Visualization Analysis & Design Full-Day Tutorial Session 4

Tamara Munzner
Department of Computer Science
University of British Columbia

Sanger Institute / European Bioinformatics Institute
June 2014, Cambridge UK

http://www.cs.ubc.ca/~tmm/talks.html#minicourse14

Outline

- Visualization Analysis Framework
  Session 1 *9:30-10:45am*
  - Introduction: Definitions
  - Analysis: What, Why, How
  - Marks and Channels
- Idiom Design Choices, Part 2
  Session 3 *1:15pm-2:45pm*
  - Manipulate: Change, Select, Navigate
  - Facet: Juxtapose, Partition, Superimpose
  - Reduce: Filter, Aggregate, Embed
- Idiom Design Choices
  Session 2 *11:00am-12:15pm*
  - Arrange Tables
  - Arrange Spatial Data
  - Arrange Networks and Trees
  - Map Color
- Guidelines and Examples
  Session 4 *3-4:30pm*
  - Rules of Thumb
  - Validation
  - BioVis Analysis Example

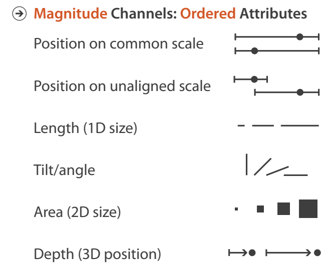
http://www.cs.ubc.ca/~tmm/talks.html#minicourse14

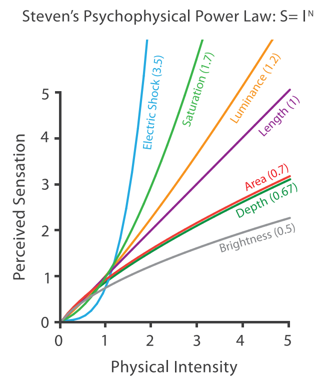
Rules of Thumb

- No unjustified 3D
  - Power of the plane, dangers of depth
  - Occlusion hides information
  - Perspective distortion loses information
  - Tilted text isn't legible
- No unjustified 2D
- Eyes beat memory
- Resolution over immersion
- Overview first, zoom and filter, details on demand
- Function first, form next

No unjustified 3D: Power of the plane

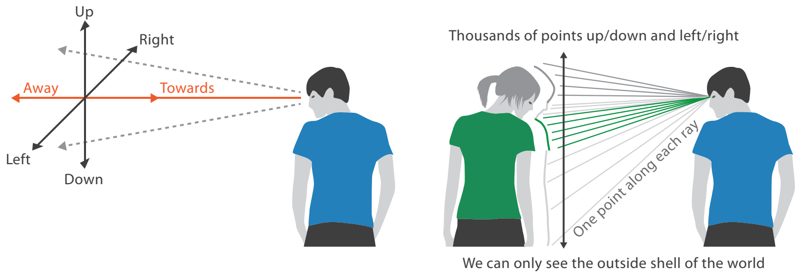
- high-ranked spatial position channels: **planar** spatial position
  - not depth!





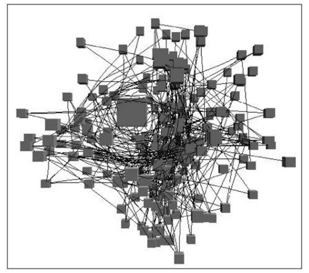
No unjustified 3D: Danger of depth

- we don't really live in 3D: we **see** in 2.05D
  - acquire more info on image plane quickly from eye movements
  - acquire more info for depth slower, from head/body motion



Occlusion hides information

- occlusion
- interaction complexity



[Distortion Viewing Techniques for 3D Data. Carpendale et al. InfoVis1996.]

Perspective distortion loses information

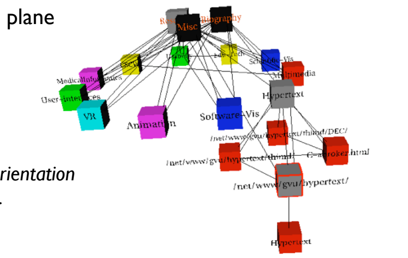
- perspective distortion
  - interferes with all size channel encodings
  - power of the plane is lost!

[Visualizing the Results of Multimedia Web Search Engines. Mukherjea, Hirata, and Hara. InfoVis 96]

Tilted text isn't legible

- text legibility
  - far worse when tilted from image plane
- further reading

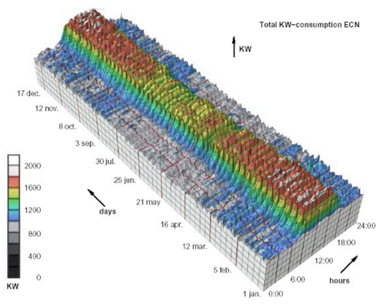
[Exploring and Reducing the Effects of Orientation on Text Readability in Volumetric Displays. Grossman et al. CHI 2007]



[Visualizing the World-Wide Web with the Navigational View Builder. Mukherjea and Foley. Computer Networks and ISDN Systems, 1995.]

No unjustified 3D example: Time-series data

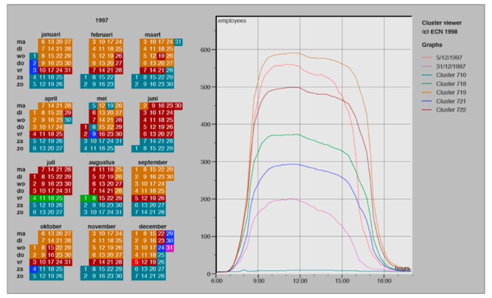
- extruded curves: detailed comparisons impossible



[Cluster and Calendar based Visualization of Time Series Data. van Wijk and van Selow, Proc. InfoVis 99.]

No unjustified 3D example: Transform for new data abstraction

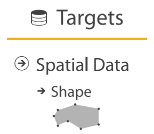
- derived data: cluster hierarchy
- juxtapose multiple views: calendar, superimposed 2D curves

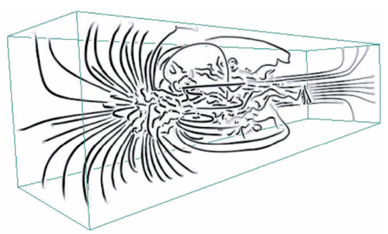


[Cluster and Calendar based Visualization of Time Series Data. van Wijk and van Selow, Proc. InfoVis 99.]

Justified 3D: shape perception

- benefits outweigh costs when task is shape perception for 3D spatial data
  - interactive navigation supports synthesis across many viewpoints

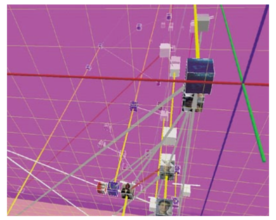




[Image-Based Streamline Generation and Rendering. Li and Shen. IEEE Trans. Visualization and Computer Graphics (TVCG) 13:3 (2007), 630–640.]

No unjustified 3D

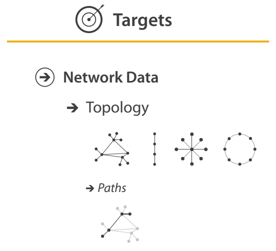
- 3D legitimate for true 3D spatial data
- 3D needs very careful justification for abstract data
  - enthusiasm in 1990s, but now skepticism
  - be especially careful with 3D for point clouds or networks



[WEBPATH-a three dimensional Web history. Frecon and Smith. Proc. InfoVis 1999]

No unjustified 2D

- consider whether network data requires 2D spatial layout
  - especially if reading text is central to task!
  - arranging as network means lower information density and harder label lookup compared to text lists
- benefits outweigh costs when topological structure/context important for task
  - be especially careful for search results, document collections, ontologies



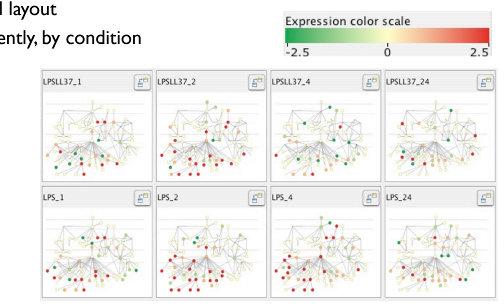
Eyes beat memory

- principle: external cognition vs. internal memory
  - easy to compare by moving eyes between side-by-side views
  - harder to compare visible item to memory of what you saw
- implications for animation
  - great for choreographed storytelling
  - great for transitions between two states
  - poor for many states with changes everywhere
    - consider small multiples instead

literal	abstract
animation	small multiples
show time with time	show time with space

Eyes beat memory example: Cerebral

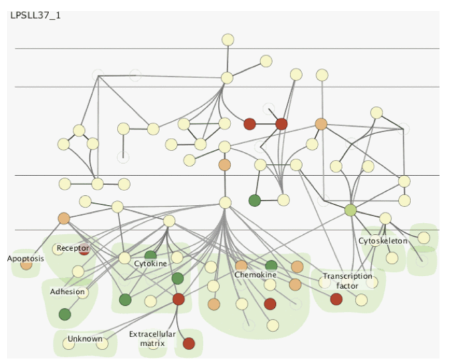
- small multiples: one graph instance per experimental condition
  - same spatial layout
  - color differently, by condition



[Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context. Barsky, Munzner, Gardy, and Kincaid. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 2008) 14:6 (2008), 1253–1260.]

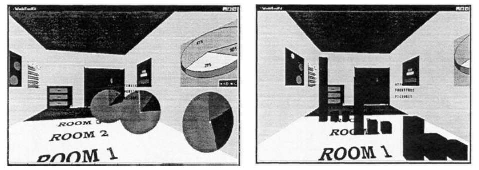
Why not animation?

- disparate frames and regions: comparison difficult
  - vs contiguous frames
  - vs small region
  - vs coherent motion of group
- change blindness
  - even major changes difficult to notice if mental buffer wiped
- safe special case
  - animated transitions



Resolution beats immersion

- immersion typically not helpful for abstract data
  - do not need sense of presence or stereoscopic 3D
- resolution much more important
  - pixels are the scarcest resource
  - desktop also better for workflow integration
- virtual reality for abstract data very difficult to justify



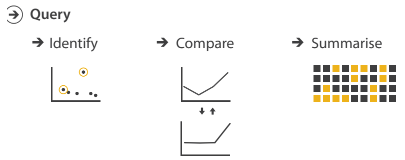
[Development of an information visualization tool using virtual reality. Kirner and Martins. Proc. Symp. Applied Computing 2000]

Overview first, zoom and filter, details on demand

- influential mantra from Shneiderman

[The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. Shneiderman. Proc. IEEE Visual Languages, pp. 336–343, 1996.]

- overview = summary
  - microcosm of full vis design problem
- nuances
  - beyond just two levels: multi-scale structure
  - difficult when scale huge: give up on overview and browse local neighborhoods?



[Search, Show Context, Expand on Demand: Supporting Large Graph Exploration with Degree-of-Interest. van Ham and Perer. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 2009) 15:6 (2009), 953–960.]

Function first, form next

- start with focus on functionality
  - straightforward to improve aesthetics later on, as refinement
  - if no expertise in-house, find good graphic designer to work with
- dangerous to start with aesthetics
  - usually impossible to add function retroactively

Further reading

- Visualization Analysis and Design. Munzner. AK Peters / CRC Press, Oct 2014.
  - Chap 6: Rules of Thumb
- Visual Thinking for Design. Ware. Morgan Kaufmann, 2008.
- Information Visualization: Perception for Design, 3rd edition. Ware. Morgan Kaufmann /Academic Press, 2004.
- *The use of 2-D and 3-D displays for shape understanding versus relative position tasks.* St. John, Cowen, Smallman, and Oonk. Human Factors 43:1 (2001), 79–98.
- *Evaluating Spatial Memory in Two and Three Dimensions.* Cockburn and McKenzie. Intl. Journal of Human-Computer Studies 61:30 (2004), 359–373.
- *Supporting and Exploiting Spatial Memory in User Interfaces.* Scarr, Cockburn, and Gutwin. Foundations and Trends in Human Computer Interaction, 6. Now, 2013.
- *Effectiveness of Animation in Trend Visualization.* Robertson, Fernandez, Fisher, Lee, and Stasko. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis08) 14:6 (2008), 1325–1332.
- *Animation: can it facilitate?* Tversky, Morrison and Betrancourt. Intl Journ Human-Computer Studies, 57(4): 247-262, 2002.
- *Current approaches to change blindness.* Simons. Visual Cognition 7:1/2/3 (2000), 1–15.
- The Non-Designer's Design Book, 3rd ed. Williams. Peachpit Press, 2008.

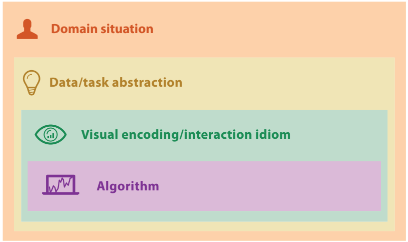
Outline

- **Visualization Analysis Framework**
  Session 1 *9:30-10:45am*
  - Introduction: Definitions
  - Analysis: What, Why, How
  - Marks and Channels
- **Idiom Design Choices, Part 2**
  Session 3 *1:15pm-2:45pm*
  - Manipulate: Change, Select, Navigate
  - Facet: Juxtapose, Partition, Superimpose
  - Reduce: Filter, Aggregate, Embed
- **Idiom Design Choices**
  Session 2 *11:00am-12:15pm*
  - Arrange Tables
  - Arrange Spatial Data
  - Arrange Networks and Trees
  - Map Color
- **Guidelines and Examples**
  Session 4 *3-4:30pm*
  - Rules of Thumb
  - Validation
  - BioVis Analysis Example

http://www.cs.ubc.ca/~tmm/talks.html#minicourse14

Four Levels of Design

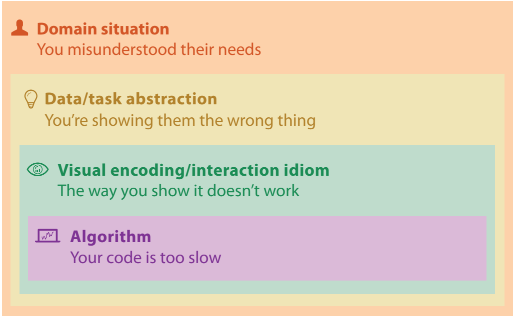
- two more levels to consider
  - domain problem: all aspects of user context
  - algorithm: efficient implementation of idioms



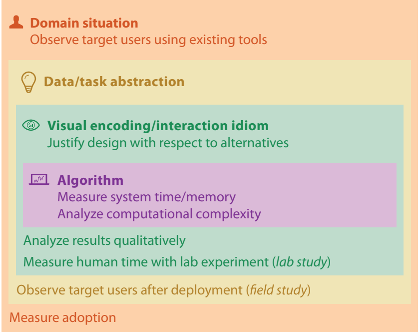
[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Four Levels of Design and Validation

- four levels of design problems
  - different threats to validity at each level



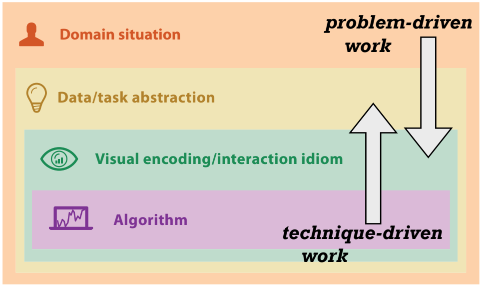
Nested Levels of Design and Validation

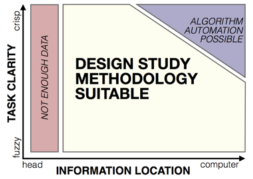


- mismatch: cannot show idiom good with system timings
- mismatch: cannot show abstraction good with lab study

Four Levels of Design

- two more levels to consider
  - domain problem: all aspects of user context
  - algorithm: efficient implementation of idioms





Design Study Methodology

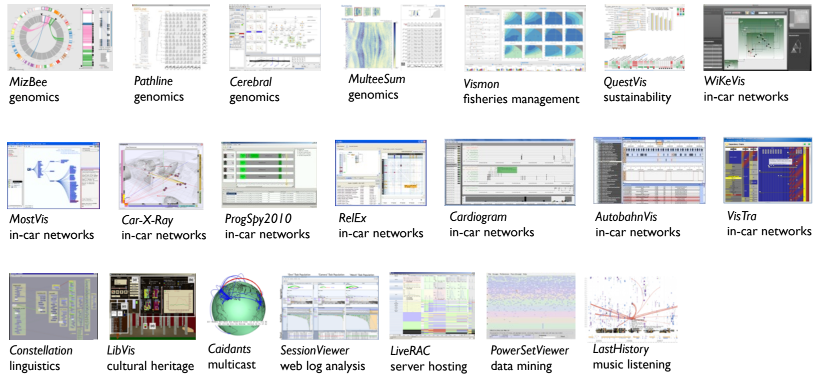
Reflections from the Trenches and from the Stacks

joint work with:
Michael Sedlmair, Miriah Meyer

http://www.cs.ubc.ca/labs/imager/tr/2012/dsm/

Design Study Methodology: Reflections from the Trenches and from the Stacks.
Sedlmair, Meyer, Munzner. IEEE TVCG 18(12): 2431-2440, 2012 (Proc. InfoVis 2012).

Design Studies: Lessons learned after 21 of them

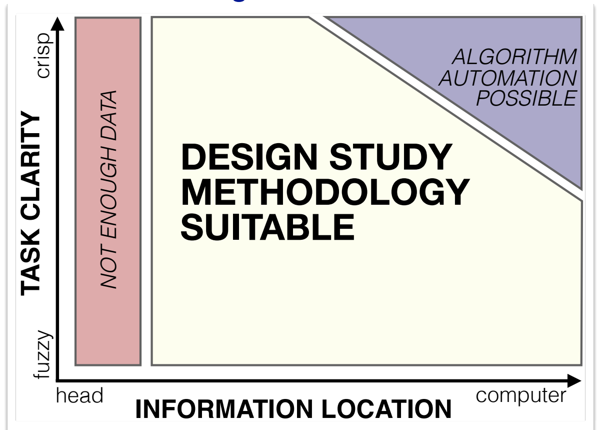


- commonality of representations cross-cuts domains!

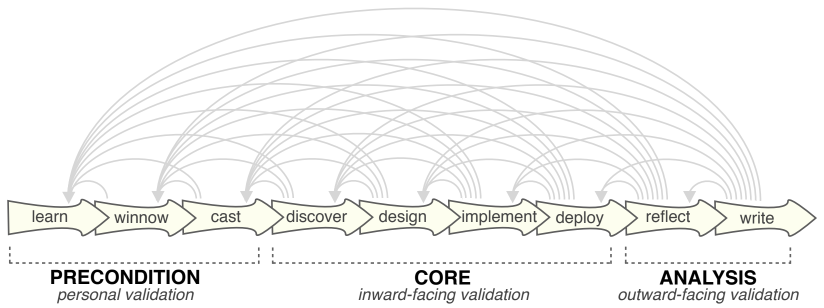
Design studies: problem-driven vis research

- a specific real-world problem
  - real users and real data,
  - collaboration is (often) fundamental
- design a visualization system
  - implications: requirements, multiple ideas
- validate the design
  - at appropriate levels
- reflect about lessons learned
  - transferable research: improve design guidelines for vis in general
    - confirm, refine, reject, propose

When To Do Design Studies

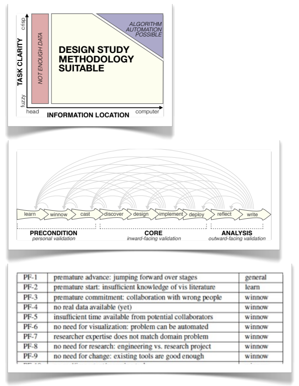


Nine-Stage Framework



How To Do Design Studies

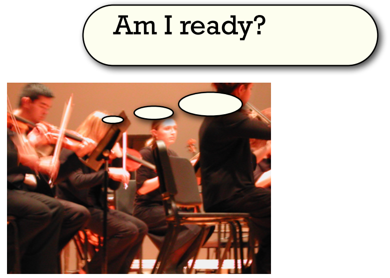
- definitions
- 9-stage framework
- 32 pitfalls and how to avoid them



Pitfall Example: Premature Publishing

algorithm innovation

design studies



http://www.prlog.org/10480334-wolverhampton-horse-racing-live-streaming-wolverhampton-handicap-8-jan-2010.html

http://www.alaineknipes.com/interests/violin_concert.jpg

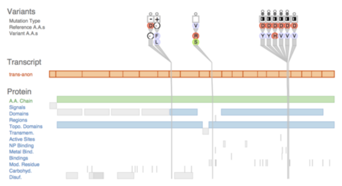
Further reading

- Visualization Analysis and Design. Munzner. AK Peters / CRC Press, Oct 2014.
  - *Chap 4: Analysis: Four Levels for Validation*
- *A Nested Model of Visualization Design and Validation.* Munzner. IEEE TVCG 15(6): 921-928, 2009 (Proc. InfoVis 2009).
- *Design Study Methodology: Reflections from the Trenches and from the Stacks.* Sedlmair, Meyer, Munzner. IEEE TVCG 18(12): 2431-2440, 2012 (Proc. InfoVis 2012).

Outline

- Visualization Analysis Framework
  Session 1 *9:30-10:45am*
  - Introduction: Definitions
  - Analysis: What, Why, How
  - Marks and Channels
- Idiom Design Choices, Part 2
  Session 3 *1:15pm-2:45pm*
  - Manipulate: Change, Select, Navigate
  - Facet: Juxtapose, Partition, Superimpose
  - Reduce: Filter, Aggregate, Embed
- Idiom Design Choices
  Session 2 *11:00am-12:15pm*
  - Arrange Tables
  - Arrange Spatial Data
  - Arrange Networks and Trees
  - Map Color
- Guidelines and Examples
  Session 4 *3-4:30pm*
  - Rules of Thumb
  - Validation
  - BioVis Analysis Example

http://www.cs.ubc.ca/~tmm/talks.html#minicourse14



Variant View

Visualizing Sequence Variants in their Gene Context

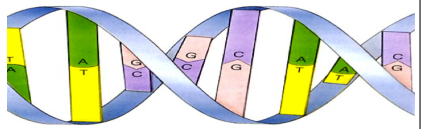
joint work with:
Joel Ferstay, Cydney Nielsen

http://www.cs.ubc.ca/labs/imager/tr/2012/VariantView/

Variant View: Visualizing Sequence Variants in their Gene Context.
Ferstay, Nielsen, Munzner. IEEE TVCG 19(12): 2546-2555, 2013 (Proc. InfoVis 2013).

Sequence Variant Definition

- Sequence variants
  - Difference between reference and given genome



Reference Genome DNA: ATA TGA TCA ACA CTT

Sample 1 Genome DNA: ATA TGG TCA ATA CTT — Harmful?

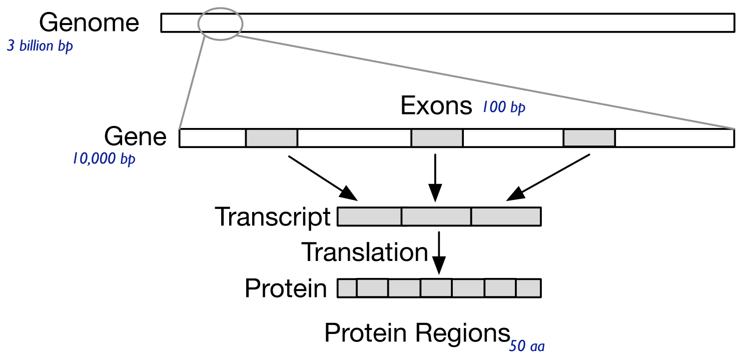
Sample 2 Genome DNA: ATA TGA TGA ACA CCT — Harmless?

Cancer Research

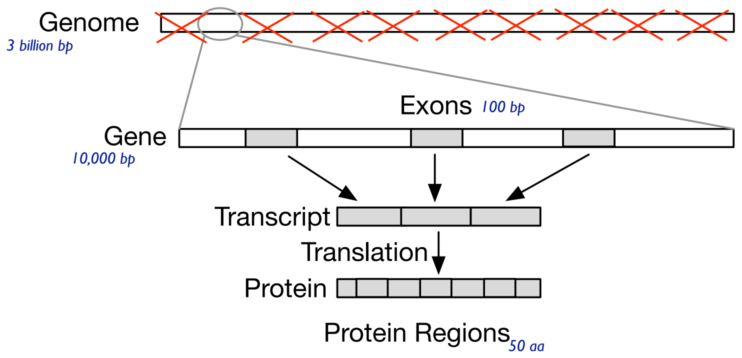
- collaboration with analysts at BC Genome Sciences Center
  - studying genetic basis of leukemia
- driving task
  - discover new candidate genes with harmful variants
- two big questions
  - what to show
    - data abstraction
    - challenge: enormous range of scales in the data
  - how to show it
    - visual encoding idiom

Abstractions

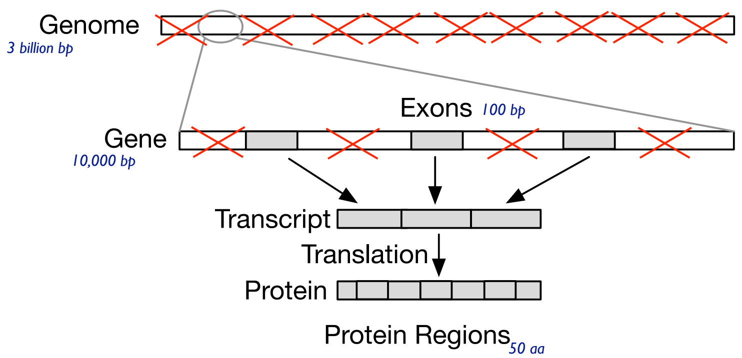
Data: Filtering to relevant biological levels and scales



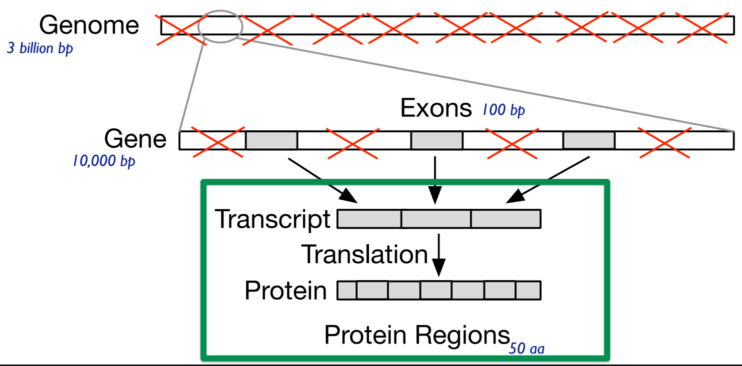
Filter out whole genome; keep genes



Filter out non-exon regions

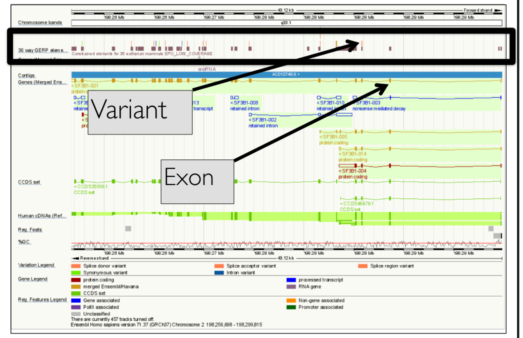


Data abstraction: highly filtered scope of *transcript coordinates*



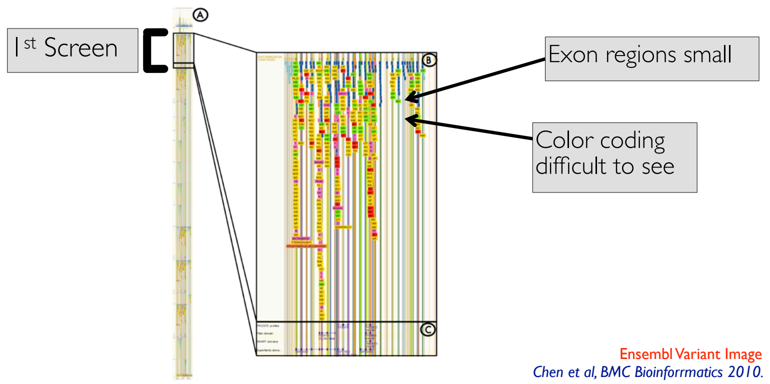
Dominant paradigm: genome browsers

- strengths: flexible and powerful
  - horizontal tracks: user data
  - shared coordinate system: *genome coordinates* (bp)
- problems
  - tiny features of interest spread out across large extent
    - must zoom far in to inspect known feature, then zoom out and pan to locate next
    - high cognitive load for interaction
    - must already know where to look



representative example: Ensembl
Chen et al, BMC Bioinformatics 2010.

Features of interest small even in variant-specific view

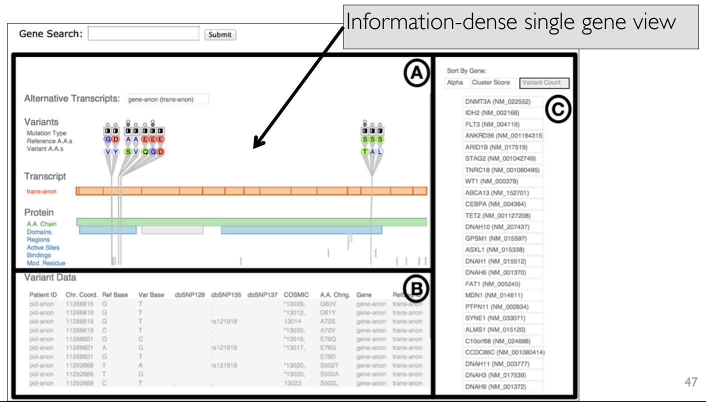


Ensembl Variant Image
Chen et al, BMC Bioinformatics 2010.

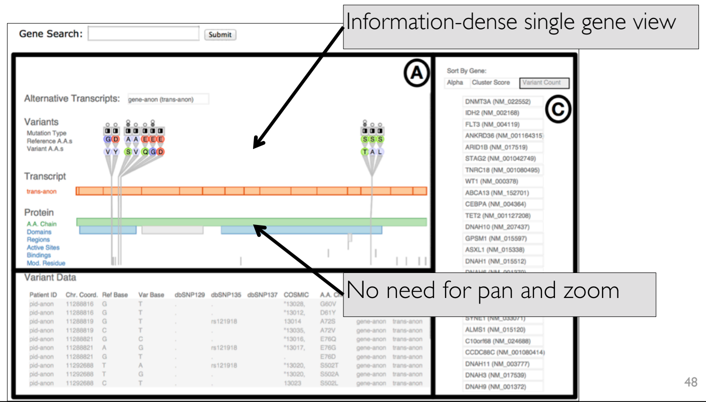
Idioms

Variant View

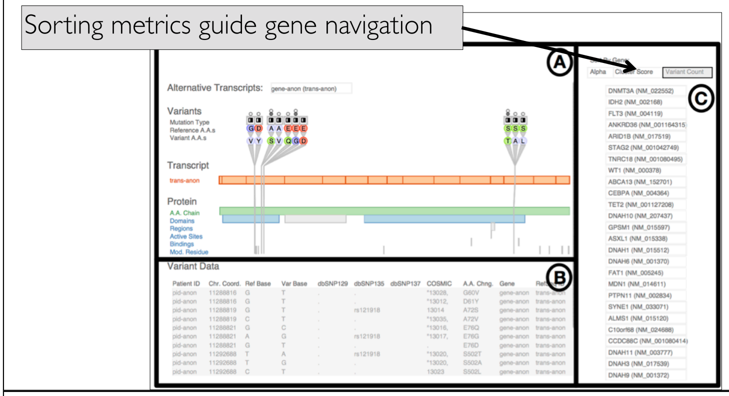
Variant View



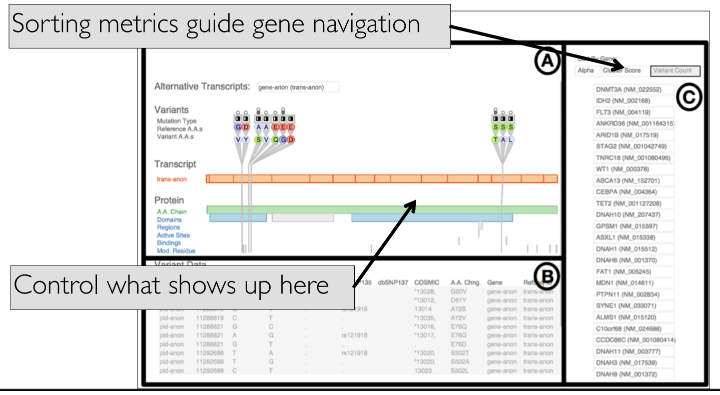
Variant View



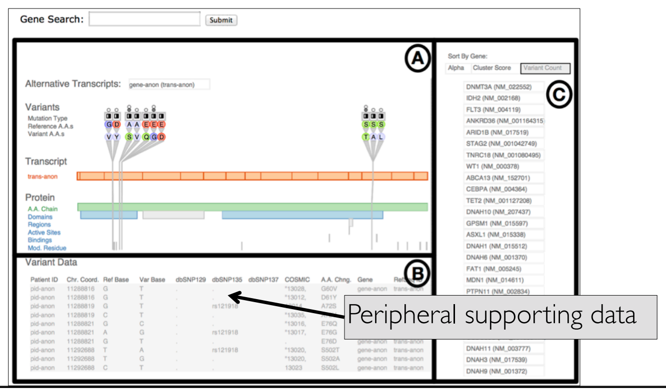
Variant View



Variant View

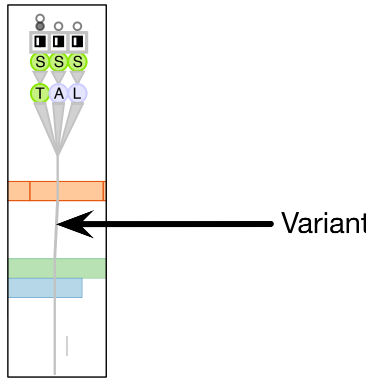


Variant View

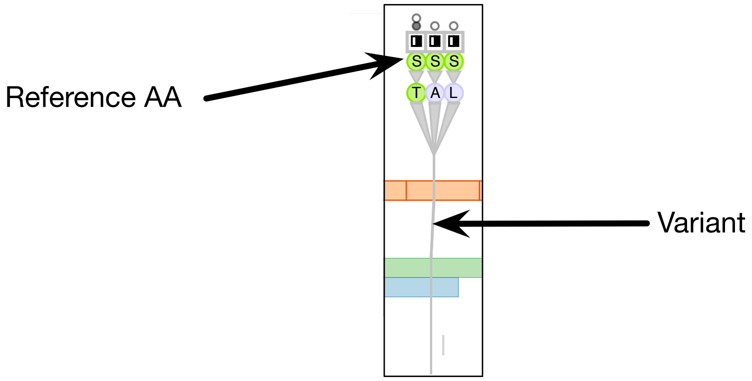


Design information-dense visual encoding

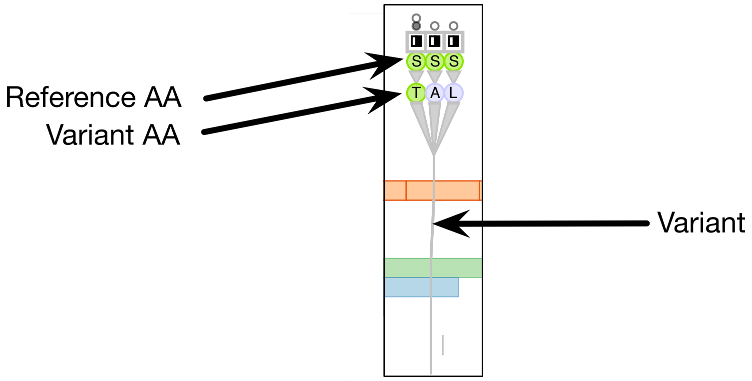
- show all attributes necessary for variant analysis
  - match salience with importance for analysis task
- variant not just a thin line!
- emphasize with high salience
  - collocated variants fan out at top
  - grey variant vertical stroke intersects horizontal colored protein regions



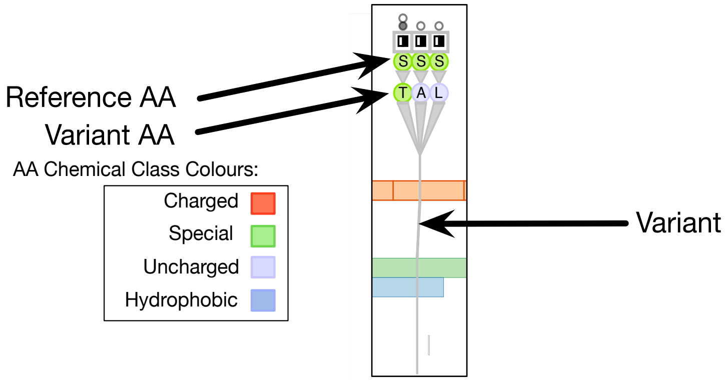
Design information-dense visual encoding



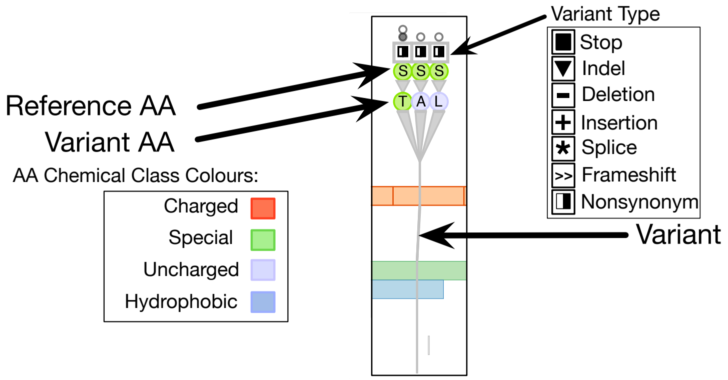
Design information-dense visual encoding



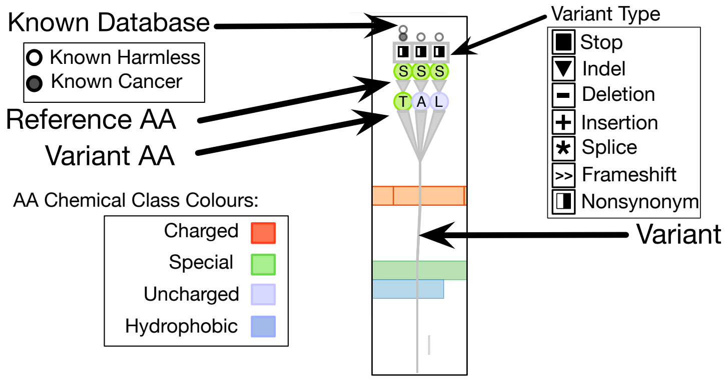
Design information-dense visual encoding



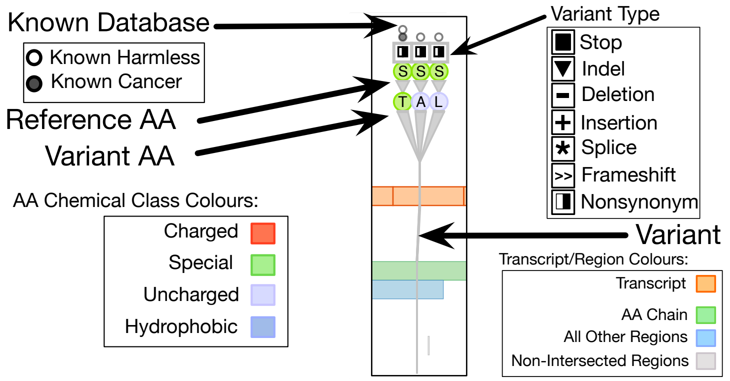
Design information-dense visual encoding



Design information-dense visual encoding

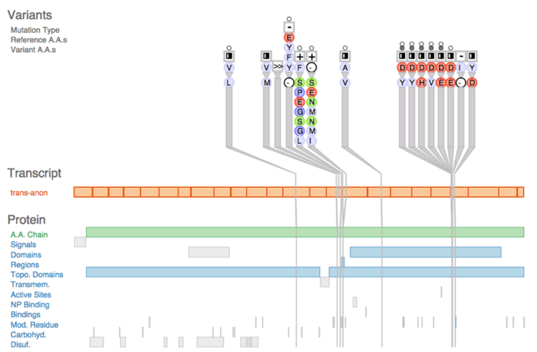


Design information-dense visual encoding

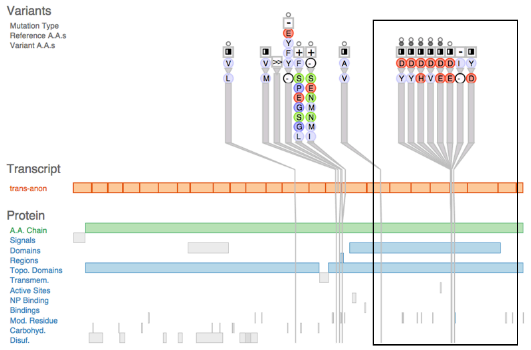


Results

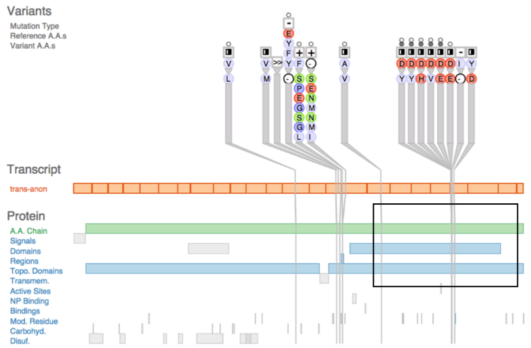
Highly scored gene by sorting metric: known leukemia gene



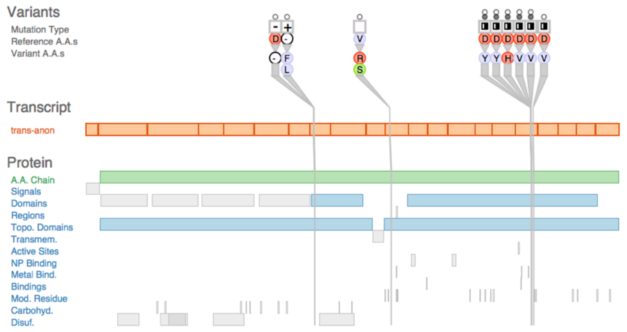
Visual inspection reveals collocation of variants



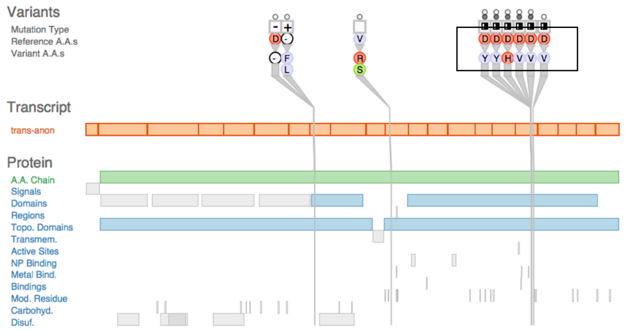
Several functional protein regions affected



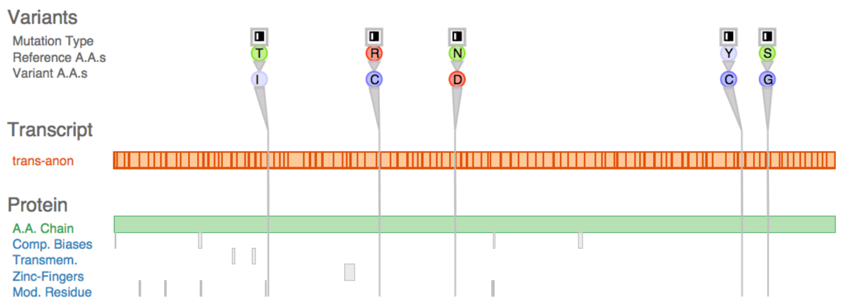
Highly scored by metric: not previously known, good candidate



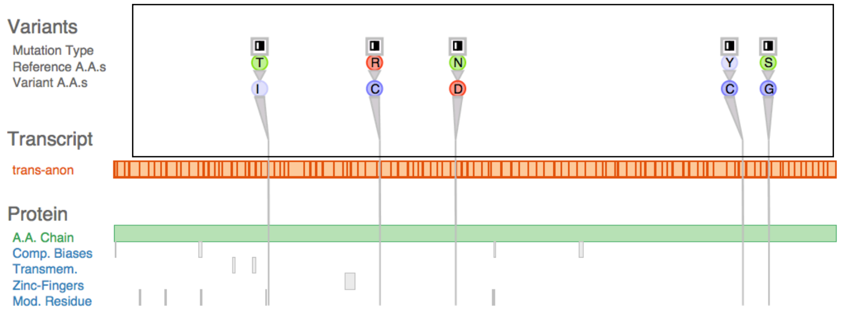
Protein chemical class change evident



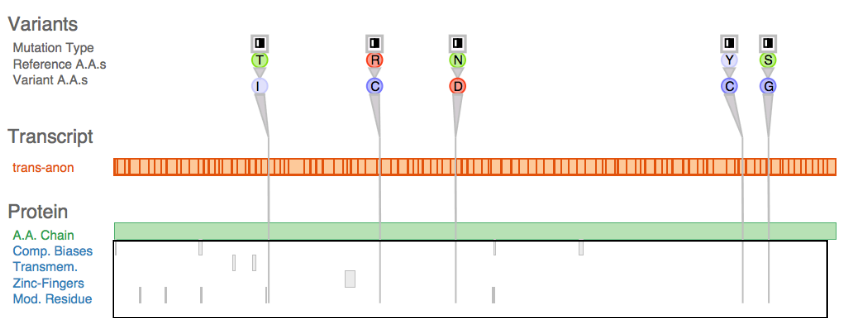
In contrast, low scoring gene



No collocation of variants



Mostly unaffected protein regions



Methods

Phase 1: Winnow and Cast

5 months

- embedded within GSC for all stages
- winnow stage
  - considered and ruled out many potential collaborators
- cast stage
  - gatekeeper (PI)
  - two front-line analysts (postdocs)

more at:
Design Study Methodology: Reflections from the Trenches and from the Stacks.
Sedlmair, Meyer, Munzner. IEEE TVCG 18(12): 2431-2440, 2012 (Proc. InfoVis 2012).

Phase 2: Core Design

5 months

- main task abstraction
  - discover gene
- semi-structured interviews
  - every week for 1 hr
- iterative refinement
  - 8 data sketches deployed

Phase 3: Two More Tasks

1 month

- two new analysts
  - connected by enthusiastic gatekeeper
- new task abstractions
  - compare patients
  - debug pipeline
- transferrable with minimal changes

Phase 4: Reflect and write

3 months

- abstraction innovation
  - data abstraction: highly filtered *transcript coordinates* (vs genome coordinates)
- guidelines
  - specialize first, generalize later
    - good for domains with complex data
  - high-level considerations
    - identifying scales of interest
    - what to visually encode directly vs what to support through interaction
    - when (and how) to eliminate navigation

Outline

- **Visualization Analysis Framework**
  Session 1 *9:30-10:45am*
  - Introduction: Definitions
  - Analysis: What, Why, How
  - Marks and Channels
- **Idiom Design Choices, Part 2**
  Session 3 *1:15pm-2:45pm*
  - Manipulate: Change, Select, Navigate
  - Facet: Juxtapose, Partition, Superimpose
  - Reduce: Filter, Aggregate, Embed
- **Idiom Design Choices**
  Session 2 *11:00am-12:15pm*
  - Arrange Tables
  - Arrange Spatial Data
  - Arrange Networks and Trees
  - Map Color
- **Guidelines and Examples**
  Session 4 *3-4:30pm*
  - Rules of Thumb
  - Validation
  - BioVis Analysis Example

http://www.cs.ubc.ca/~tmm/talks.html#minicourse14

More Information

- book
- this tutorial
  http://www.cs.ubc.ca/~tmm/talks.html#minicourse14
- papers, videos, software, talks, courses
  http://www.cs.ubc.ca/~tmm
- conferences
  - VIS: VAST, InfoVis, SciVis http://ieeevis.org
    - 2014: Paris, Nov 9-14
  - EuroVis
    - 2014: Swansea, Jun 9-13
  - BioVis
    - 2014: Boston, Jul 11-12 (w/ ISMB)
  - VizBi
    - 2015: Boston, March 25-27