

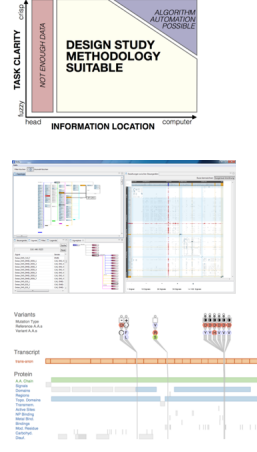
A Trio of Visualization Design Studies

Tamara Munzner
Department of Computer Science
University of British Columbia

Segal Design Institute, Northwestern University
8 April 2014, Chicago IL
<http://www.cs.ubc.ca/~tmm/talks.html#chicago14>

Trio

- Design Study Methodology
– meta-paper: how to do design studies
- RelEx
– overlay network optimization for in-car networks
- Variant View
– sequence variant analysis in gene context



Defining Visualization

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

Defining Visualization

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

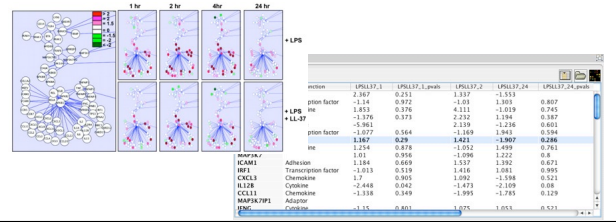
- human in the loop needs the details
– doesn't know exactly what questions to ask in advance

Identical statistics	
x mean	9.0
x variance	10.0
y mean	7.50
y variance	3.75
x/y correlation	0.816

Defining Visualization

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- human in the loop needs the details
– doesn't know exactly what questions to ask in advance
- external representation: replace cognition with perception



Defining Visualization

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- human in the loop needs the details
– doesn't know exactly what questions to ask in advance
- external representation: perception vs cognition
- intended task

Defining Visualization

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- human in the loop needs the details
– doesn't know exactly what questions to ask in advance
- external representation: perception vs cognition
- intended task
- measurable definitions of effectiveness

more at:
Visualization Analysis and Design, Chapter 1.
Munzner, AK Peters, 2014, to appear.

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).

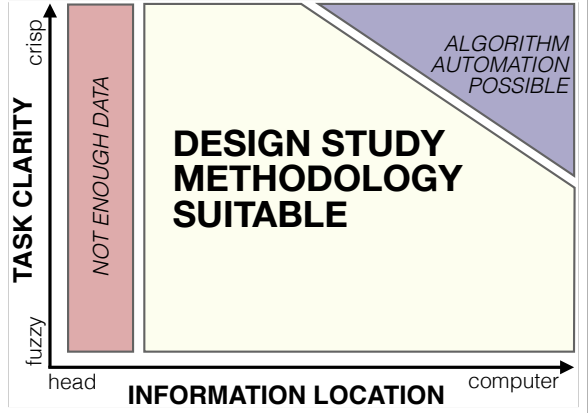
Defining Design Study

- 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

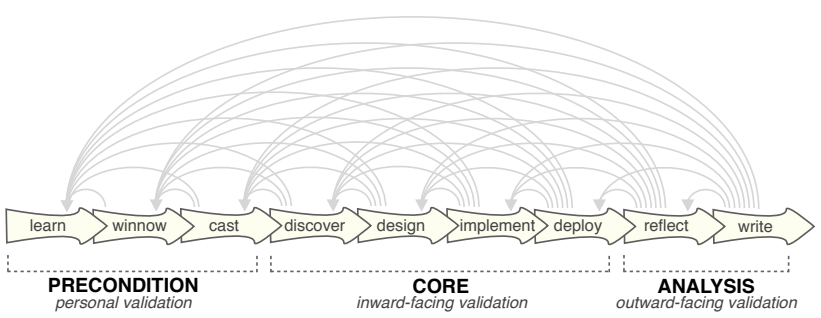
more at:
A Nested Model of Visualization Design and Validation.
Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).

more at:
The Nested Blocks and Guidelines Model.
Meyer, Sedlmair, Quinan, Munzner. Information Visualization Journal, 2014, to appear.

When To Do Design Studies

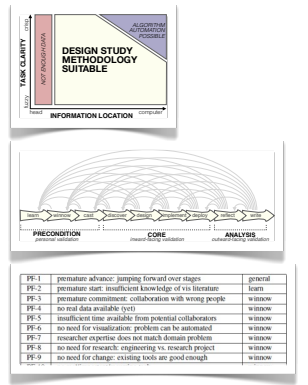


Nine-Stage Framework



How To Do Design Studies

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

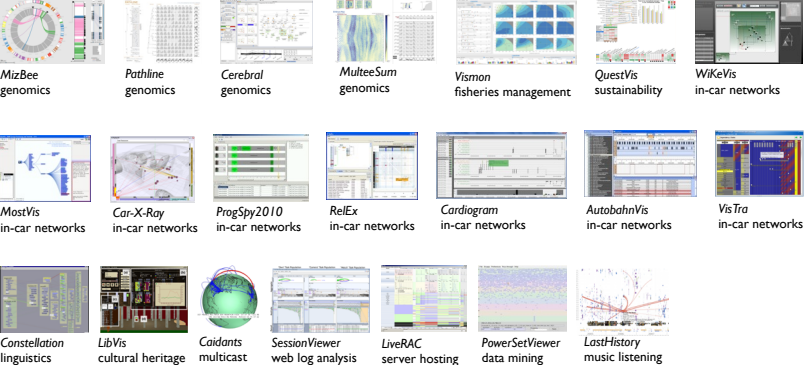


Pitfall Example: Premature Publishing

algorithm innovation design studies

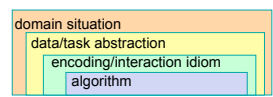
Must be first! **Am I ready?**

Design Studies: Lessons learned after 21 of them



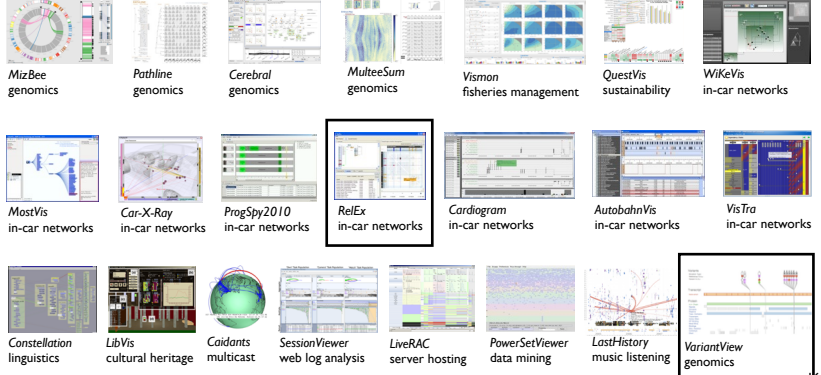
Abstractions and Idioms

- abstractions
– **translate** from specifics of domain to vocabulary of vis
• task abstraction: **why** they're looking at it
• data abstraction: **what** to draw
– **transform** data into form useful for task at hand
• don't just draw what you're given; decide what is the right thing!
- idioms
– visual encoding idiom: **how** to draw
– interaction idiom: **how** to manipulate
- focus today: two mappings
– from domain to abstraction
– from abstraction to idiom



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

Today's Focus

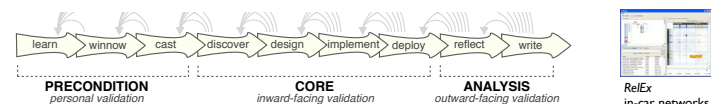


http://www.pinfo.org/10480334_wolverhampton_horse_racing_live_streaming_wolverhampton_handslap_8_jan-2010.html

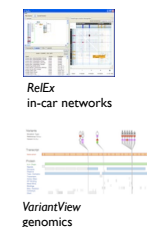
http://www.silberstein.com/infocenter/visual_concept.jpg

Themes

- task and data abstraction
 - both cases: complex and tricky
 - clear description in final talk/paper is end of a long, long road
 - writing as research: refine during reflection even after vis tool is finalized...



- visual encoding and interaction idioms
 - RelEx: reduce memory load with interaction
 - VariantView: reduce interaction load with better visual encoding

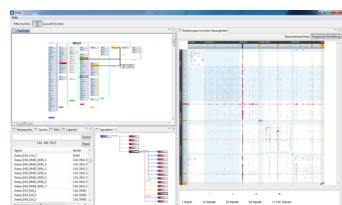


RelEx

Visualization for Actively Changing Overlay Network Specifications

joint work with:
Michael Sedlmair, Annika Frank, Andreas Butz
<http://www.cs.ubc.ca/labs/imager/tr/2012/relex/>

RelEx: Visualization for Actively Changing Overlay Network Specifications.
Sedlmair, Frank, Butz, Munzner. IEEE TVCG 18(12): 2729-2738, 2012 (Proc. InfoVis 2012).

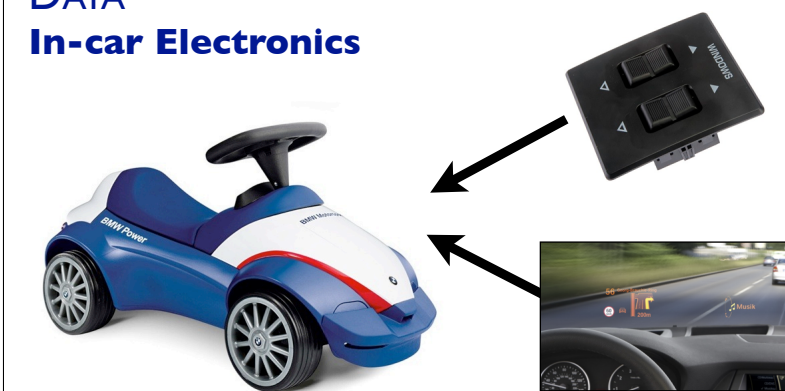


Domain: In-car network engineering



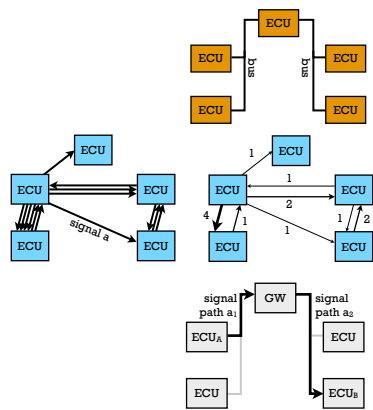
Abstractions

DATA In-car Electronics



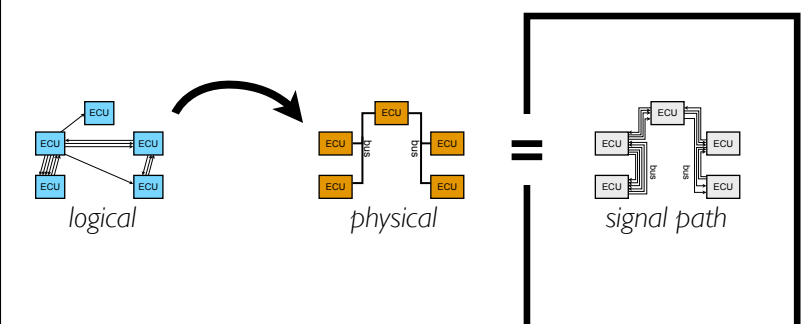
Data Abstraction: 3 Networks

- physical network**
 - 100 nodes: Electronic Control Units
 - 10-15 hyperedges: bus systems
 - hardware engineers
- logical network**
 - same nodes
 - 10,000 multigraph edges: signals
 - 1,000 weighted edges: signal counts
 - software engineers
- overlay network**
 - maps logical onto physical
 - 30,000 edges: signal paths
 - target engineers



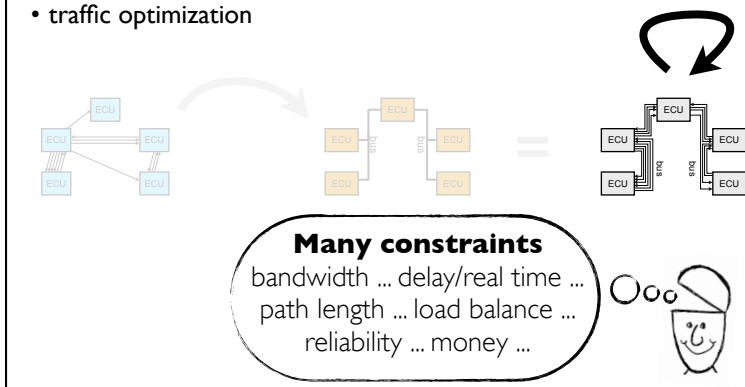
Task Abstraction: Mapping

- specify overlay network that maps logical onto physical



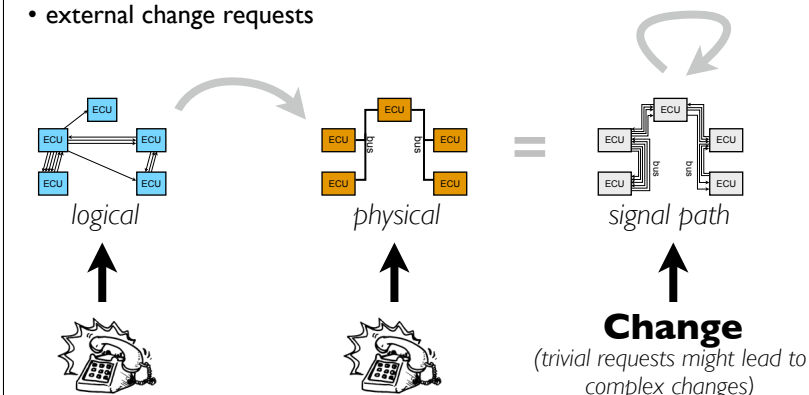
Task Abstraction: Optimizing

- traffic optimization



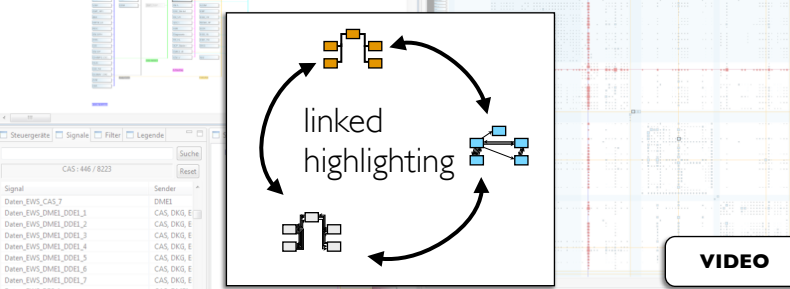
Task Abstraction: Changing

- external change requests

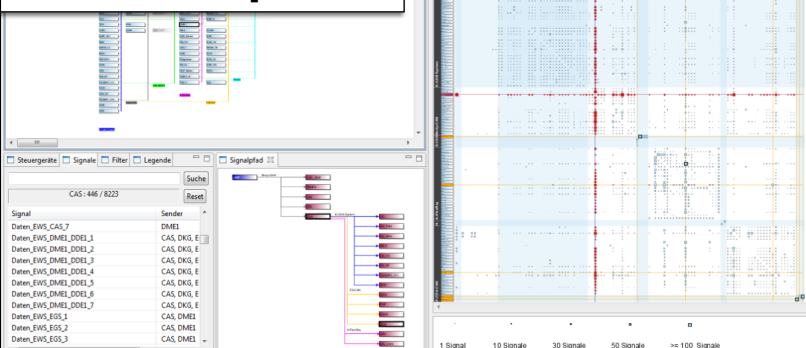


Idioms

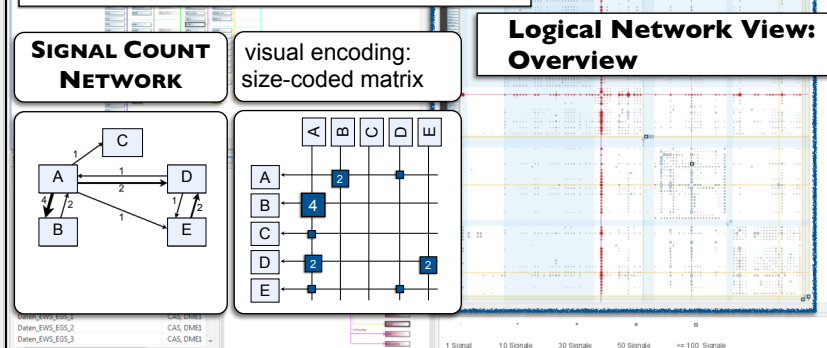
INTERACTION IDIOM: Cross-Network Relations



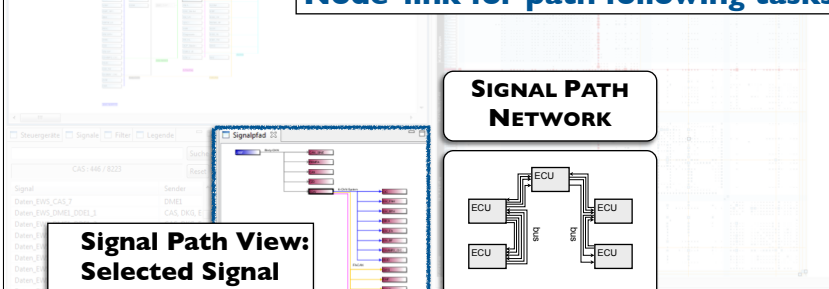
RELEX: Relation Explorer



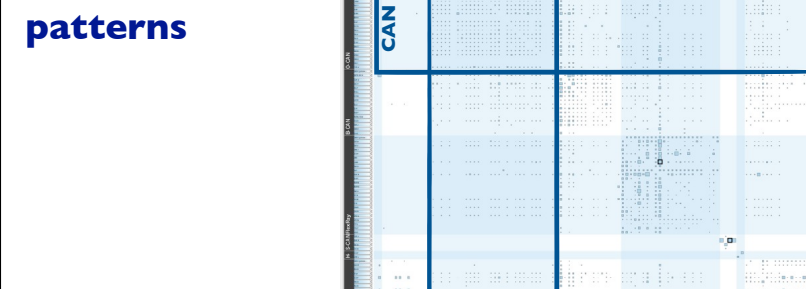
Vis Guideline [Ghoniem 2005] Matrix for dense network data



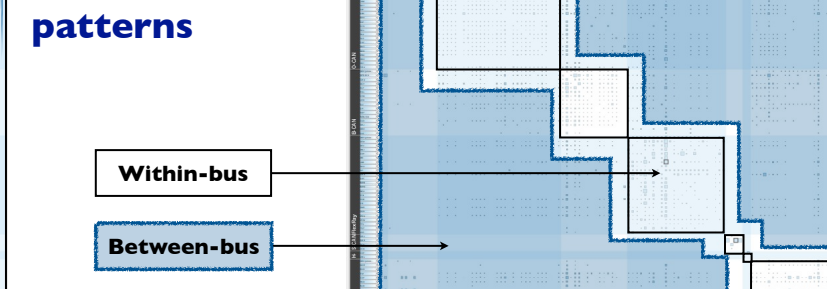
INTERESTS Signal Path View: Selected Signal



INTERESTS Bus communication patterns



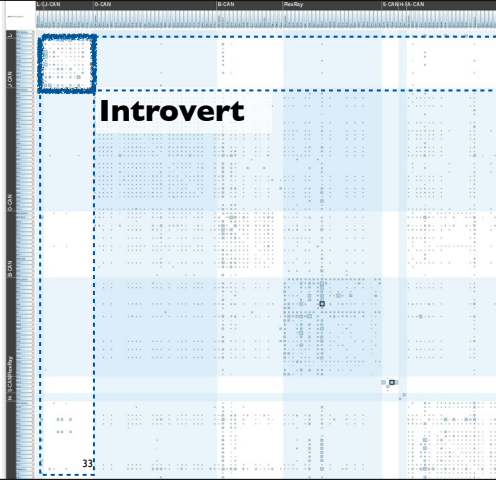
INTERESTS Bus communication patterns



INTERESTS

Bus communication patterns

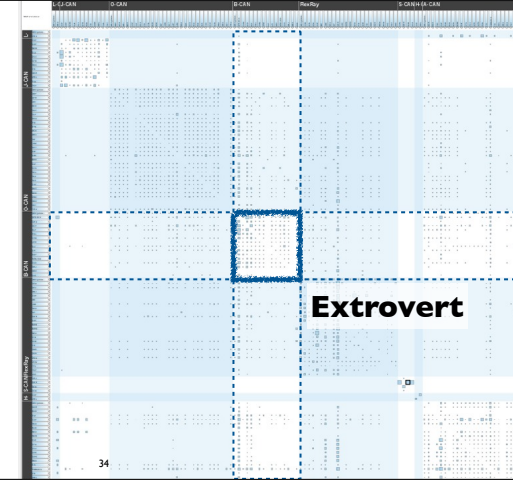
introvert vs. extrovert



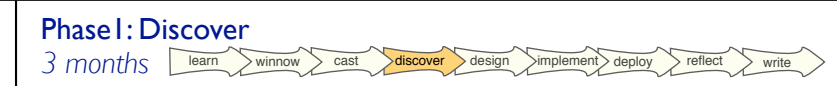
INTERESTS

Bus communication patterns

introvert vs. extrovert



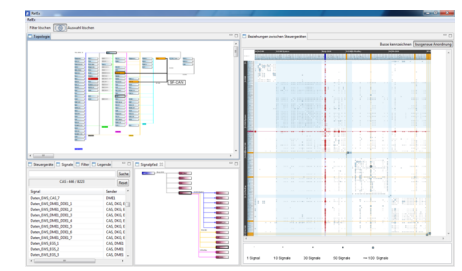
Methods



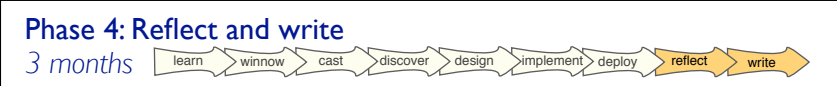
- embedded within BMW
 - phases 1, 2, 3
- contextual inquiry
- abstracting
- deriving design requirements



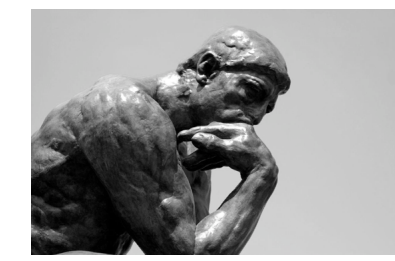
- iterative paper prototyping
- agile software development
 - 3 lead users (engineers)
 - 6 deployed releases
- usability engineering
 - domain experts
 - HCI students



- field study
 - 7 engineers
 - 5 weeks
- think aloud study
 - 10 engineers
 - ~1 hour each session
- adoption
 - 15+ users, 3 months post-study



- revisit abstractions
- relate to other design studies
- write up



Abstraction Innovation

Previous Work

Focus on social network analysis

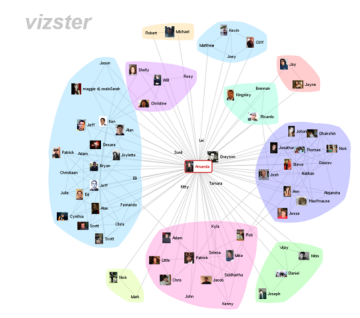
- radically different task and data abstractions



Task Abstraction

Social Network Analysis Domain

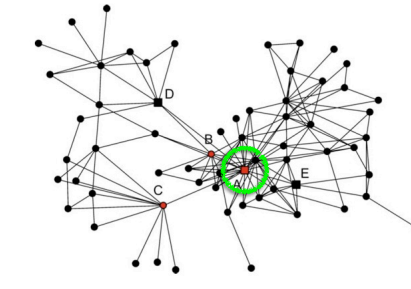
- find clusters



Task Abstraction

Social Network Analysis Domain

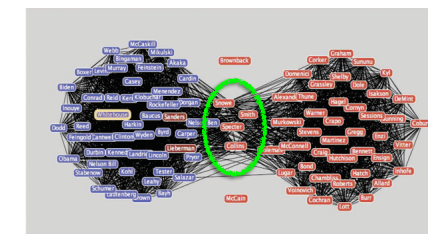
- find clusters
- find high-degree nodes



Task Abstraction

Social Network Analysis Domain

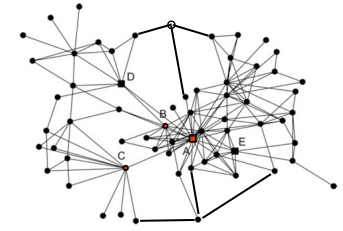
- find clusters
- find high-degree nodes
- find bridge nodes



Task Abstraction

Social Network Analysis Domain

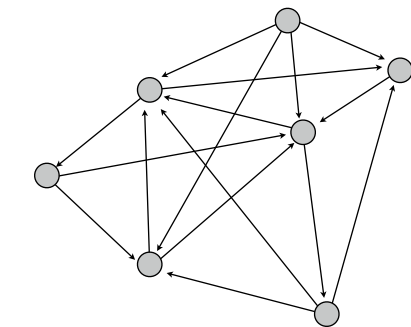
- find clusters
- find high-degree nodes
- find bridge nodes
- understand temporal dynamics
 - passively notice changes



Data Abstraction

Social Network Analysis Domain

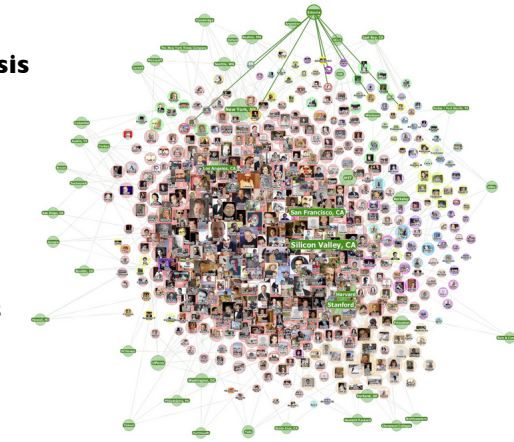
- single graph



Data Abstraction

Social Network Analysis

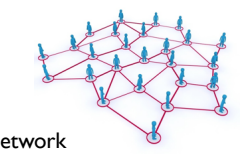
- single graph
- scalability challenge: nodes



Abstraction Differences

Social Network Analysis vs Overlay Network Optimization

- data
 - single network
 - node scalability
 - sparse edges
- task
 - find clusters, high-degree nodes, bridge nodes
 - passive changes



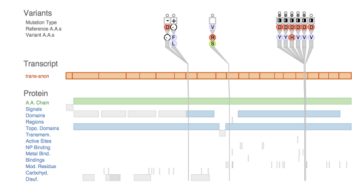
- data
 - three related networks
 - physical, logical, overlay
 - path scalability
 - dense edges, few nodes
- task
 - traffic optimization
 - active changes



Variant View

Visualizing Sequence Variants in their Gene Context

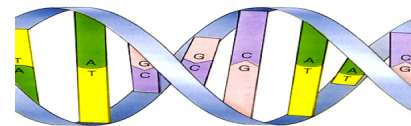
joint work with:
Joel Ferstay, Cydney Nielsen
<http://www.cs.ubc.ca/labs/imager/tr/2012/VariantView/>



VariantView: 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?**

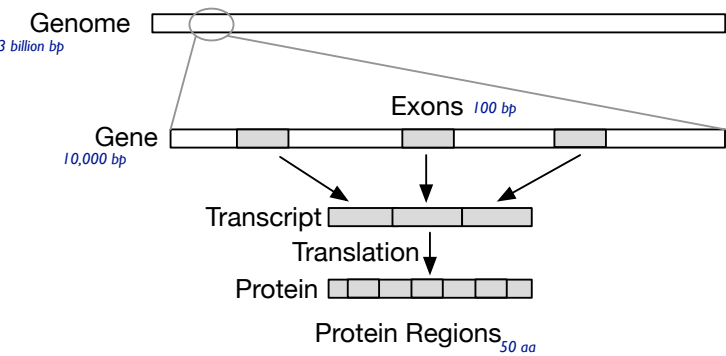
Sample 2 Genome DNA: ATA TGA TGA ACA CCT **Harmless?**

Cancer Research

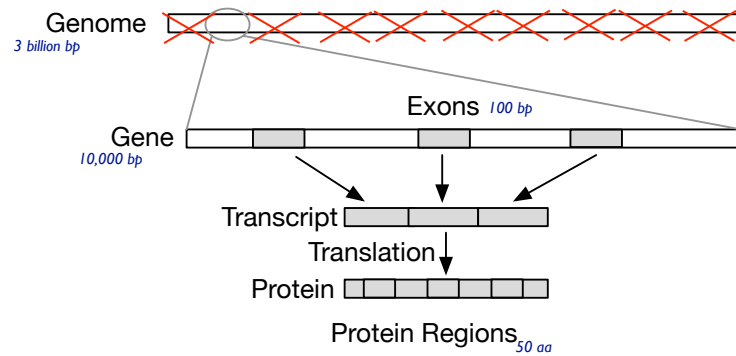
- collaboration with analysts at BC Genome Sciences Center
 - studying genetic basis of leukemia
- driving task
 - discover new candidate genes with harmful variants
- two big questions
 - what to show
 - data abstraction
 - challenge: enormous range of scales in the data
 - how to show it
 - visual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

Abstractions

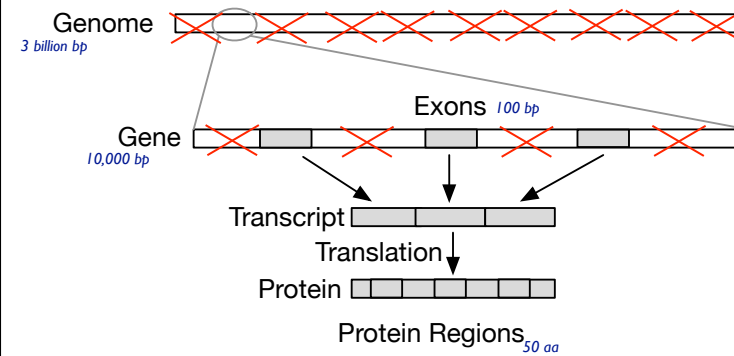
Data: Filtering to relevant biological levels and scales



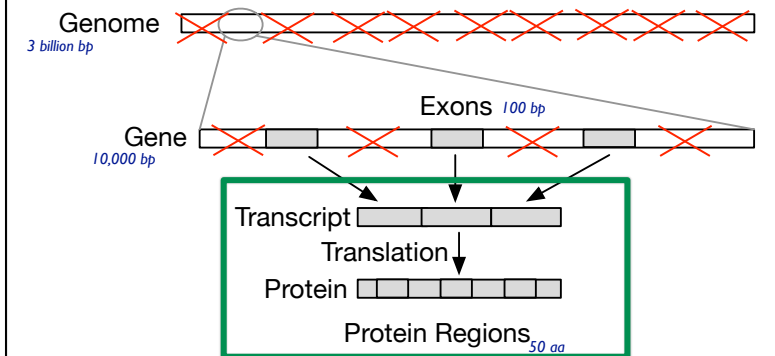
Filter out whole genome; keep genes



Filter out non-exon regions

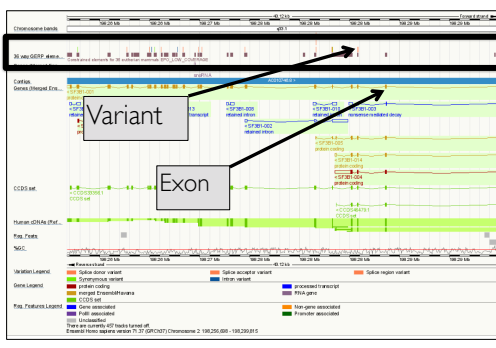


Data abstraction: highly filtered scope of transcript coordinates



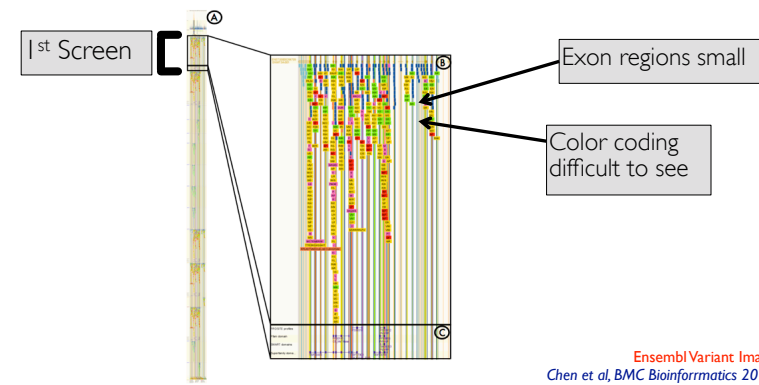
Dominant paradigm: genome browsers

- 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



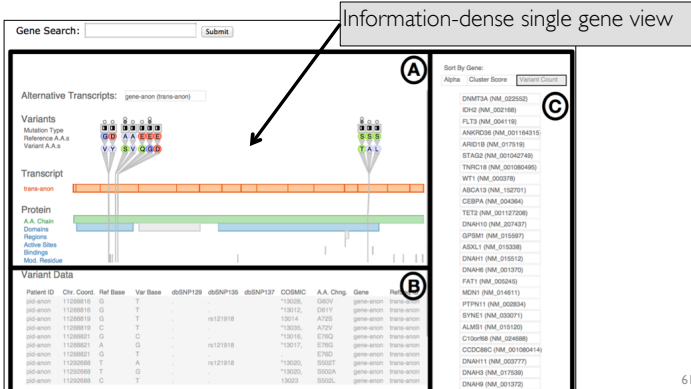
Ensembl Variant Image
Chen et al, BMC Bioinformatics 2010.

Idioms

Variant View



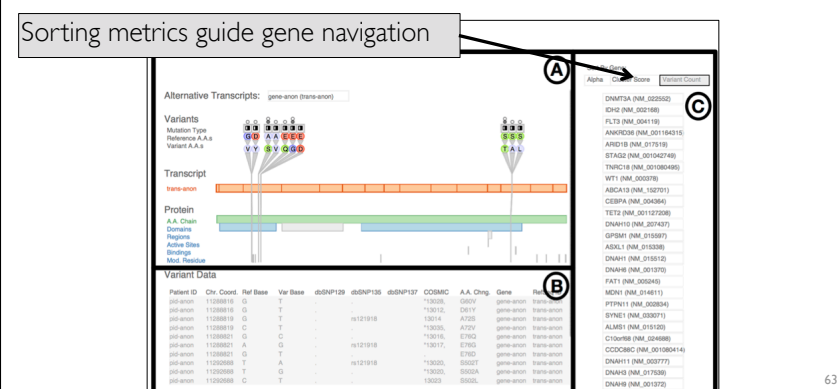
Variant View



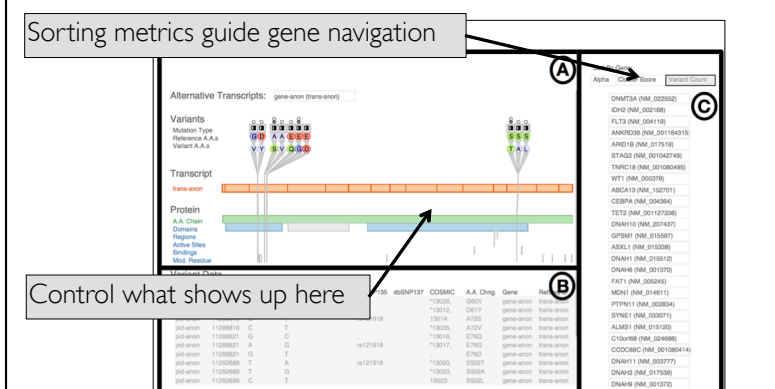
Variant View



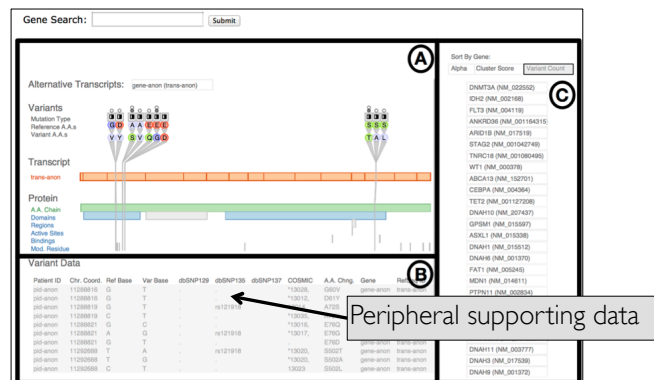
Variant View



Variant View

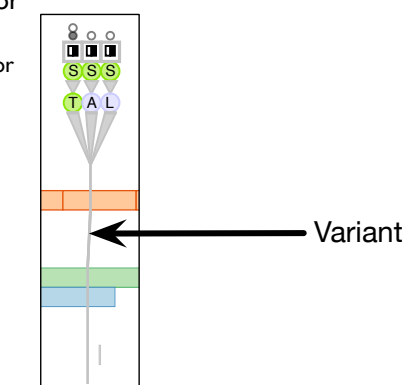


Variant View



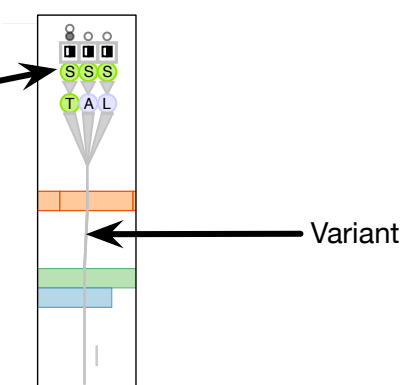
Design information-dense visual encoding

- 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



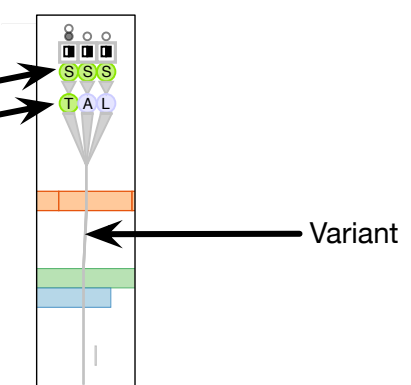
Design information-dense visual encoding

- Reference AA
- Variant AA

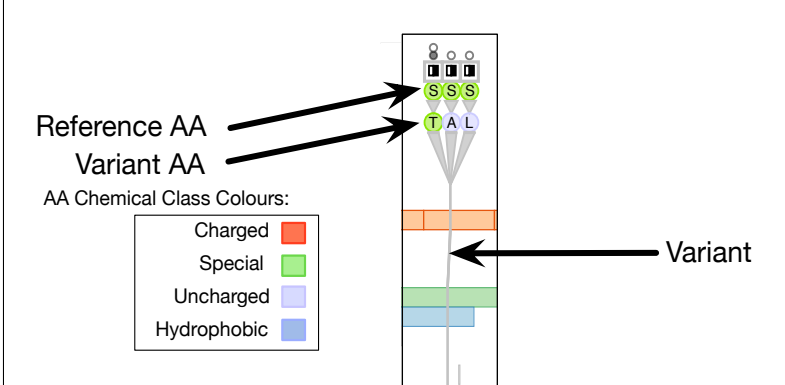


Design information-dense visual encoding

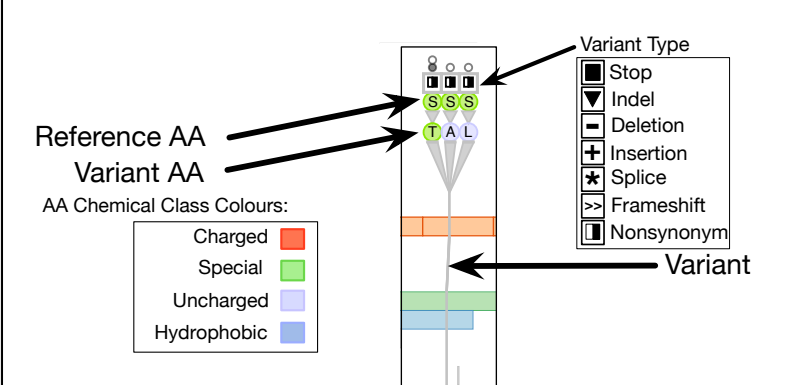
- Reference AA
- Variant AA



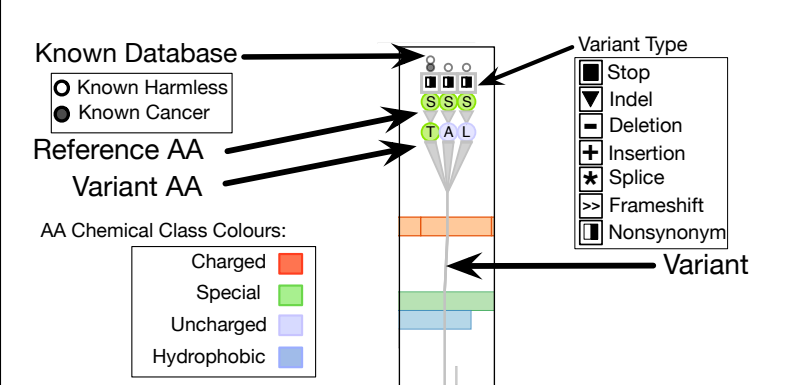
Design information-dense visual encoding



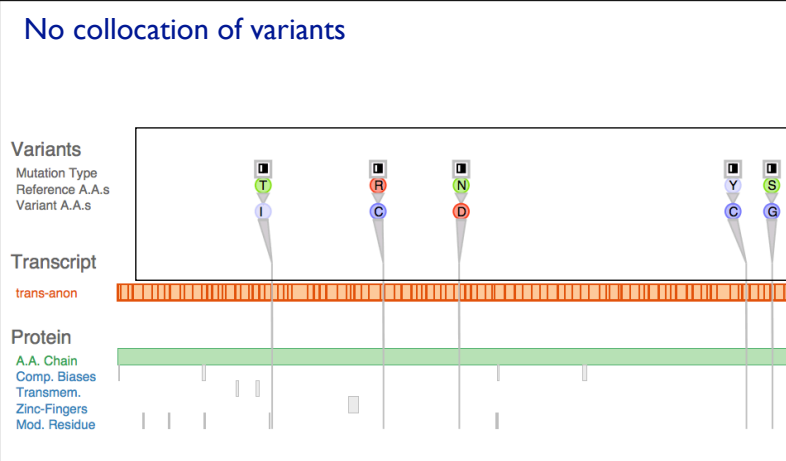
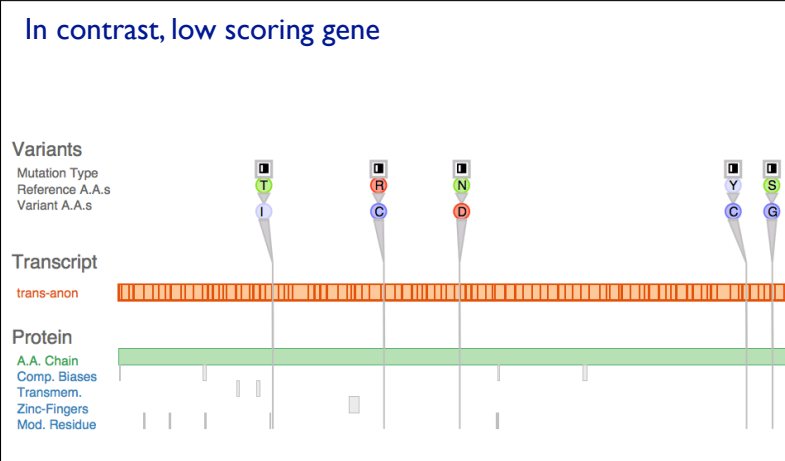
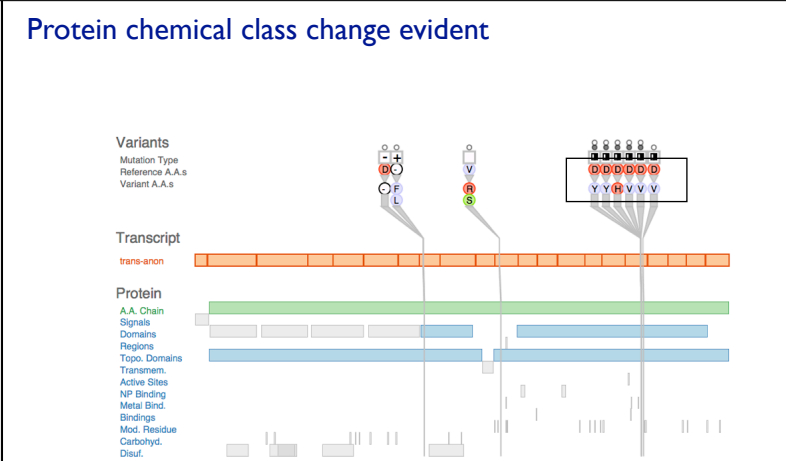
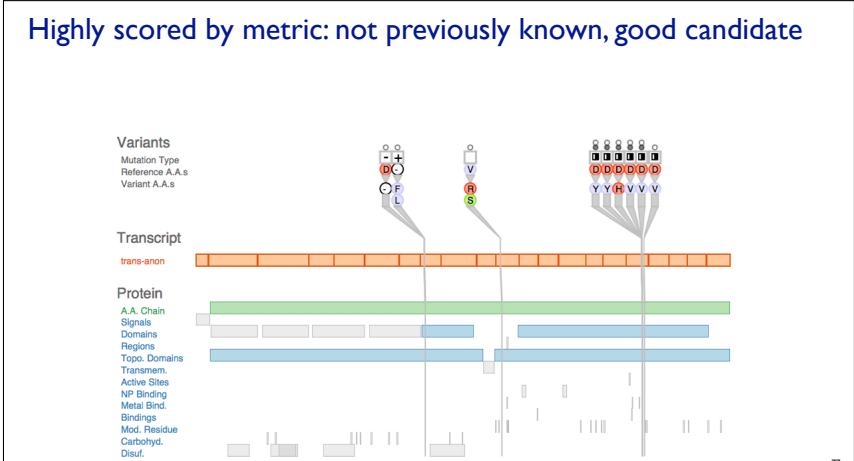
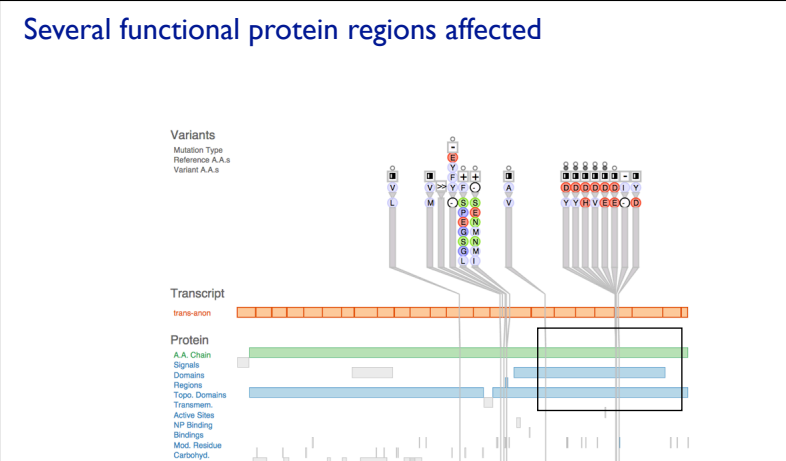
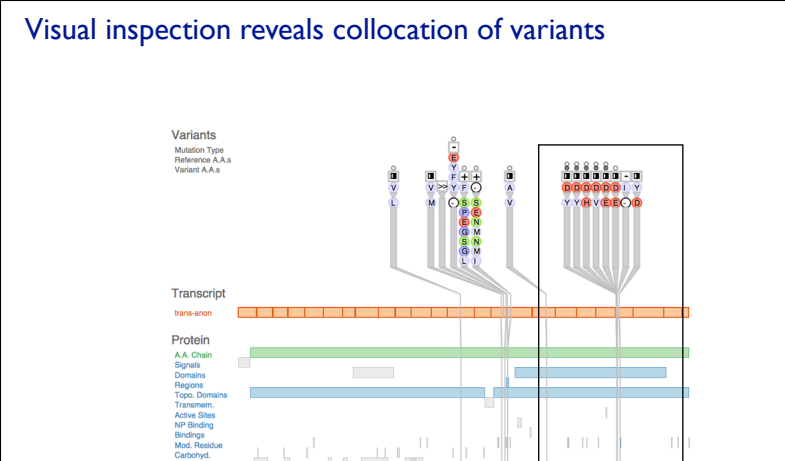
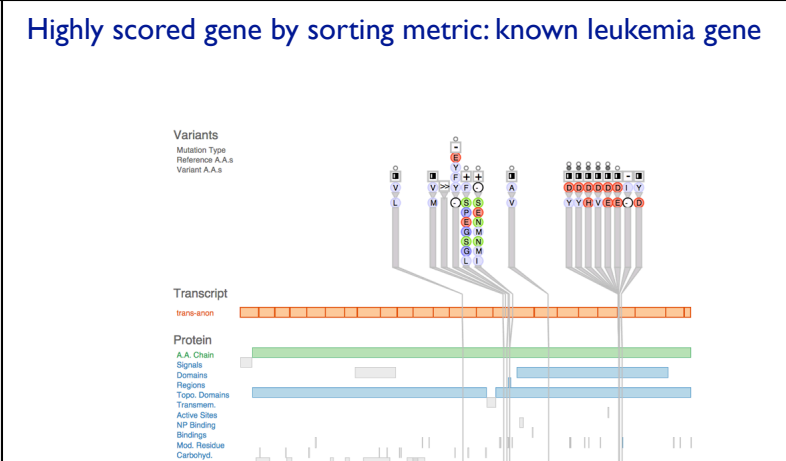
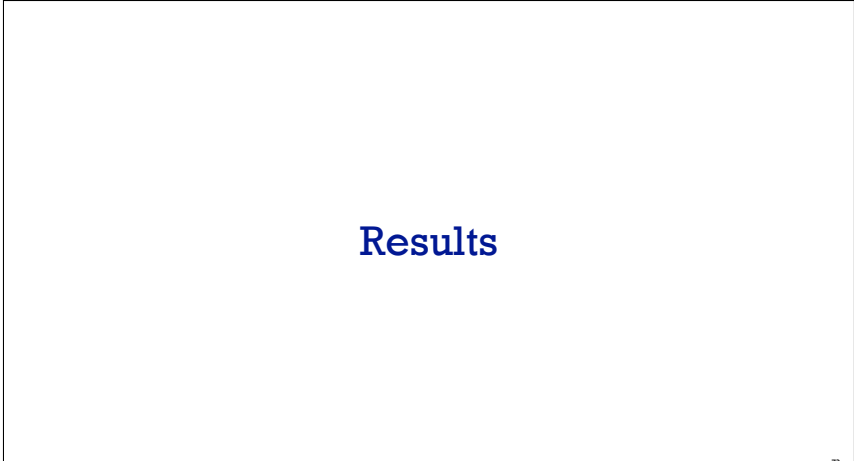
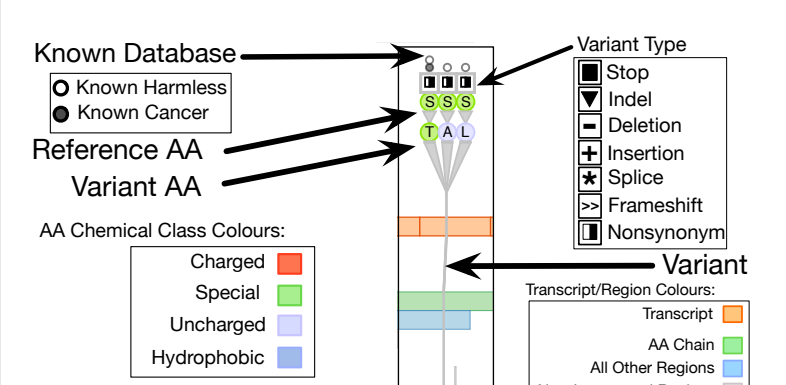
Design information-dense visual encoding



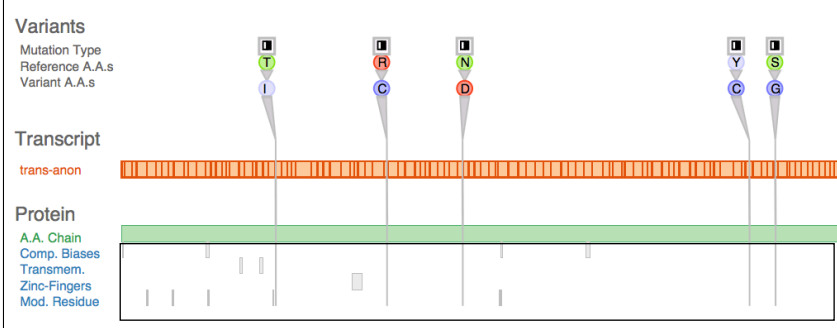
Design information-dense visual encoding



Design information-dense visual encoding



Mostly unaffected protein regions



81

Phase I: Winnow and Cast



- 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).

82

Methods

Phase 2: Core Design



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



83

Phase 3: Two More Tasks



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



85

Phase 4: Reflect and write



- 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

86

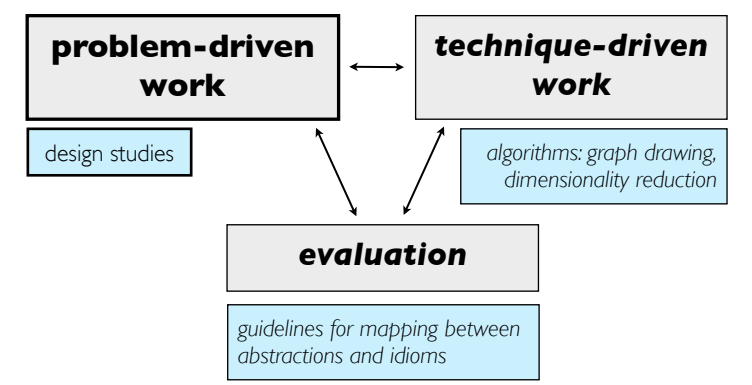
Themes, Revisited

- what and why to show: task and data abstraction
 - task and data commonalities cross-cut domains
- how to show: visual encoding and interaction idioms
 - RelEx: reduce memory load with interaction
 - VariantView: reduce interaction load with better visual encoding
- transferability from design studies
 - DSM: reflection to confirm/refute/refine/propose guidelines



87

A Different Trio: Research Interests



88

Further Information

- further info
 - <http://www.cs.ubc.ca/~tmm/talks.html#chicago14> (this talk, and many others)
 - <http://www.cs.ubc.ca/group/infovis> (papers, software, videos)
 - <http://www.cs.ubc.ca/~tmm/courses/infovis> (course: readings, lectures)
 - <http://www.cs.ubc.ca/~tmm/courses/infovis/book> (book: to appear)
- open source software downloads
 - <http://www.cs.ubc.ca/labs/imager/tr/2013/VariantView/VariantViewSoftware/>
- acknowledgements
 - funding: NSERC, NSF
 - joint work: all co-authors
 - Andreas Butz, Annika Frank, Joel Ferstay, Miriah Meyer, Cydney Nielsen, Michael Sedlmair
 - feedback on this talk
 - Matthew Brehmer, Stephen Ingram

89