Data Visualization in Genomics and In-Car Network Engineering

Tamara Munzner
Department of Computer Science
University of British Columbia

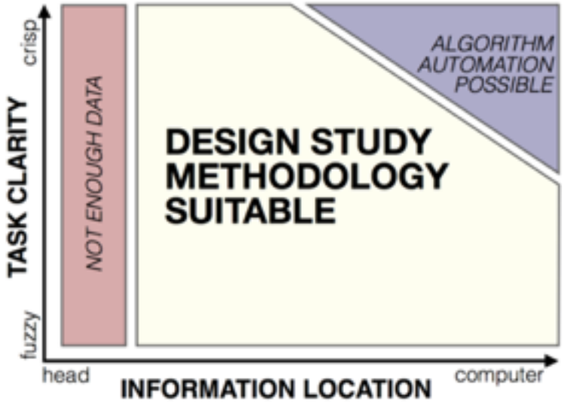
City University of London, Computer Science Department Seminar

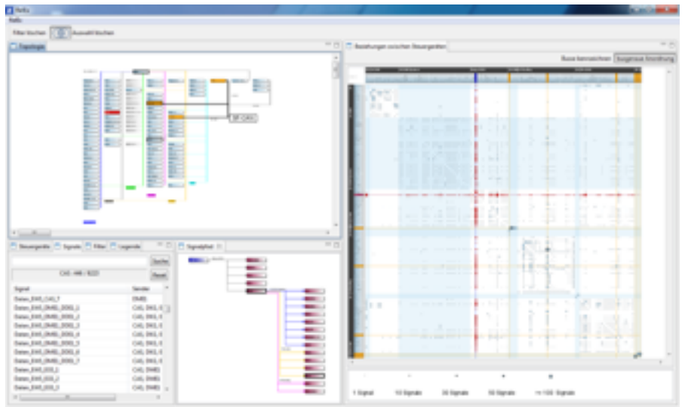
1 July 2014, London UK

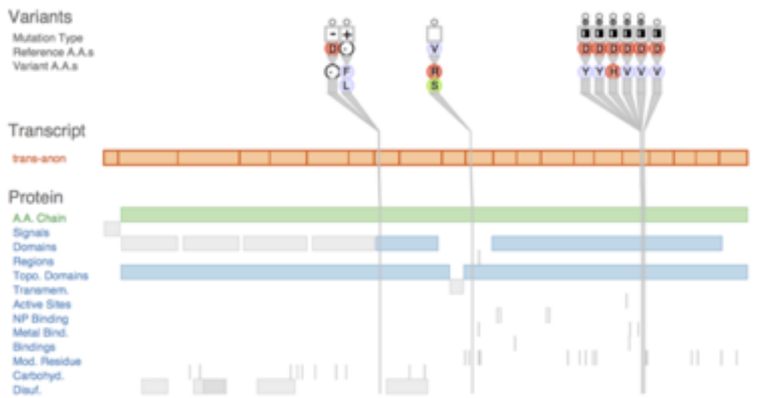
http://www.cs.ubc.ca/~tmm/talks.html#london14

Outline

- 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

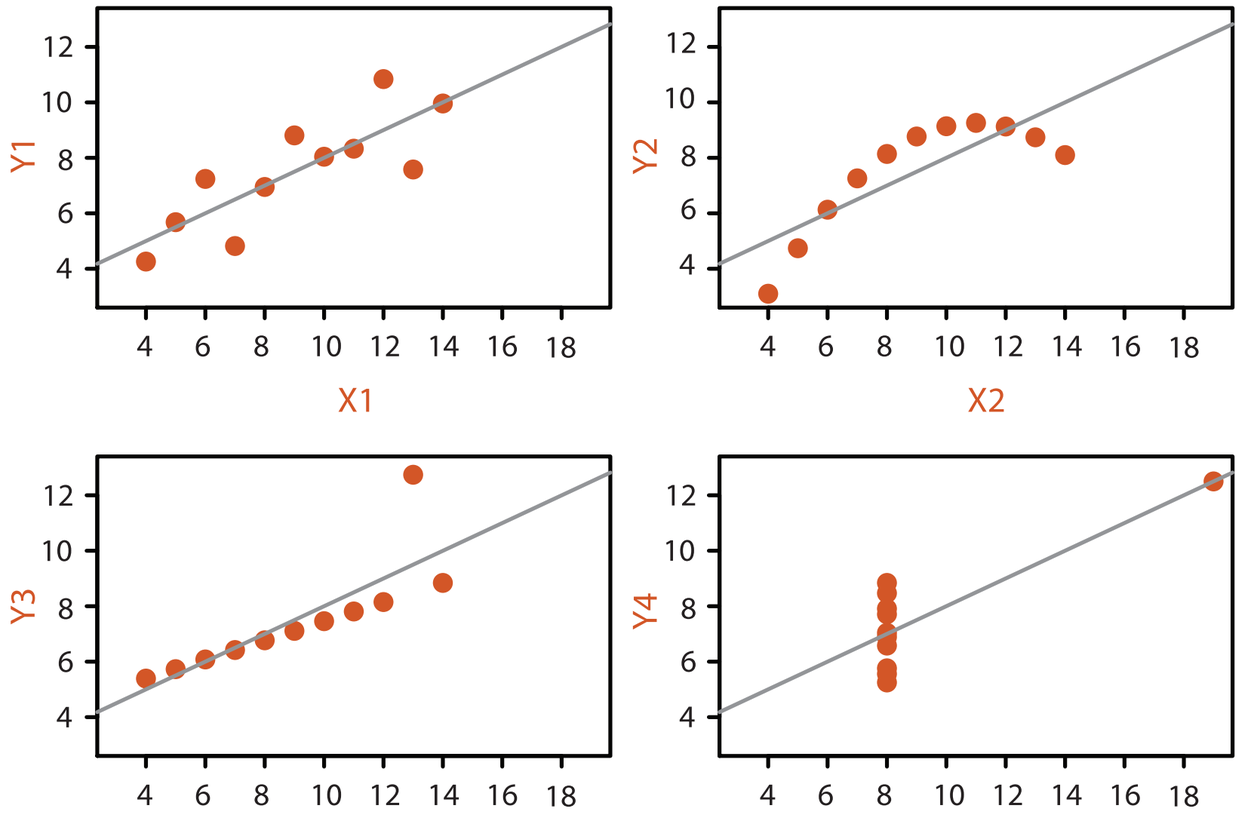
Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

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

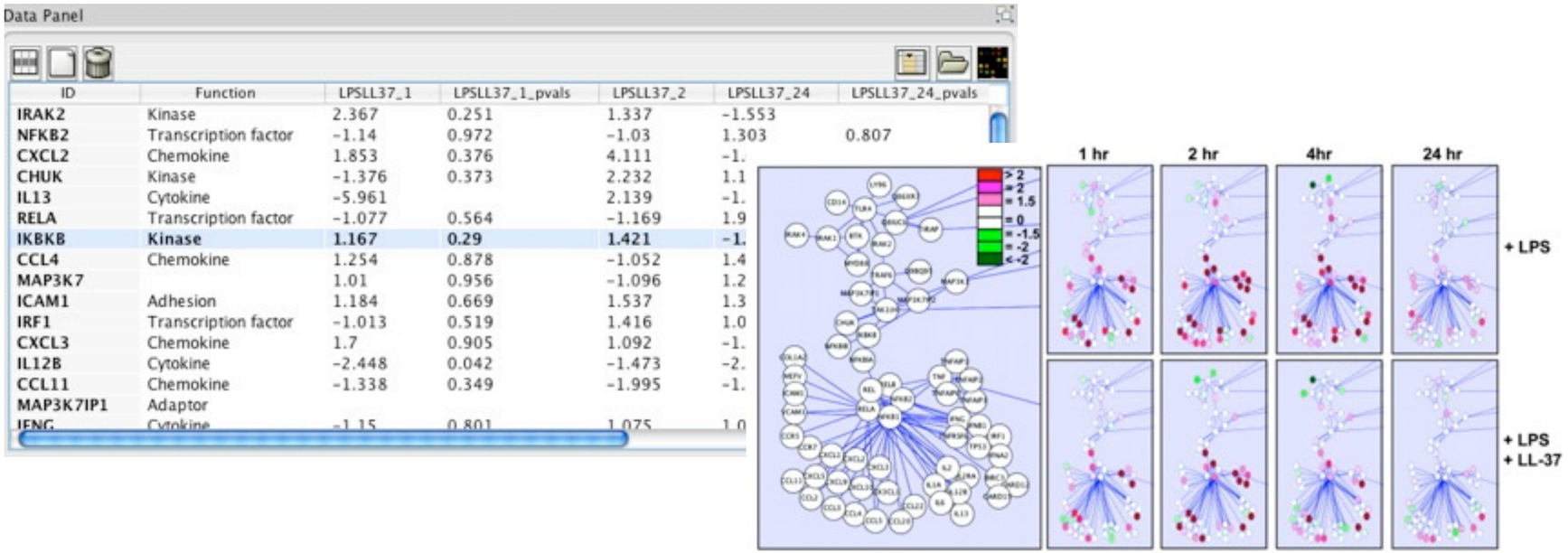


http://upload.wikimedia.org/wikipedia/commons/b/b6/Anscombe.svg

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.

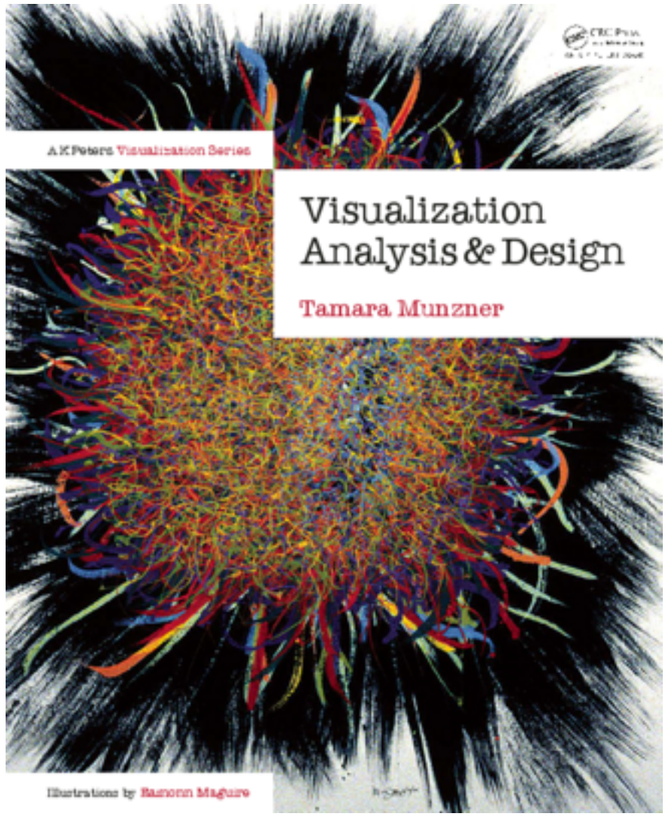
- human in the loop needs the details
  - doesn't know exactly what questions to ask in advance
- external representation: perception vs cognition
- intended task

http://upload.wikimedia.org/wikipedia/commons/b/b6/Anscombe.svg

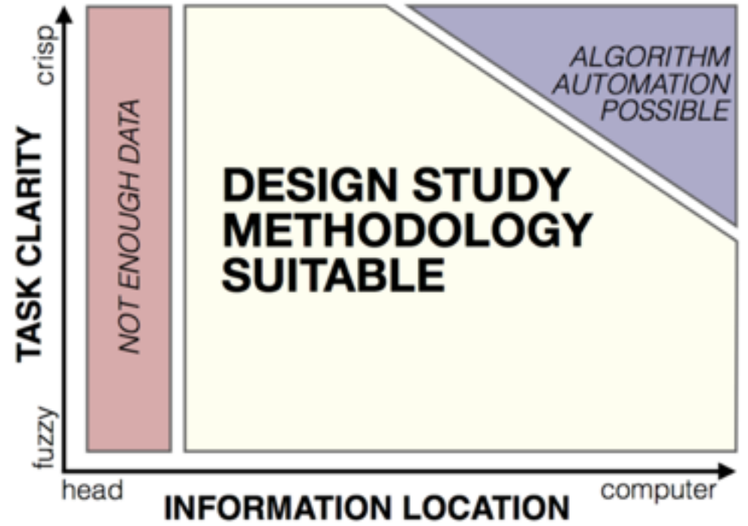
Defining Visualization

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
- measureable 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).

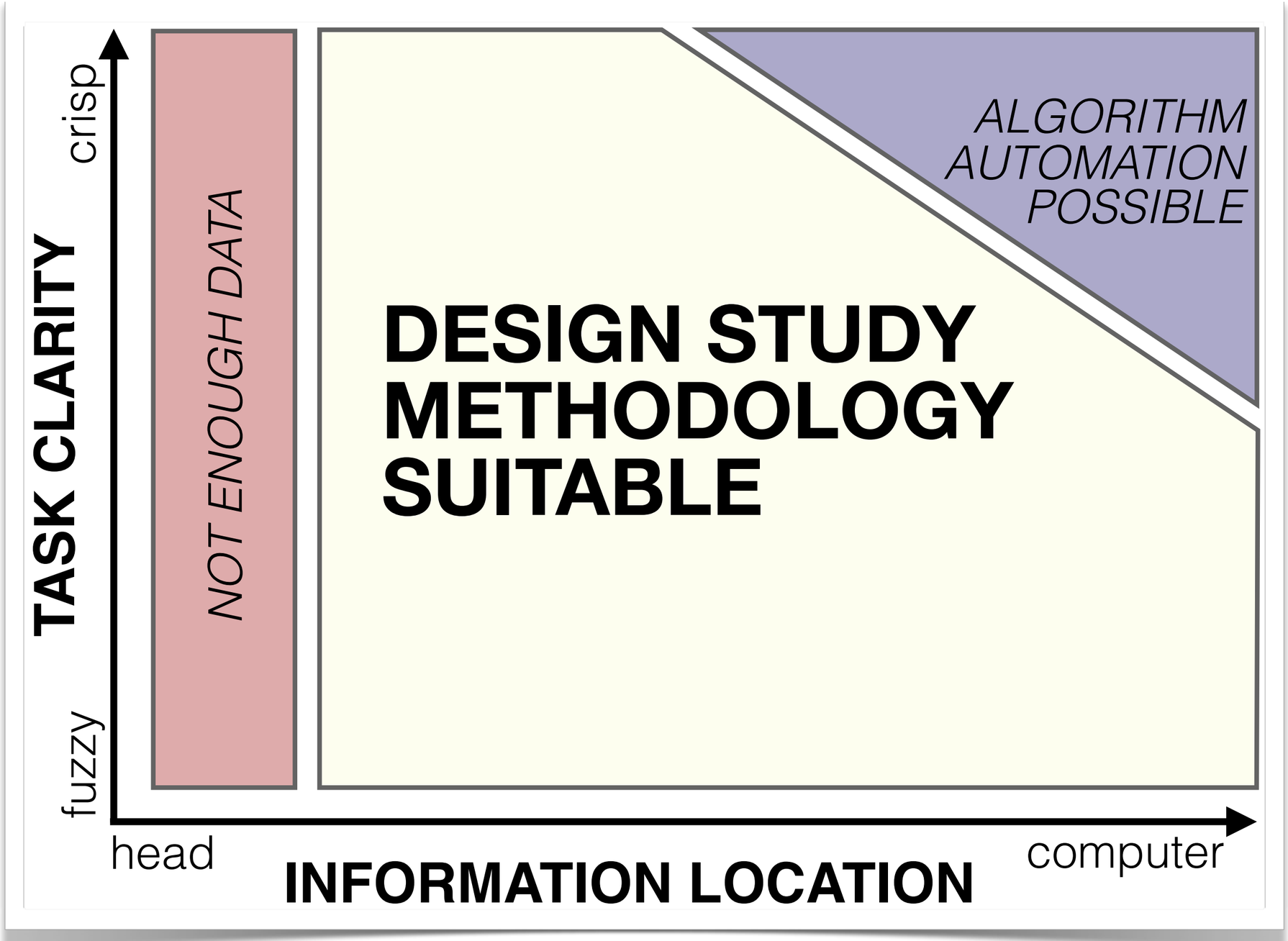
Defining Design Study

- 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

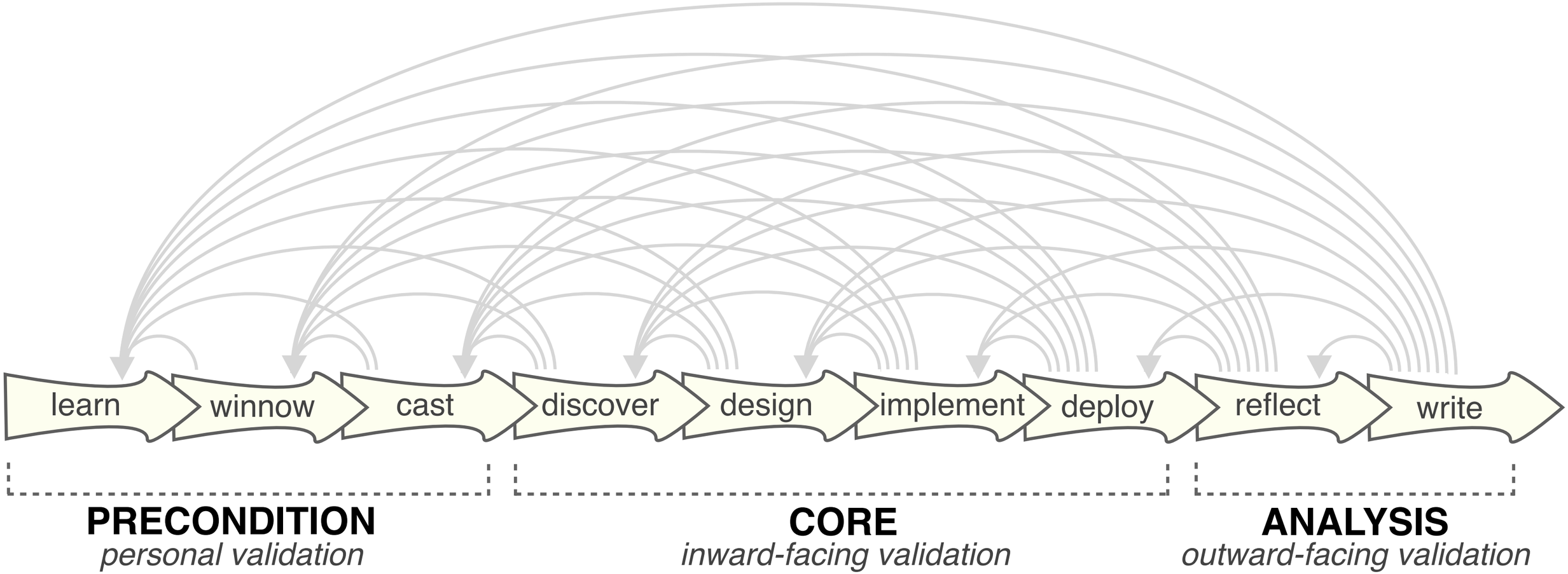
more at:
A Nested Model of Visualization Design and Validation.
Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).

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

When To Do Design Studies

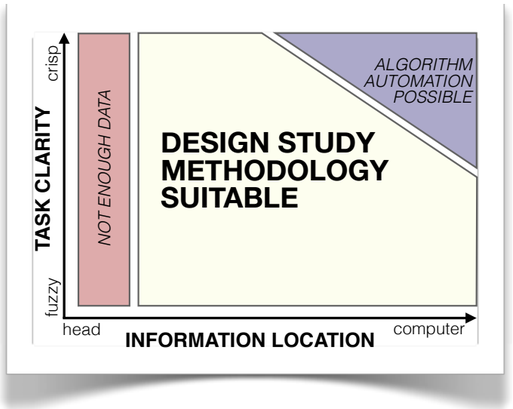


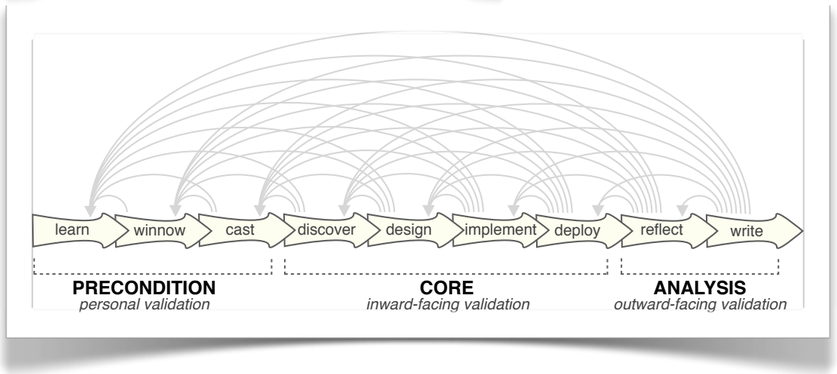
Nine-Stage Framework



How To Do Design Studies

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





PF-1	premature advance: jumping forward over stages	general
PF-2	premature start: insufficient knowledge of vis literature	learn
PF-3	premature commitment: collaboration with wrong people	winnow
PF-4	no real data available (yet)	winnow
PF-5	insufficient time available from potential collaborators	winnow
PF-6	no need for visualization: problem can be automated	winnow
PF-7	researcher expertise does not match domain problem	winnow
PF-8	no need for research: engineering vs. research project	winnow
PF-9	no need for change: existing tools are good enough	winnow

Pitfall Example: Premature Publishing

algorithm innovation

design studies

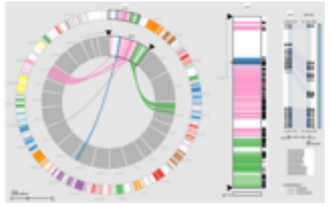
Must be first!

Am I ready?

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

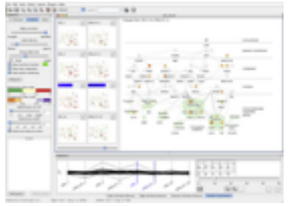
http://www.alaineknipes.com/interests/violin_concert.jpg

Design Studies: Lessons learned after 21 of them

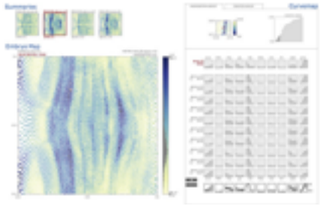


MizBee
genomics

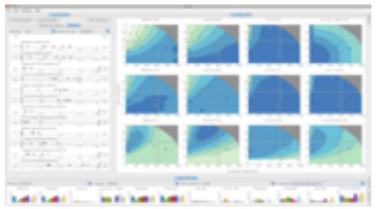
Pathline
genomics



Cerebral
genomics



MulteeSum
genomics

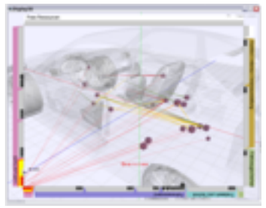


Vismon
fisheries management

QuestVis
sustainability

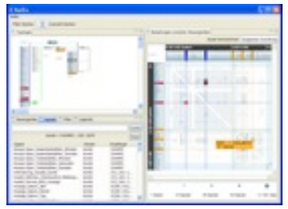
WiKeVis
in-car networks

MostVis
in-car networks

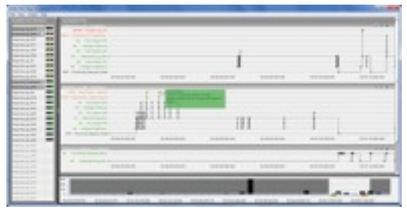


Car-X-Ray
in-car networks

ProgSpy2010
in-car networks



RelEx
in-car networks

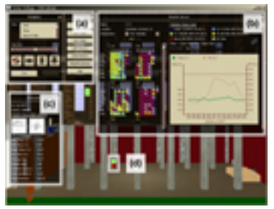


Cardiogram
in-car networks

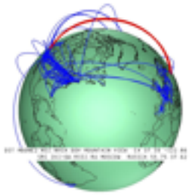
AutobahnVis
in-car networks

VisTra
in-car networks

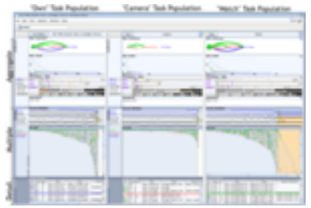
Constellation
linguistics



LibVis
cultural heritage

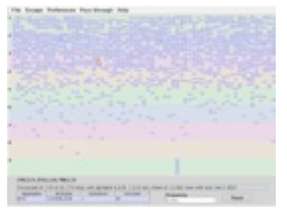


Caidants
multicast

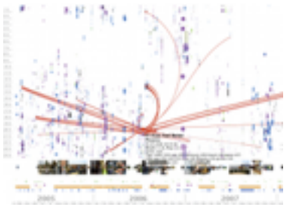


SessionViewer
web log analysis

LiveRAC
server hosting



PowerSetViewer
data mining

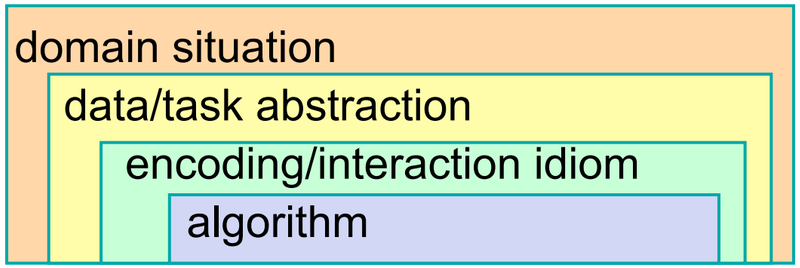


LastHistory
music listening

- commonality of representations cross-cuts domains!

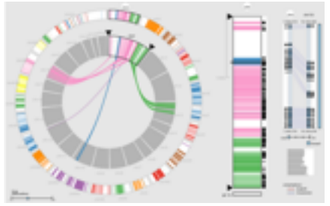
Abstractions and Idioms

- 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



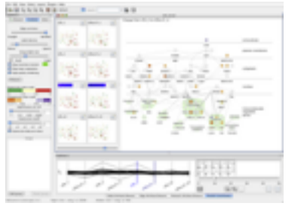
A Nested Model of Visualization Design and Validation.
Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).

Today's Focus

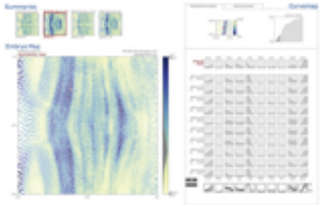


MizBee
genomics

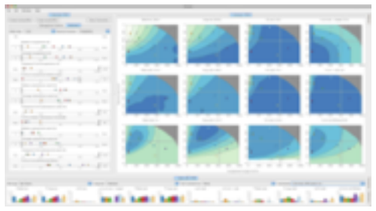
Pathline
genomics



Cerebral
genomics



MulteeSum
genomics

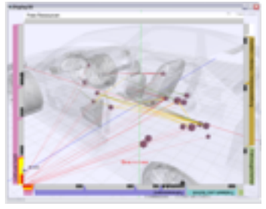


Vismon
fisheries management

QuestVis
sustainability

WiKeVis
in-car networks

MostVis
in-car networks



Car-X-Ray
in-car networks

ProgSpy2010
in-car networks

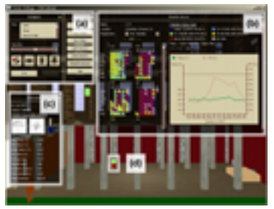
RelEx
in-car networks

Cardiogram
in-car networks

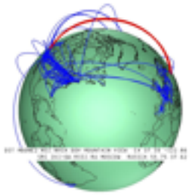
AutobahnVis
in-car networks

VisTra
in-car networks

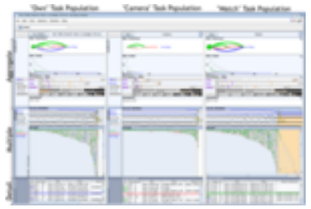
Constellation
linguistics



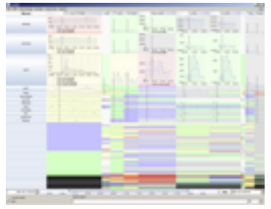
LibVis
cultural heritage



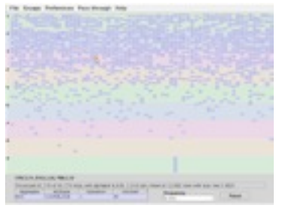
Caidants
multicast



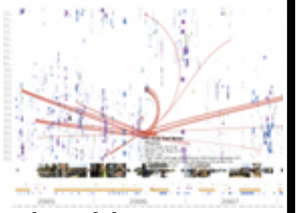
SessionViewer
web log analysis



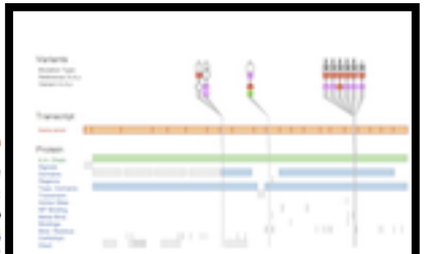
LiveRAC
server hosting



PowerSetViewer
data mining



LastHistory
music listening



VariantView
genomics

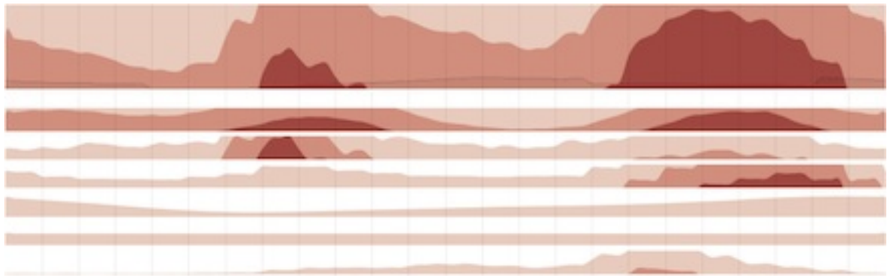
Design Studies: giCentre Context

- **methodology**

  Human-centered approaches in geovisualization design: investigating multiple methods through a long-term case study. *Lloyd and Dykes. IEEE Transactions on Visualization and Computer Graphics, 17(12):2498–2507, 2011.*

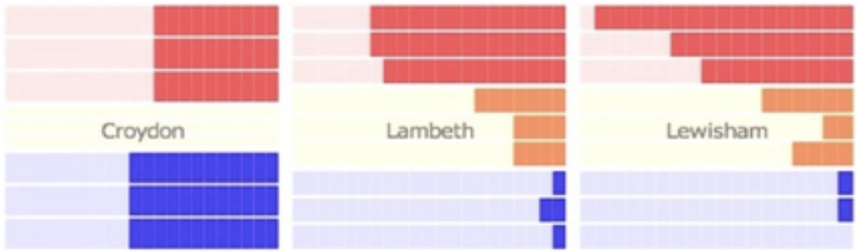
- **energy analysis**

  Creative user-centered visualization design for energy analysts and modelers. *Goodwin, Dykes, Jones, Dillingham, Dove, Duffy, Kachkaev, Slingsby, Wood. IEEE Transactions on Visualization and Computer Graphics, 19(12), pp. 2516-2525, 2013.*



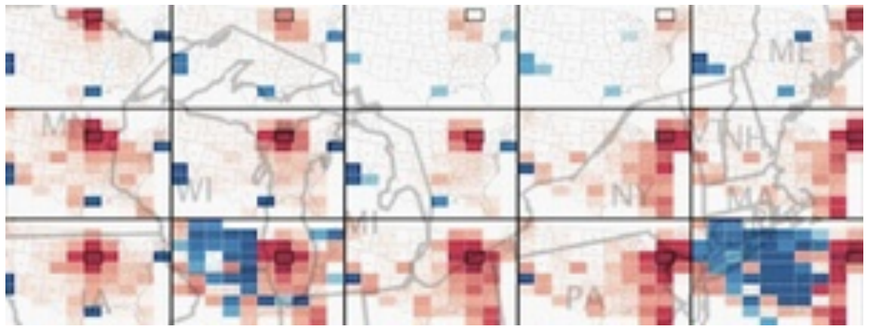
- **BallotMaps**

  BallotMaps: Detecting name bias in alphabetically ordered ballot papers. *Wood, Badawood, Dykes, Slingsby. IEEE Transactions on Visualization and Computer Graphics, 17(12), pp. 2384-2391, 2011*



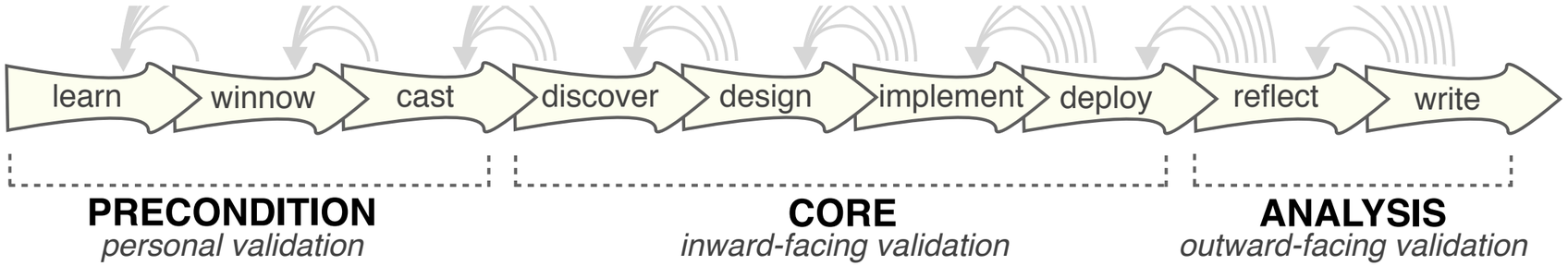
- **ODMaps**

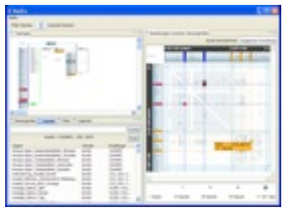
  Visualisation of origins, destinations and flows with OD maps. *Wood, Dykes, Slingsby. The Cartographic Journal, 47(2), pp. 117-129, 2010.*



Themes

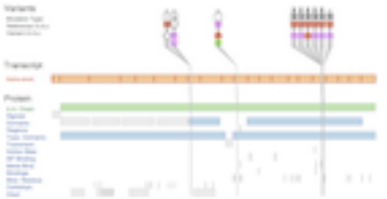
- task and data abstraction
  - both cases: complex and tricky
  - clear description in final talk/paper is end of a long, long road
    - writing as research: refine during reflection even after vis tool is finalized...



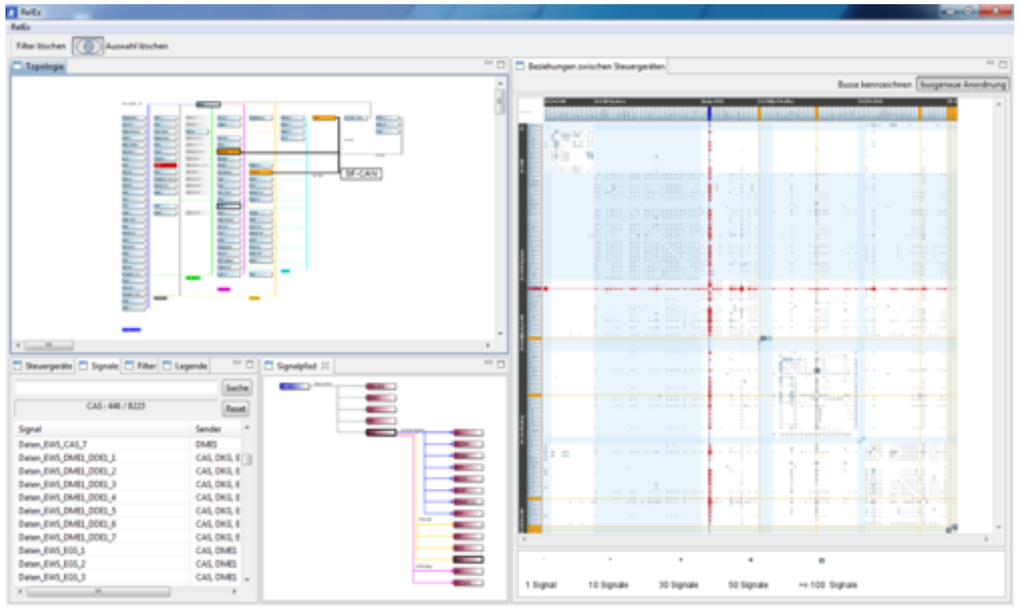


RelEx
in-car networks

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



VariantView
genomics



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**

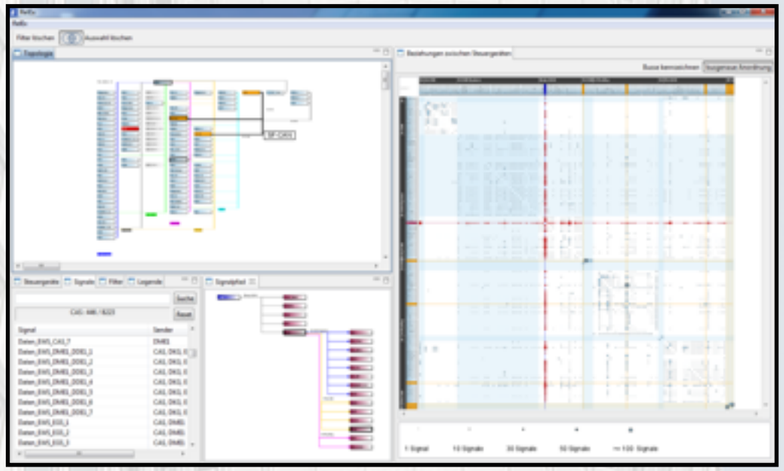


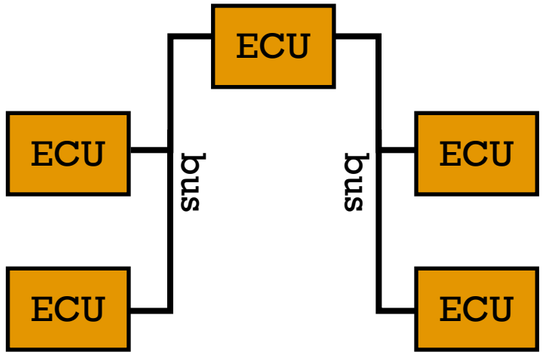
Image: http://www.bmwblog.com/wp-content/uploads/bmw-hq-munich-4.jpg

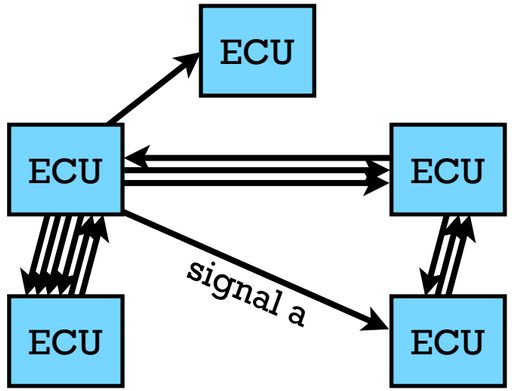
Abstractions

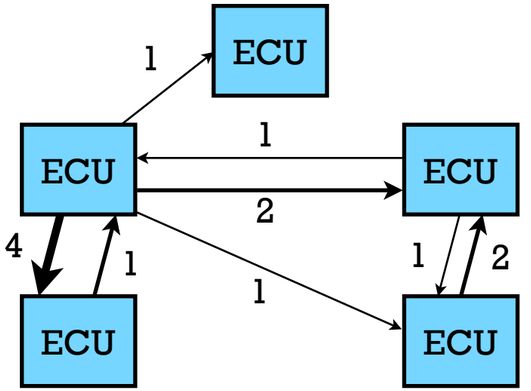
DATA
In-car Electronics

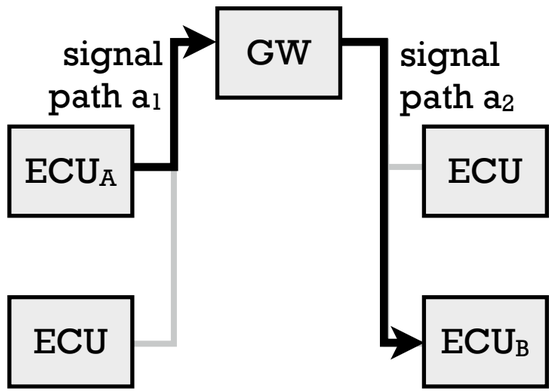
Data Abstraction: 3 Networks

- **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



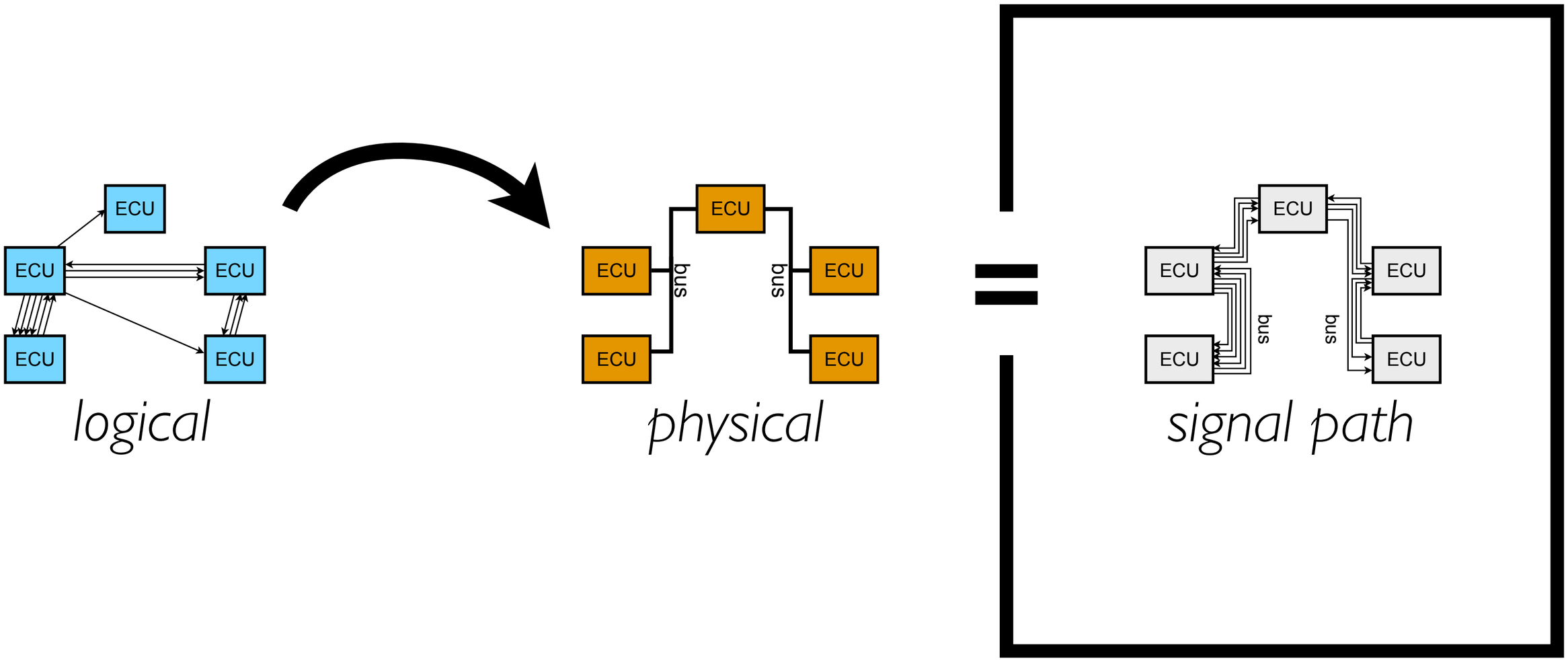






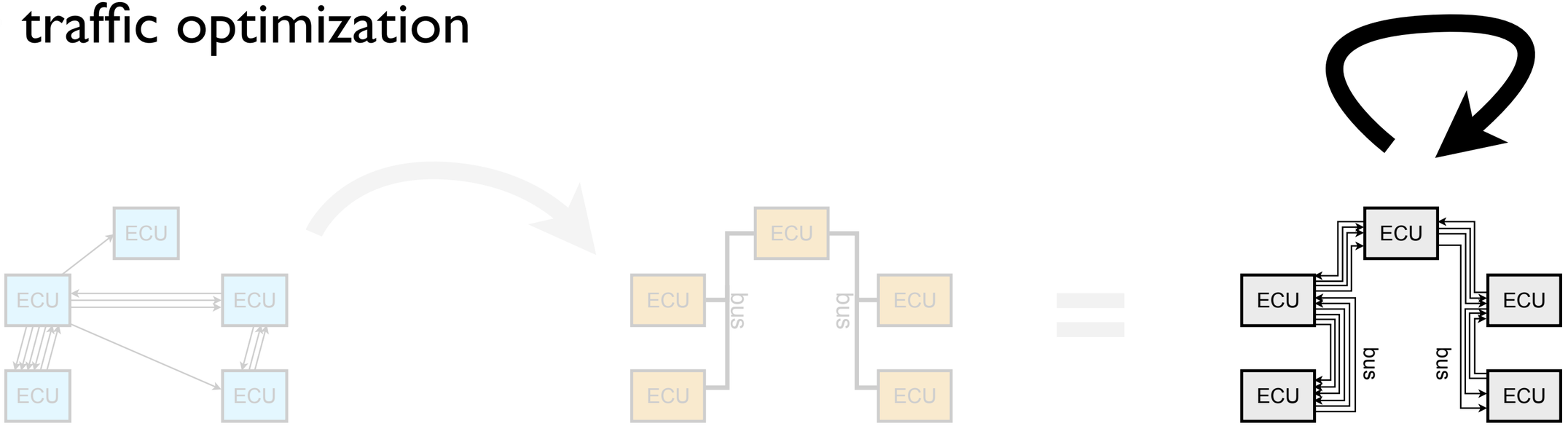
Task Abstraction: Mapping

- specify overlay network that maps logical onto physical



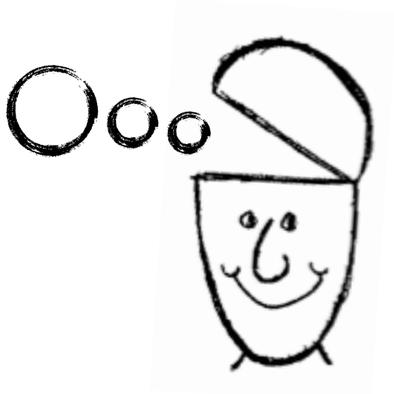
Task Abstraction: Optimizing

- traffic optimization



Many constraints

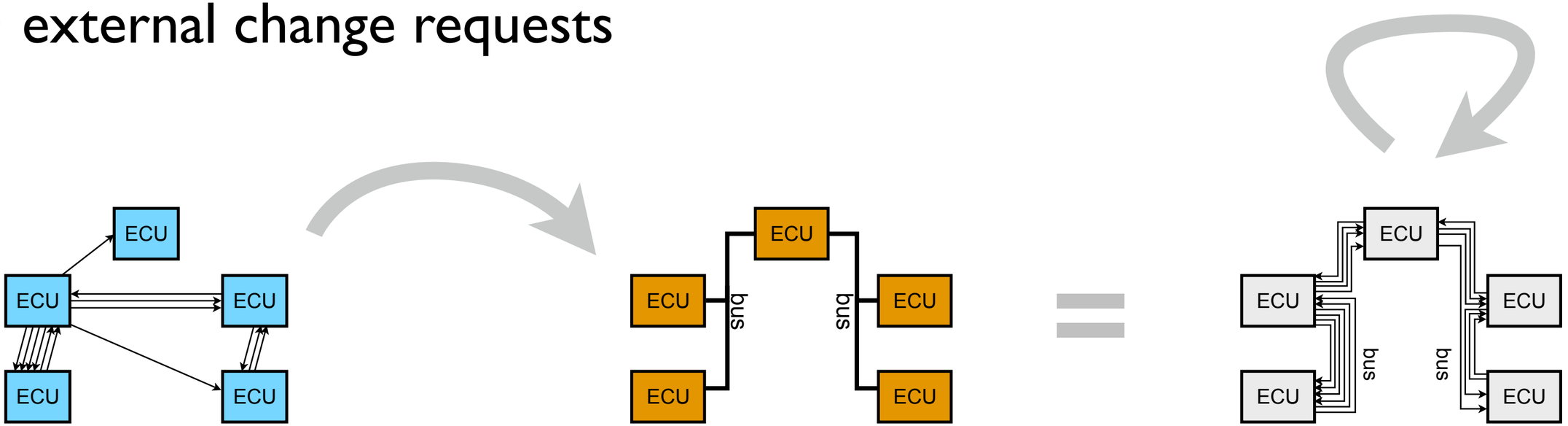
bandwidth ... delay/real time ...
path length ... load balance ...
reliability ... money ...



-- engineer, BMW --

Task Abstraction: Changing

- external change requests

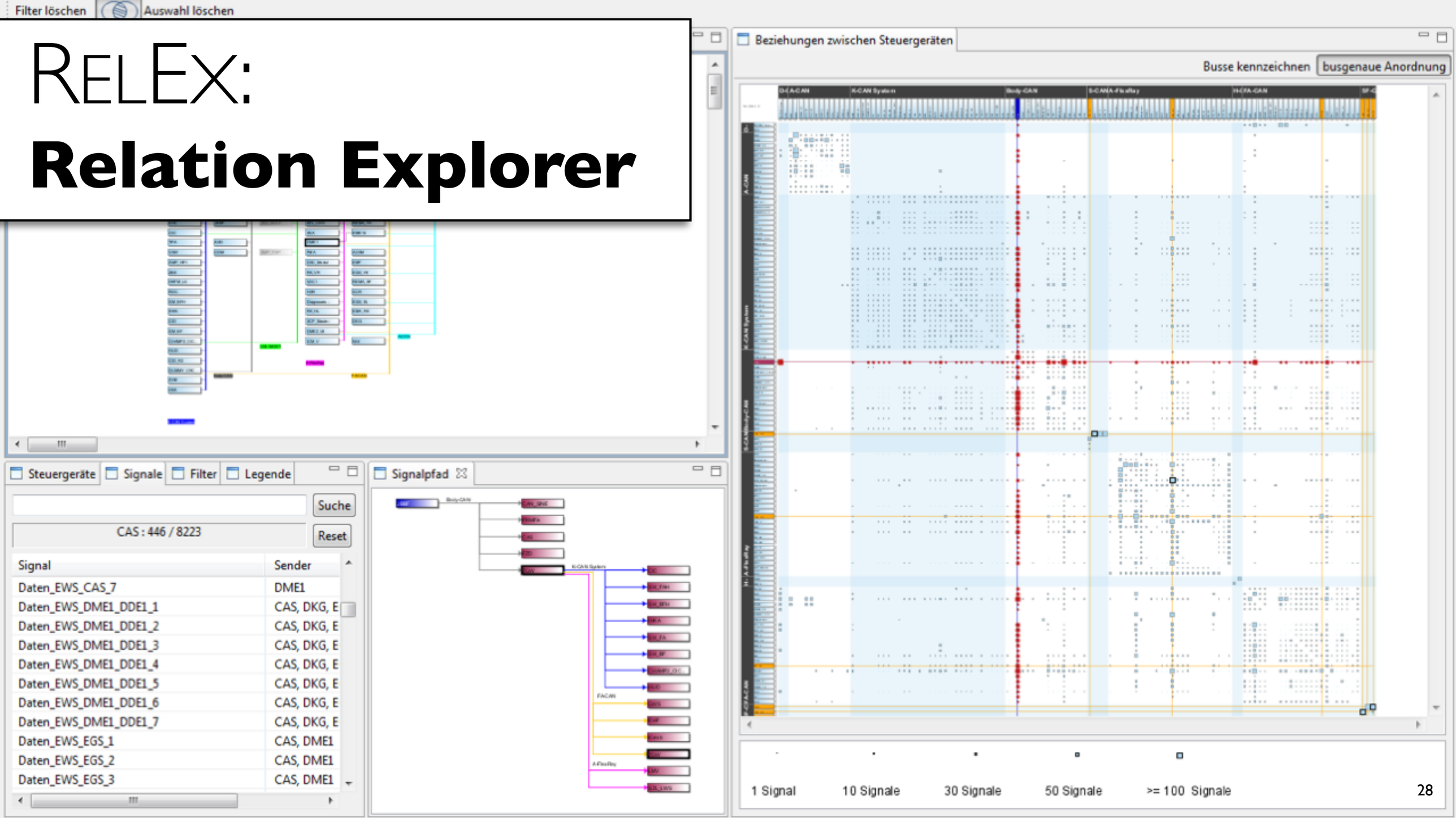


Change

(trivial requests might lead to complex changes)

Idioms

RELEX:
Relation Explorer

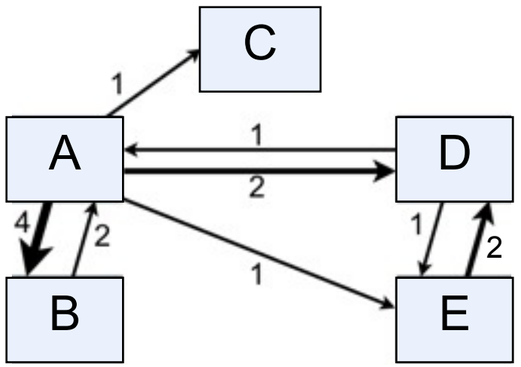


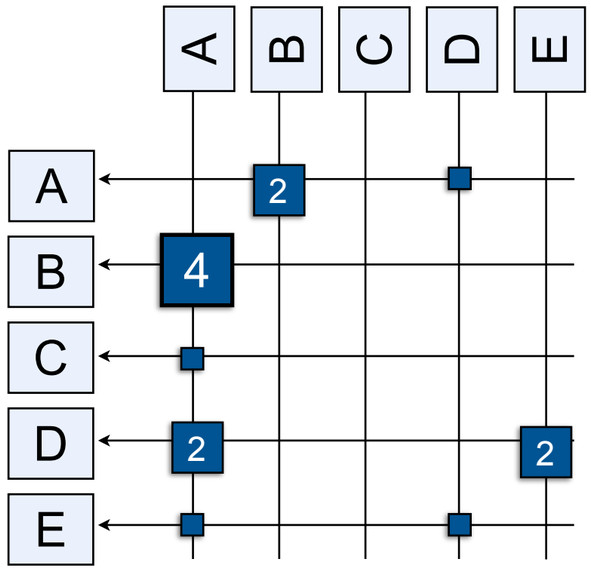
Vis Guideline [Ghoniem 2005]

Matrix for dense network data

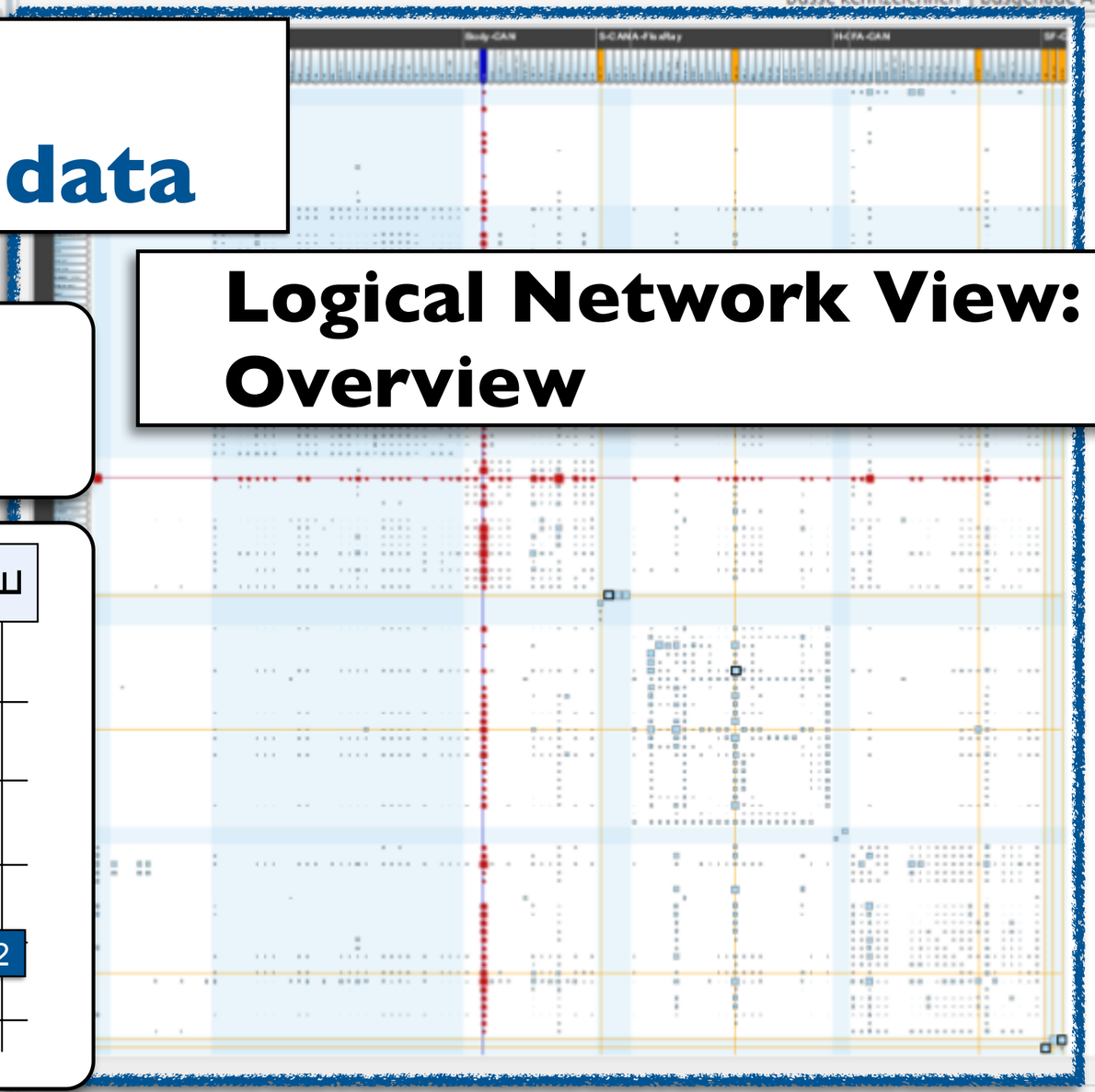
Signal Count Network

visual encoding:
size-coded matrix





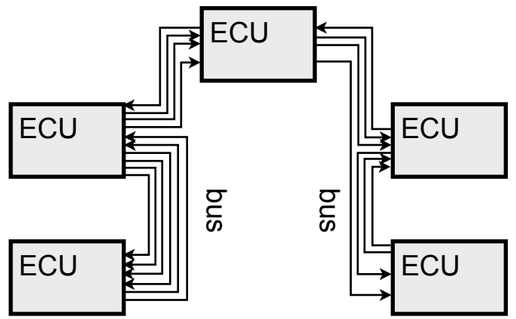
Logical Network View: Overview



Vis Guideline [Ghoniem 2005]

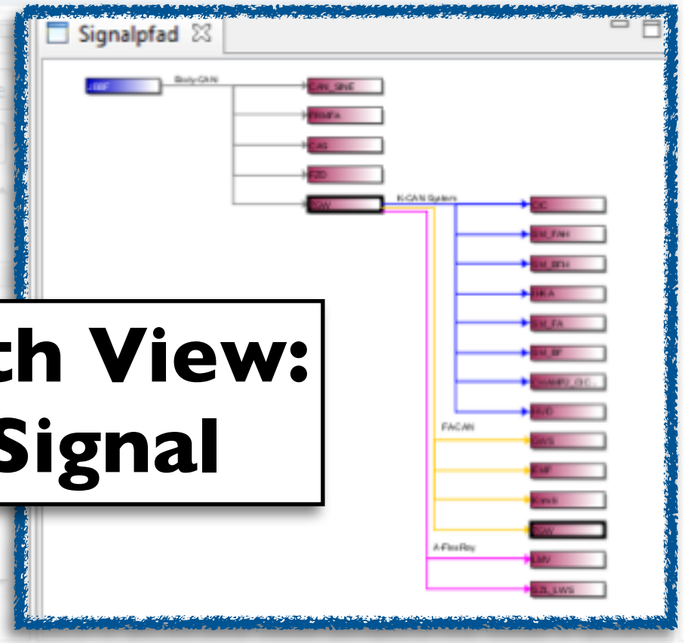
Node-link for path following tasks

Signal Path Network

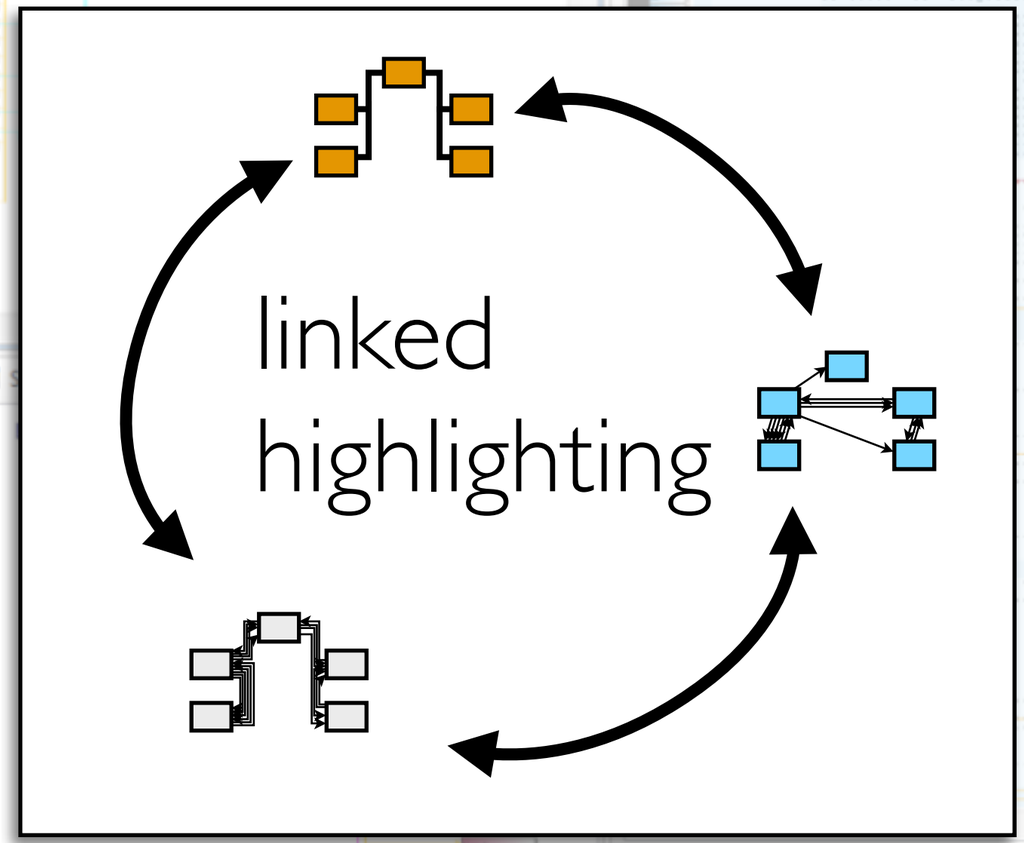


filtered by signal

Signal Path View: Selected Signal

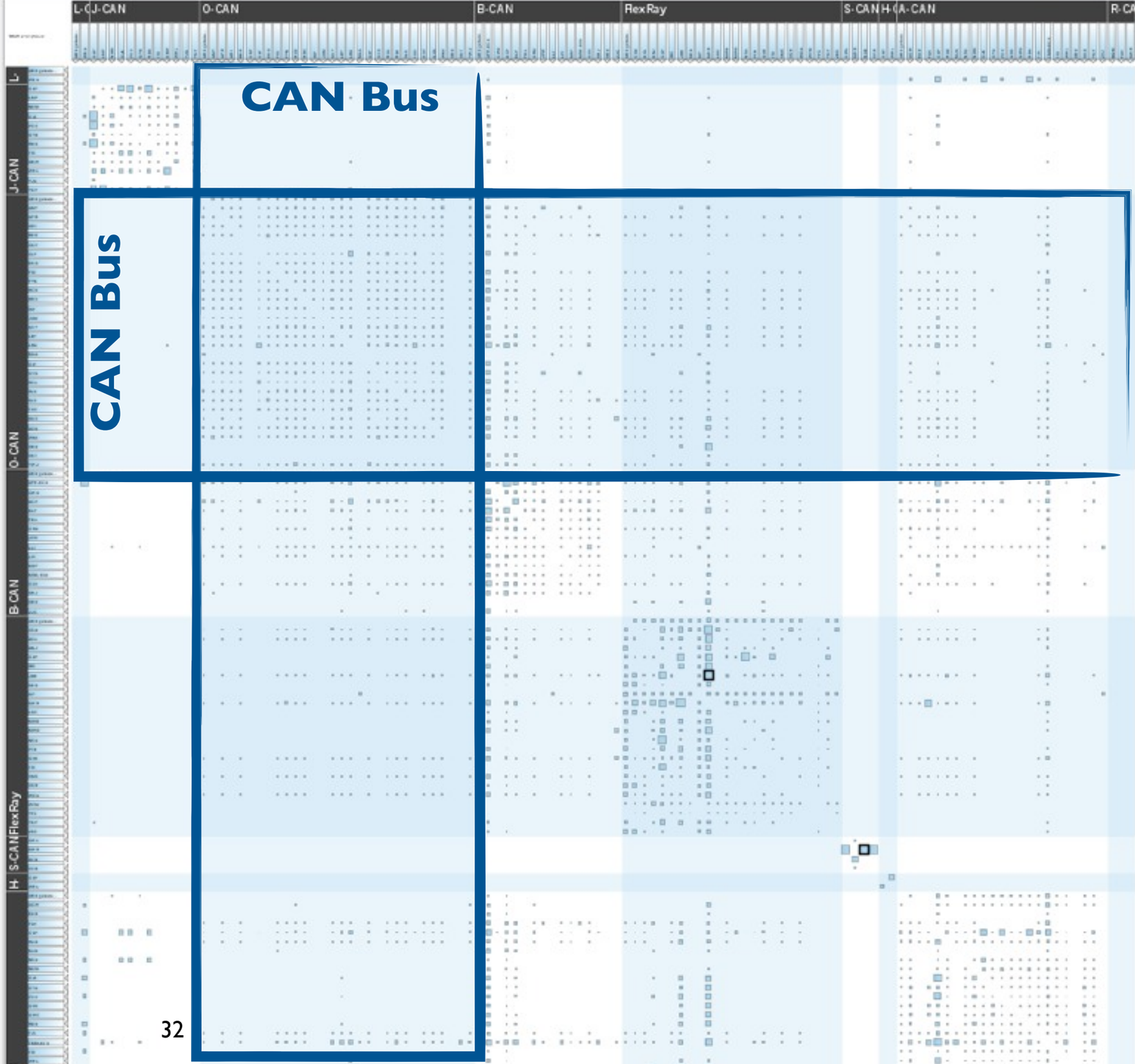


Interaction Idiom:
Cross-Network Relations



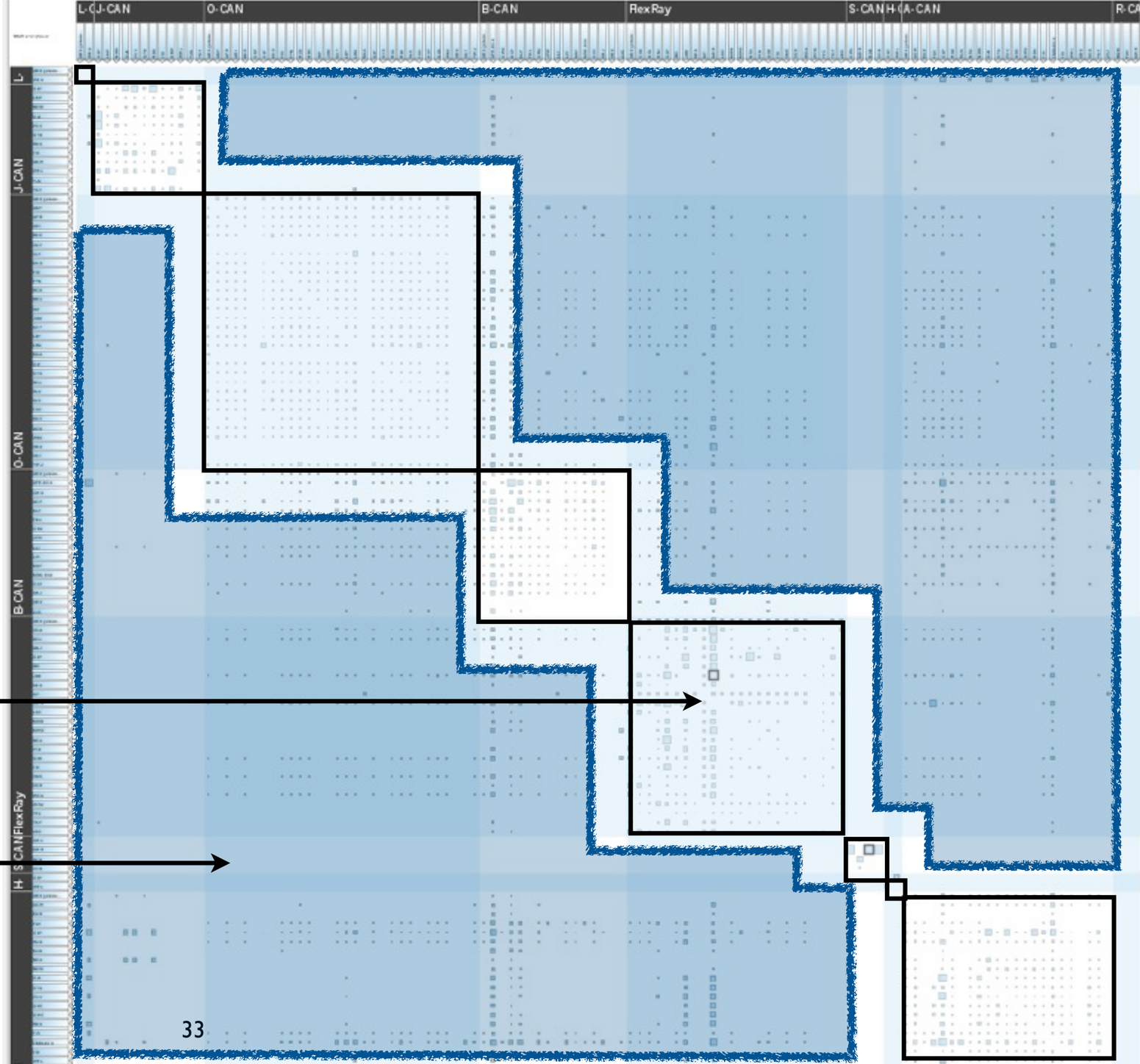
VIDEO

INTERESTS
Bus communication patterns



INTERESTS

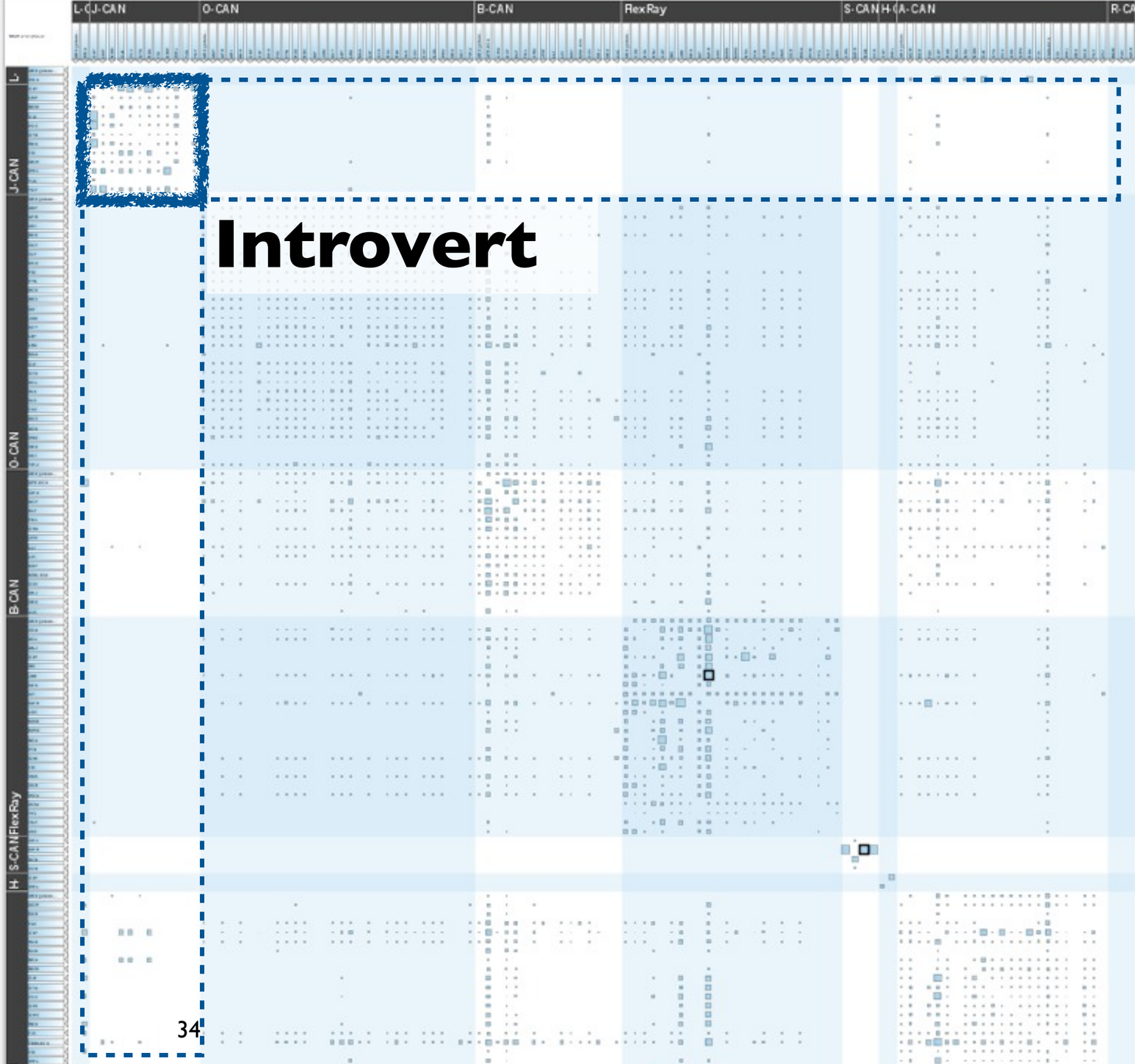
Bus communication patterns



INTERESTS

Bus communication patterns

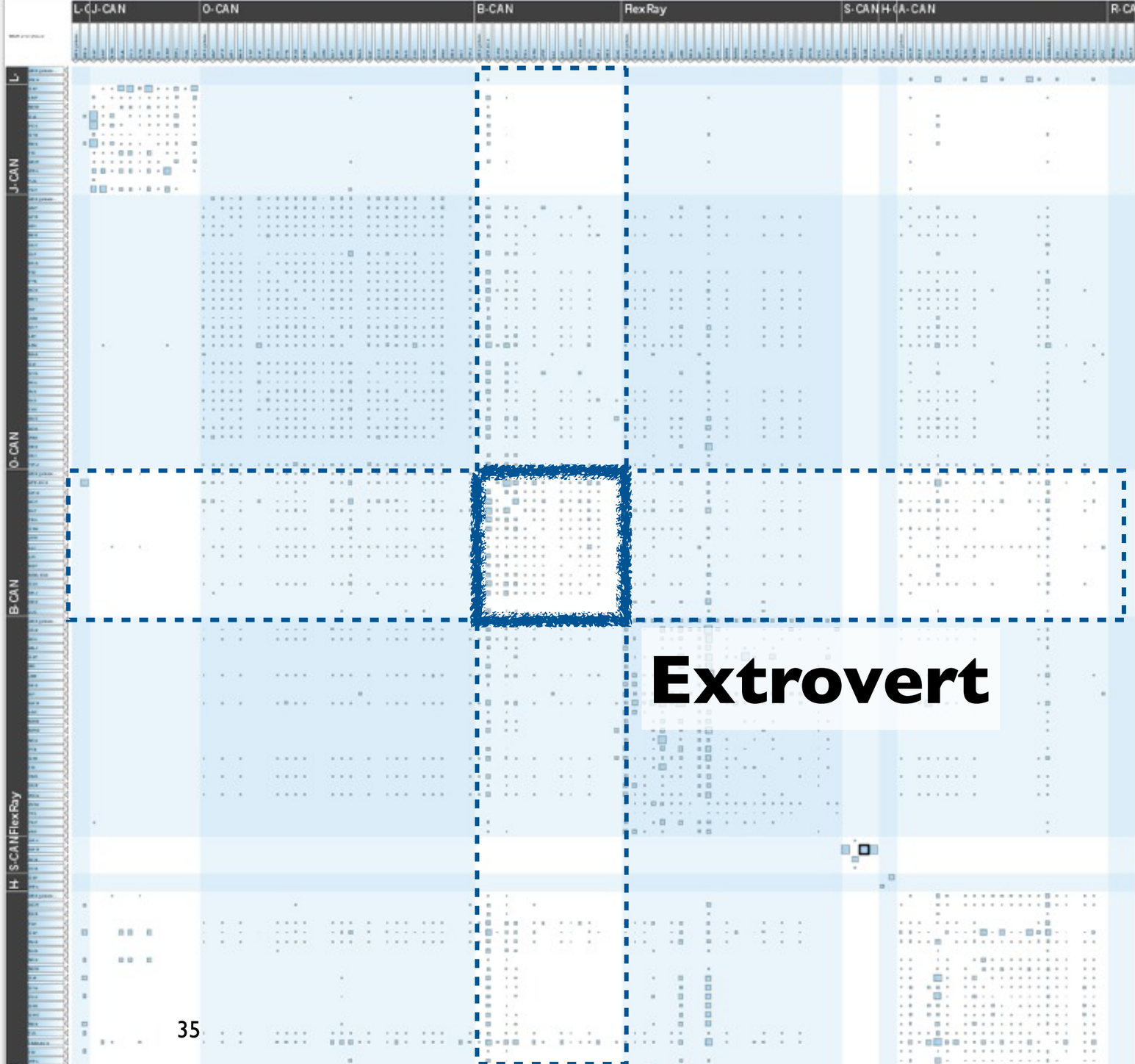
introvert
vs.
extrovert



INTERESTS

Bus communication patterns

introvert
vs.
extrovert



Methods

Phase I: Discover

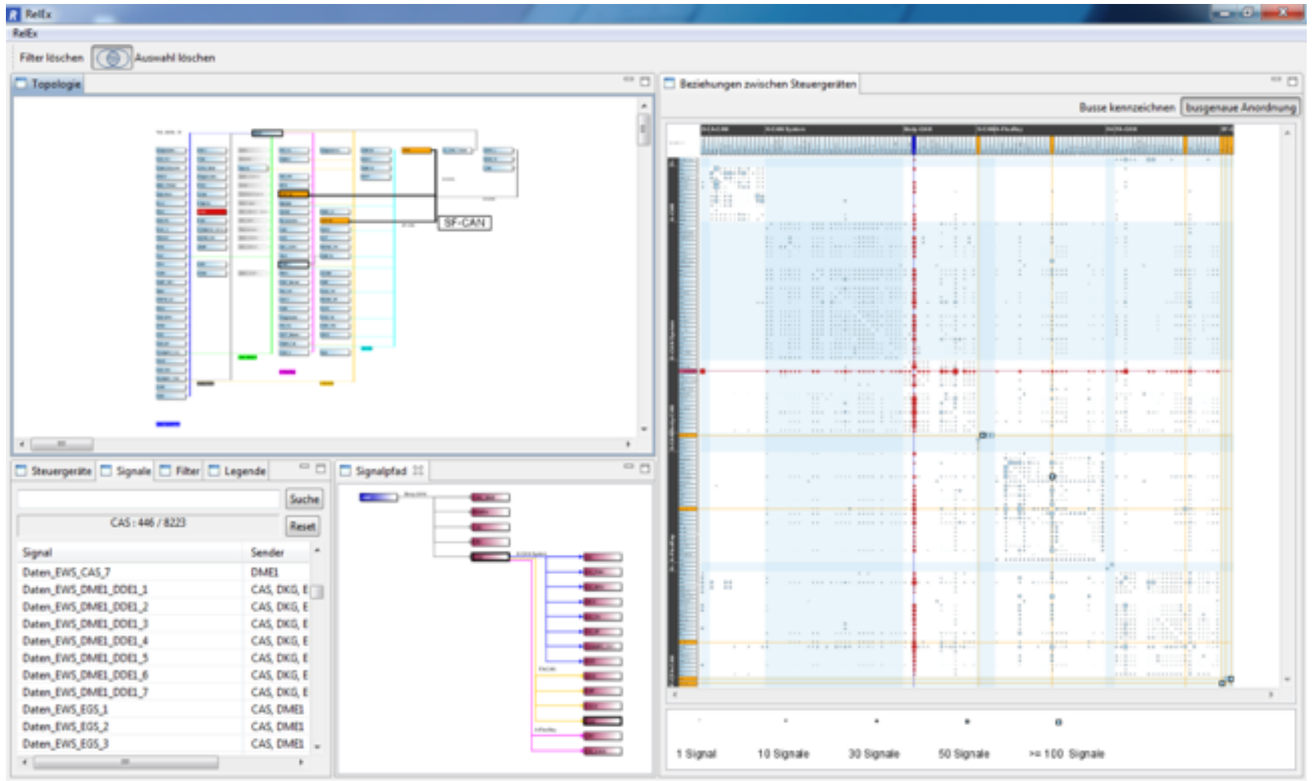
3 months

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

Phase 2: Design, implement, deploy

4 months

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



Phase 3: Summative evaluation

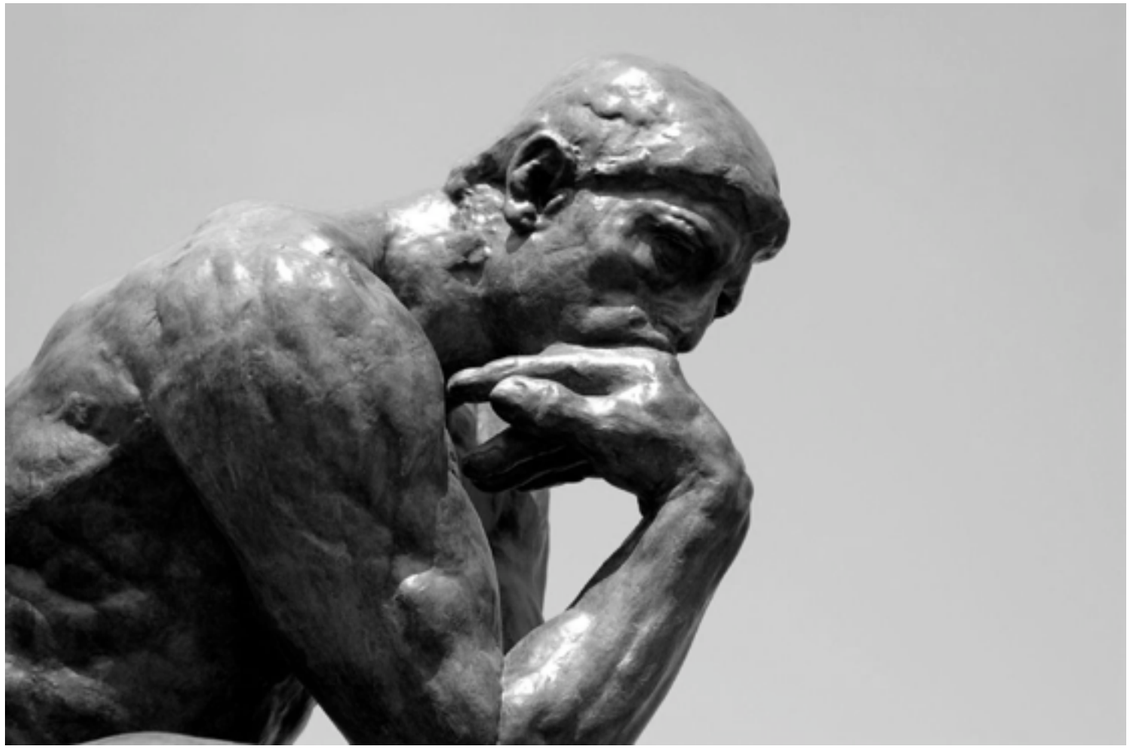
2 months

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

Phase 4: Reflect and write

3 months

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

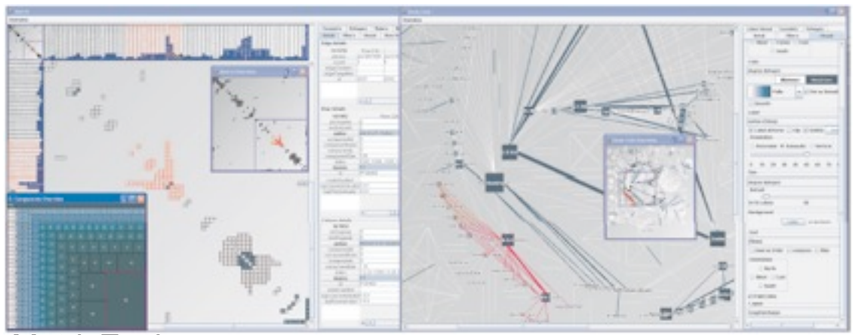


Abstraction Innovation

Previous Work

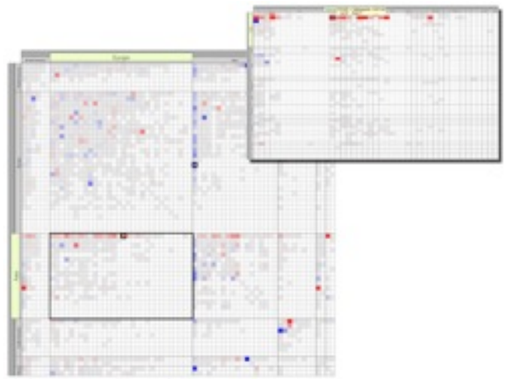
Focus on social network analysis

- radically different task and data abstractions



MatrixExplorer

SocialAction



Honeycomb

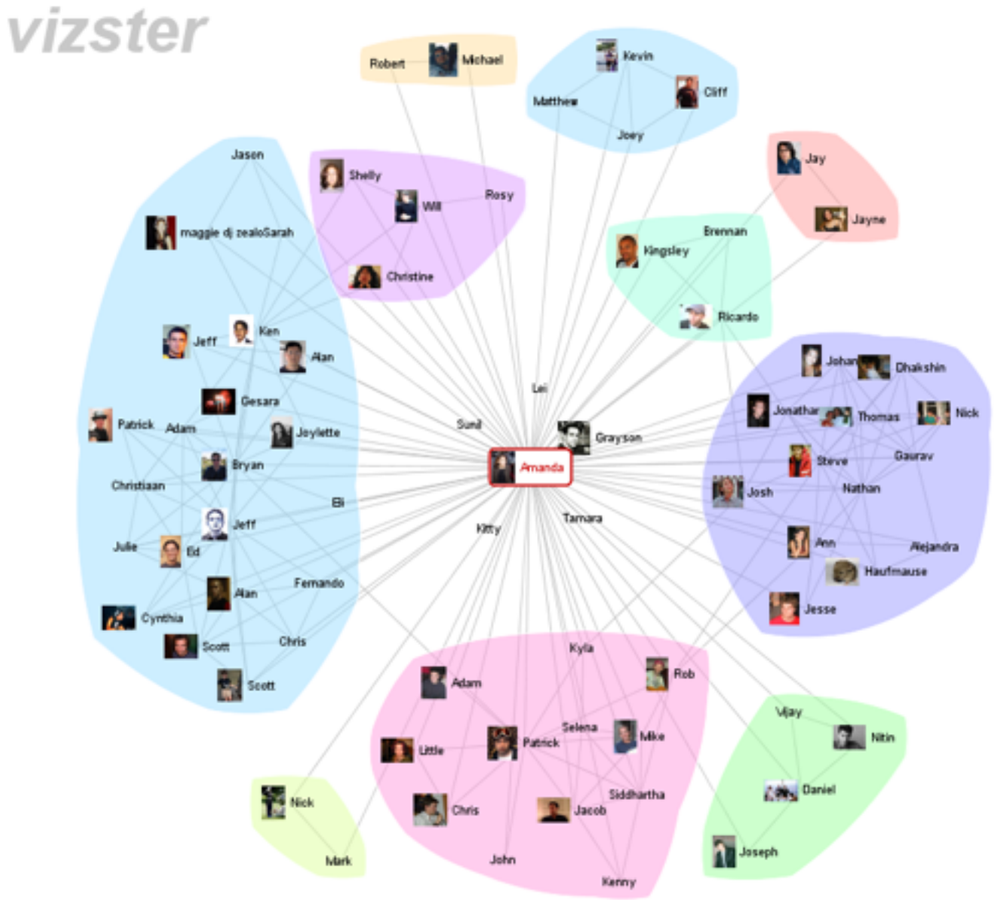
vizster

Image: http://www.orbitalrpm.com/wp-content/uploads/2007/10/sna-graphic-7.jpg

Task Abstraction

Social Network Analysis Domain

- find clusters

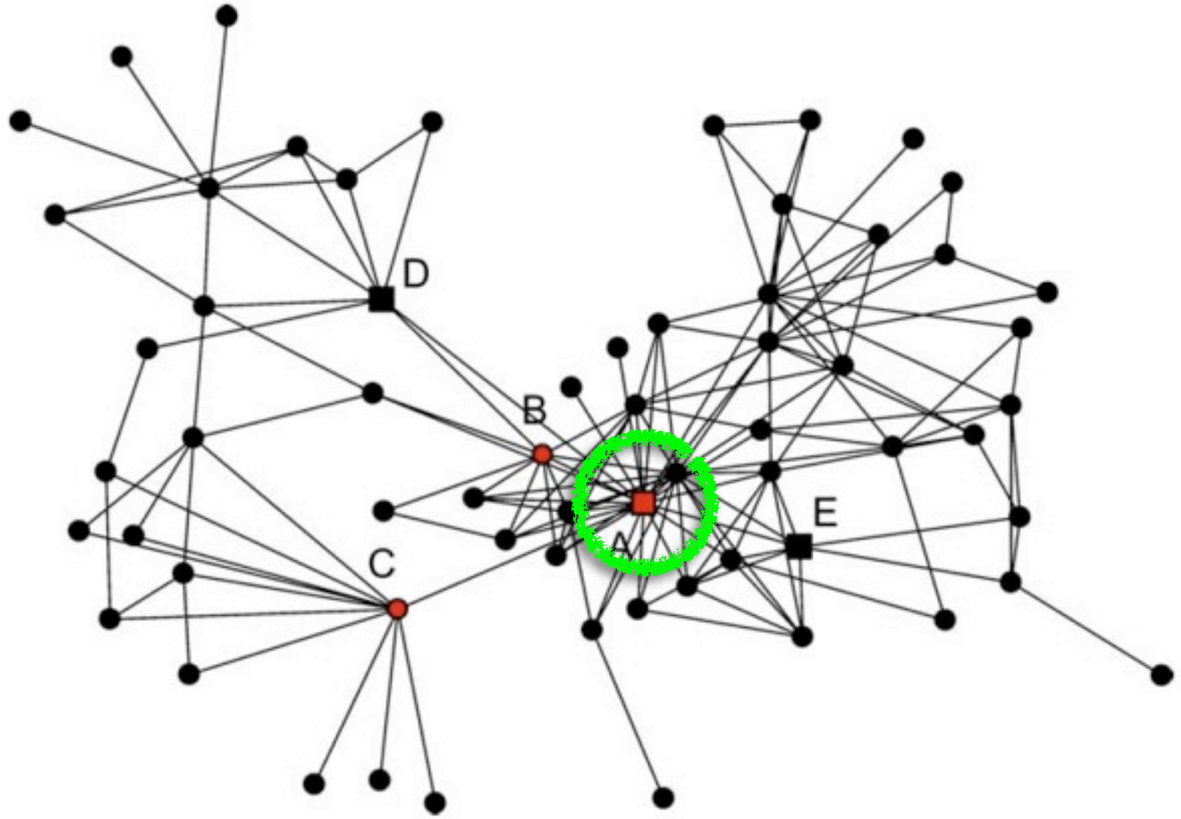


Task Abstraction

Social Network Analysis Domain

- find clusters

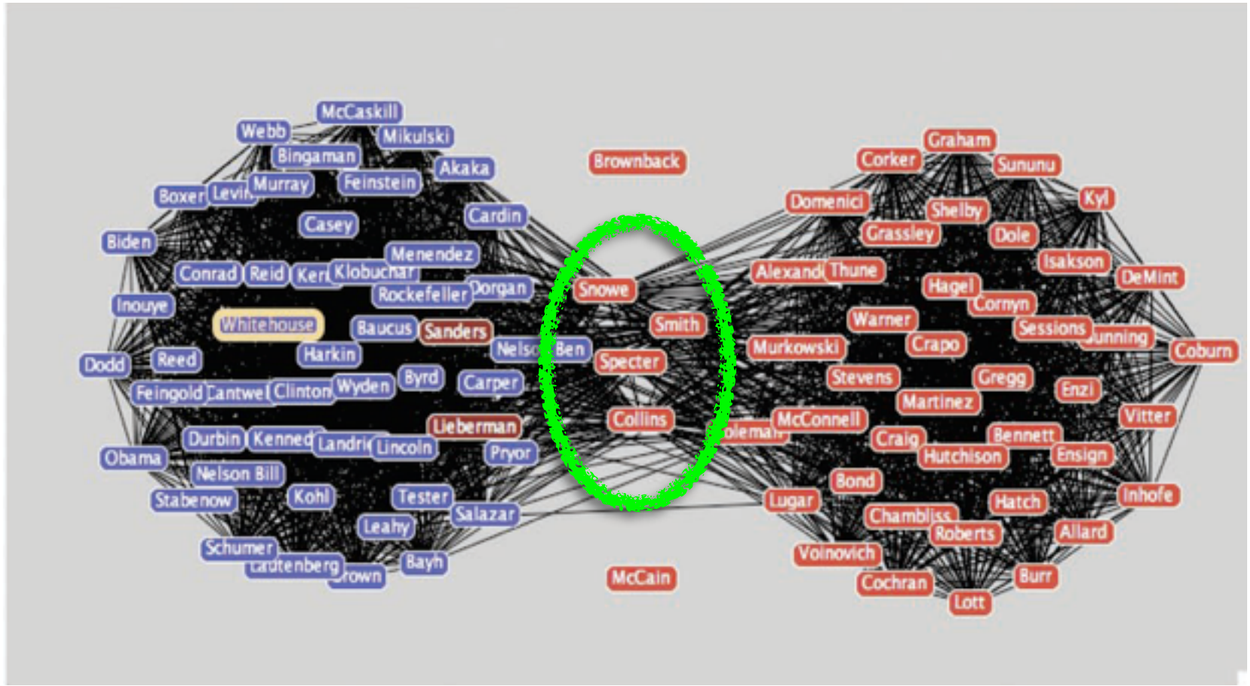
- find high-degree nodes



Task Abstraction

Social Network Analysis Domain

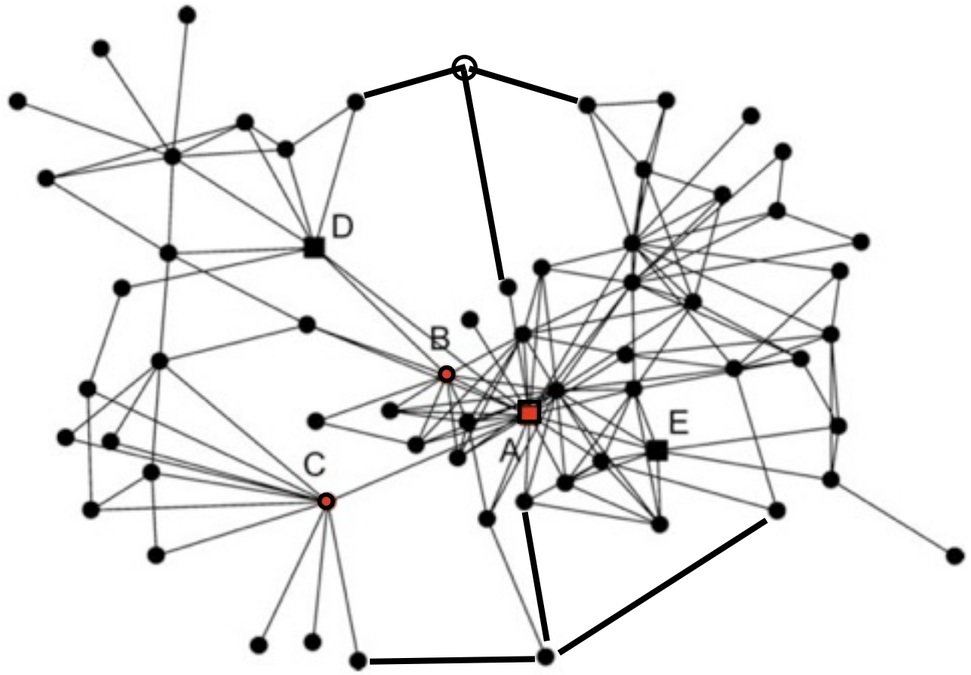
- find clusters
- find high-degree nodes
- find bridge nodes



Task Abstraction

Social Network Analysis Domain

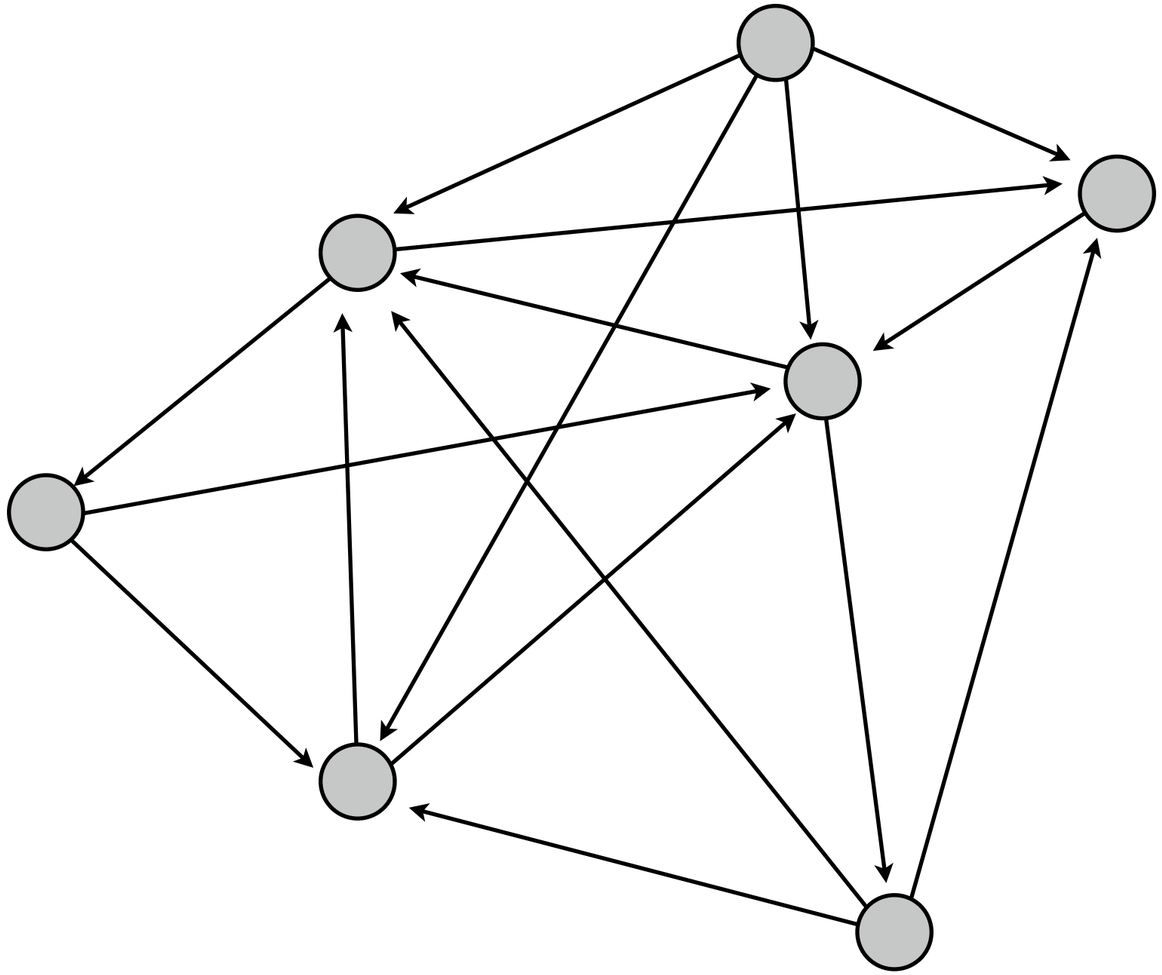
- find clusters
- find high-degree nodes
- find bridge nodes
- understand temporal dynamics
  - passively notice changes



Data Abstraction

Social Network Analysis Domain

- single graph

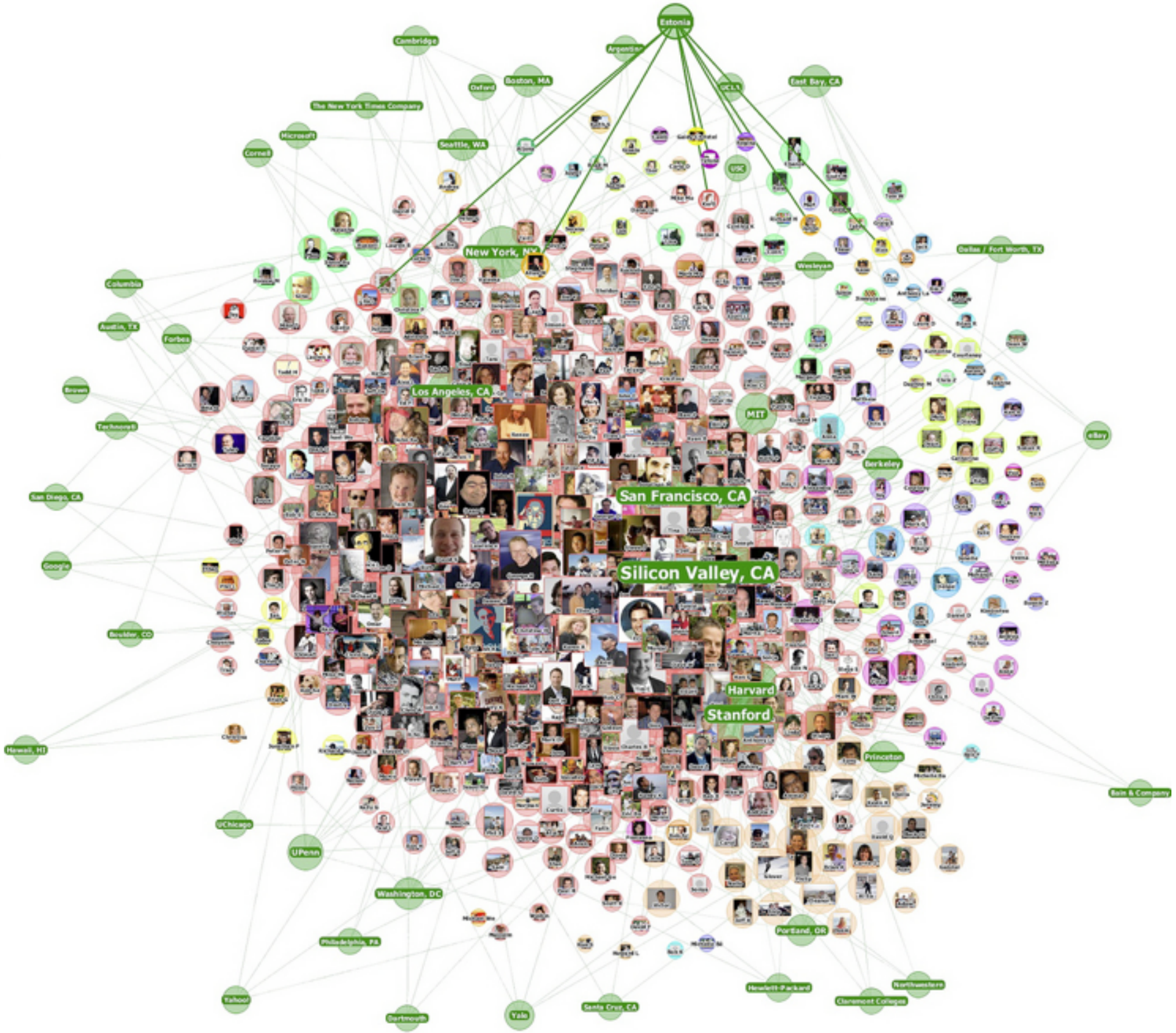


Data Abstraction

Social Network Analysis

- single graph

- scalability challenge: nodes

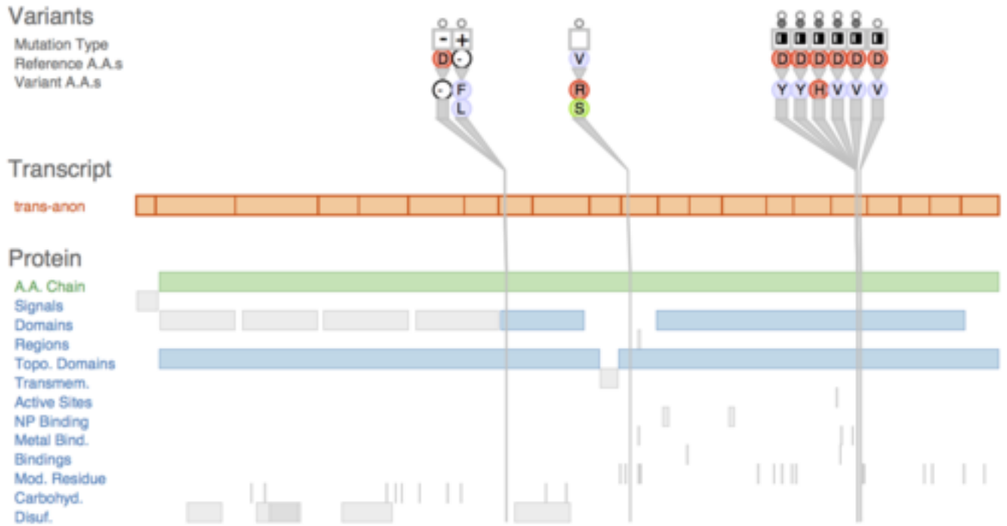


Abstraction Differences

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:

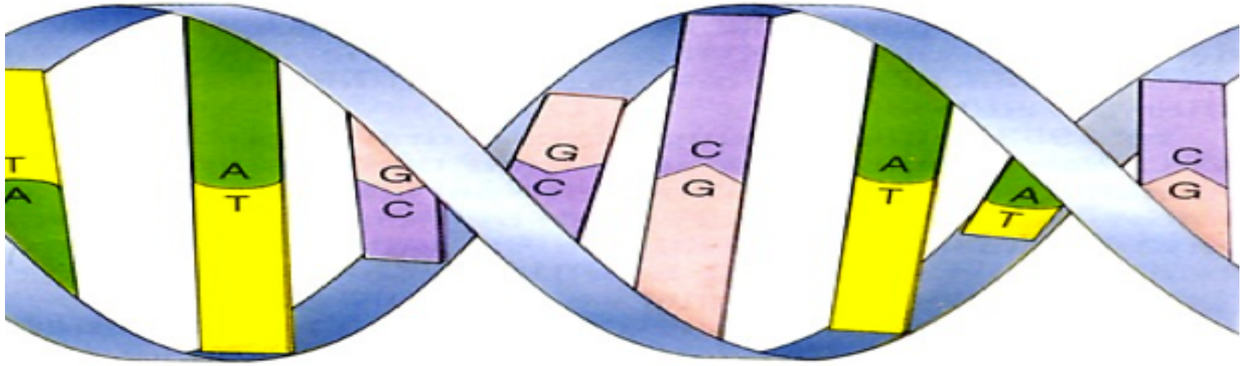
Joel Ferstay, Cydney Nielsen

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

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

Sample 2 Genome DNA: ATA TGA TGA ACA CCT

Harmful?

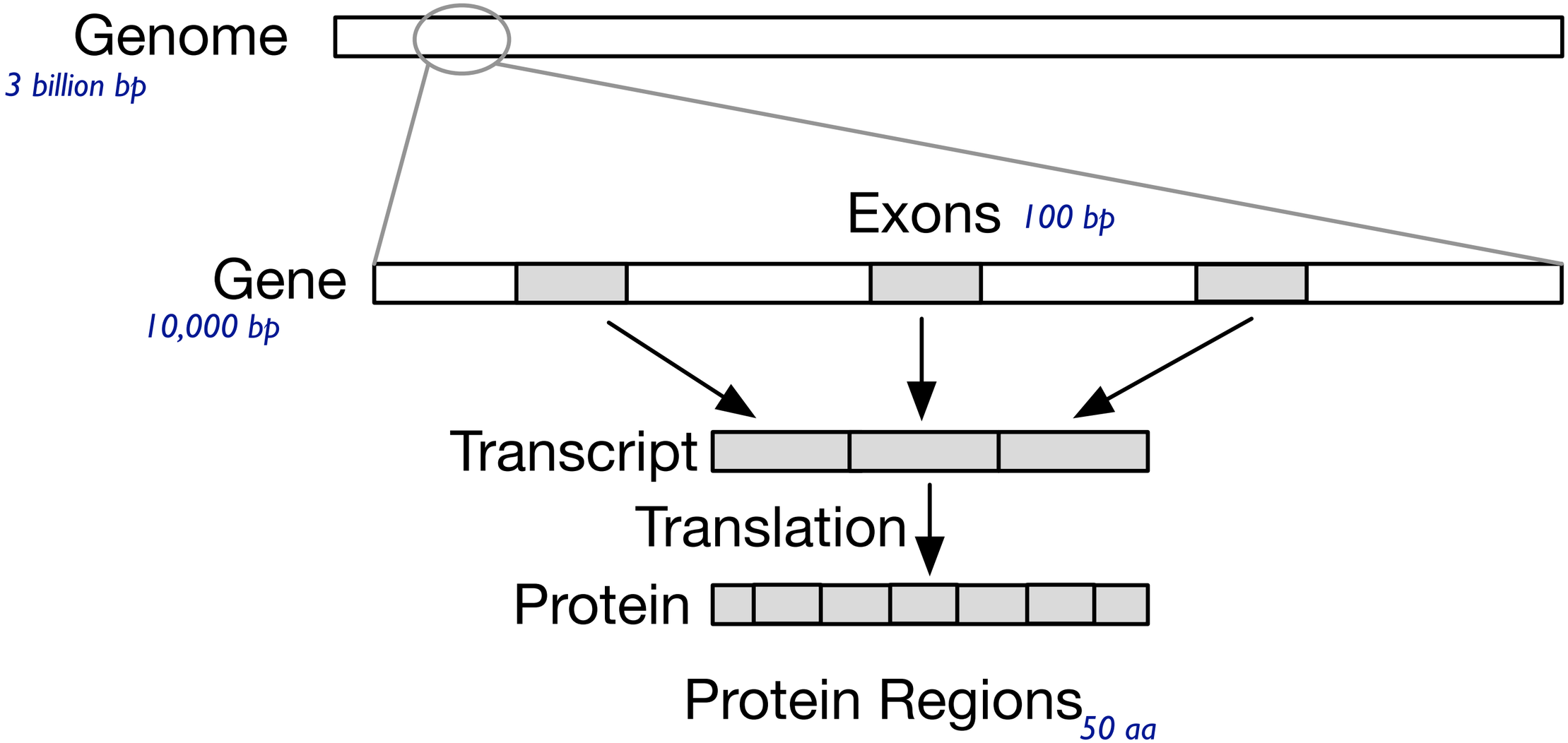
Harmless?

Cancer Research

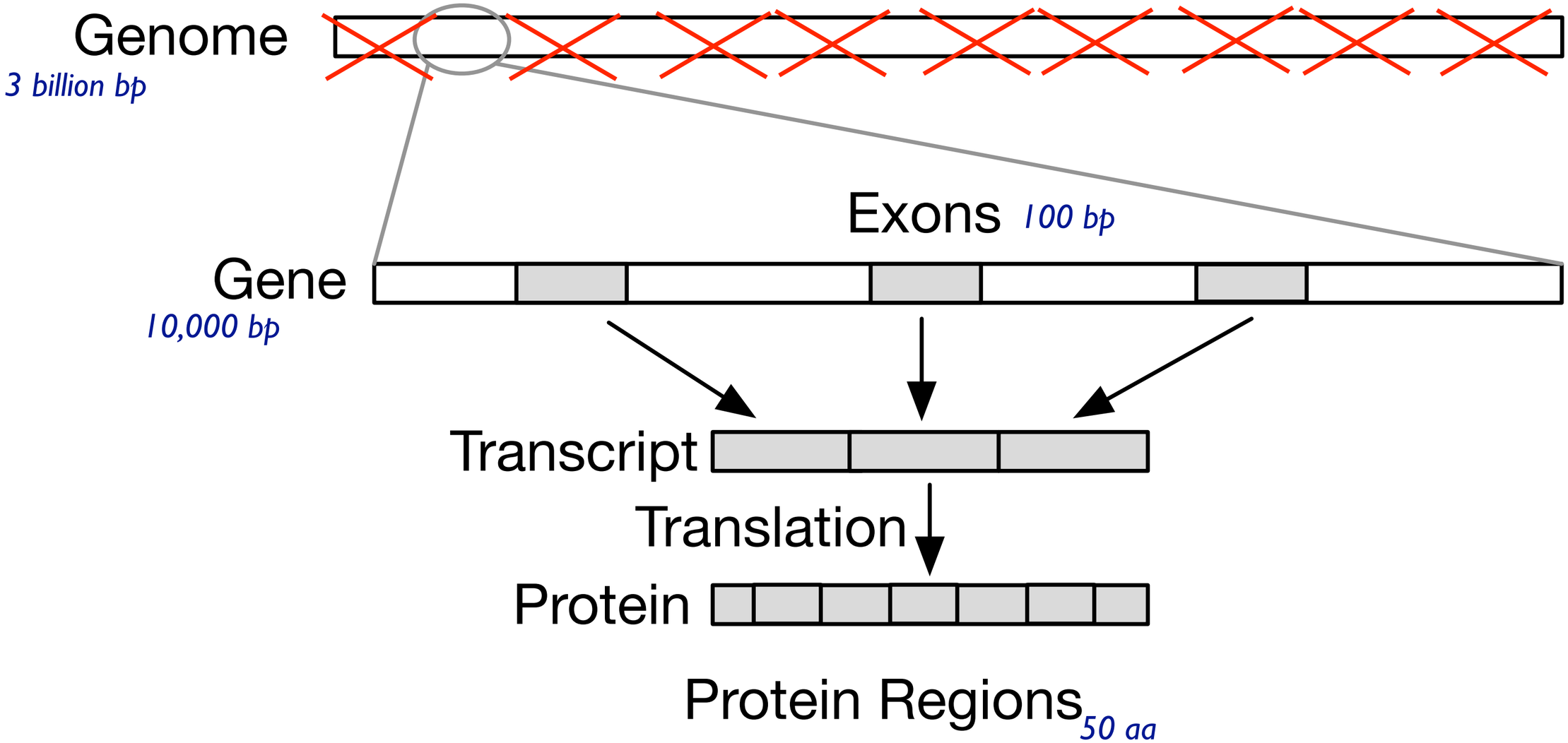
- collaboration with analysts at BC Genome Sciences Center
  - studying genetic basis of leukemia
- driving task
  - discover new candidate genes with harmful variants
- two big questions
  - what to show
    - data abstraction
    - challenge: enormous range of scales in the data
  - how to show it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

Abstractions

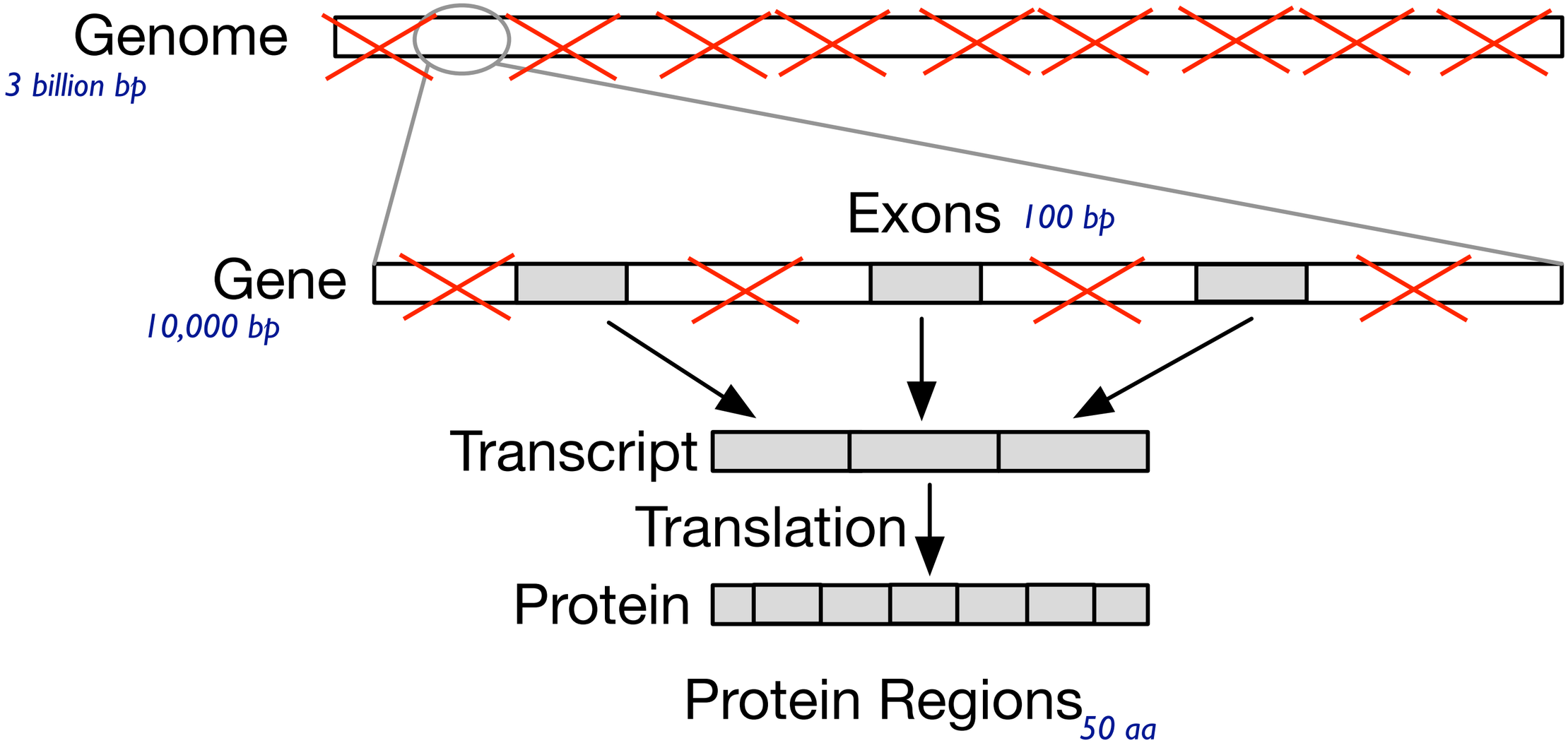
Data: Filtering to relevant biological levels and scales



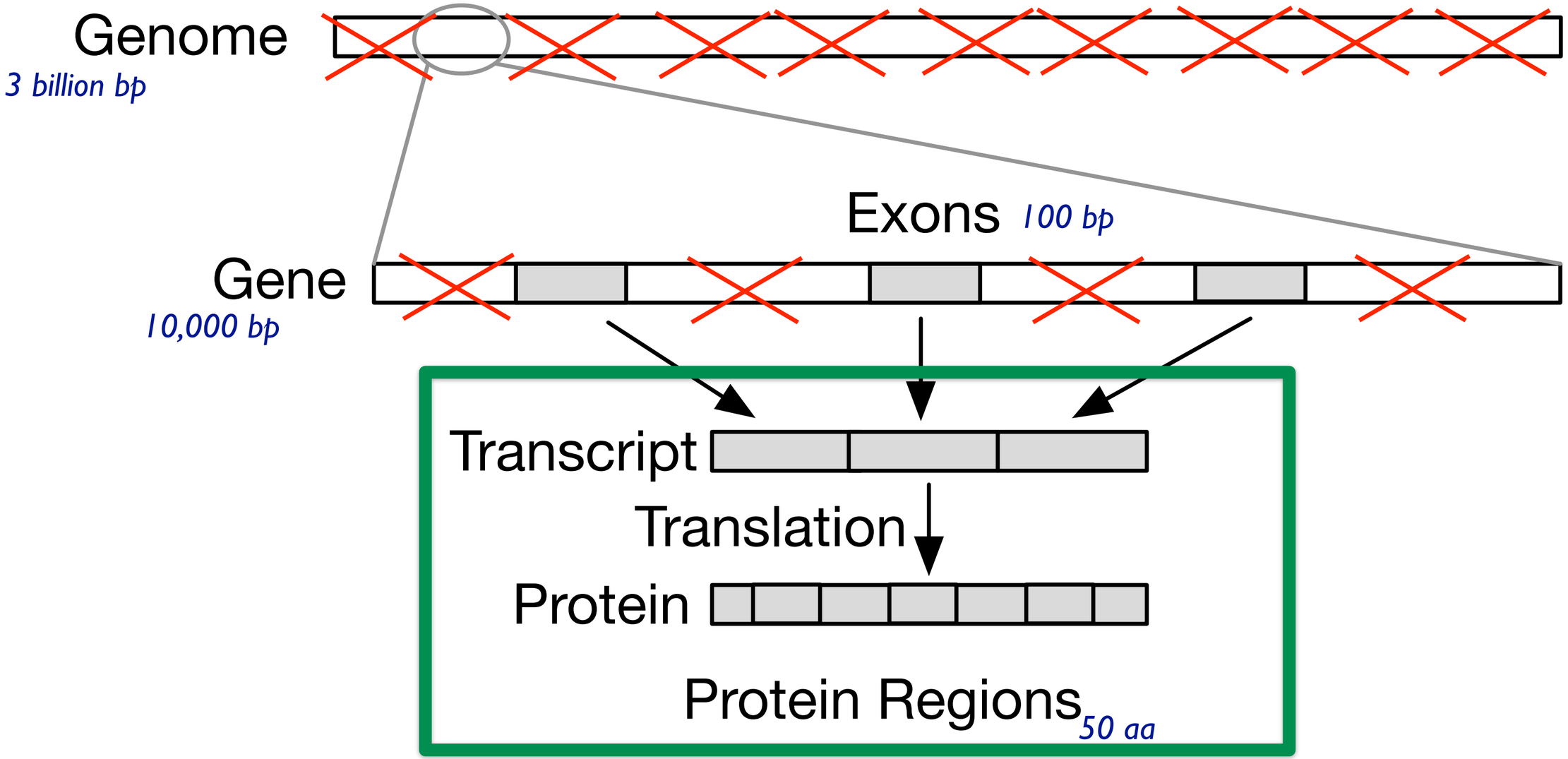
Filter out whole genome; keep genes



Filter out non-exon regions

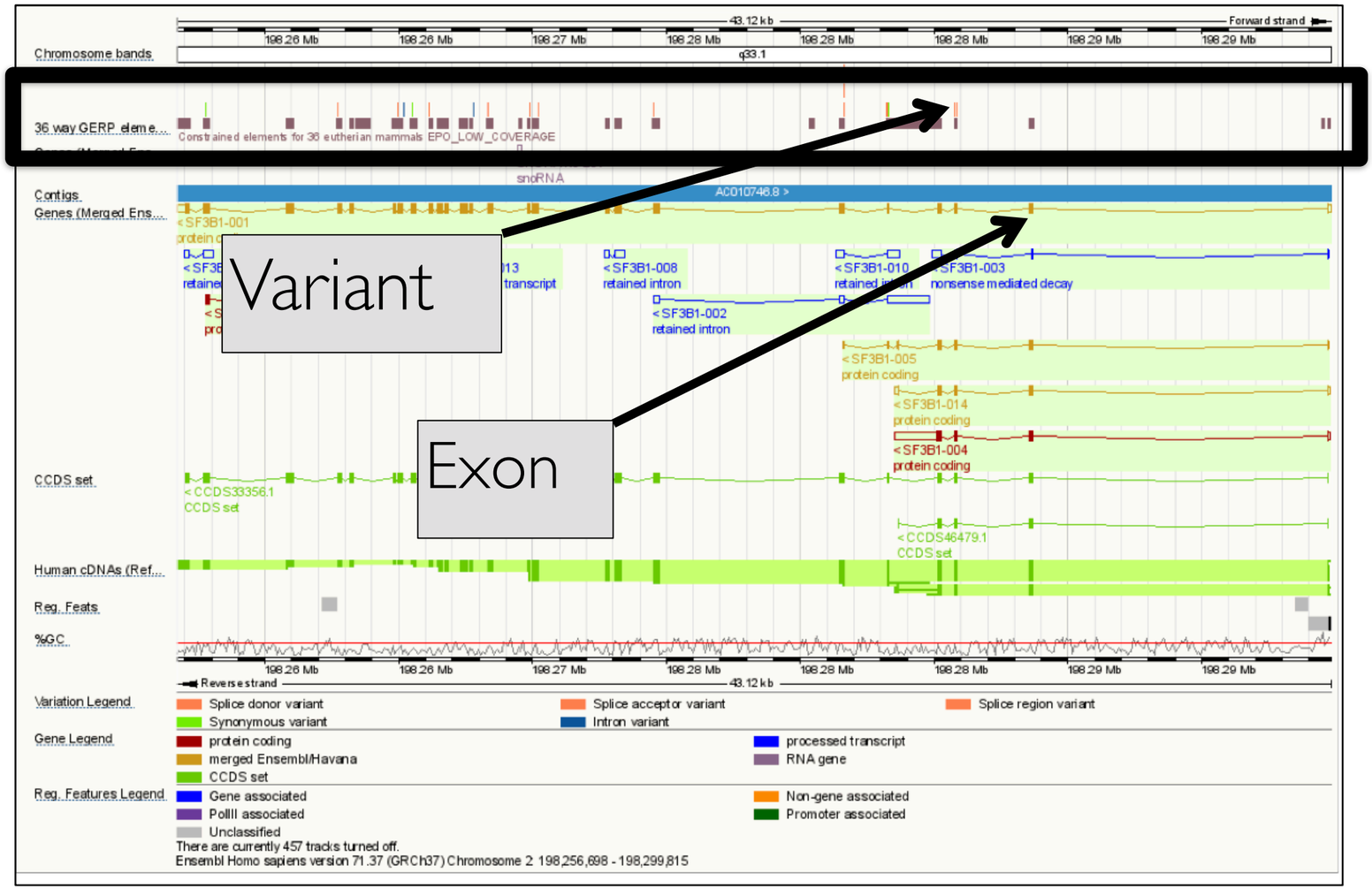


Data abstraction: highly filtered scope of *transcript coordinates*



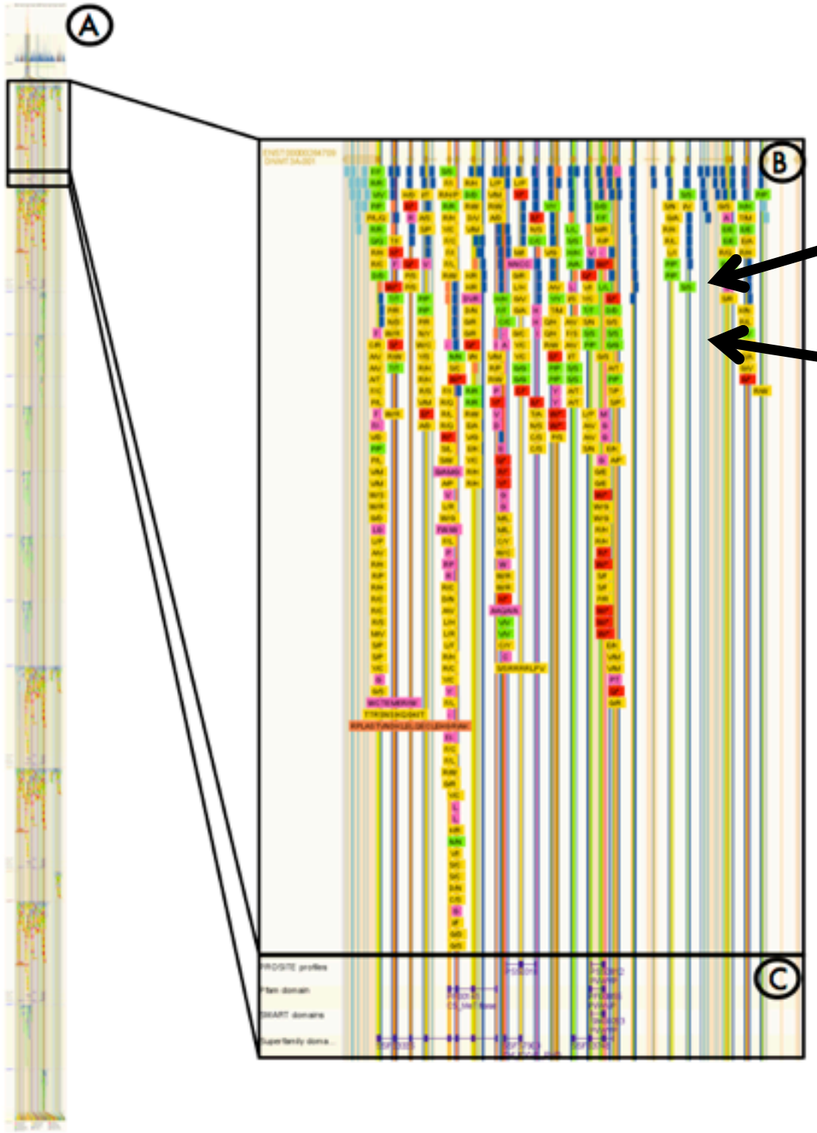
Dominant paradigm: genome browsers

- 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



1st Screen

Exon regions small

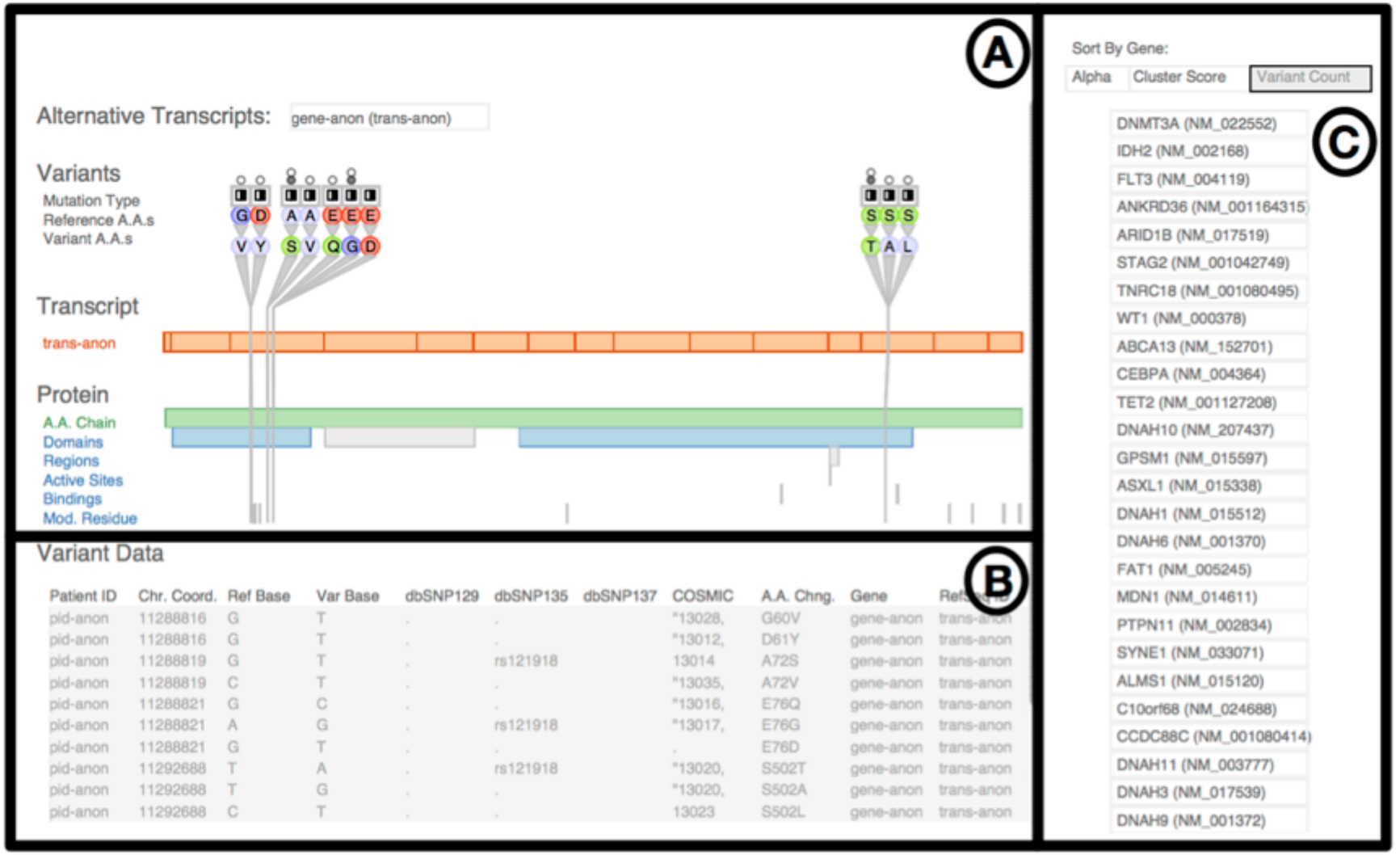
Color coding difficult to see

Ensembl Variant Image

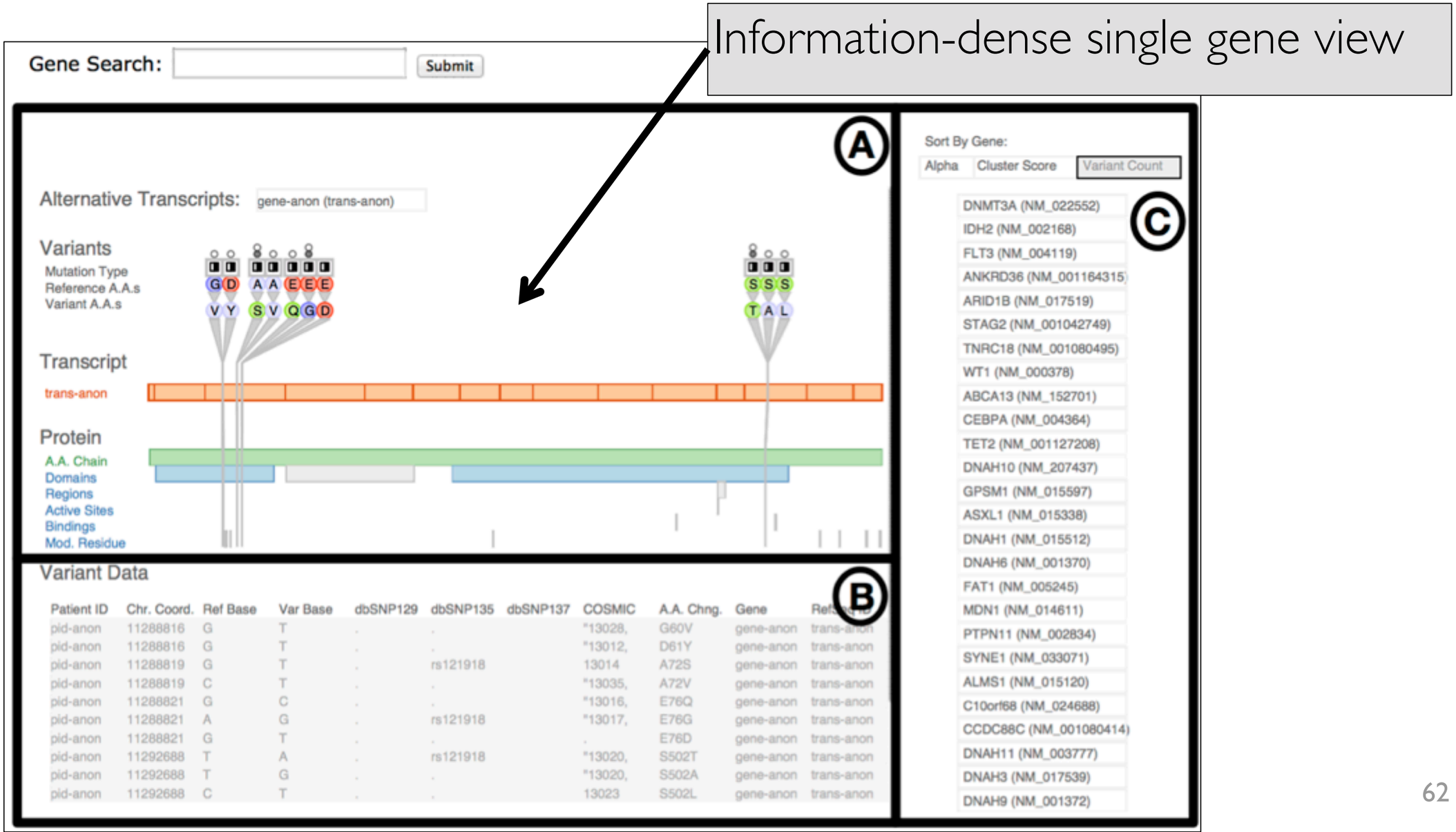
Chen et al, BMC Bioinforrmatics 2010.

Idioms

Variant View



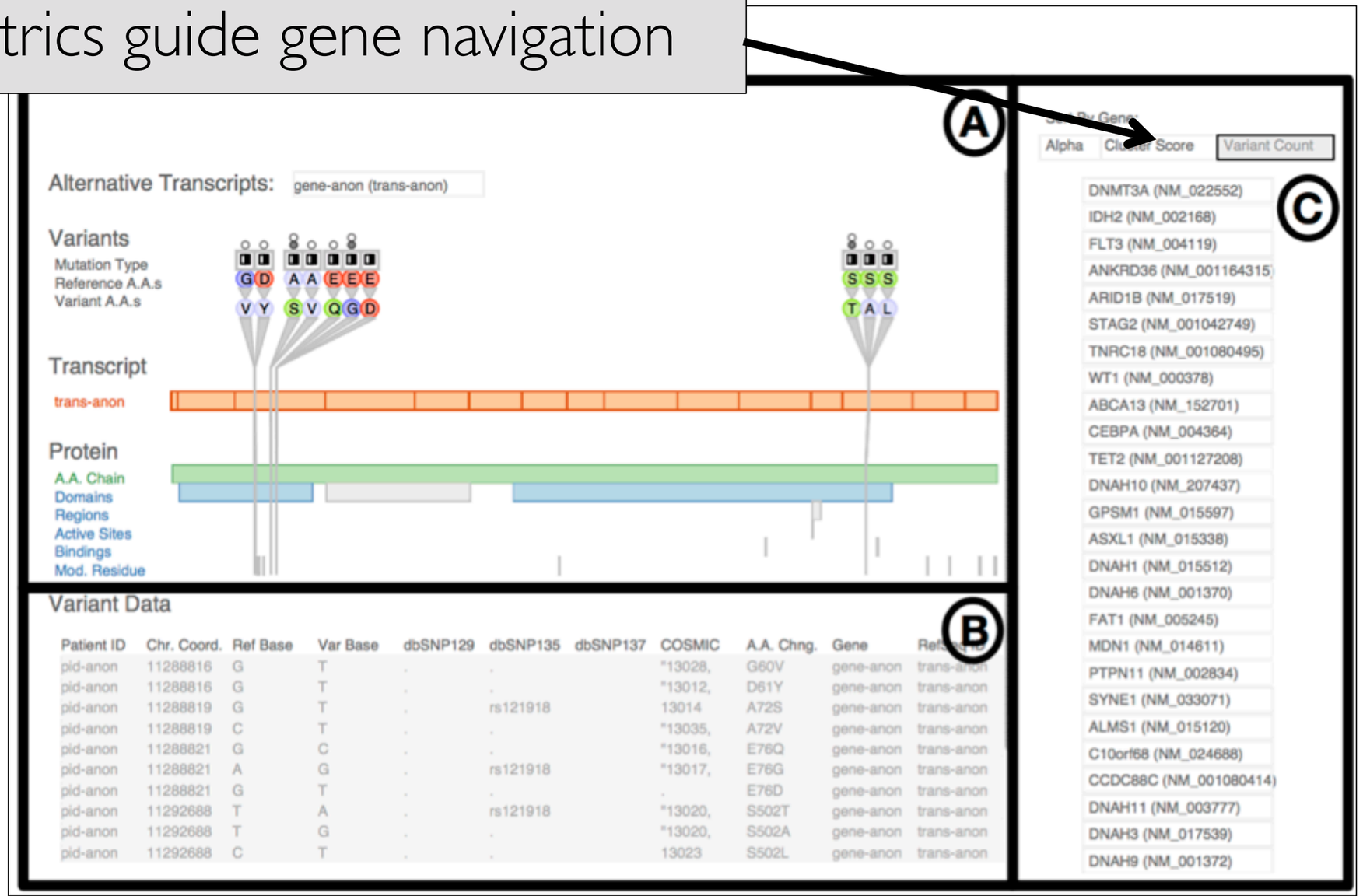
Variant View



Variant View

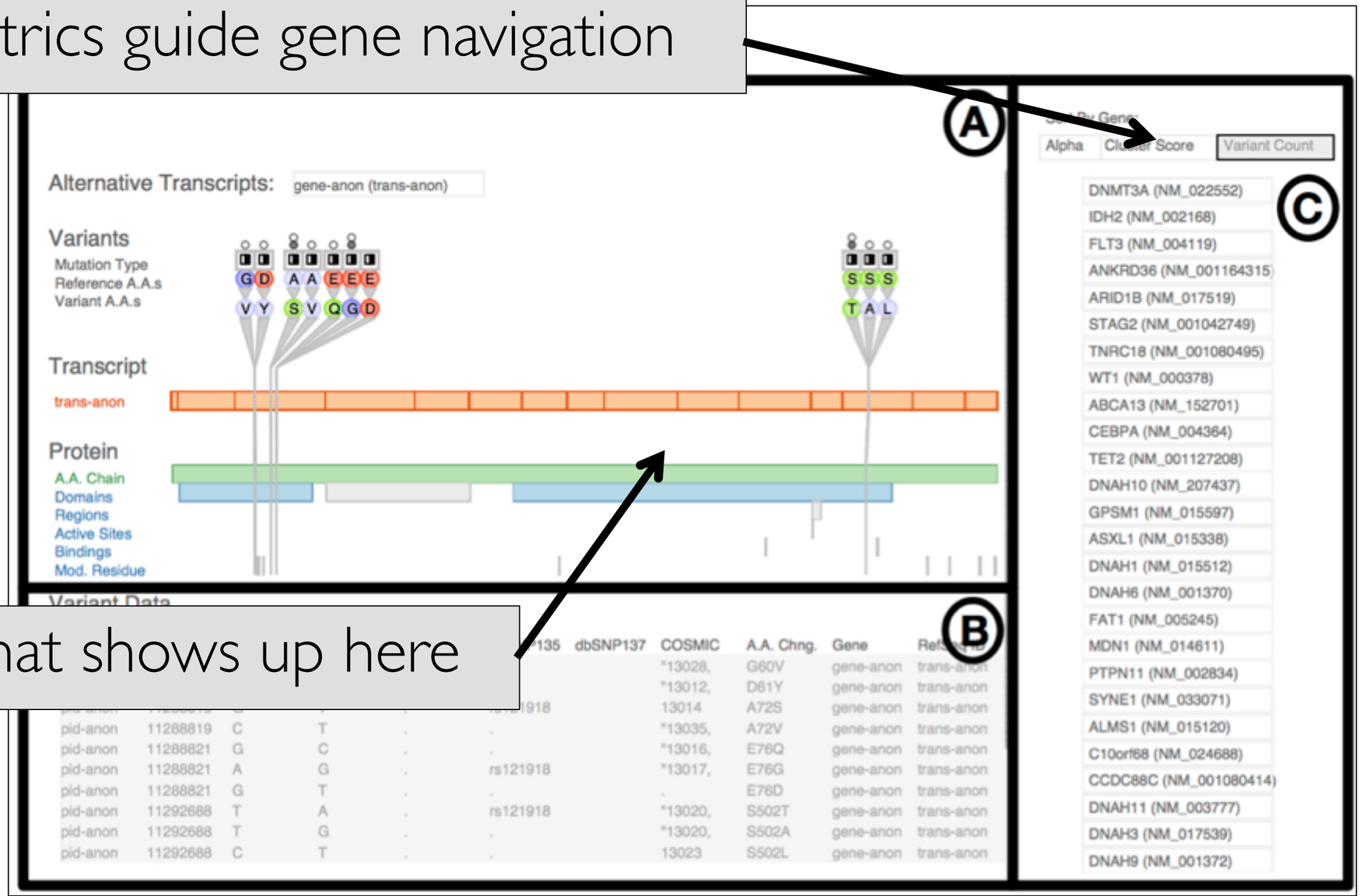
Variant View

Sorting metrics guide gene navigation



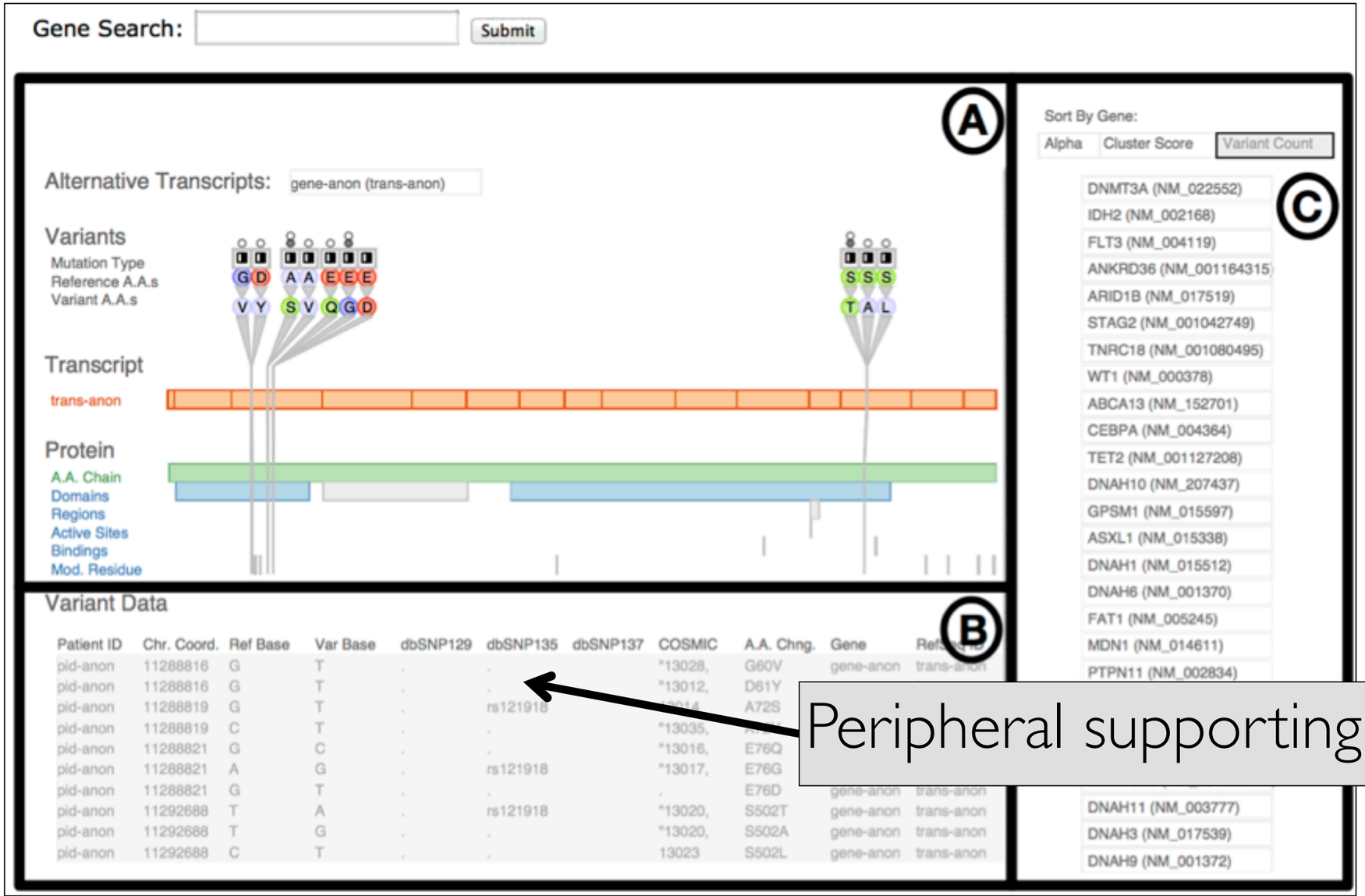
Variant View

Sorting metrics guide gene navigation



Control what shows up here

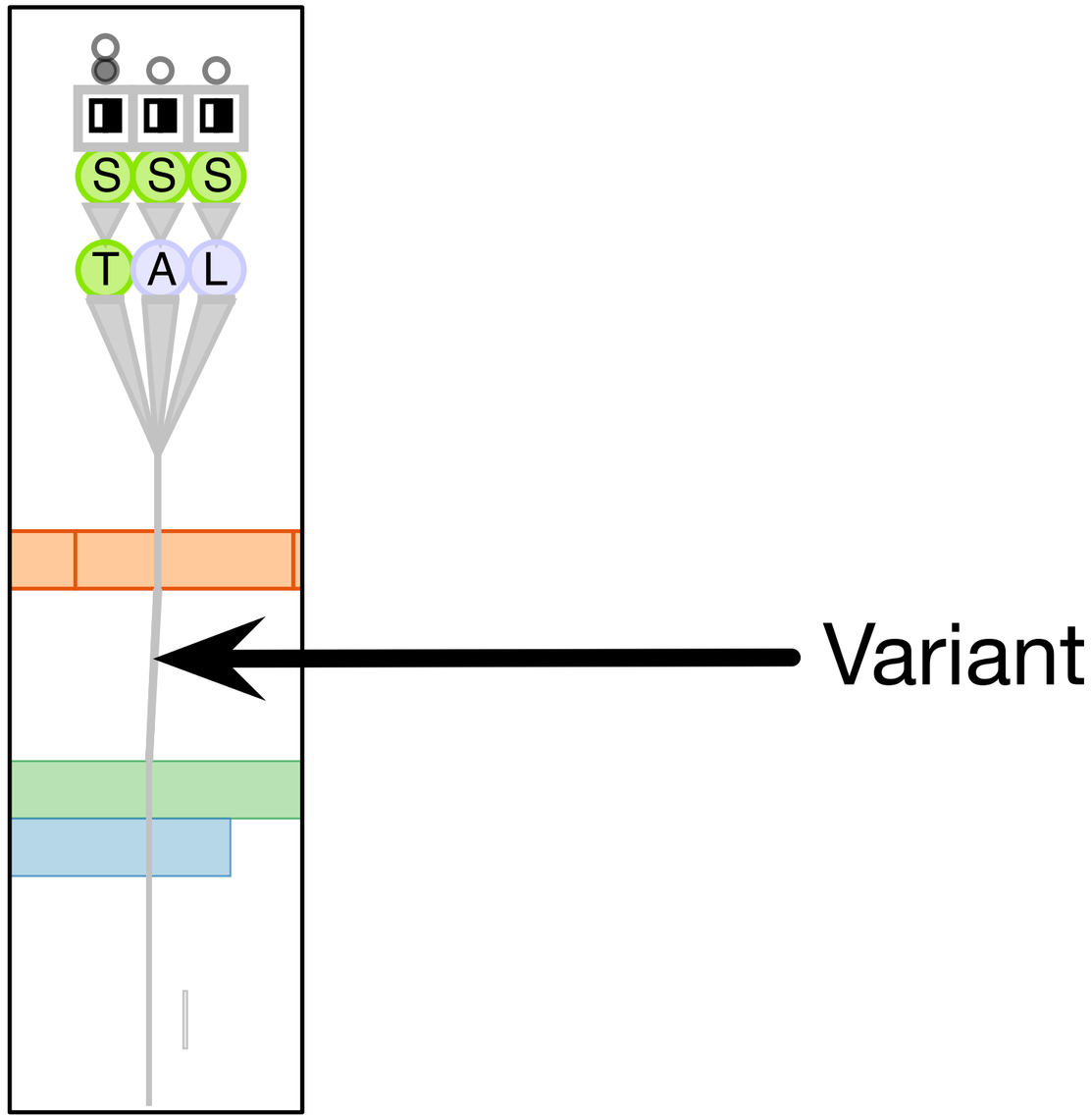
Variant View



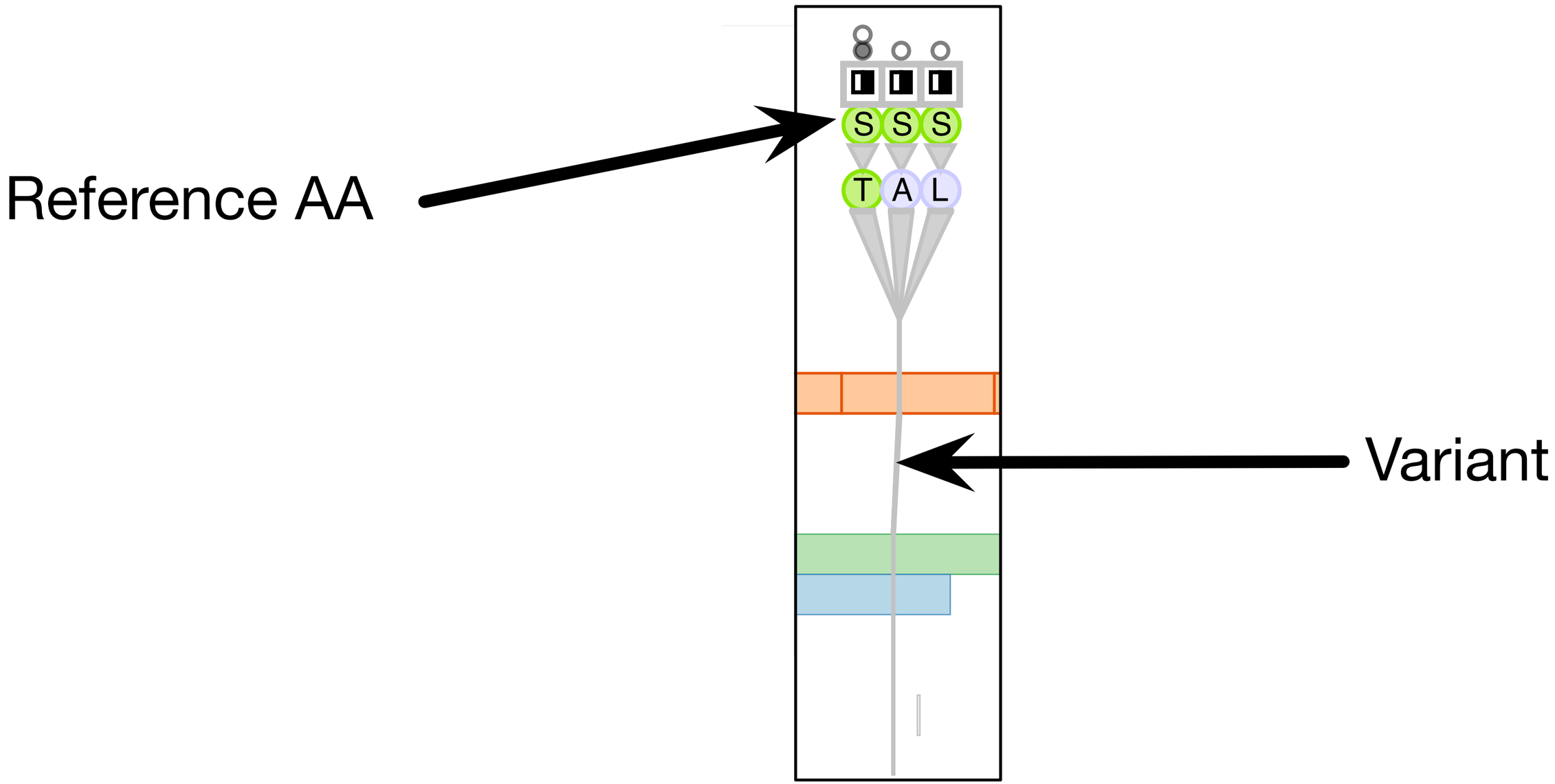
Peripheral supporting data

Design information-dense visual encoding

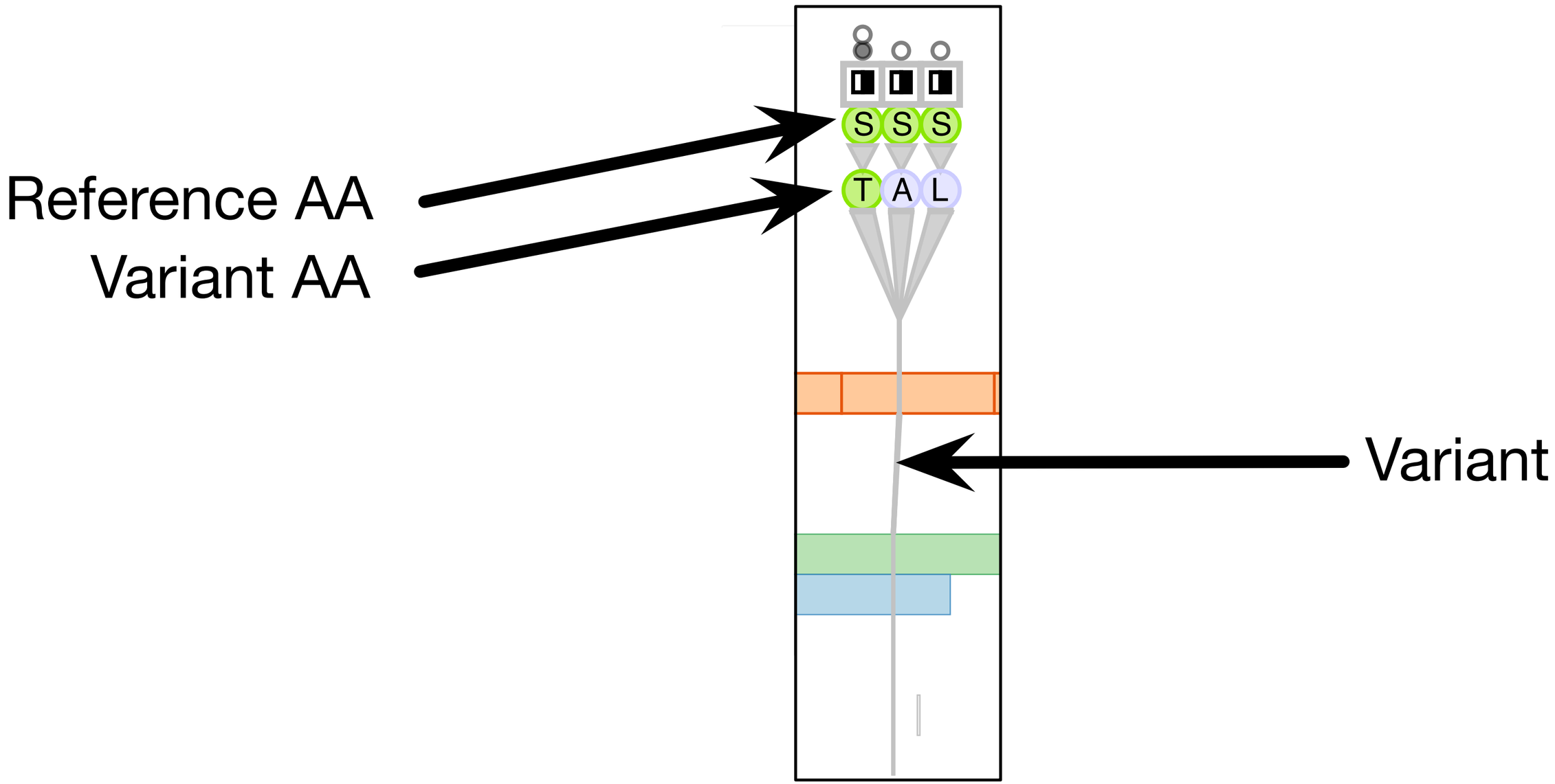
- show all attributes necessary for variant analysis
  - match salience with importance for analysis task
- variant not just a thin line!
- emphasize with high salience
  - collocated variants fan out at top
  - grey variant vertical stroke intersects horizontal colored protein regions



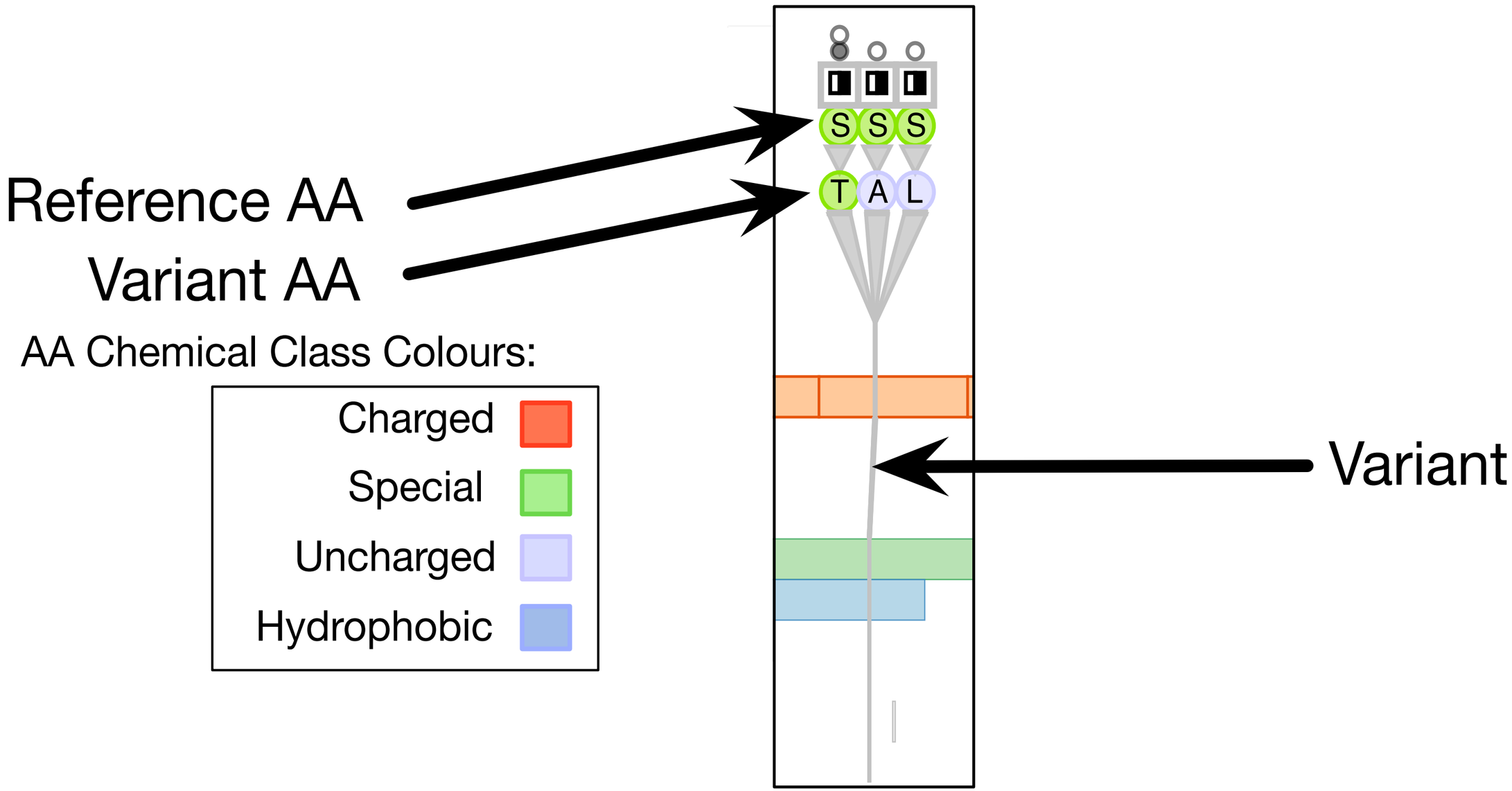
Design information-dense visual encoding



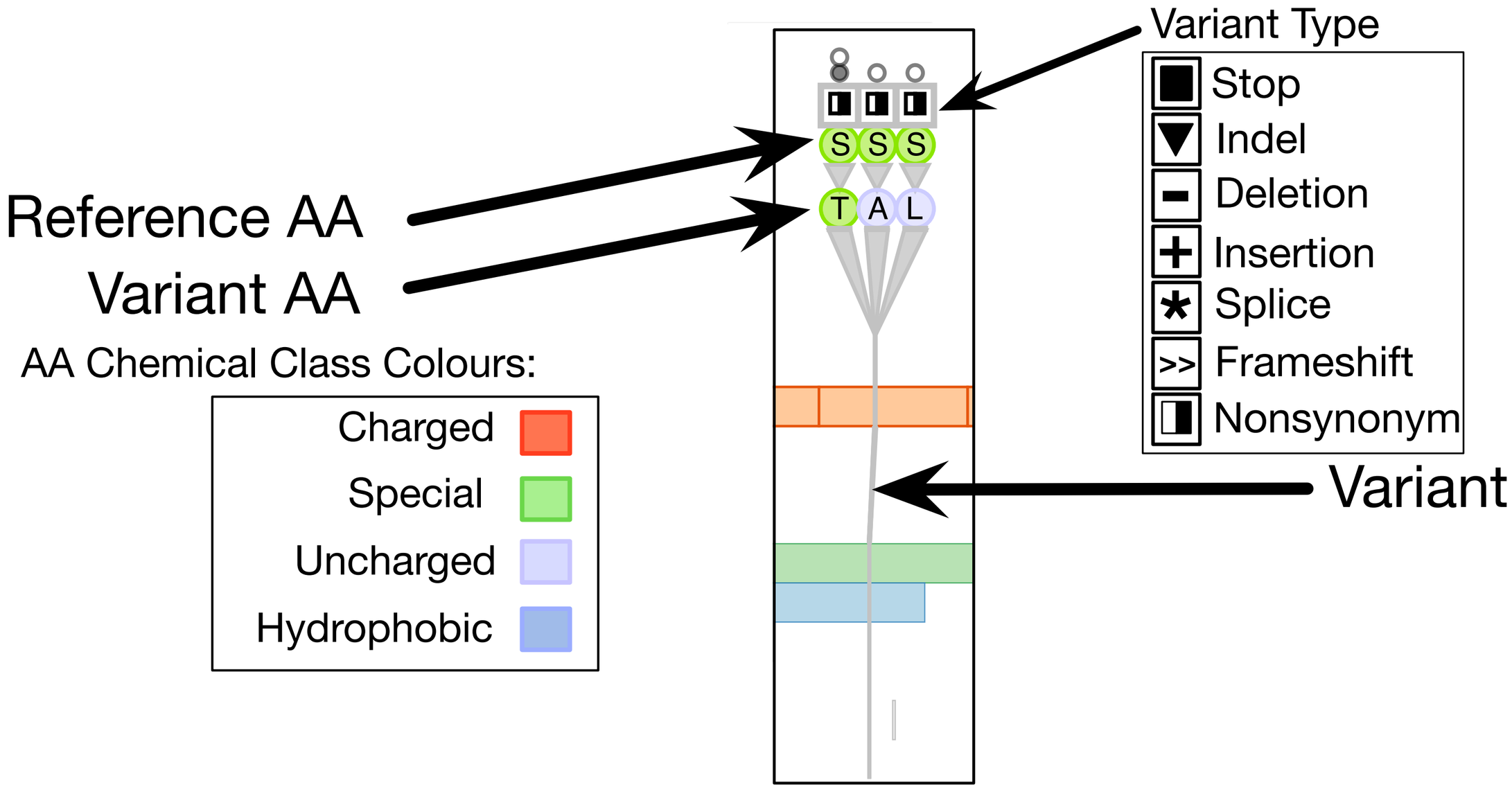
Design information-dense visual encoding



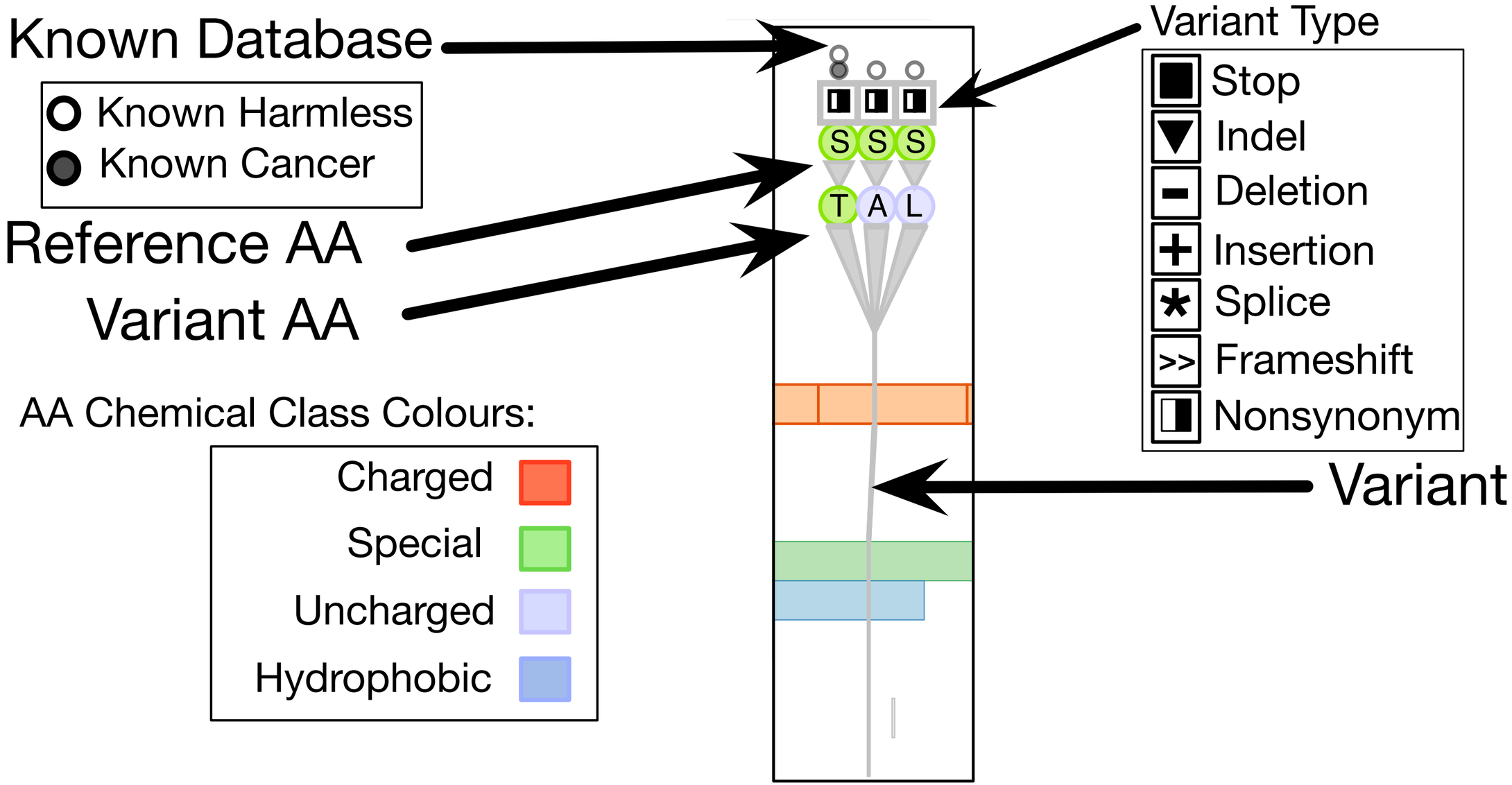
Design information-dense visual encoding



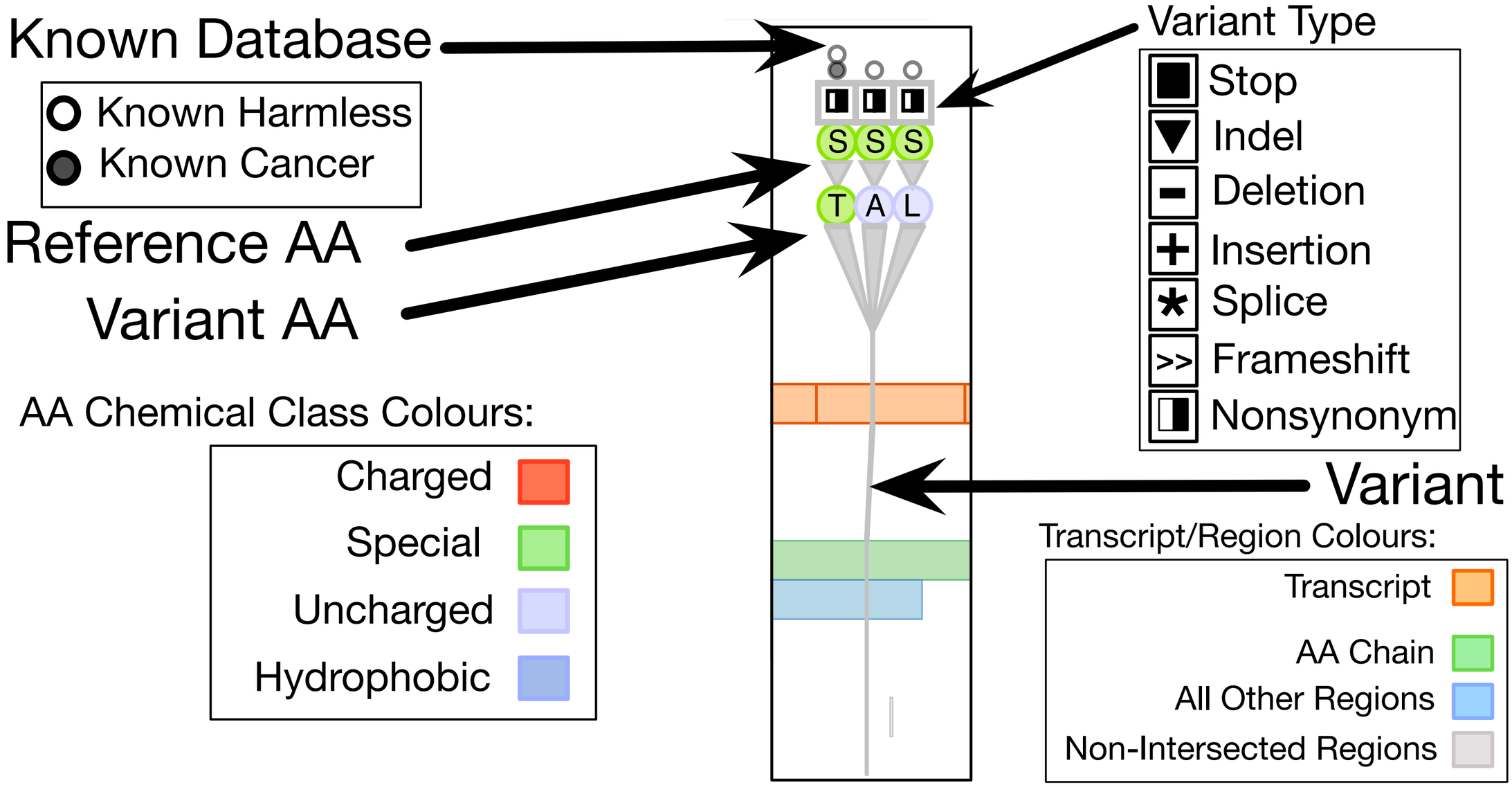
Design information-dense visual encoding



Design information-dense visual encoding

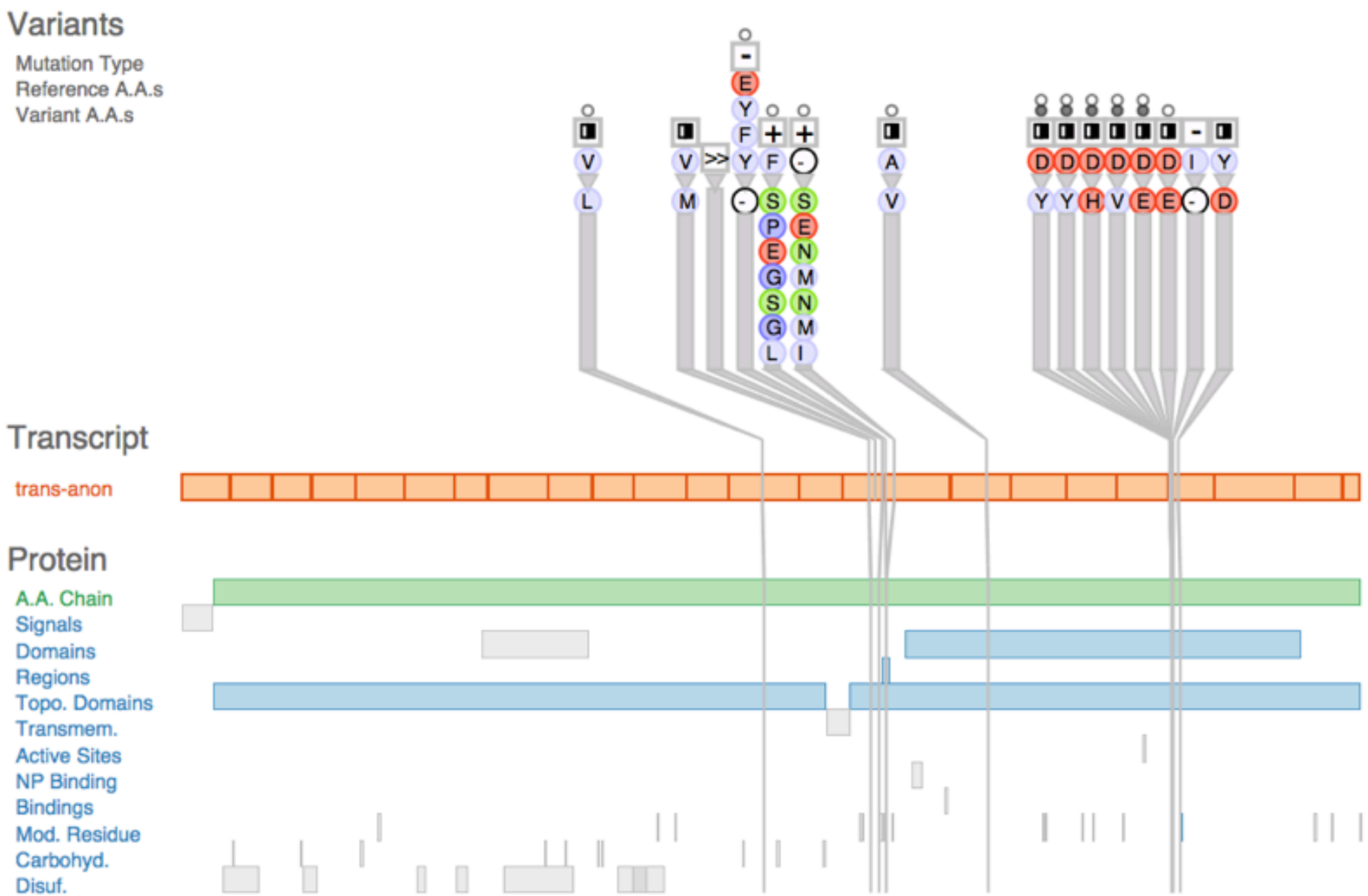


Design information-dense visual encoding

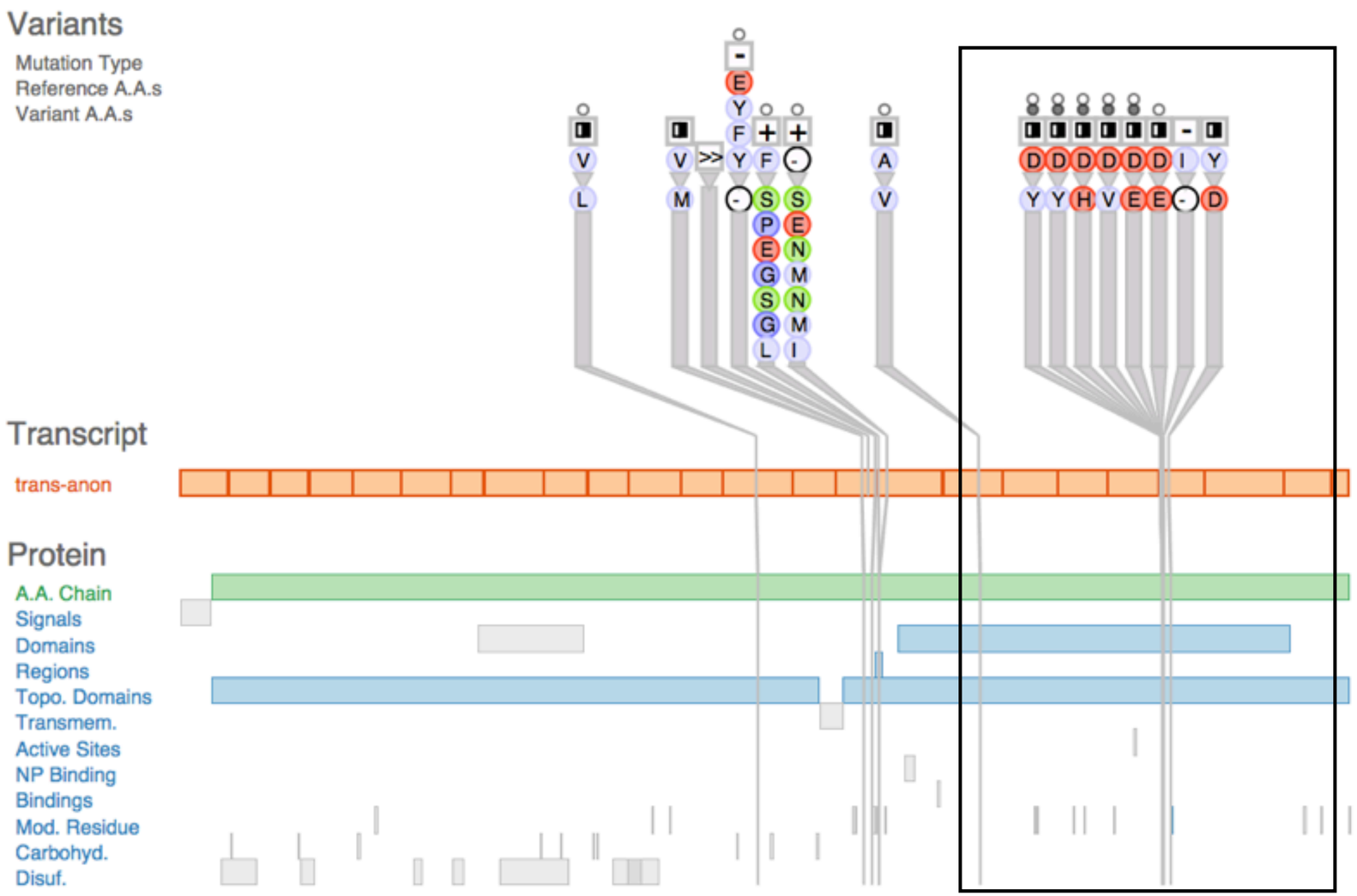


Results

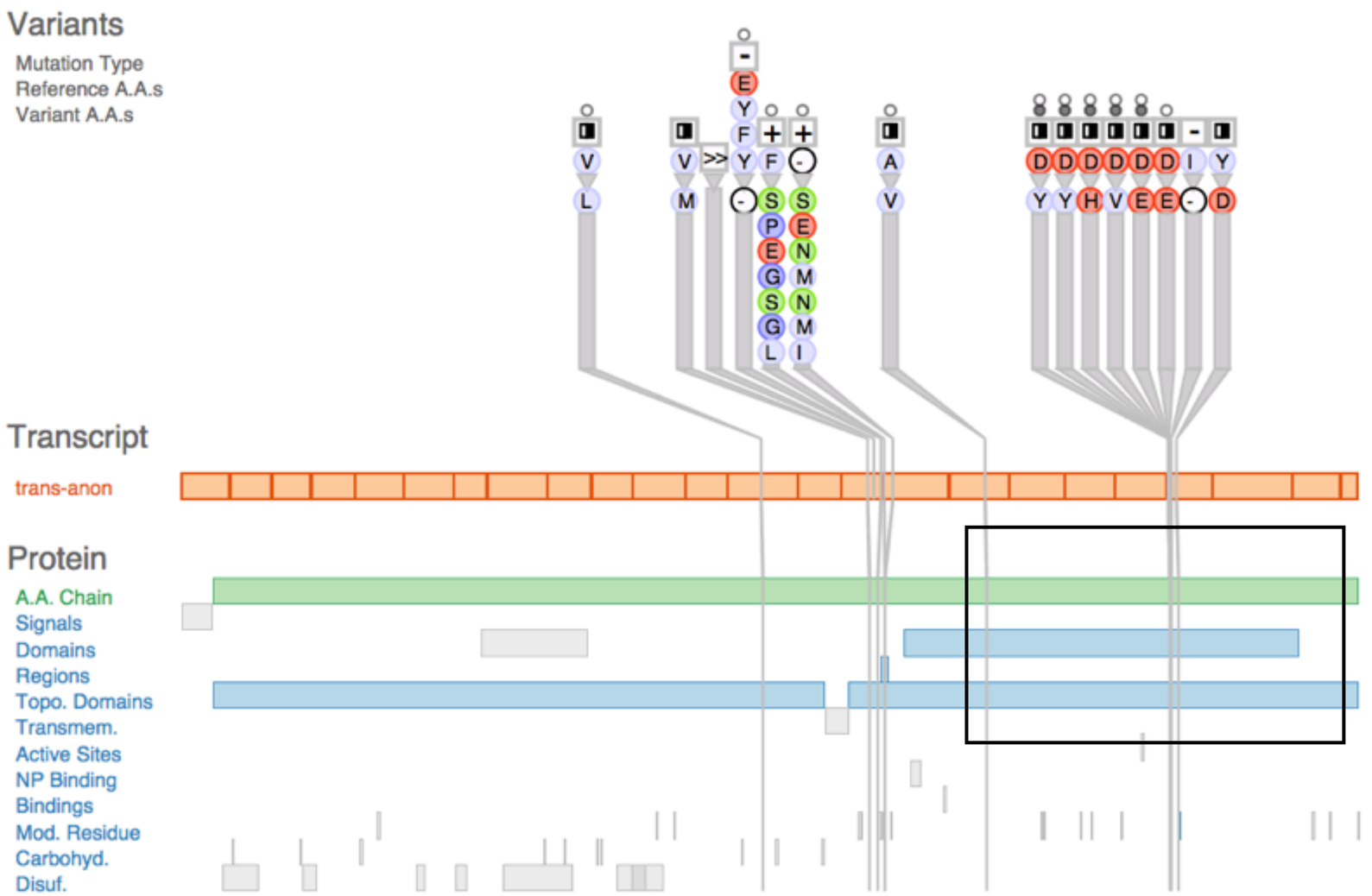
Highly scored gene by sorting metric: known leukemia gene



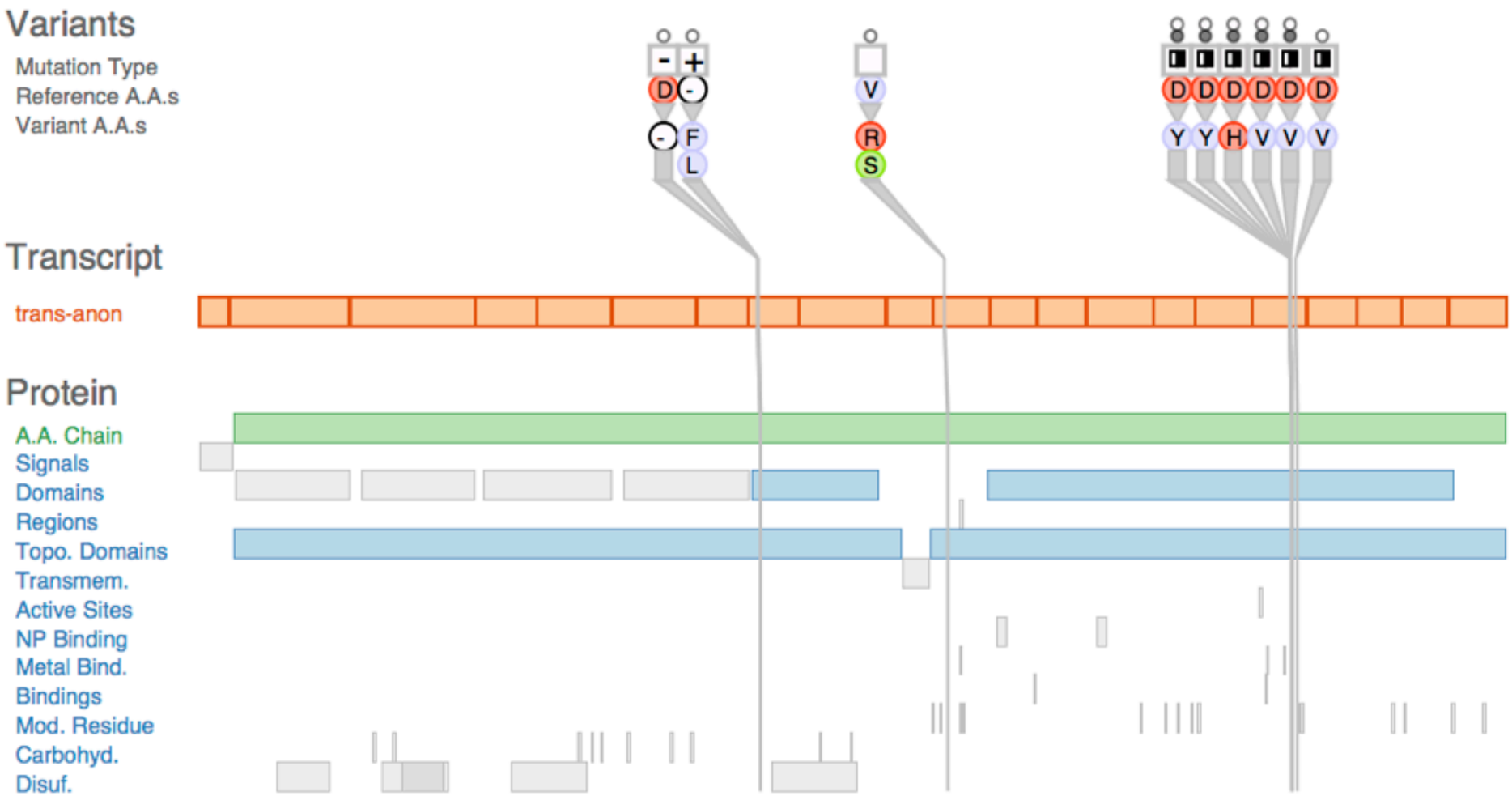
Visual inspection reveals collocation of variants



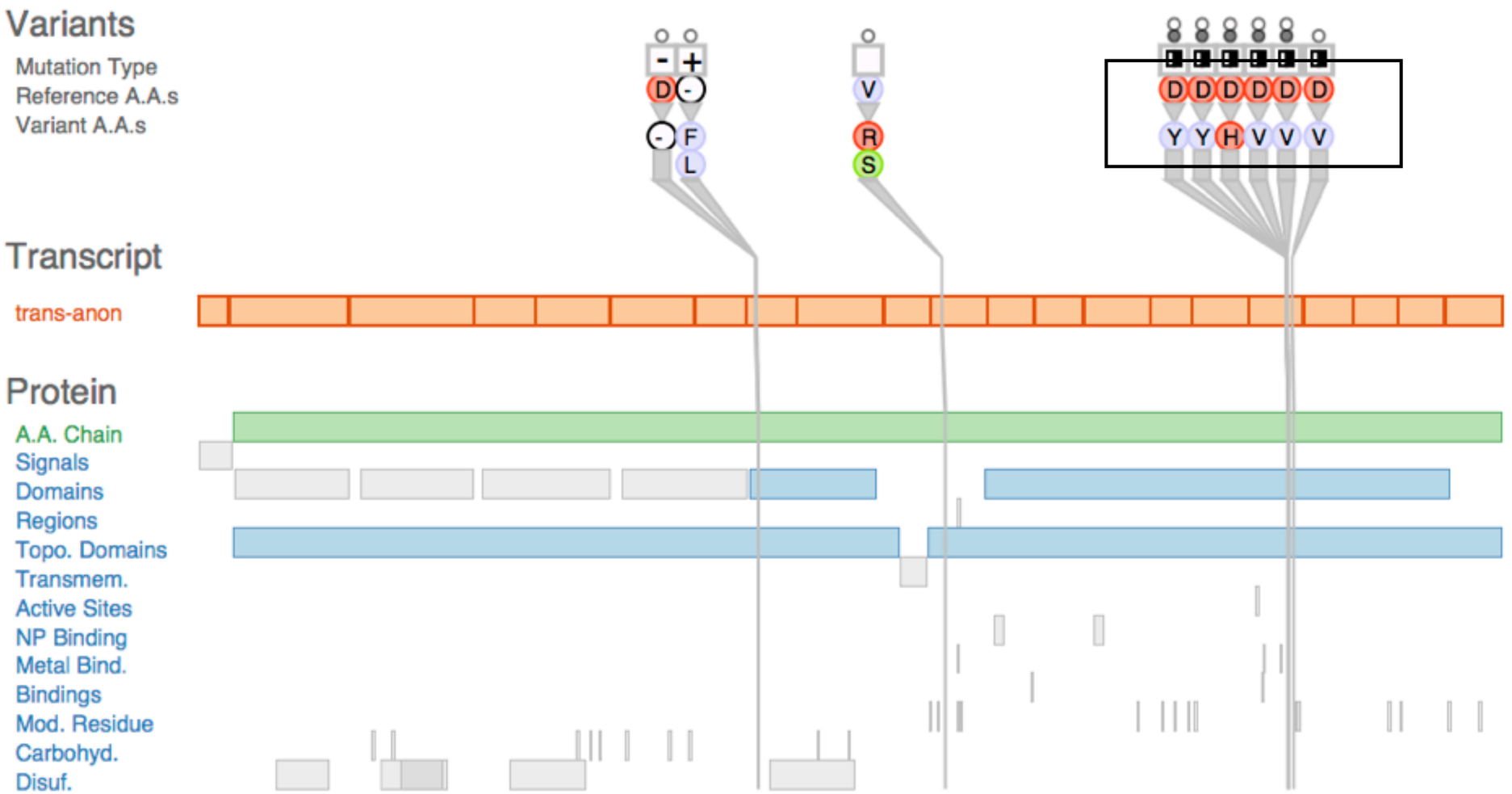
Several functional protein regions affected



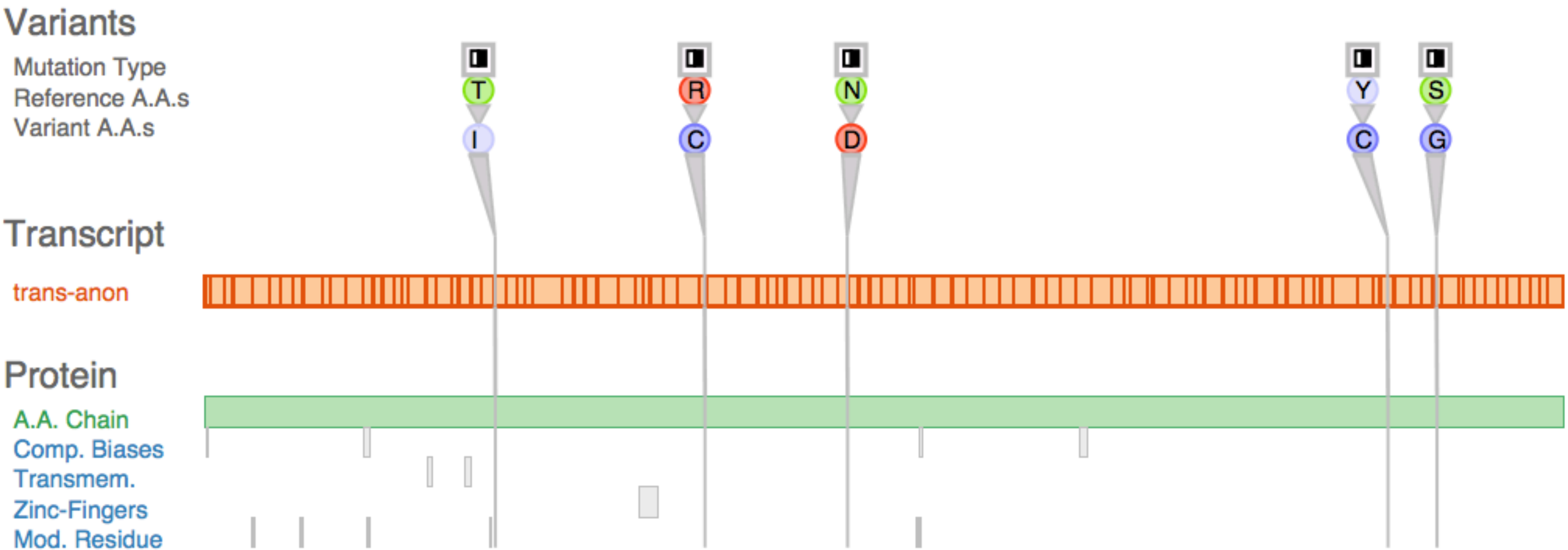
Highly scored by metric: not previously known, good candidate



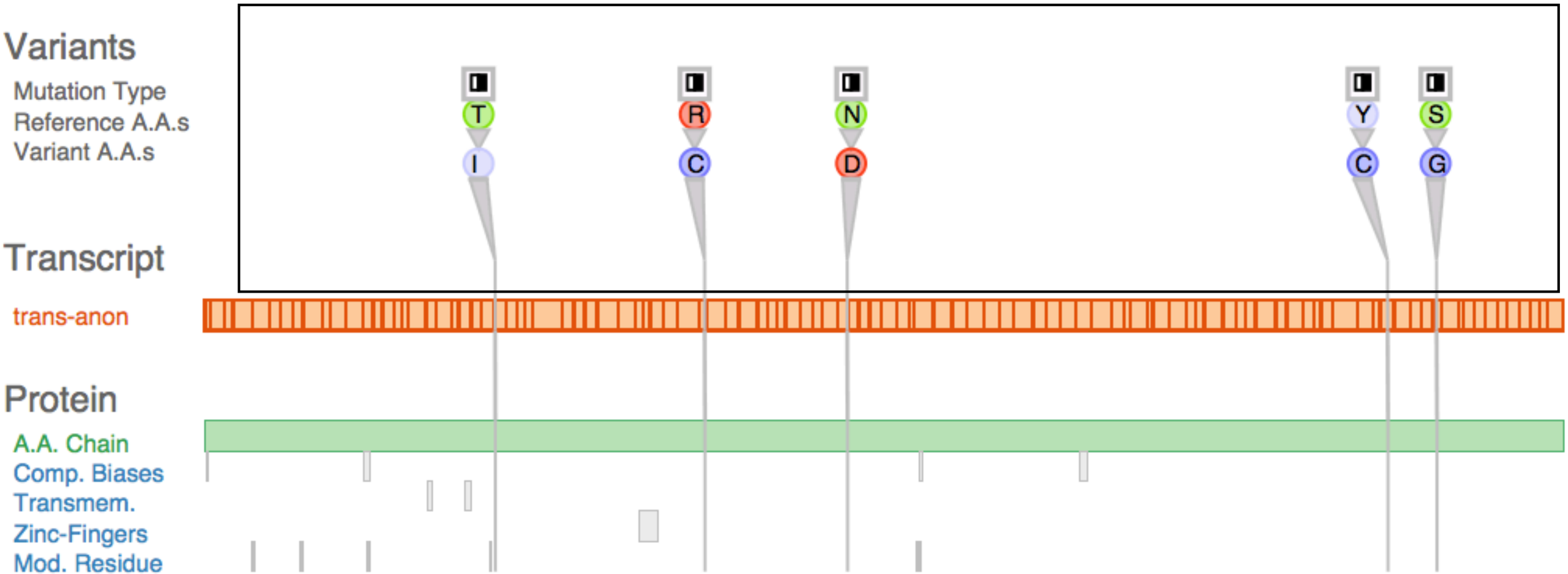
Protein chemical class change evident



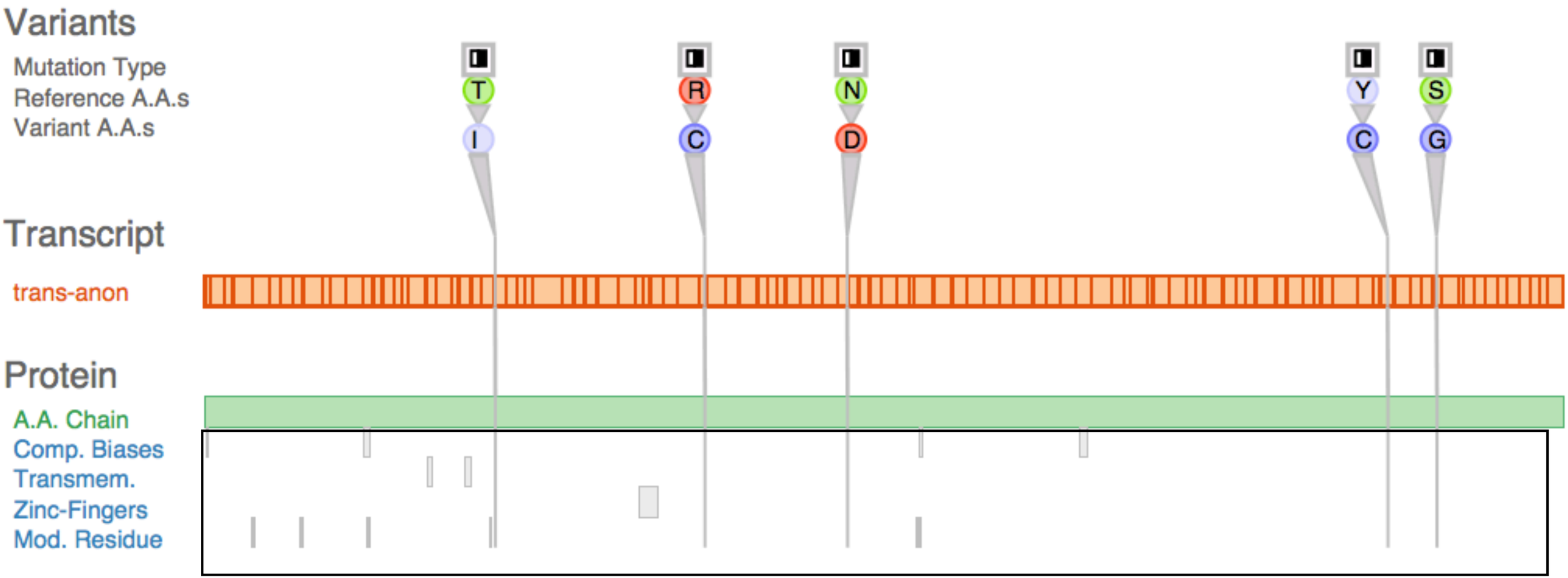
In contrast, low scoring gene



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:

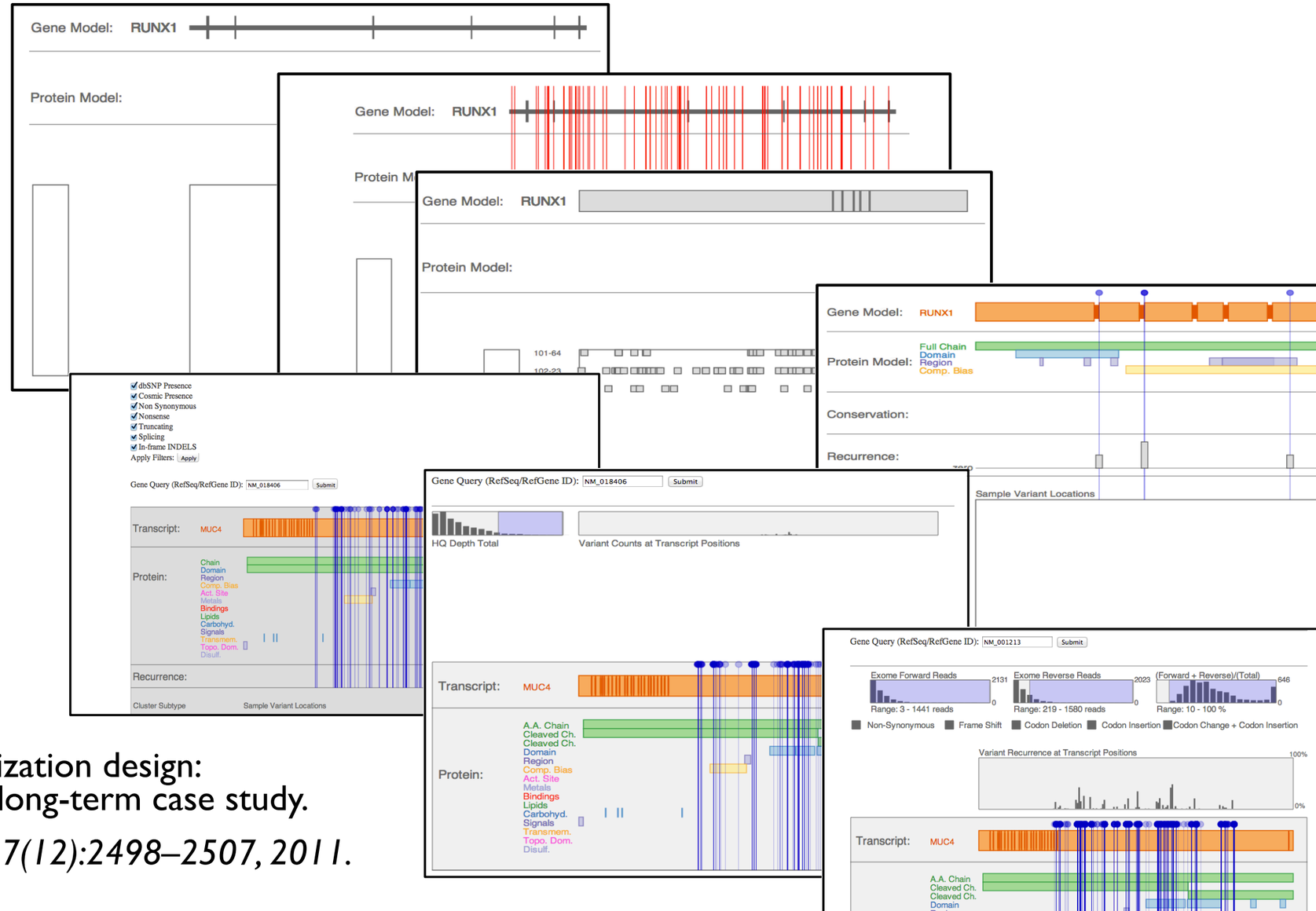
Design Study Methodology: Reflections from the Trenches and from the Stacks.
Sedlmair, Meyer, Munzner. IEEE TVCG 18(12): 2431-2440, 2012 (Proc. InfoVis 2012).

Phase 2: Core Design

5 months

learn → winnow → cast → **discover** → design → implement → deploy → reflect → write

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



Human-centered approaches in geovisualization design: investigating multiple methods through a long-term case study.

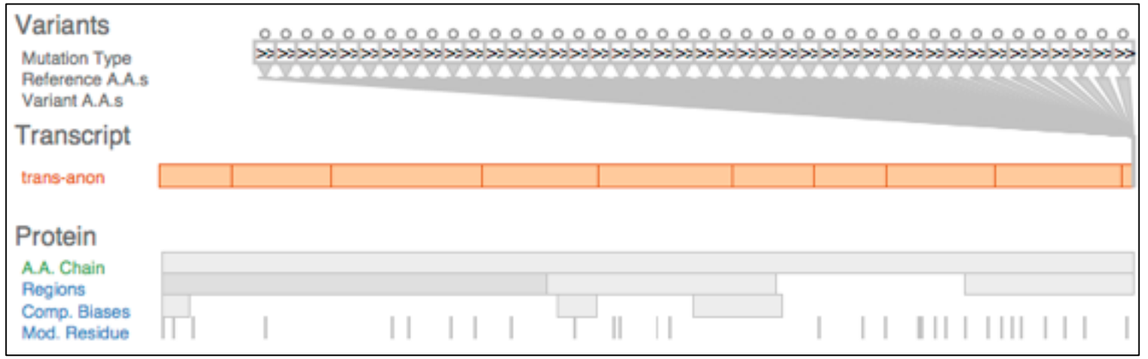
Lloyd and Dykes. IEEE TVCG (Proc. InfoVis), 17(12):2498–2507, 2011.

Phase 3: Two More Tasks

1 month

learn → winnow → cast → **discover** → design → implement → deploy → reflect → write

- 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

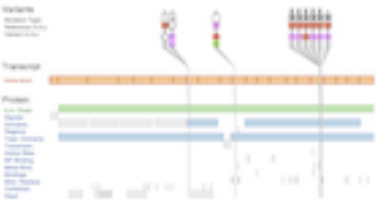
Themes, Revisited

- 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
  - Variant View: reduce interaction load with better visual encoding

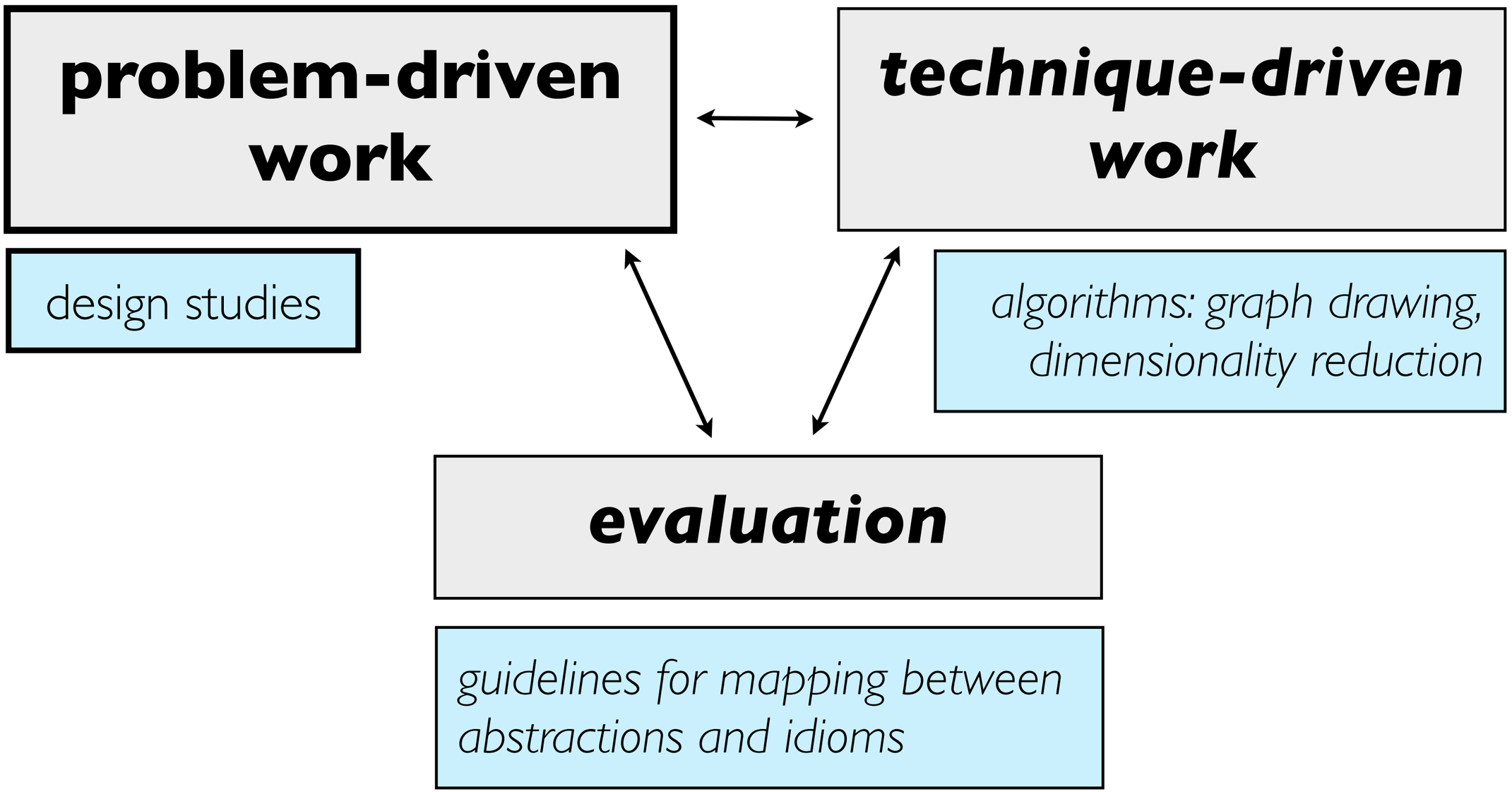
- transferability from design studies
  - DSM: reflection to confirm/refute/refine/propose guidelines

RelEx
in-car networks



VariantView
genomics

Research Interests



Research Interests: giCentre Context

problem-driven work ⟷ ***technique-driven work***

design studies

evaluation

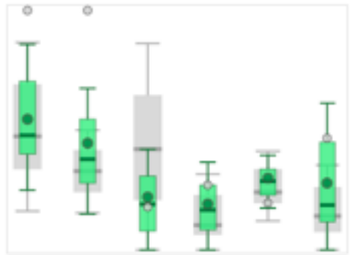
guidelines for mapping between abstractions and idioms

hierarchical layouts
IEEE TVCG 2009

sketchy rendering
IEEE TVCG 2012

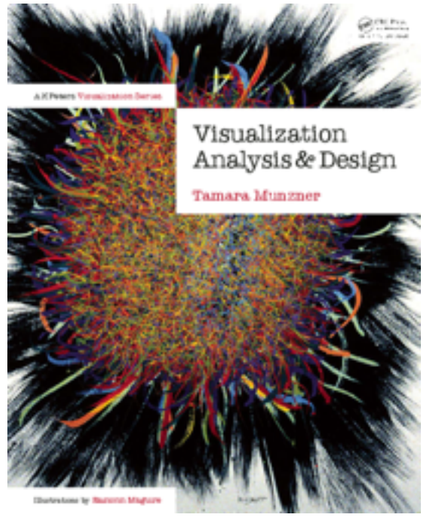
spatially ordered treemaps
IEEE TVCG 2012

geographically weighted boxplots
IEEE TVCG 2007



Further Information

- further info
  - http://www.cs.ubc.ca/~tmm/talks.html#london14 (this talk, and many others)
  - http://www.cs.ubc.ca/group/infovis (papers, software, videos)
  - http://www.cs.ubc.ca/~tmm/courses/infovis/book (book: to appear)
    - Visualization Analysis and Design. Munzner. AK Peters 2014



- 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