

Variant View

Visualizing Sequence Variants in their Gene Context

Tamara Munzner
 Department of Computer Science
 University of British Columbia

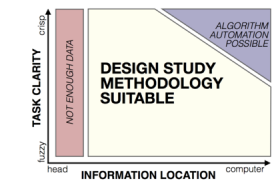
BioIT World, Data Visualization and Exploration Tools Track
 30 April 2014, Boston MA
<http://www.cs.ubc.ca/~tmm/talks.html#bioit14>

Variant View: Visualization 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.



Design Study Methodology

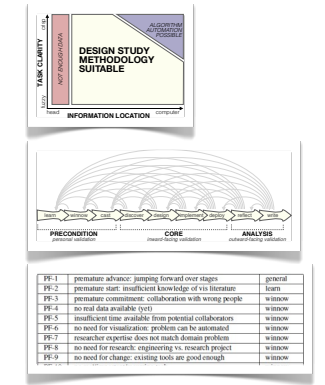
Reflections from the Trenches and from the Stacks

joint work with:
 Michael Sedlmair, Miriah Meyer
<http://www.cs.ubc.ca/labs/imager/tr/2012/dsm/>

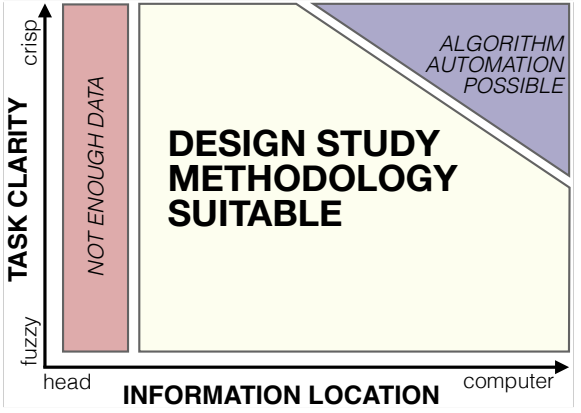
Design Study Methodology: Reflections from the Trenches and from the Stacks.
 Sedlmair, Meyer, Munzner. IEEE TVCG 18(12): 2431-2440, 2012 (Proc. InfoVis 2012).

How To Do Design Studies

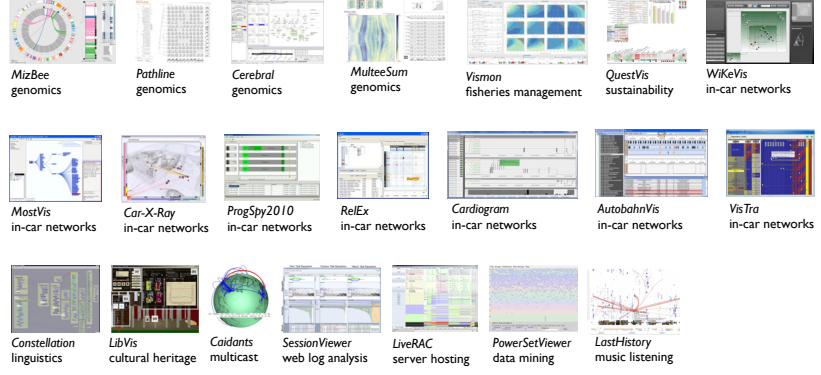
- definitions
- 9-stage framework
- 32 pitfalls and how to avoid them



When To Do Design Studies



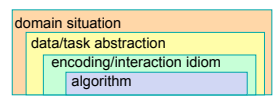
Design Studies: Lessons learned after 21 of them



• commonality of representations cross-cuts domains!

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

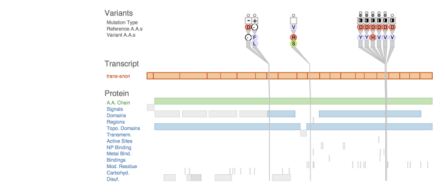


A Nested Model of Visualization Design and Validation.
 Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).

Variant View

Visualizing Sequence Variants in their Gene Context

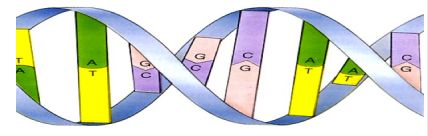
joint work with:
 Joel Ferstay, Cydney Nielsen
<http://www.cs.ubc.ca/labs/imager/tr/2012/VariantView/>



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



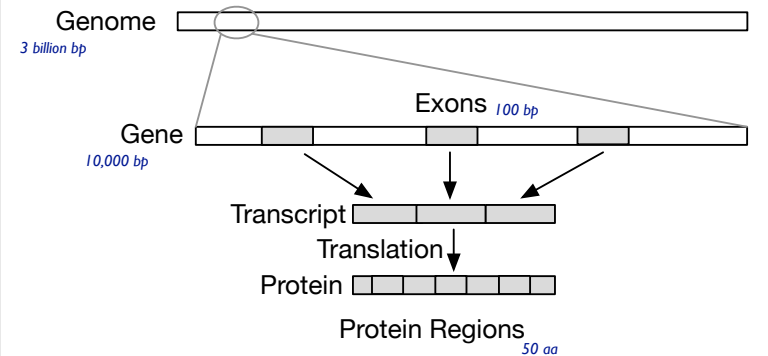
Reference Genome DNA:	ATA TGA TCA ACA CTT	
Sample 1 Genome DNA:	ATA TGG TCA ATA CTT	Harmful?
Sample 2 Genome DNA:	ATA TGA TGA ACA CCT	Harmless?

Cancer Research

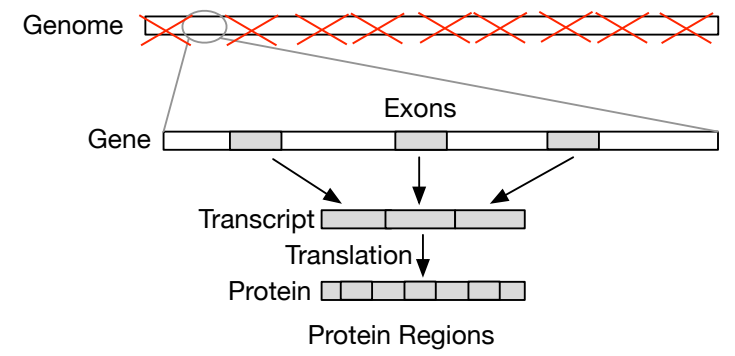
- collaboration with analysts at BC Genome Sciences Center
 - studying genetic basis of leukemia
- driving task
 - discover new candidate genes with harmful variants
- two big questions
 - what to show
 - data abstraction
 - challenge: enormous range of scales in the data
 - how to show it
 - visual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

Abstractions

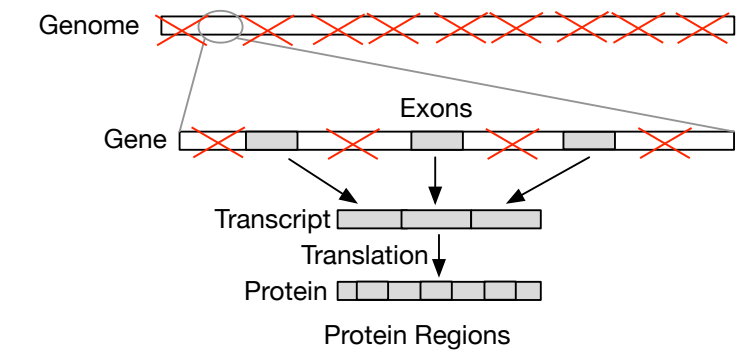
Data: Filtering to relevant biological levels and scales



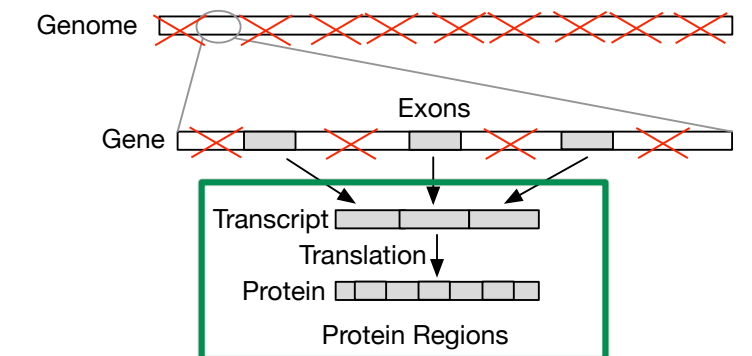
Filter out whole genome; keep genes



Filter out non-exon regions

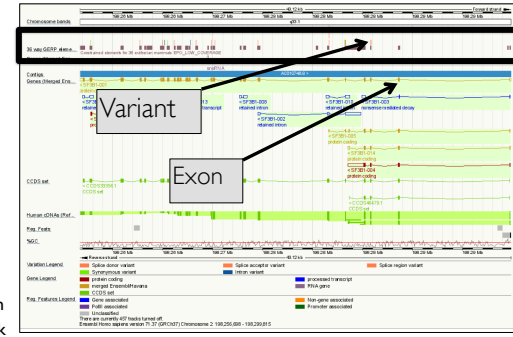


Data abstraction: highly filtered scope



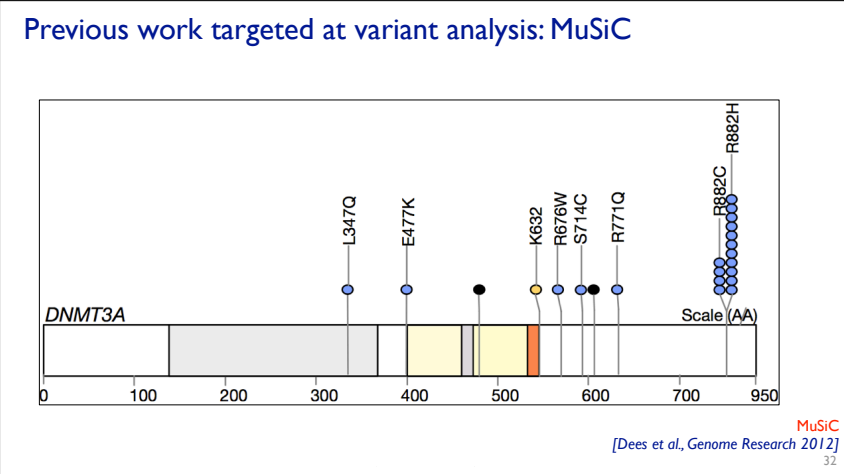
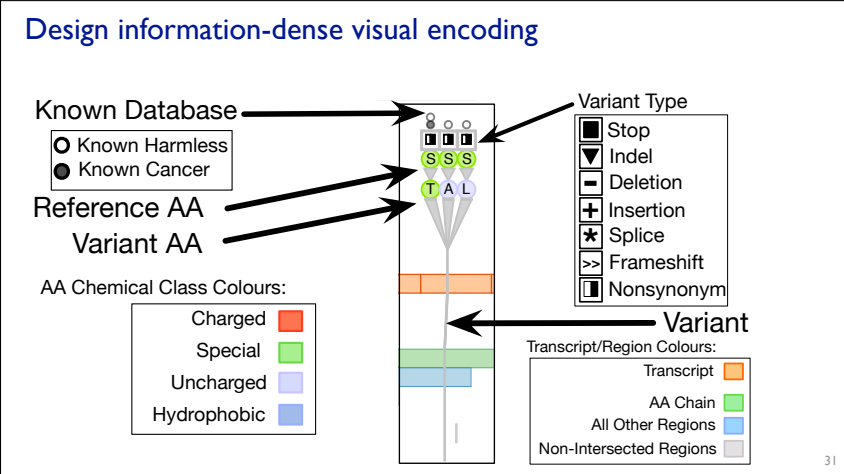
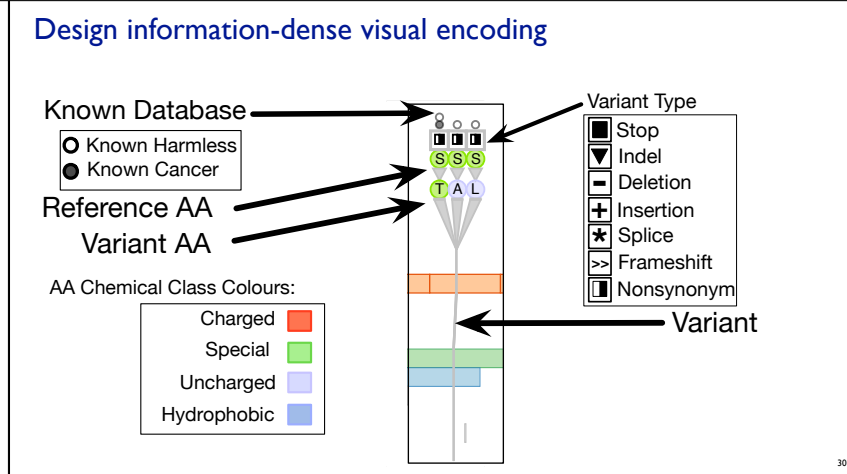
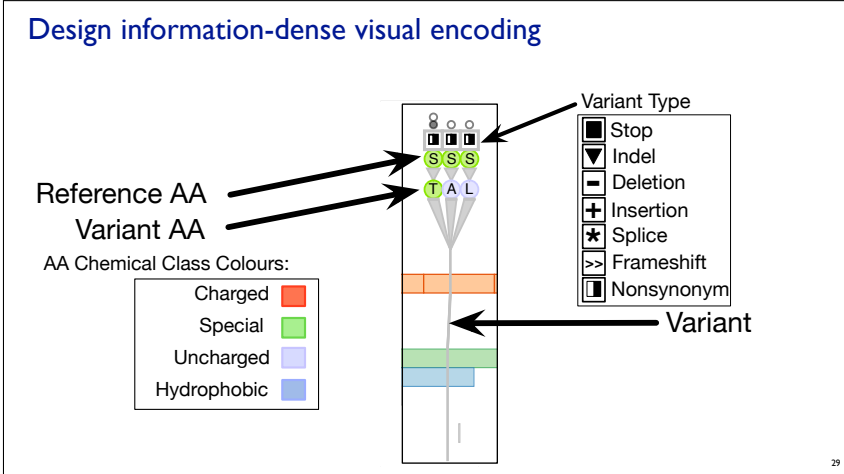
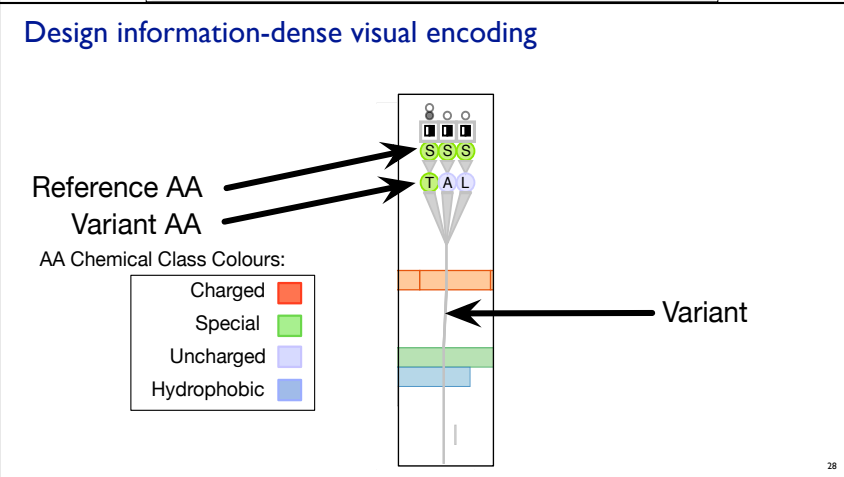
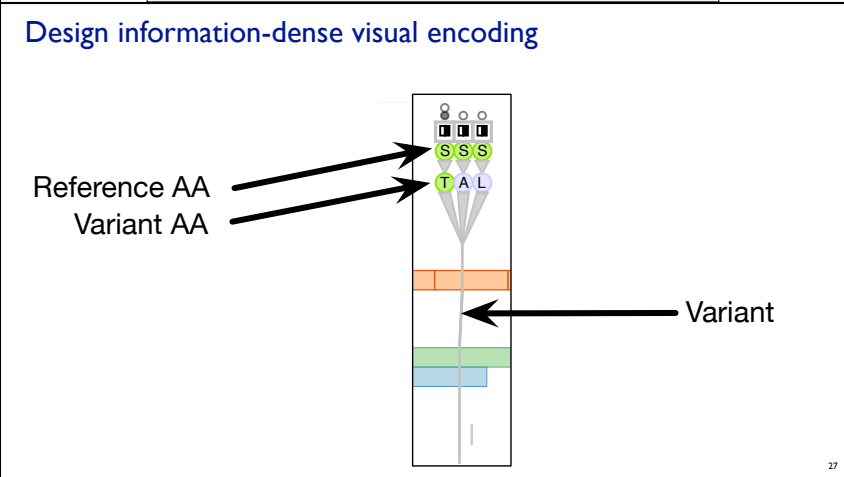
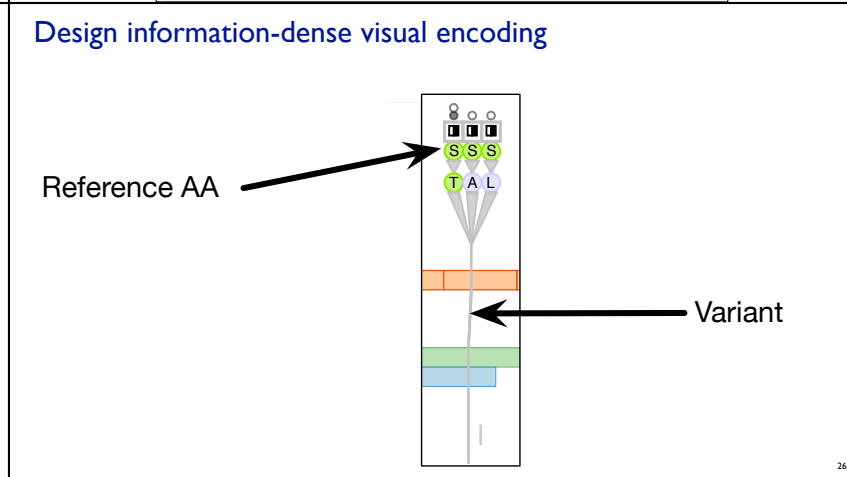
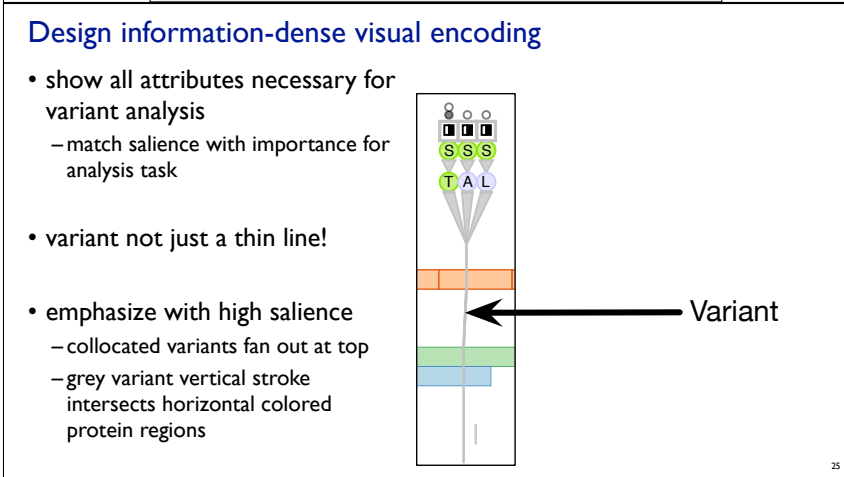
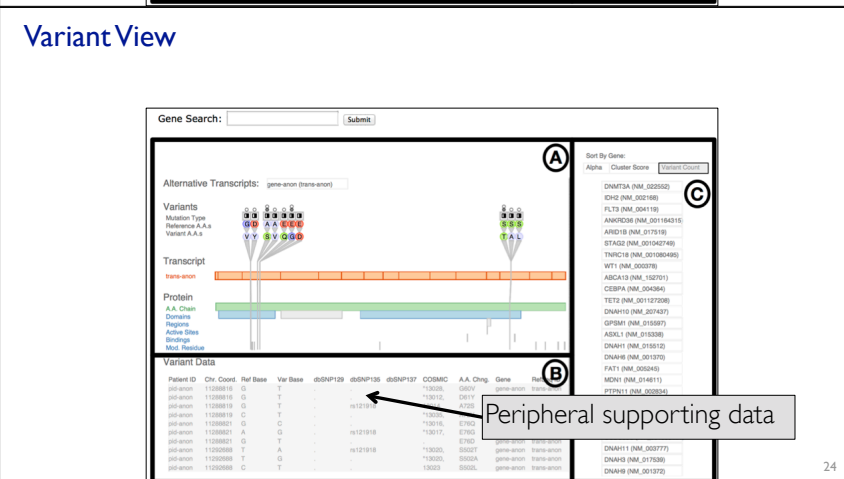
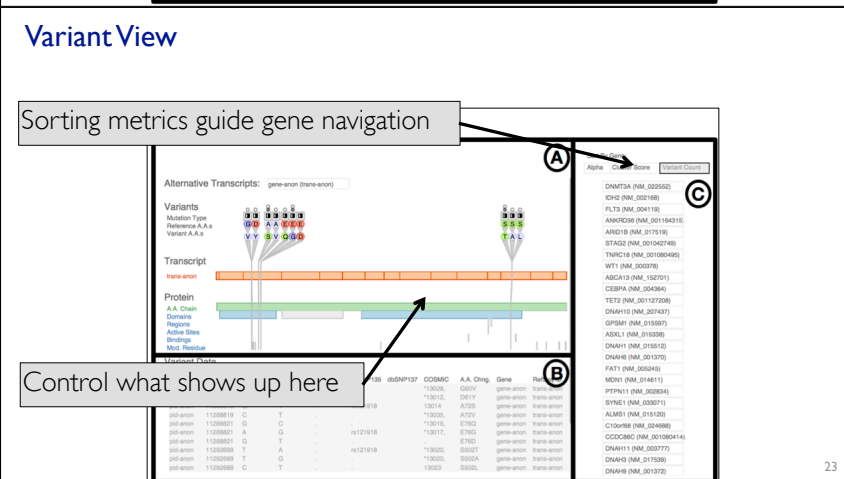
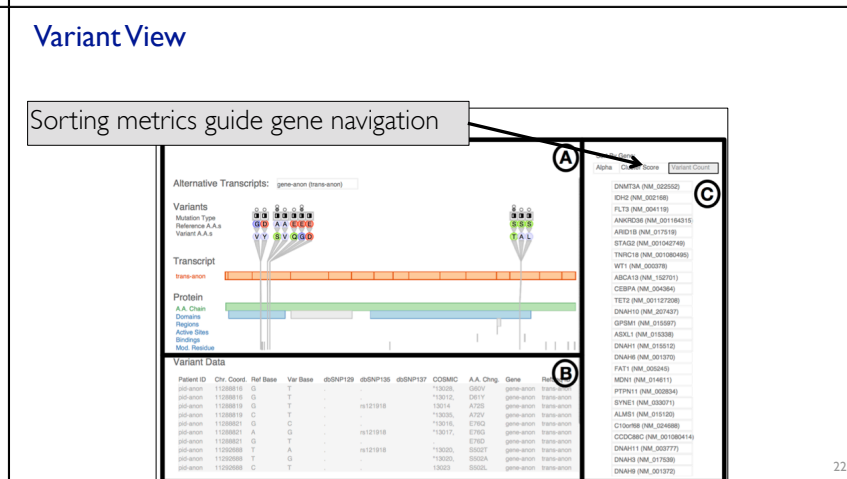
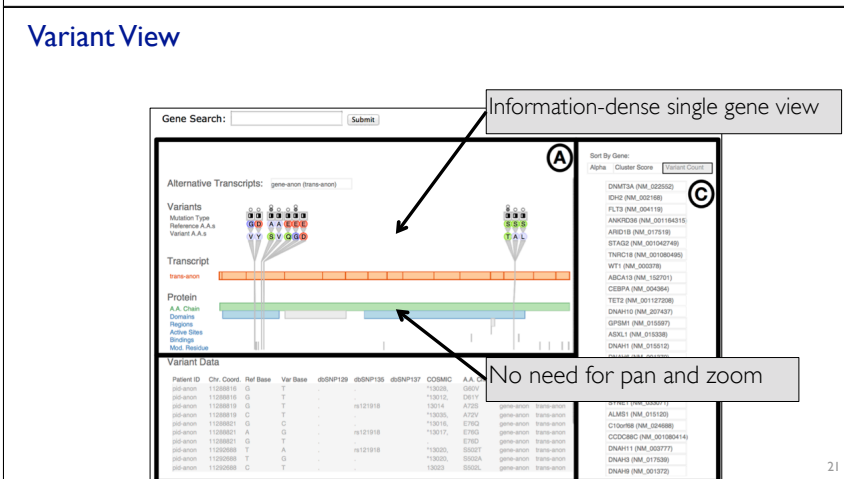
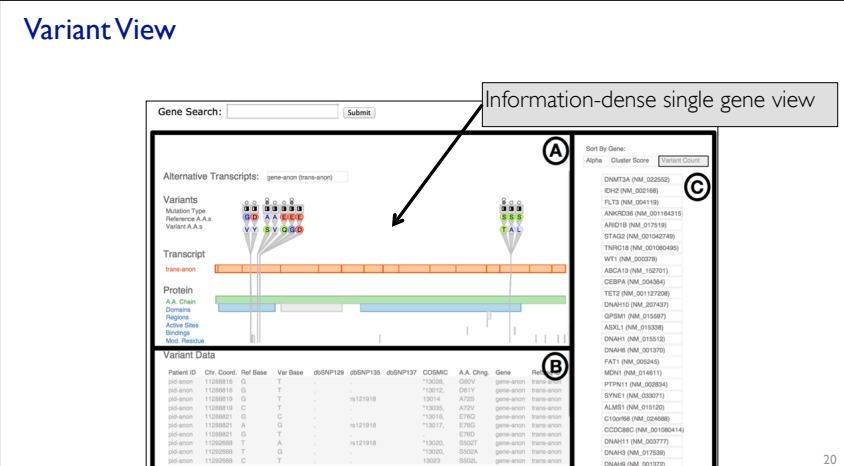
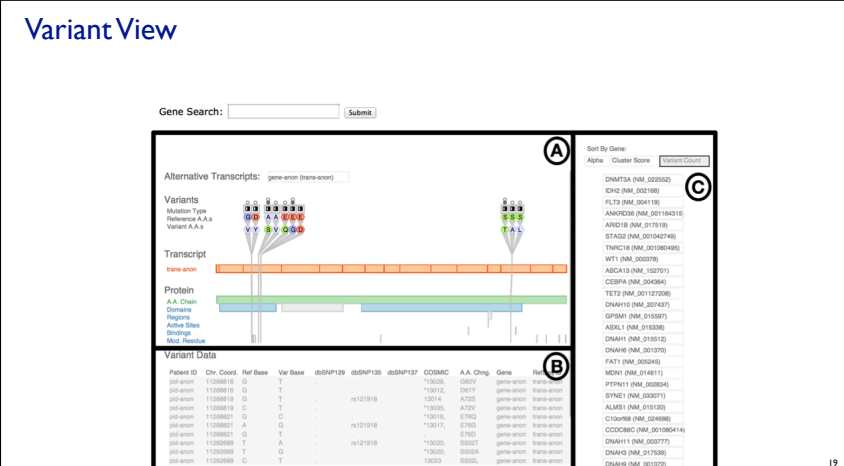
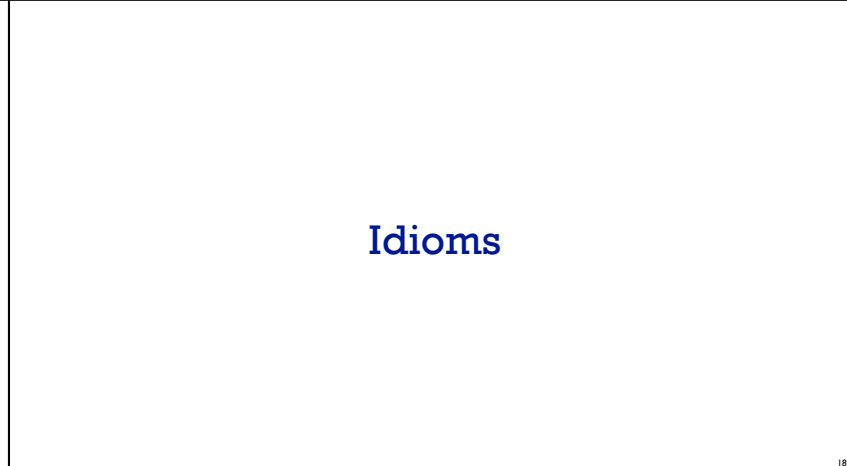
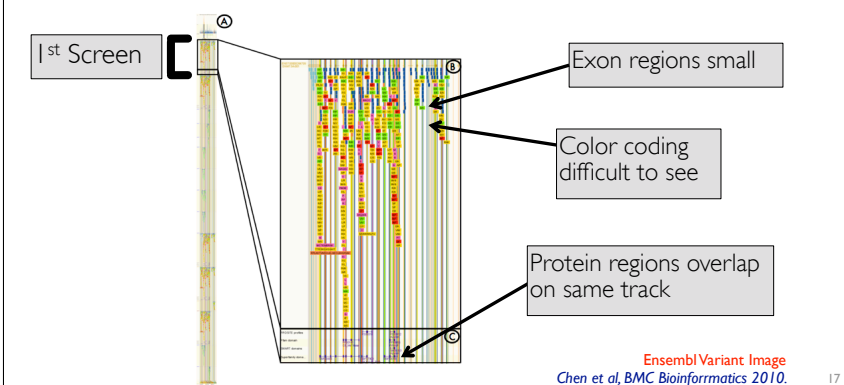
Dominant paradigm: genome browsers

- 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

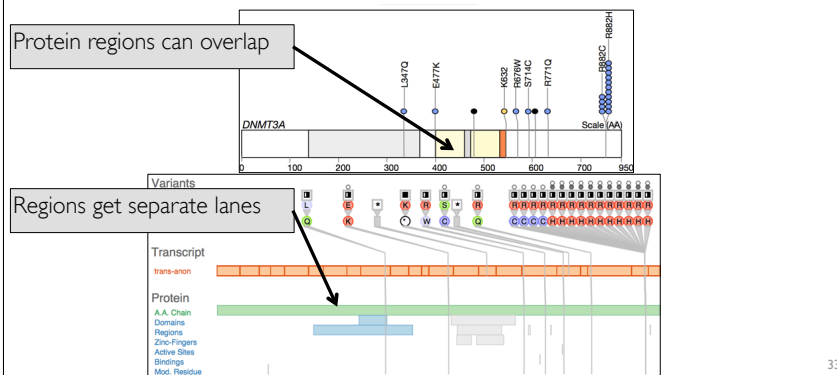


representative example: Ensembl
 Chen et al, BMC Bioinformatics 2010.

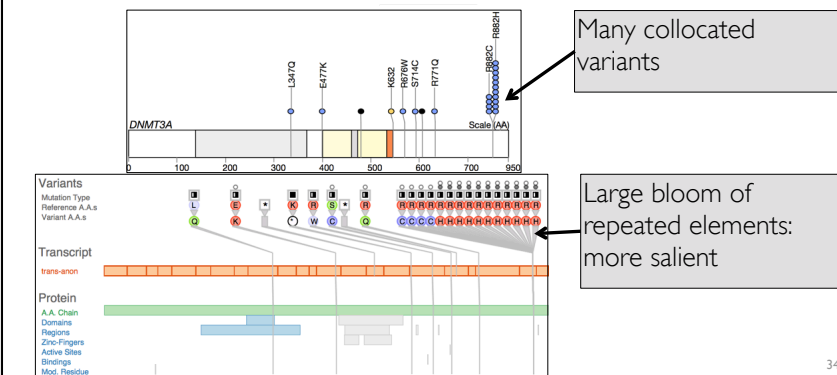
Features of interest small even in variant-specific view



Side-by-side comparison: MuSiC vs Variant View



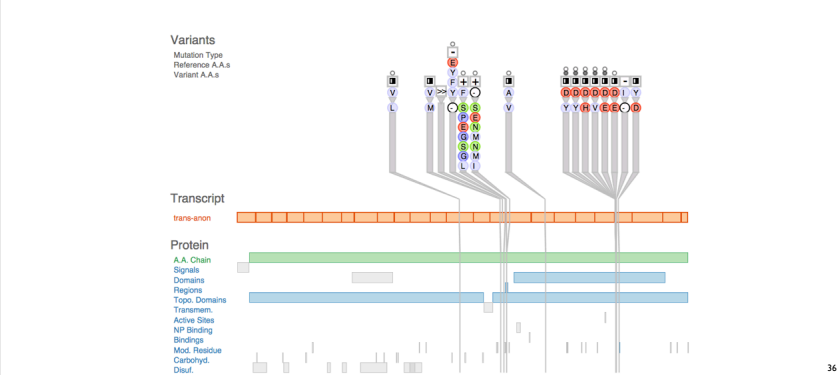
Side-by-side comparison: MuSiC vs Variant View



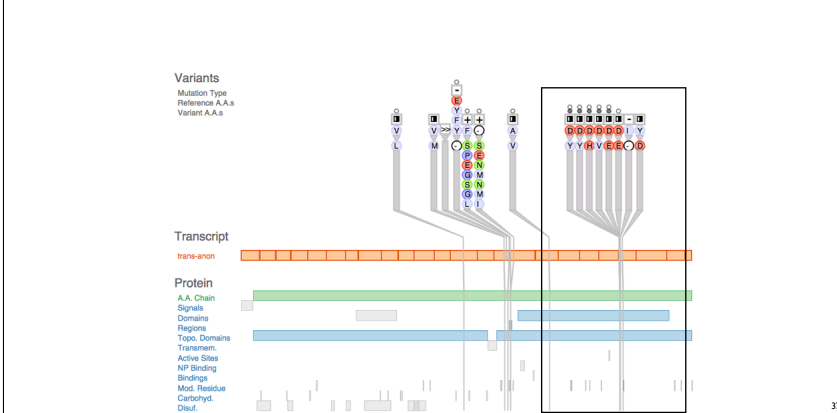
Verify known leukemia gene: Highly scored by sorting metric



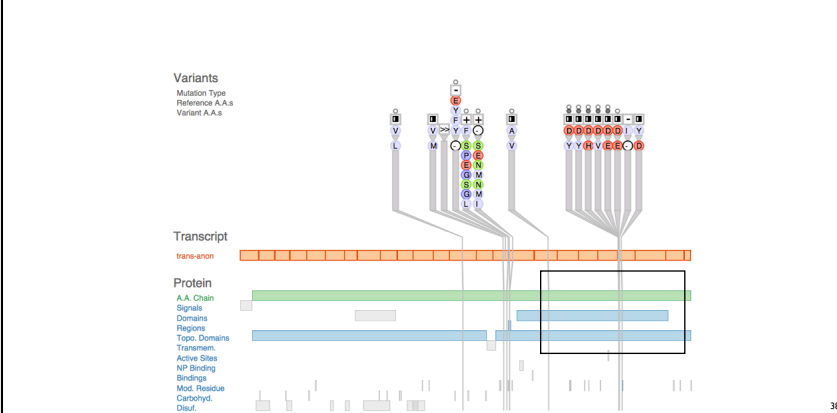
Verify known leukemia gene: Highly scored by sorting metric



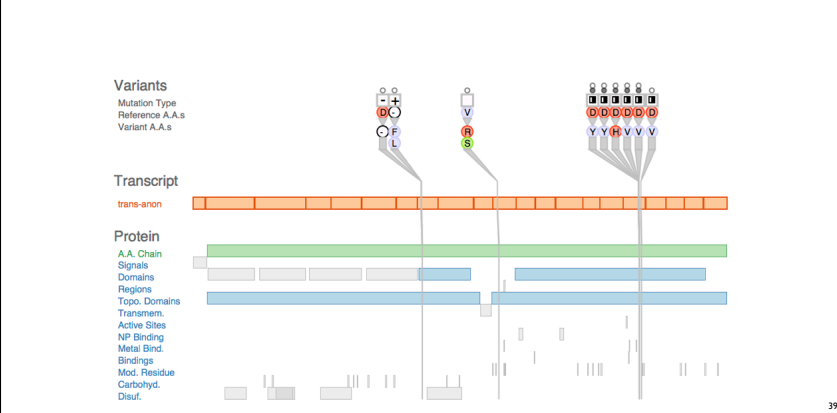
Visual inspection reveals collocation of variants



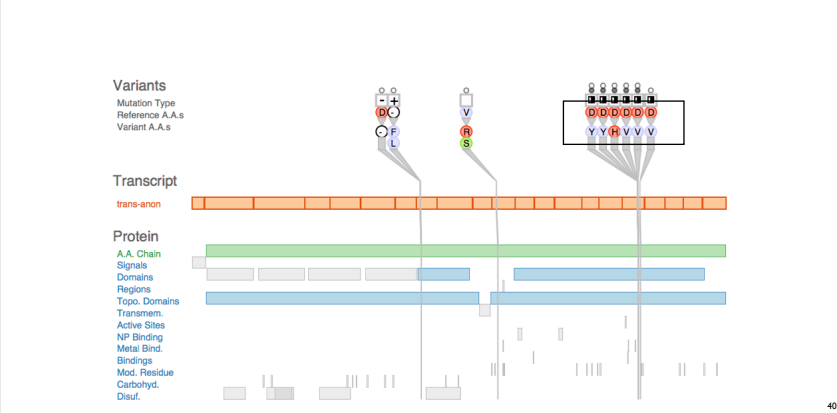
Several functional protein regions affected



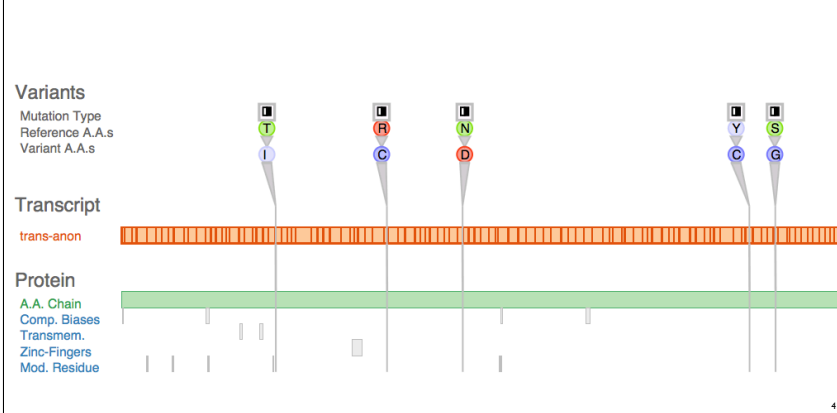
Highly scored by metric: not previously known, good candidate



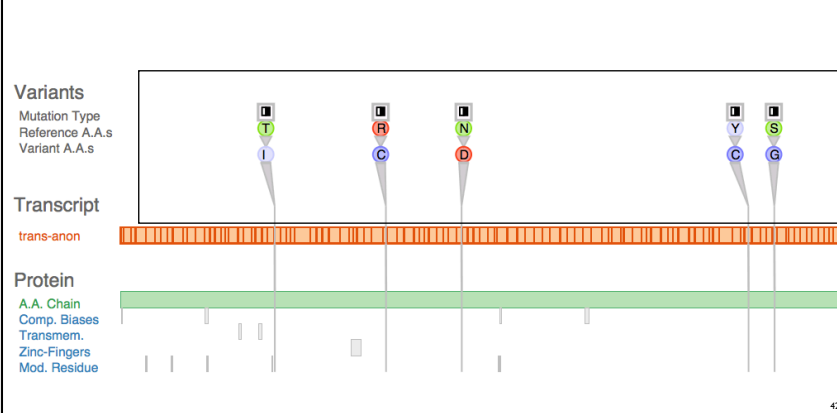
Protein chemical class change evident



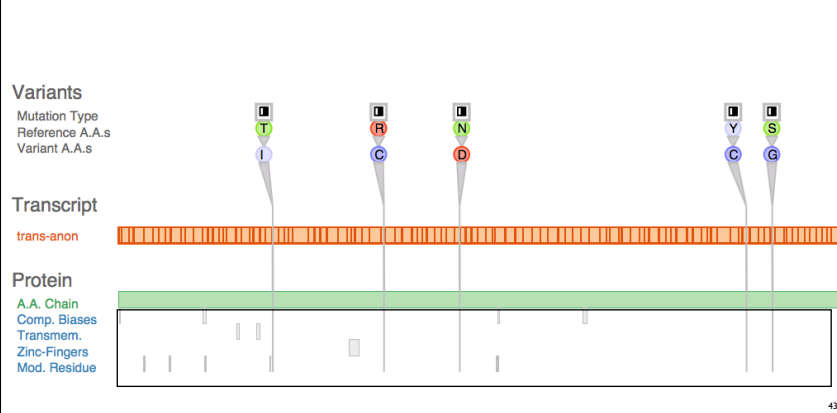
In contrast, low scoring gene



No collocation of variants



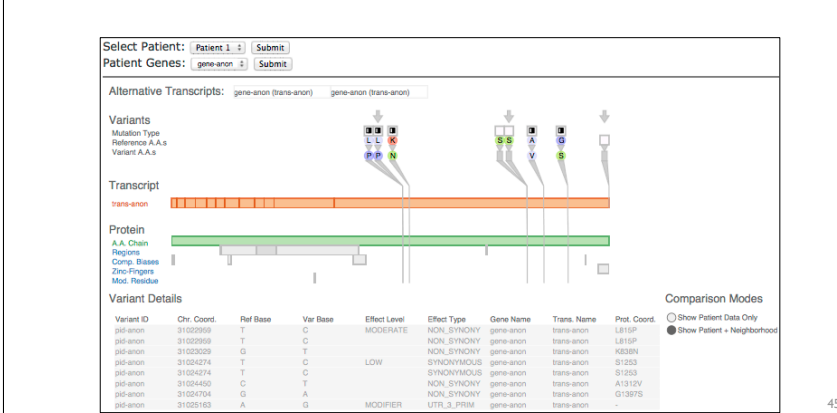
Mostly unaffected protein regions



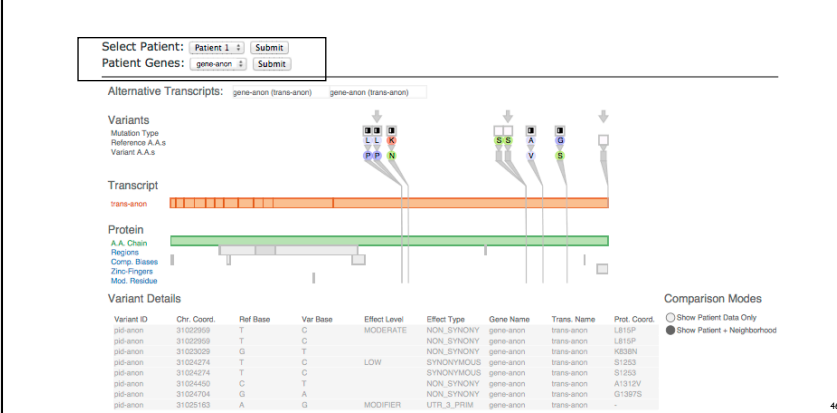
Additional tasks

- task 2: compare patients
 - clinical setting application
 - compare patient data to known harmful variants
 - challenge
 - similarity is loosely understood rather than fully characterized
 - visual inspection for what constitutes a match
- 44

Adapted Variant View with minimal changes



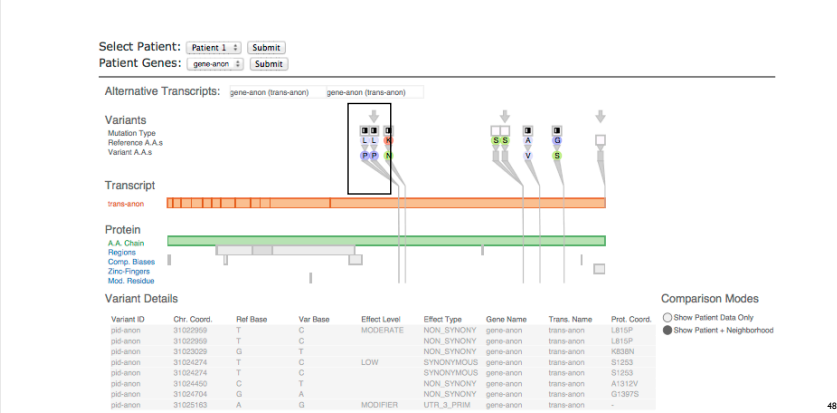
Navigate through patient data with list



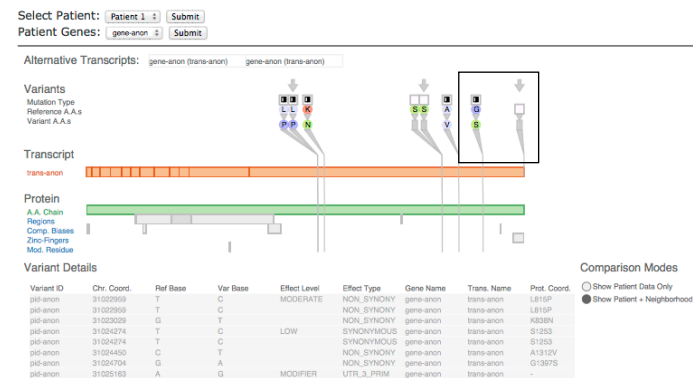
Patient data emphasized with arrows



Patient has same harmful L to P mutation



Nonmatching variants



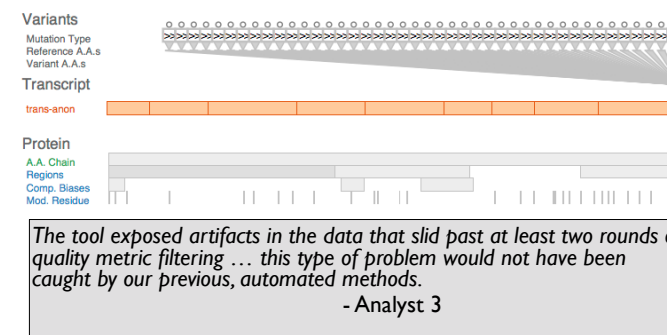
49

Additional tasks

- task 3: debug pipeline
 - data cleansing before analysis
 - analysts originally thought pipeline fully debugged
 - no perceived need for vis support

50

Tool revealed errors in the data



51

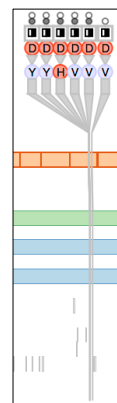
Reflections: vis design guidelines

- transferrable to other domains
 - specialize first, generalize later
 - good for domains where with complex, multi-scale data
 - difficult to judge a priori which design elements will generalize
 - high-level considerations
 - identifying scales of interest
 - what to visually encode directly vs what to support through interaction
 - when (and how) to eliminate navigation

52

Conclusions

- visual variant impact assessment
 - designed, implemented, and deployed tool for
- originally designed for Discover Genes task
 - adapted to two others with minimal changes
- features
 - navigation-free main overview at gene level
 - reveal genes of interest through sorting by new derived metrics
- major considerations
 - what to show
 - filtering data scope
 - how to show it
 - carefully selected visual encodings



53

Further Information

- paper page
 - <http://www.cs.ubc.ca/labs/imager/tr/2012/VariantView/>
- open source software download
 - <http://www.cs.ubc.ca/labs/imager/tr/2013/VariantView/VariantViewSoftware/>
- further info
 - <http://www.cs.ubc.ca/~tmm/talks.html#bioit14> (this talk, and many others)
 - <http://www.cs.ubc.ca/group/infovis> (papers, software, videos)
- acknowledgements
 - funding: Vancouver Institute for Visual Analytics (VIVA), Aeroinfo/Boeing, Mitacs
 - collaborators at the GSC
 - Dr. Aly Karsan
 - Rod Docking
 - Dr. Linda Chang
 - Dr. Gerben Duns
 - Simon Chang

54