Interactive Visual Analysis of Biological Data

Tutorial: Interactive Visual Analysis of Scientific Data

Gunther H. Weber

Berkeley Drosophila Transcription Network Project

VisWeek 2012

SCIVIS • INFOVIS • VAST
BIOVIS • LDAV

Interactive Exploration of 3D Gene Expression Data

G.H. Weber, O. Rübel, M.-Y. Huang, A. H. DePace, C. C. Fowlkes, S. V. E. Keränen, C. L. Luengo Hendriks, H. Hagen, D. W. Knowles, J. Malik, M. D. Biggin and B. Hamann: *Visual Exploration of Three-dimensional Gene Expression Using Physical Views and Linked Abstract Views .* IEEE Trans. Comp. Bio. Bioinform., 6(2): 296-309 (2009)

Analysis of 3D Gene Expression Patterns

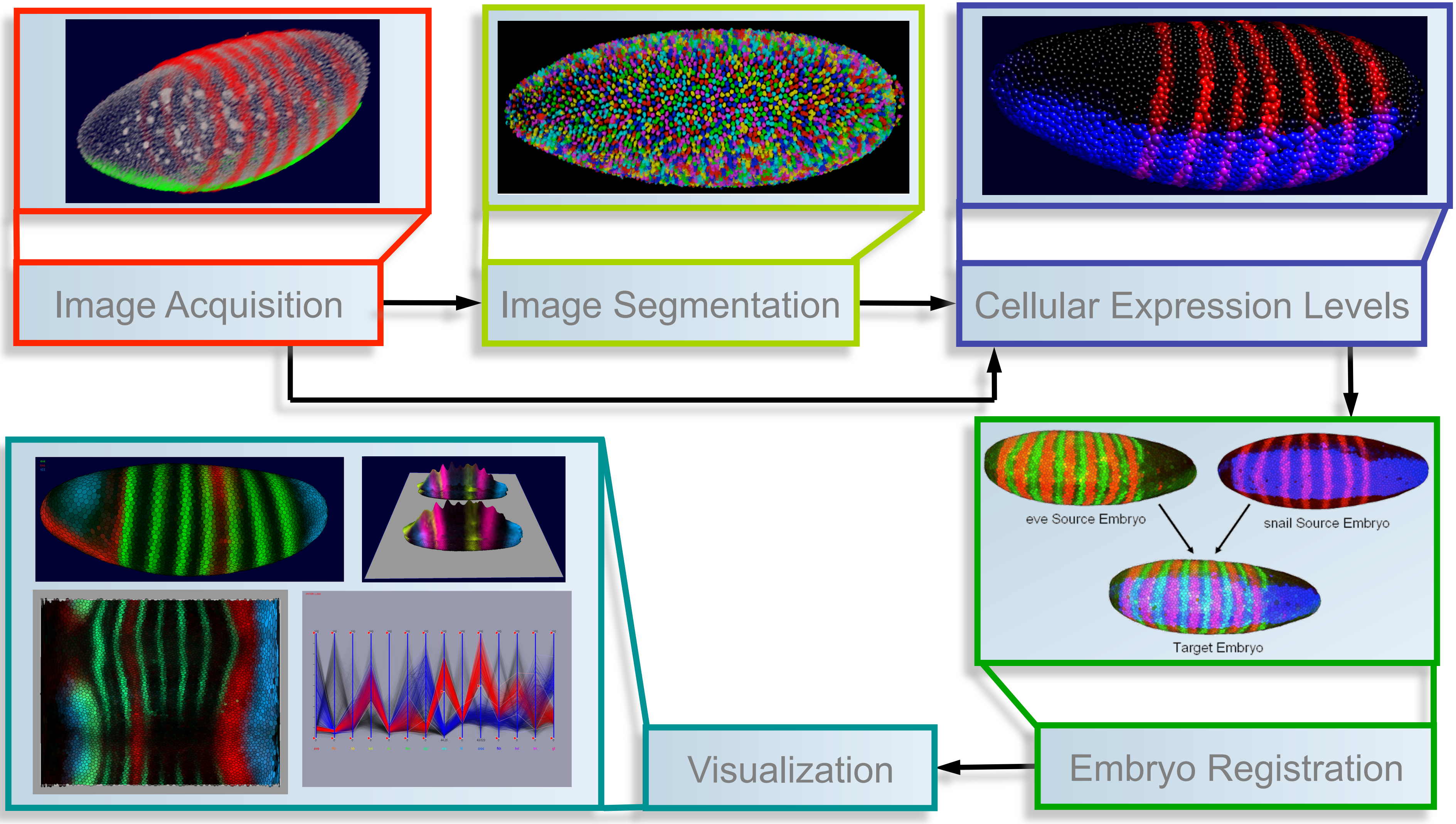


Image Acquisition

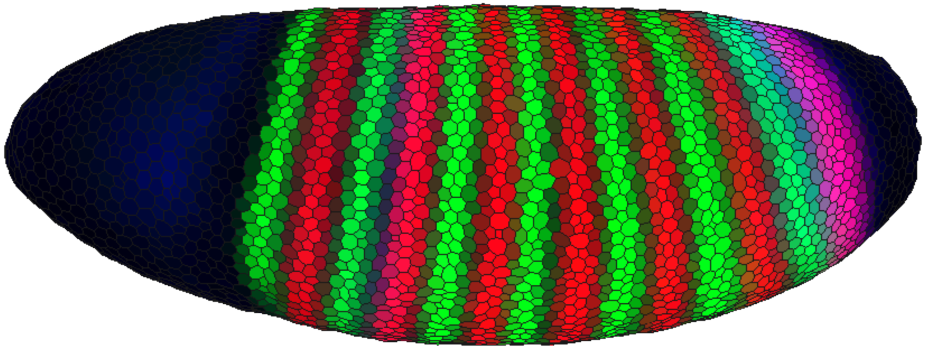
Image Segmentation

Cellular Expression Levels

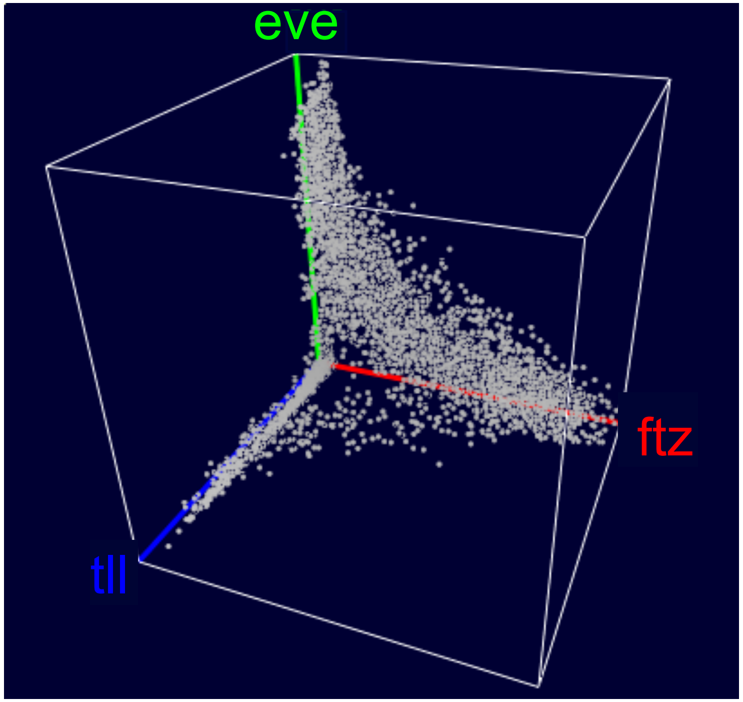
Embryo Registration

Visualization

Multiple Views For Different Aspects of Gene Expression

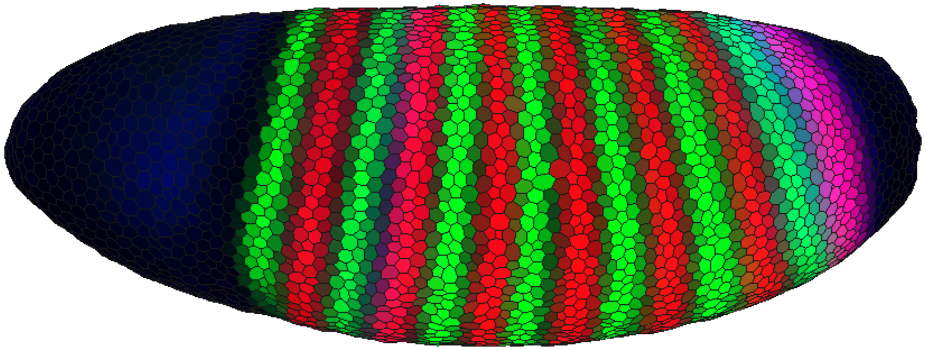


Physical views show gene expression as spatial patterns in relation to the embryo

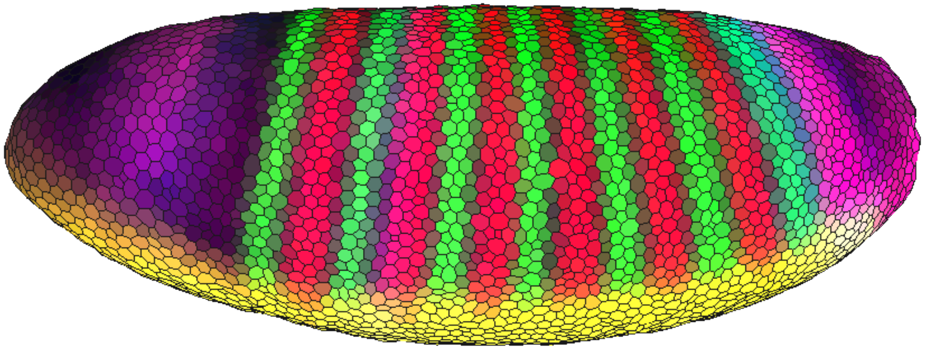


Expression space views show the relationship between expression of different genes

3D Embryo View



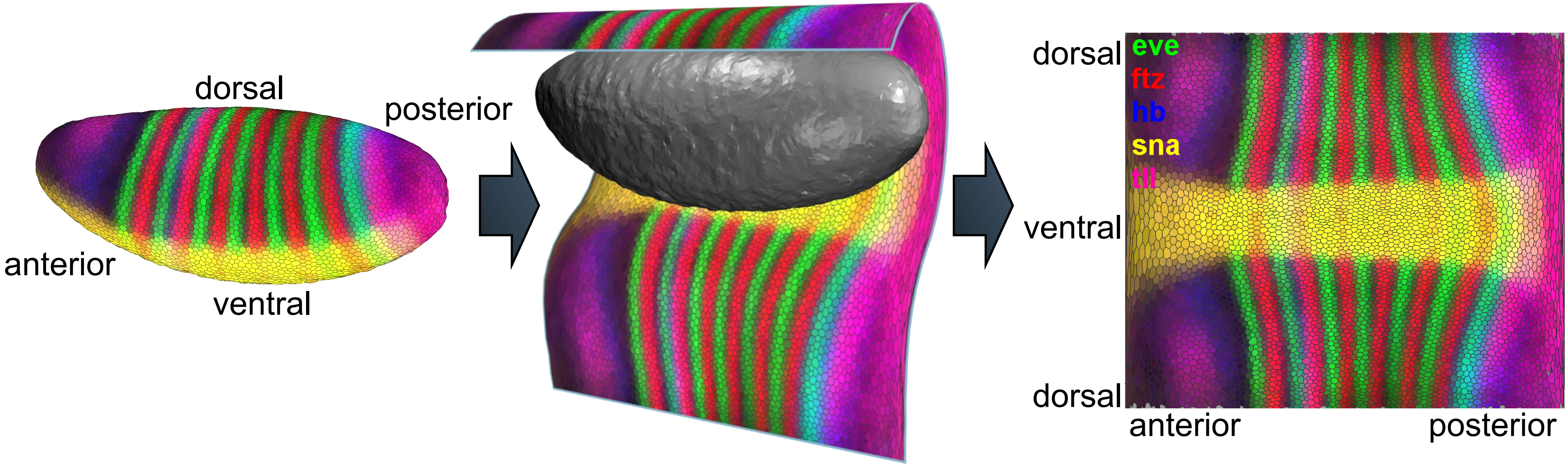
ftz eve hb



ftz eve hb sna tll

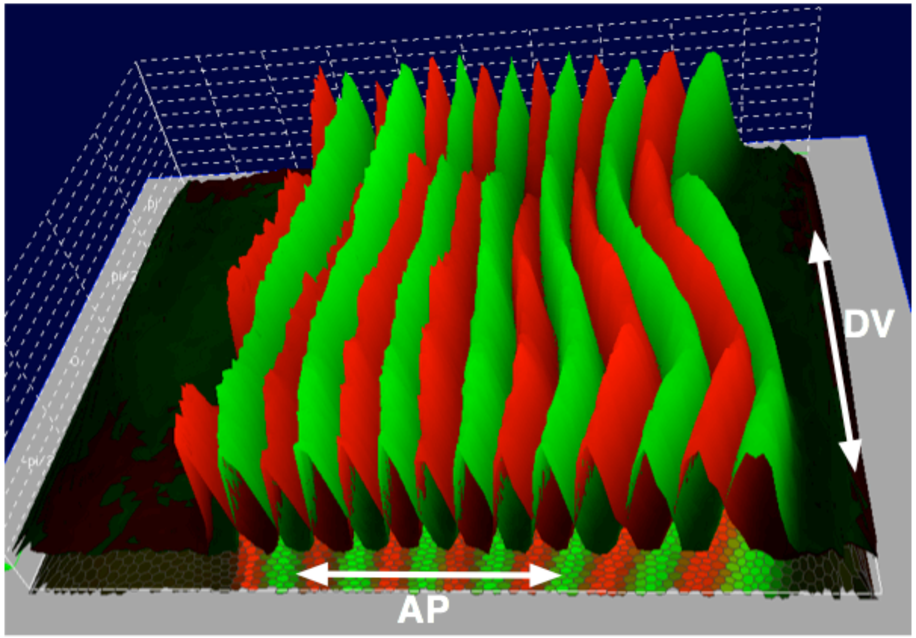
- Each cell represented by a 3D object at the physical cell location
- Usage:
  - Qualitative analysis of gene expression levels
  - Analysis and comparison of gene expression patterns

Unrolled View

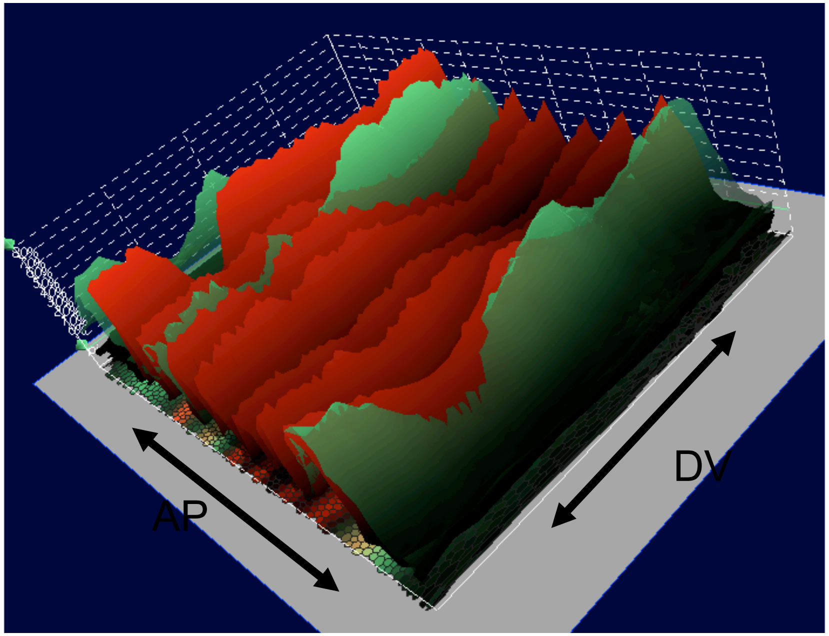


- Entire embryo in one cylindrical projection
- Third dimension can be used for display of additional information

3D Offset Surfaces



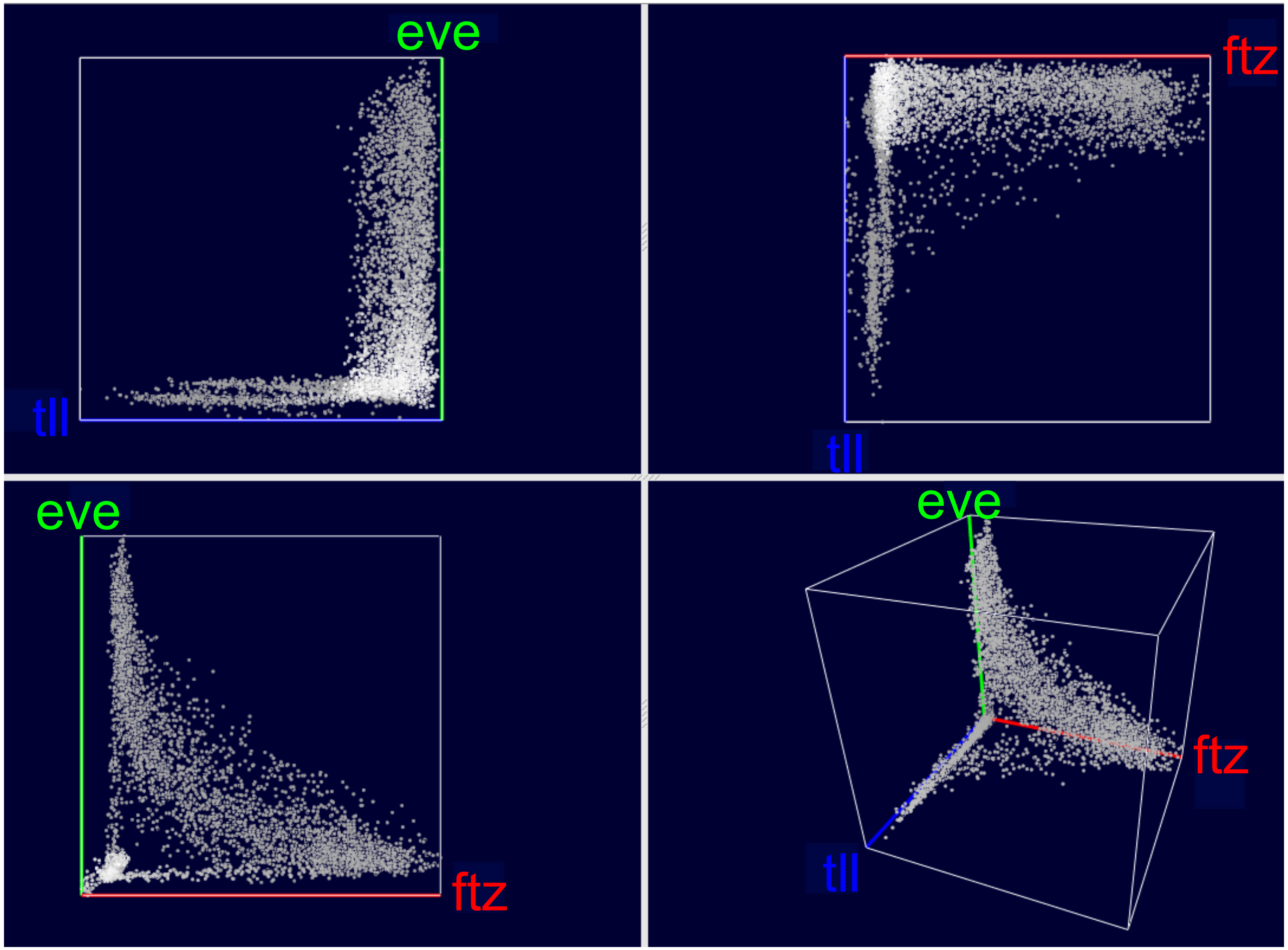
ftz eve



eve hb

- Small differences in expression visible
- Quantitative analysis of gene expression

2D/3D Scatter Plots



- Identify relationships between gene expression levels

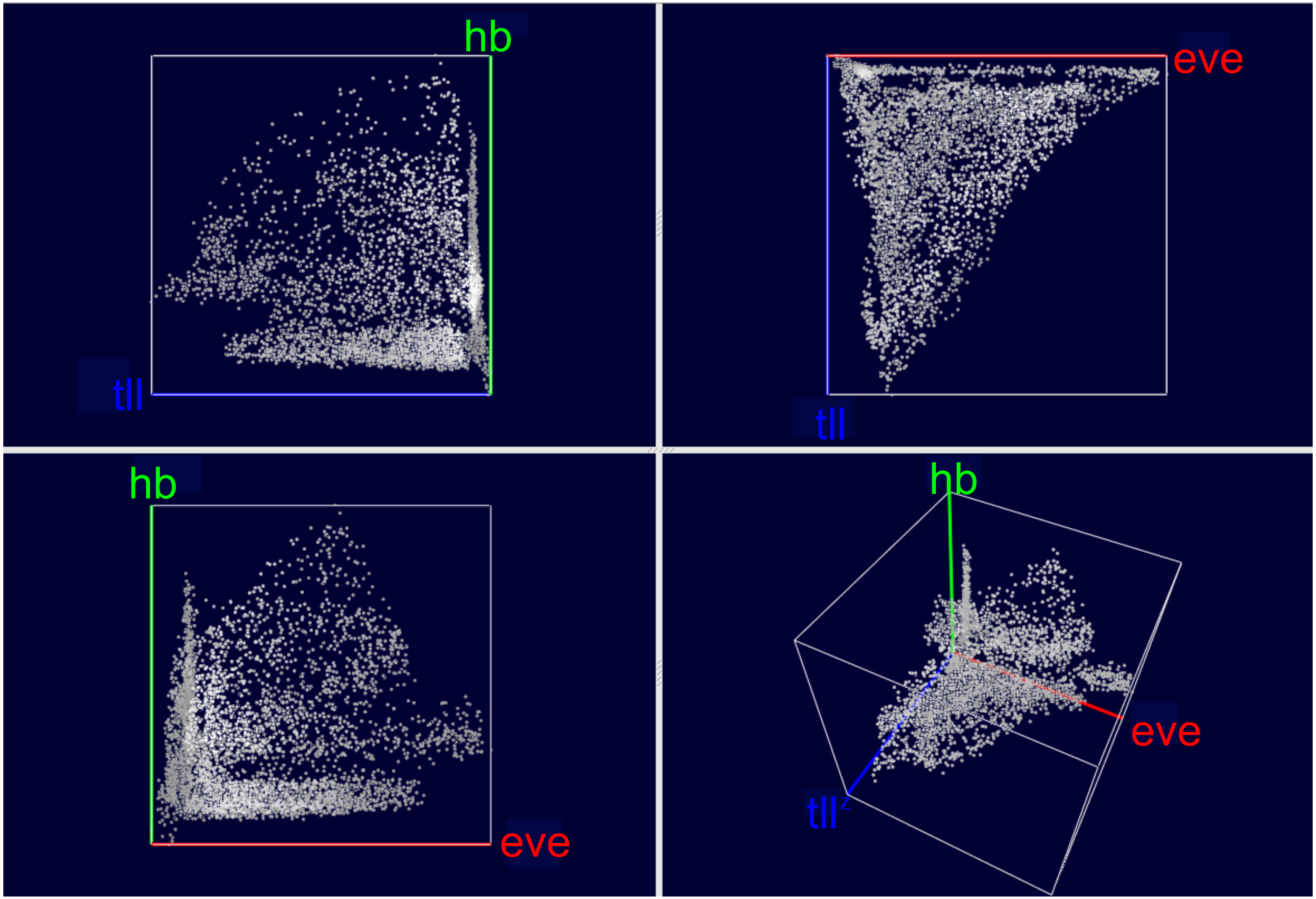
Correlate Views vis “Show & Brush”

- *Brushing*: Select a subset of cells in one view
  - By “painting” on embryo in physical view
  - By choosing expression level ranges in expression space views

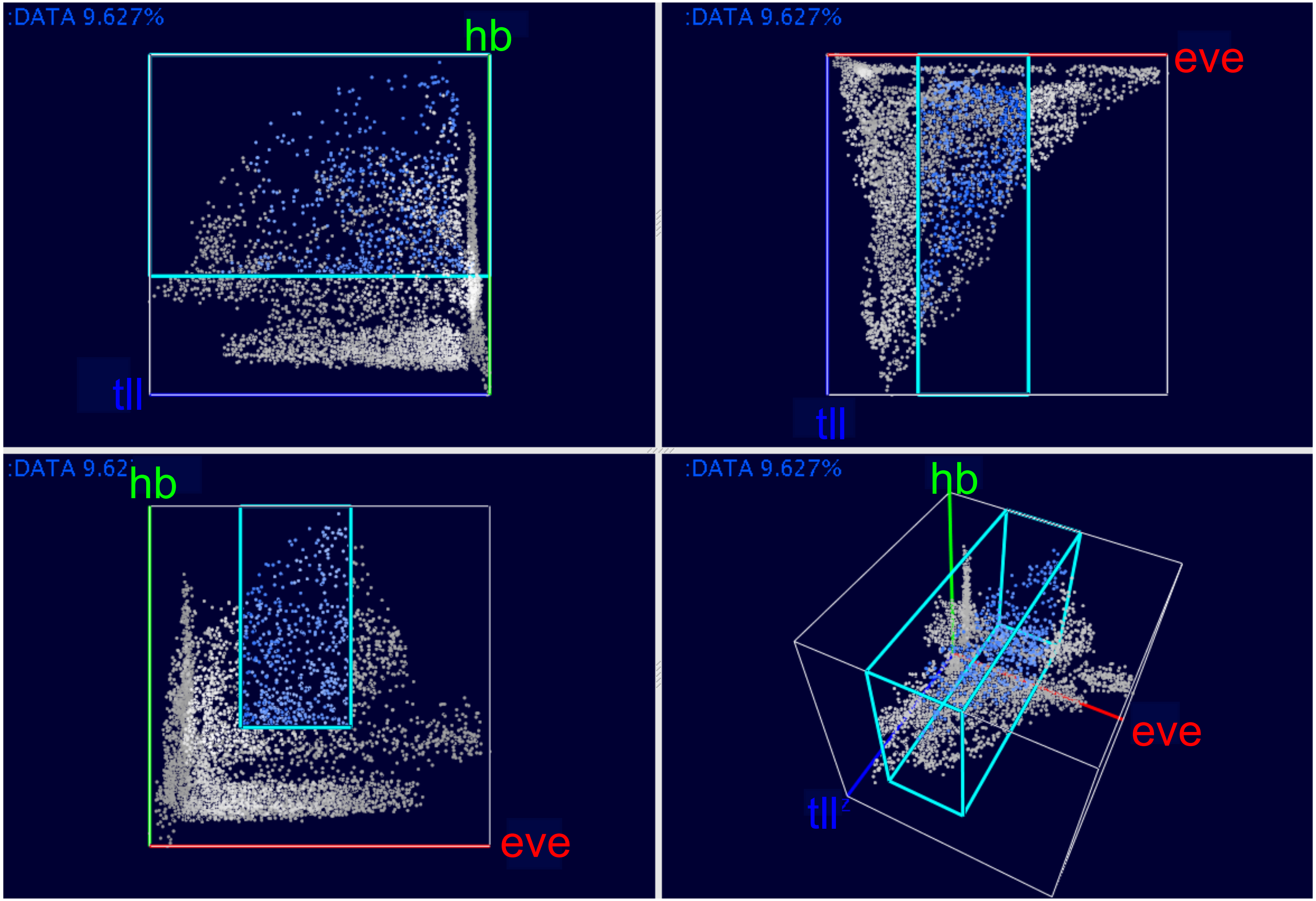
- *Linking:* Highlight selected cells in all other views

- **Fundamental Idea:**
  - Select “interesting expression properties” in one view and examine properties of this subset in other views.

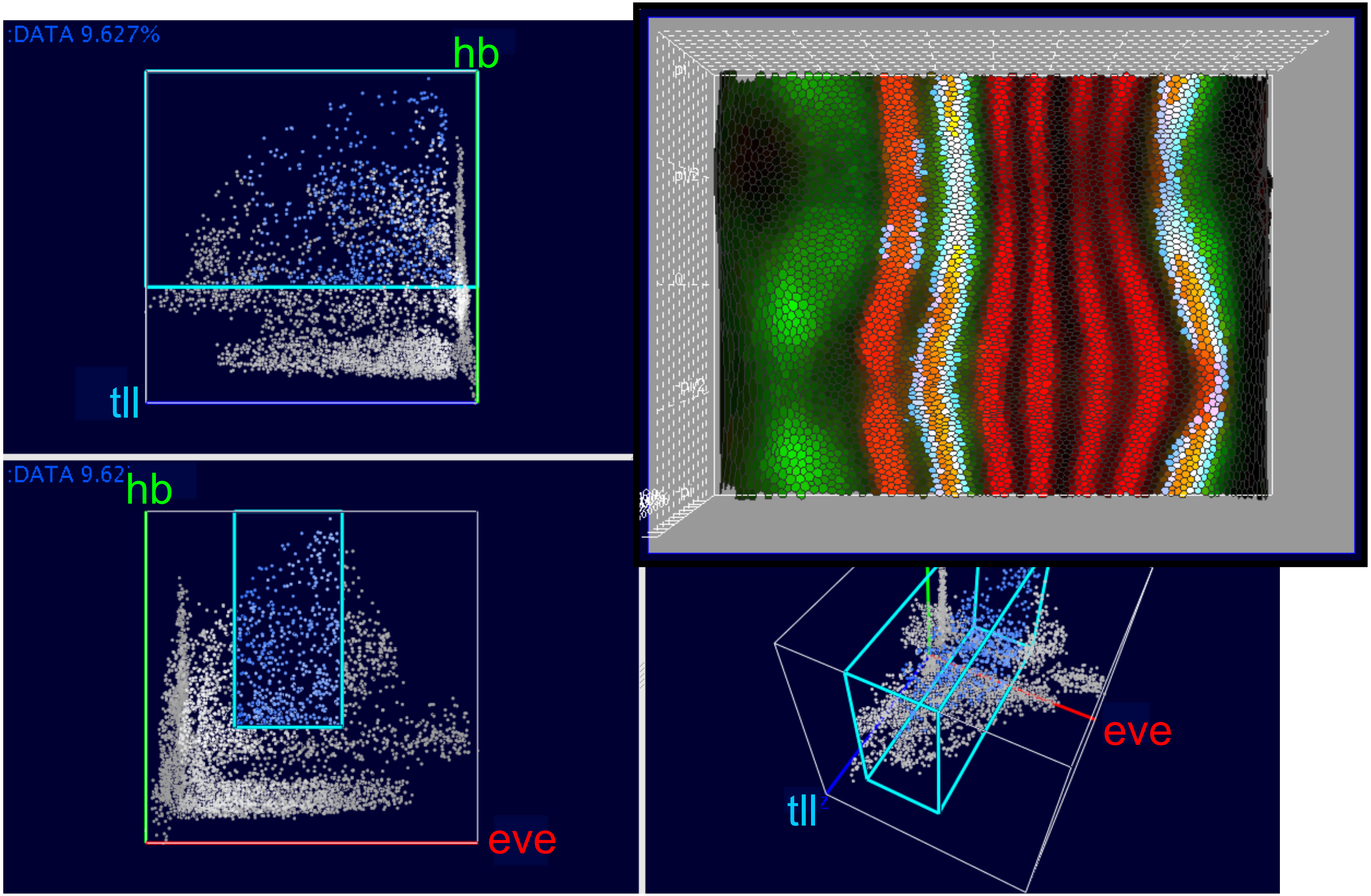
Brushing and Linking Example – Scatterplot



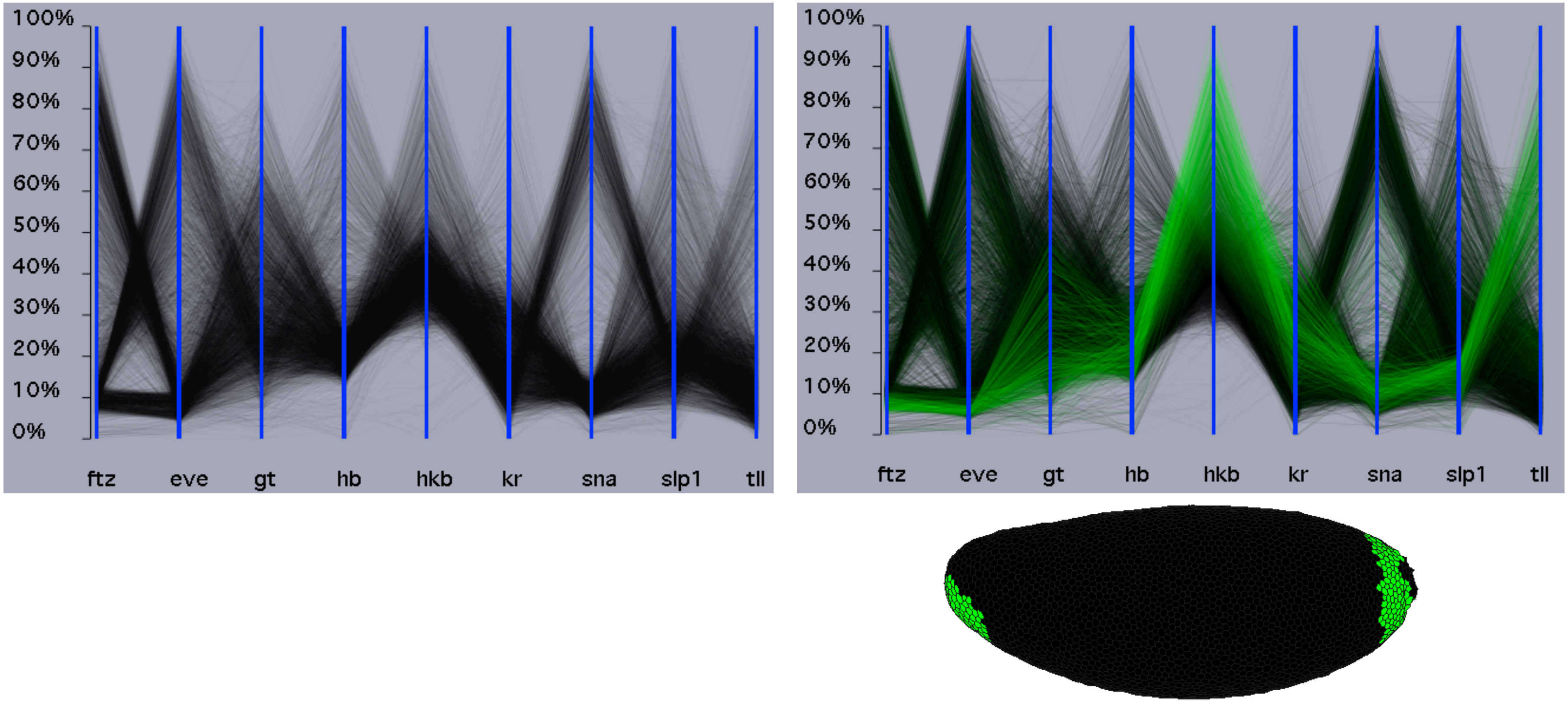
Brushing and Linking Example – Brushing



Brushing and Linking Example – Linking



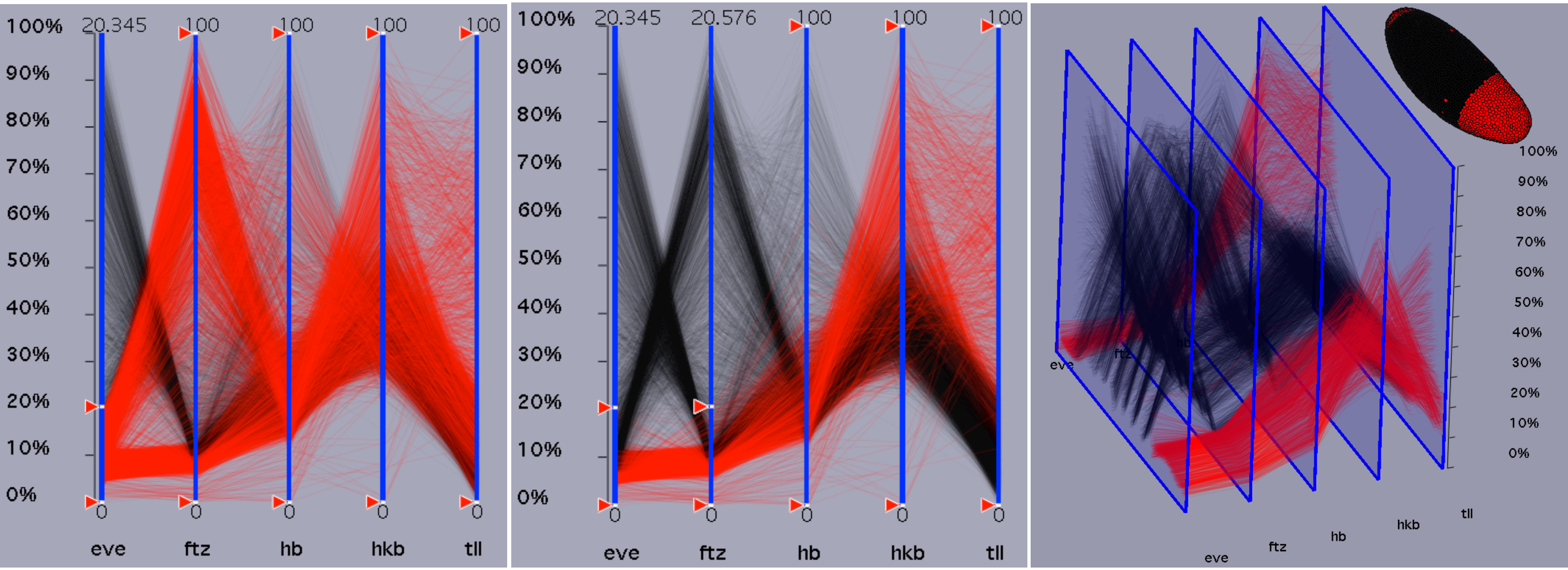
Parallel Coordinates



- Support analysis of relationship of more than three genes

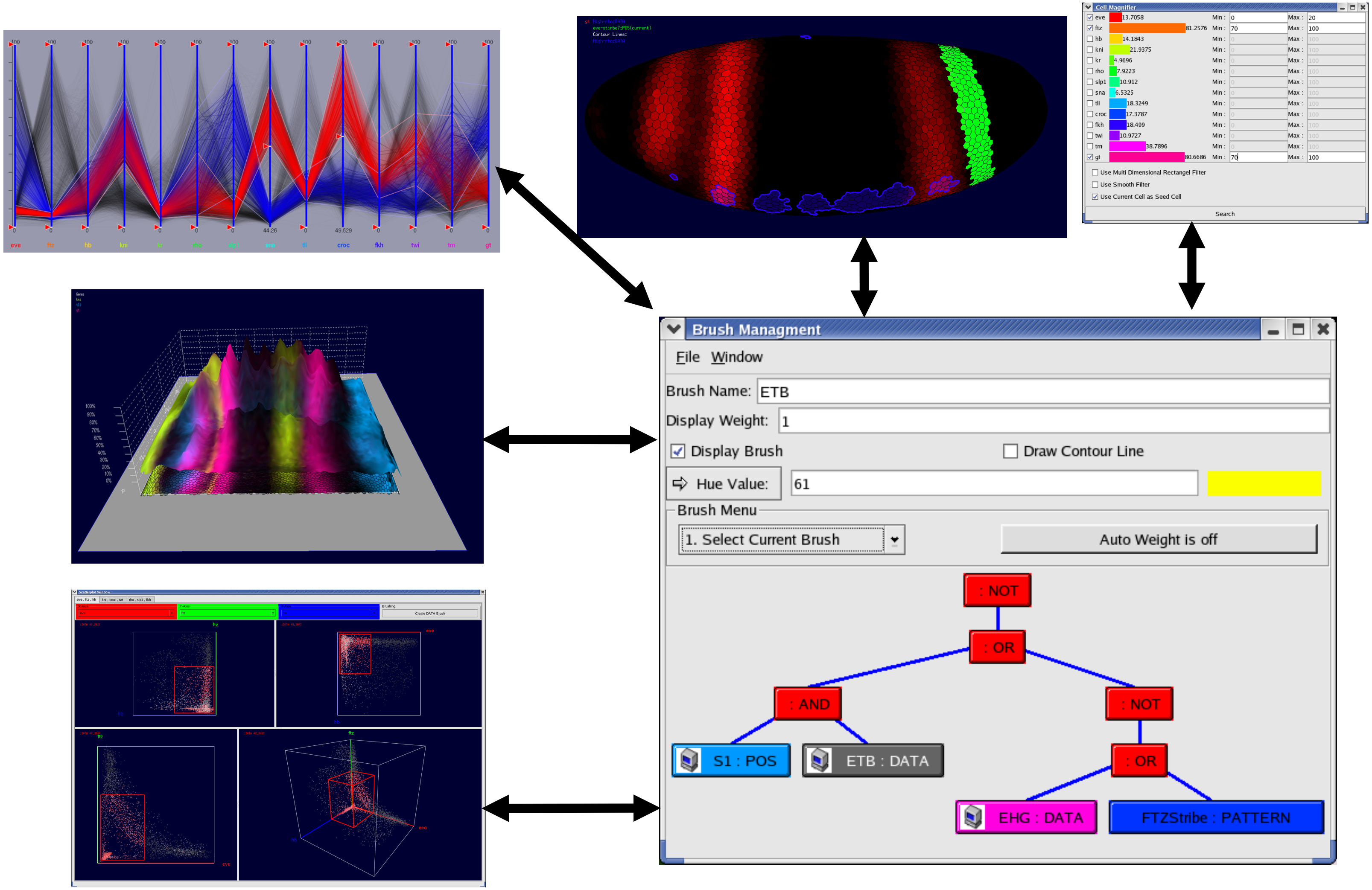
[Rübel et al., Eurographics/IEEE VGTC EuroVis Conference, Lisbon, Portugal, 2006]

Brushing in Parallel Coordinates



- Selection of expression range along each parallel coordinate axis possible

Linked Views – Combining Selections



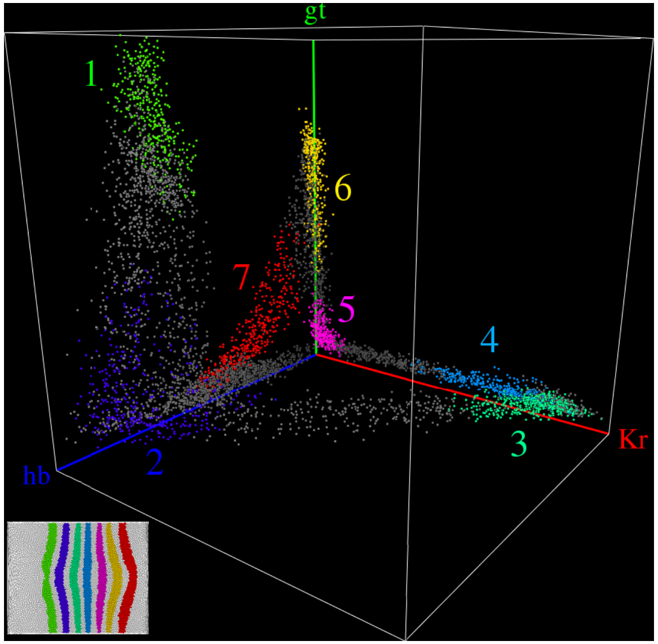
☑/☐	Gene	Value	Min :	Max :
☑	eve	13.7058	0	20
☑	ftz	81.2576	70	100
☐	hb	14.1843	0	100
☐	kni	21.9375	0	100
☐	kr	4.9696	0	100
☐	rho	7.9223	0	100
☐	slp1	10.912	0	100
☐	sna	6.5325	0	100
☐	tll	18.3249	0	100
☐	croc	17.3787	0	100
☐	fkh	18.499	0	100
☐	twi	10.9727	0	100
☐	trn	38.7896	0	100
☑	gt	80.6686	70	100

☐ Use Multi Dimensional Rectangel Filter
☐ Use Smooth Filter
☑ Use Current Cell as Seed Cell
Search

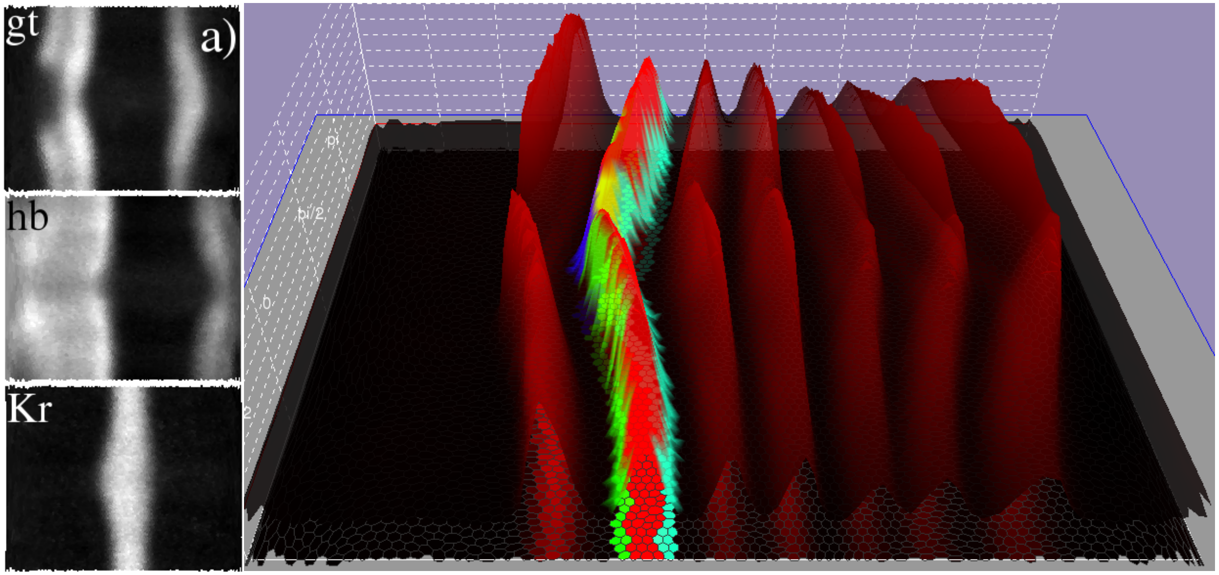
Brush Name: ETB
Display Weight: 1
☑ Display Brush ☐ Draw Contour Line
Hue Value: 61
1. Select Current Brush
Auto Weight is off

[Weber et al., IEEE TCBB, 2009]

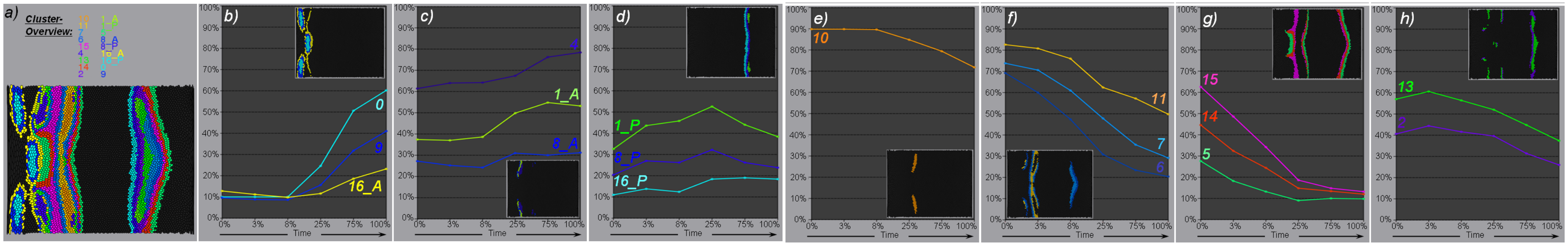
Clustering of 3D Gene Expression



How does the expression pattern of a single gene vary?



How do gene regulatory networks define the expression pattern of a gene?



What is the temporal variation of a gene expression pattern?

[Rübel et al., IEEE TCBB, 2009]

Analysis of ChIP-chip Data

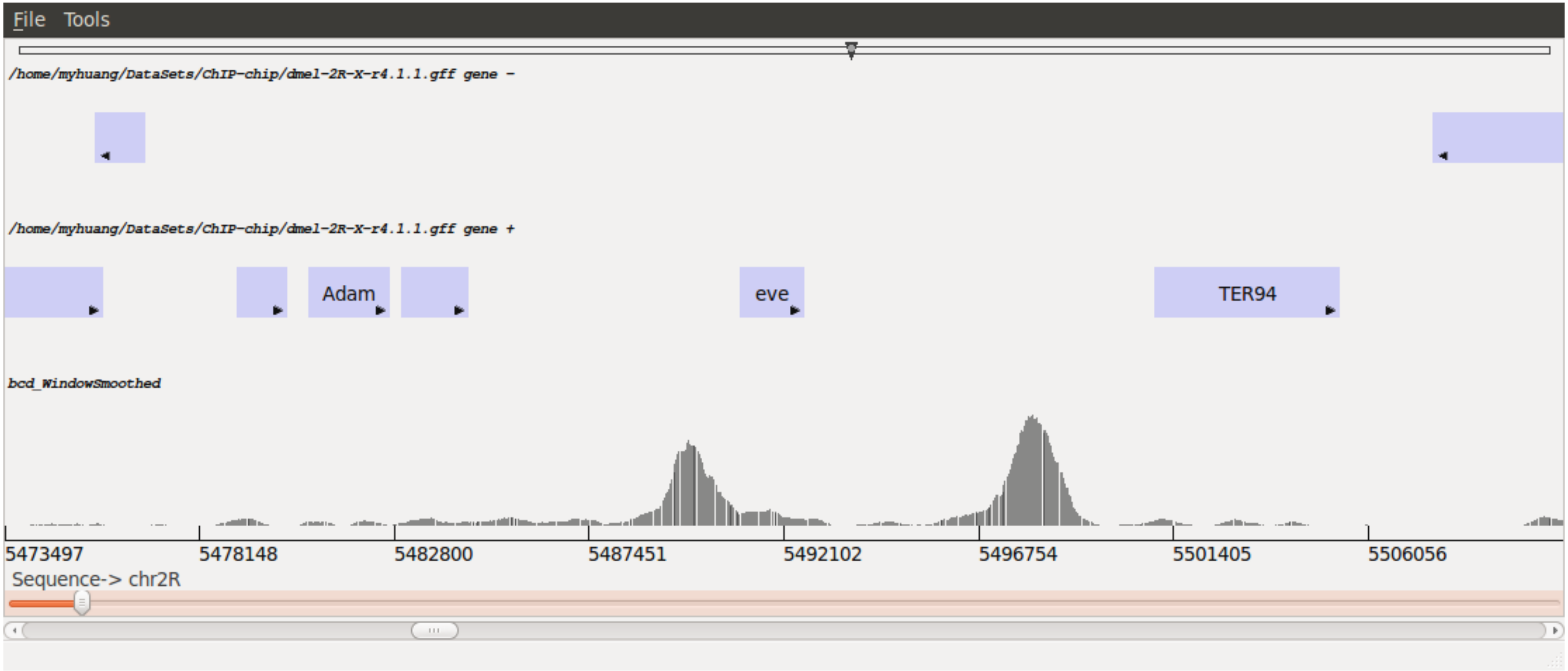
M.-Y. Huang, G.H. Weber, X.-Y. Li, M.D. Biggin, and B. Hamann. *Quantitative Visualization of ChIP-chip Data by Using Linked Views.* In: Proc. IEEE BIBM 2010 Workshops, Workshop on Integrative Data Analysis in Systems Biology (IDASB), pp. 195-200. (2010)

Motivation

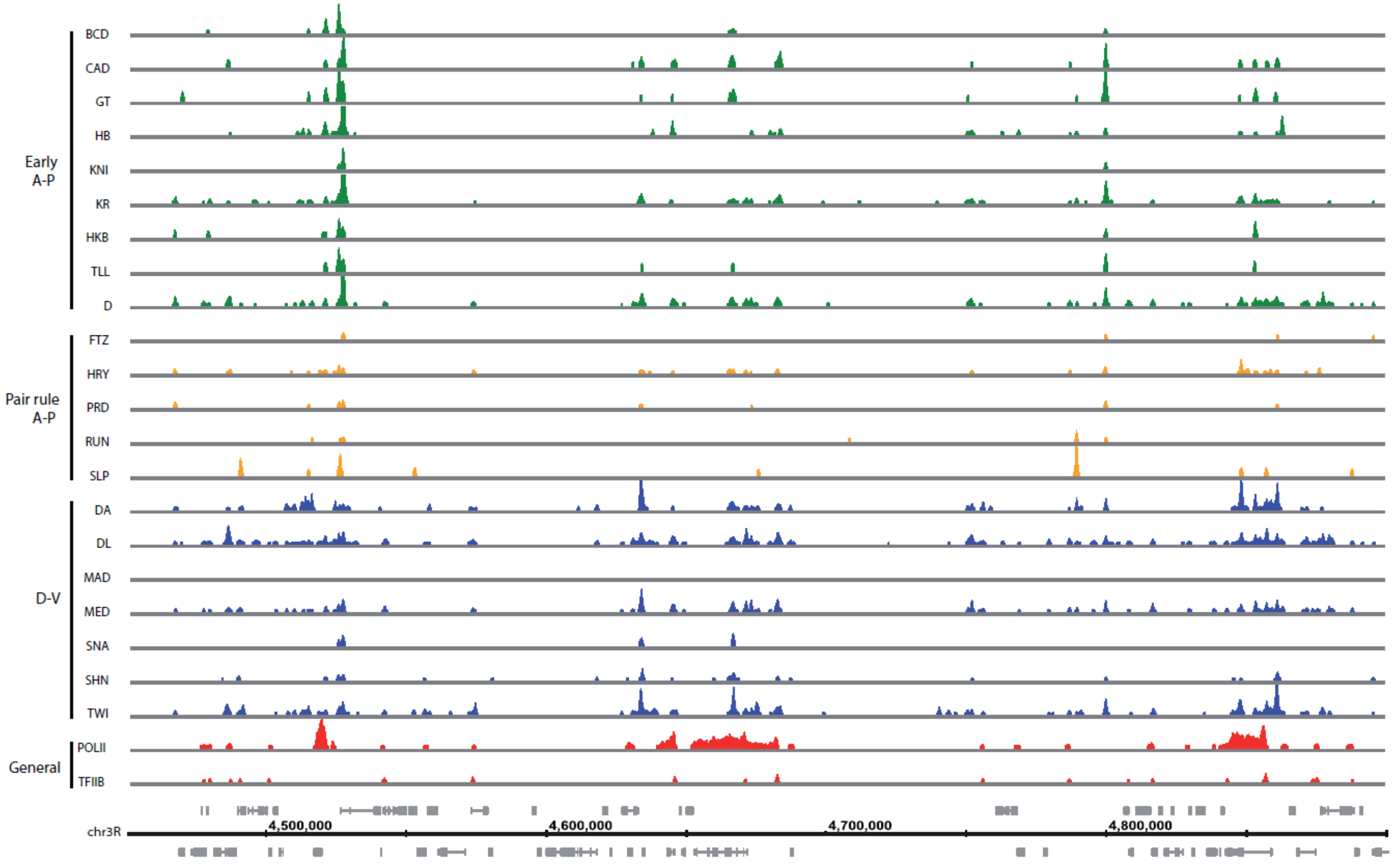
- Most analyses of ChIP-chip *in vivo* binding focused on qualitative description: region bound by certain factor or not
- Increasing evidence suggests that
  - Factors bind in highly overlapping manner on same genomic regions
  - Quantitative differences critical determinants of different biological specificity of factors
- Need tool to facilitate quantitative visualization of differences between transcription factors and genomic regions they bind to understand each factor's unique roles in the network
- Build this tool based on IVA
- Interesting example since no "physical" / 3D views

ChIP-chip Data

- ChIP-chip data describe interaction between protein and DNA *in vivo*, e.g., location of binding



ChIP-chip Data

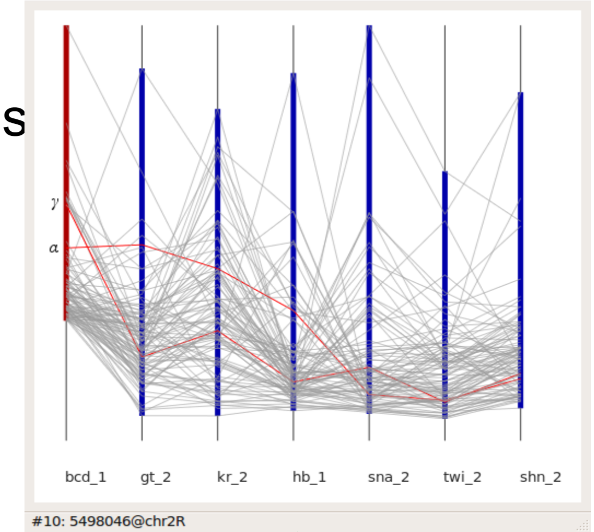


Input and Data Processing

- **Input:**
  - ChIP-chip data
  - Peak locations
- **Processing:**
  - Data smoothing
  - Correlation computation
    1. Choose a transcription factor (TF), fa, as the base
    2. Pick a subset of its primary peaks
    3. Look up corresponding ChIP-chip scores of other TFs at these peak locations
    4. Pick another TF as the new base and repeat step i~iii

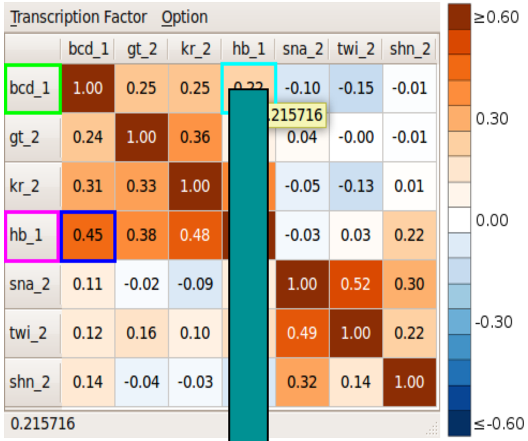
Overview

Parallel Coordinates



#10: 5498046@chr2R

Correlation Table

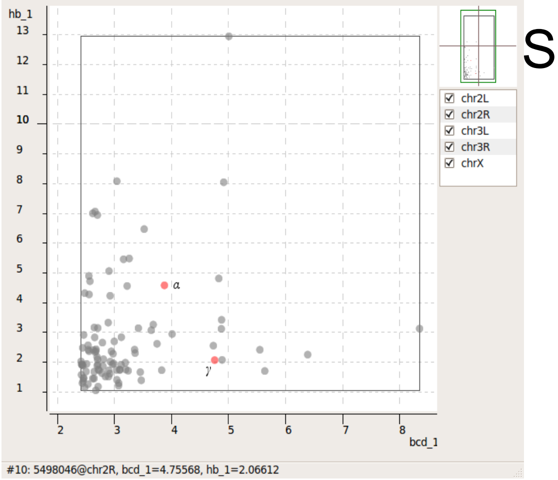


	bcd_1	gt_2	kr_2	hb_1	sna_2	twi_2	shn_2
bcd_1	1.00	0.25	0.25	0.22	-0.10	-0.15	-0.01
gt_2	0.24	1.00	0.36	[illegible]	0.04	-0.00	-0.01
kr_2	0.31	0.33	1.00	[illegible]	-0.05	-0.13	0.01
hb_1	0.45	0.38	0.48	[illegible]	-0.03	0.03	0.22
sna_2	0.11	-0.02	-0.09	[illegible]	1.00	0.52	0.30
twi_2	0.12	0.16	0.10	[illegible]	0.49	1.00	0.22
shn_2	0.14	-0.04	-0.03	[illegible]	0.32	0.14	1.00

215716

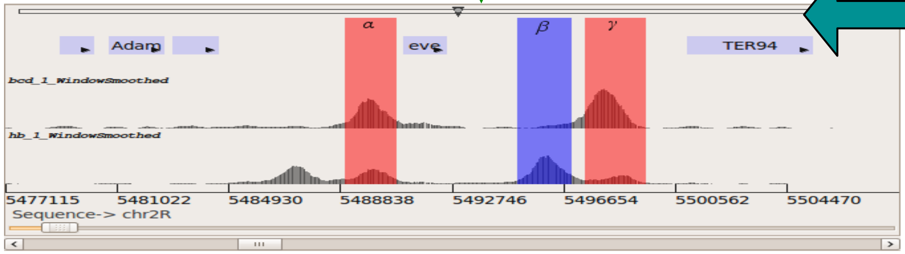
0.215716

Scatter Plots



#10: 5498046@chr2R, bcd_1=4.75568, hb_1=2.06612

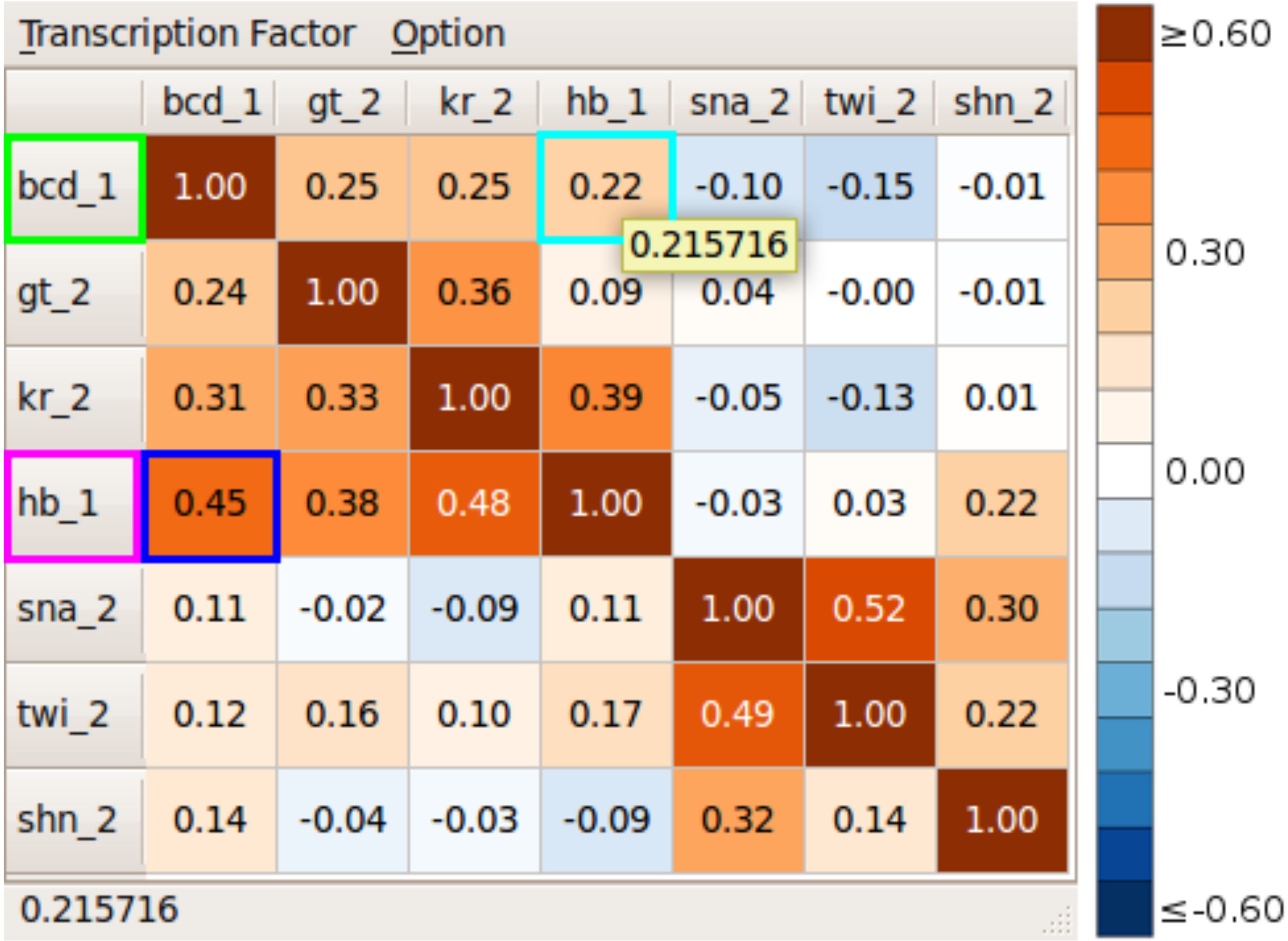
Track View



View creation

Brushing-and-Linking

Correlation table

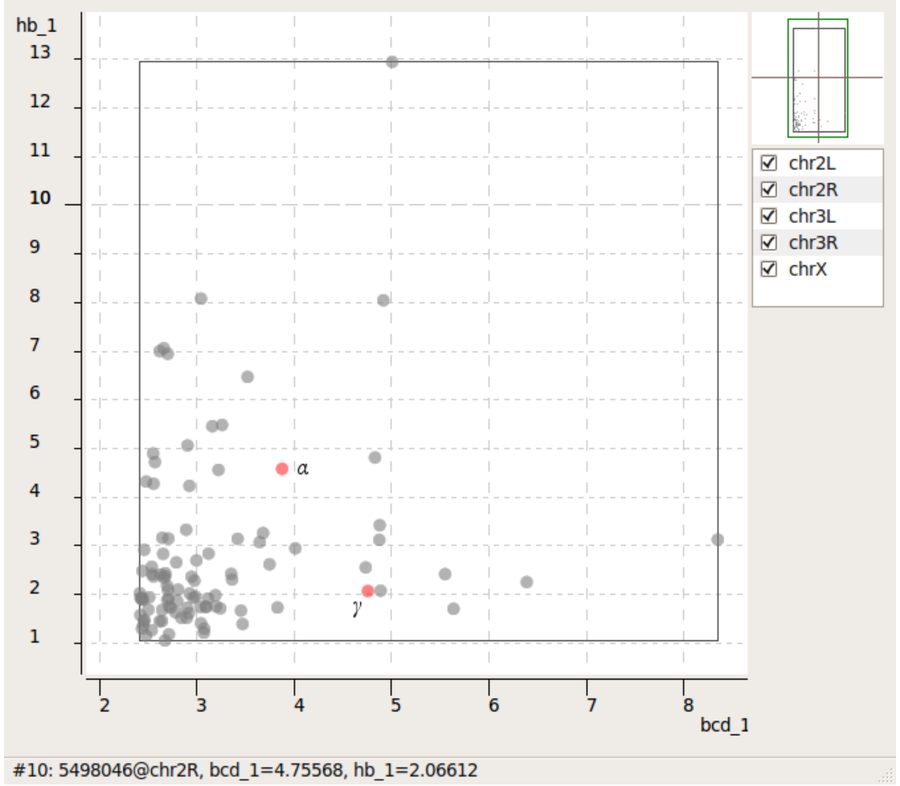


	bcd_1	gt_2	kr_2	hb_1	sna_2	twi_2	shn_2
bcd_1	1.00	0.25	0.25	0.22	-0.10	-0.15	-0.01
gt_2	0.24	1.00	0.36	0.09	0.04	-0.00	-0.01
kr_2	0.31	0.33	1.00	0.39	-0.05	-0.13	0.01
hb_1	0.45	0.38	0.48	1.00	-0.03	0.03	0.22
sna_2	0.11	-0.02	-0.09	0.11	1.00	0.52	0.30
twi_2	0.12	0.16	0.10	0.17	0.49	1.00	0.22
shn_2	0.14	-0.04	-0.03	-0.09	0.32	0.14	1.00

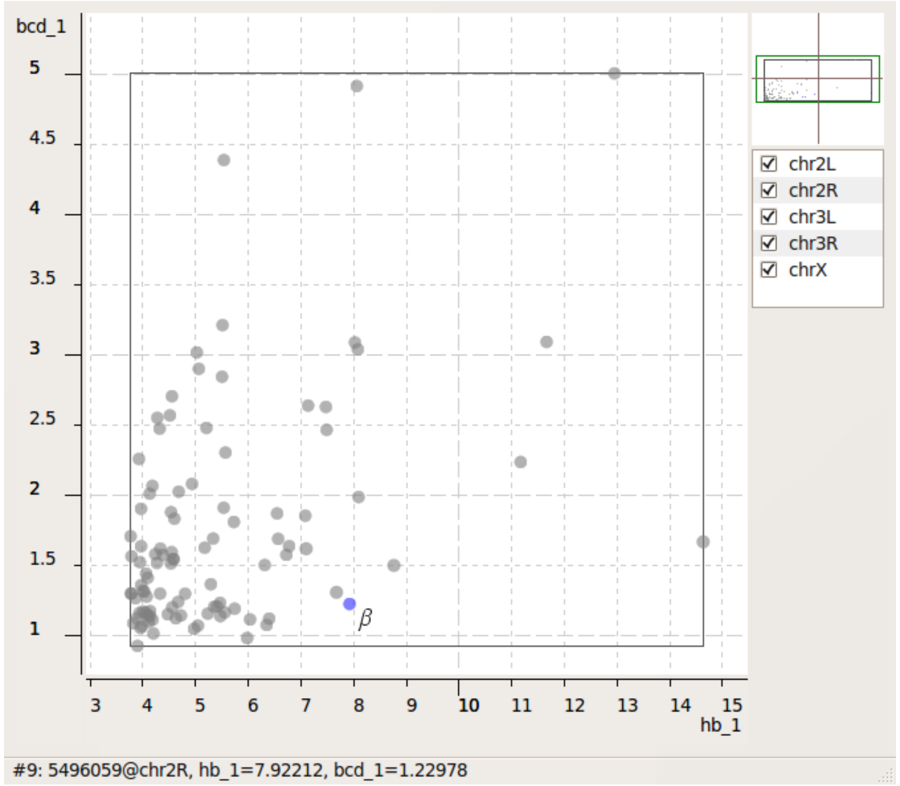
0.215716

0.215716

Scatter Plots

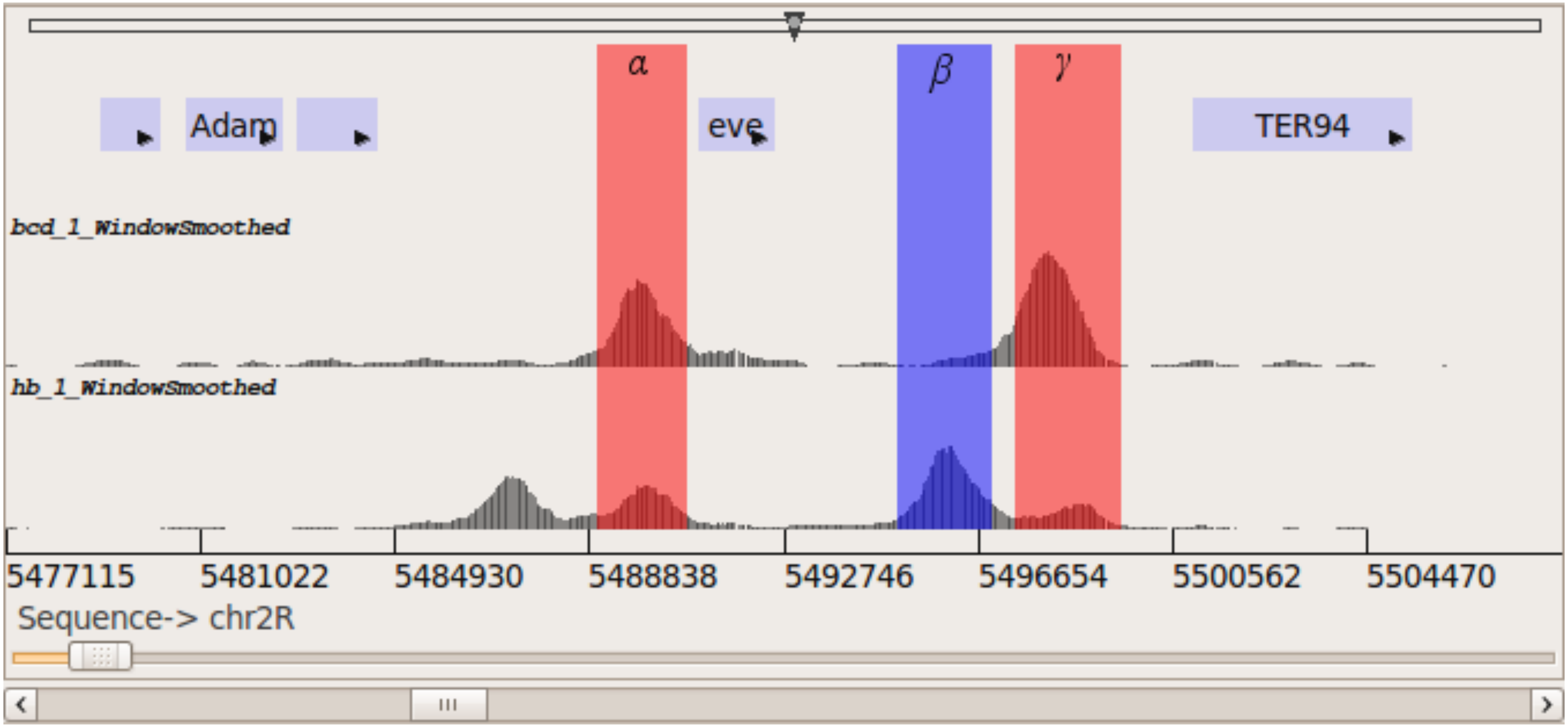


#10: 5498046@chr2R, bcd_1=4.75568, hb_1=2.06612

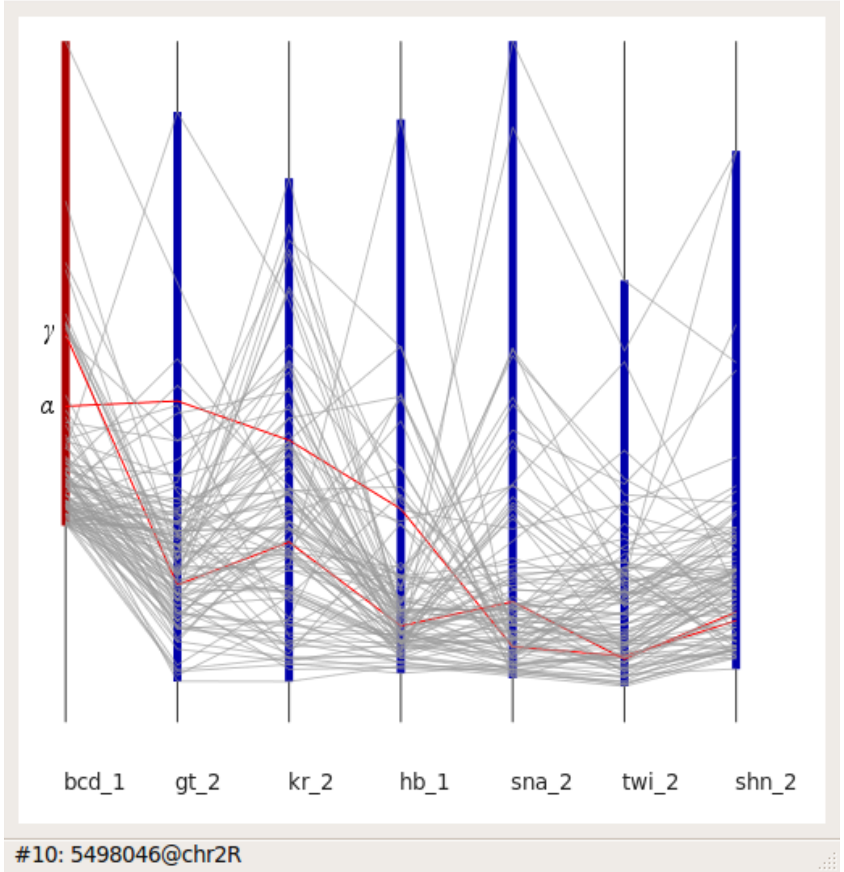


#9: 5496059@chr2R, hb_1=7.92212, bcd_1=1.22978

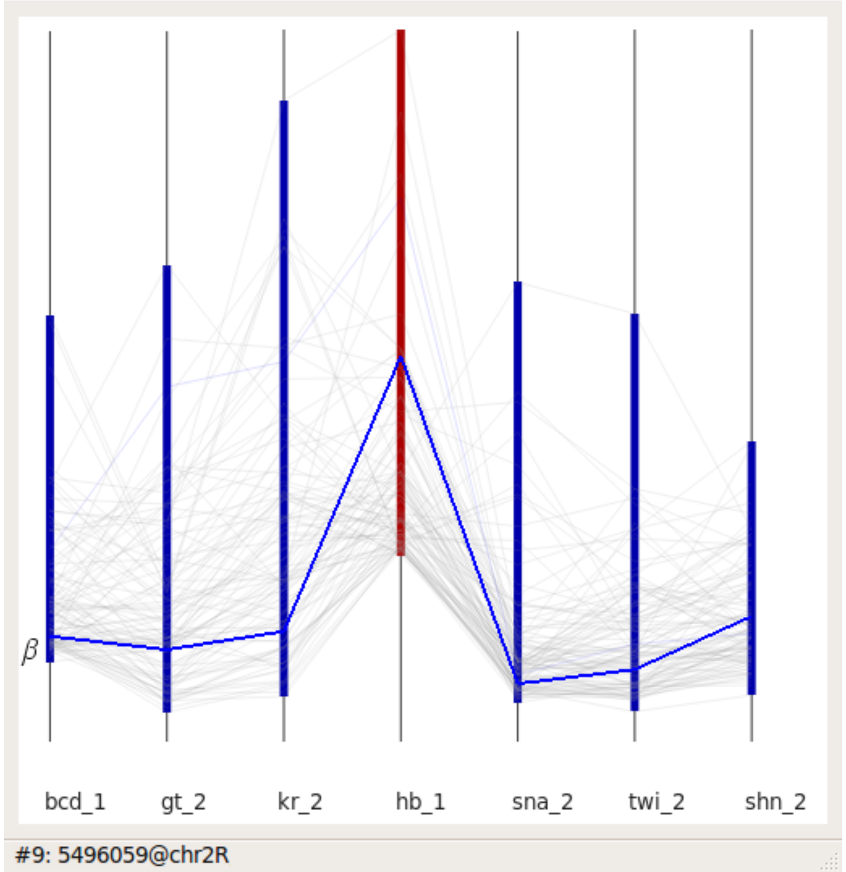
Track View


α
β
γ
Adam
eve
TER94
bcd_1_WindowSmoothed
hb_1_WindowSmoothed
5477115
5481022
5484930
5488838
5492746
5496654
5500562
5504470
Sequence-> chr2R

Parallel Coordinates

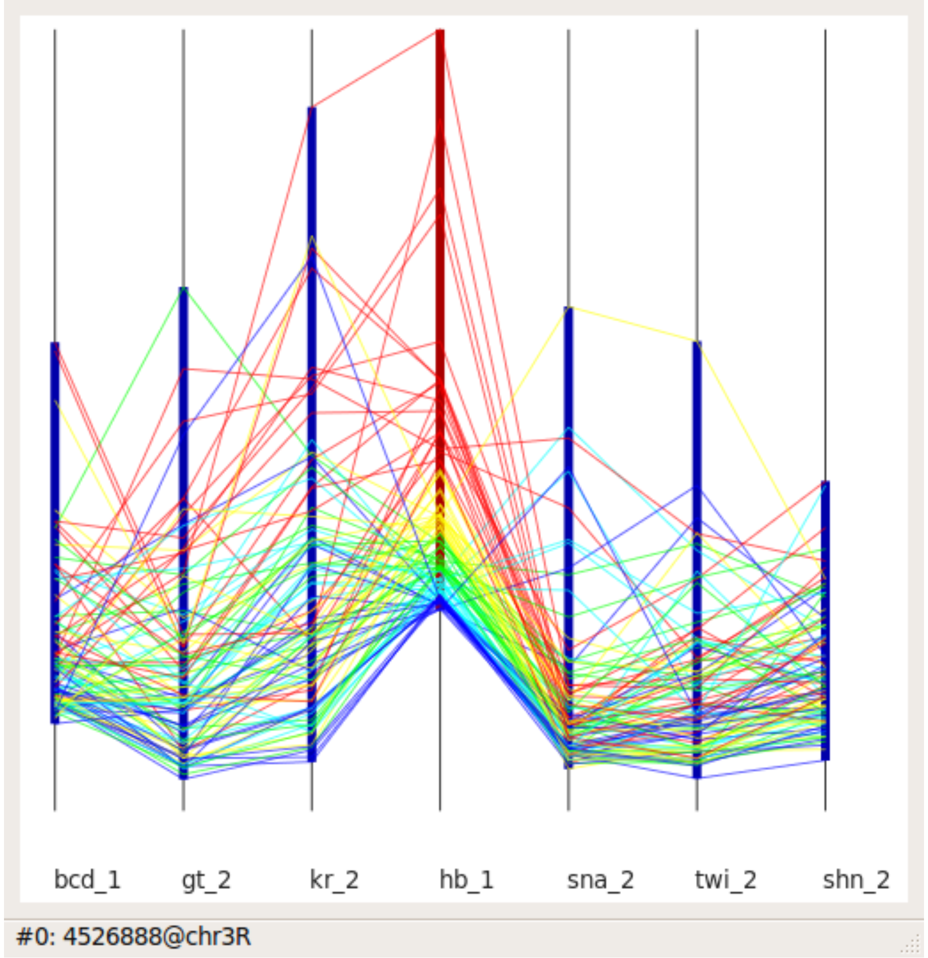


#10: 5498046@chr2R



#9: 5496059@chr2R

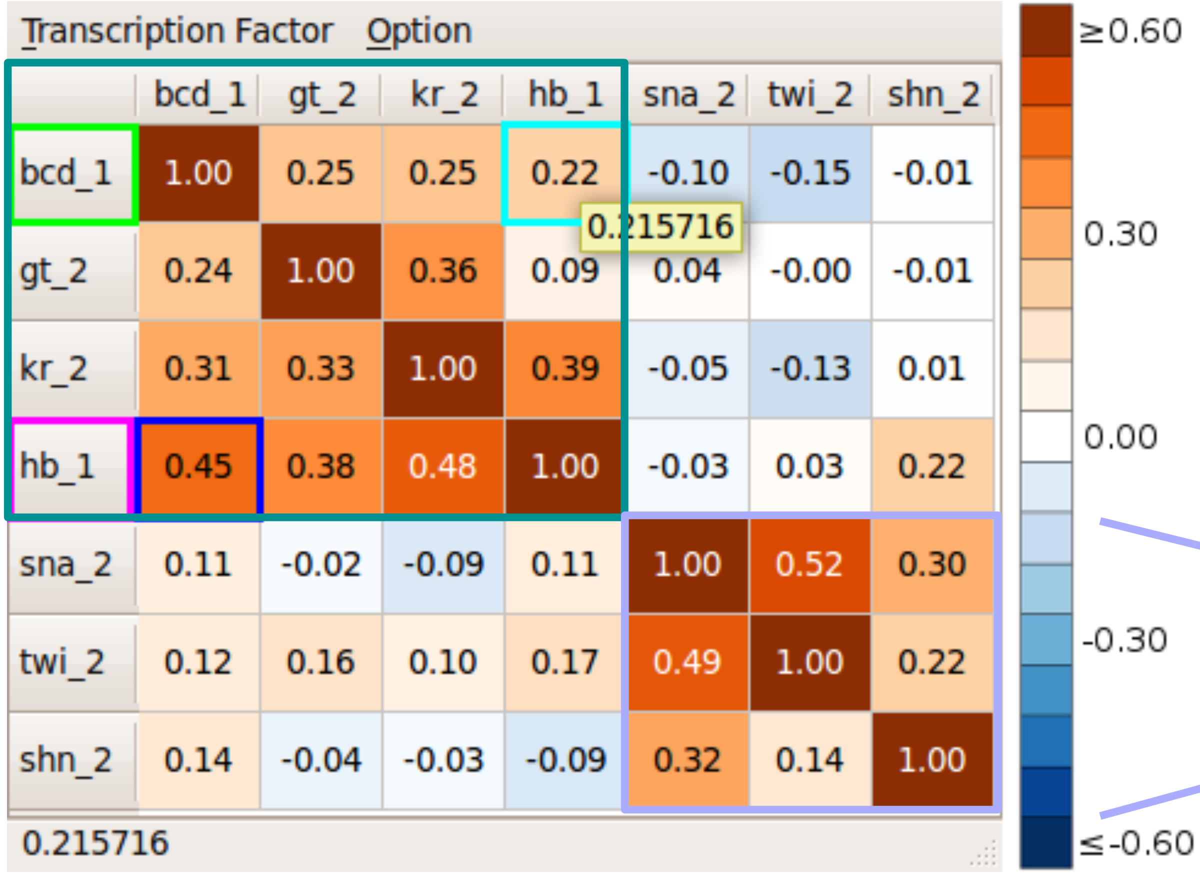
Parallel Coordinates



#0: 4526888@chr3R

Case Study (1/5)

A-P Factors



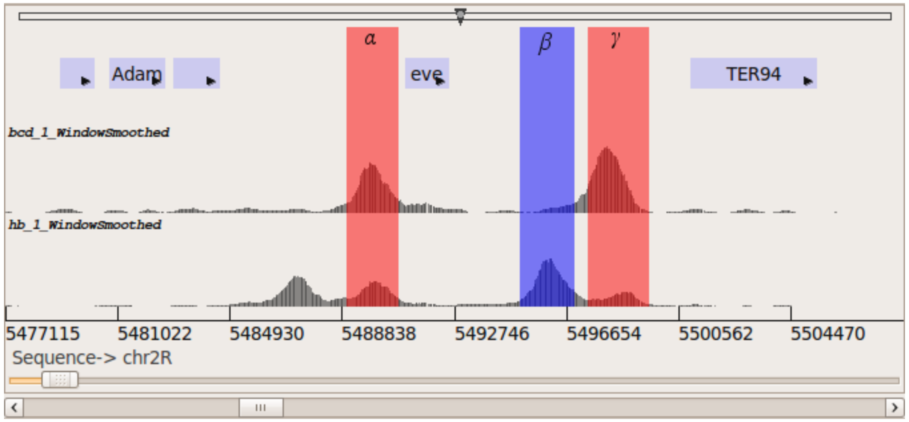
	bcd_1	gt_2	kr_2	hb_1	sna_2	twi_2	shn_2
bcd_1	1.00	0.25	0.25	0.22	-0.10	-0.15	-0.01
gt_2	0.24	1.00	0.36	0.09	0.04	-0.00	-0.01
kr_2	0.31	0.33	1.00	0.39	-0.05	-0.13	0.01
hb_1	0.45	0.38	0.48	1.00	-0.03	0.03	0.22
sna_2	0.11	-0.02	-0.09	0.11	1.00	0.52	0.30
twi_2	0.12	0.16	0.10	0.17	0.49	1.00	0.22
shn_2	0.14	-0.04	-0.03	-0.09	0.32	0.14	1.00

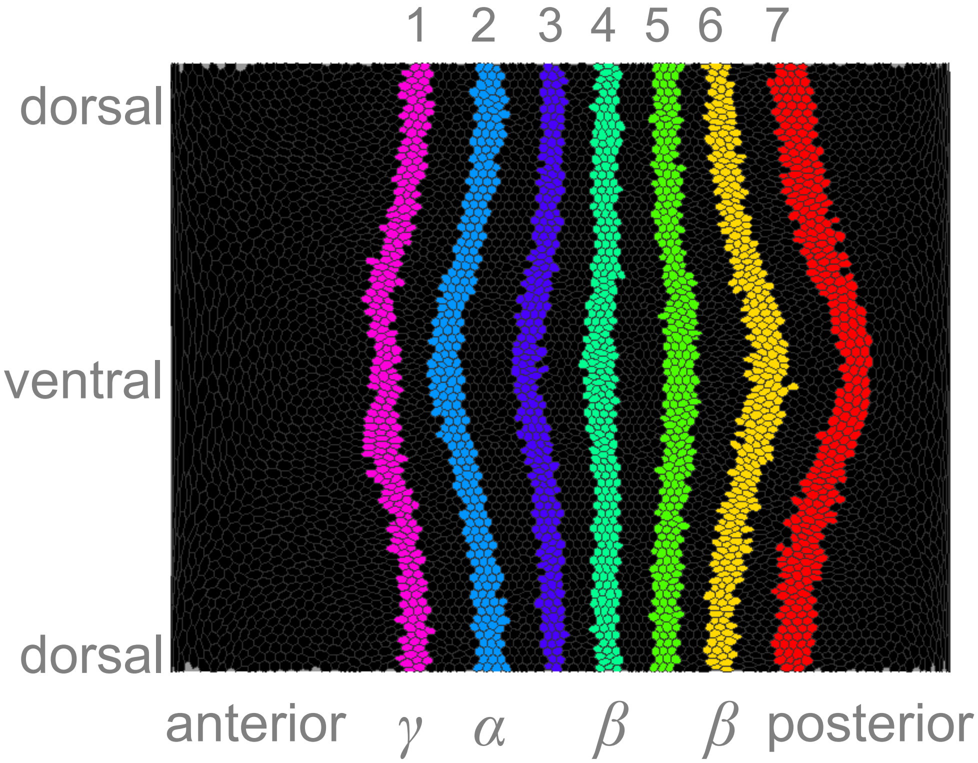
0.215716

D-V Factors

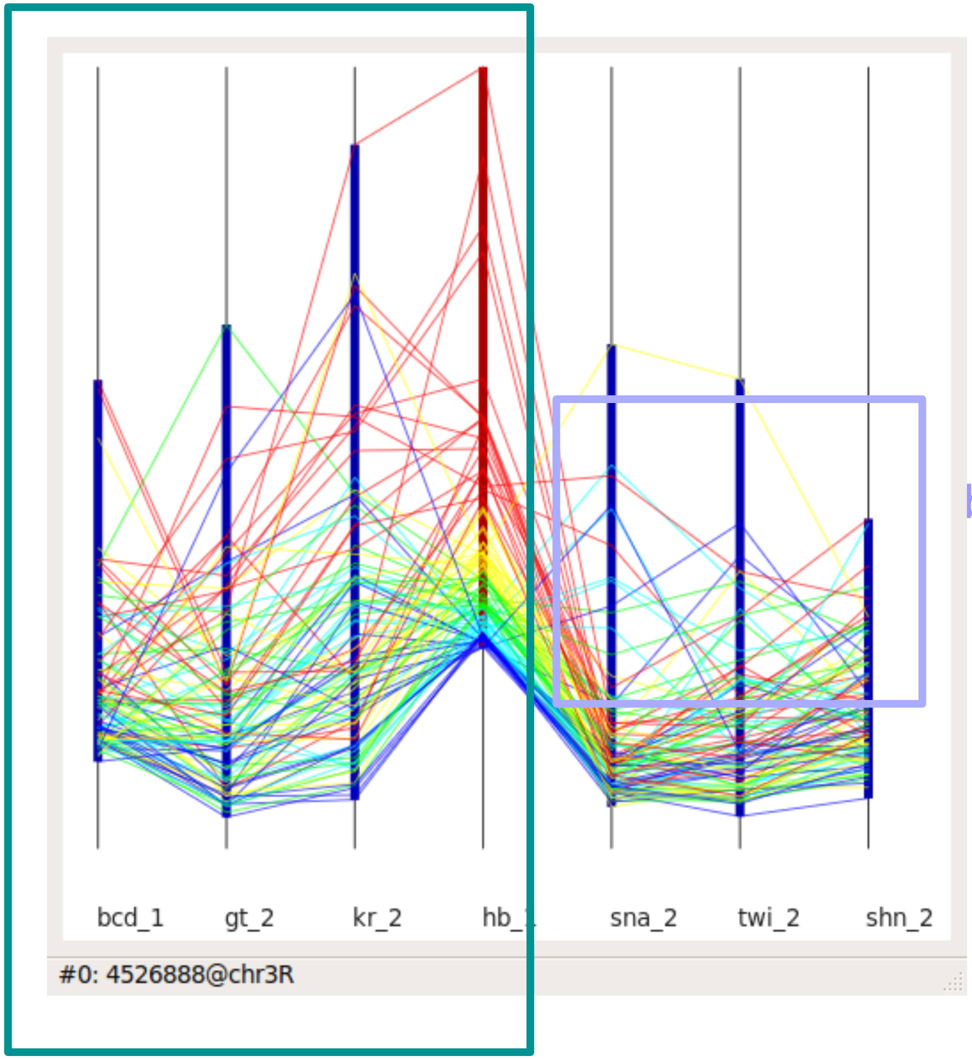
Case Study (2/5)

- *Eve*
  - α: stripe 2, *bcd* & *hb* activate
  - β: stripe 4/6, *hb* represses
  - γ: stripe 1, *bcd* activates





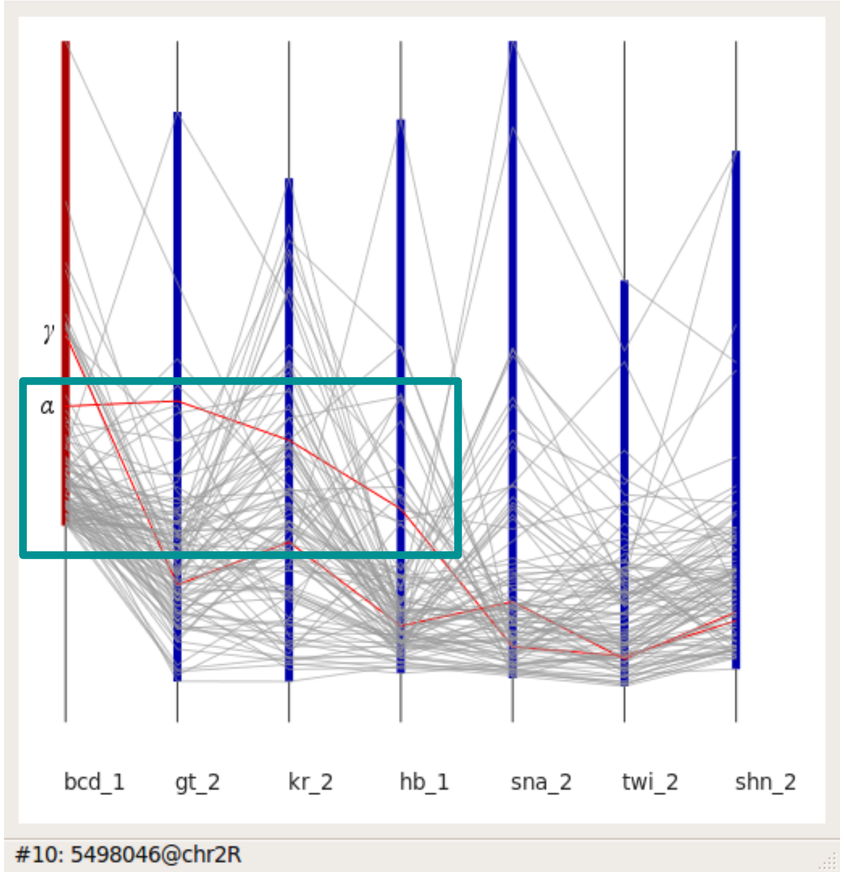
Case Study (3/5)



Higher correlation with other A-P factors

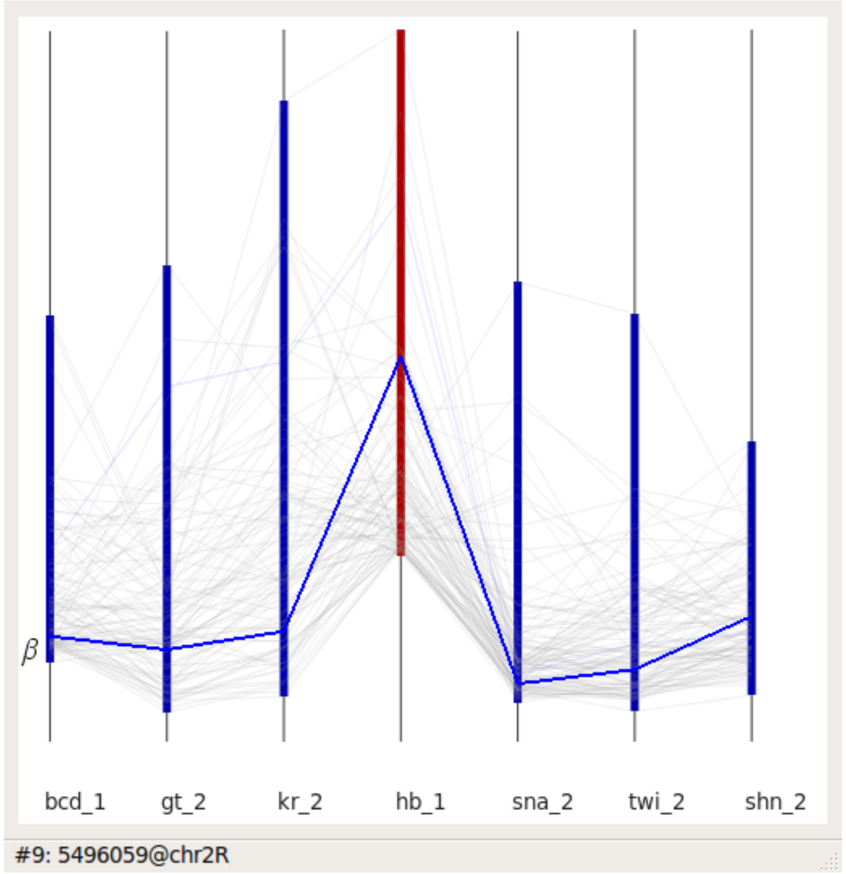
Still have regions bound to high *hb* and moderate D-V factors

Case Study (4/5)



- α is bound strongly not only by bcd and hb, but also by kr and gt (two factors known to be important for defining the expression boundaries of eve stripe 2)
- γ (eve stripe 1 enhancer) is only strongly bound by bcd

Case Study (5/5)



#9: 5496059@chr2R

- β (*eve* stripe 4/6) is only strongly bound by *hb*

Literature

- O. Rübel, S.V.E. Keränen, M.D. Biggin, D.W. Knowles, G.H. Weber, H. Hagen, B. Hamann, and E.W. Bethel, "Linking Advanced Visualization and MATLAB for the Analysis of 3D Gene Expression Data," Series: Mathematics and Visualization, Visualization in Medicine and Life Sciences II, Progress and New Challenges, Springer Verlag, Heidelberg, Germany, 2012, pp 267-285. LBNL-4891E.
- Min-Yu Huang, Gunther H. Weber, Xiao-Yong Li, Mark D. Biggin, and Bernd Hamann. "Quantitative Visualization of ChIP-chip Data by Using Linked Views." In: Proceedings IEEE International Conference on Bioinformatics and Biomedicine 2010 (IEEE BIBM 2010) Workshops, Workshop on Integrative Data Analysis in Systems Biology (IDASB), pp. 195-200. LBNL-4491E.
- O. Rübel, G. H. Weber, M-Y Huang, E. W. Bethel, M. D. Biggin, C. C. Fowlkes, C. Luengo Hendriks, S. V. E. Keränen, M. Eisen, D. Knowles, J. Malik, H. Hagen and B. Hamann, "Integrating Data Clustering and Visualization for the Analysis of 3D Gene Expression Data." IEEE Transactions on Computational Biology and Bioinformatics, Vol 7, No.1, Pages 64-79, Jan/Mar 2010. LBNL-382E.
- G. H. Weber, O. Rübel, M.-Y. Huang, A. H. DePace, C. C. Fowlkes, S. V. E. Keränen, C. L. Luengo Hendriks, H. Hagen, D. W. Knowles, J. Malik, M. D. Biggin and B. Hamann. "Visual exploration of three-dimensional gene expression using physical views and linked abstract views." In *IEEE Transactions on Computational Biology and Bioinformatics.*, 6(2), April-June, pp. 296-309, 2009. LBNL-63776.
- O. Rübel, G.H. Weber, S.V.E. Keränen, C.C. Fowlkes, C.L. Luengo Hendriks, L. Simirenko, N.Y. Shah, M.B. Eisen, M.D. Biggin, H. Hagen, J.D. Sudar, J. Malik, D.W. Knowles, and B. Hamann. PointCloudXplore: Visual analysis of 3D gene expression data using physical views and parallel coordinates. In: B. Sousa Santos, T. Ertl, and K.I. Joy, eds., *Data Visualization 2006 (Proceedings of EuroVis 2006)*, Eurographics Association, Aire-la-Ville, Switzerland, pp. 203-210. May, 2006. LBNL-60005.