional
Metabolites
Metrics
PearsonALL


Pathline
PATHLINE
A TOOL FOR COMPARATIVE FUNCTIONAL GENOMICS
PATHWAY
METRIC OVERVIEW
similarity scores
OVERLAYS
0.00
1.00
P1
P2
P3
B1
m1
m2
a
b
c
d
m7
m10
m18
g30
m28
m33
g4
g8
g16
g17
g18
g19
g20
g21
g22
g23
g1
g2
g3
g5
g6
g7
g9
g10
g11
g12
g13
g14
g15
s1
s2
s3
s4
s5
s6
s7
s8
s9
s10
s11
s12
s13
s14
KEY
Genes
forward
reverse
bidirectional
Metabolites
Metrics
PearsonALL

Pathline

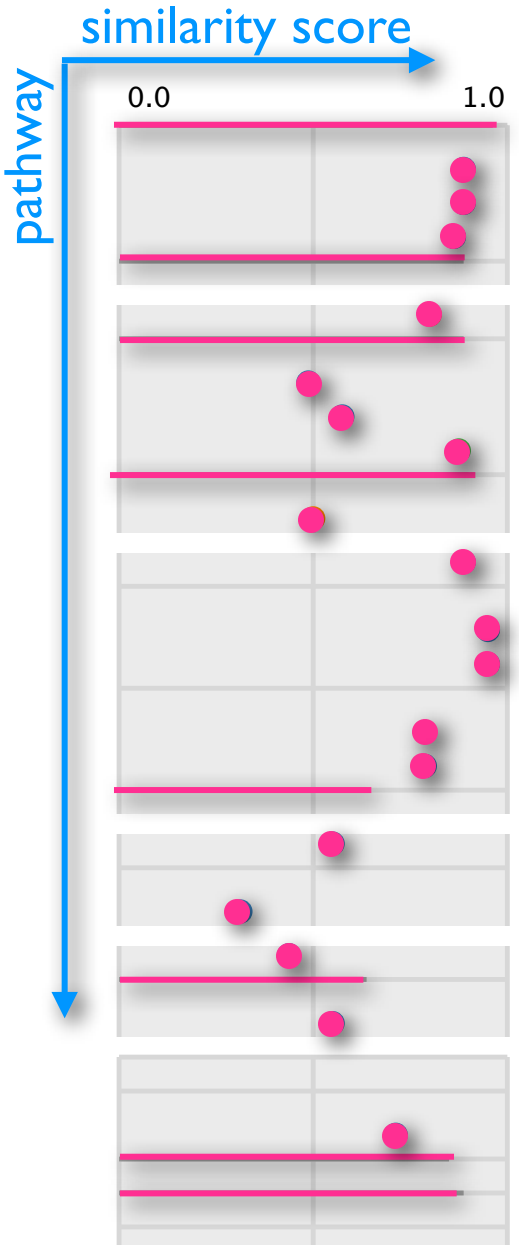
gene expression phylogeny



linearized pathway representation

common axes to compare similarity scores

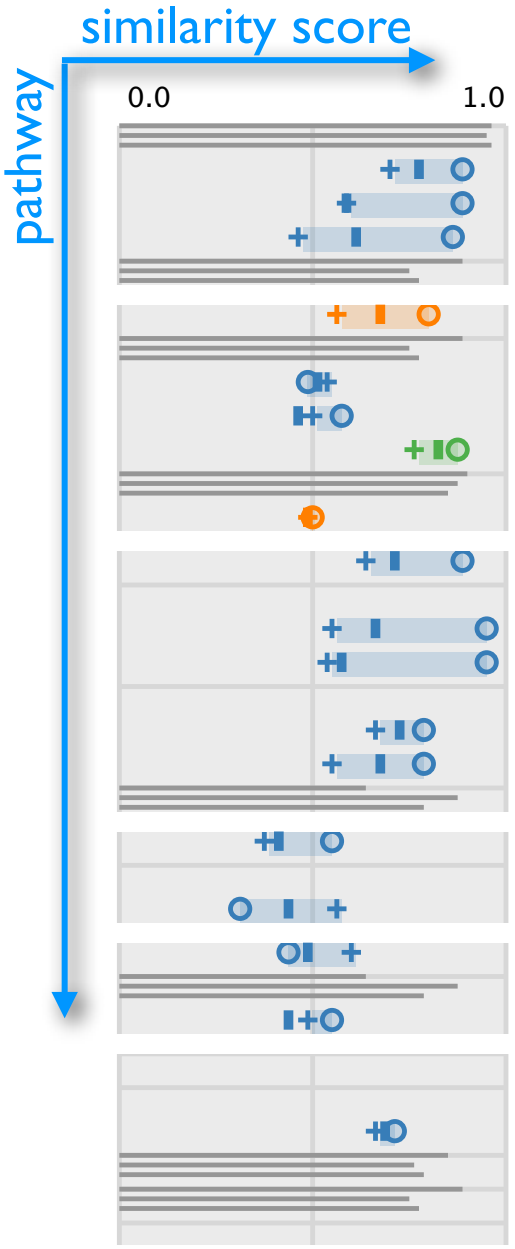
- bars and circles
  - visual layers for selective attention
  - color-code gene direction



linearized pathway representation

common axes to compare similarity scores

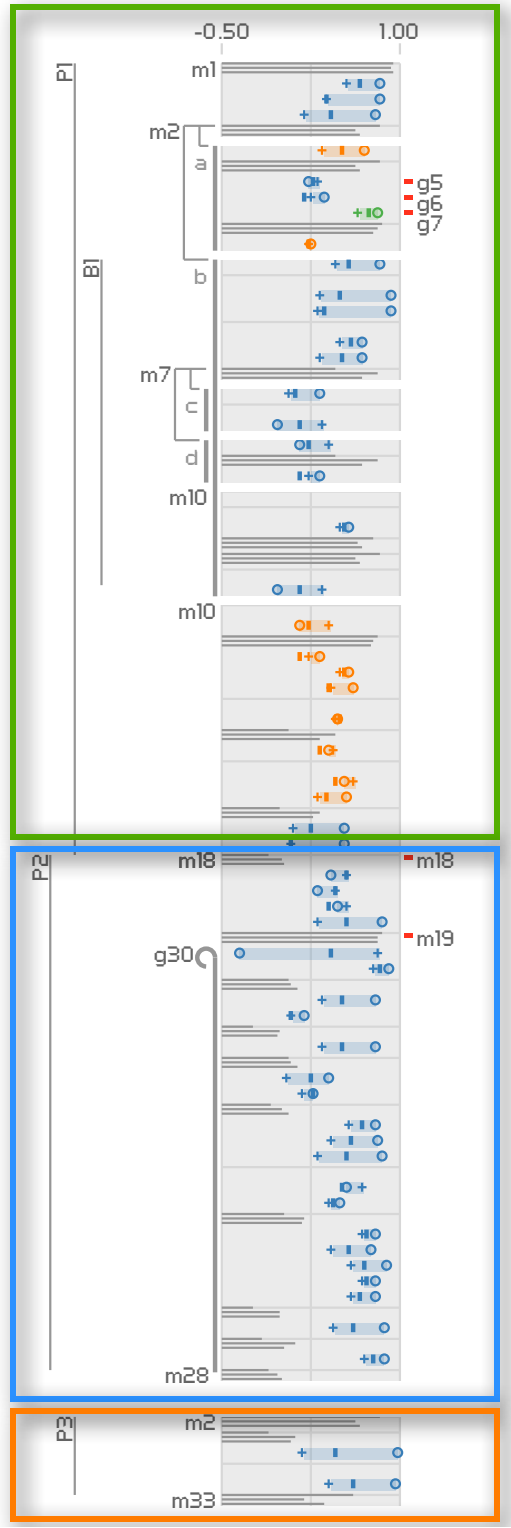
- bars and circles
  - visual layers for selective attention
  - color-code gene direction
- multiple similarity scores



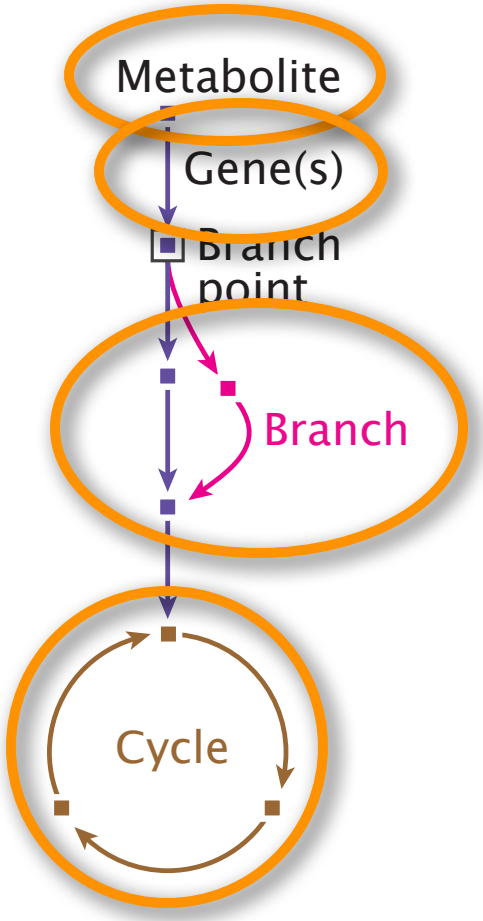
linearized pathway representation

common axes to compare similarity scores

- bars and circles
  - visual layers for selective attention
  - color-code gene direction
- multiple similarity scores
- multiple pathways



pathway to ordered list of nodes

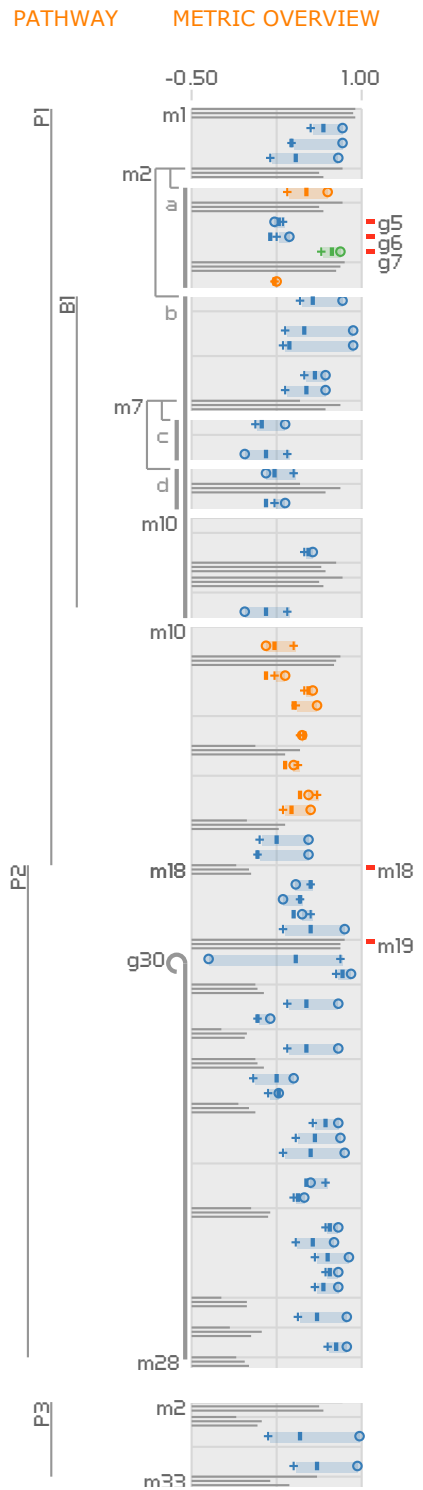


unroll and cut **reinsert** **shared coordinate frame**

linearized pathway representation

putting it together . . .

- use spatial position for similarity scores
- topology is secondary



curvemap

inspired by heatmaps

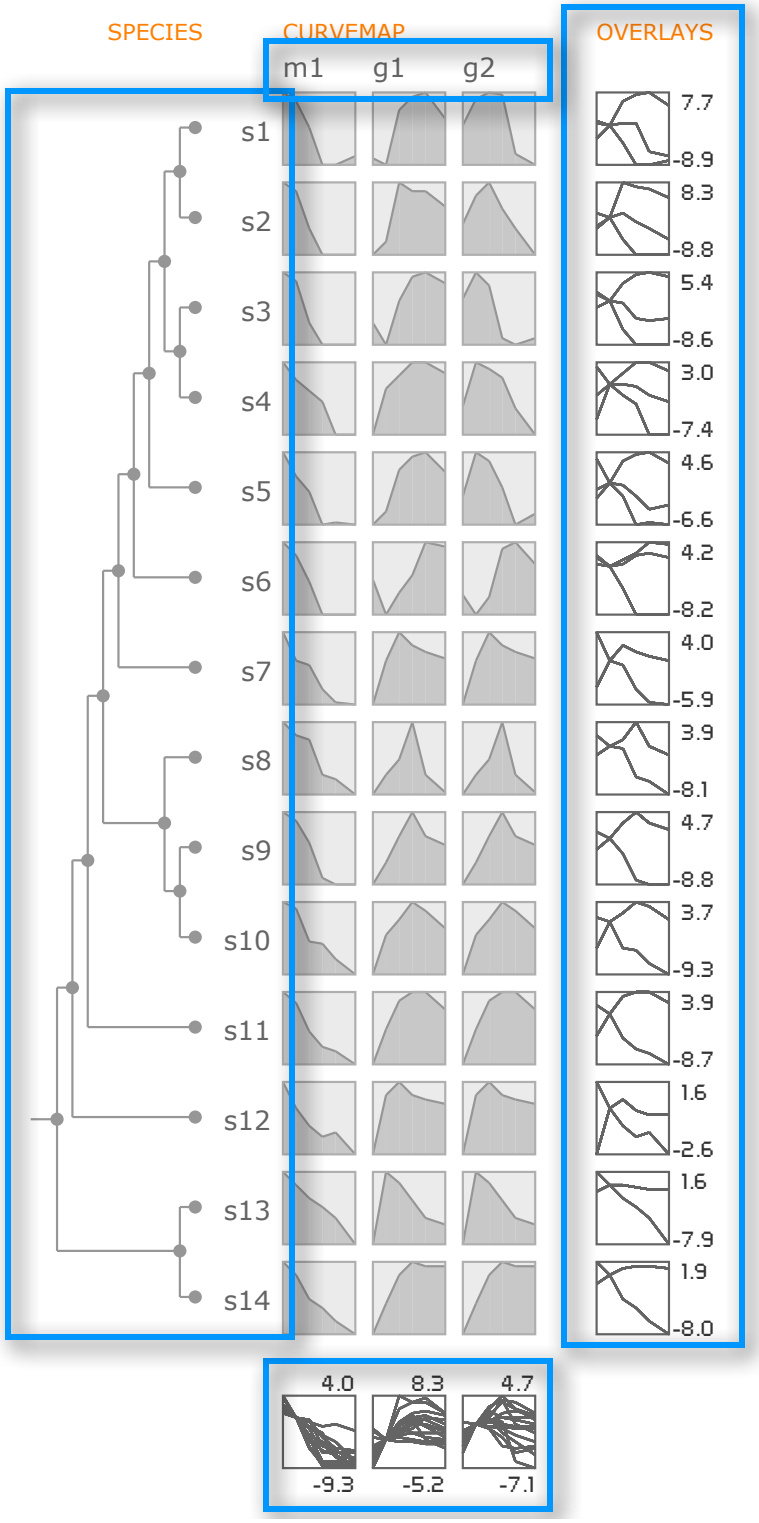
- base visual unit is a curve

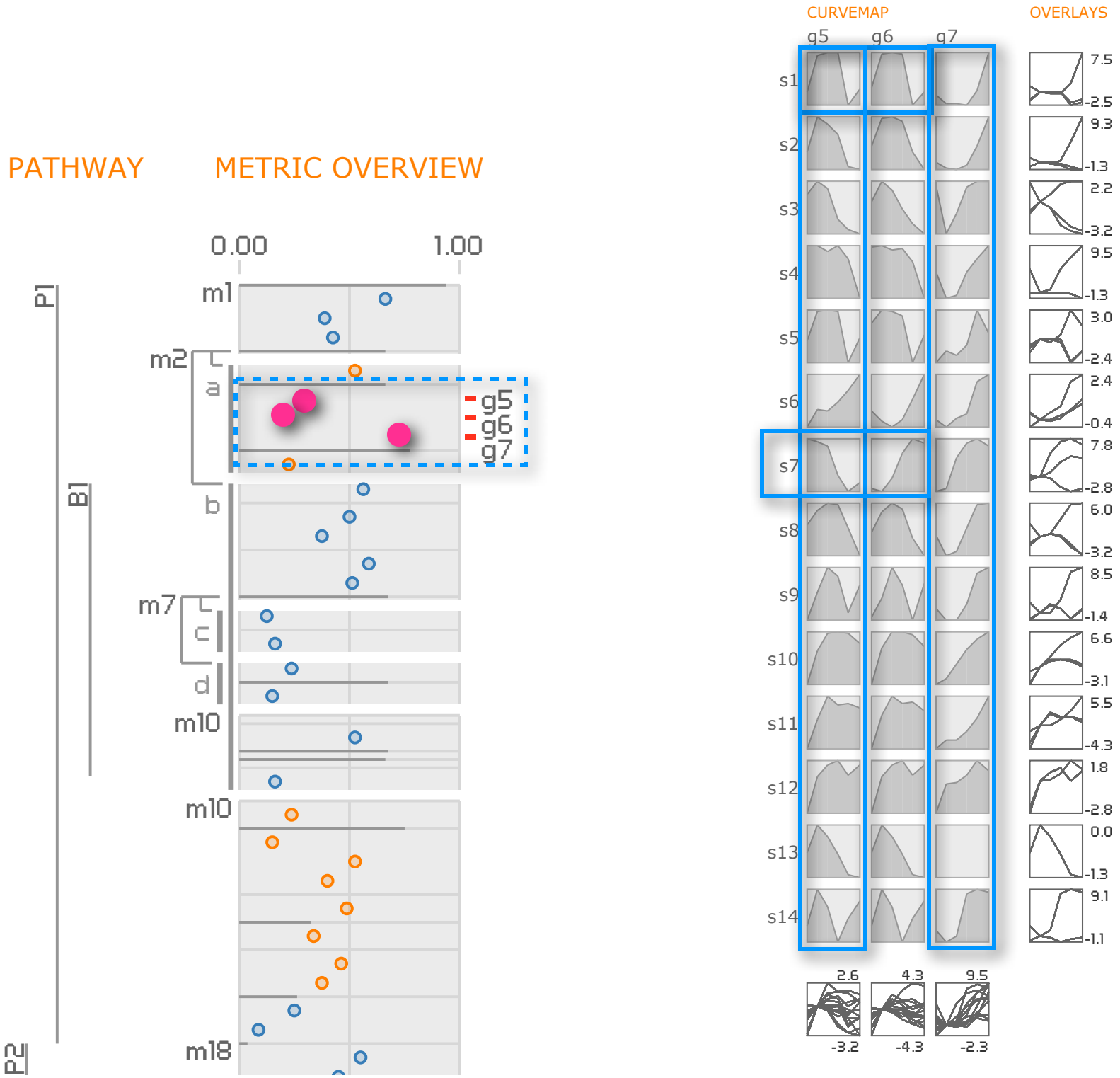
- filled, framed line charts to enhance shape perception

curvemap

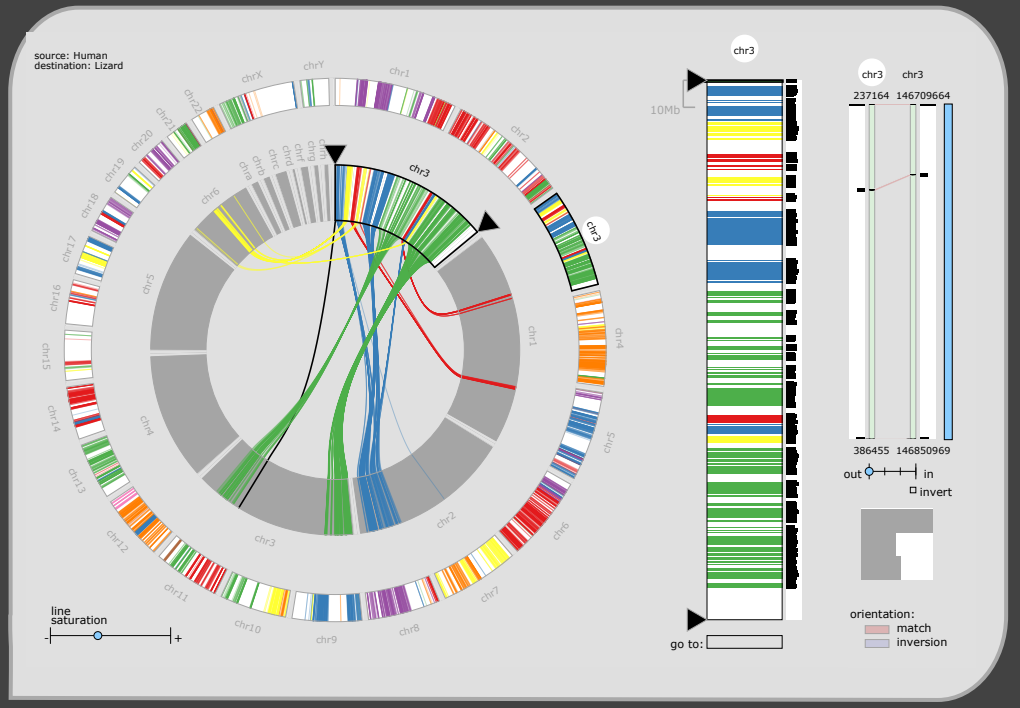
inspired by heatmaps

- base visual unit is a curve
- filled, framed line charts to enhance shape perception
- rows are species
- columns are genes/metabolites
- overlays to enhance trends




PATHWAY
METRIC OVERVIEW
0.00
1.00
P1
B1
P2
m1
m2
a
b
m7
c
d
m10
m10
m18
g5
g6
g7
CURVEMAP
OVERLAYS
g5
g6
g7
s1
s2
s3
s4
s5
s6
s7
s8
s9
s10
s11
s12
s13
s14
7.5
-2.5
9.3
-1.3
2.2
-3.2
9.5
-1.3
3.0
-2.4
2.4
-0.4
7.8
-2.8
6.0
-3.2
8.5
-1.4
6.6
-3.1
5.5
-4.3
1.8
-2.8
0.0
-1.3
9.1
-1.1
2.6
-3.2
4.3
-4.3
9.5
-2.3

www.pathline.org



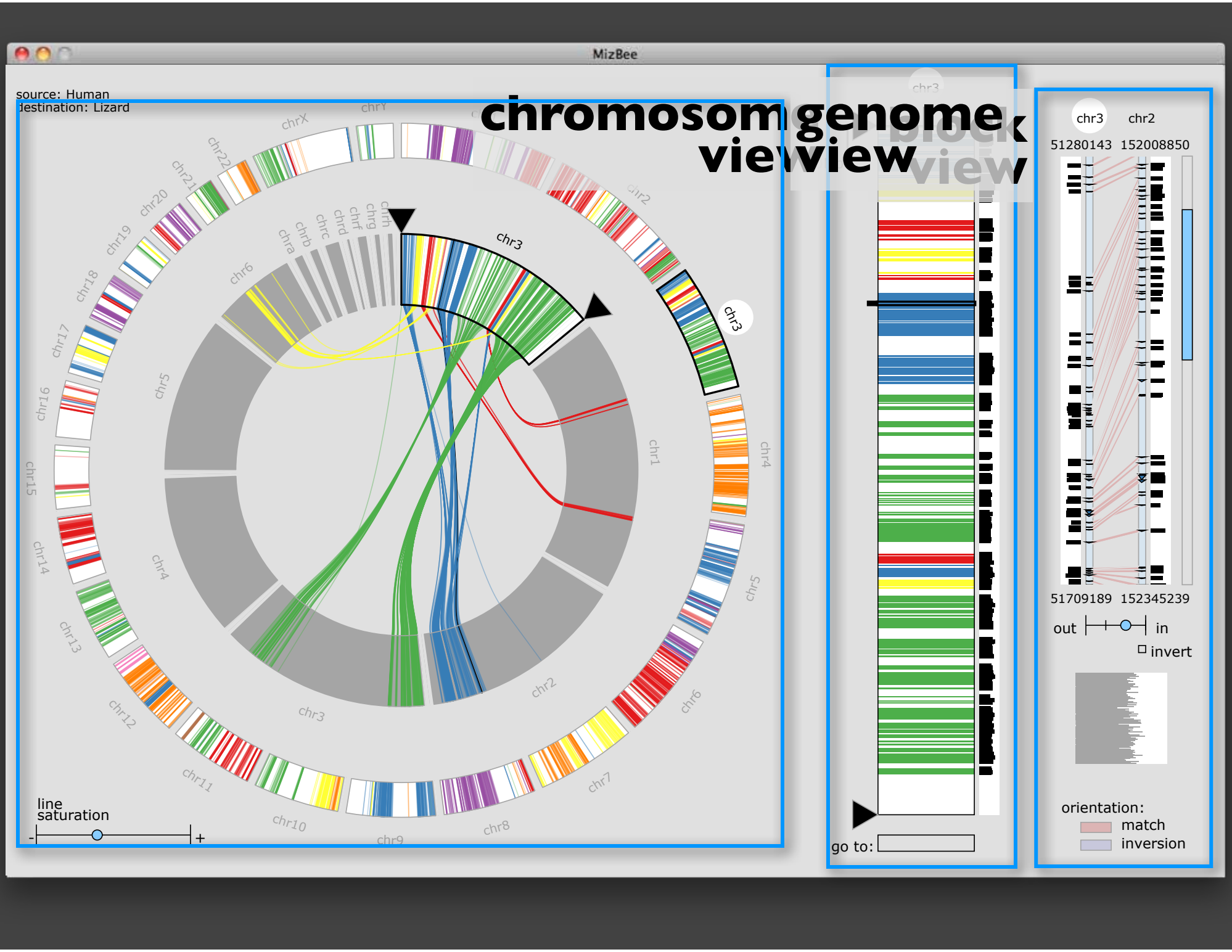
MizBee

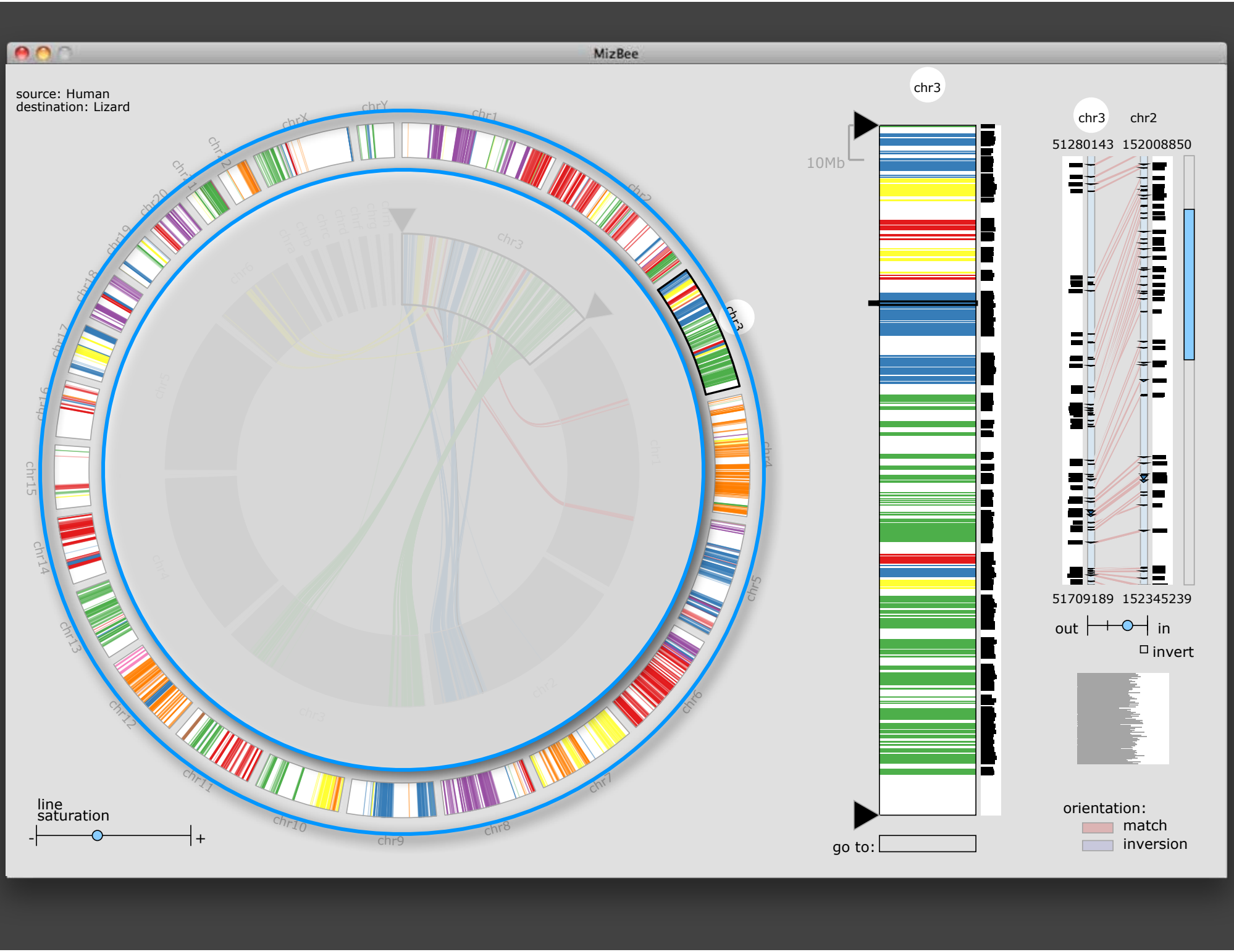
A Browser for Comparative Genomics Data

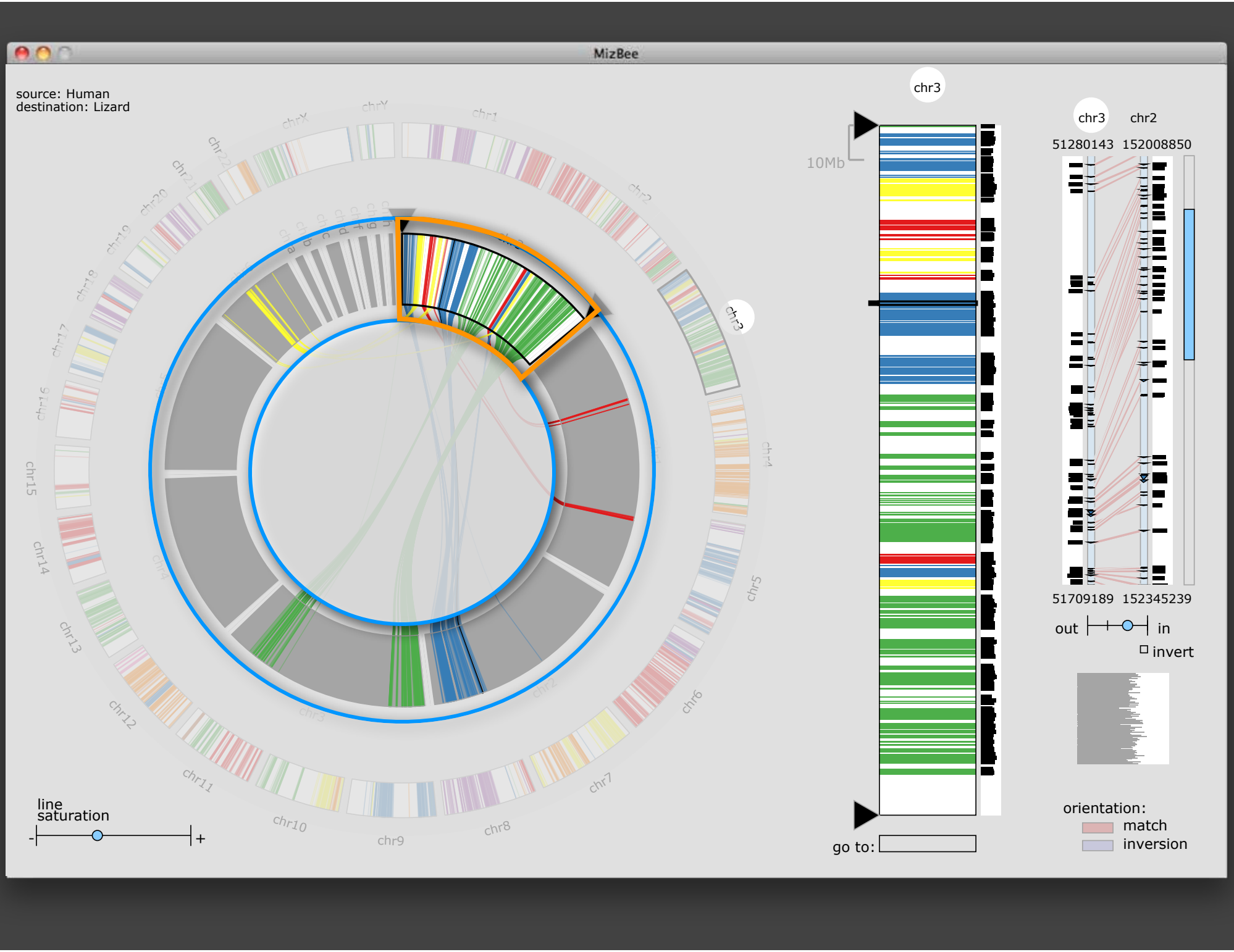
Miriah Meyer, Tamara Munzner, Hanspeter Pfister

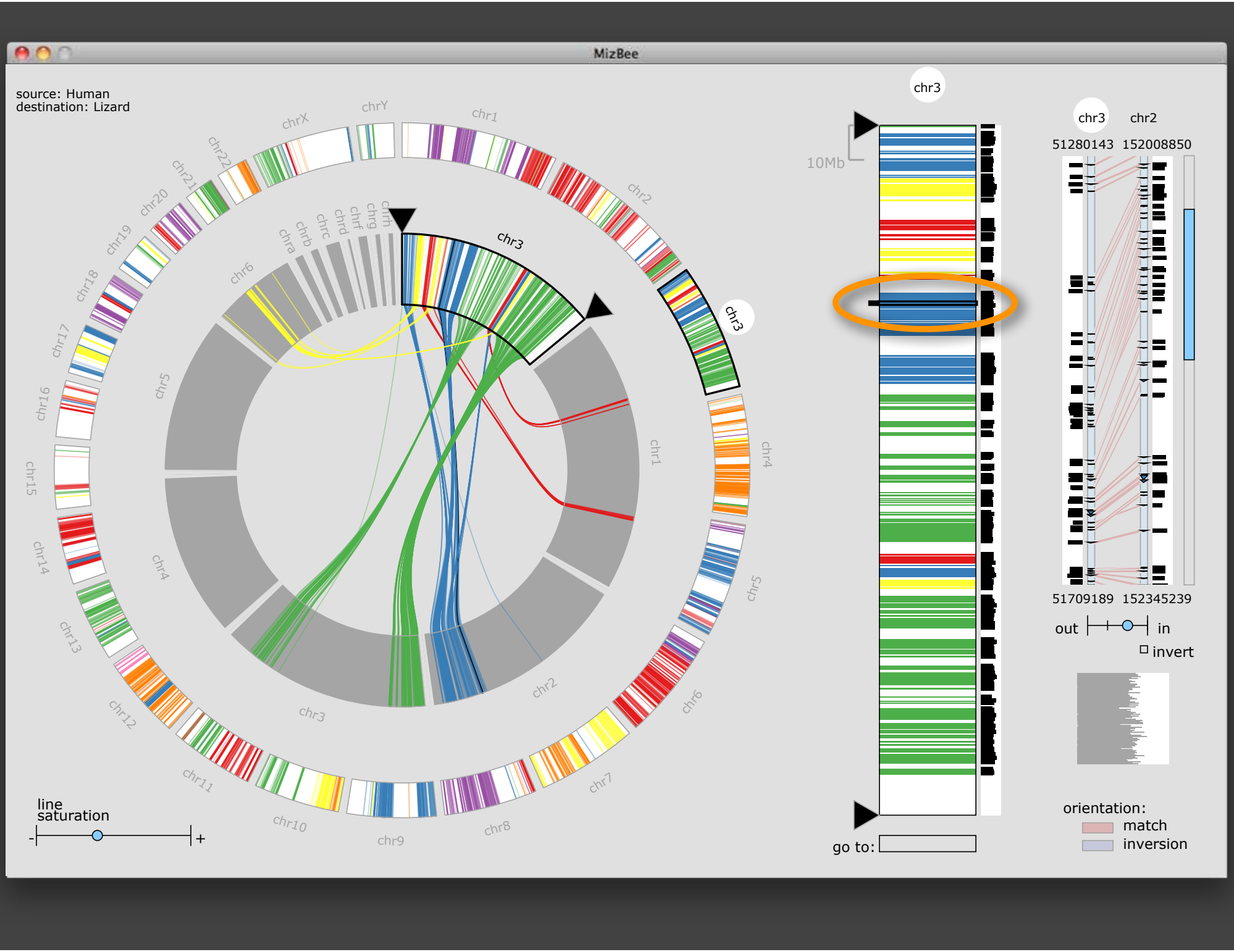
MizBee: A Multiscale Synteny Browser.
M. Meyer et al., IEEE InfoVis 2009.

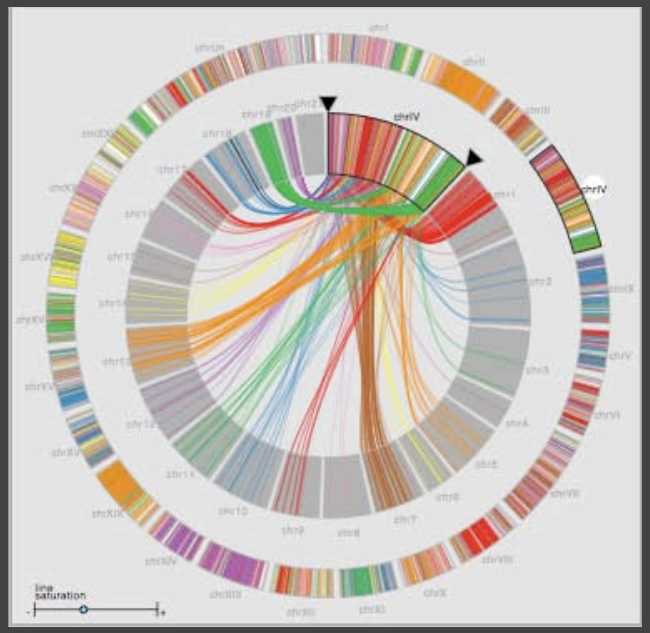
- comparative genomics
- compare regions of similar genomic sequences between genomes
- at multiple scales, from genomes to genes


MizBee
source: Human
destination: Lizard
chromosome view
genome view
block view
chr3
chr2
51280143
152008850
51709189
152345239
out
in
invert
orientation:
match
inversion
go to:
line saturation
-
+


MizBee
source: Human
destination: Lizard
chrX
chrY
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
line
saturation
-
+
chr3
10Mb
go to:
chr3
chr2
51280143
152008850
51709189
152345239
out
in
invert
orientation:
match
inversion


MizBee
source: Human
destination: Lizard
chrX
chrY
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
line
saturation
-
+
chr3
10Mb
go to:
chr3
chr2
51280143
152008850
51709189
152345239
out
in
invert
orientation:
match
inversion

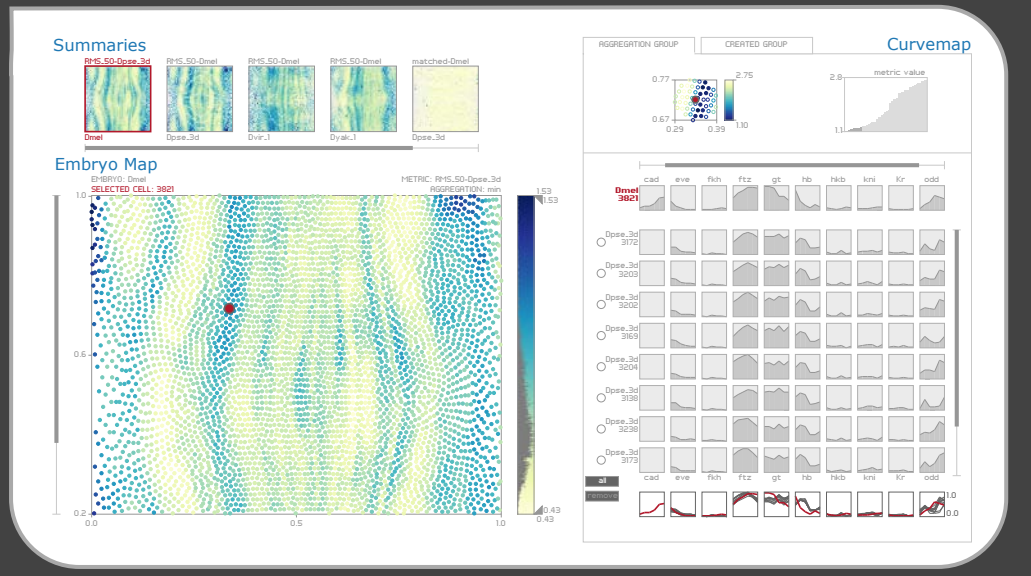

MizBee
source: Human
destination: Lizard
chrX
chrY
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
chra
chrb
chrc
chrd
chrf
chrg
chrh
line
saturation
-
+
chr3
10Mb
go to:
chr3
chr2
51280143
152008850
51709189
152345239
out
in
invert
orientation:
match
inversion



"The first time I saw my data in [MizBee] I was totally disappointed. The data was very noisy, and there were many small blocks that went to different chromosomes."

"Honestly, I don't know. I don't think I would ever have gotten here. The noise was very hard see in the scatter plots while [MizBee] is much more unforgiving."

Genome-wide Synteny Through Highly Sensitive Sequence Alignment: Satsuma
M. Grabherr et al., Bioinformatics 2010.



MulteeSum

A Tool for Exploring Space-Time Expression Data

Miriah Meyer, Tamara Munzner, Angela DePace, Hanspeter Pfister

MulteeSum: A Tool for Comparative Spatial and Temporal Gene Expression Data
M. Meyer et al., IEEE InfoVis 2010.

collaborators: DePace lab at Harvard Medical School

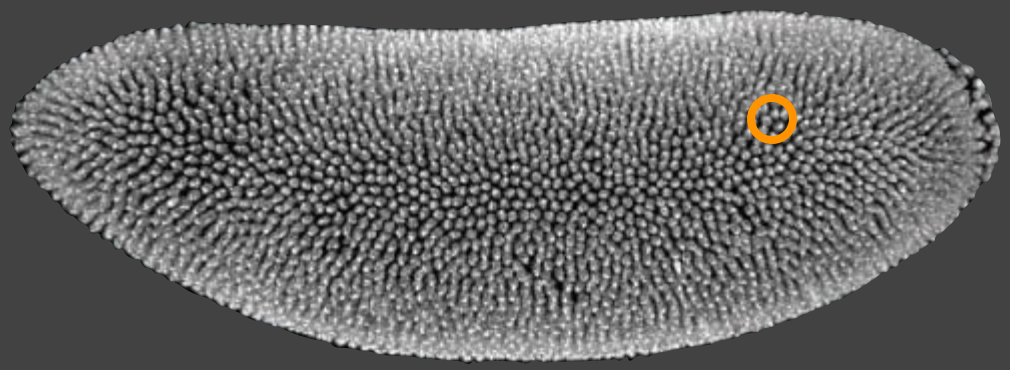
model species: fruit fly

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

requires: characterize differences in gene expression patterns between species

MulteeSum

gene expressions for every cell in a single fruit fly embryo.



correlate gene expression with spatial location.

what are the differences between species?

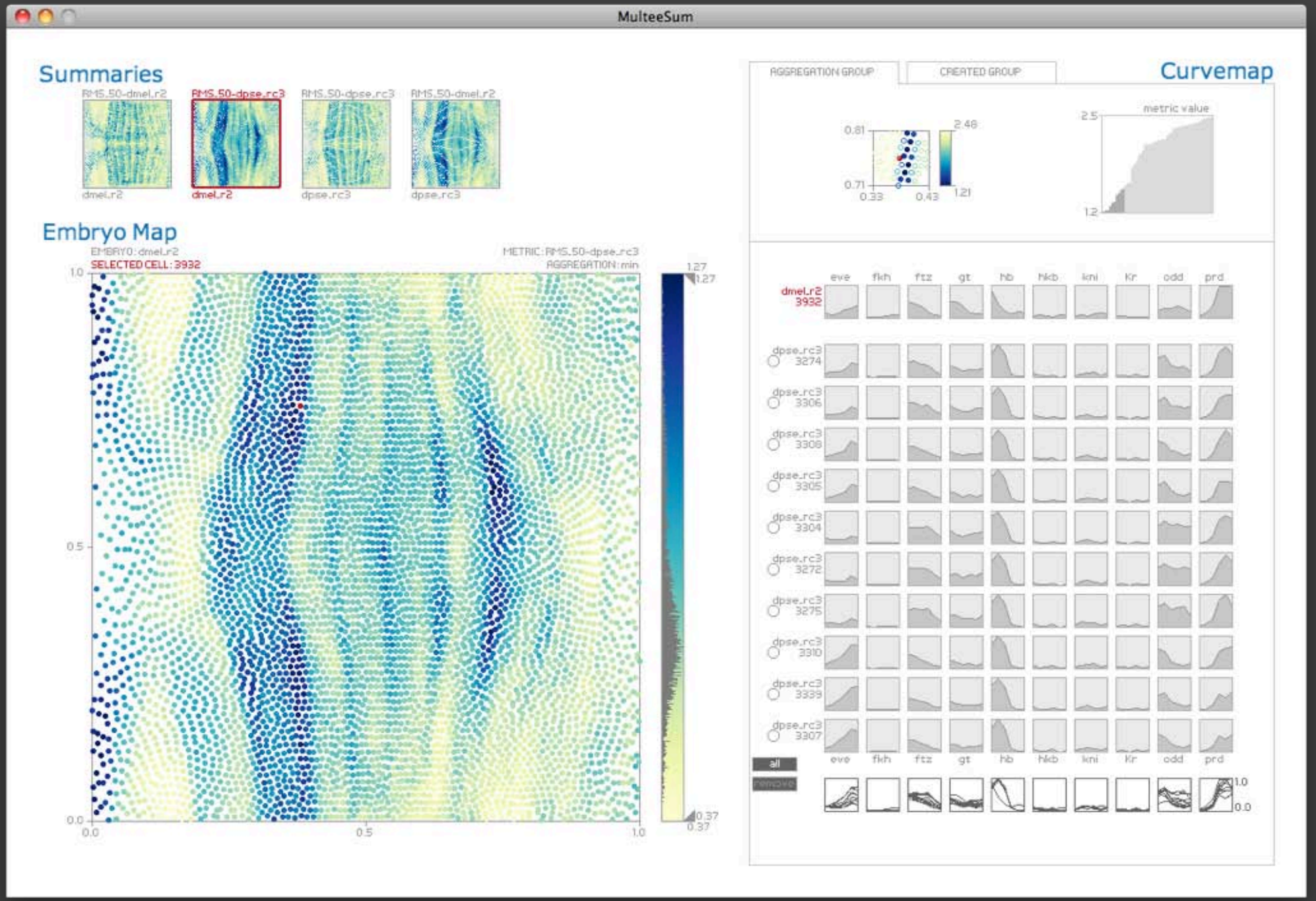
species

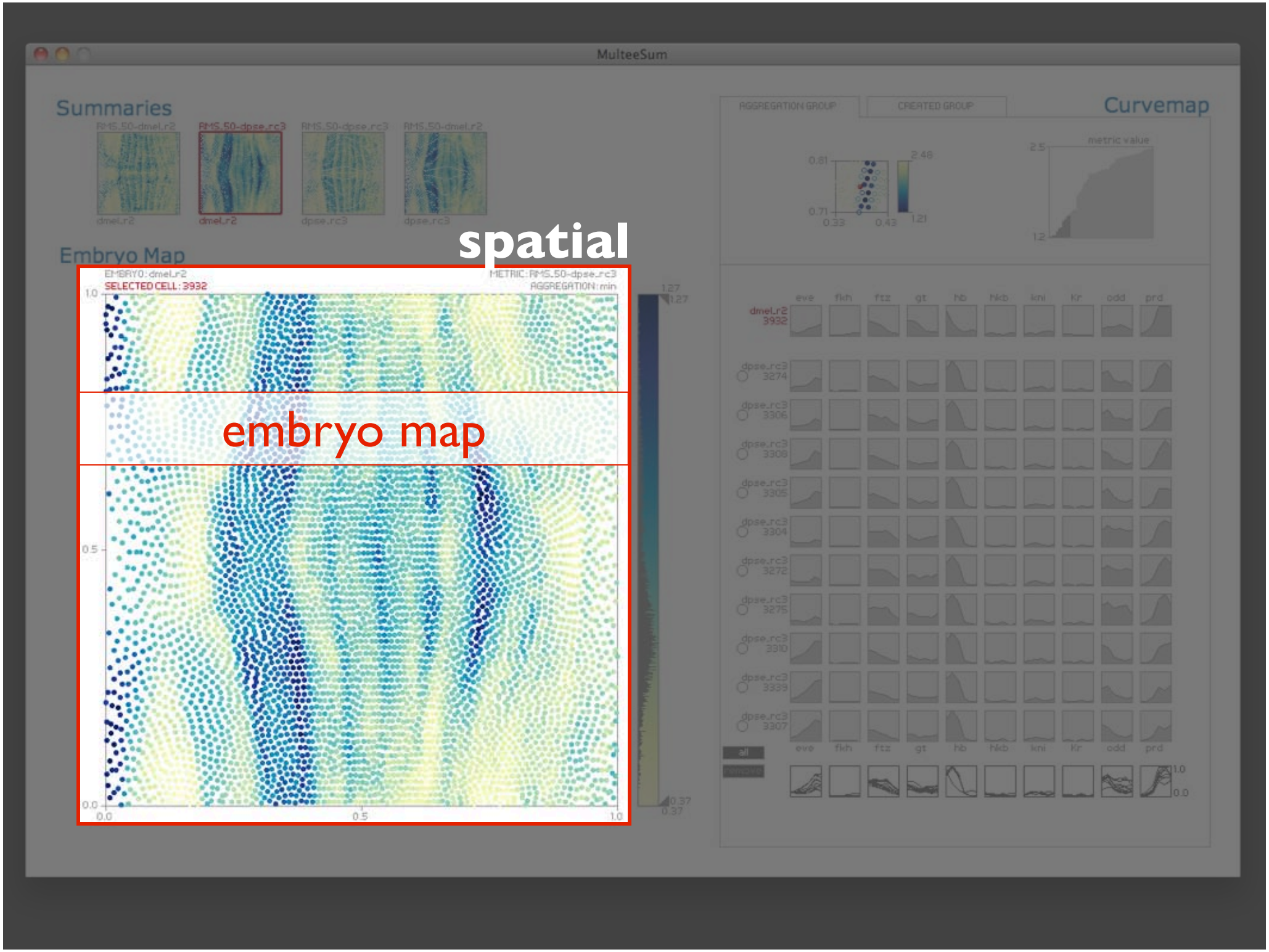
cells

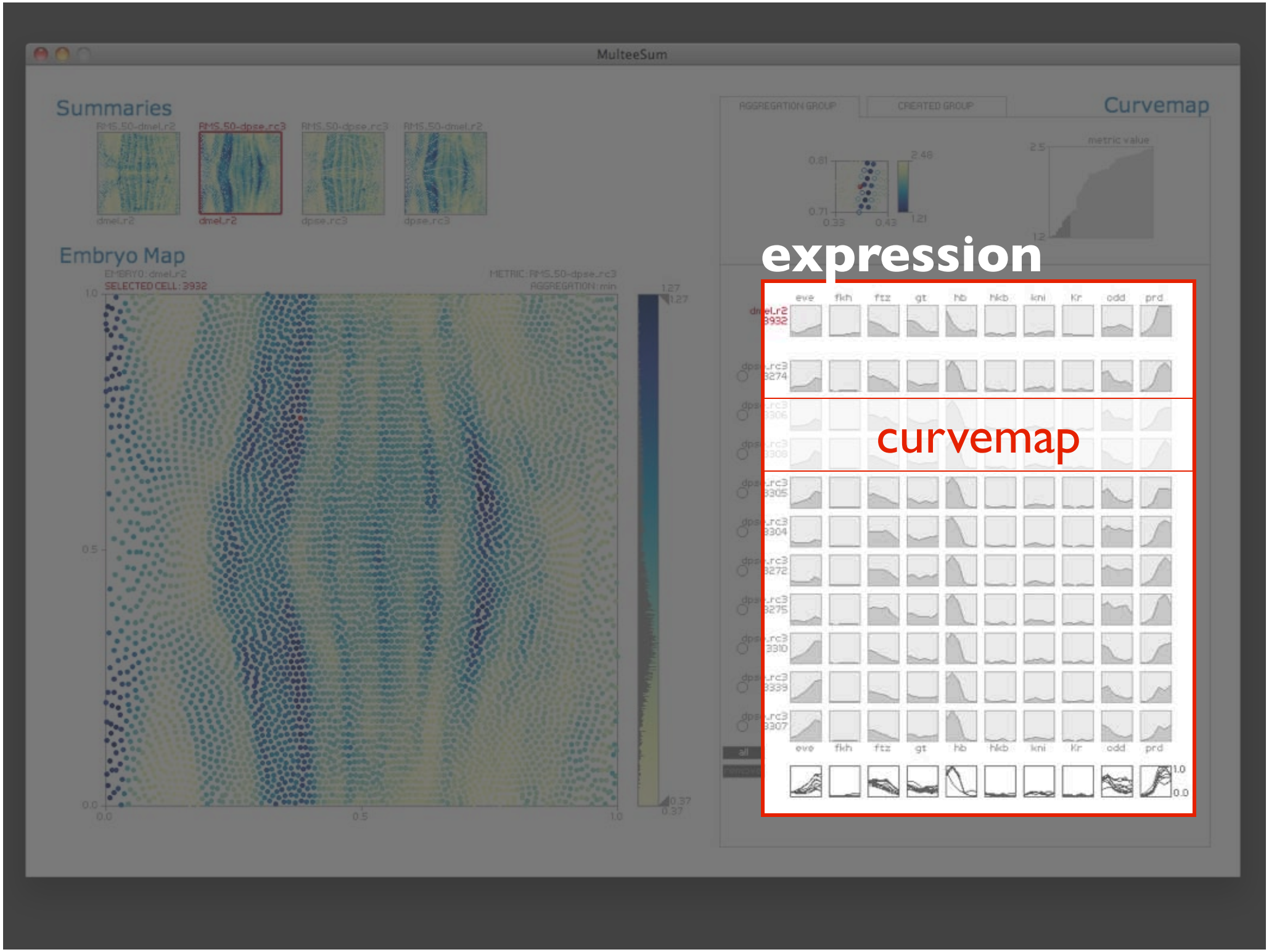
time

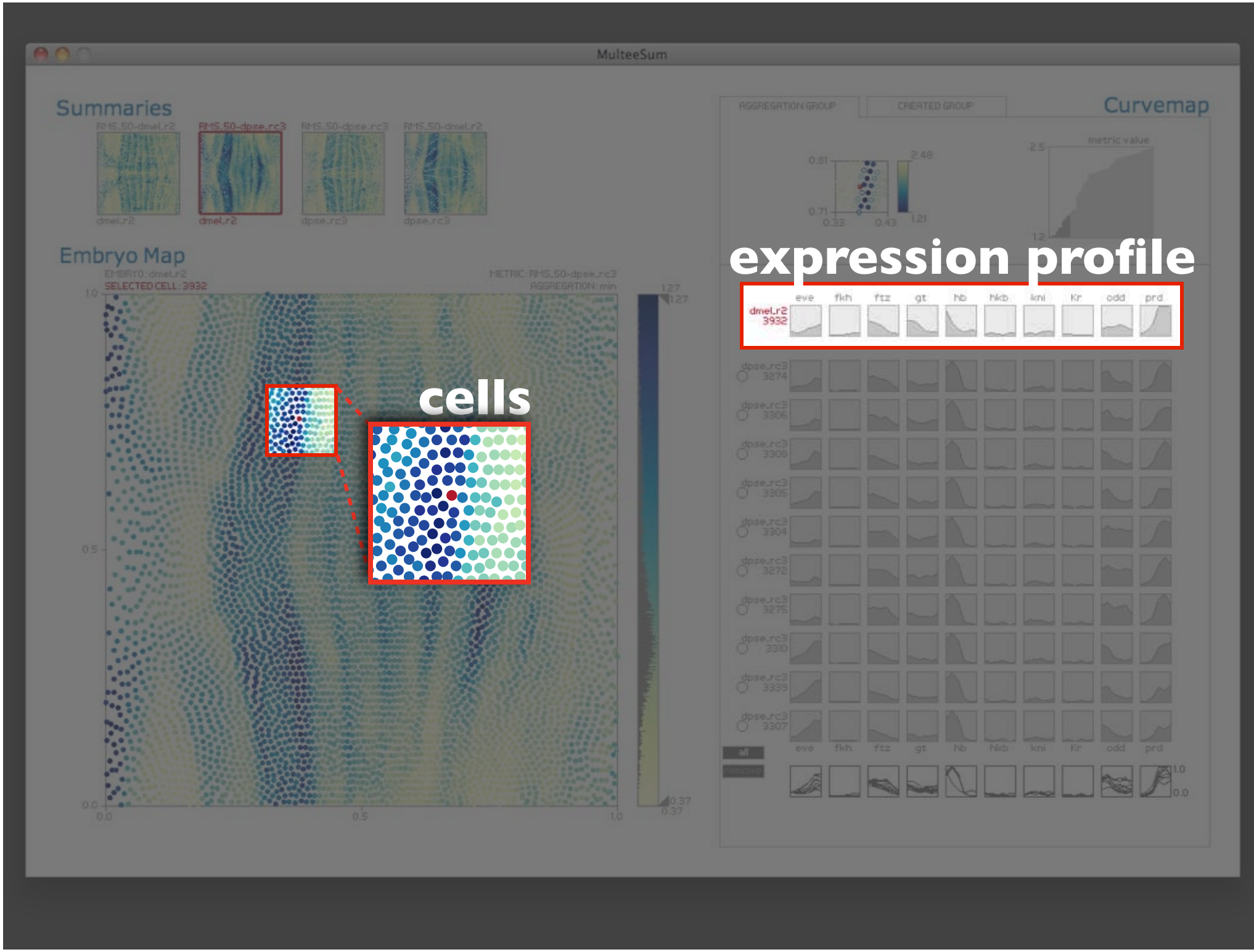
genes

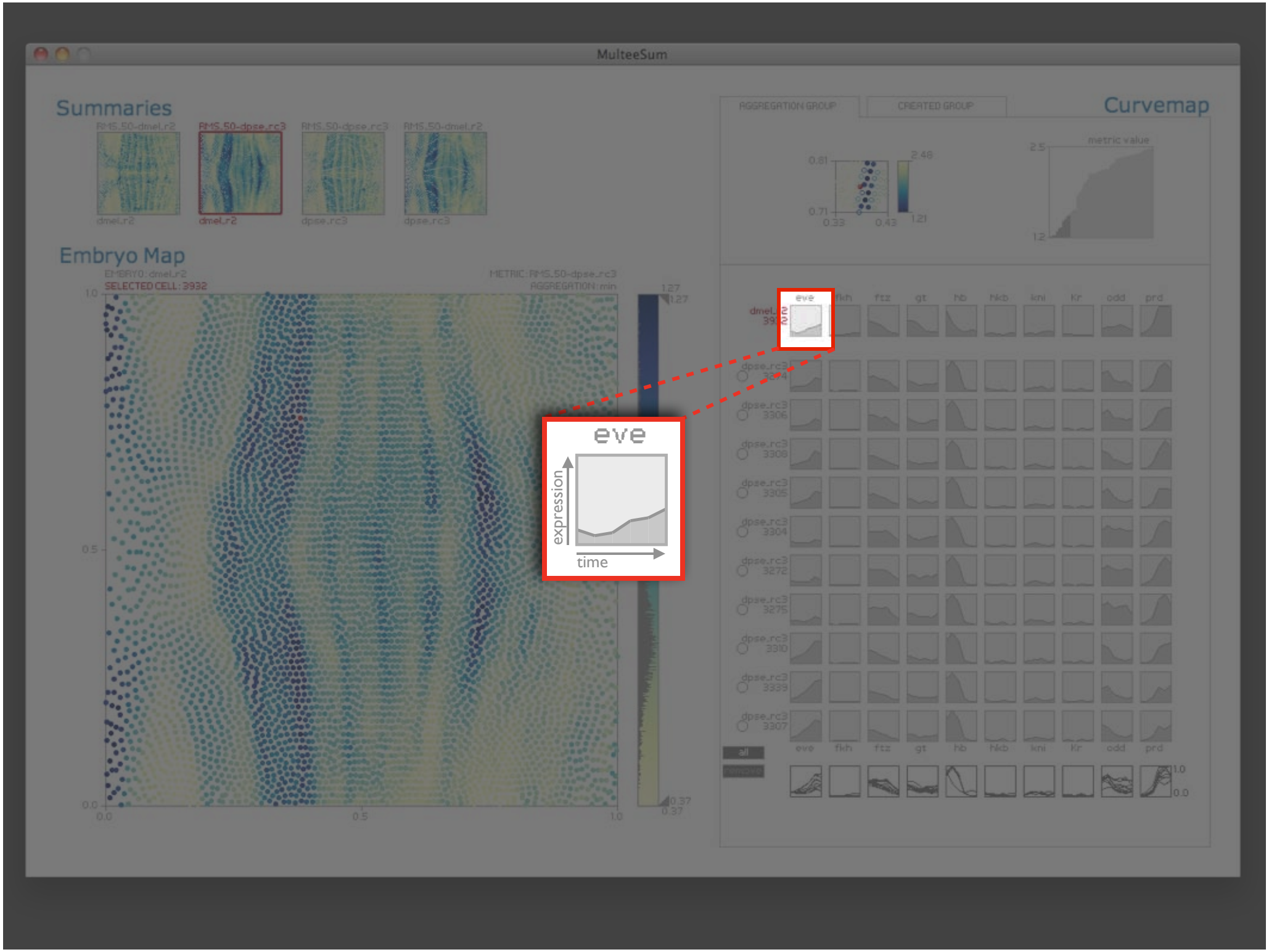
0.2	0.4	1.0	1.0	1.0	0.8
1.0	0.0	0.0	0.0	1.0	1.0
0.7	0.8	1.0	1.0	0.8	0.6
1.0	0.0	0.2	0.5	1.0	1.0
0.5	0.8	0.5	0.3	0.5	0.8
0.7	0.5	0.8	0.7	1.0	1.0
1.0	0.3	0.4	1.0	1.0	1.0

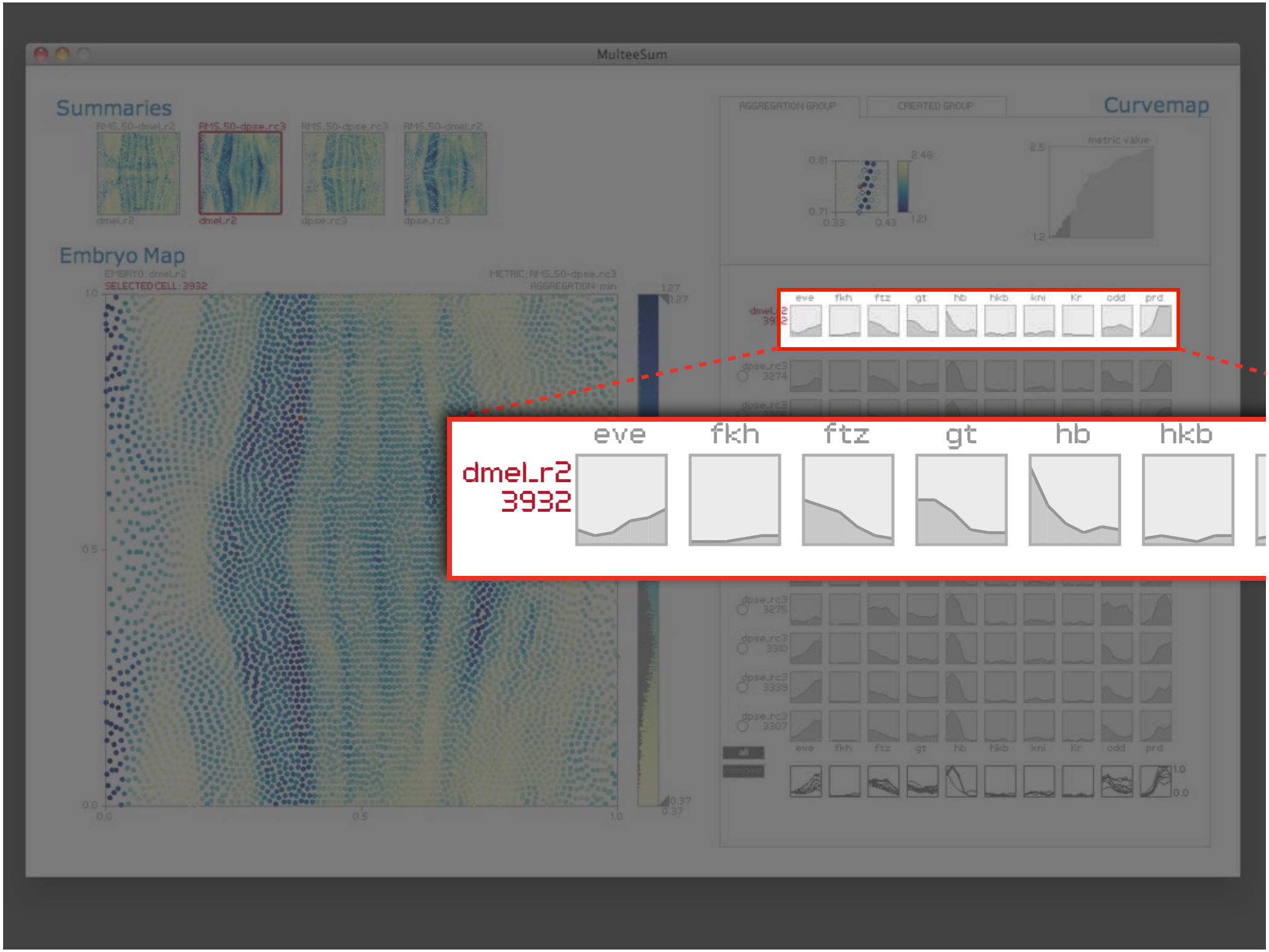

MulteeSum
Summaries
RMS_50-dmel_r2
RMS_50-dpse_rc3
RMS_50-dpse_rc3
RMS_50-dmel_r2
dmel_r2
dmel_r2
dpse_rc3
dpse_rc3
Embryo Map
EMBRYO: dmel_r2
SELECTED CELL: 3932
METRIC: RMS_50-dpse_rc3
AGGREGATION: min
1.0
0.5
0.0
0.0
0.5
1.0
1.27
1.27
0.37
0.37
AGGREGATION GROUP
CREATED GROUP
Curvemap
0.81
0.71
0.33
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
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
all
removed
1.0
0.0

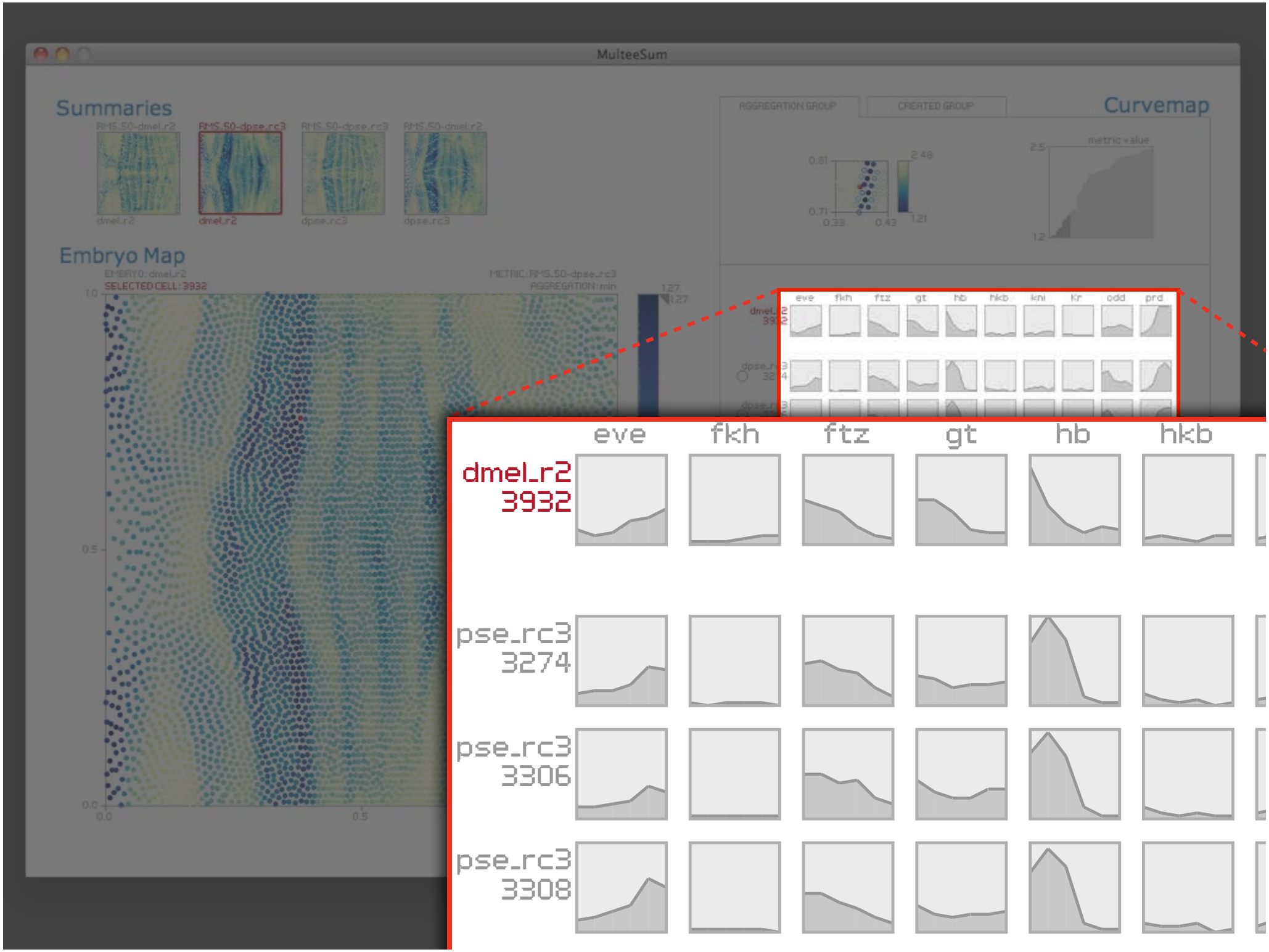

MulteeSum
Summaries
RMS.50-dmel.r2
RMS.50-dpse.rc3
RMS.50-dpse.rc3
RMS.50-dmel.r2
dmel.r2
dmel.r2
dpse.rc3
dpse.rc3
Embryo Map
spatial
EMBRYO: dmel.r2
SELECTED CELL: 3932
METRIC: RMS.50-dpse.rc3
AGGREGATION: min
embryo map
1.0
0.5
0.0
0.0
0.5
1.0
1.27
1.27
0.37
0.37
AGGREGATION GROUP
CREATED GROUP
Curvemap
0.81
2.48
0.71
0.33
0.43
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
all
selected
1.0
0.0

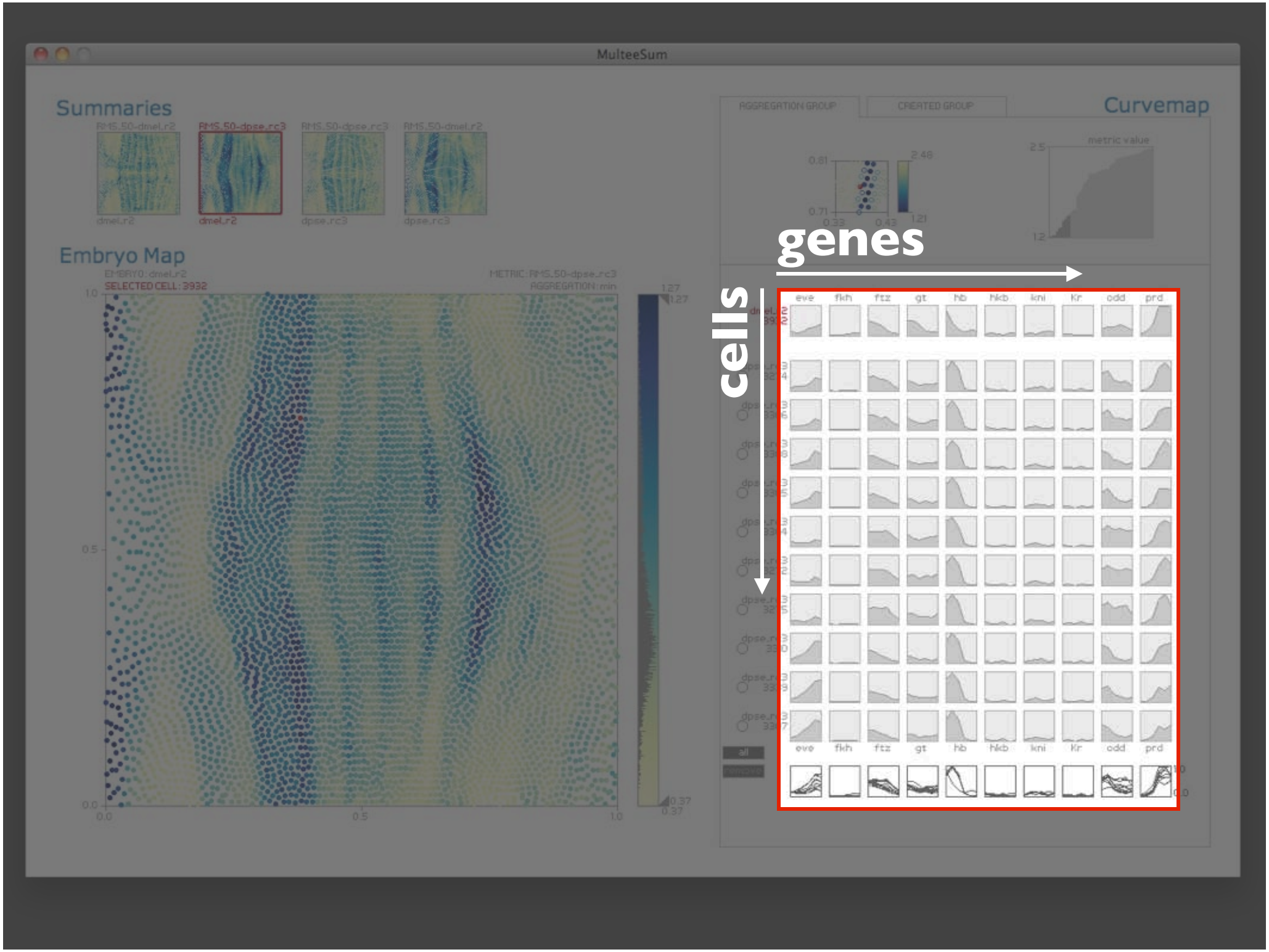

MulteeSum
Summaries
RMS.50-dmel.r2
RMS.50-dpse.rc3
RMS.50-dpse.rc3
RMS.50-dmel.r2
dmel.r2
dmel.r2
dpse.rc3
dpse.rc3
Embryo Map
EMBRYO: dmel.r2
SELECTED CELL: 3932
METRIC: RMS.50-dpse.rc3
AGGREGATION: min
1.27
1.27
0.37
0.37
0.0
0.5
1.0
0.0
0.5
1.0
AGGREGATION GROUP
CREATED GROUP
Curvemap
0.81
0.71
0.33
0.43
2.48
1.21
metric value
2.5
1.2
expression
curvemap
eve
fkh
ftz
gt
hb
hkb
kni
Kr
odd
prd
all
1.0
0.0

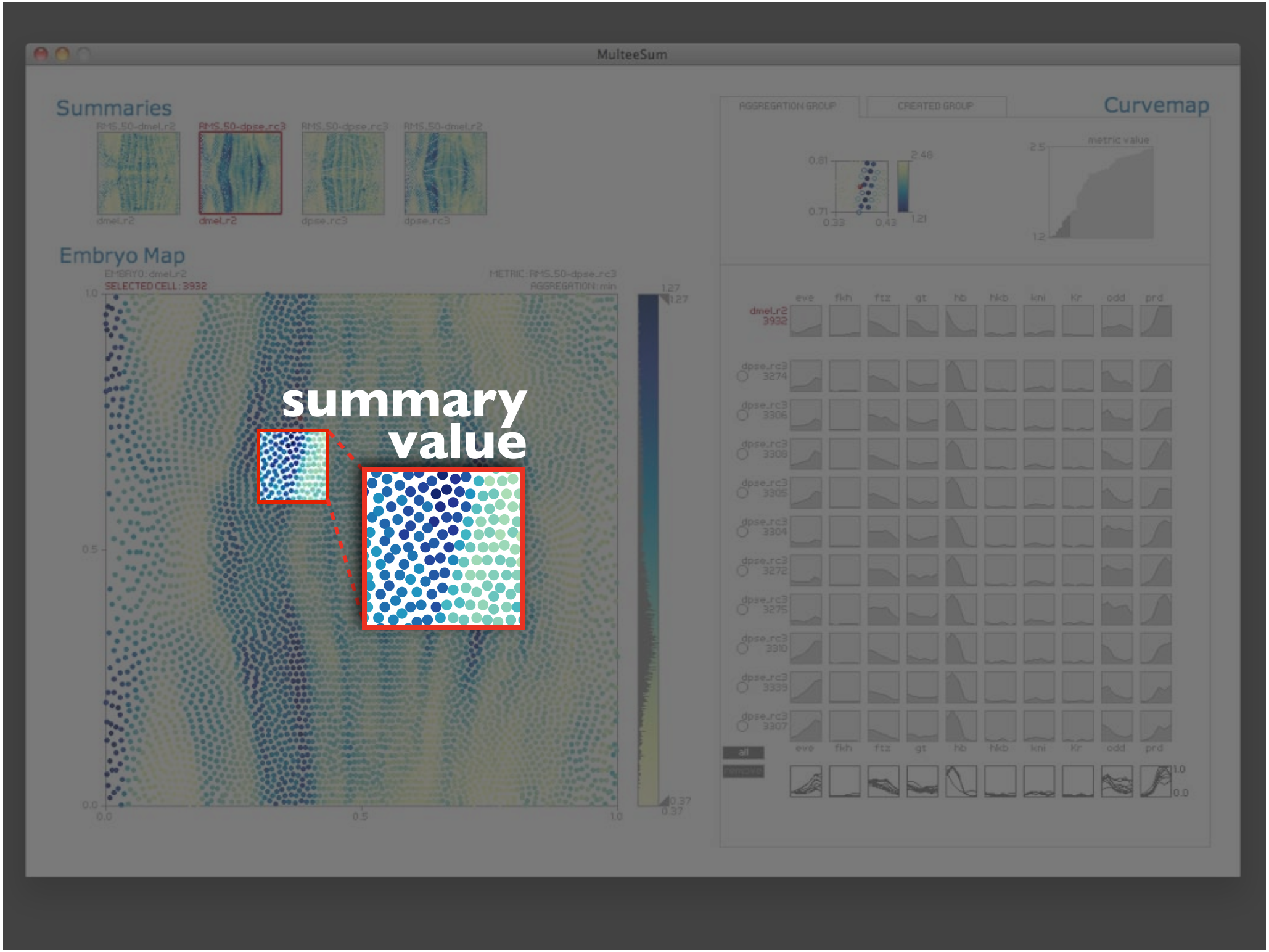

MulteeSum
Summaries
RMS.50-dmel.r2
RMS.50-dpse.rc3
RMS.50-dpse.rc3
RMS.50-dmel.r2
dmel.r2
dmel.r2
dpse.rc3
dpse.rc3
Embryo Map
EMBRYO: dmel.r2
SELECTED CELL: 3932
METRIC: RMS.50-dpse.rc3
AGGREGATION: min
cells
AGGREGATION GROUP
CREATED GROUP
Curvemap
metric value
expression profile
dmel.r2 3932
eve fkh ftz gt hb hkb kni Kr odd prd
all
removed

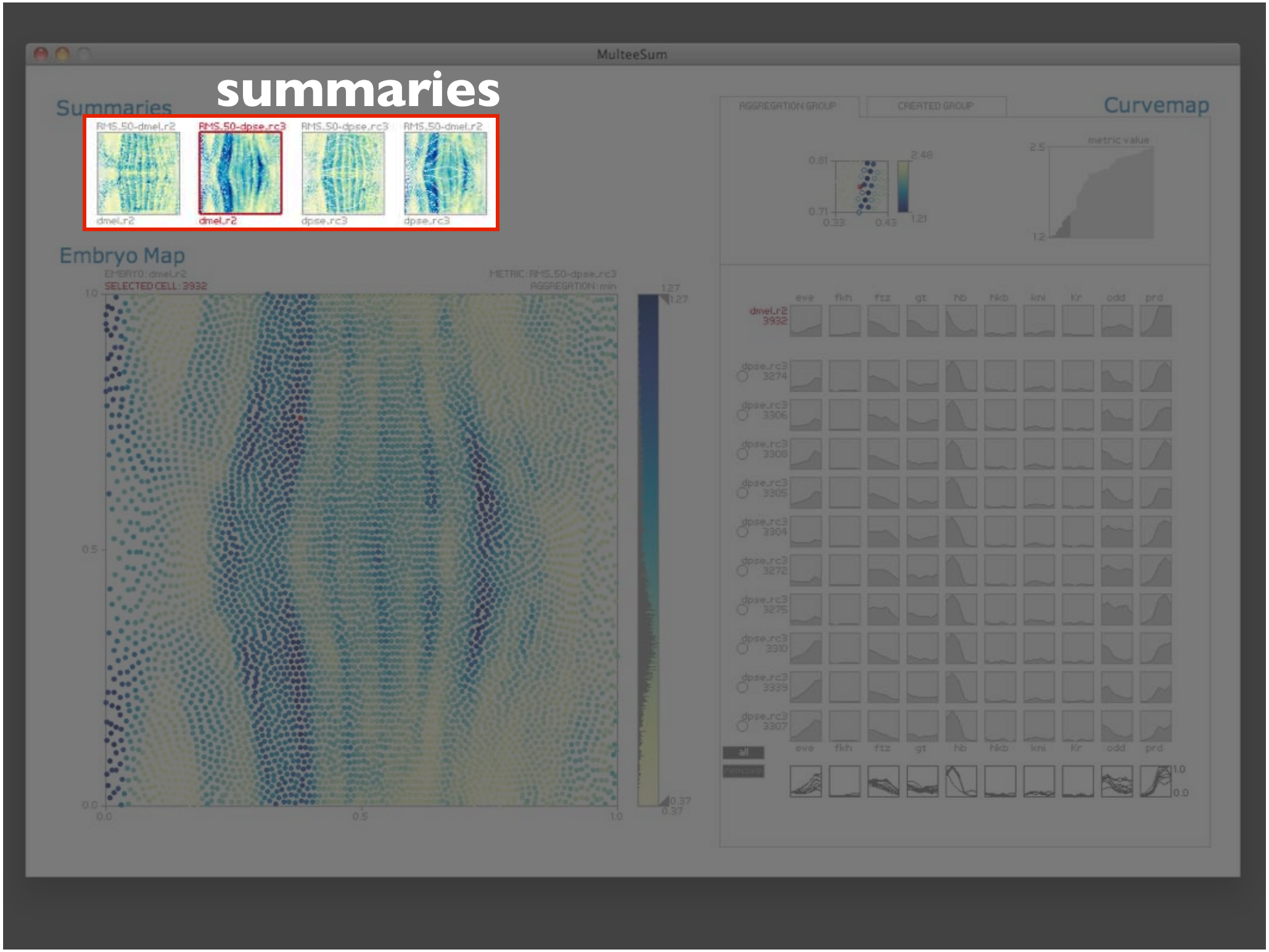

MulteeSum
Summaries
RMS.50-dmel.r2
RMS.50-dpse.rc3
RMS.50-dpse.rc3
RMS.50-dmel.r2
dmel.r2
dmel.r2
dpse.rc3
dpse.rc3
Embryo Map
EMBRYO: dmel.r2
SELECTED CELL: 3932
METRIC: RMS.50-dpse.rc3
AGGREGATION: min
1.27
1.27
0.37
0.37
1.0
0.5
0.0
0.0
0.5
1.0
eve
expression
time
AGGREGATION GROUP
CREATED GROUP
Curvemap
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
dpse.rc3
3306
3308
3305
3304
3272
3275
3310
3339
3307
all
1.0
0.0

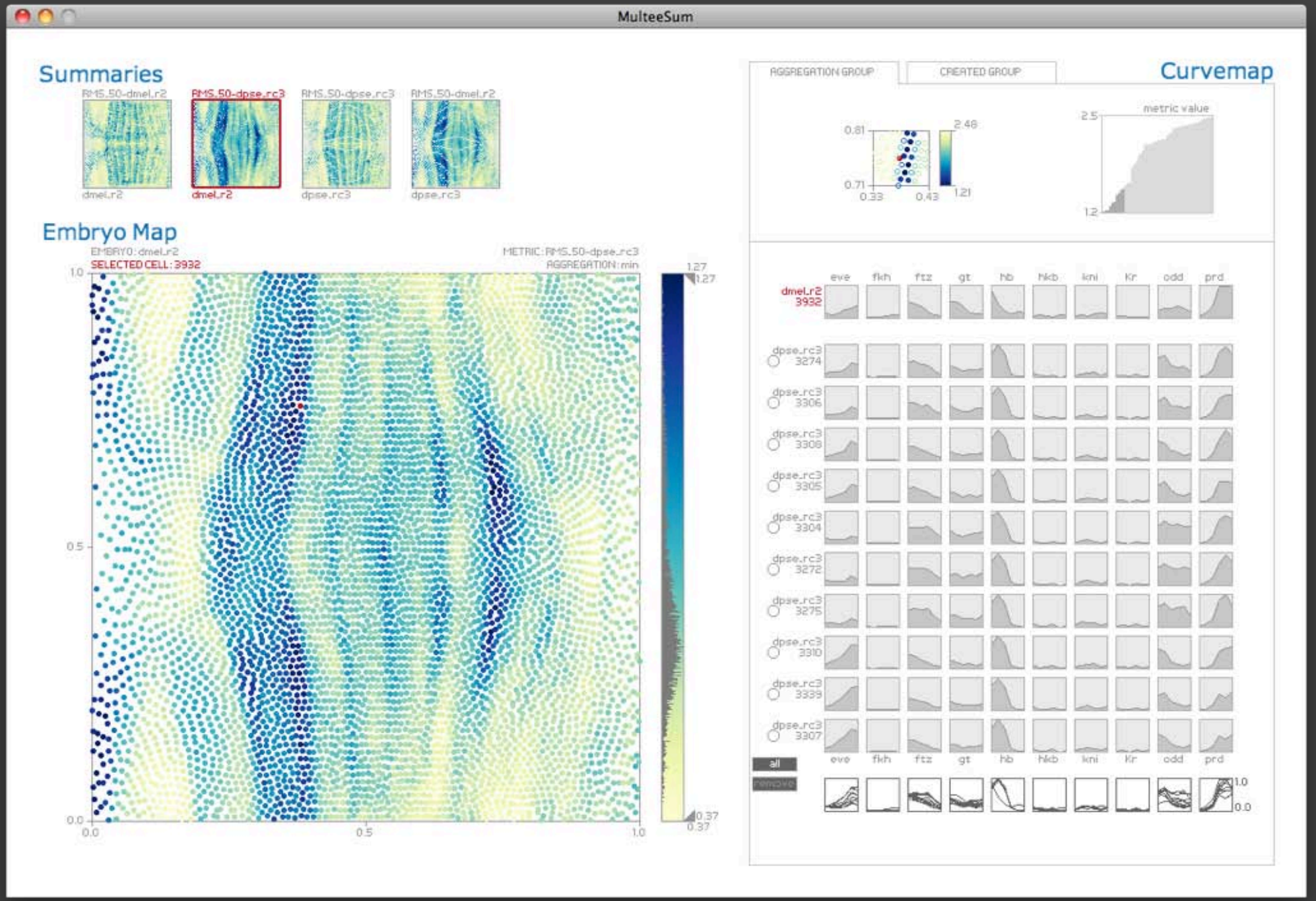

MulteeSum
Summaries
RMS.50-dmel.r2
RMS.50-dpse.rc3
RMS.50-dpse.rc3
RMS.50-dmel.r2
dmel.r2
dmel.r2
dpse.rc3
dpse.rc3
Embryo Map
EMBRYO: dmel.r2
SELECTED CELL: 3932
METRIC: RMS.50-dpse.rc3
AGGREGATION: min
1.27
1.27
0.37
0.37
1.0
0.5
0.0
0.0
0.5
1.0
AGGREGATION GROUP
CREATED GROUP
Curvemap
metric value
2.5
1.2
0.81
0.71
0.33
0.43
2.48
1.21
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
removed
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
1.27
1.27
1.0
0.5
0.0
0.0
0.5
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
eve
fkh
ftz
gt
hb
hkb
dmel_r2
3932
pse_rc3
3274
pse_rc3
3306
pse_rc3
3308

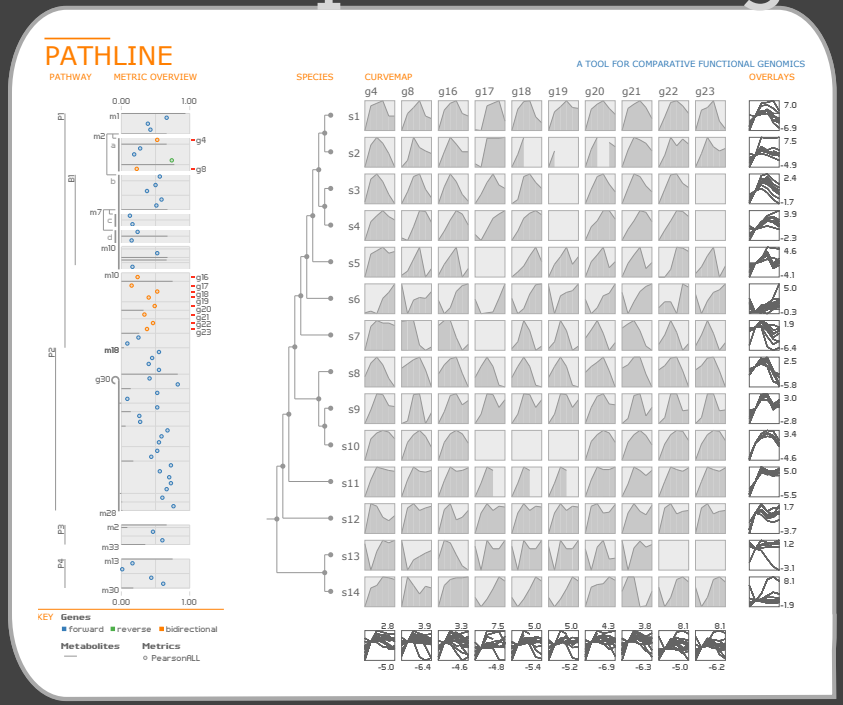

MulteeSum
Summaries
RMS.50-dmel.r2
RMS.50-dpse.rc3
RMS.50-dpse.rc3
RMS.50-dmel.r2
dmel.r2
dmel.r2
dpse.rc3
dpse.rc3
Embryo Map
EMBRYO: dmel.r2
SELECTED CELL: 3932
METRIC: RMS.50-dpse.rc3
AGGREGATION: min
1.27
1.27
0.37
0.37
1.0
0.5
0.0
0.0
0.5
1.0
AGGREGATION GROUP
CREATED GROUP
Curvemap
metric value
2.5
1.2
0.81
0.71
0.33
0.43
2.48
1.21
genes
cells
eve
fkh
ftz
gt
hb
hkb
kni
Kr
odd
prd
all
selected


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
1.0
1.27
1.27
0.37
0.37
summary
value
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
dpse.rc3
3305
dpse.rc3
3304
dpse.rc3
3272
dpse.rc3
3275
dpse.rc3
3310
dpse.rc3
3339
dpse.rc3
3307
all
1.0
0.0

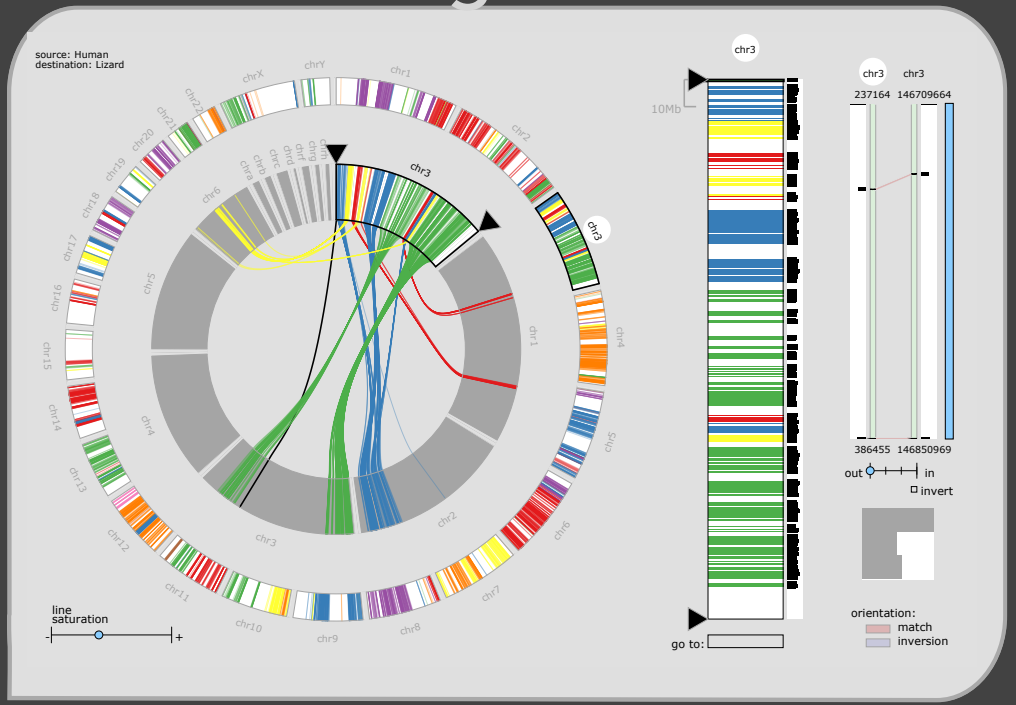

MulteeSum
summaries
Summaries
RMS.50-dmel.r2
RMS.50-dpse.rc3
RMS.50-dpse.rc3
RMS.50-dmel.r2
dmel.r2
dmel.r2
dpse.rc3
dpse.rc3
Embryo Map
EMBRYO: dmel.r2
SELECTED CELL: 3932
METRIC: RMS.50-dpse.rc3
AGGREGATION: min
1.0
0.5
0.0
0.0
0.5
1.0
1.27
1.27
0.37
0.37
AGGREGATION GROUP
CREATED GROUP
Curvemap
0.81
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
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
all
removed
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
1.0
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
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
all
removed
1.0
0.0

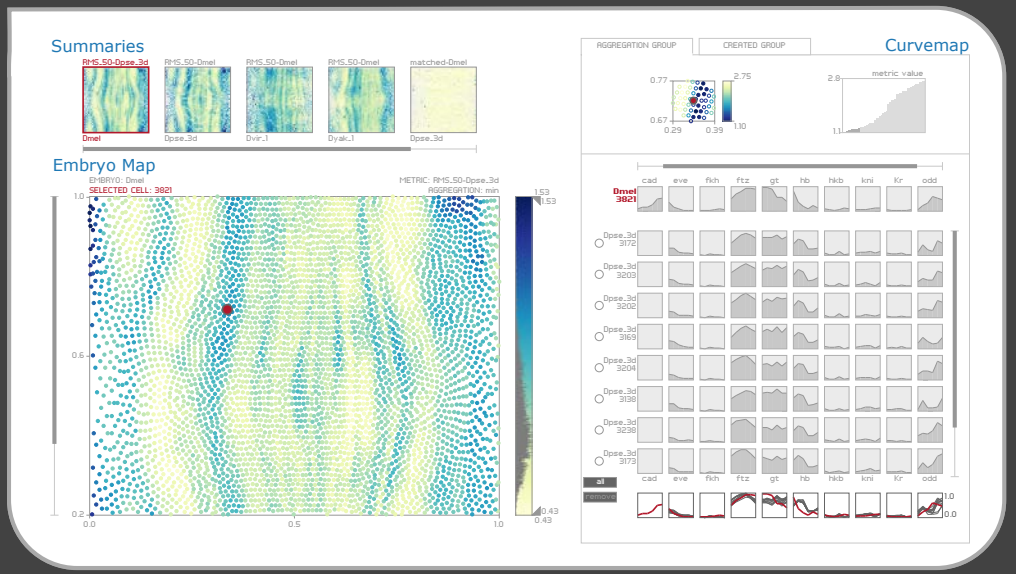
pathline.org



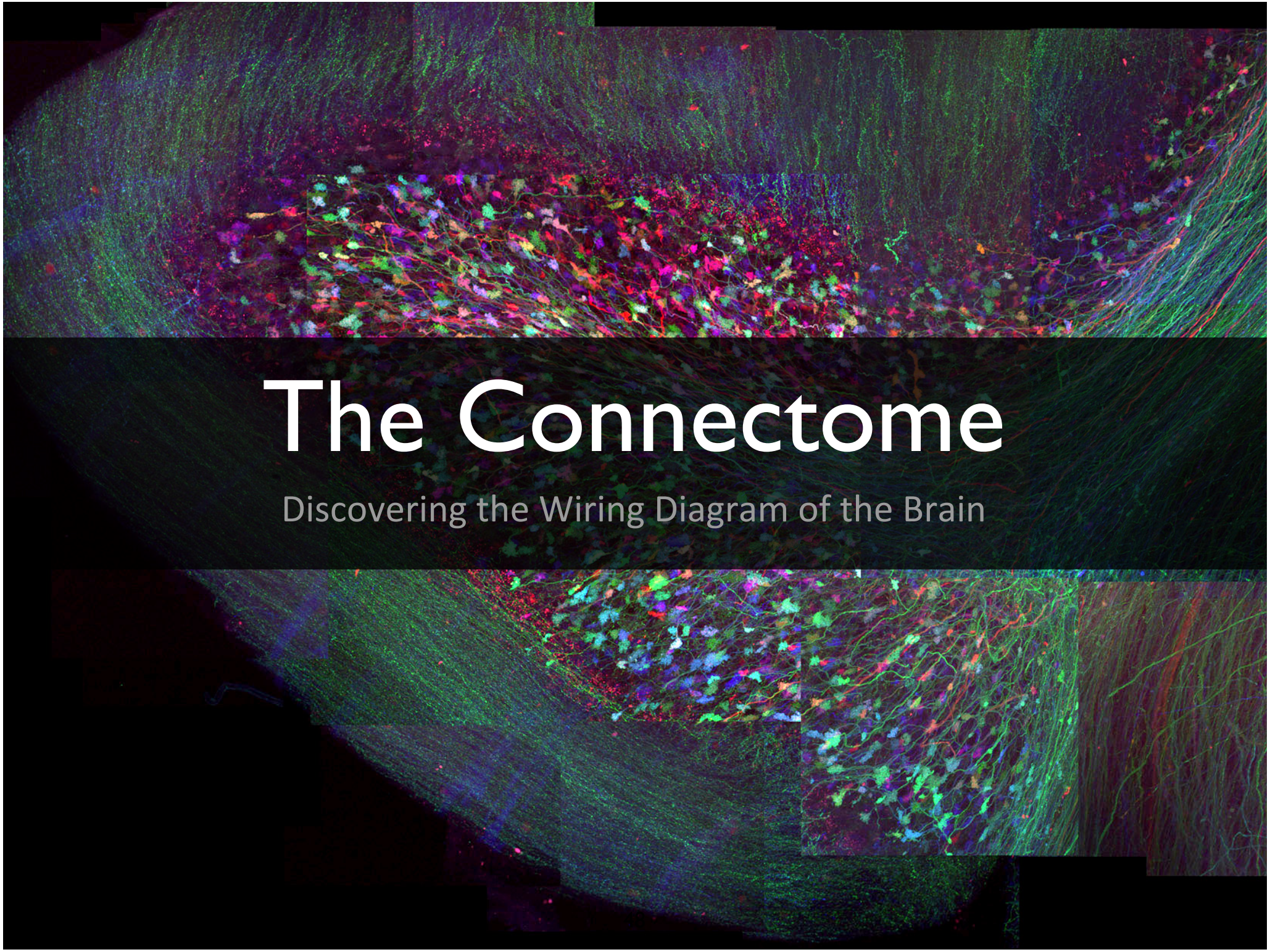
mizbee.org



characterization validation process



multeesum.org



The Connectome

Discovering the Wiring Diagram of the Brain

Collaborators

Harvard Center for Brain Science

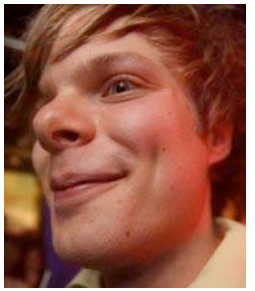
- Prof. Jeff Lichtman & Prof. Clay Reid

KAUST

- Prof. Markus Hadwiger, Dr. Johanna Beyer

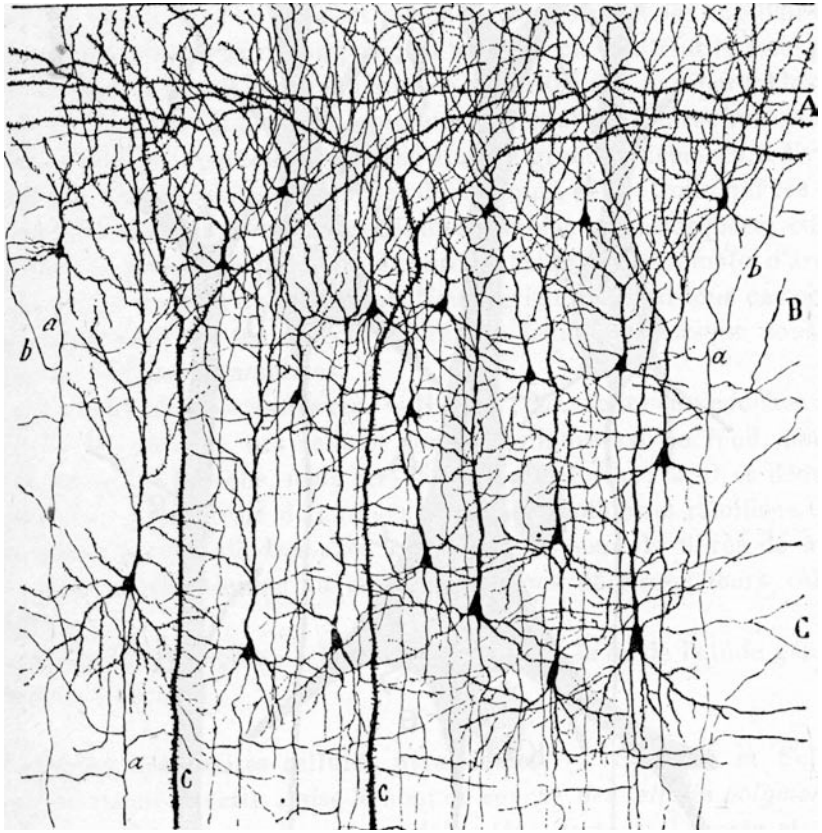
SEAS

- Dr. Won-Ki Jeong
  Dr. Verena Kaynig-Fittkau
  Amelio Vazquez
  Mike Roberts

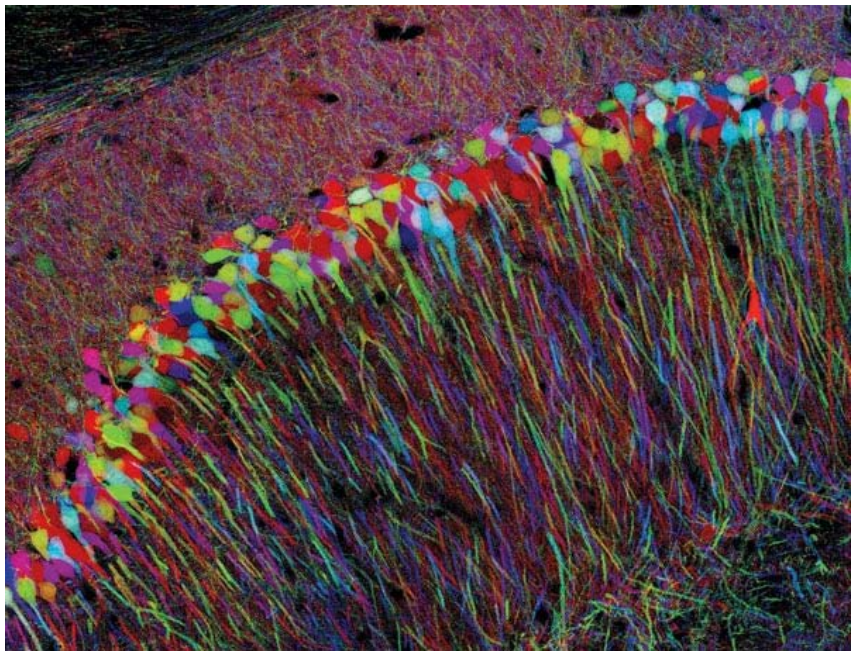


The Scientific Challenge

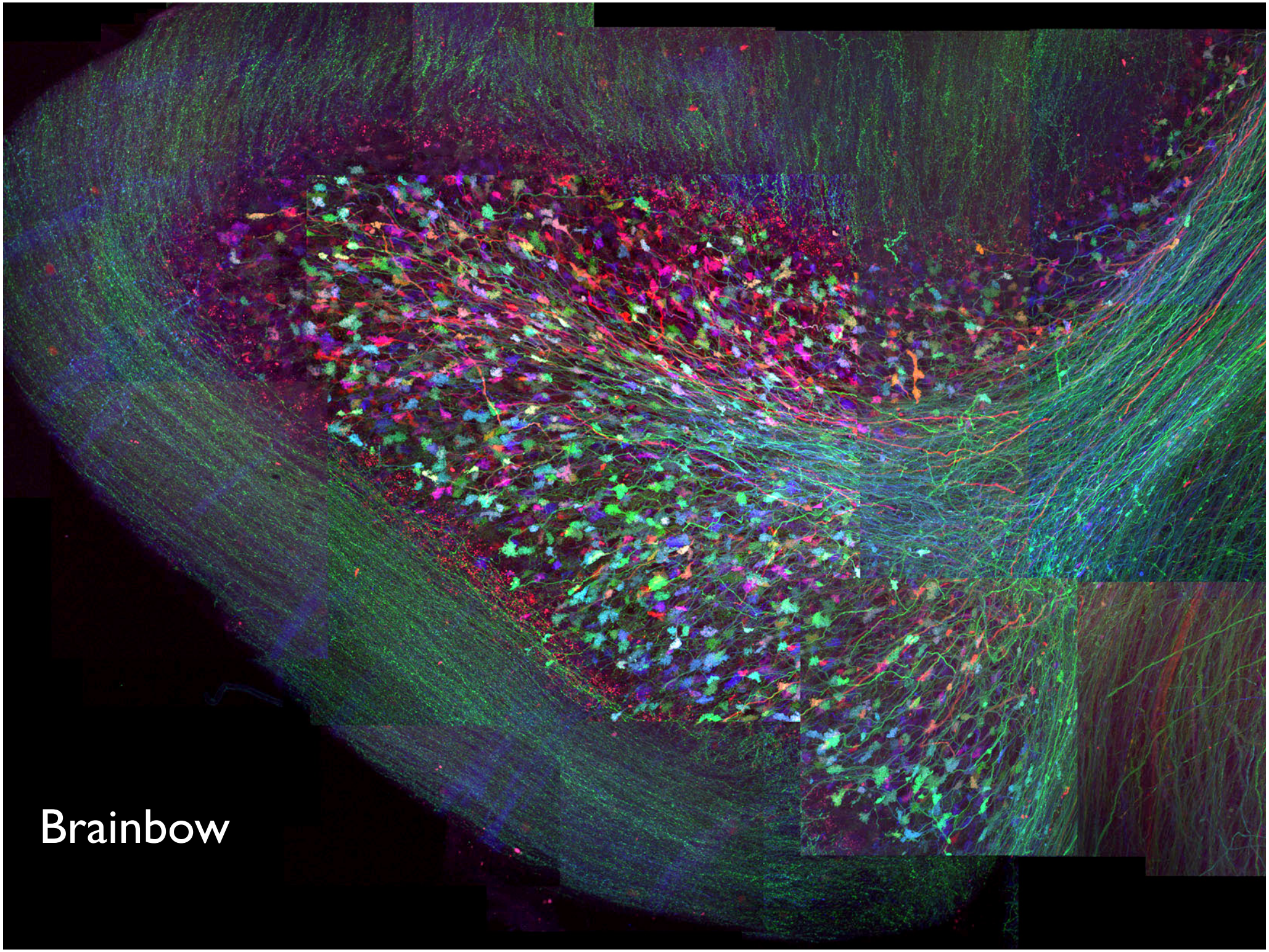
How is the mammalian brain wired?



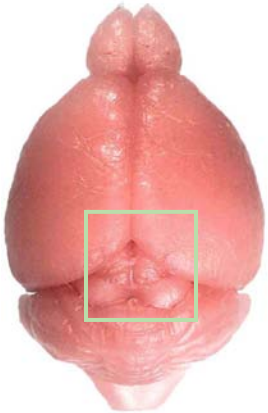
Ramón y Cajal, 1905

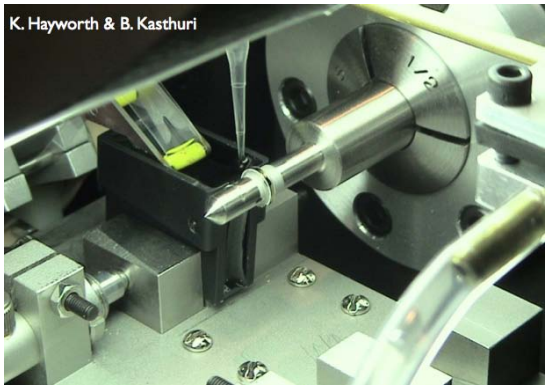


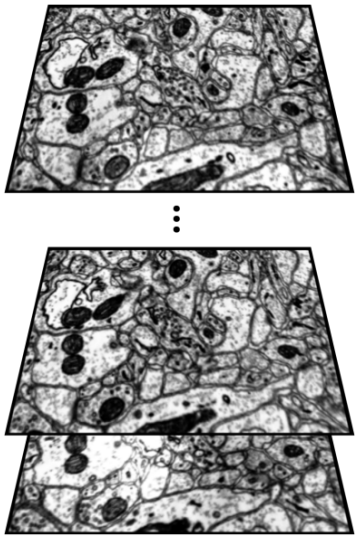
Harvard Center for Brain Science

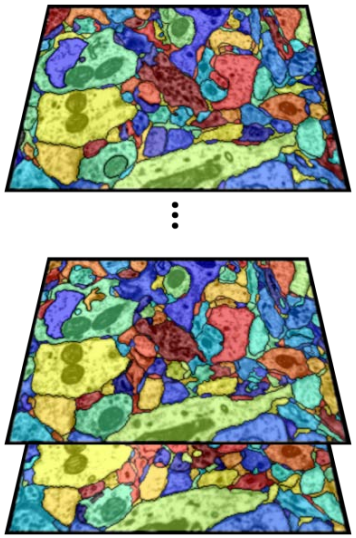

Brainbow

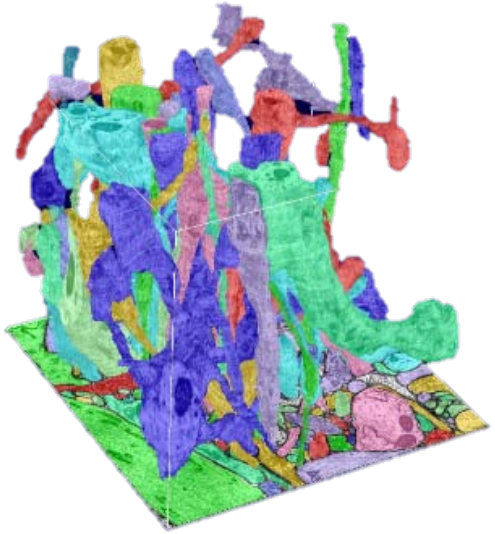
Connectome Workflow



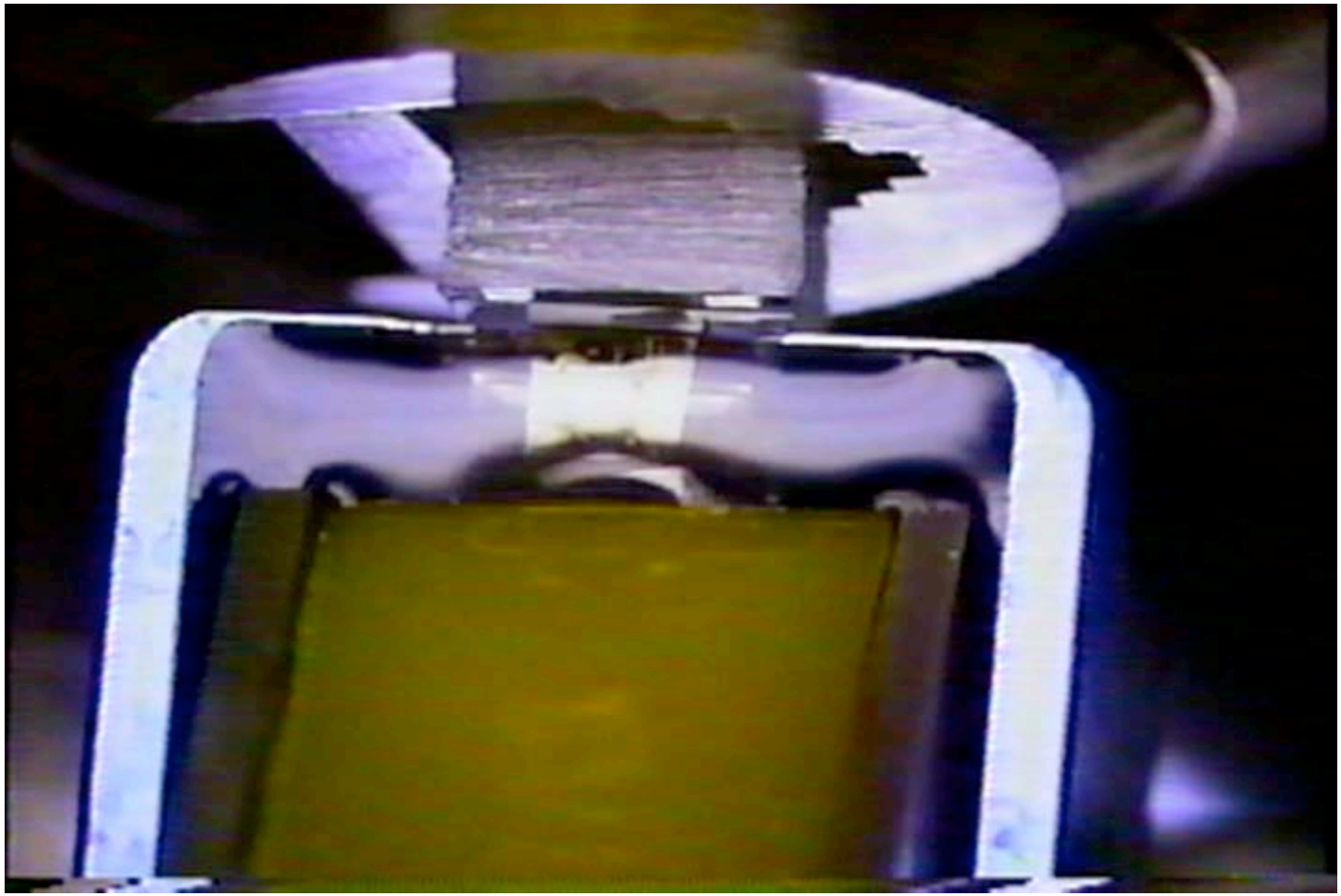






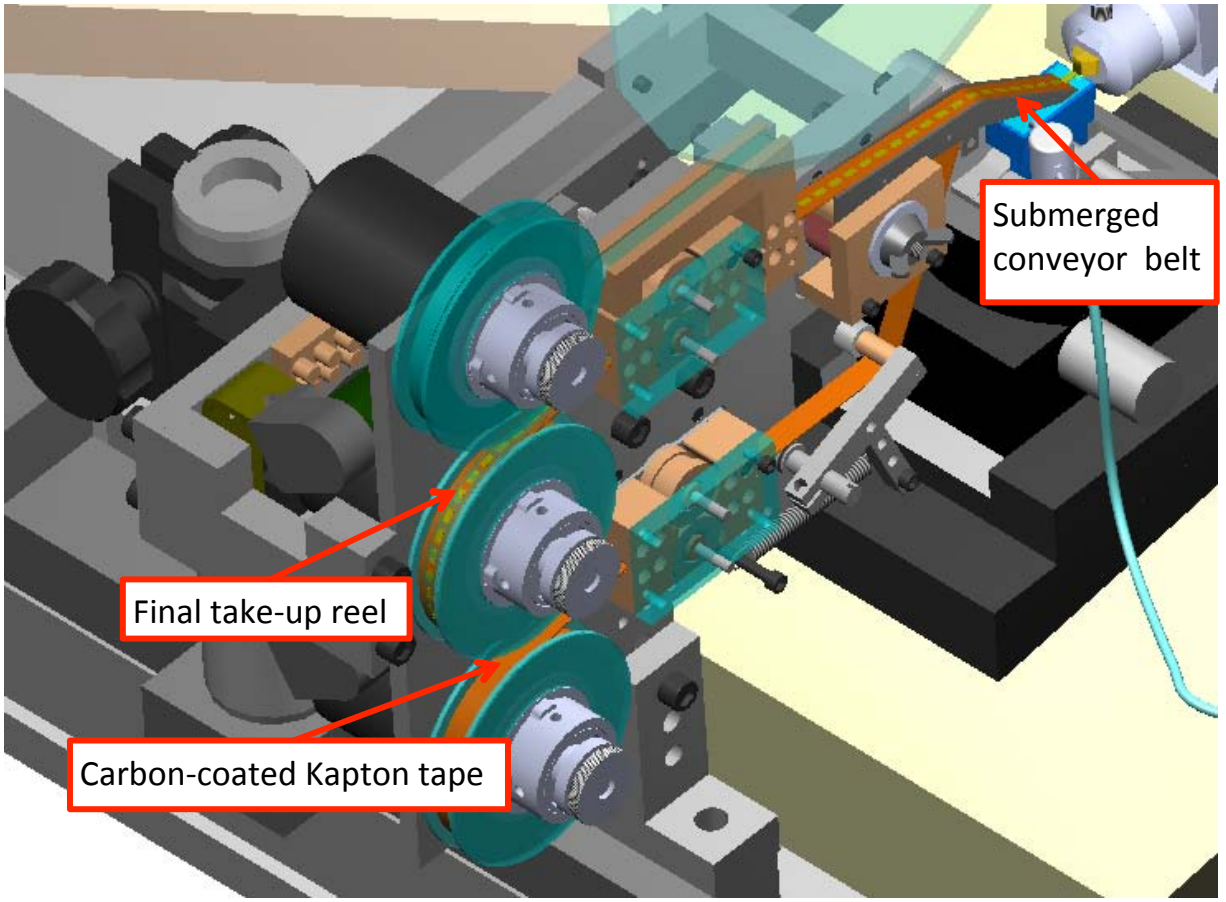


Microtome Cutting

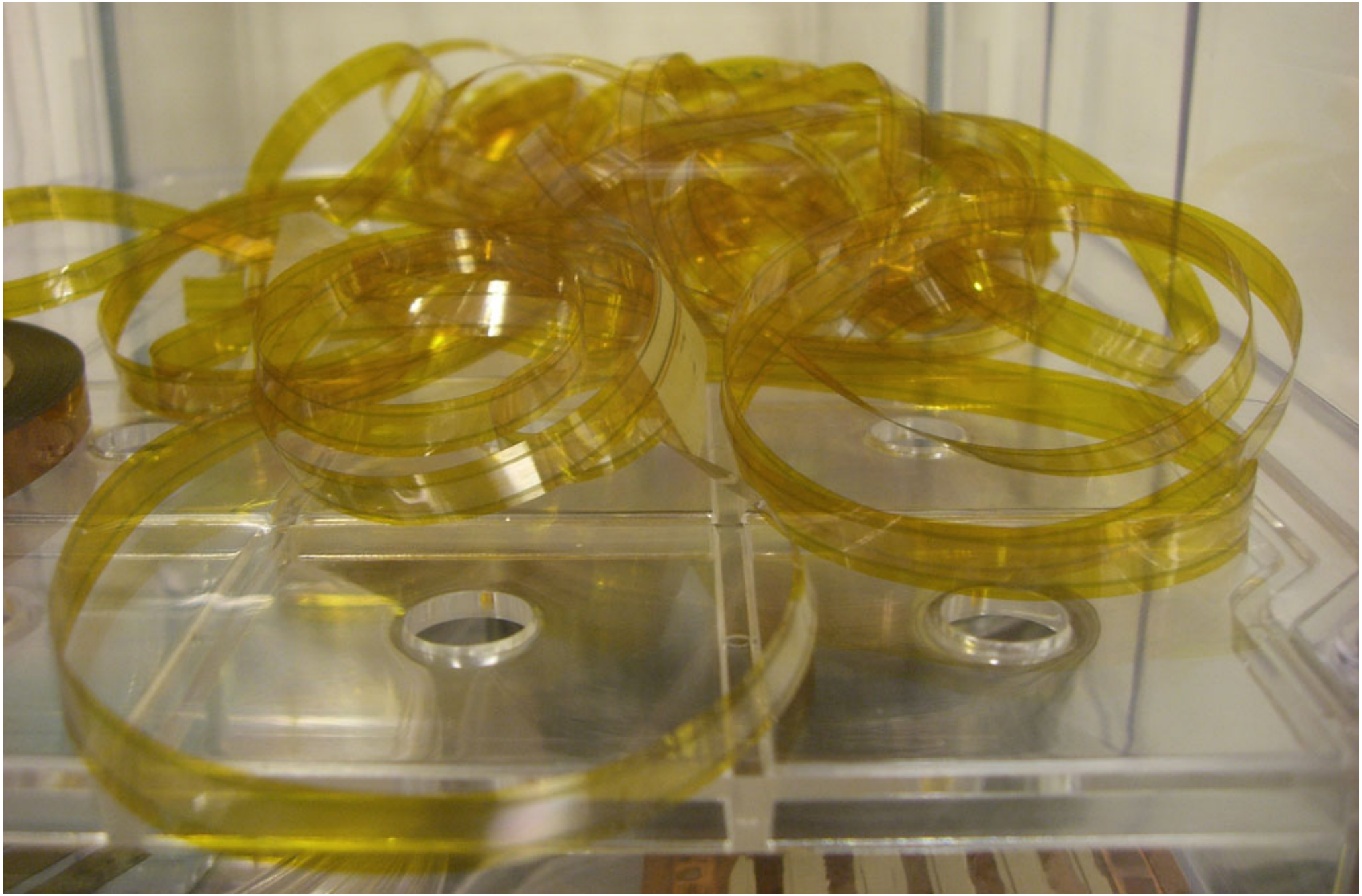


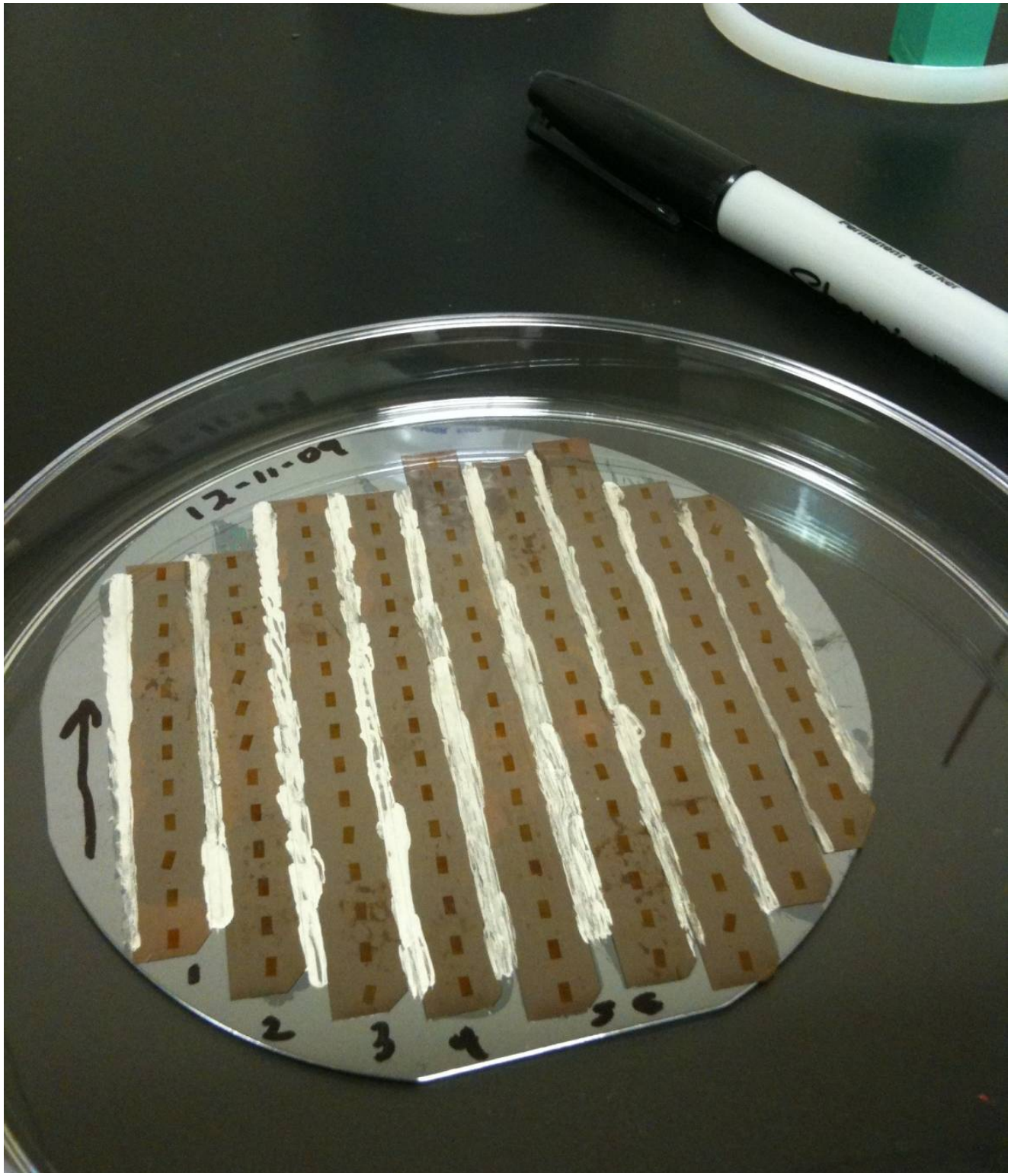
Ken Hayworth

Automatic Tape Collection Mechanism

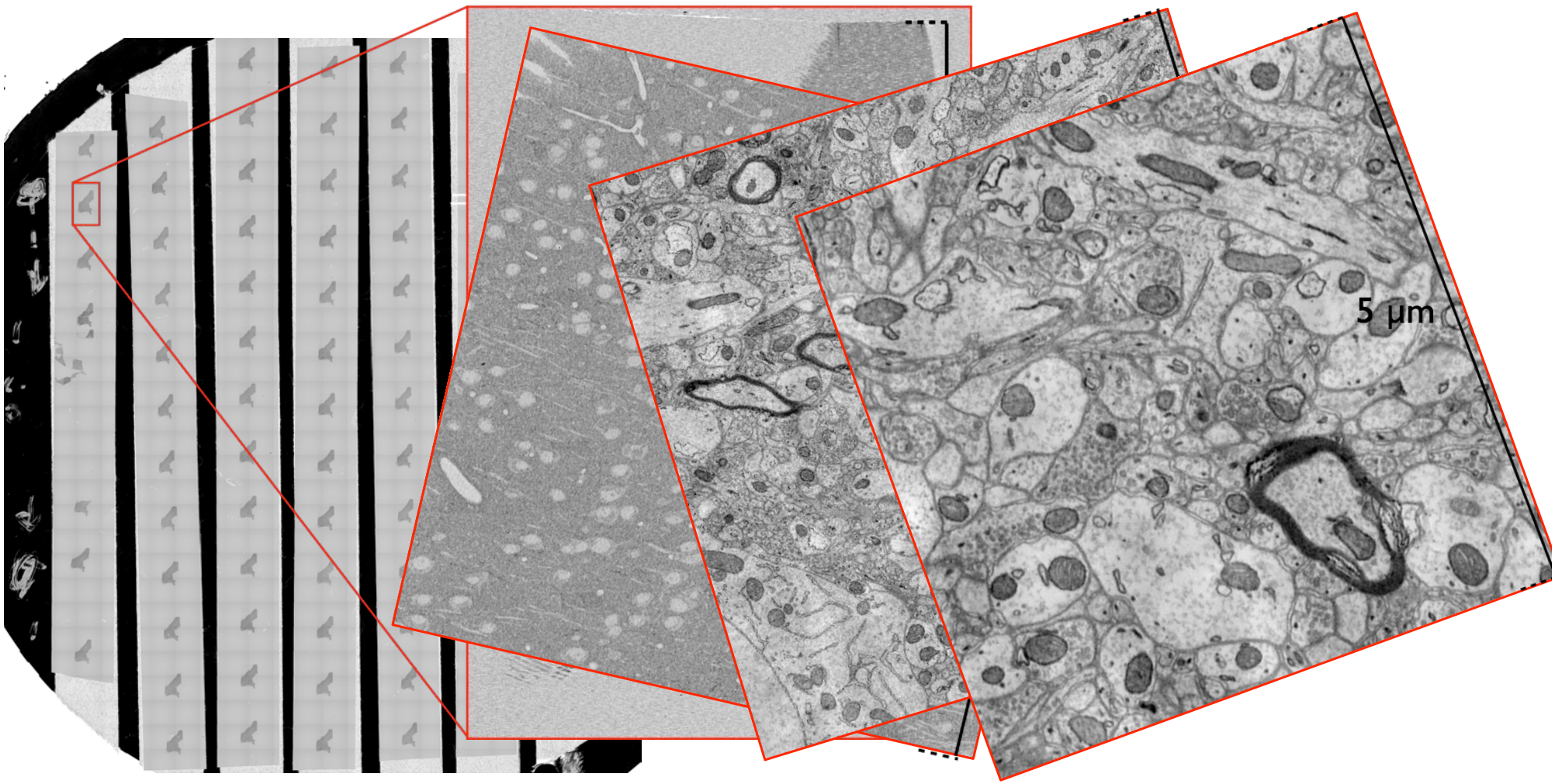


Ken Hayworth

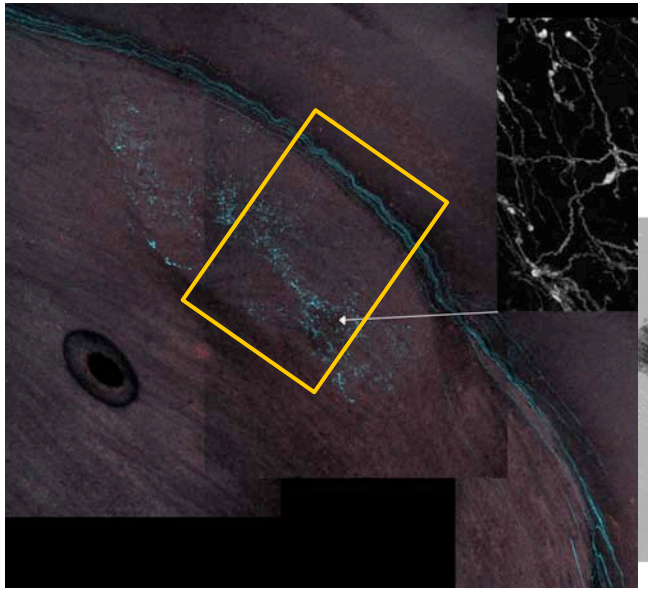


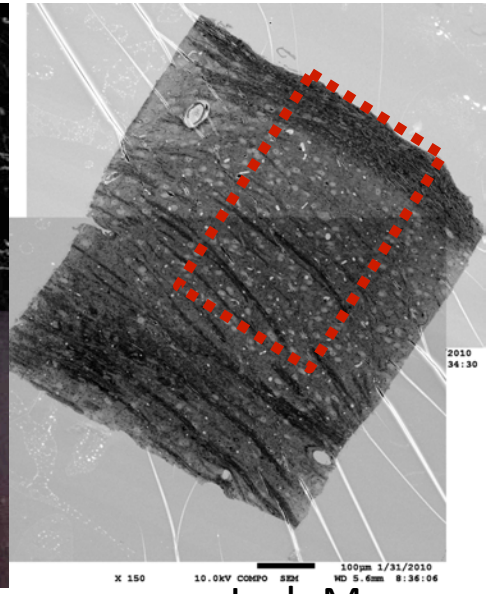

12-11-09
1
2
3
4
5
6

Electron Microscopy



The Data Challenge





Josh Morgan

- Pixel resolution: 3-5 nm
- Slice thickness: 30-50 nm
- 1 mm^3: 200k x 200k pixels x 20k slices

  40 Gpixels x 20k = 800 TB

CS Challenges

- Stitching and alignment of overlapping tiles
- 3D registration of sections
- Visualization
- 3D reconstruction
- Feature detection (e.g., synapses)
- Network analysis

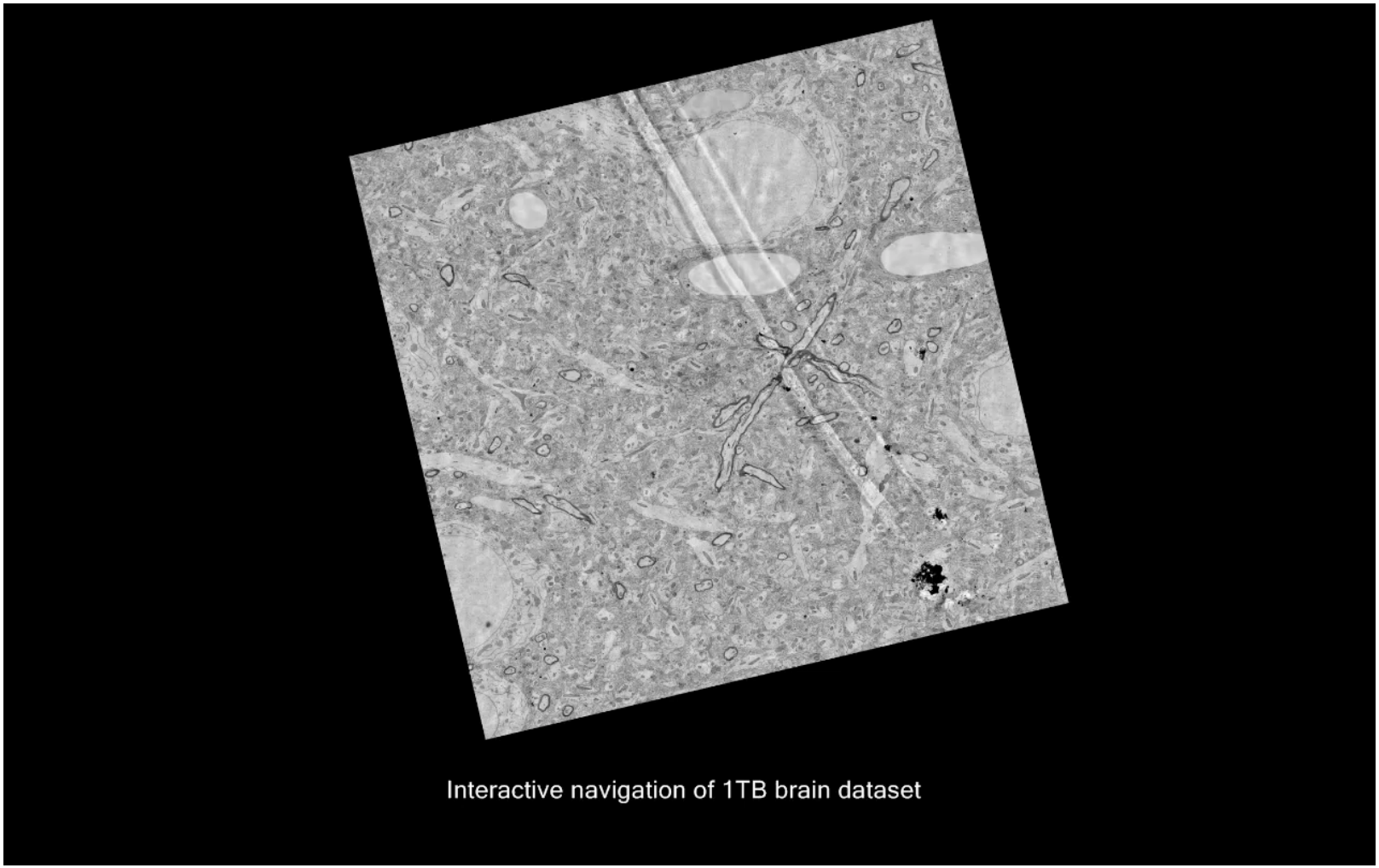
CS Challenges

- Stitching and alignment of overlapping tiles
- 3D registration of sections
- **Visualization**
- **3D reconstruction**

This Talk

- Feature detection (e.g., synapses)
- Network analysis

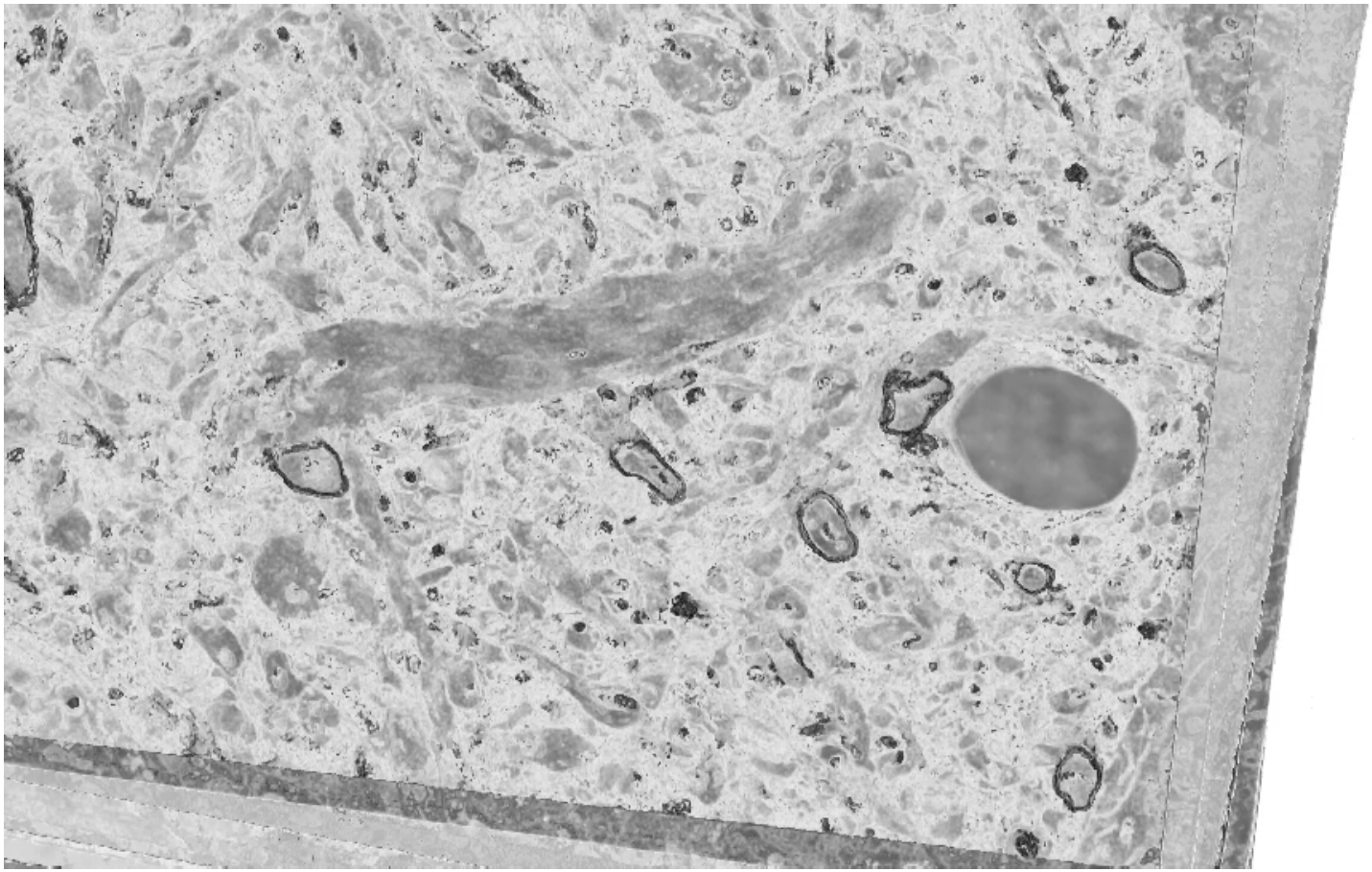
Neurotrace



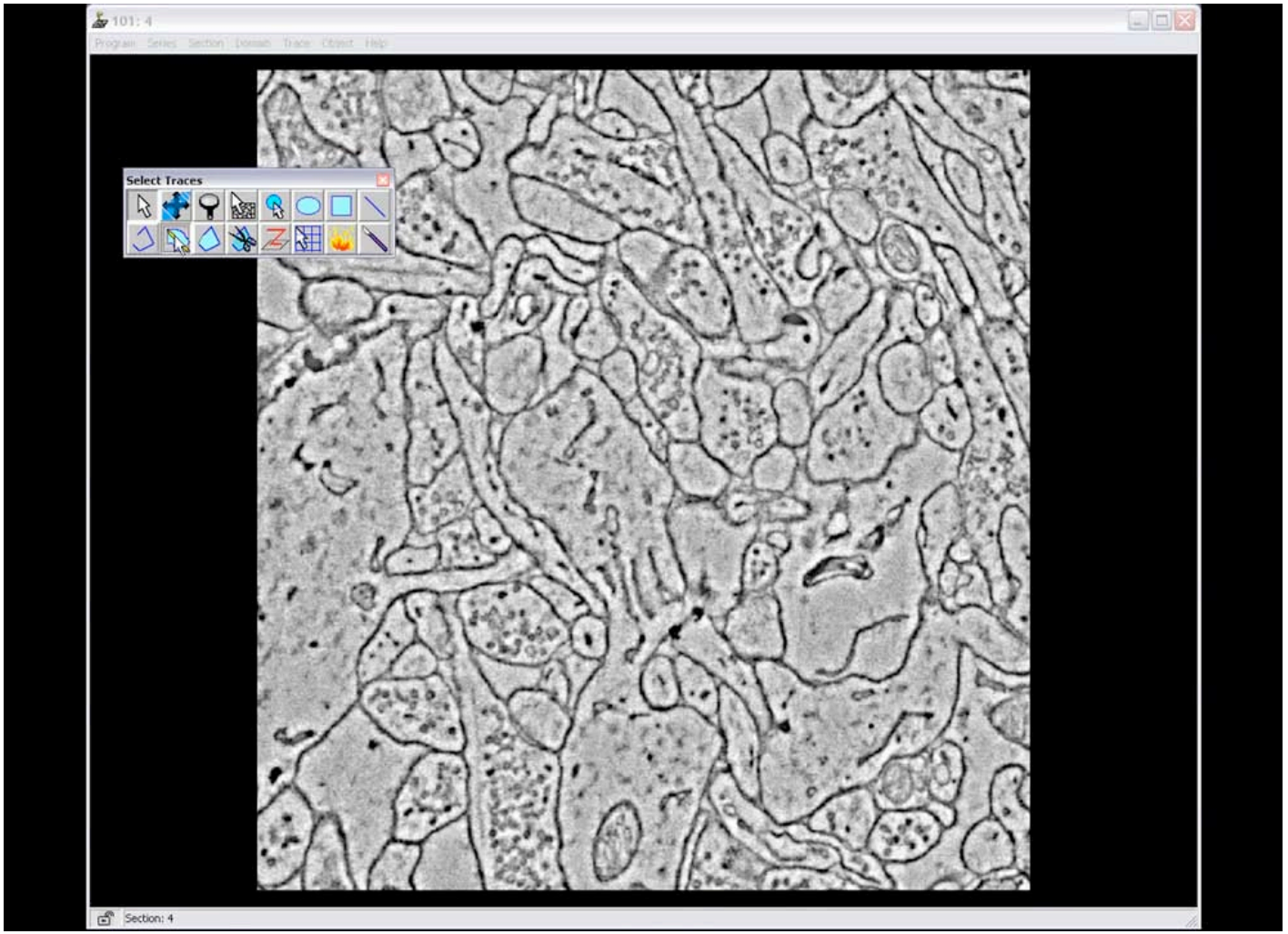
[Jeong et al., TVCG 2009]

Volume Visualization

21,494 X 25,790 X 1850 = 955 GB



Manual Reconstruction



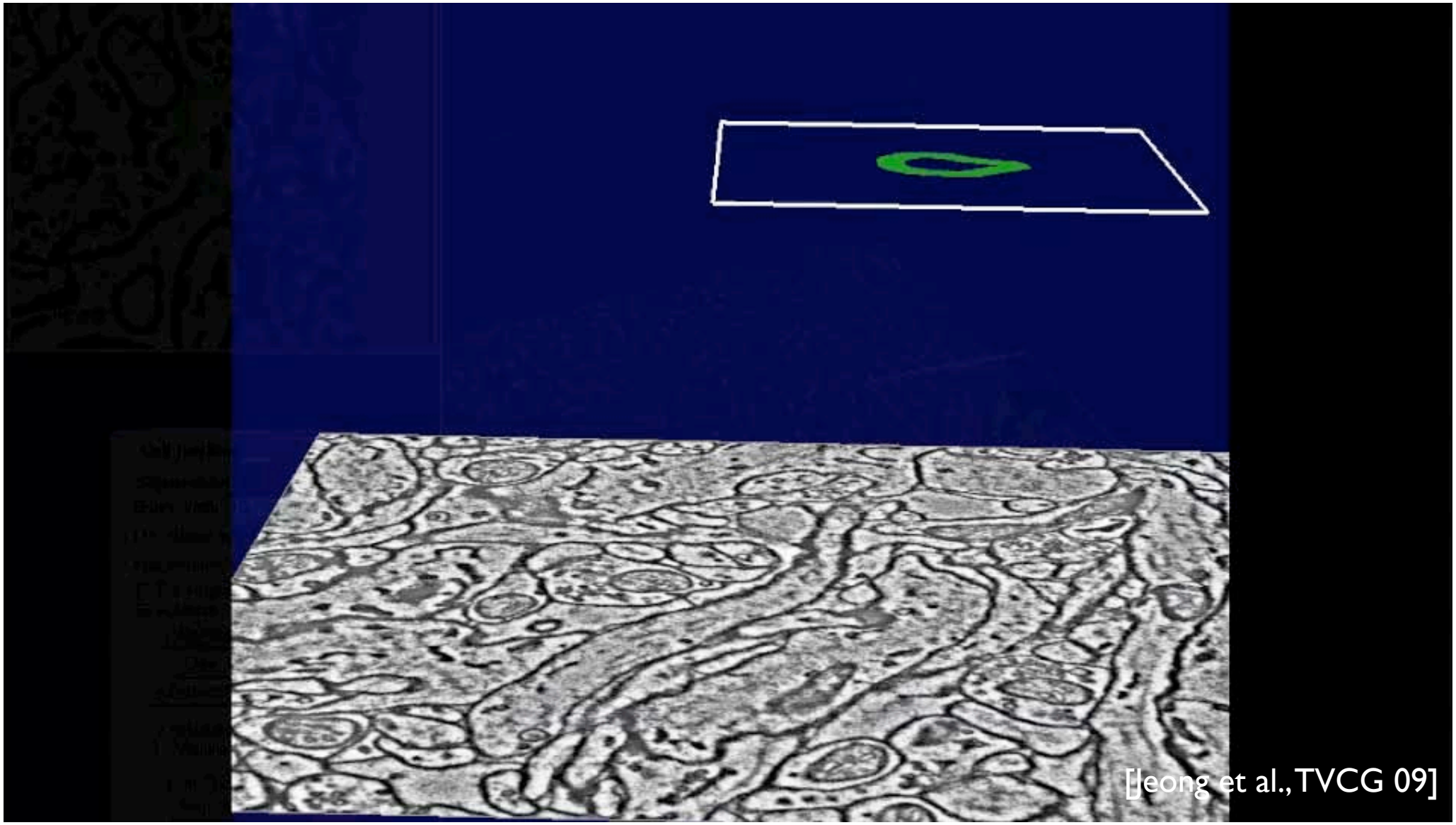
[Reconstruct]

Active Ribbons

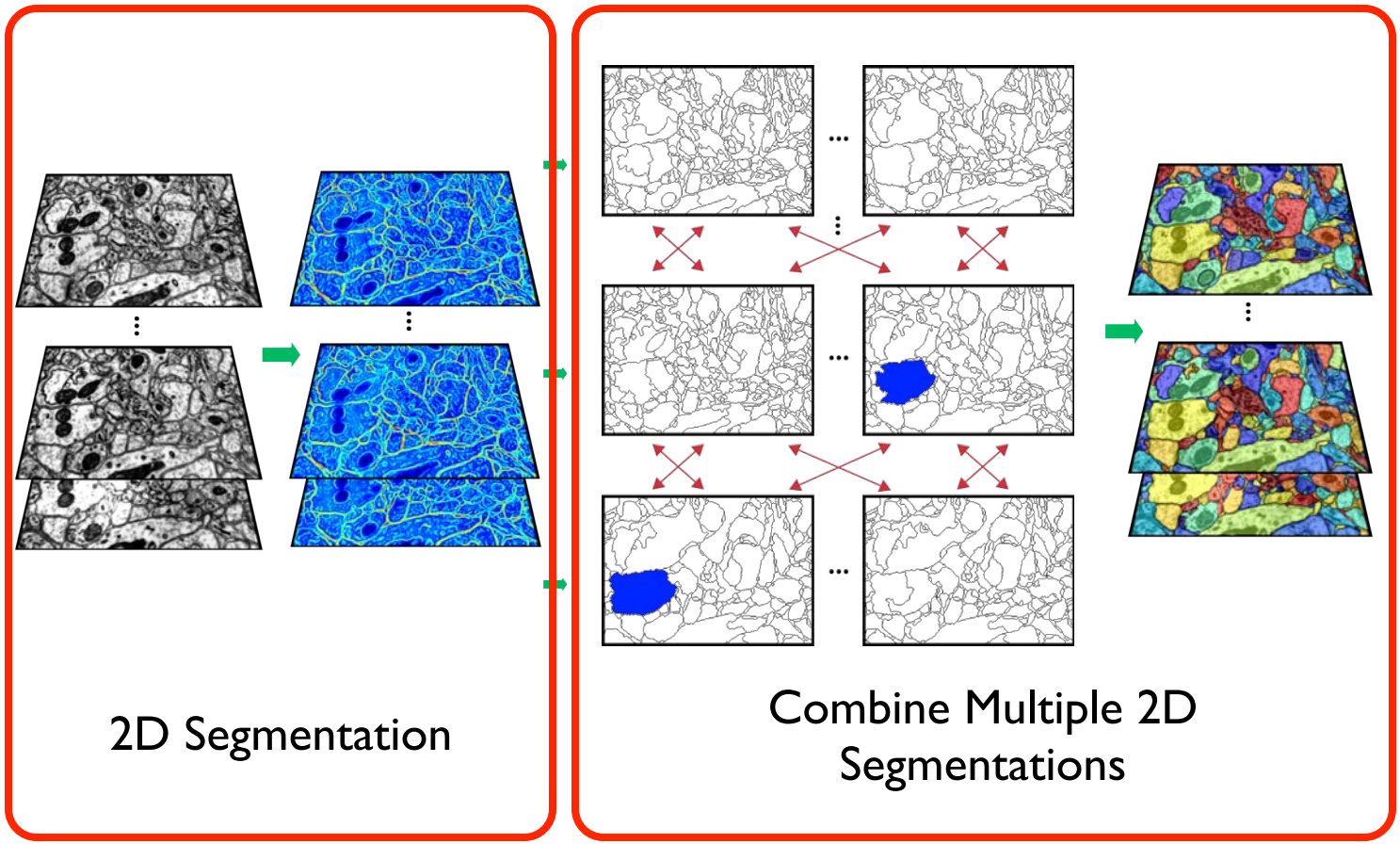
Section 1

[Vazquez et al., CVPR 09]

User-Guided Reconstruction



Automatic Reconstruction



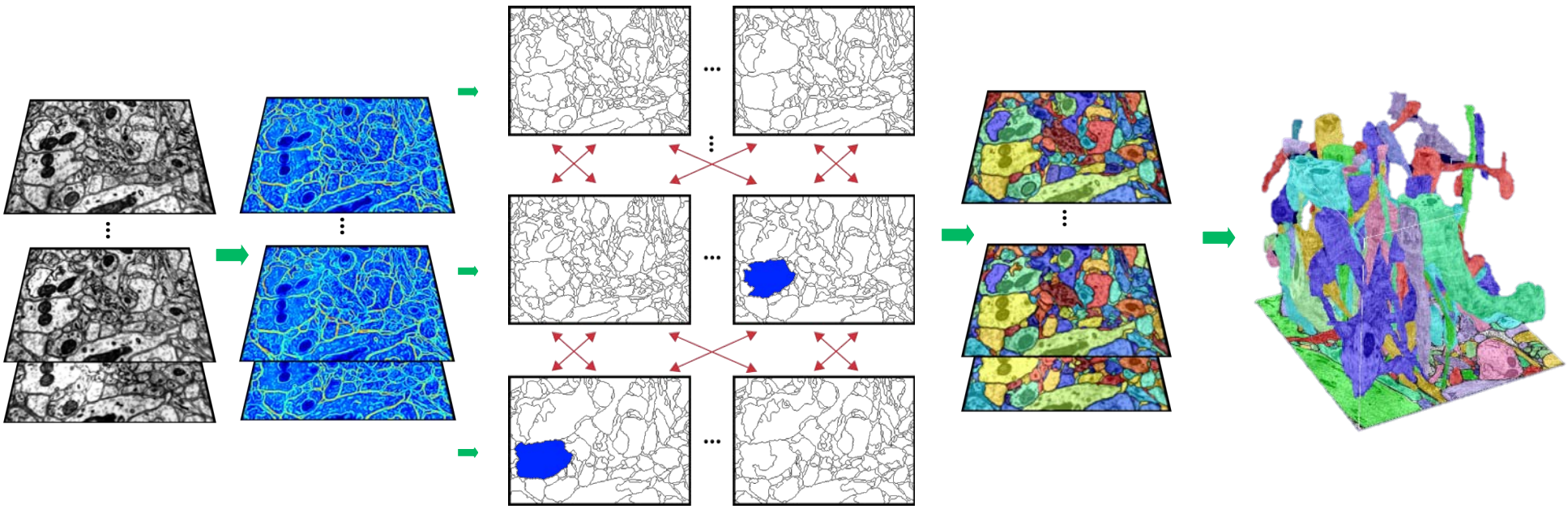
2D Segmentation

Combine Multiple 2D Segmentations

3D Reconstruction

[Kaynig et al., CVPR 10]
[Vazquez et al., ICCV 2011]

Automatic Reconstruction



Membrane probability maps

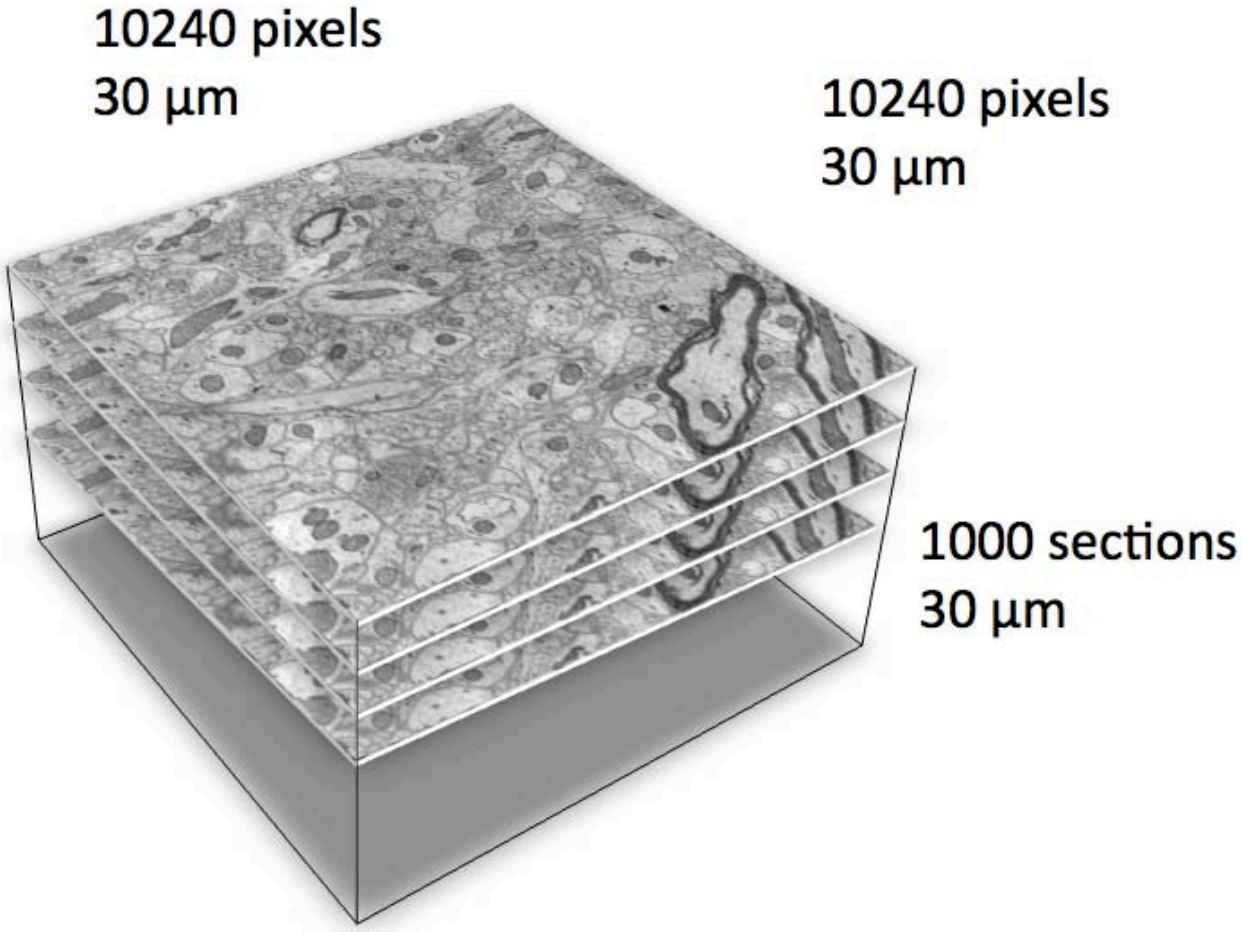
Enumerate and evaluate combinations

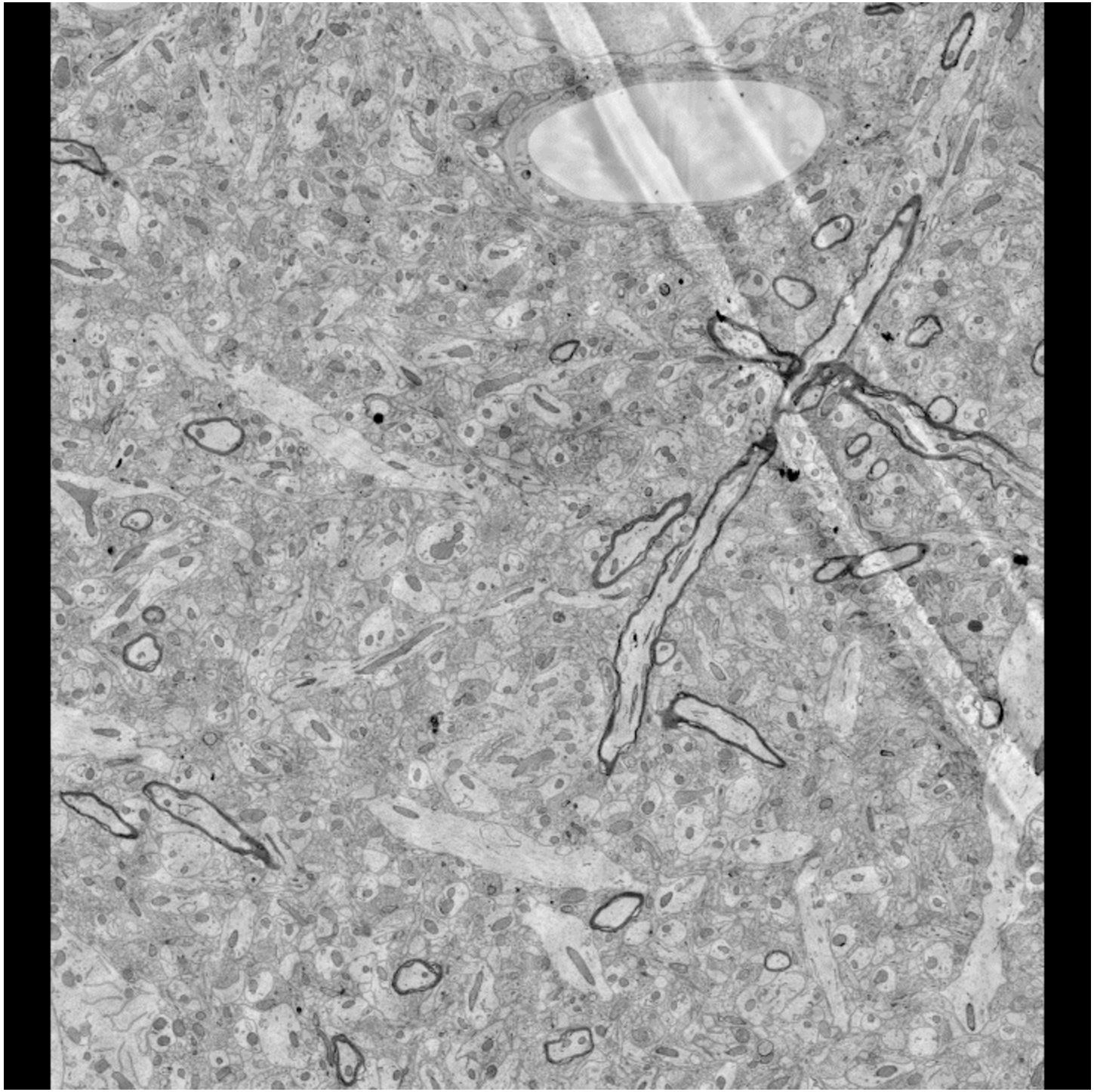
Globally consistent 3D reconstruction

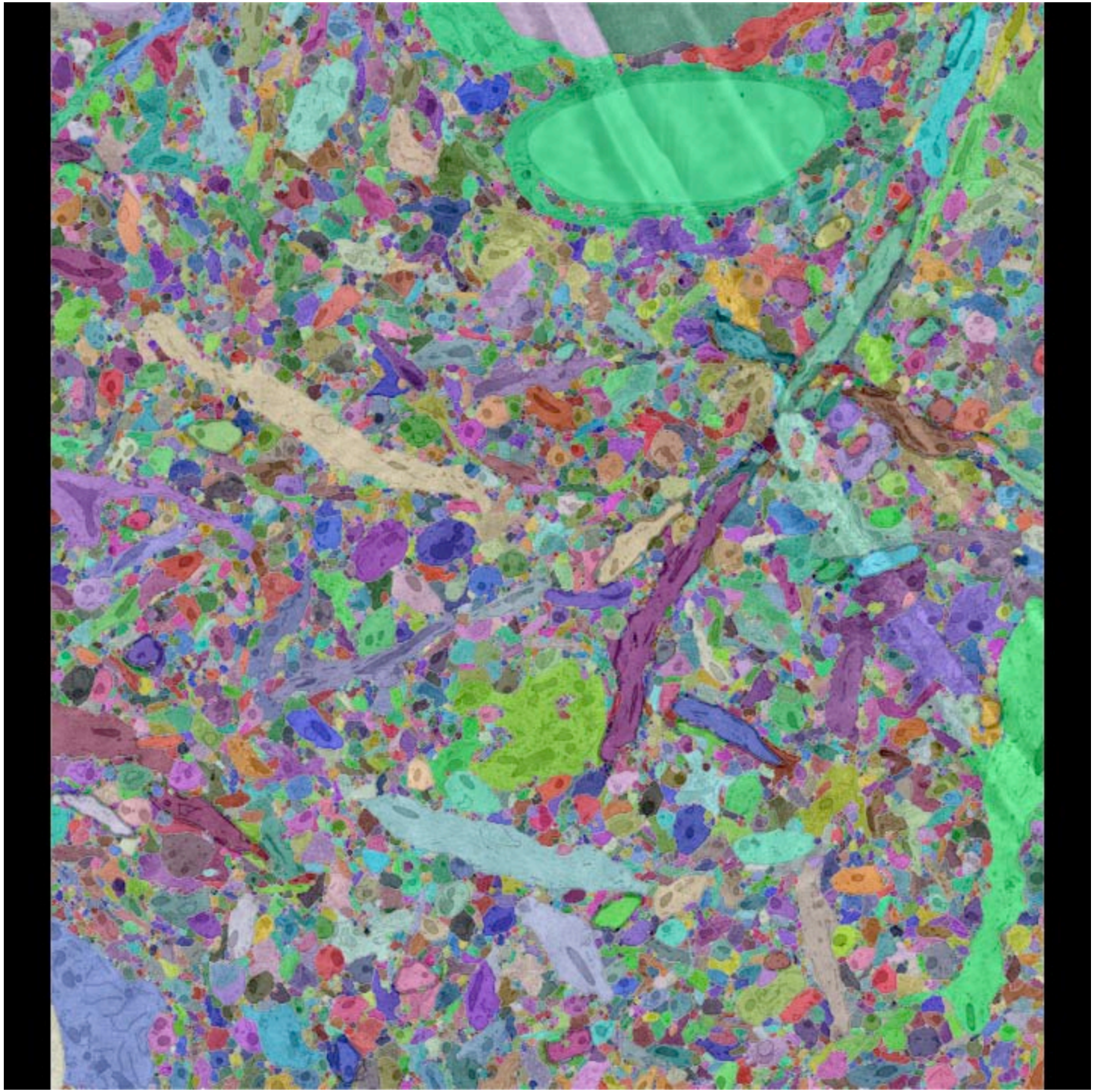
[Kaynig et al., CVPR 10]
[Vazquez et al., ICCV 2011]

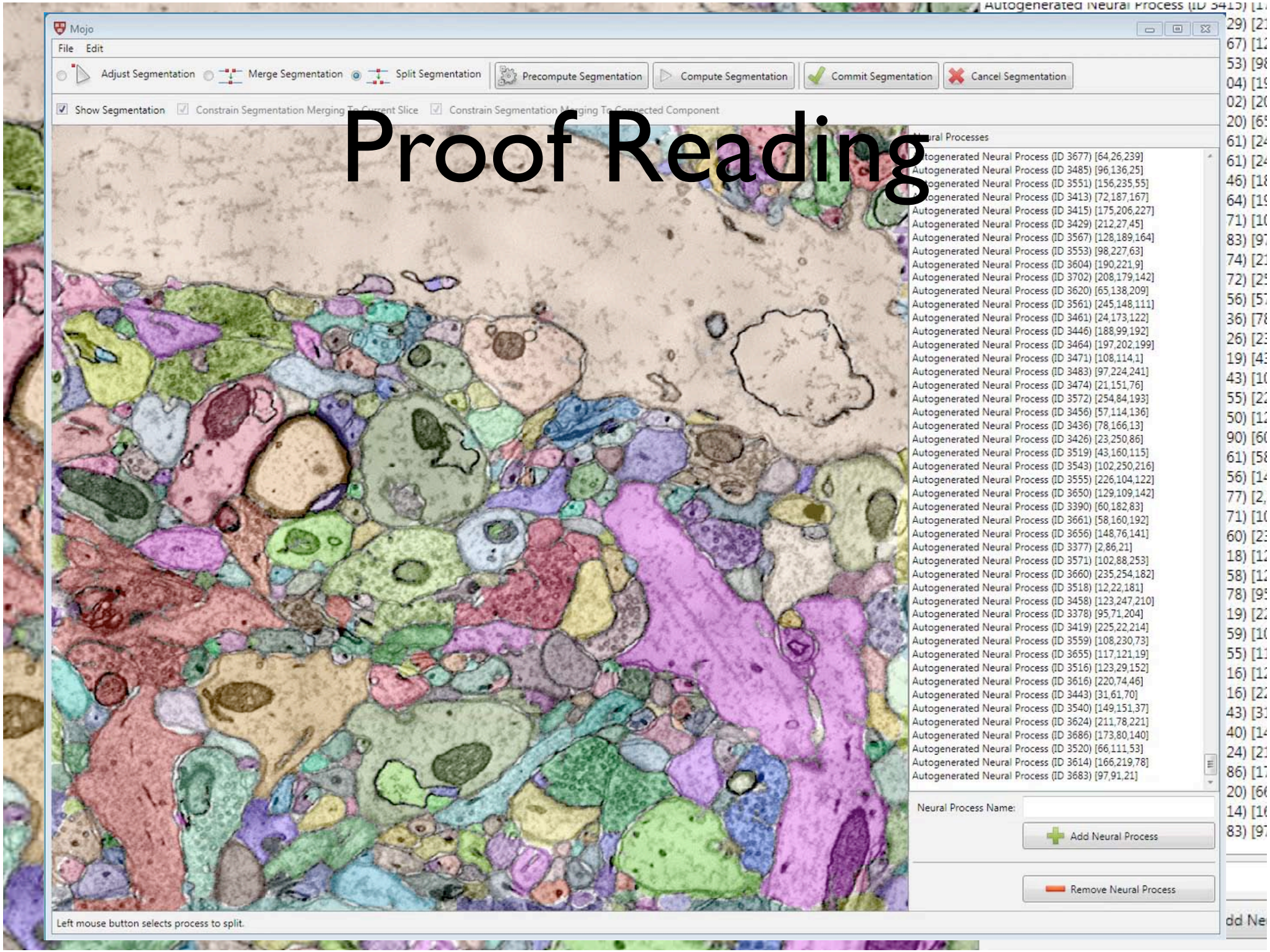
Jeff's Challenge

Automatic 3D reconstruction for largest EM stack **ever**

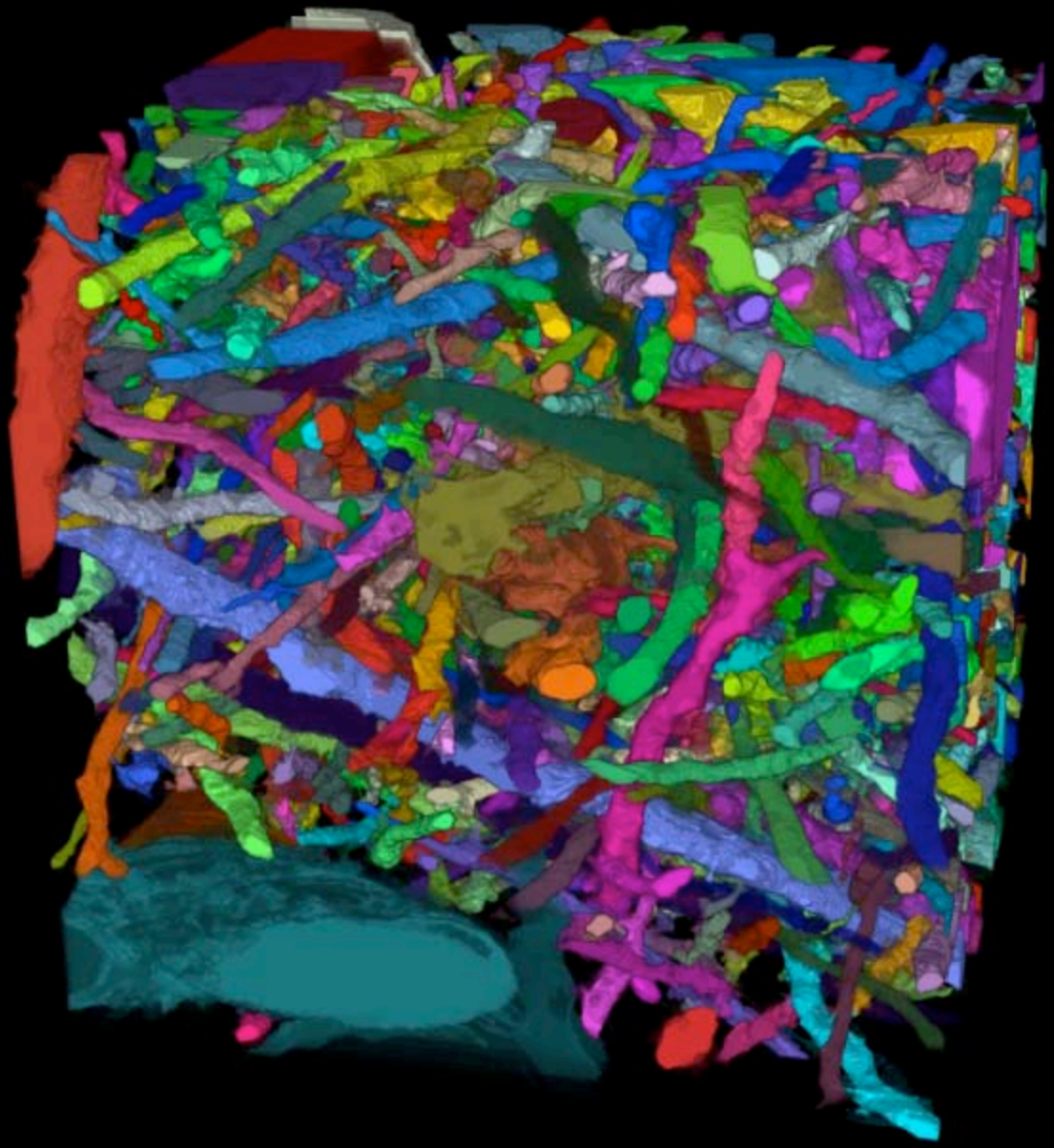


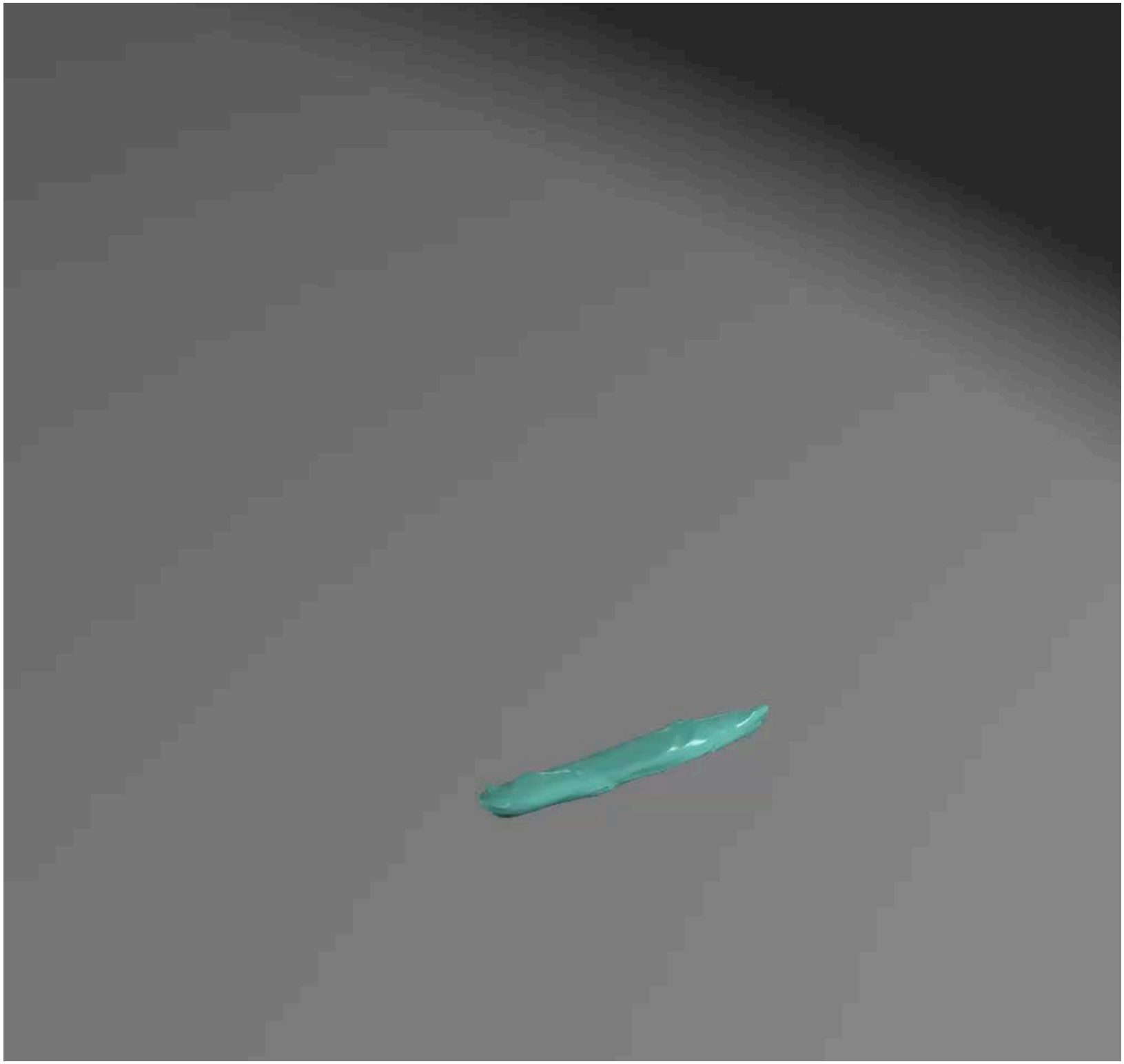


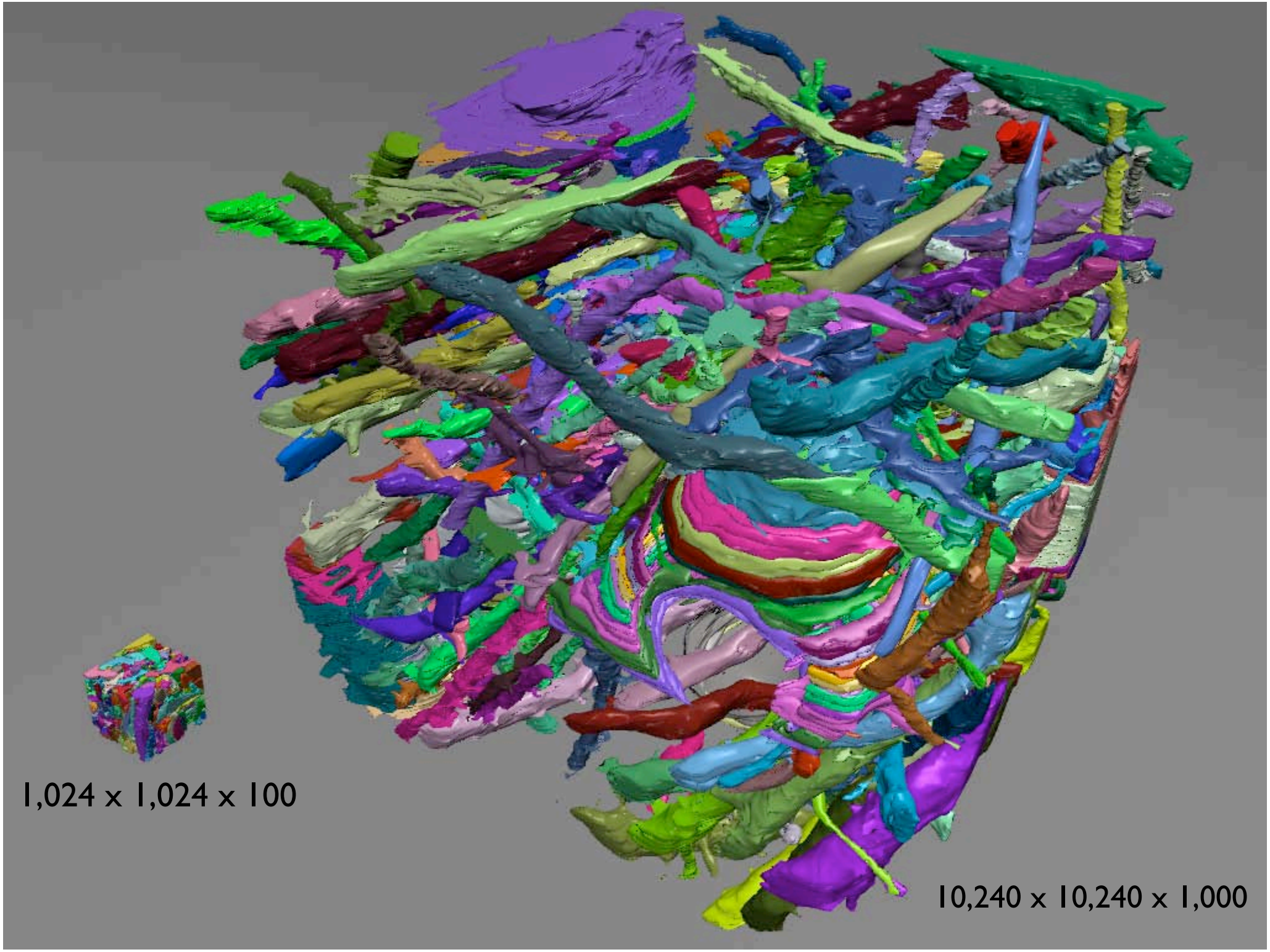



Mojo
File Edit
Adjust Segmentation
Merge Segmentation
Split Segmentation
Precompute Segmentation
Compute Segmentation
Commit Segmentation
Cancel Segmentation
Show Segmentation
Constrain Segmentation Merging To Current Slice
Constrain Segmentation Merging To Connected Component
Proof Reading
Autogenerated Neural Process (ID 3677) [64,26,239]
Autogenerated Neural Process (ID 3485) [96,136,25]
Autogenerated Neural Process (ID 3551) [156,235,55]
Autogenerated Neural Process (ID 3413) [72,187,167]
Autogenerated Neural Process (ID 3415) [175,206,227]
Autogenerated Neural Process (ID 3429) [212,27,45]
Autogenerated Neural Process (ID 3567) [128,189,164]
Autogenerated Neural Process (ID 3553) [98,227,63]
Autogenerated Neural Process (ID 3604) [190,221,9]
Autogenerated Neural Process (ID 3702) [208,179,142]
Autogenerated Neural Process (ID 3620) [65,138,209]
Autogenerated Neural Process (ID 3561) [245,148,111]
Autogenerated Neural Process (ID 3461) [24,173,122]
Autogenerated Neural Process (ID 3446) [188,99,192]
Autogenerated Neural Process (ID 3464) [197,202,199]
Autogenerated Neural Process (ID 3471) [108,114,1]
Autogenerated Neural Process (ID 3483) [97,224,241]
Autogenerated Neural Process (ID 3474) [21,151,76]
Autogenerated Neural Process (ID 3572) [254,84,193]
Autogenerated Neural Process (ID 3456) [57,114,136]
Autogenerated Neural Process (ID 3436) [78,166,13]
Autogenerated Neural Process (ID 3426) [23,250,86]
Autogenerated Neural Process (ID 3519) [43,160,115]
Autogenerated Neural Process (ID 3543) [102,250,216]
Autogenerated Neural Process (ID 3555) [226,104,122]
Autogenerated Neural Process (ID 3650) [129,109,142]
Autogenerated Neural Process (ID 3390) [60,182,83]
Autogenerated Neural Process (ID 3661) [58,160,192]
Autogenerated Neural Process (ID 3656) [148,76,141]
Autogenerated Neural Process (ID 3377) [2,86,21]
Autogenerated Neural Process (ID 3571) [102,88,253]
Autogenerated Neural Process (ID 3660) [235,254,182]
Autogenerated Neural Process (ID 3518) [12,22,181]
Autogenerated Neural Process (ID 3458) [123,247,210]
Autogenerated Neural Process (ID 3378) [95,71,204]
Autogenerated Neural Process (ID 3419) [225,22,214]
Autogenerated Neural Process (ID 3559) [108,230,73]
Autogenerated Neural Process (ID 3655) [117,121,19]
Autogenerated Neural Process (ID 3516) [123,29,152]
Autogenerated Neural Process (ID 3616) [220,74,46]
Autogenerated Neural Process (ID 3443) [31,61,70]
Autogenerated Neural Process (ID 3540) [149,151,37]
Autogenerated Neural Process (ID 3624) [211,78,221]
Autogenerated Neural Process (ID 3686) [173,80,140]
Autogenerated Neural Process (ID 3520) [66,111,53]
Autogenerated Neural Process (ID 3614) [166,219,78]
Autogenerated Neural Process (ID 3683) [97,91,21]
Neural Process Name:
Add Neural Process
Remove Neural Process
Left mouse button selects process to split.

All objects containing at least 100 sections



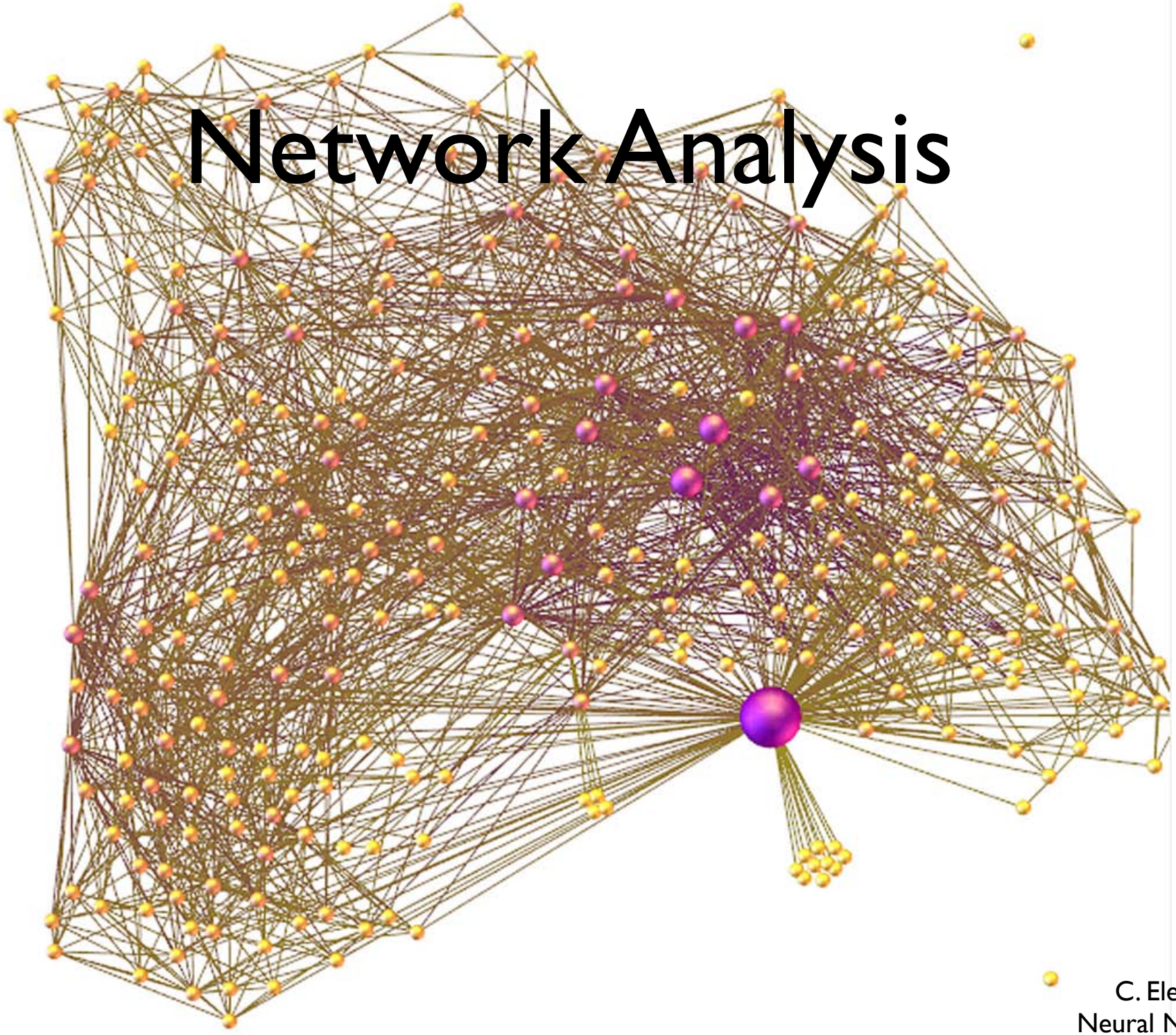



1,024 x 1,024 x 100
10,240 x 10,240 x 1,000

33,000 x 33,000 x 10,000

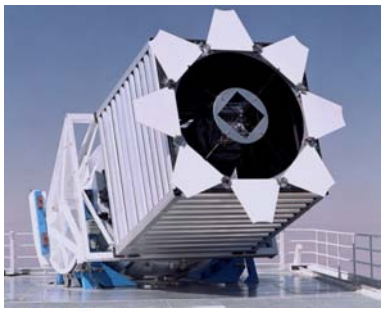
Future Work

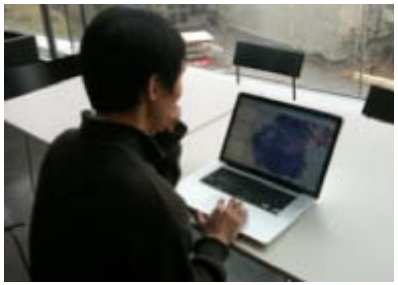
Network Analysis



C. Elegans
Neural Network

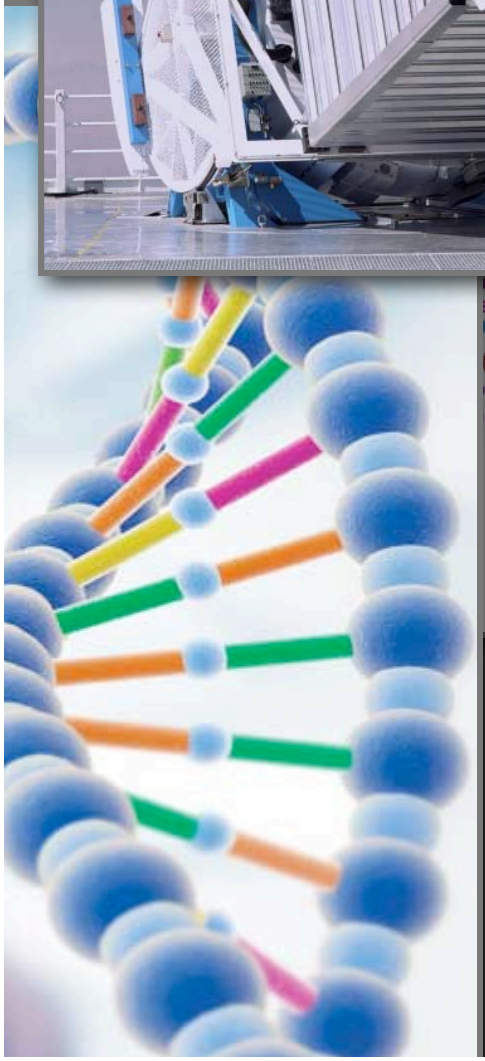
Distributed Image Analysis & Visualization

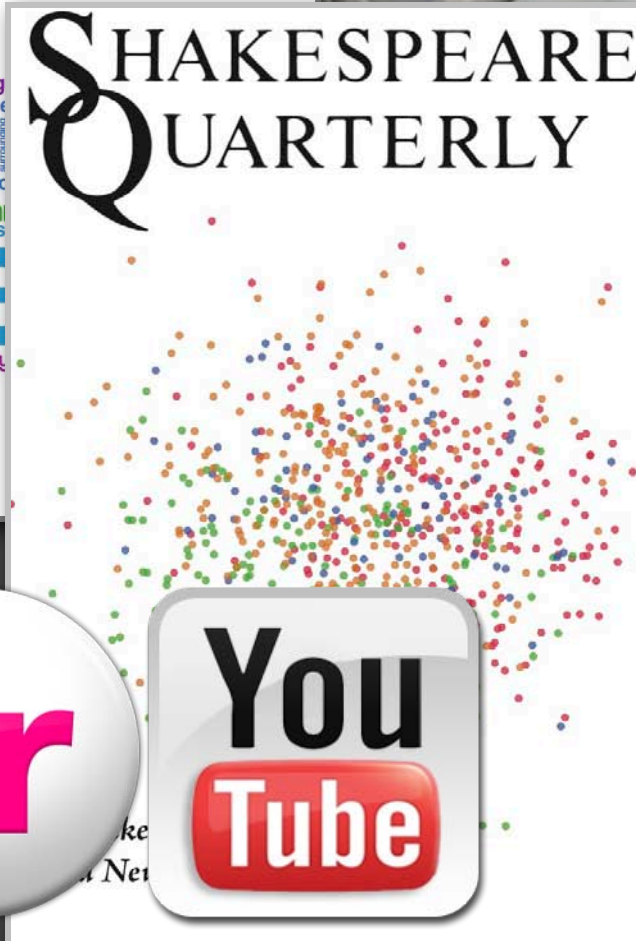




Human Analysis

Data = Opportunities





Thank you!
http://gvi.seas.harvard.edu/pfister