

Data Visualization in Genomics and In-Car Network Engineering

Tamara Munzner

Department of Computer Science

University of British Columbia

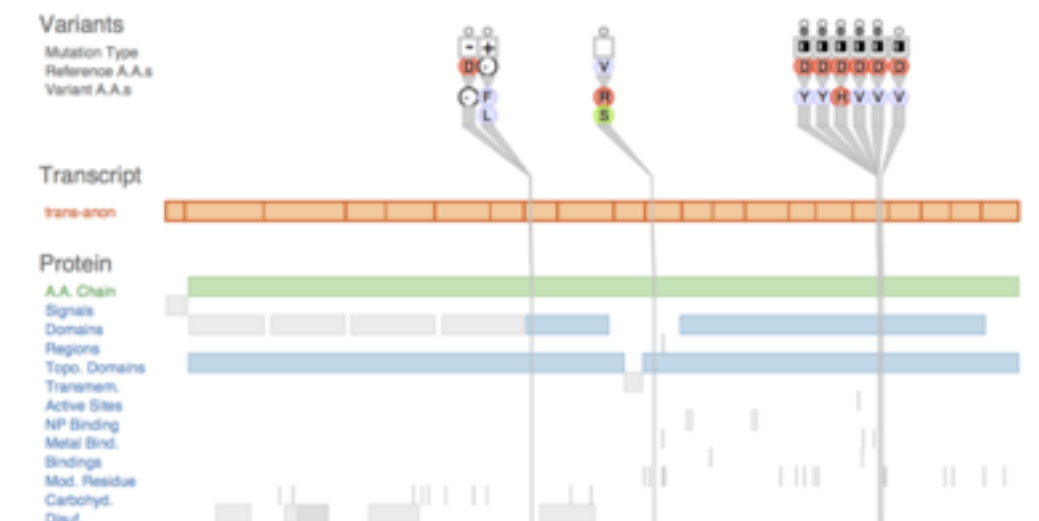
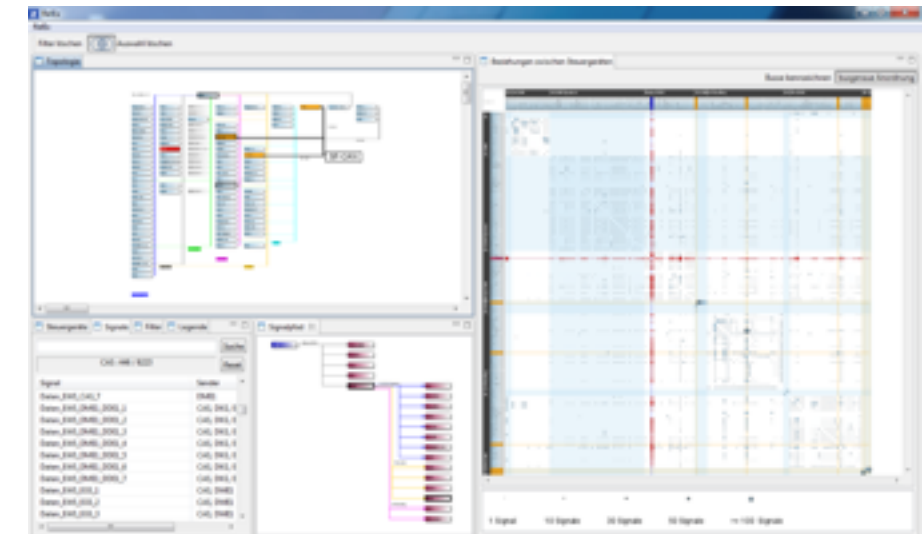
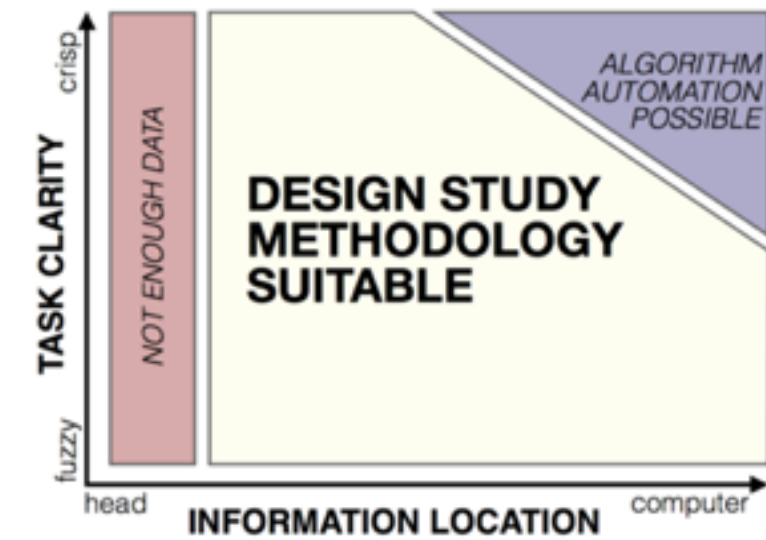
City University of London, Computer Science Department Seminar

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.

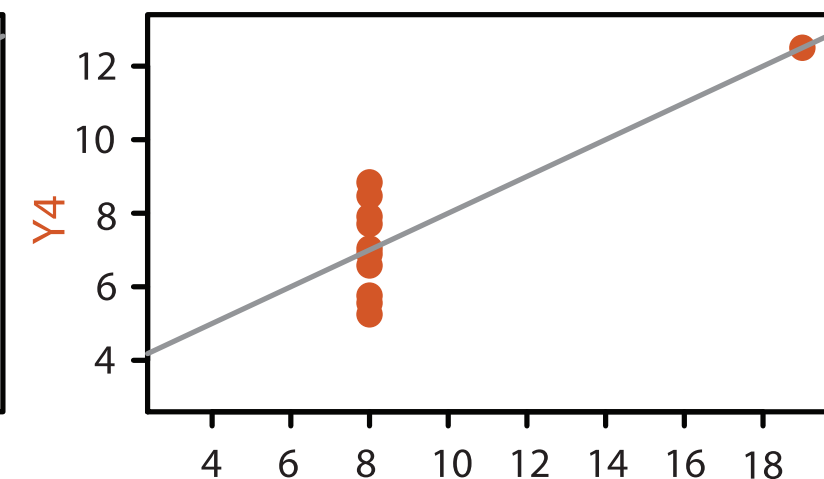
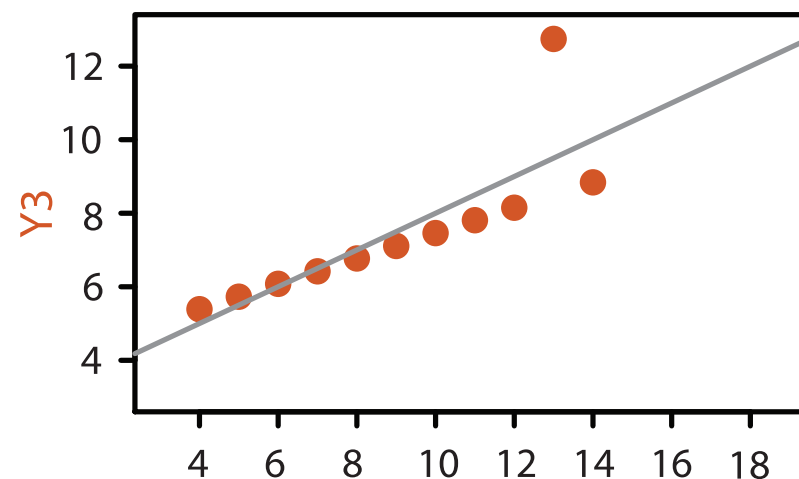
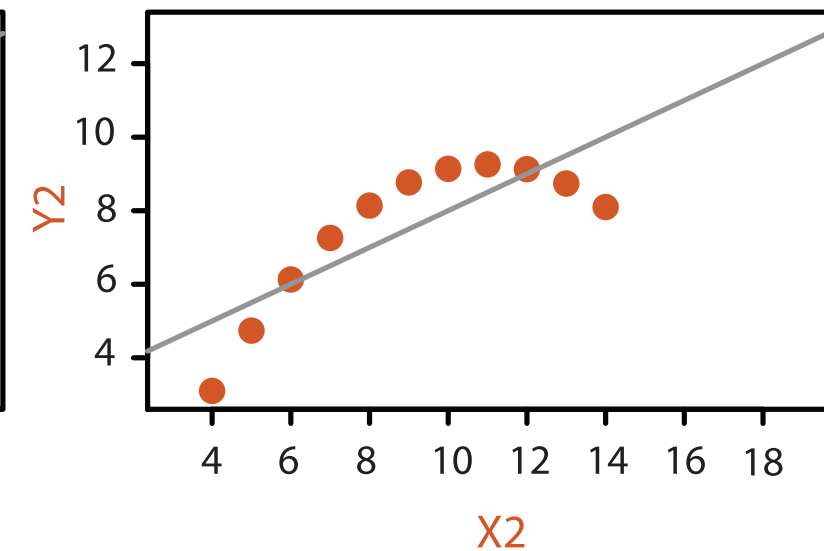
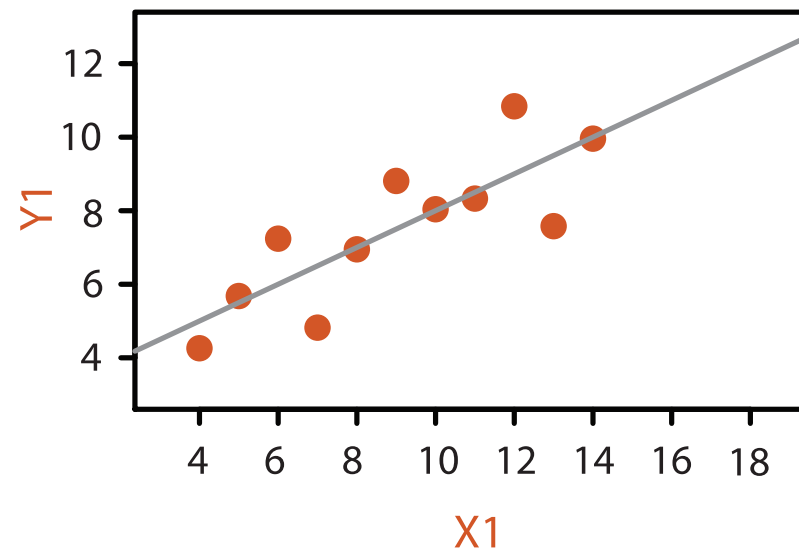
Defining Visualization

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

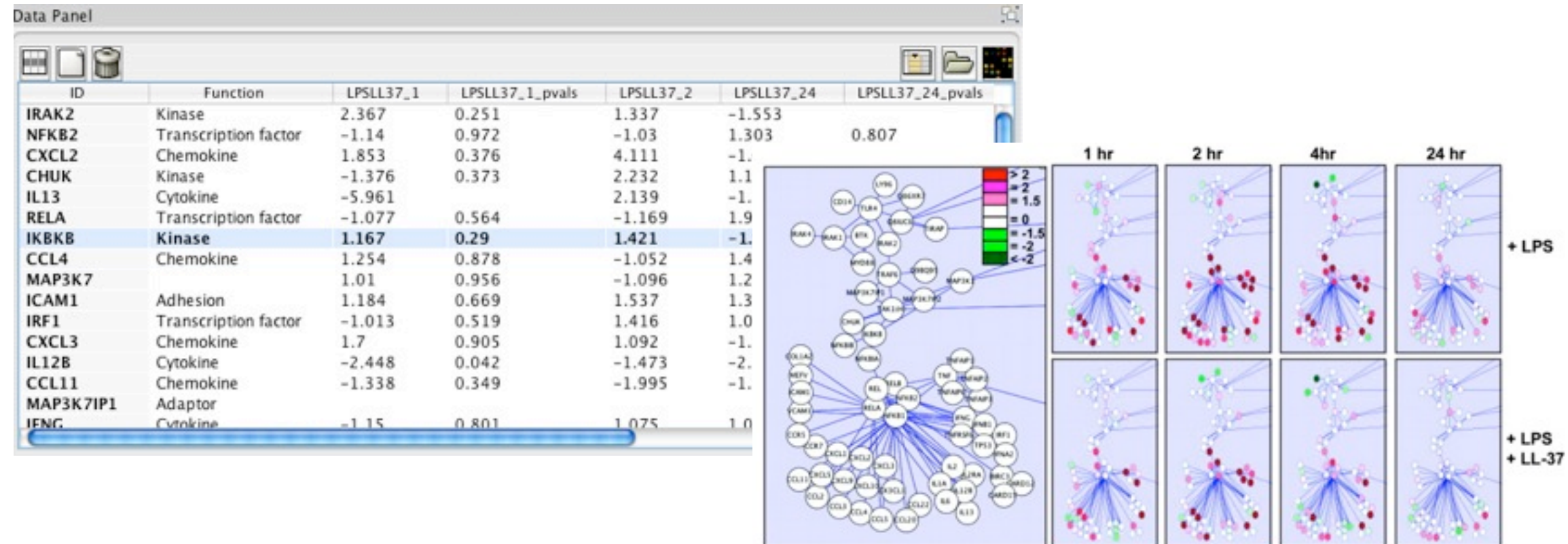
x mean	9.0
x variance	10.0
y mean	7.50
y variance	3.75
x/y correlation	0.816



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

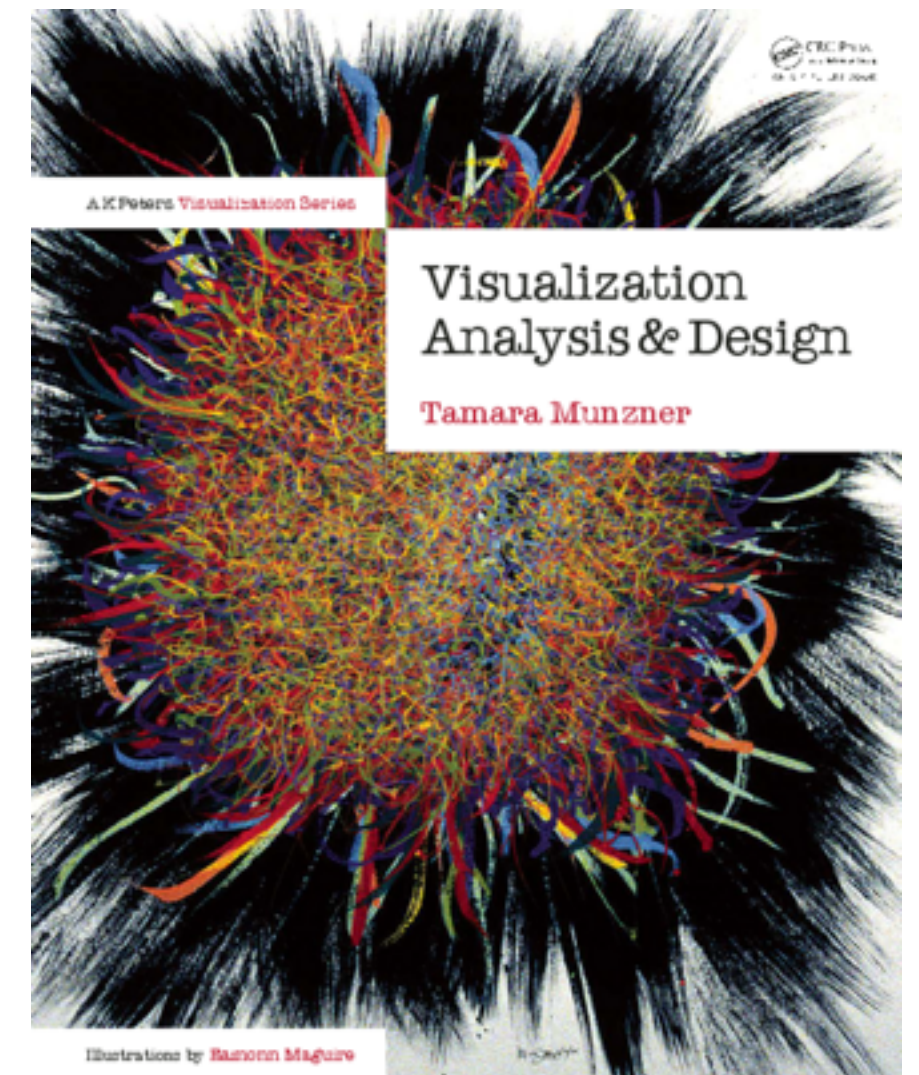
Defining Visualization

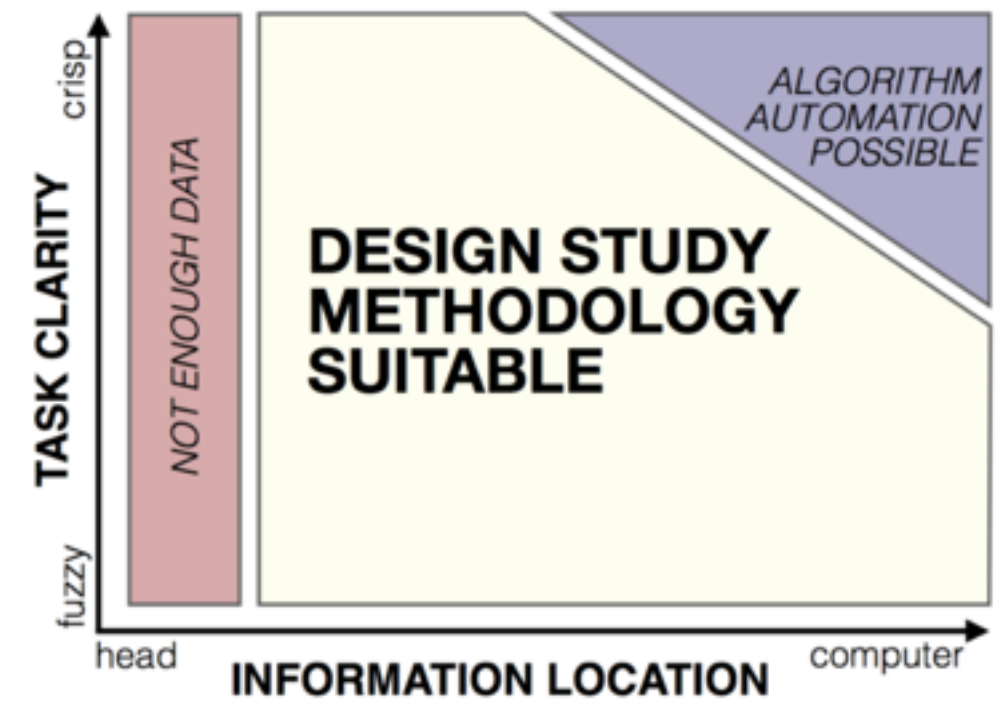
Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- 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

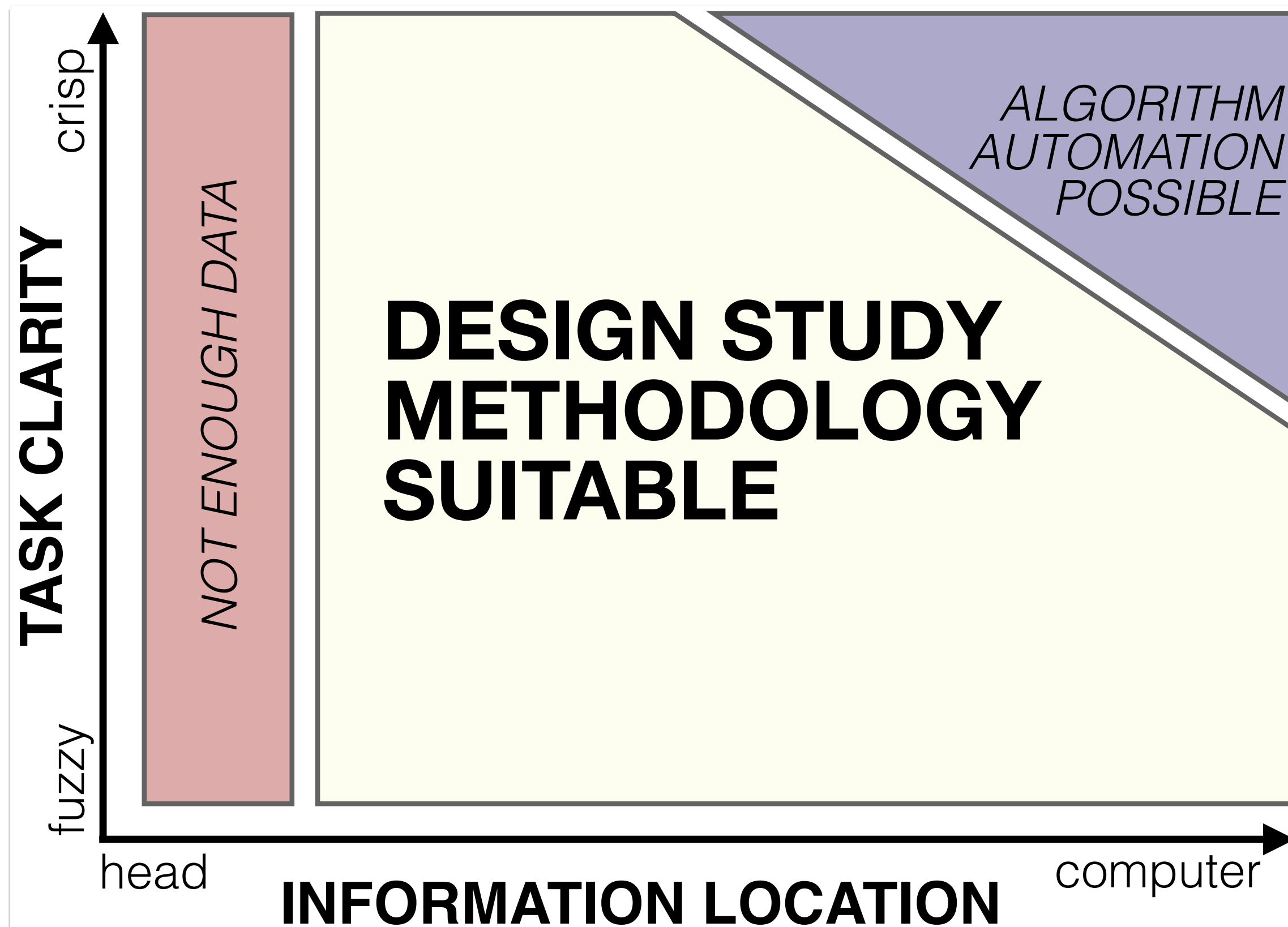
more at:

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

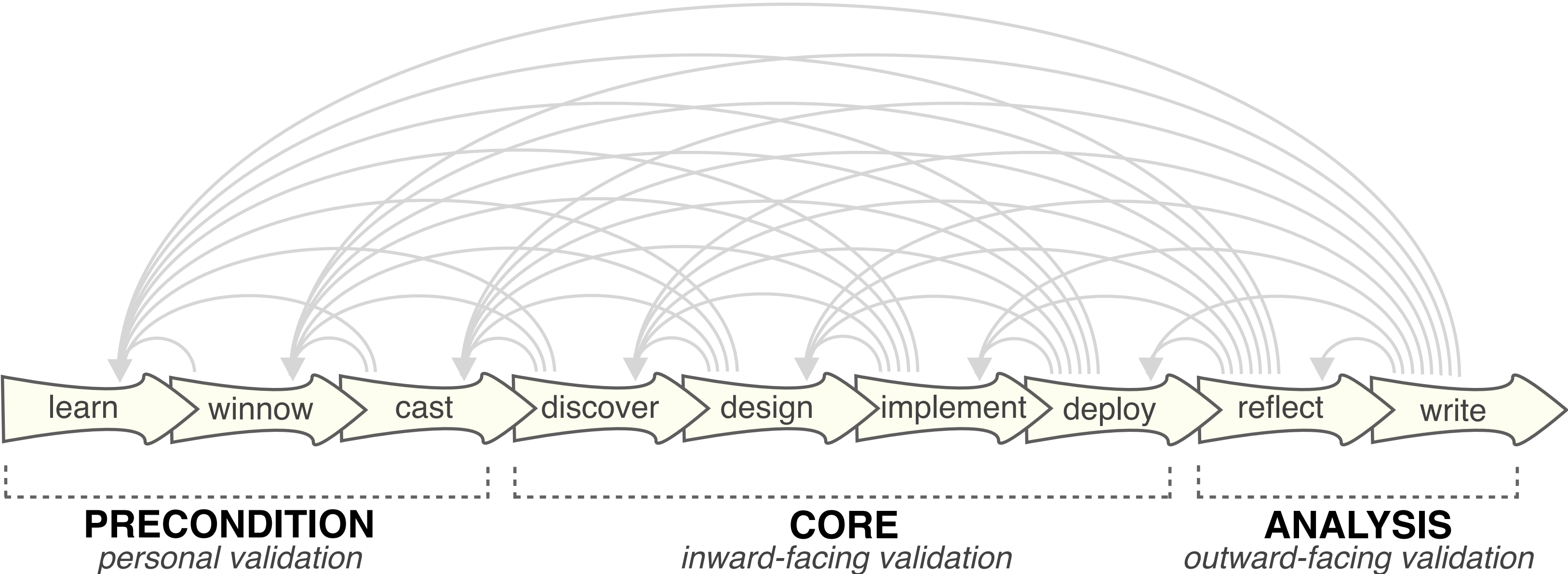
more at:

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

When To Do Design Studies

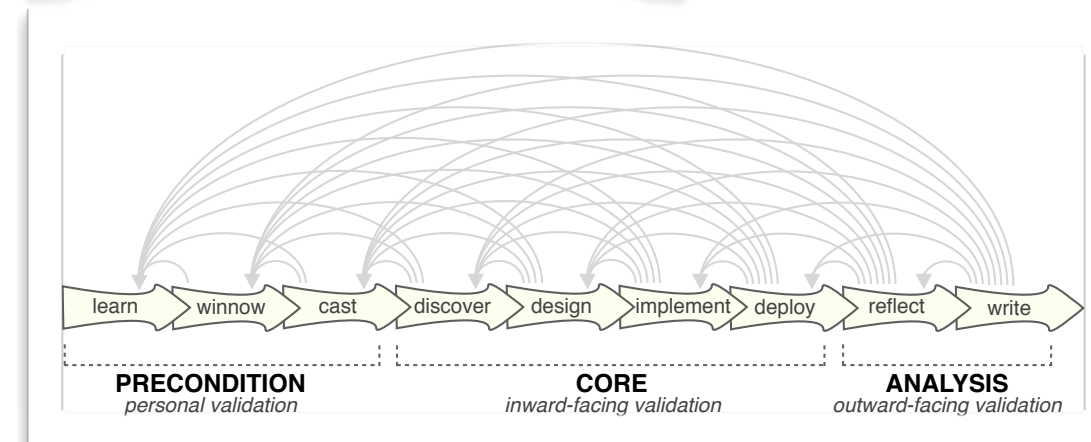
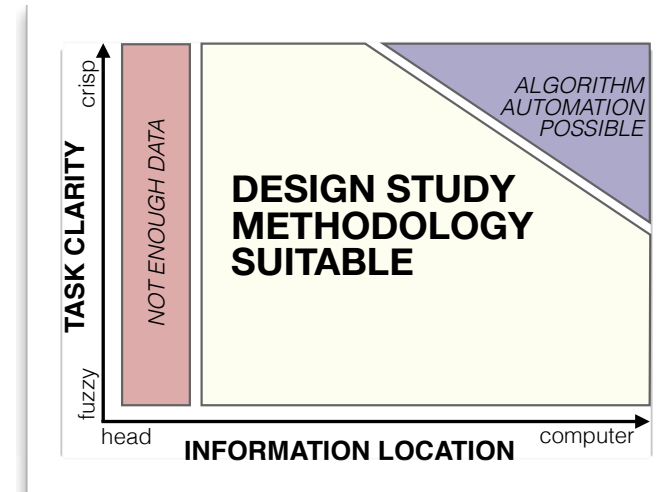


Nine-Stage Framework



How To Do Design Studies

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



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

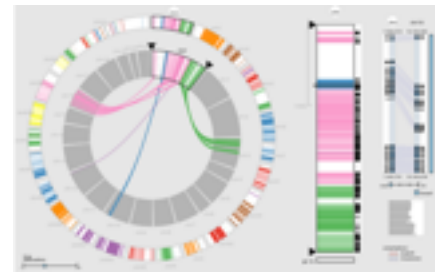
design studies

Must be first!

Am I ready?



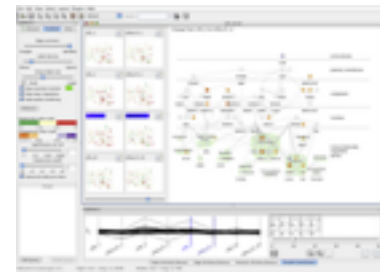
Design Studies: Lessons learned after 21 of them



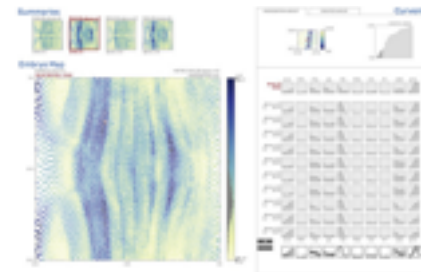
MizBee
genomics



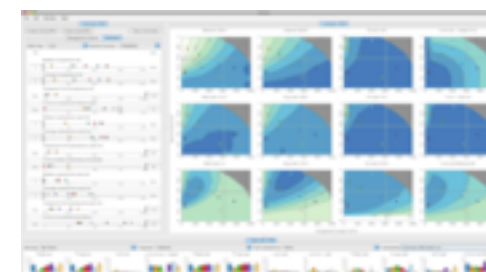
Pathline
genomics



Cerebral
genomics



MulteeSum
genomics



Vismon
fisheries management



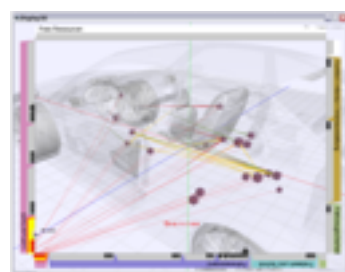
QuestVis
sustainability



WiKeVis
in-car networks



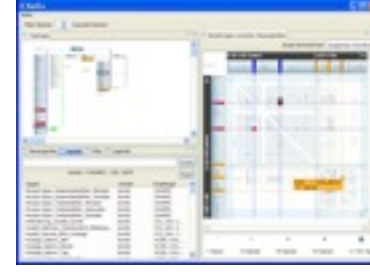
MostVis
in-car networks



Car-X-Ray
in-car networks



ProgSpy2010
in-car networks



ReEx
in-car networks



Cardiogram
in-car networks



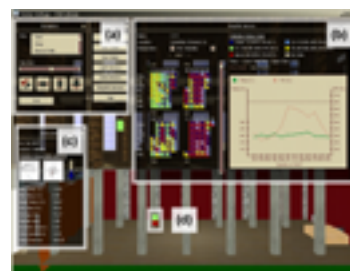
AutobahnVis
in-car networks



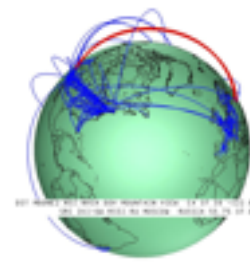
VisTra
in-car networks



Constellation
linguistics



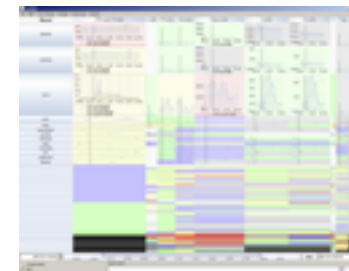
LibVis
cultural heritage



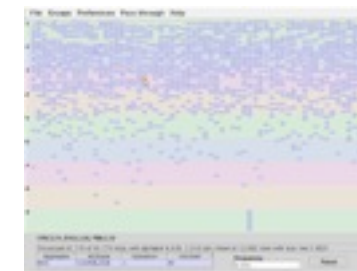
Caidants
multicast



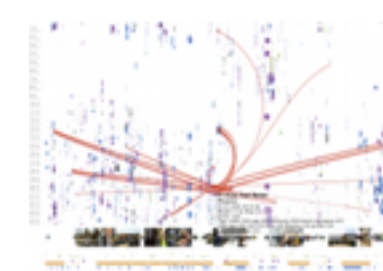
SessionViewer
web log analysis



LiveRAC
server hosting



PowerSetViewer
data mining

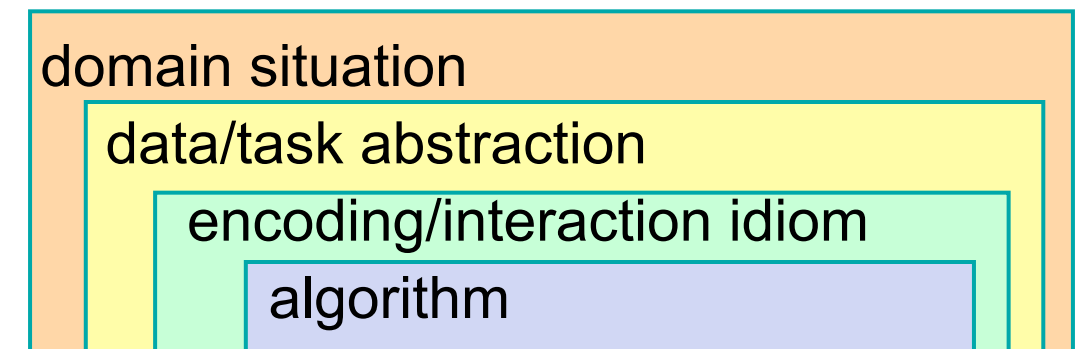


LastHistory
music listening

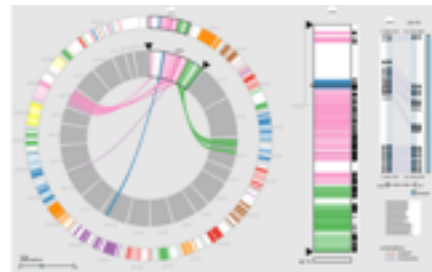
- commonality of representations cross-cuts domains!

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



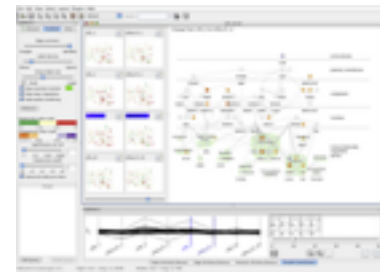
Today's Focus



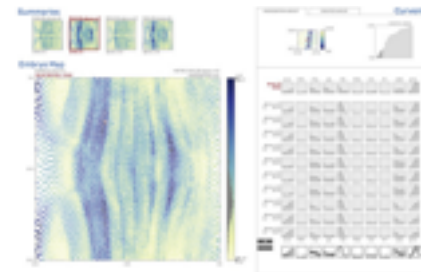
MizBee
genomics



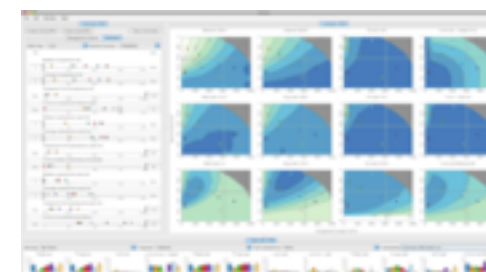
Pathline
genomics



Cerebral
genomics



MulteeSum
genomics



Vismon
fisheries management



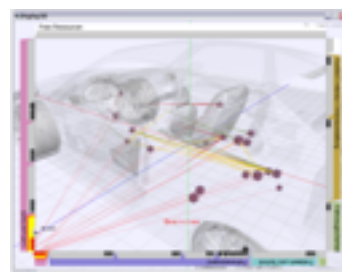
QuestVis
sustainability



WiKeVis
in-car networks



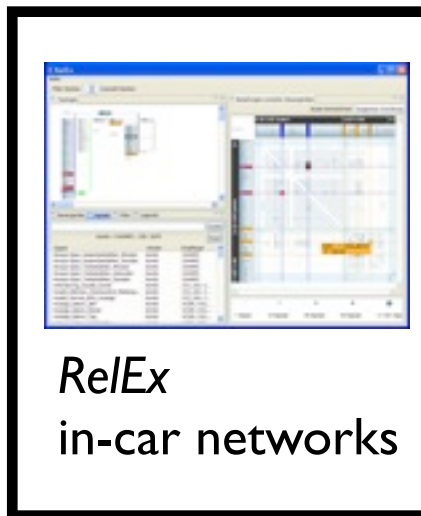
MostVis
in-car networks



Car-X-Ray
in-car networks



ProgSpy2010
in-car networks



RelEx
in-car networks



Cardiogram
in-car networks



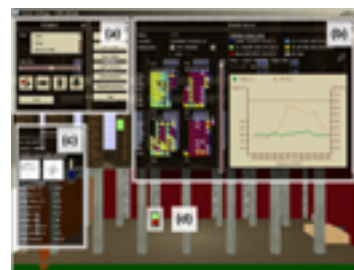
AutobahnVis
in-car networks



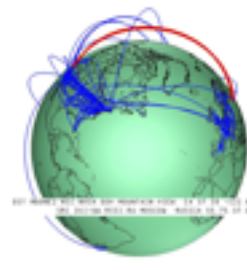
VisTra
in-car networks



Constellation
linguistics



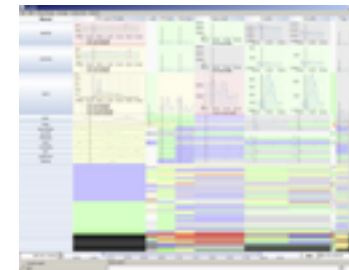
LibVis
cultural heritage



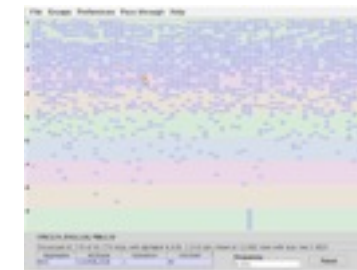
Caidants
multicast



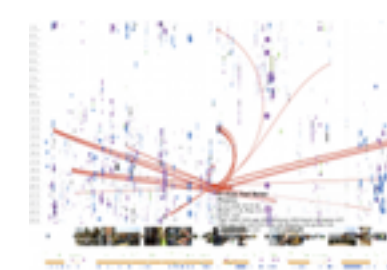
SessionViewer
web log analysis



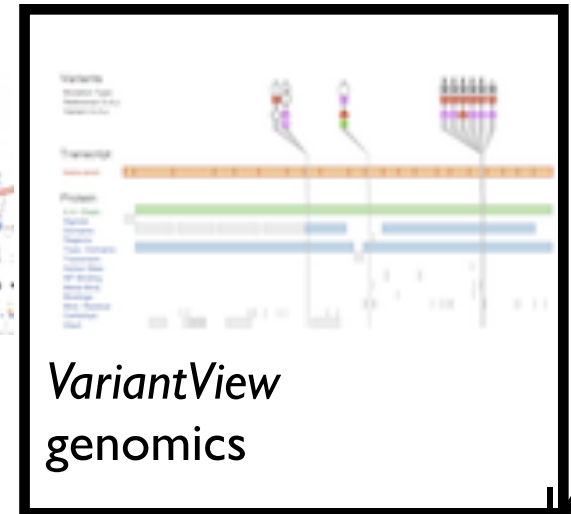
LiveRAC
server hosting



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

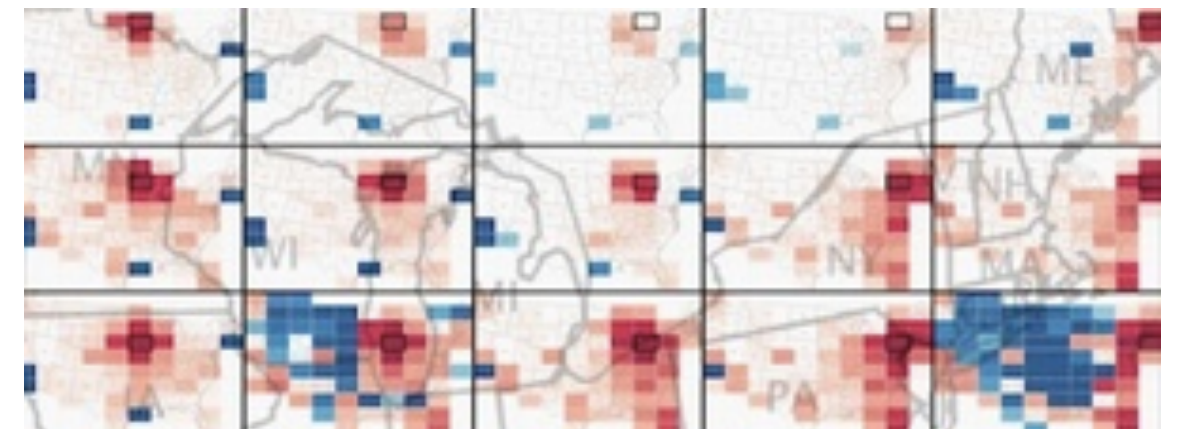
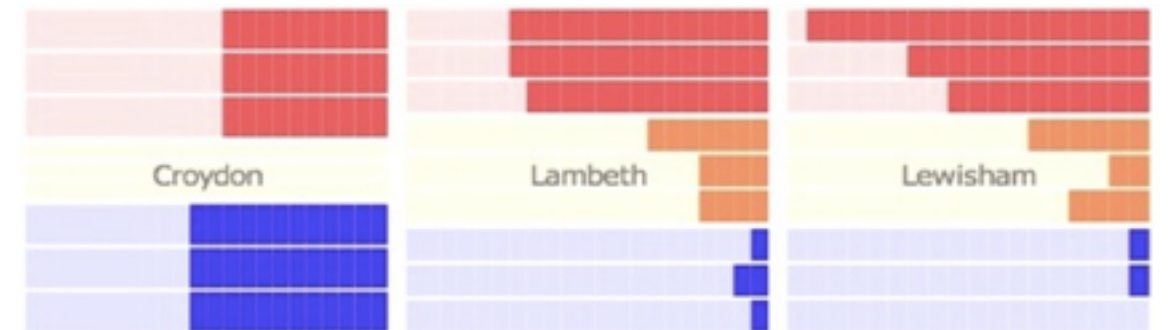
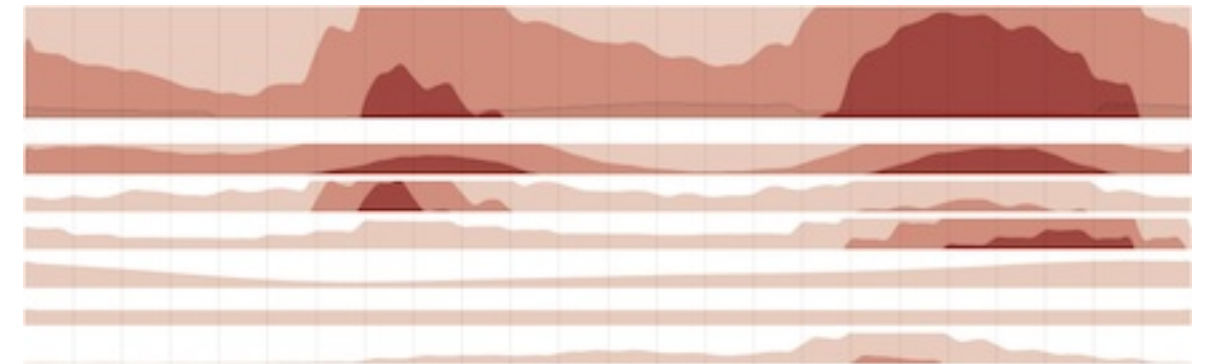
Creative user-centered visualization design for energy analysts and modelers. Goodwin, Dykes, Jones, Dillingham, Dove, Duffy, Kachkaev, Slingsby, Wood. *IEEE Transactions on Visualization and Computer Graphics*, 19(12), pp. 2516-2525, 2013.

- 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

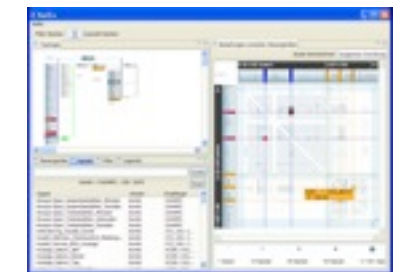
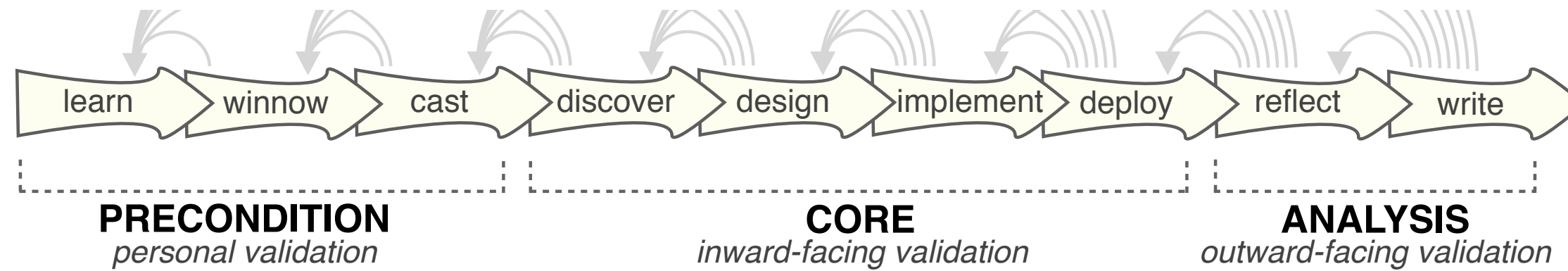
- ODMaps

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

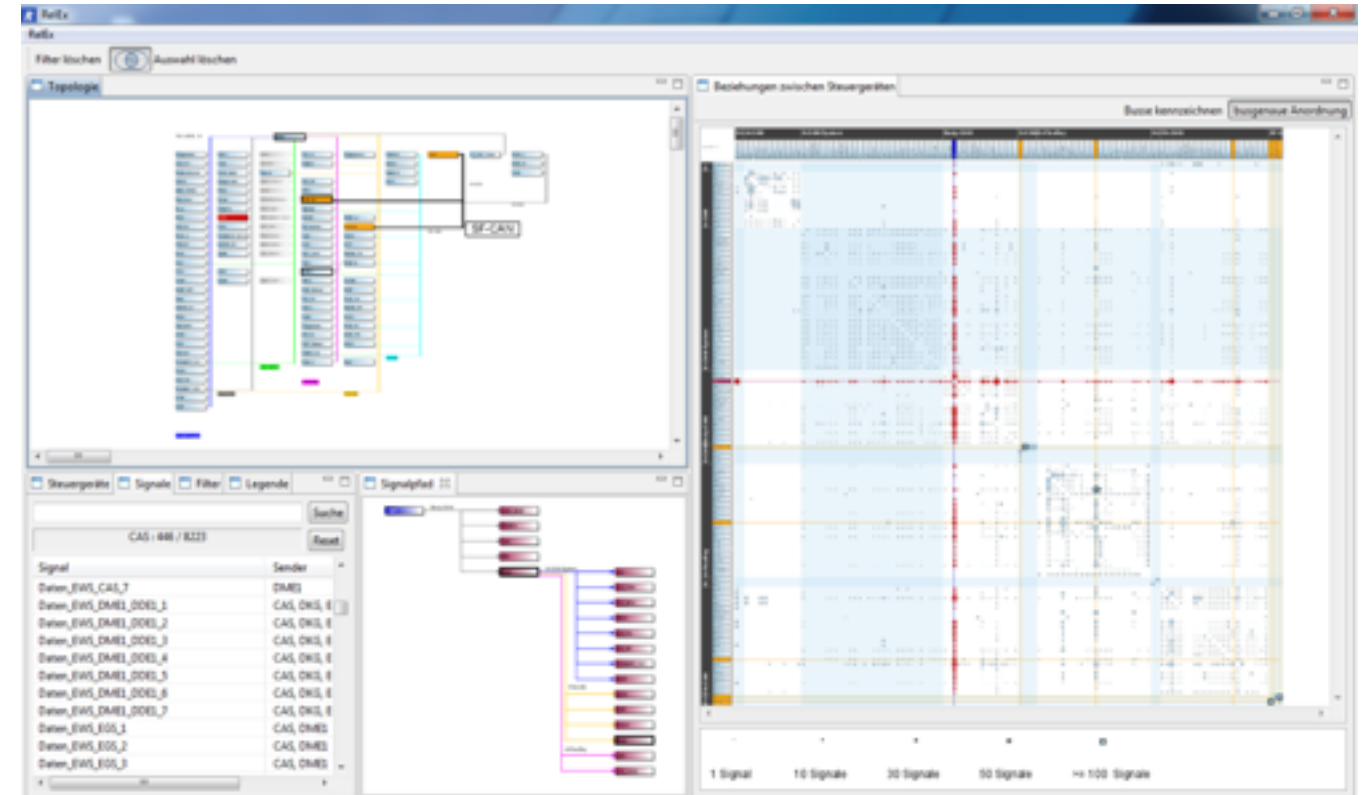
RelEx

Visualization for Actively Changing Overlay Network Specifications

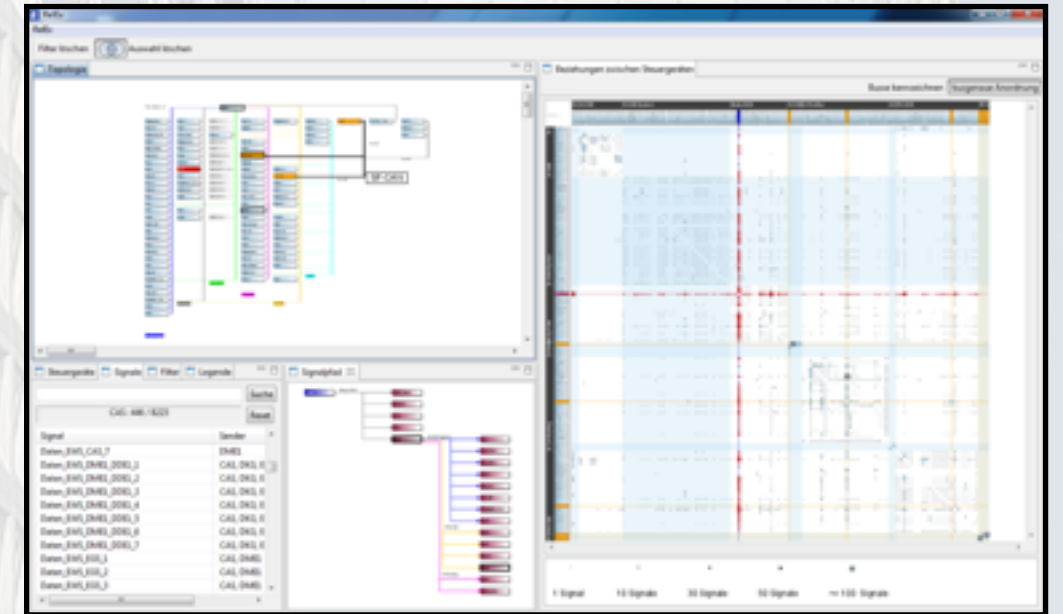
joint work with:

Michael Sedlmair, Annika Frank, Andreas Butz

<http://www.cs.ubc.ca/labs/imager/tr/2012/relex/>



Domain: **In-car network engineering**



Abstractions

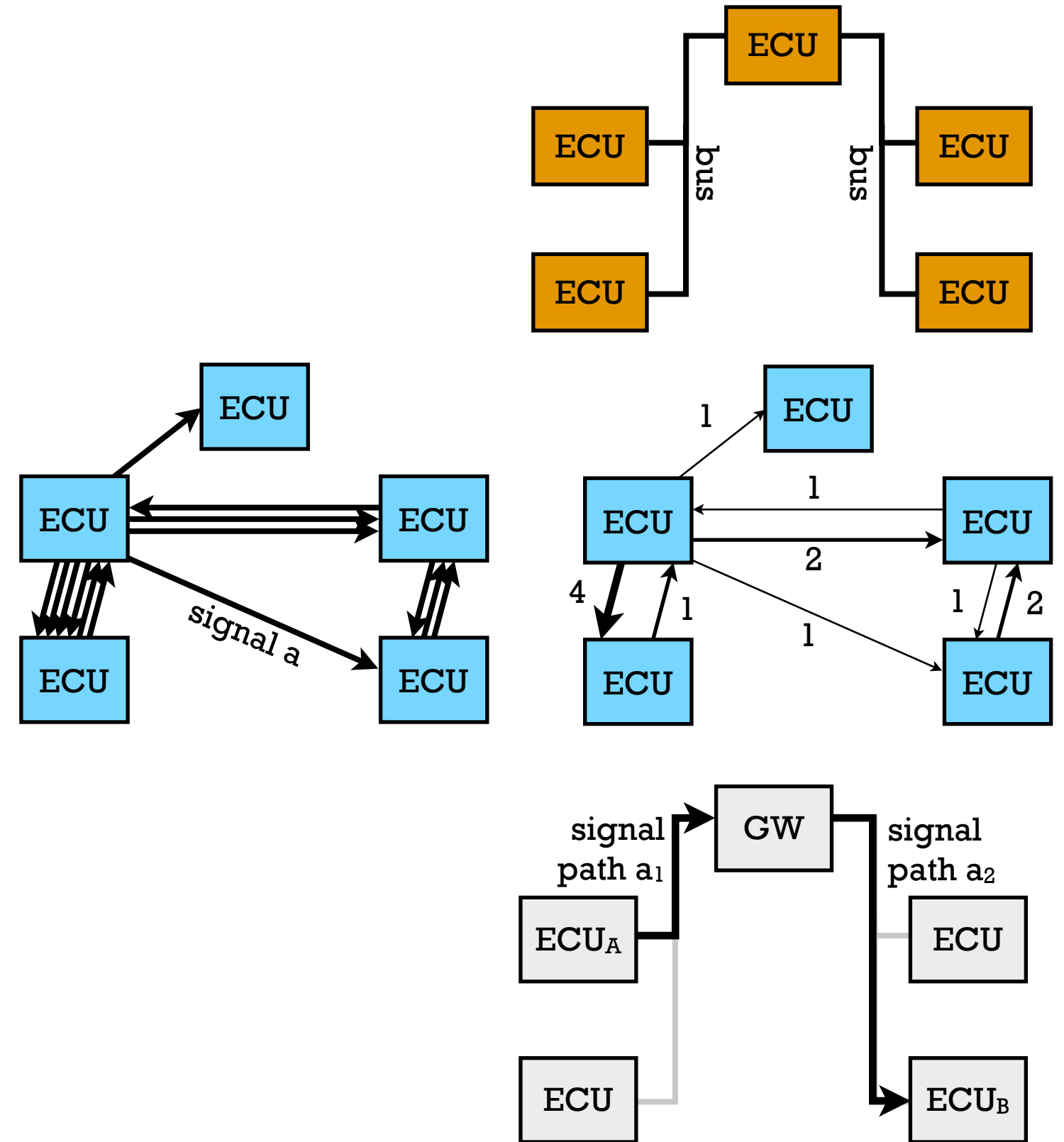
DATA

In-car Electronics



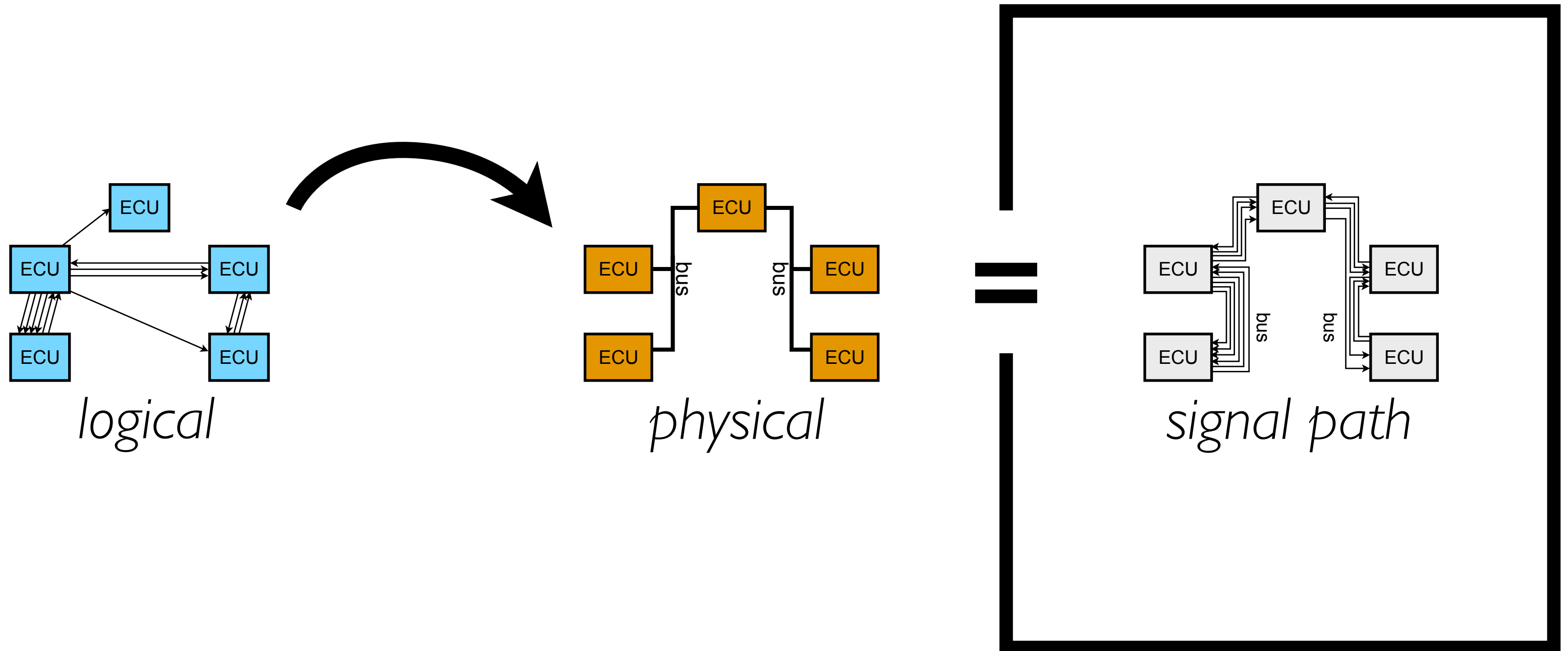
Data Abstraction: 3 Networks

- **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



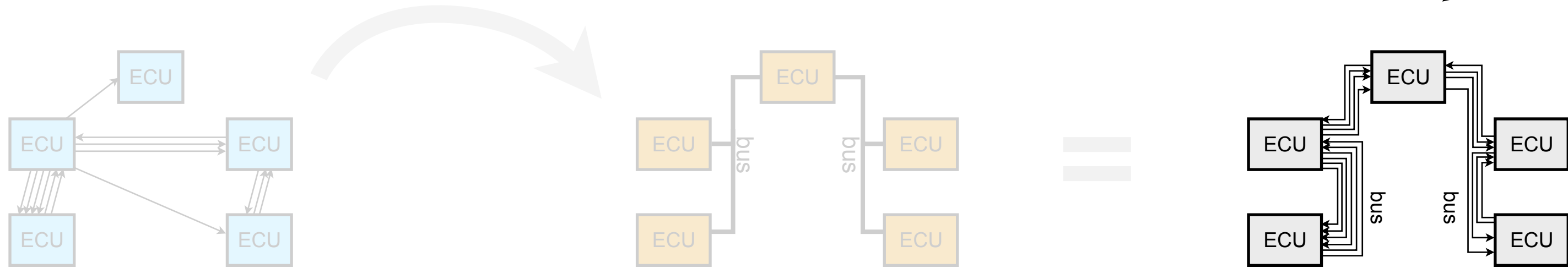
Task Abstraction: Mapping

- specify overlay network that maps logical onto physical



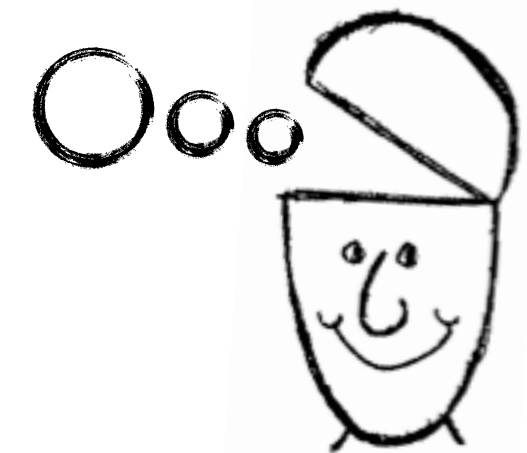
Task Abstraction: Optimizing

- traffic optimization



Many constraints

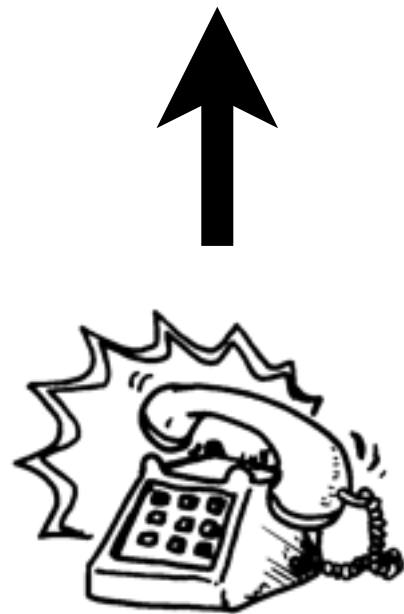
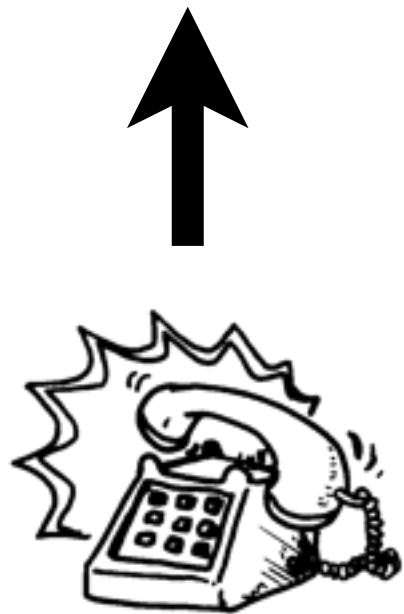
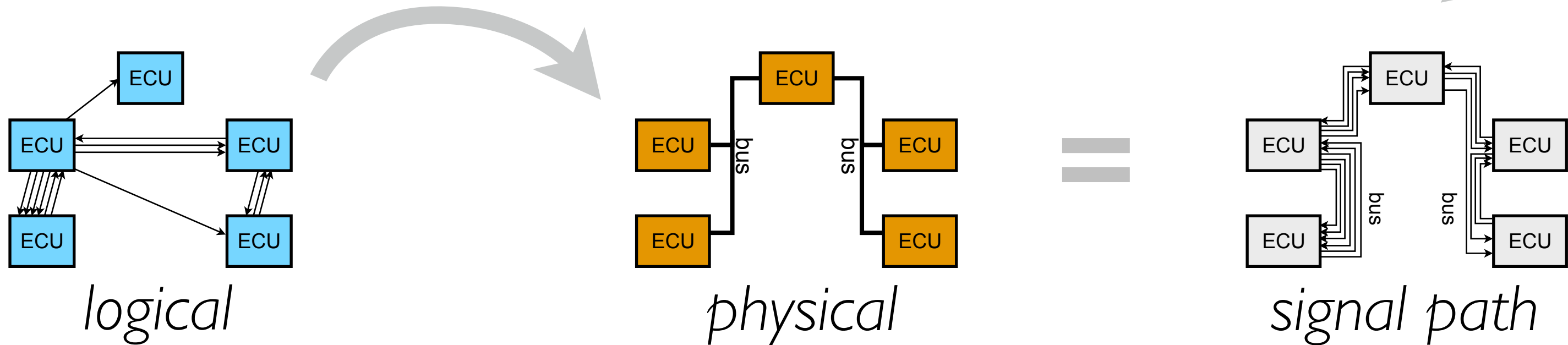
bandwidth ... delay/real time ...
path length ... load balance ...
reliability ... money ...



-- engineer, BMW --

Task Abstraction: Changing

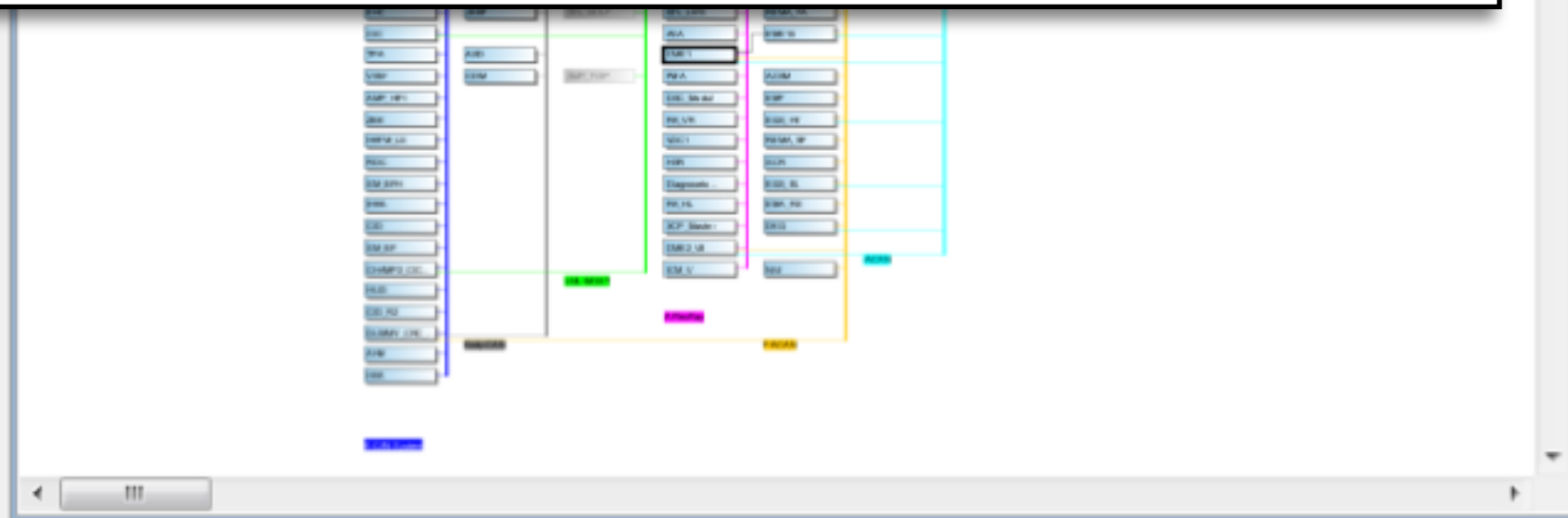
- external change requests



Change
(trivial requests might lead to complex changes)

Idioms

RELEX: Relation Explorer



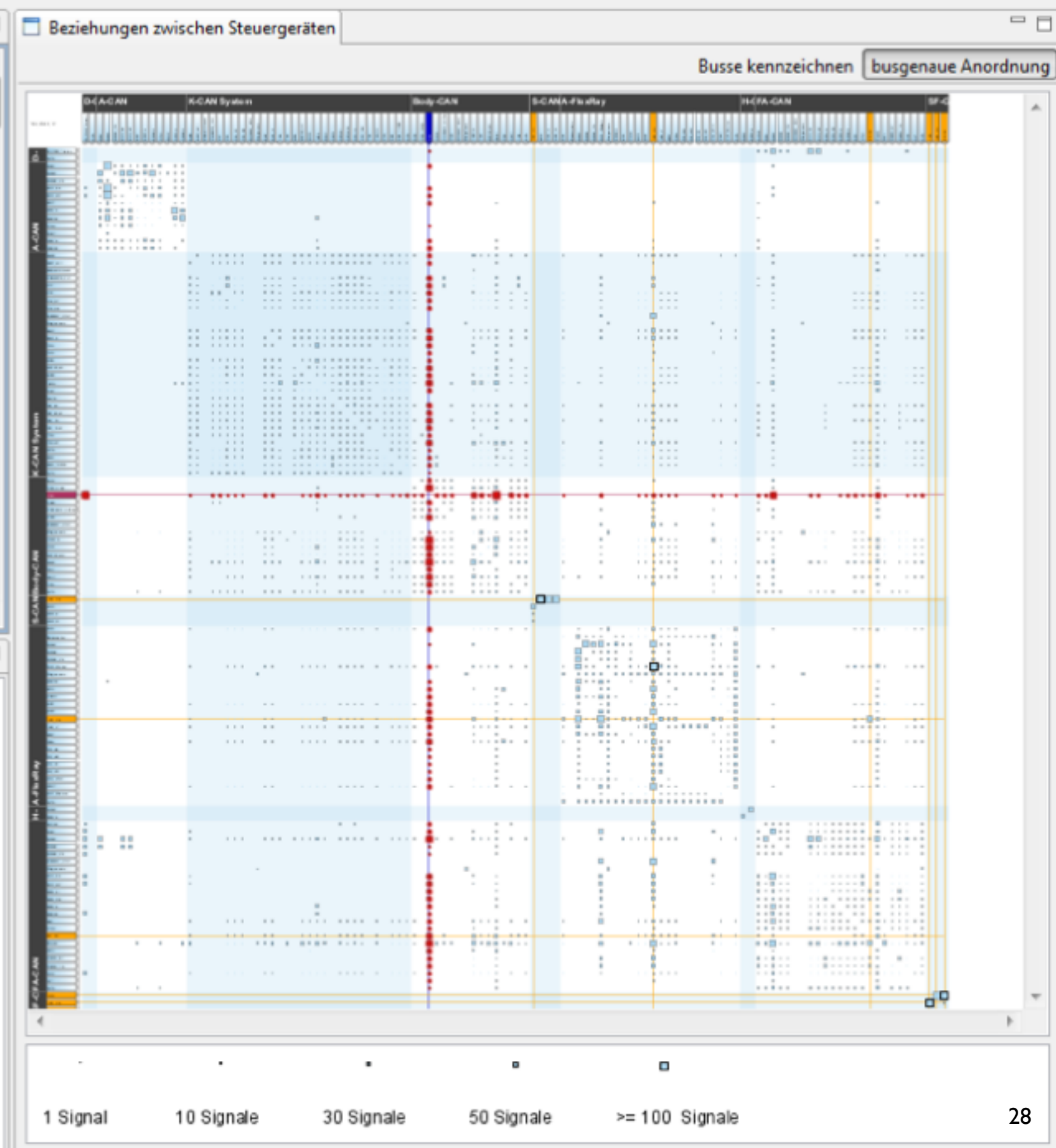
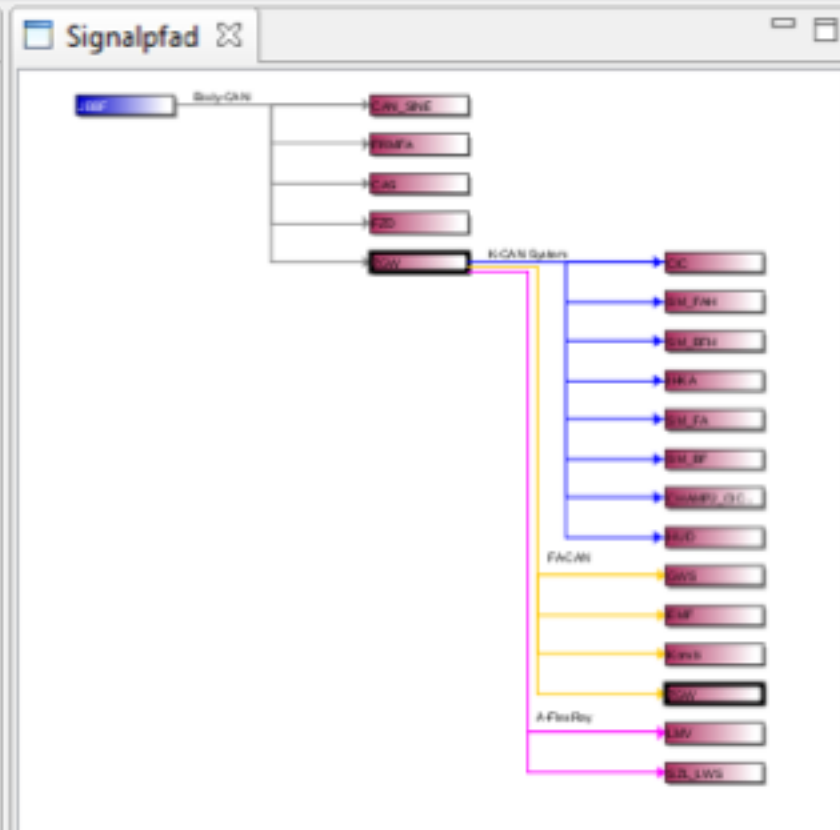
Steuergeräte Signale Filter Legende

Suche

CAS: 446 / 8223

Reset

Signal	Sender
Daten_EWS_CAS_7	DME1
Daten_EWS_DME1_DDE1_1	CAS, DKG, E
Daten_EWS_DME1_DDE1_2	CAS, DKG, E
Daten_EWS_DME1_DDE1_3	CAS, DKG, E
Daten_EWS_DME1_DDE1_4	CAS, DKG, E
Daten_EWS_DME1_DDE1_5	CAS, DKG, E
Daten_EWS_DME1_DDE1_6	CAS, DKG, E
Daten_EWS_DME1_DDE1_7	CAS, DKG, E
Daten_EWS_EGS_1	CAS, DME1
Daten_EWS_EGS_2	CAS, DME1
Daten_EWS_EGS_3	CAS, DME1

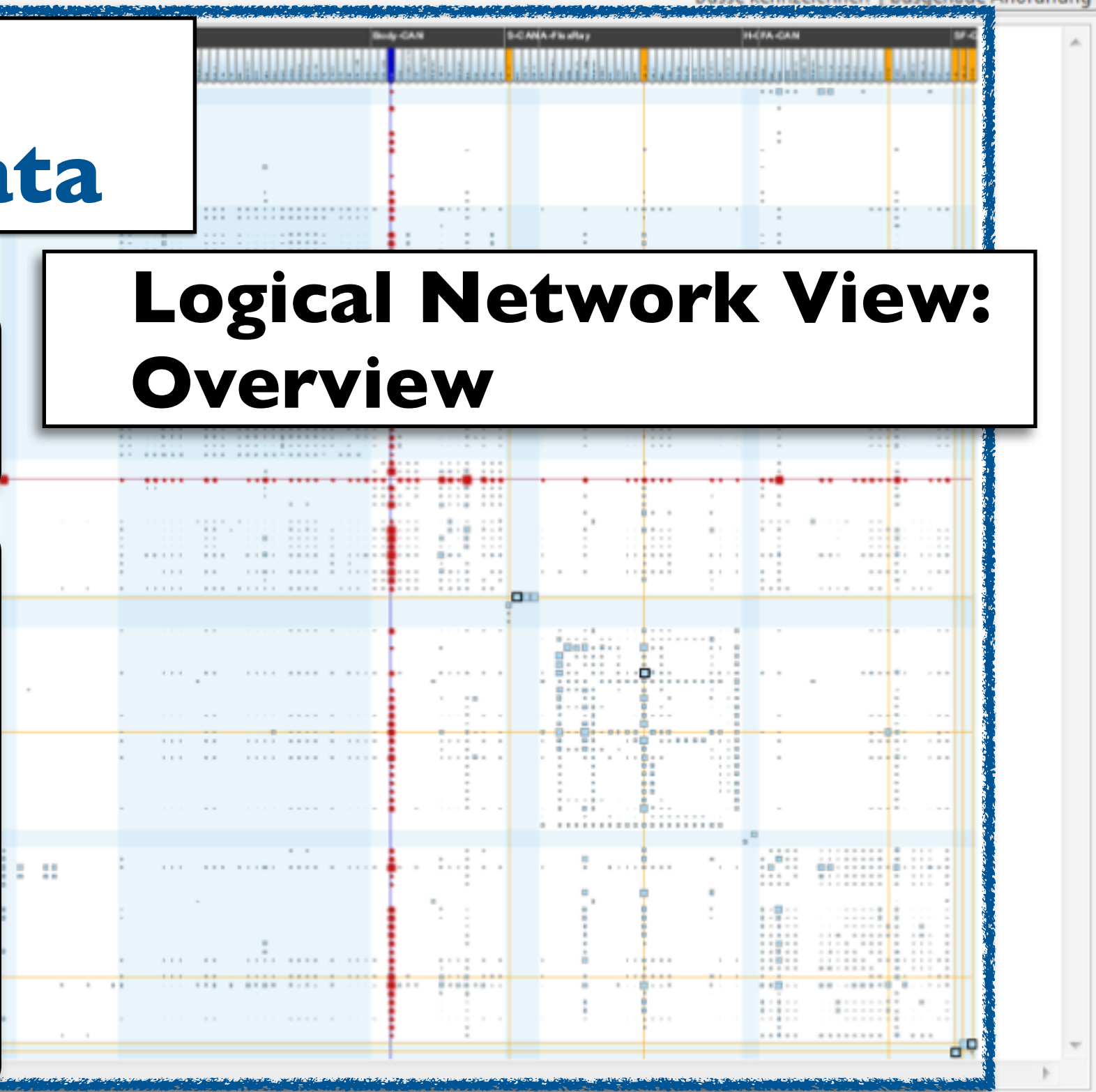
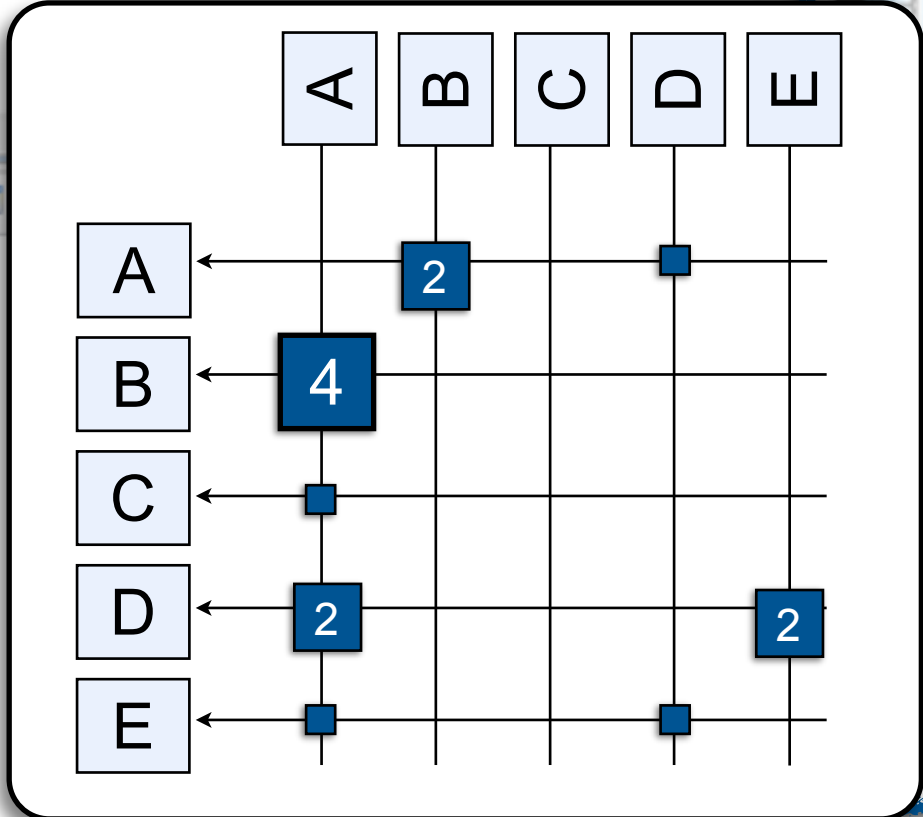
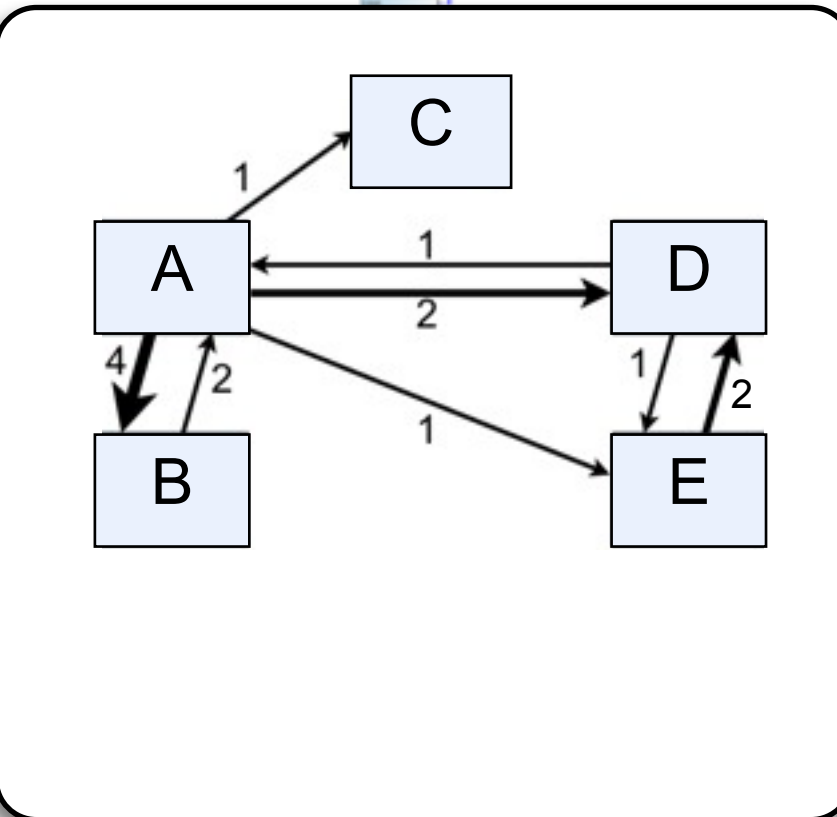


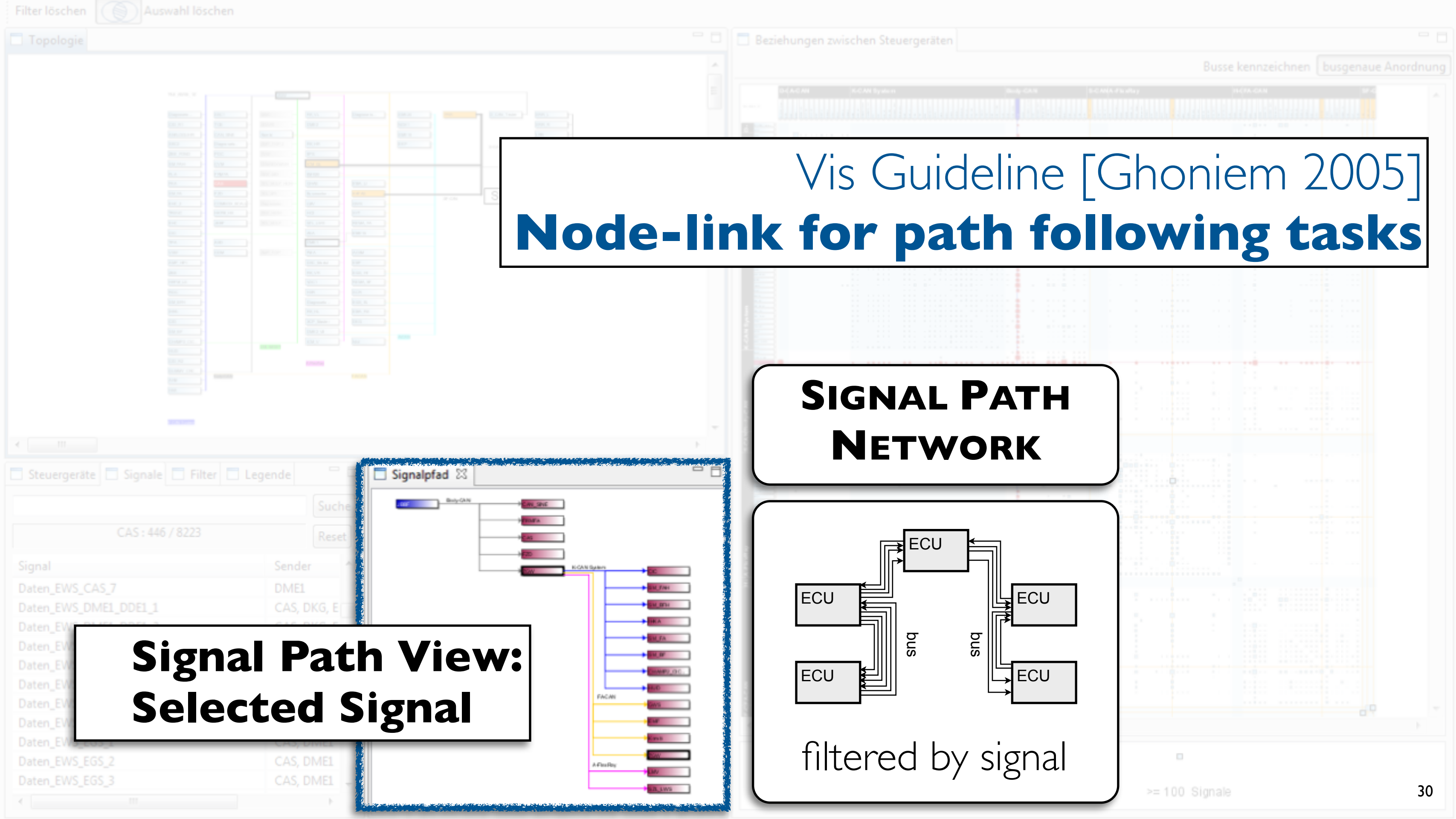
Vis Guideline [Ghoniem 2005] Matrix for dense network data

Logical Network View: Overview

SIGNAL COUNT NETWORK

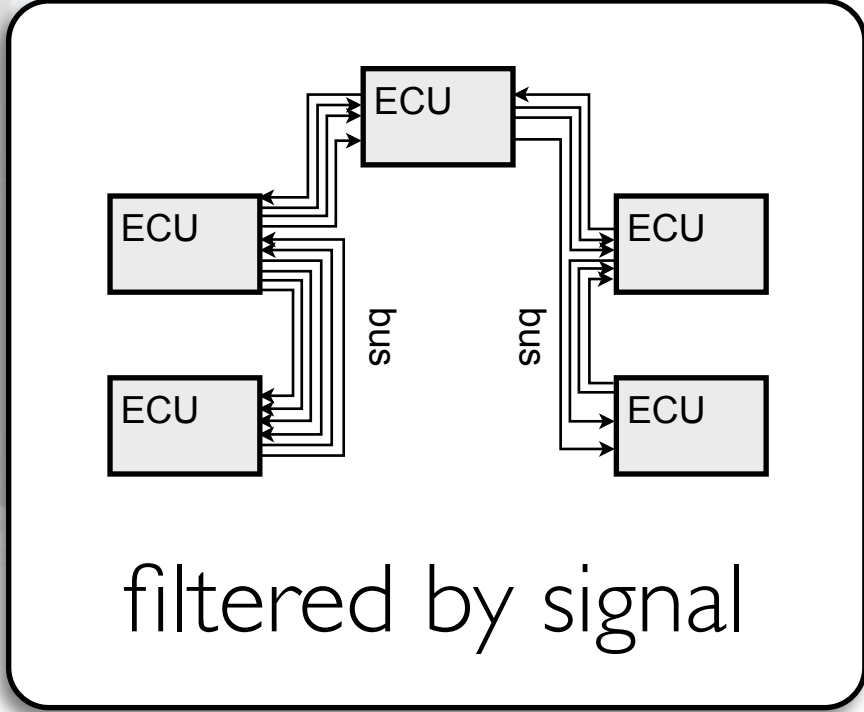
visual encoding: size-coded matrix





Vis Guideline [Ghoniem 2005]
Node-link for path following tasks

SIGNAL PATH NETWORK



**Signal Path View:
Selected Signal**

Signal	Sender
Daten_EWS_CAS_7	DME1
Daten_EWS_DME1_DDE1_1	CAS, DKG, E
Daten_EWS_DME1_DDE1_2	CAS, DME1
Daten_EWS_DME1_DDE1_3	CAS, DME1
Daten_EWS_DME1_DDE1_4	CAS, DME1
Daten_EWS_DME1_DDE1_5	CAS, DME1
Daten_EWS_DME1_DDE1_6	CAS, DME1
Daten_EWS_DME1_DDE1_7	CAS, DME1
Daten_EWS_DME1_DDE1_8	CAS, DME1
Daten_EWS_DME1_DDE1_9	CAS, DME1
Daten_EWS_DME1_DDE1_10	CAS, DME1
Daten_EWS_DME1_DDE1_11	CAS, DME1
Daten_EWS_DME1_DDE1_12	CAS, DME1
Daten_EWS_DME1_DDE1_13	CAS, DME1
Daten_EWS_DME1_DDE1_14	CAS, DME1
Daten_EWS_DME1_DDE1_15	CAS, DME1
Daten_EWS_DME1_DDE1_16	CAS, DME1
Daten_EWS_DME1_DDE1_17	CAS, DME1
Daten_EWS_DME1_DDE1_18	CAS, DME1
Daten_EWS_DME1_DDE1_19	CAS, DME1
Daten_EWS_DME1_DDE1_20	CAS, DME1
Daten_EWS_DME1_DDE1_21	CAS, DME1
Daten_EWS_DME1_DDE1_22	CAS, DME1
Daten_EWS_DME1_DDE1_23	CAS, DME1
Daten_EWS_DME1_DDE1_24	CAS, DME1
Daten_EWS_DME1_DDE1_25	CAS, DME1
Daten_EWS_DME1_DDE1_26	CAS, DME1
Daten_EWS_DME1_DDE1_27	CAS, DME1
Daten_EWS_DME1_DDE1_28	CAS, DME1
Daten_EWS_DME1_DDE1_29	CAS, DME1
Daten_EWS_DME1_DDE1_30	CAS, DME1
Daten_EWS_DME1_DDE1_31	CAS, DME1
Daten_EWS_DME1_DDE1_32	CAS, DME1
Daten_EWS_DME1_DDE1_33	CAS, DME1
Daten_EWS_DME1_DDE1_34	CAS, DME1
Daten_EWS_DME1_DDE1_35	CAS, DME1
Daten_EWS_DME1_DDE1_36	CAS, DME1
Daten_EWS_DME1_DDE1_37	CAS, DME1
Daten_EWS_DME1_DDE1_38	CAS, DME1
Daten_EWS_DME1_DDE1_39	CAS, DME1
Daten_EWS_DME1_DDE1_40	CAS, DME1
Daten_EWS_DME1_DDE1_41	CAS, DME1
Daten_EWS_DME1_DDE1_42	CAS, DME1
Daten_EWS_DME1_DDE1_43	CAS, DME1
Daten_EWS_DME1_DDE1_44	CAS, DME1
Daten_EWS_DME1_DDE1_45	CAS, DME1
Daten_EWS_DME1_DDE1_46	CAS, DME1
Daten_EWS_DME1_DDE1_47	CAS, DME1
Daten_EWS_DME1_DDE1_48	CAS, DME1
Daten_EWS_DME1_DDE1_49	CAS, DME1
Daten_EWS_DME1_DDE1_50	CAS, DME1
Daten_EWS_DME1_DDE1_51	CAS, DME1
Daten_EWS_DME1_DDE1_52	CAS, DME1
Daten_EWS_DME1_DDE1_53	CAS, DME1
Daten_EWS_DME1_DDE1_54	CAS, DME1
Daten_EWS_DME1_DDE1_55	CAS, DME1
Daten_EWS_DME1_DDE1_56	CAS, DME1
Daten_EWS_DME1_DDE1_57	CAS, DME1
Daten_EWS_DME1_DDE1_58	CAS, DME1
Daten_EWS_DME1_DDE1_59	CAS, DME1
Daten_EWS_DME1_DDE1_60	CAS, DME1
Daten_EWS_DME1_DDE1_61	CAS, DME1
Daten_EWS_DME1_DDE1_62	CAS, DME1
Daten_EWS_DME1_DDE1_63	CAS, DME1
Daten_EWS_DME1_DDE1_64	CAS, DME1
Daten_EWS_DME1_DDE1_65	CAS, DME1
Daten_EWS_DME1_DDE1_66	CAS, DME1
Daten_EWS_DME1_DDE1_67	CAS, DME1
Daten_EWS_DME1_DDE1_68	CAS, DME1
Daten_EWS_DME1_DDE1_69	CAS, DME1
Daten_EWS_DME1_DDE1_70	CAS, DME1
Daten_EWS_DME1_DDE1_71	CAS, DME1
Daten_EWS_DME1_DDE1_72	CAS, DME1
Daten_EWS_DME1_DDE1_73	CAS, DME1
Daten_EWS_DME1_DDE1_74	CAS, DME1
Daten_EWS_DME1_DDE1_75	CAS, DME1
Daten_EWS_DME1_DDE1_76	CAS, DME1
Daten_EWS_DME1_DDE1_77	CAS, DME1
Daten_EWS_DME1_DDE1_78	CAS, DME1
Daten_EWS_DME1_DDE1_79	CAS, DME1
Daten_EWS_DME1_DDE1_80	CAS, DME1
Daten_EWS_DME1_DDE1_81	CAS, DME1
Daten_EWS_DME1_DDE1_82	CAS, DME1
Daten_EWS_DME1_DDE1_83	CAS, DME1
Daten_EWS_DME1_DDE1_84	CAS, DME1
Daten_EWS_DME1_DDE1_85	CAS, DME1
Daten_EWS_DME1_DDE1_86	CAS, DME1
Daten_EWS_DME1_DDE1_87	CAS, DME1
Daten_EWS_DME1_DDE1_88	CAS, DME1
Daten_EWS_DME1_DDE1_89	CAS, DME1
Daten_EWS_DME1_DDE1_90	CAS, DME1
Daten_EWS_DME1_DDE1_91	CAS, DME1
Daten_EWS_DME1_DDE1_92	CAS, DME1
Daten_EWS_DME1_DDE1_93	CAS, DME1
Daten_EWS_DME1_DDE1_94	CAS, DME1
Daten_EWS_DME1_DDE1_95	CAS, DME1
Daten_EWS_DME1_DDE1_96	CAS, DME1
Daten_EWS_DME1_DDE1_97	CAS, DME1
Daten_EWS_DME1_DDE1_98	CAS, DME1
Daten_EWS_DME1_DDE1_99	CAS, DME1
Daten_EWS_DME1_DDE1_100	CAS, DME1

INTERACTION IDIOM: Cross-Network Relations

Busse kennzeichnen busgenaue Anordnung

Steuergeräte Signale Filter Legende

Suche

CAS: 446 / 8223

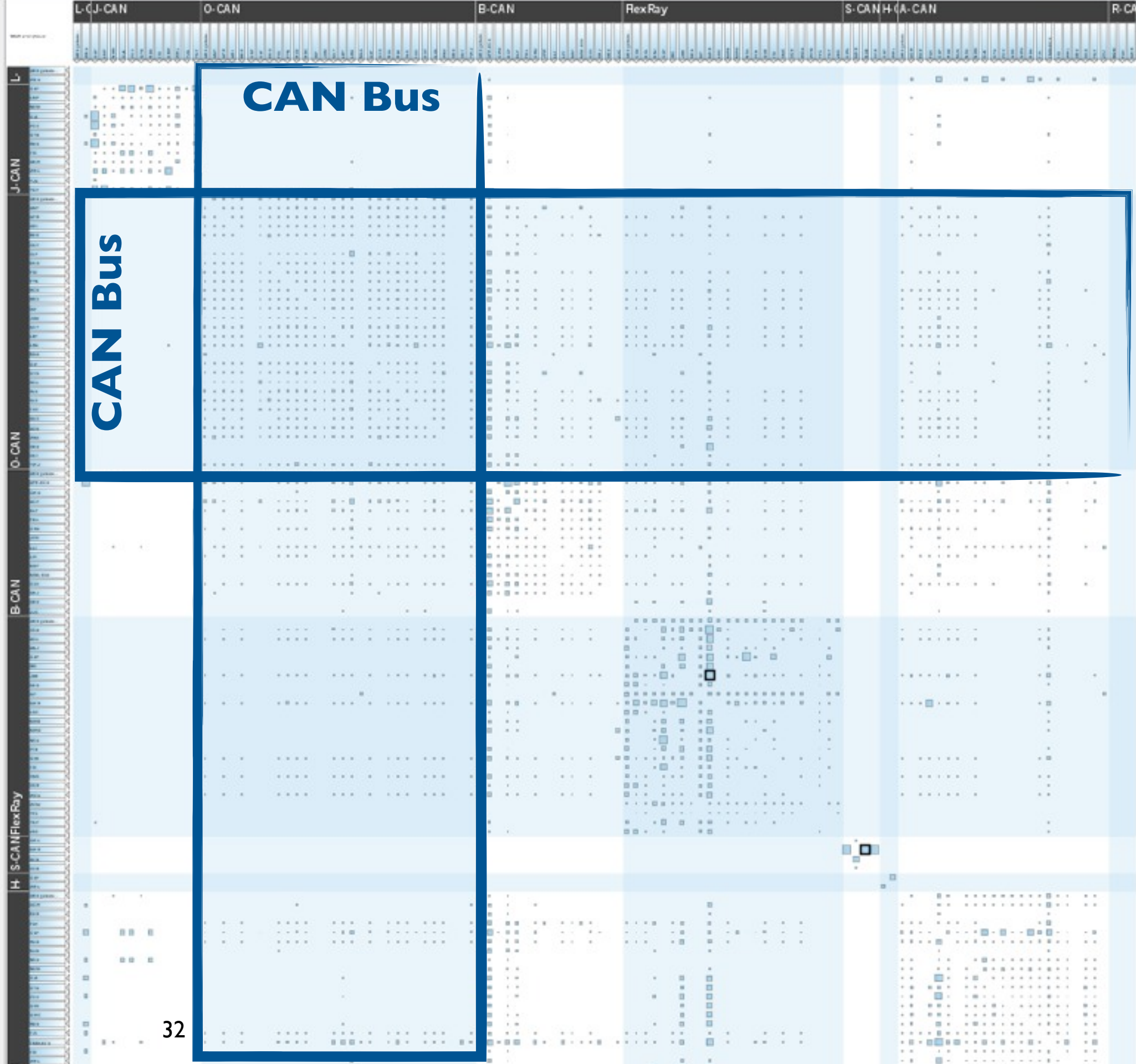
Reset

Signal	Sender
Daten_EWS_CAS_7	DME1
Daten_EWS_DME1_DDE1_1	CAS, DKG, E
Daten_EWS_DME1_DDE1_2	CAS, DKG, E
Daten_EWS_DME1_DDE1_3	CAS, DKG, E
Daten_EWS_DME1_DDE1_4	CAS, DKG, E
Daten_EWS_DME1_DDE1_5	CAS, DKG, E
Daten_EWS_DME1_DDE1_6	CAS, DKG, E
Daten_EWS_DME1_DDE1_7	CAS, DKG, E
Daten_EWS_EGS_1	CAS, DME1
Daten_EWS_EGS_2	CAS, DME1
Daten_EWS_EGS_3	CAS, DME1

1 Signal 10 Signale 30 Signale 50 Signale >= 100 Signale

VIDEO

INTERESTS Bus communication patterns

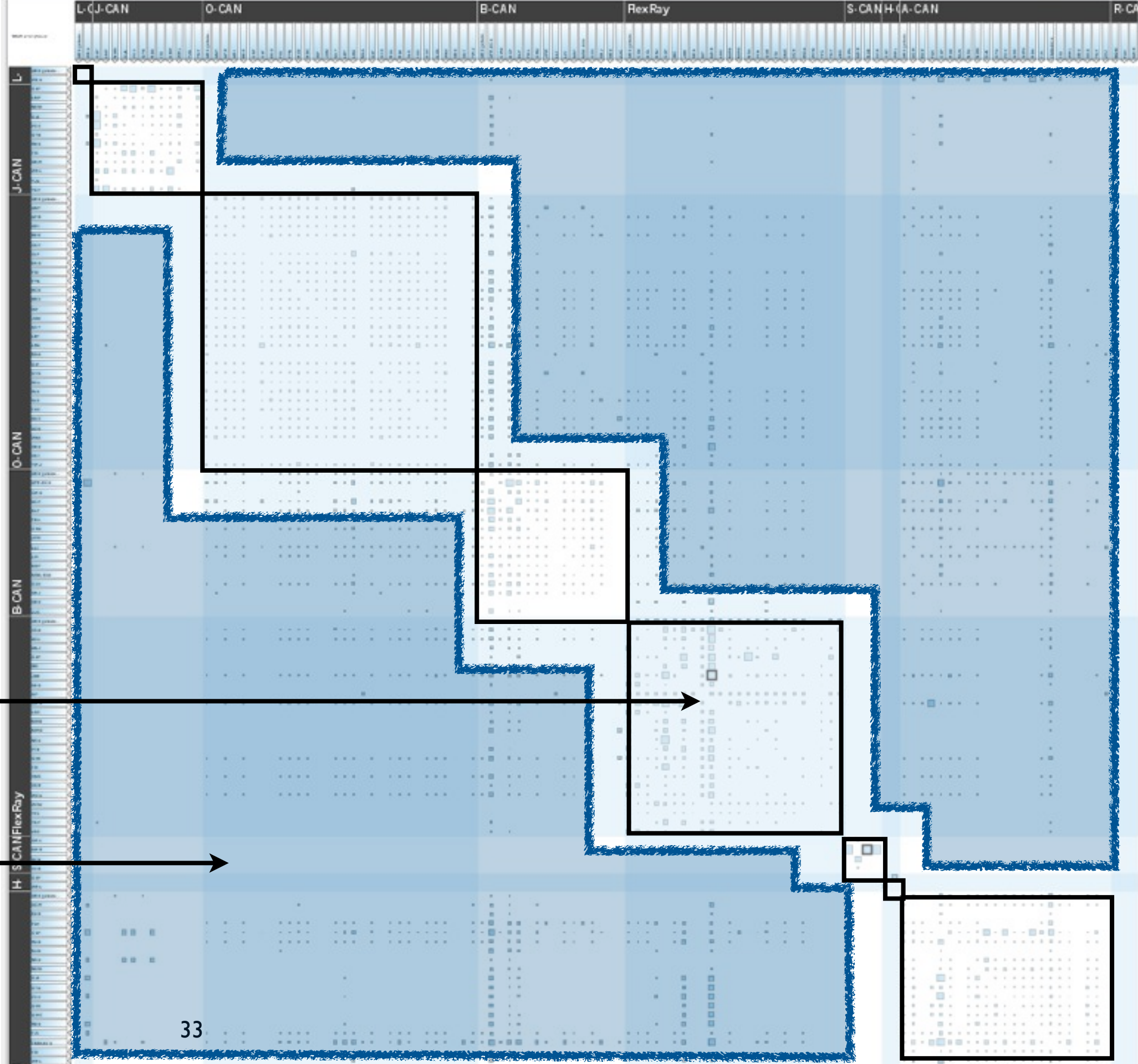


INTERESTS

Bus communication patterns

Within-bus

Between-bus



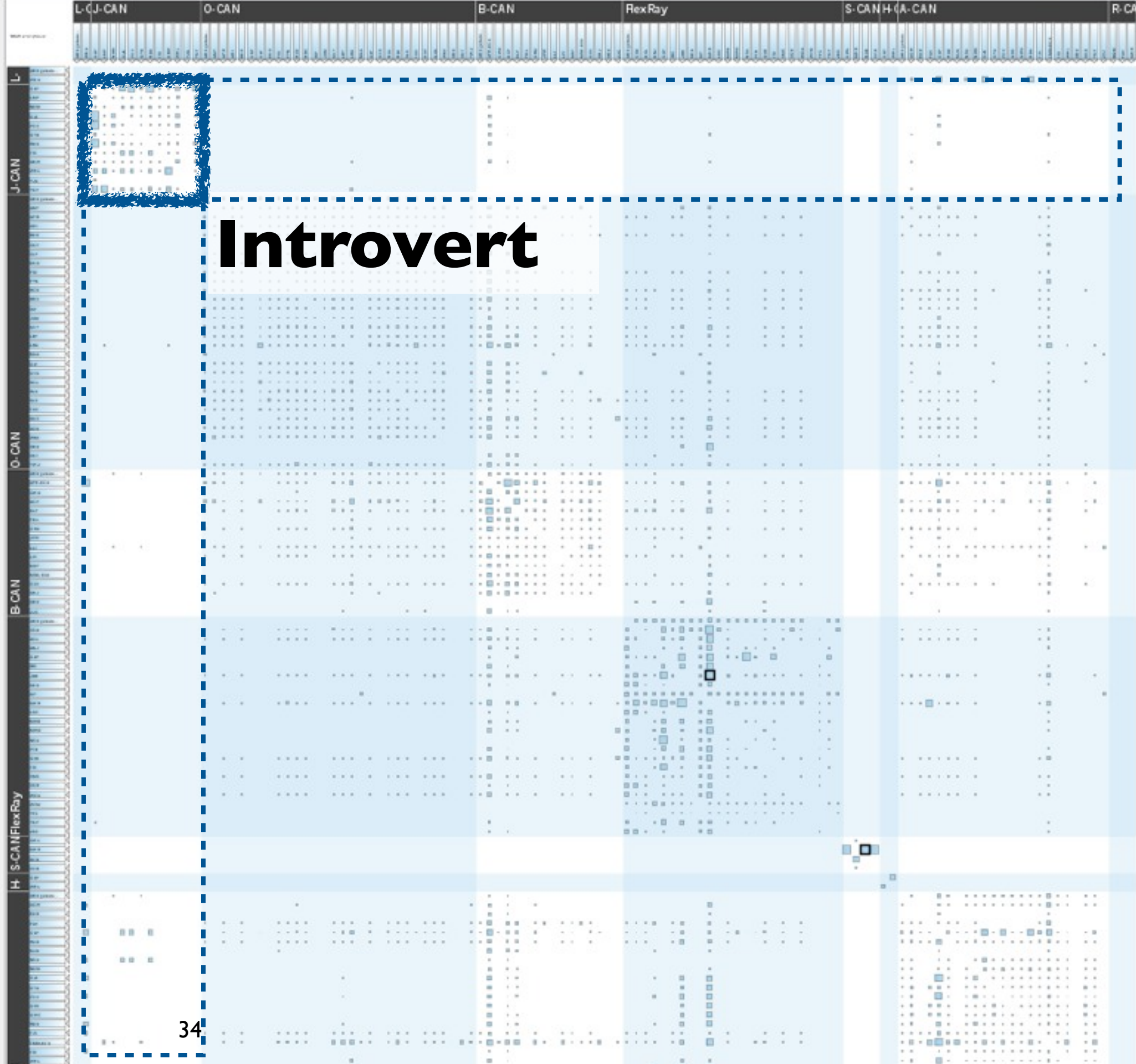
INTERESTS

Bus communication patterns

introvert

vs.

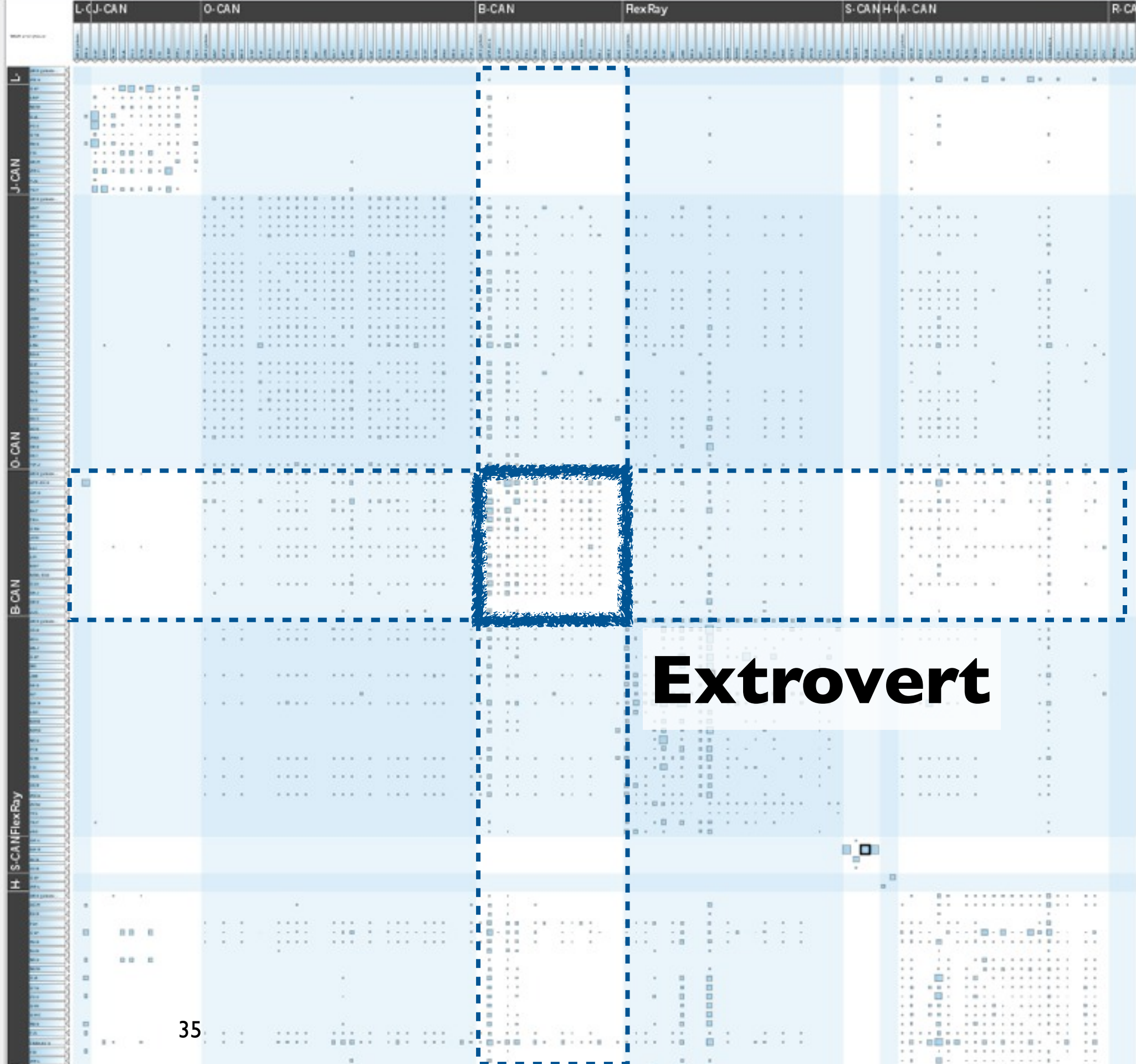
extrovert



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

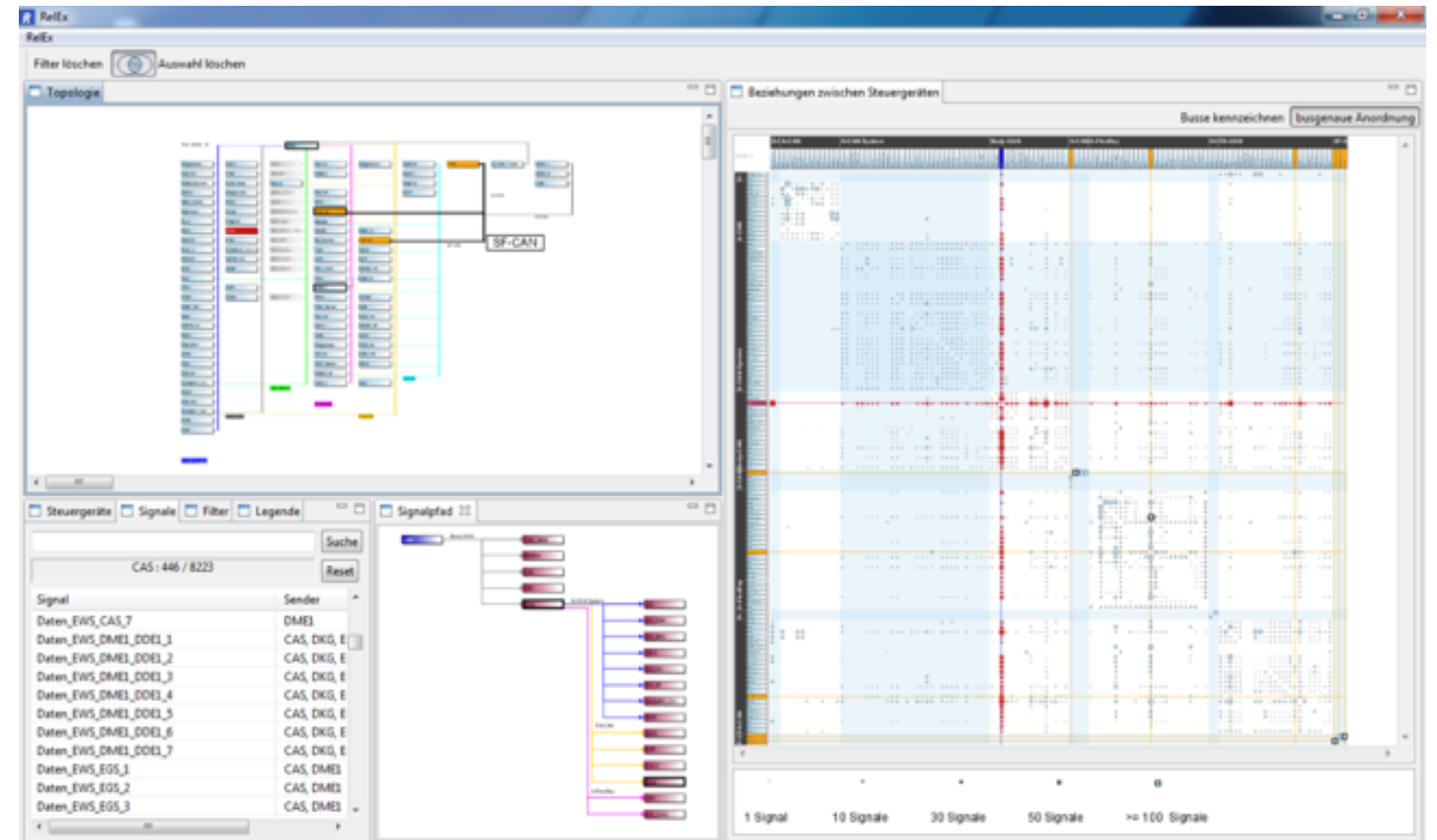


Phase 2: Design, implement, deploy

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

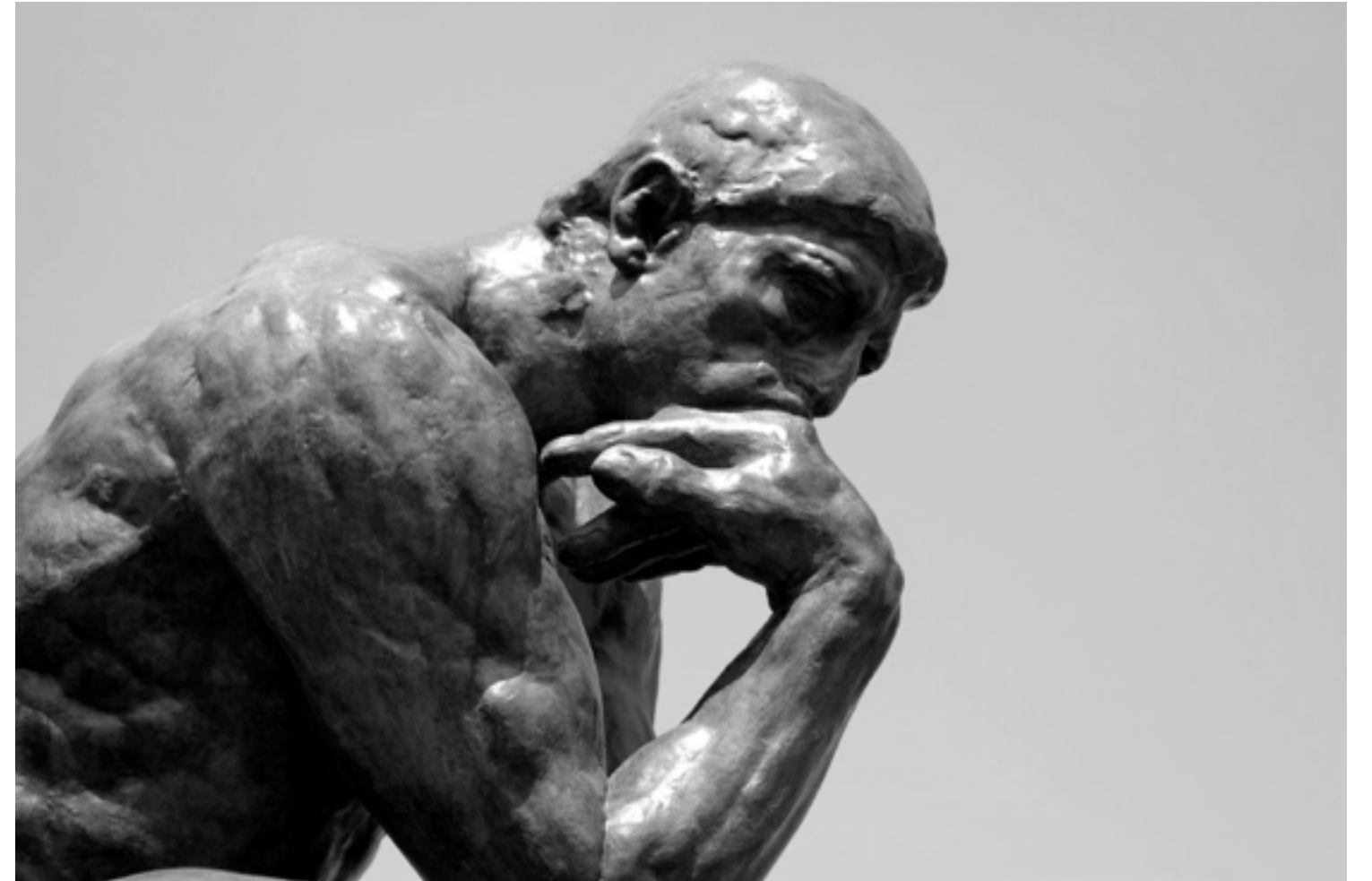


Phase 4: Reflect and write

3 months



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

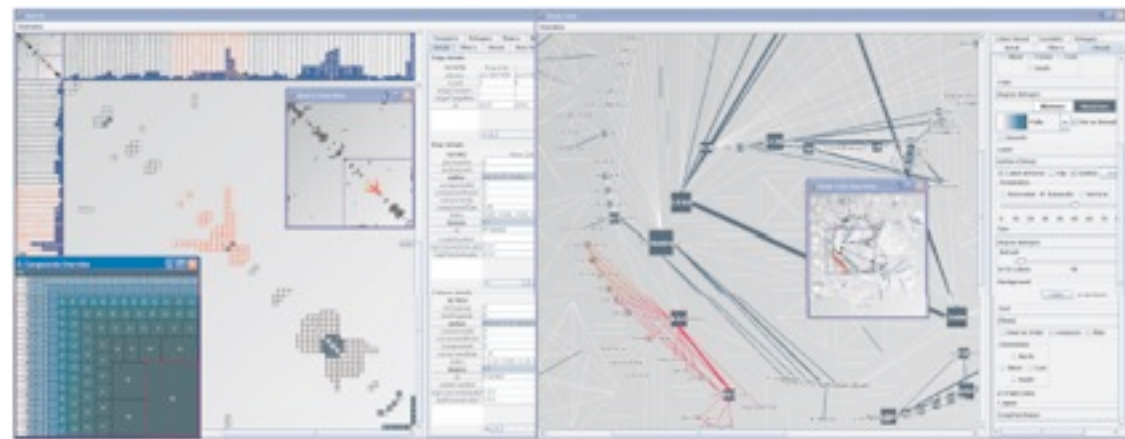


Abstraction Innovation

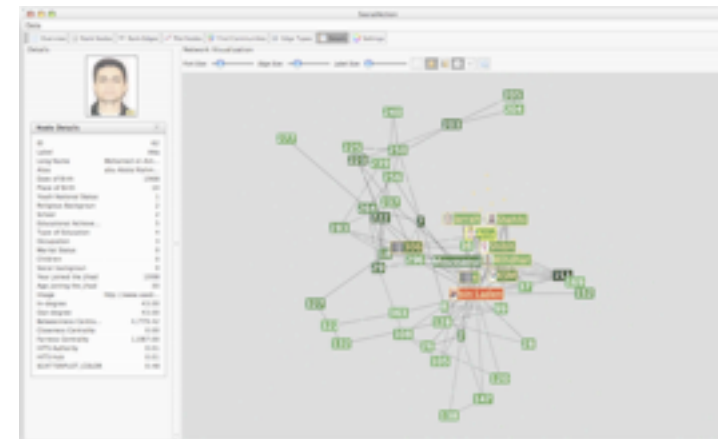
Previous Work

Focus on social network analysis

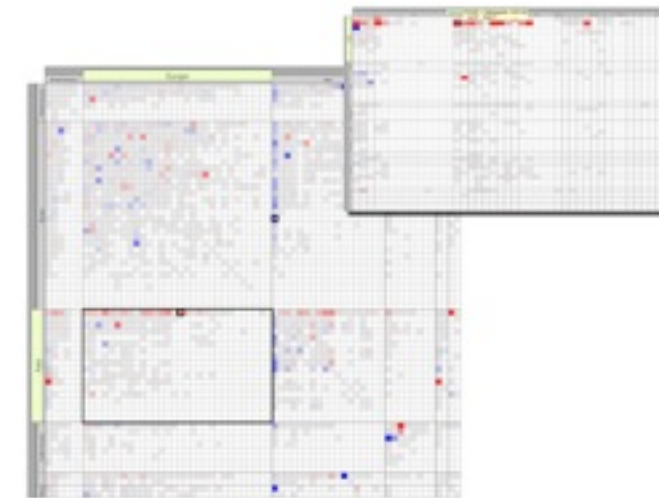
- radically different task and data abstractions



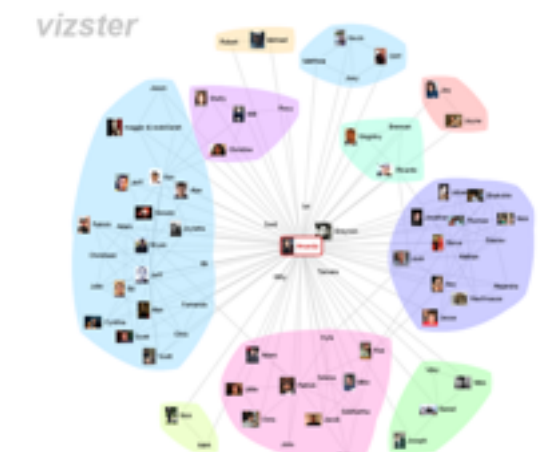
MatrixExplorer



SocialAction



Honeycomb



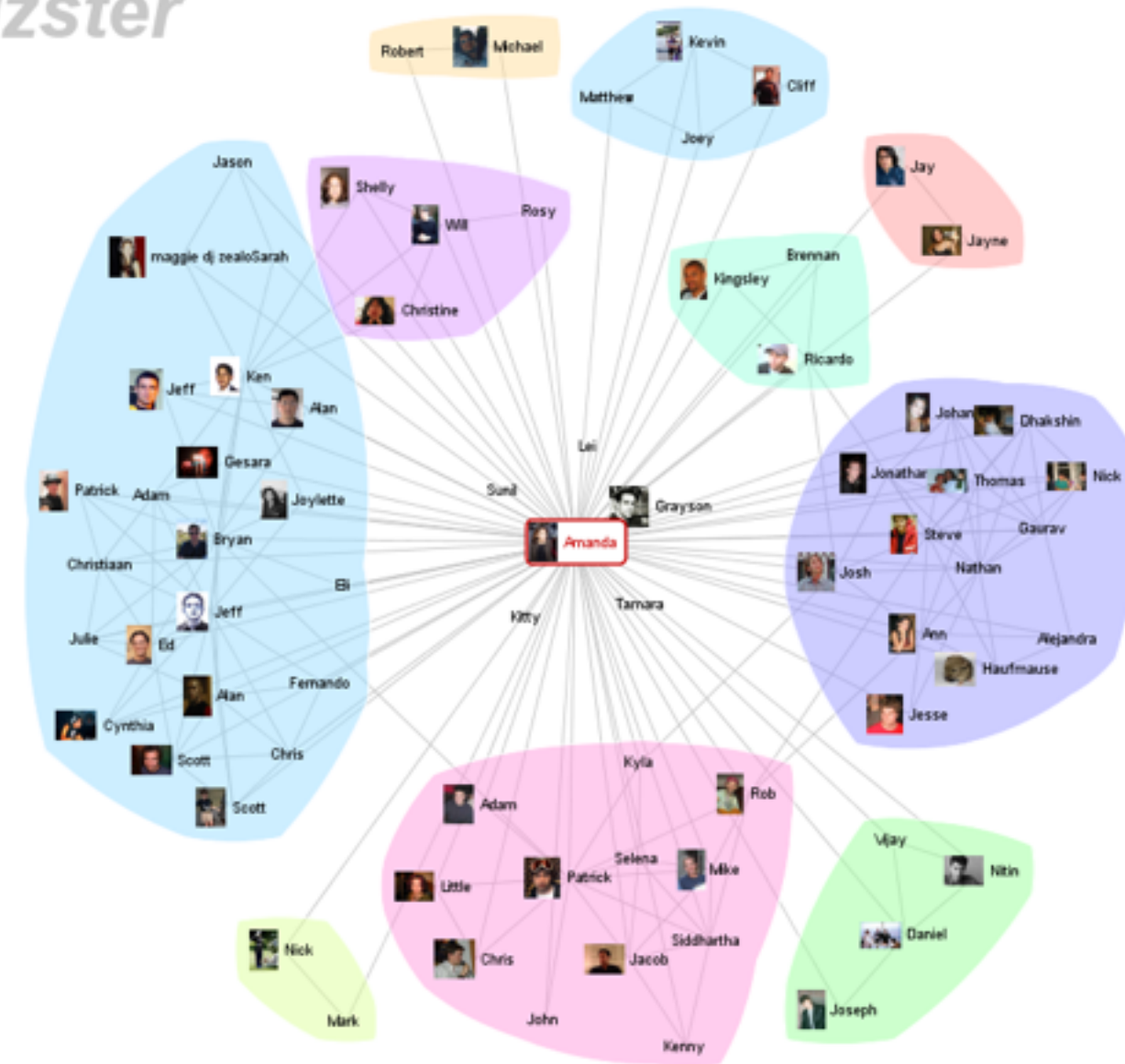
vizster

Task Abstraction

Social Network Analysis Domain

- find clusters

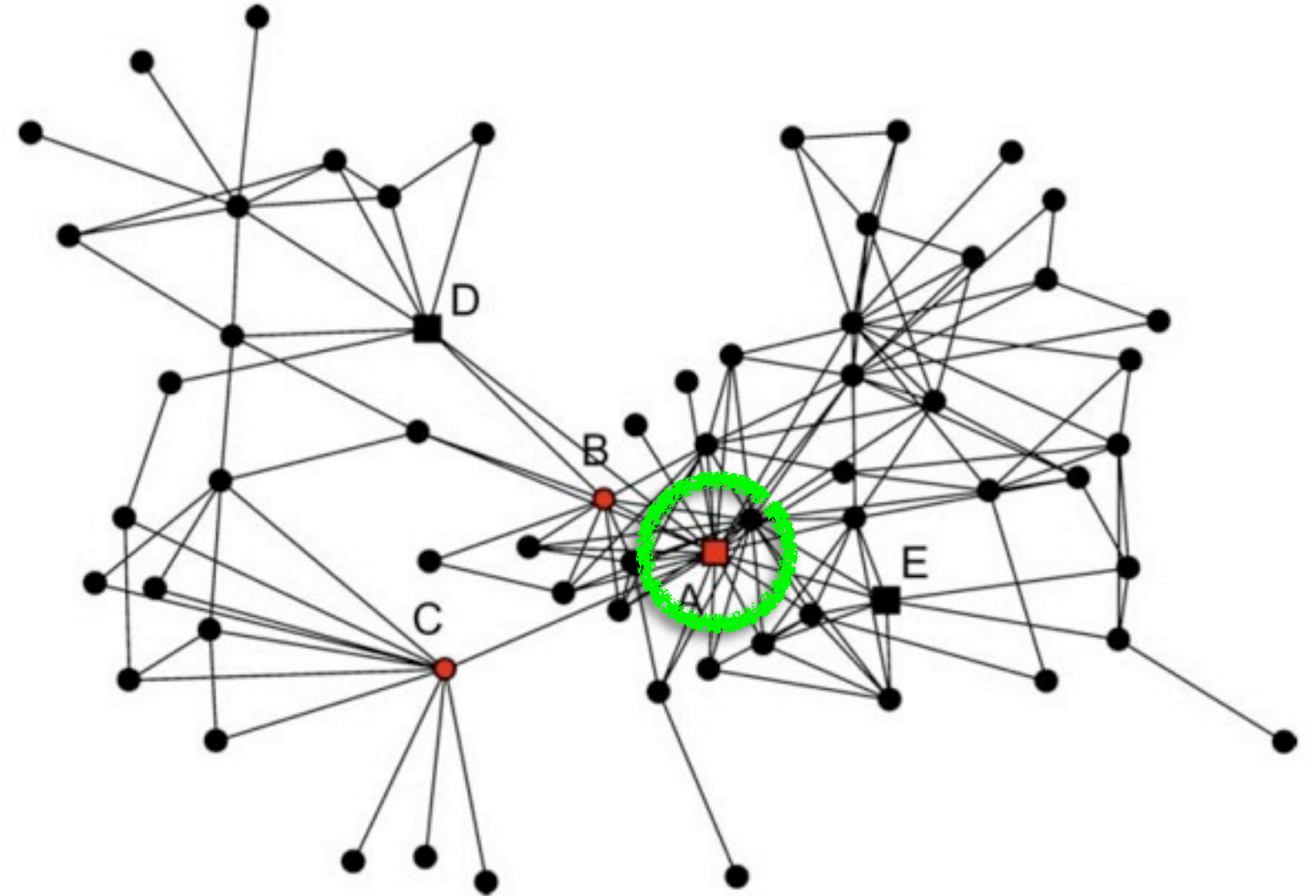
vizster



Task Abstraction

Social Network Analysis Domain

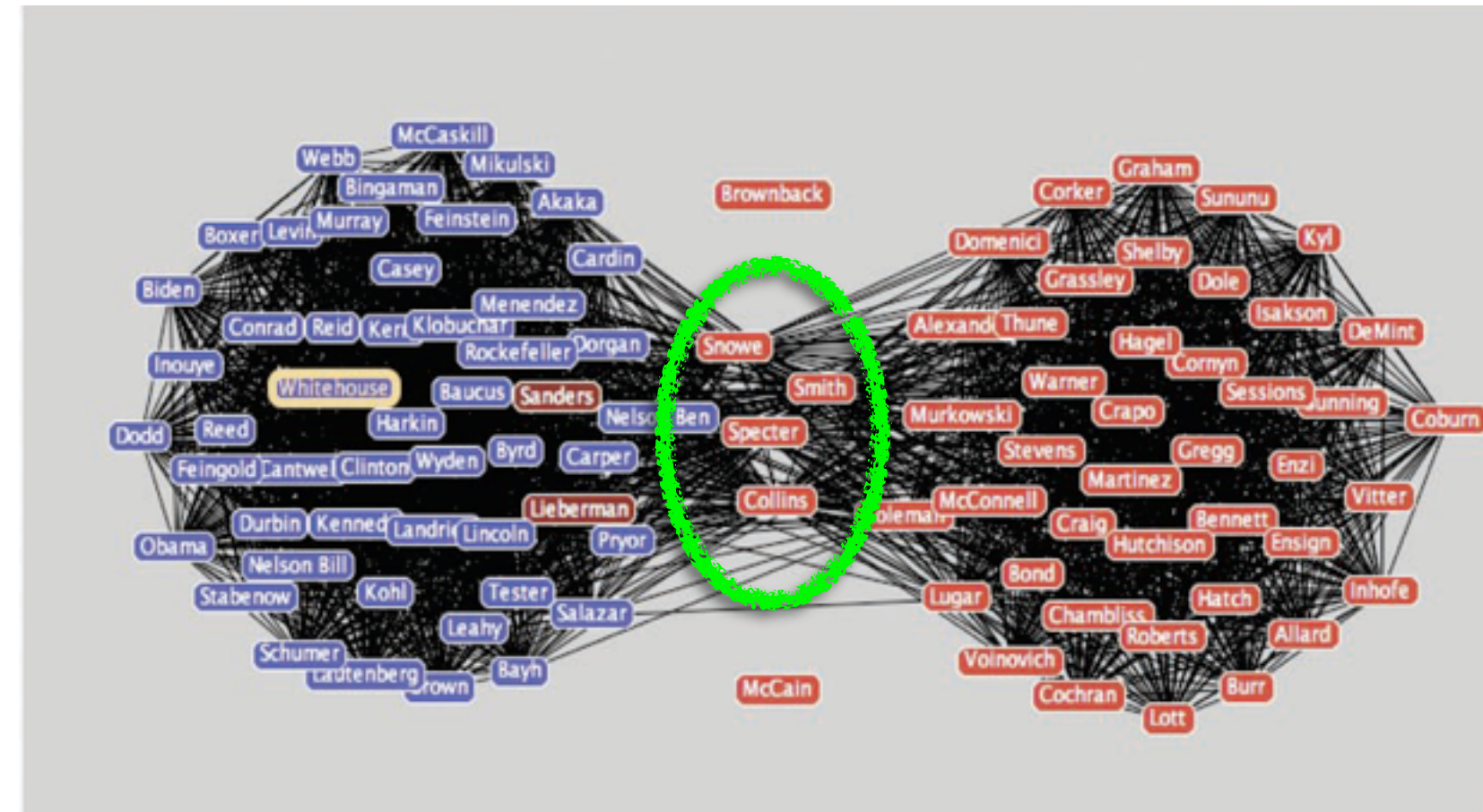
- find clusters
- find high-degree nodes



Task Abstraction

Social Network Analysis Domain

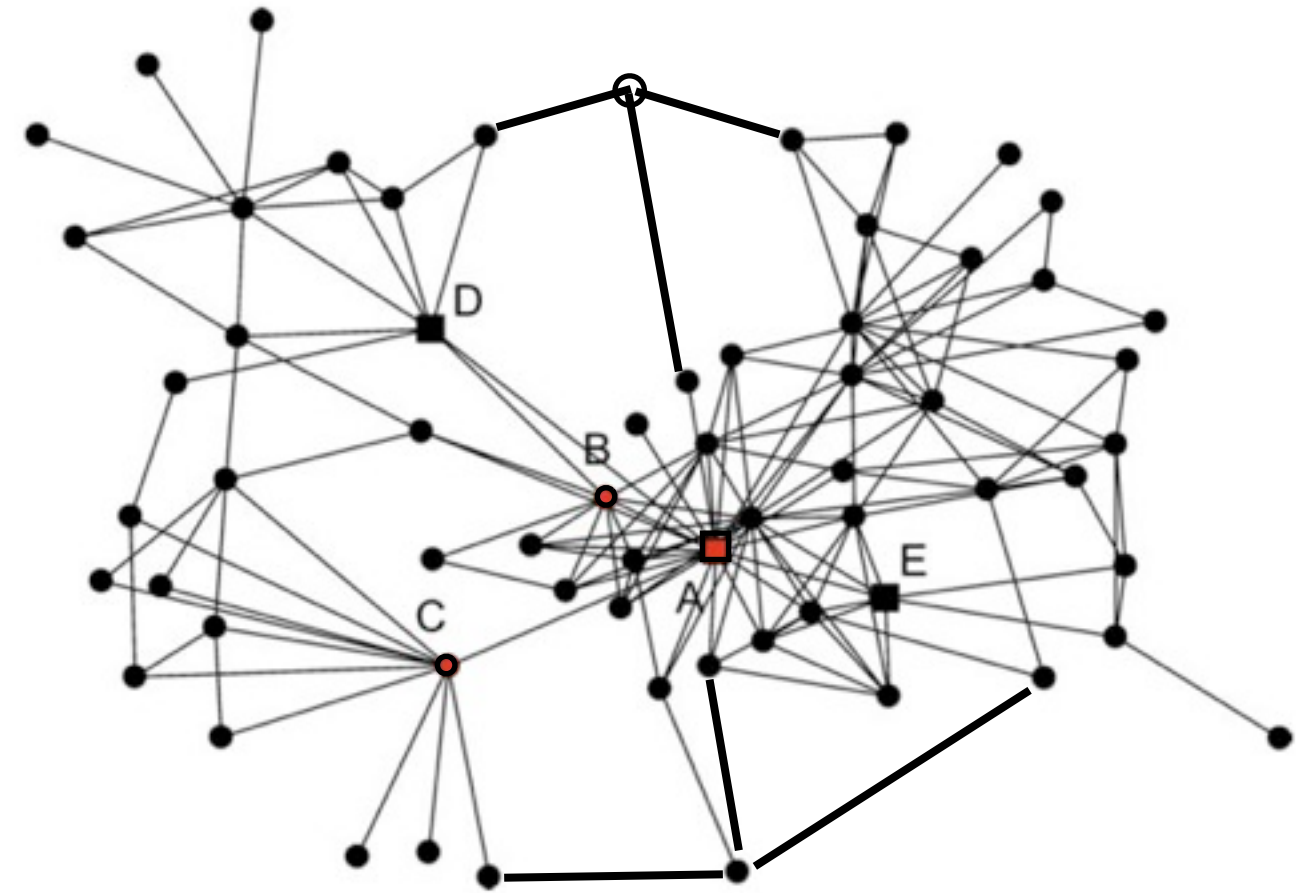
- find clusters
- find high-degree nodes
- find bridge nodes



Task Abstraction

Social Network Analysis Domain

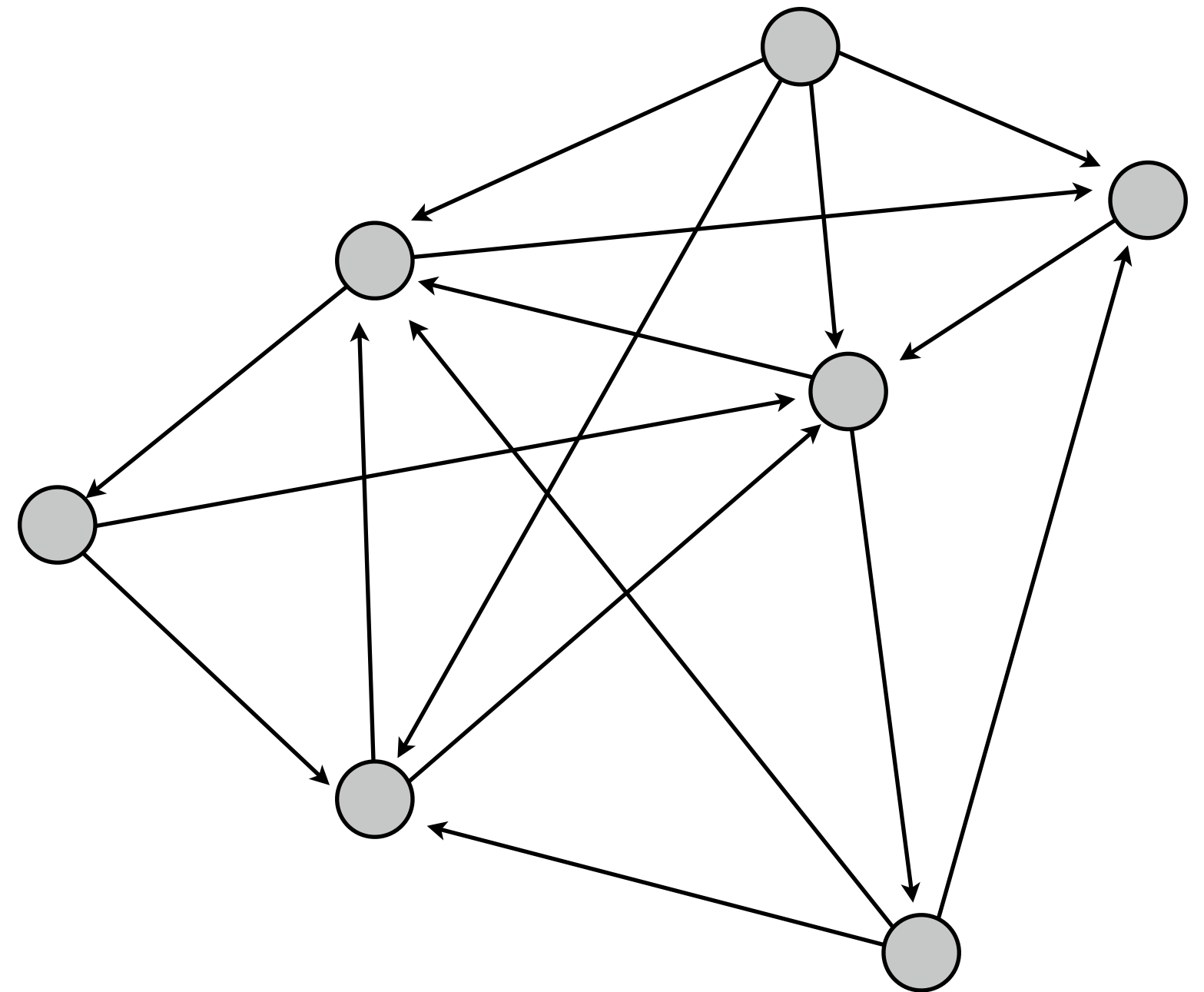
- find clusters
- find high-degree nodes
- find bridge nodes
- understand temporal dynamics
 - passively notice changes



Data Abstraction

Social Network Analysis Domain

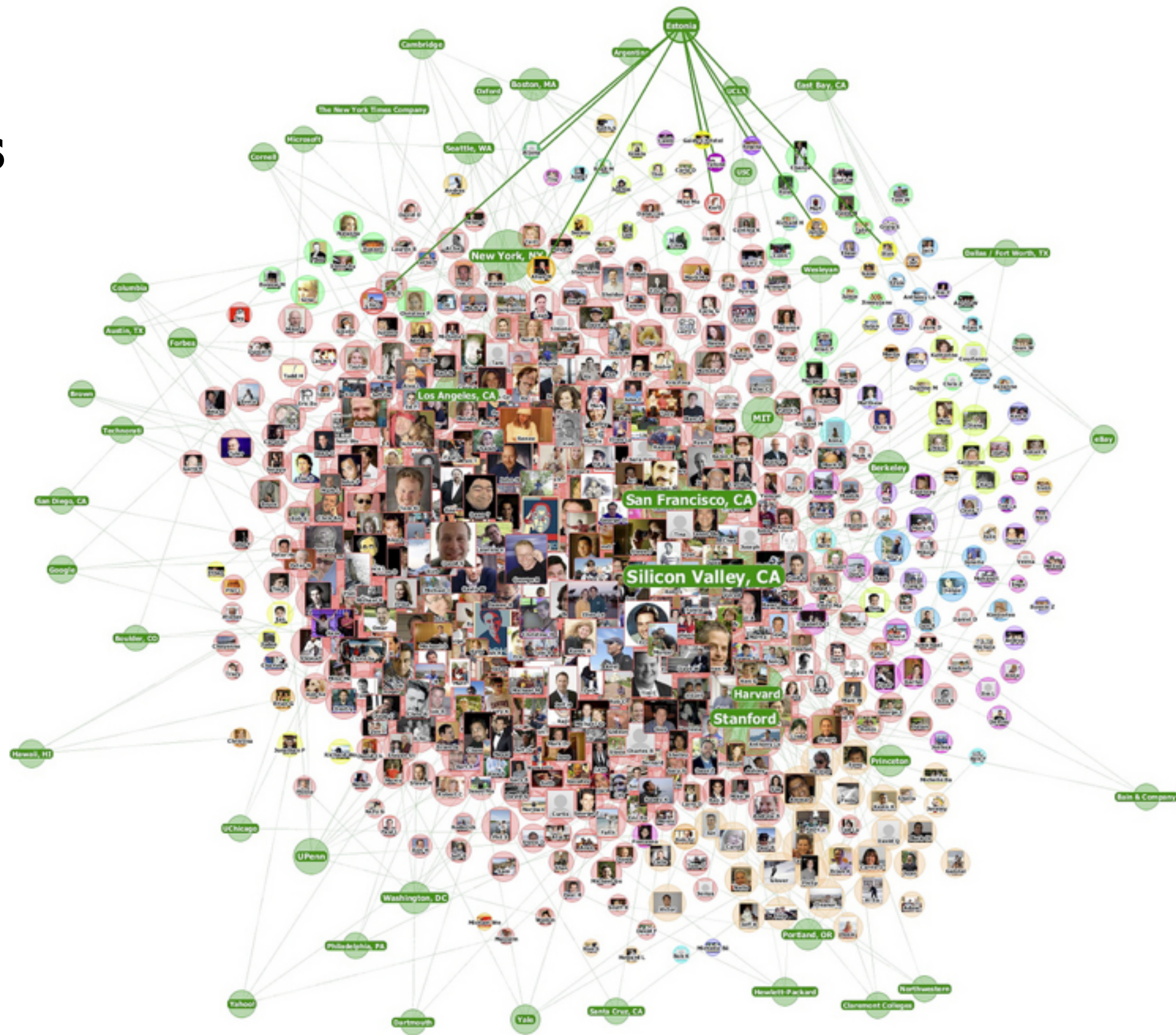
- single graph



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 

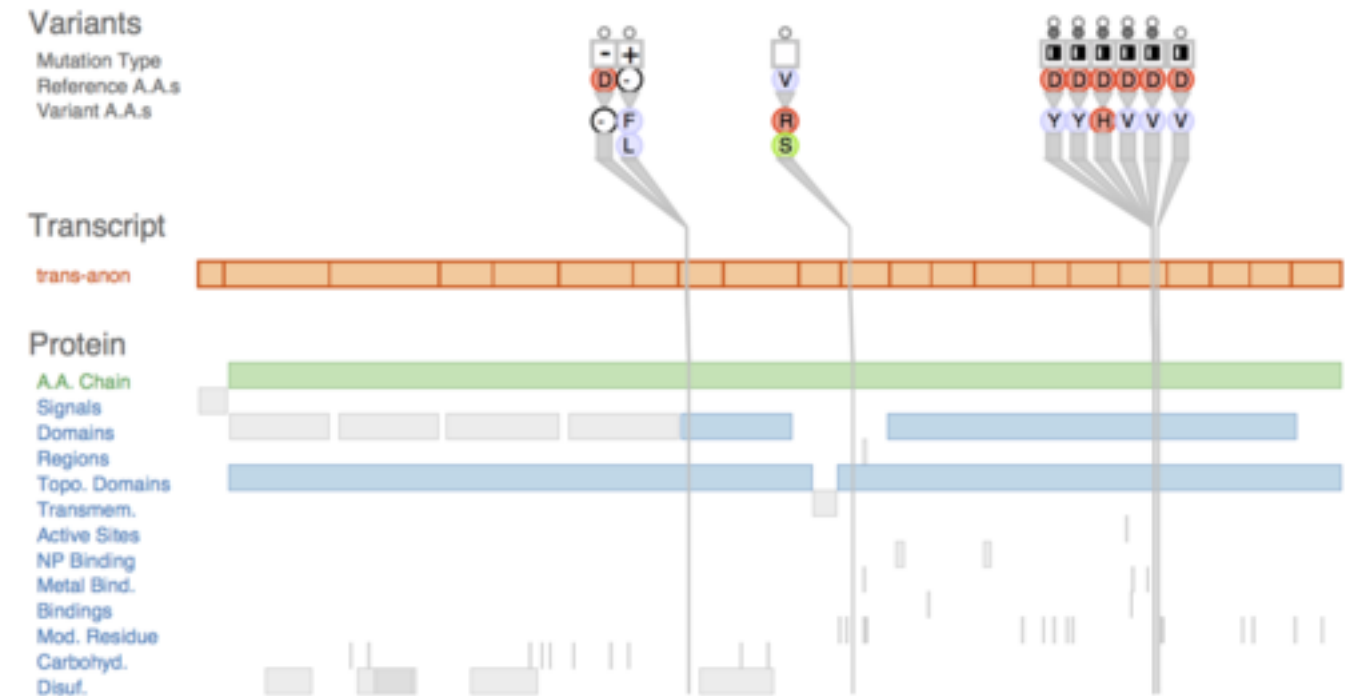
Variant View

Visualizing Sequence Variants in their Gene Context

joint work with:

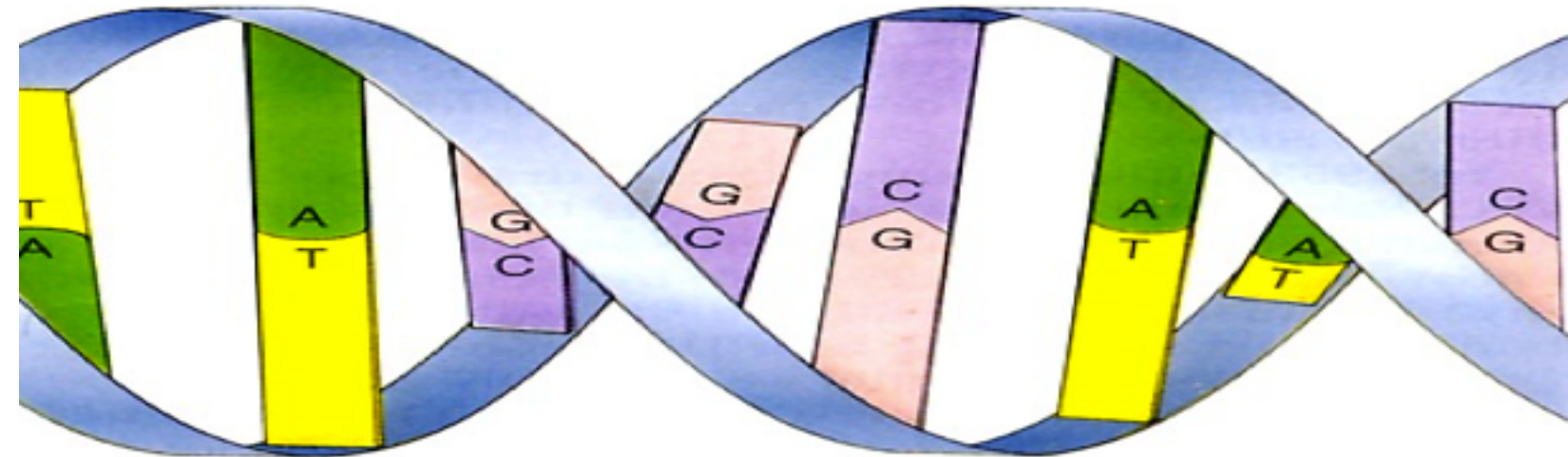
Joel Ferstay, Cydney Nielsen

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



Sequence Variant Definition

- Sequence variants
 - Difference between reference and given genome



Reference Genome DNA: ATA TGA TCA ACA CTT

Sample 1 Genome DNA: ATA TG**G** TCA **ATA** CTT

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

Harmful?

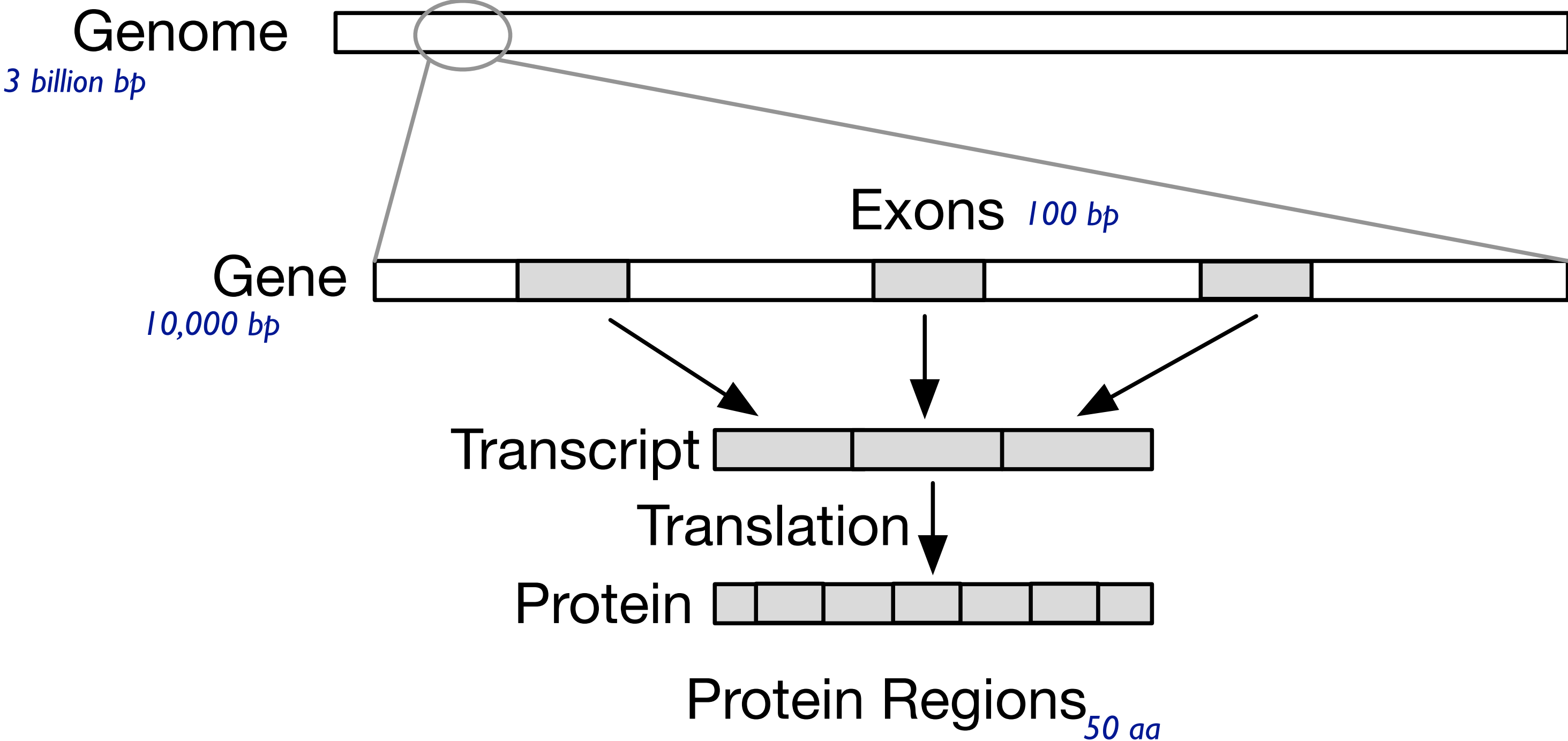
Harmless?

Cancer Research

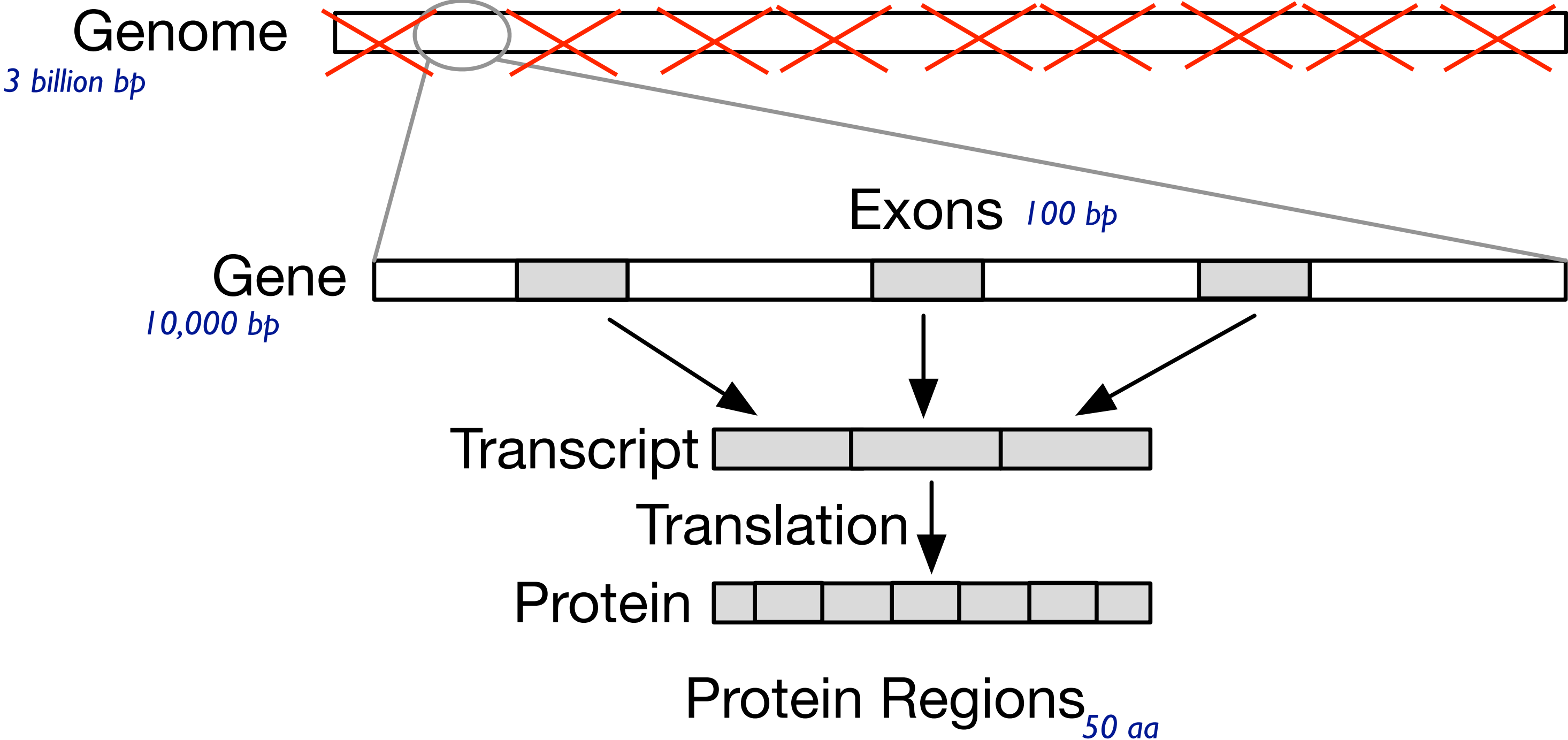
- collaboration with analysts at BC Genome Sciences Center
 - studying genetic basis of leukemia
- driving task
 - discover new candidate genes with harmful variants
- two big questions
 - what to show
 - data abstraction
 - challenge: enormous range of scales in the data
 - how to show it
 - visual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

Abstractions

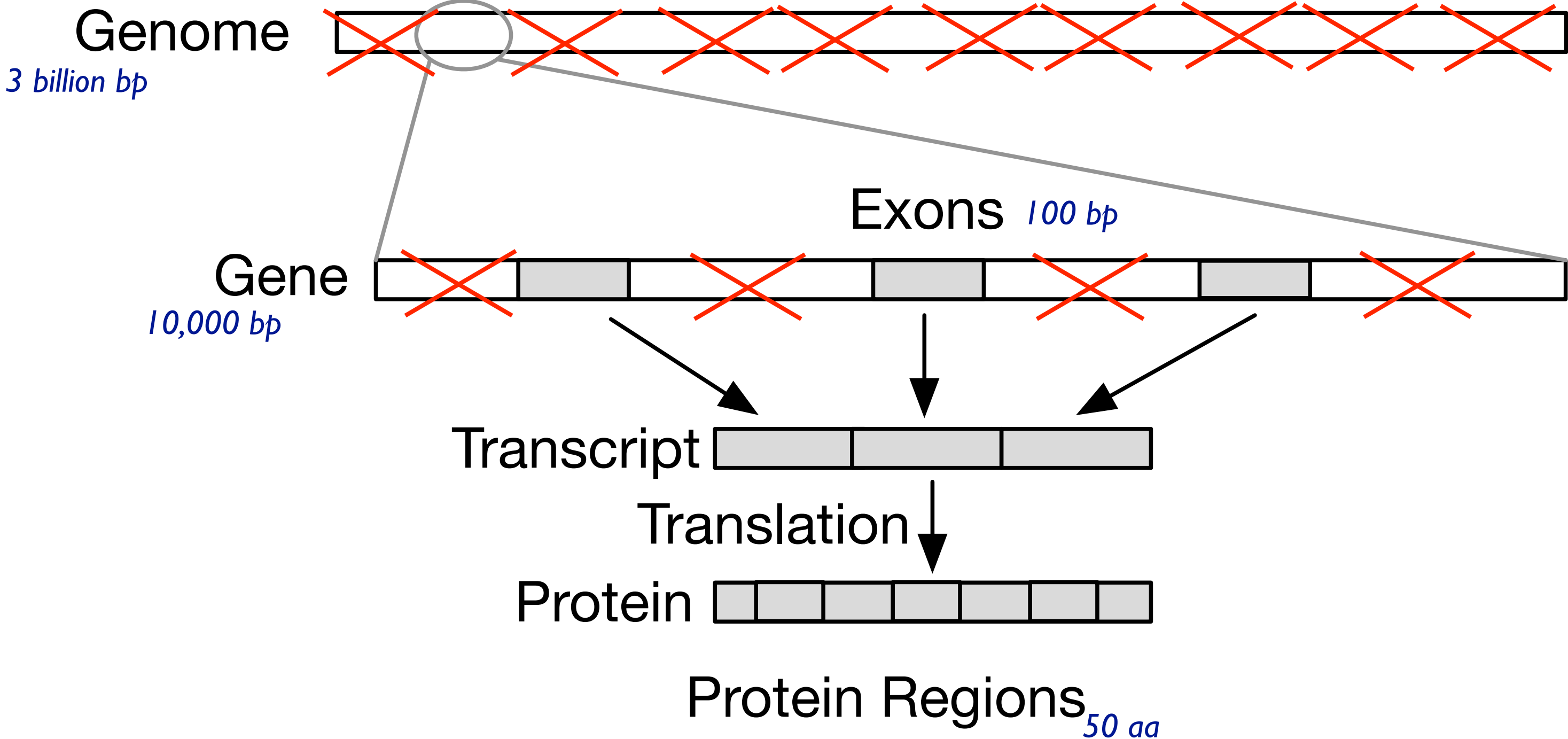
Data: Filtering to relevant biological levels and scales



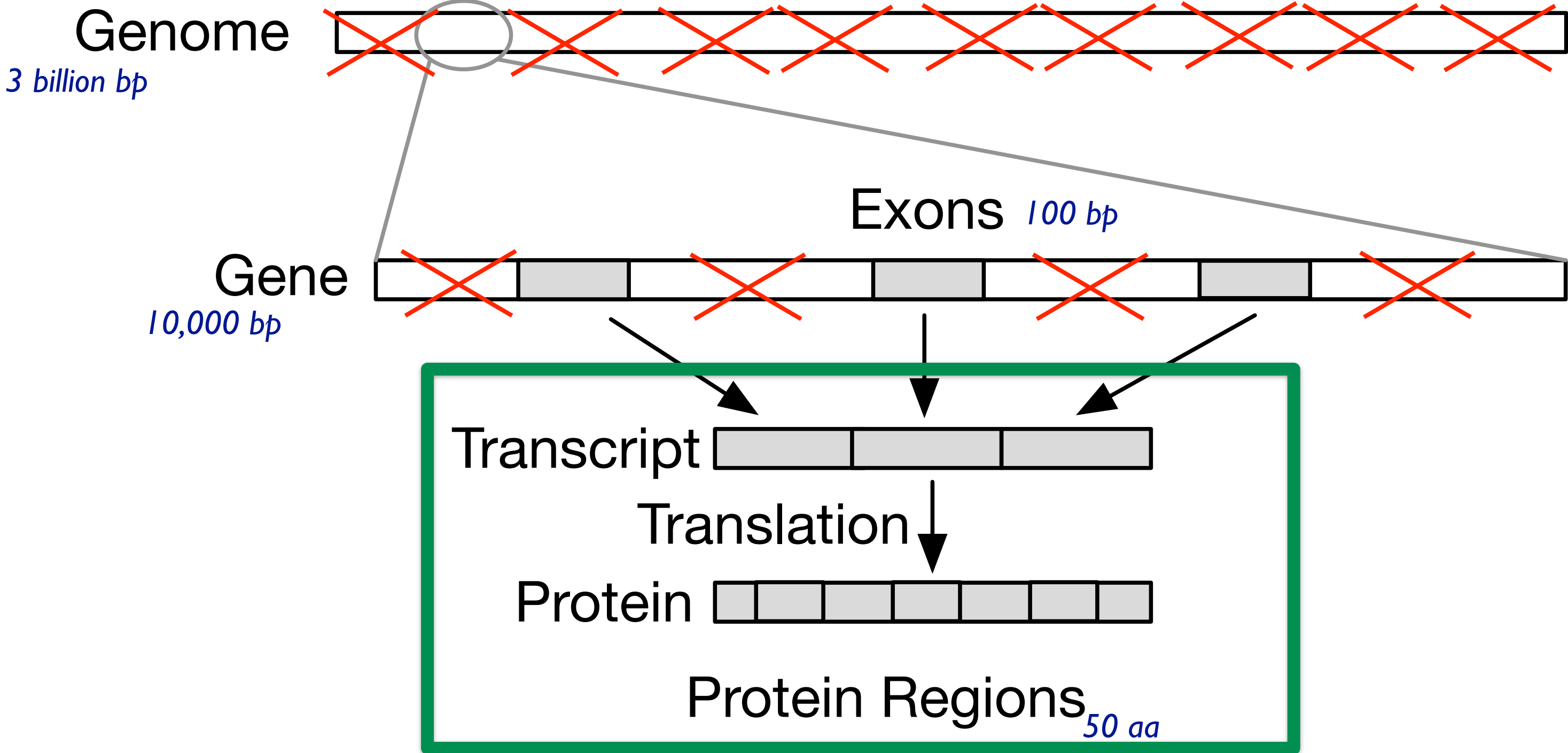
Filter out whole genome; keep genes



Filter out non-exon regions

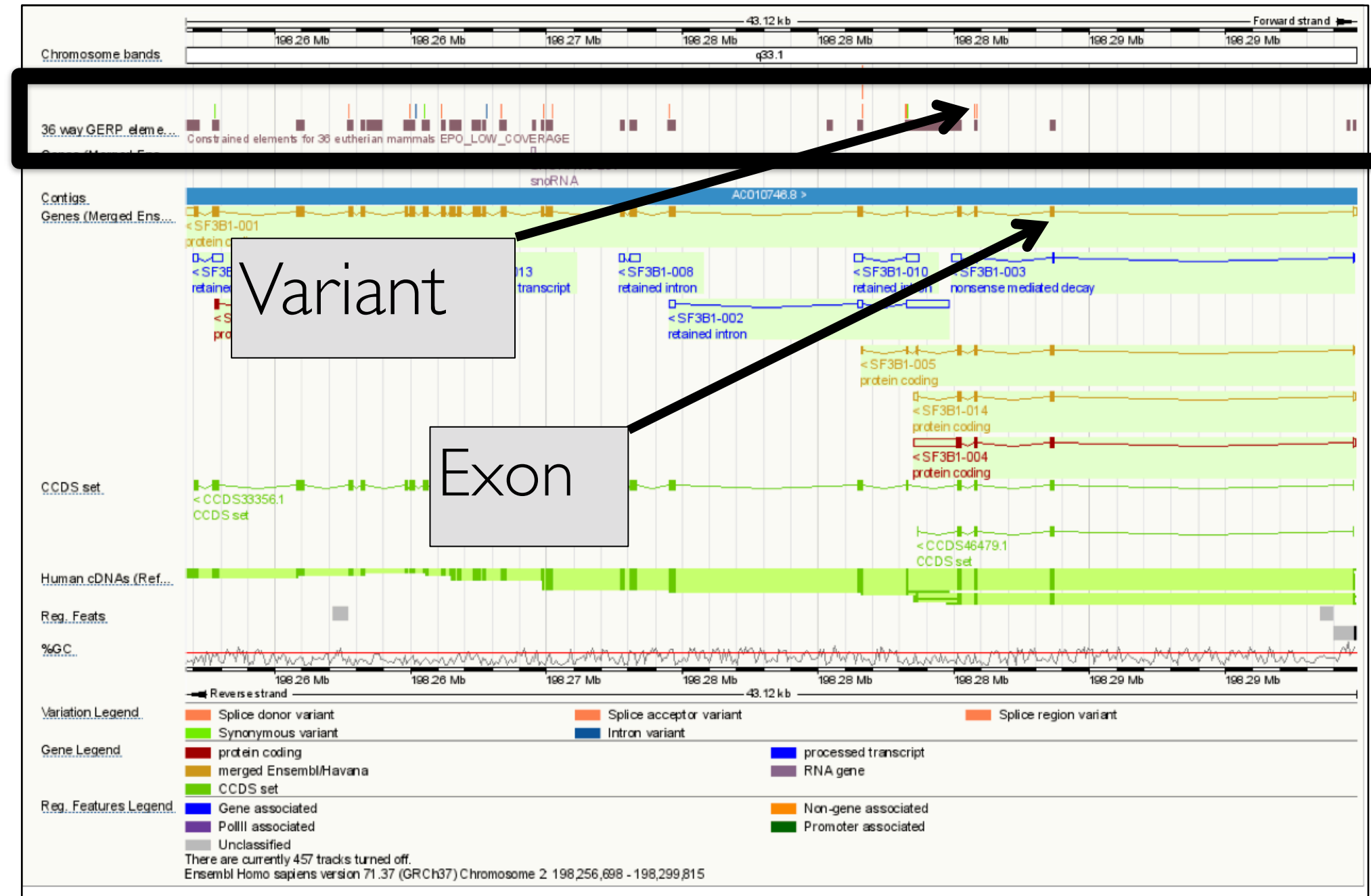


Data abstraction: highly filtered scope of *transcript coordinates*



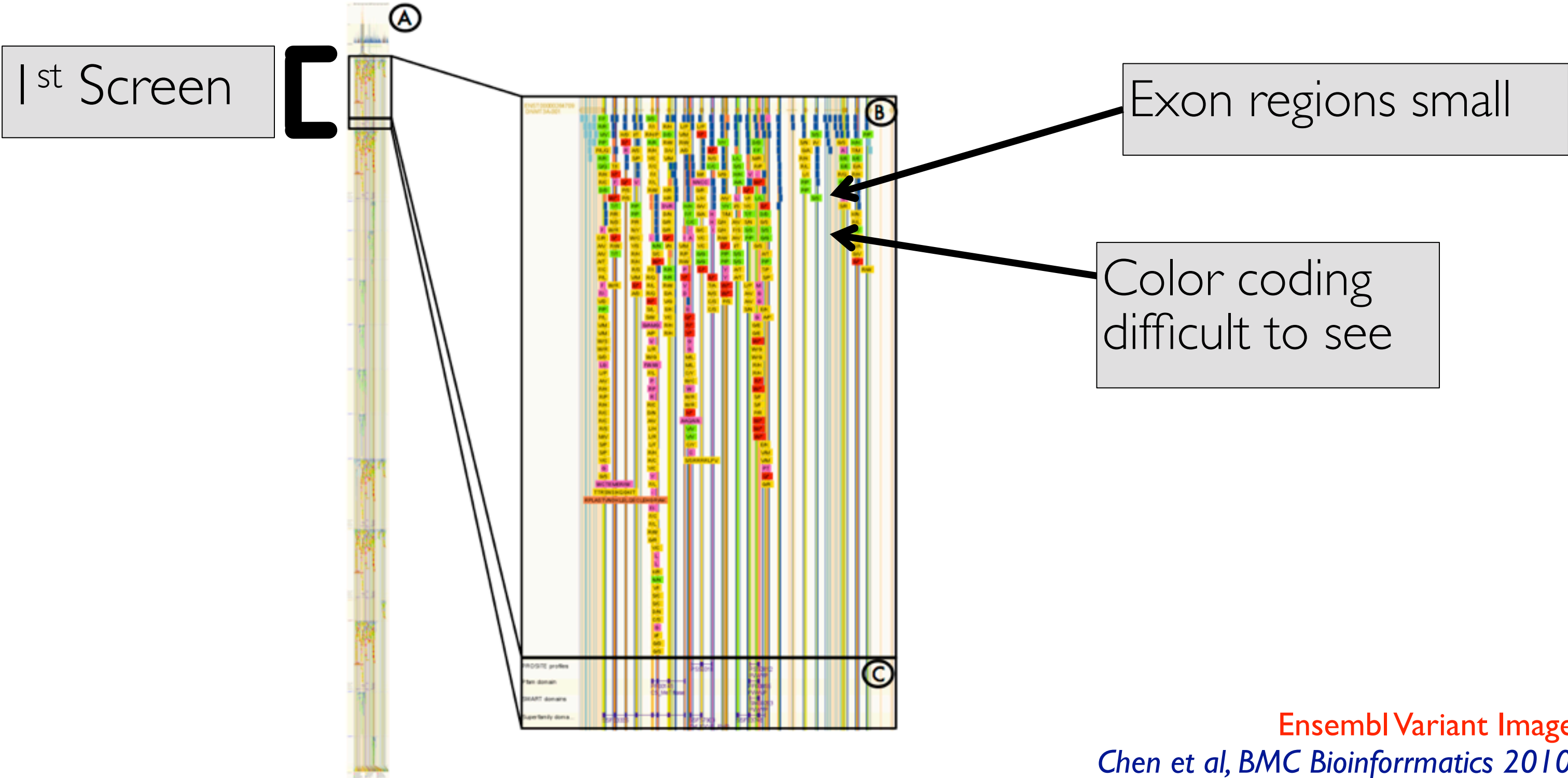
Dominant paradigm: genome browsers

- 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



Idioms

Variant View

Gene Search:

Alternative Transcripts:

Variants

Mutation Type
Reference A.A.s
Variant A.A.s

Transcript

trans-anon

Protein

A.A. Chain
Domains
Regions
Active Sites
Bindings
Mod. Residue

Sort By Gene:

Alpha Cluster Score **Variant Count**

C

- DNMT3A (NM_022552)
- IDH2 (NM_002168)
- FLT3 (NM_004119)
- ANKRD36 (NM_001164315)
- ARID1B (NM_017519)
- STAG2 (NM_001042749)
- TNRC18 (NM_001080495)
- WT1 (NM_000378)
- ABCA13 (NM_152701)
- CEBPA (NM_004364)
- TET2 (NM_001127208)
- DNAH10 (NM_207437)
- GPSM1 (NM_015597)
- ASXL1 (NM_015338)
- DNAH1 (NM_015512)
- DNAH6 (NM_001370)
- FAT1 (NM_005245)
- MDN1 (NM_014611)
- PTPN11 (NM_002834)
- SYNE1 (NM_033071)
- ALMS1 (NM_015120)
- C10orf68 (NM_024688)
- CCDC88C (NM_001080414)
- DNAH11 (NM_003777)
- DNAH3 (NM_017539)
- DNAH9 (NM_001372)

Variant Data

Patient ID	Chr. Coord.	Ref Base	Var Base	dbSNP129	dbSNP135	dbSNP137	COSMIC	A.A. Chng.	Gene	Ref. Gene
pid-anon	11288816	G	T	.	.	.	*13028,	G60V	gene-anon	trans-anon
pid-anon	11288816	G	T	.	.	.	*13012,	D61Y	gene-anon	trans-anon
pid-anon	11288819	G	T	.	rs121918	.	13014	A72S	gene-anon	trans-anon
pid-anon	11288819	C	T	.	.	.	*13035,	A72V	gene-anon	trans-anon
pid-anon	11288821	G	C	.	.	.	*13016,	E76Q	gene-anon	trans-anon
pid-anon	11288821	A	G	.	rs121918	.	*13017,	E76G	gene-anon	trans-anon
pid-anon	11288821	G	T	E76D	gene-anon	trans-anon
pid-anon	11292688	T	A	.	rs121918	.	*13020,	S502T	gene-anon	trans-anon
pid-anon	11292688	T	G	.	.	.	*13020,	S502A	gene-anon	trans-anon
pid-anon	11292688	C	T	.	.	.	13023	S502L	gene-anon	trans-anon

B

Variant View

Information-dense single gene view

Gene Search:

Alternative Transcripts:

Variants

Mutation Type
Reference A.A.s
Variant A.A.s

Transcript

trans-anon

Protein

A.A. Chain
Domains
Regions
Active Sites
Bindings
Mod. Residue

Variant Data

Patient ID	Chr. Coord.	Ref Base	Var Base	dbSNP129	dbSNP135	dbSNP137	COSMIC	A.A. Chng.	Gene	Ref. Gene
pid-anon	11288816	G	T	.	.	.	*13028,	G60V	gene-anon	trans-anon
pid-anon	11288816	G	T	.	.	.	*13012,	D61Y	gene-anon	trans-anon
pid-anon	11288819	G	T	.	rs121918	.	13014	A72S	gene-anon	trans-anon
pid-anon	11288819	C	T	.	.	.	*13035,	A72V	gene-anon	trans-anon
pid-anon	11288821	G	C	.	.	.	*13016,	E76Q	gene-anon	trans-anon
pid-anon	11288821	A	G	.	rs121918	.	*13017,	E76G	gene-anon	trans-anon
pid-anon	11288821	G	T	E76D	gene-anon	trans-anon
pid-anon	11292688	T	A	.	rs121918	.	*13020,	S502T	gene-anon	trans-anon
pid-anon	11292688	T	G	.	.	.	*13020,	S502A	gene-anon	trans-anon
pid-anon	11292688	C	T	.	.	.	13023	S502L	gene-anon	trans-anon

Sort By Gene:
Alpha Cluster Score **Variant Count**

- DNMT3A (NM_022552)
- IDH2 (NM_002168)
- FLT3 (NM_004119)
- ANKRD36 (NM_001164315)
- ARID1B (NM_017519)
- STAG2 (NM_001042749)
- TNRC18 (NM_001080495)
- WT1 (NM_000378)
- ABCA13 (NM_152701)
- CEBPA (NM_004364)
- TET2 (NM_001127208)
- DNAH10 (NM_207437)
- GPSM1 (NM_015597)
- ASXL1 (NM_015338)
- DNAH1 (NM_015512)
- DNAH6 (NM_001370)
- FAT1 (NM_005245)
- MDN1 (NM_014611)
- PTPN11 (NM_002834)
- SYNE1 (NM_033071)
- ALMS1 (NM_015120)
- C10orf68 (NM_024688)
- CCDC88C (NM_001080414)
- DNAH11 (NM_003777)
- DNAH3 (NM_017539)
- DNAH9 (NM_001372)

Variant View

Gene Search:

Alternative Transcripts:

Information-dense single gene view

Variants

Mutation Type
Reference A.A.s
Variant A.A.s

Transcript

trans-anon

Protein

A.A. Chain
Domains
Regions
Active Sites
Bindings
Mod. Residue

Variant Data

Patient ID	Chr. Coord.	Ref Base	Var Base	dbSNP129	dbSNP135	dbSNP137	COSMIC	A.A. Ch		
pid-anon	11288816	G	T	.	.	.	*13028,	G60V		
pid-anon	11288816	G	T	.	.	.	*13012,	D61Y		
pid-anon	11288819	G	T	.	rs121918	.	13014	A72S	gene-anon	trans-anon
pid-anon	11288819	C	T	.	.	.	*13035,	A72V	gene-anon	trans-anon
pid-anon	11288821	G	C	.	.	.	*13016,	E76Q	gene-anon	trans-anon
pid-anon	11288821	A	G	.	rs121918	.	*13017,	E76G	gene-anon	trans-anon
pid-anon	11288821	G	T	E76D	gene-anon	trans-anon
pid-anon	11292688	T	A	.	rs121918	.	*13020,	S502T	gene-anon	trans-anon
pid-anon	11292688	T	G	.	.	.	*13020,	S502A	gene-anon	trans-anon
pid-anon	11292688	C	T	.	.	.	13023	S502L	gene-anon	trans-anon

Sort By Gene:
Alpha Cluster Score Variant Count

- DNMT3A (NM_022552)
- IDH2 (NM_002168)
- FLT3 (NM_004119)
- ANKRD36 (NM_001164315)
- ARID1B (NM_017519)
- STAG2 (NM_001042749)
- TNRC18 (NM_001080495)
- WT1 (NM_000378)
- ABCA13 (NM_152701)
- CEBPA (NM_004364)
- TET2 (NM_001127208)
- DNAH10 (NM_207437)
- GPSM1 (NM_015597)
- ASXL1 (NM_015338)
- DNAH1 (NM_015512)
- DNAH6 (NM_001370)
- SYNET (NM_033071)
- ALMS1 (NM_015120)
- C10orf68 (NM_024688)
- CCDC88C (NM_001080414)
- DNAH11 (NM_003777)
- DNAH3 (NM_017539)
- DNAH9 (NM_001372)

No need for pan and zoom

Variant View

Sorting metrics guide gene navigation

Alternative Transcripts:

Variants

Mutation Type
Reference A.A.s
Variant A.A.s

Transcript

trans-anon

Protein

A.A. Chain
Domains
Regions
Active Sites
Bindings
Mod. Residue

Variant Data

Patient ID	Chr. Coord.	Ref Base	Var Base	dbSNP129	dbSNP135	dbSNP137	COSMIC	A.A. Chng.	Gene	Ref. Gene
pid-anon	11288816	G	T	.	.	.	*13028,	G60V	gene-anon	trans-anon
pid-anon	11288816	G	T	.	.	.	*13012,	D61Y	gene-anon	trans-anon
pid-anon	11288819	G	T	.	rs121918	.	13014	A72S	gene-anon	trans-anon
pid-anon	11288819	C	T	.	.	.	*13035,	A72V	gene-anon	trans-anon
pid-anon	11288821	G	C	.	.	.	*13016,	E76Q	gene-anon	trans-anon
pid-anon	11288821	A	G	.	rs121918	.	*13017,	E76G	gene-anon	trans-anon
pid-anon	11288821	G	T	E76D	gene-anon	trans-anon
pid-anon	11292688	T	A	.	rs121918	.	*13020,	S502T	gene-anon	trans-anon
pid-anon	11292688	T	G	.	.	.	*13020,	S502A	gene-anon	trans-anon
pid-anon	11292688	C	T	.	.	.	13023	S502L	gene-anon	trans-anon

Sort By Gene: Alpha Cluster Score Variant Count

DNMT3A (NM_022552) ©
IDH2 (NM_002168)
FLT3 (NM_004119)
ANKRD36 (NM_001164315)
ARID1B (NM_017519)
STAG2 (NM_001042749)
TNRC18 (NM_001080495)
WT1 (NM_000378)
ABCA13 (NM_152701)
CEBPA (NM_004364)
TET2 (NM_001127208)
DNAH10 (NM_207437)
GPSM1 (NM_015597)
ASXL1 (NM_015338)
DNAH1 (NM_015512)
DNAH6 (NM_001370)
FAT1 (NM_005245)
MDN1 (NM_014611)
PTPN11 (NM_002834)
SYNE1 (NM_033071)
ALMS1 (NM_015120)
C10orf68 (NM_024688)
CCDC88C (NM_001080414)
DNAH11 (NM_003777)
DNAH3 (NM_017539)
DNAH9 (NM_001372)

Variant View

Sorting metrics guide gene navigation

The screenshot displays a 'Variant View' interface. At the top, there's a search bar with 'gene-anon (trans-anon)'. Below it, the 'Variants' section shows mutation types, reference A.A.s, and variant A.A.s. The 'Transcript' section shows a 'trans-anon' transcript. The 'Protein' section shows an 'A.A. Chain' with domains, regions, active sites, bindings, and mod. residue. The 'Variant Data' table is at the bottom.

Variant ID	dbSNP137	COSMIC	A.A. Chng.	Gene	Ref. Gene
pid-anon	1128819	C	T	gene-anon	trans-anon
pid-anon	1128821	G	C	gene-anon	trans-anon
pid-anon	1128821	A	G	gene-anon	trans-anon
pid-anon	1128821	G	T	gene-anon	trans-anon
pid-anon	11292688	T	A	gene-anon	trans-anon
pid-anon	11292688	T	G	gene-anon	trans-anon
pid-anon	11292688	C	T	gene-anon	trans-anon

On the right side, there's a list of genes sorted by 'Variant Count'. The list includes: DNMT3A (NM_022552), IDH2 (NM_002168), FLT3 (NM_004119), ANKRD36 (NM_001164315), ARID1B (NM_017519), STAG2 (NM_001042749), TNRC18 (NM_001080495), WT1 (NM_000378), ABCA13 (NM_152701), CEBPA (NM_004364), TET2 (NM_001127208), DNAH10 (NM_207437), GPSM1 (NM_015597), ASXL1 (NM_015338), DNAH1 (NM_015512), DNAH6 (NM_001370), FAT1 (NM_005245), MDN1 (NM_014611), PTPN11 (NM_002834), SYNE1 (NM_033071), ALMS1 (NM_015120), C10orf68 (NM_024688), CCDC88C (NM_001080414), DNAH11 (NM_003777), DNAH3 (NM_017539), and DNAH9 (NM_001372).

Control what shows up here

Variant View

Gene Search:

Alternative Transcripts:

Variants

Mutation Type
Reference A.A.s
Variant A.A.s

Transcript

trans-anon

Protein

A.A. Chain
Domains
Regions
Active Sites
Bindings
Mod. Residue

Variant Data

Patient ID	Chr.	Coord.	Ref Base	Var Base	dbSNP129	dbSNP135	dbSNP137	COSMIC	A.A. Chng.	Gene	Ref. Gene
pid-anon	11288816	G	T	.	.	.	*13028,	G60V	gene-anon	trans-anon	
pid-anon	11288816	G	T	.	.	.	*13012,	D61Y	gene-anon	trans-anon	
pid-anon	11288819	G	T	.	rs121918	.	*13014,	A72S	gene-anon	trans-anon	
pid-anon	11288819	C	T	.	.	.	*13035,	E76Q	gene-anon	trans-anon	
pid-anon	11288821	G	C	.	.	.	*13016,	E76Q	gene-anon	trans-anon	
pid-anon	11288821	A	G	.	rs121918	.	*13017,	E76G	gene-anon	trans-anon	
pid-anon	11288821	G	T	E76D	gene-anon	trans-anon	
pid-anon	11292688	T	A	.	rs121918	.	*13020,	S502T	gene-anon	trans-anon	
pid-anon	11292688	T	G	.	.	.	*13020,	S502A	gene-anon	trans-anon	
pid-anon	11292688	C	T	.	.	.	13023	S502L	gene-anon	trans-anon	

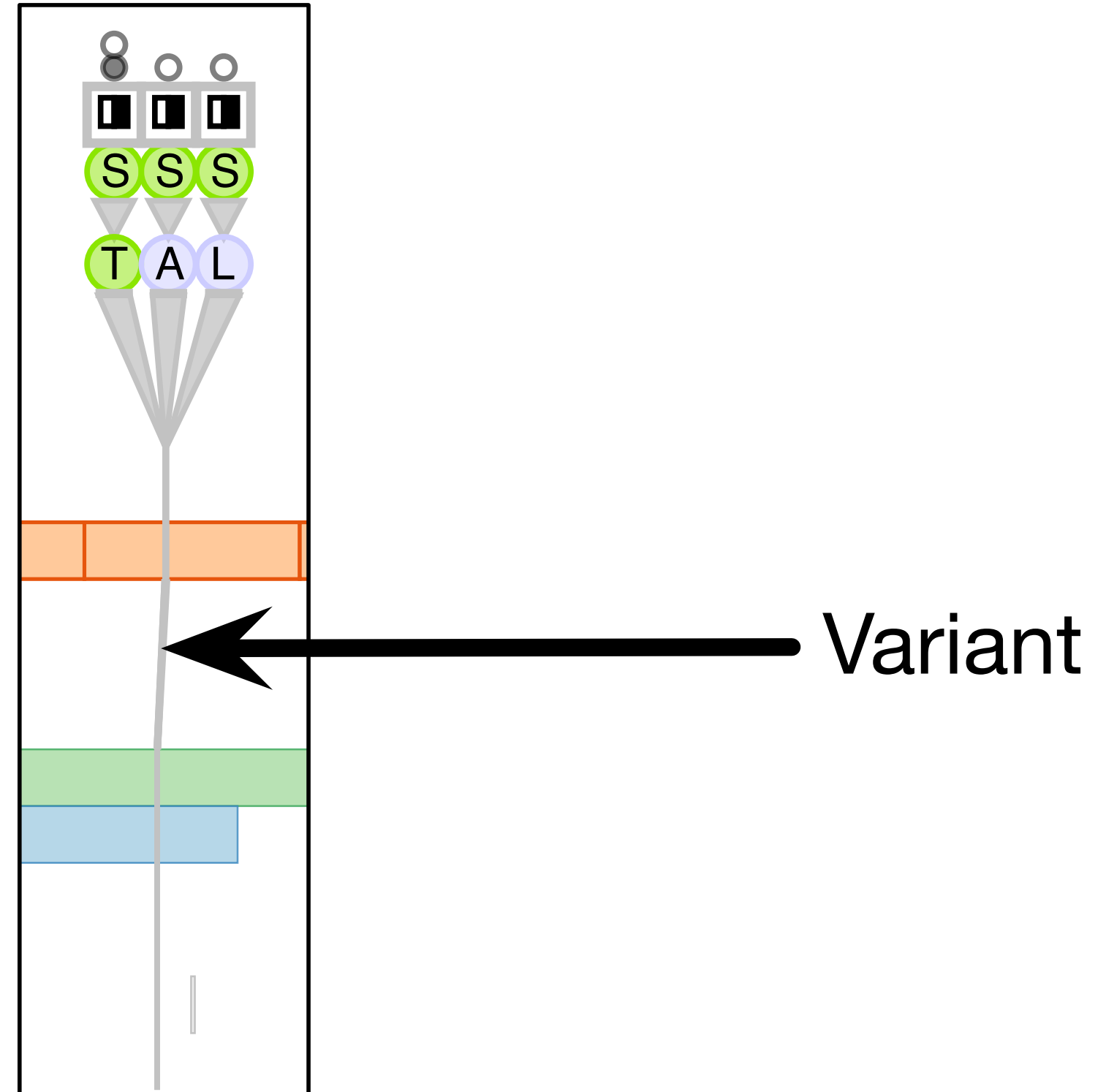
Sort By Gene:
Alpha Cluster Score Variant Count

- DNMT3A (NM_022552)
- IDH2 (NM_002168)
- FLT3 (NM_004119)
- ANKRD36 (NM_001164315)
- ARID1B (NM_017519)
- STAG2 (NM_001042749)
- TNRC18 (NM_001080495)
- WT1 (NM_000378)
- ABCA13 (NM_152701)
- CEBPA (NM_004364)
- TET2 (NM_001127208)
- DNAH10 (NM_207437)
- GPSM1 (NM_015597)
- ASXL1 (NM_015338)
- DNAH1 (NM_015512)
- DNAH6 (NM_001370)
- FAT1 (NM_005245)
- MDN1 (NM_014611)
- PTPN11 (NM_002834)
- DNAH11 (NM_003777)
- DNAH3 (NM_017539)
- DNAH9 (NM_001372)

Peripheral supporting data

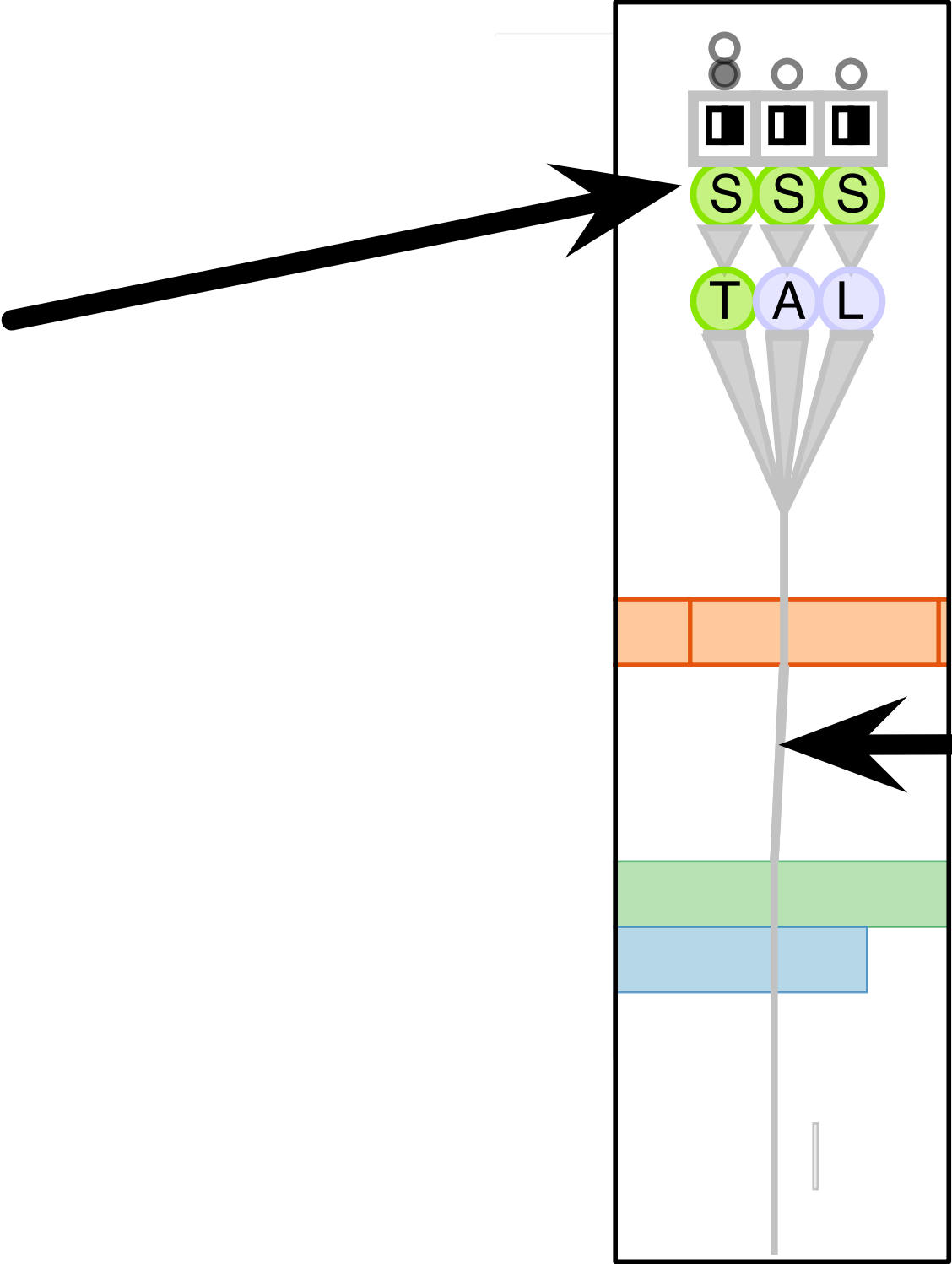
Design information-dense visual encoding

- 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

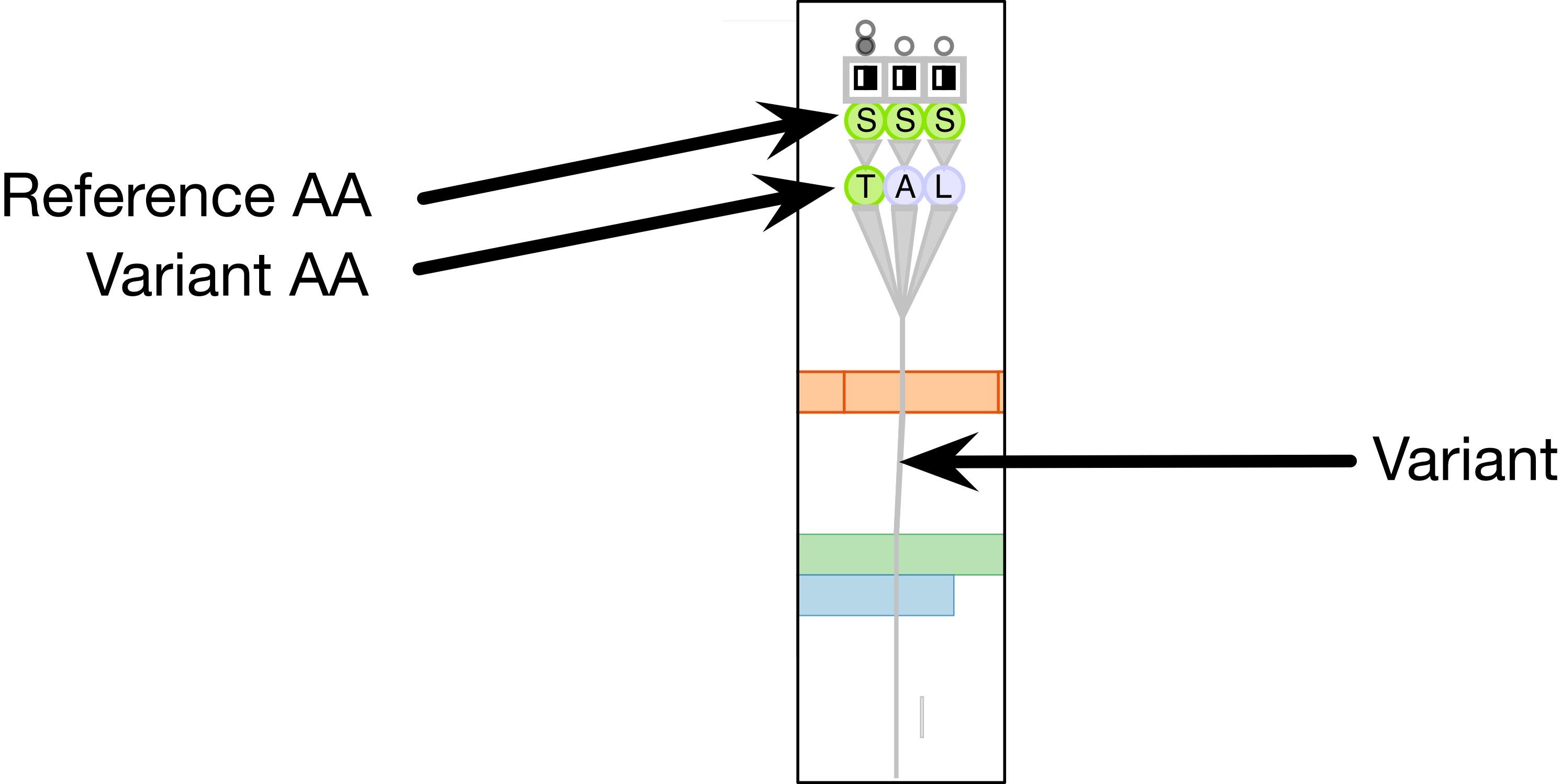
Reference AA



Variant



Design information-dense visual encoding



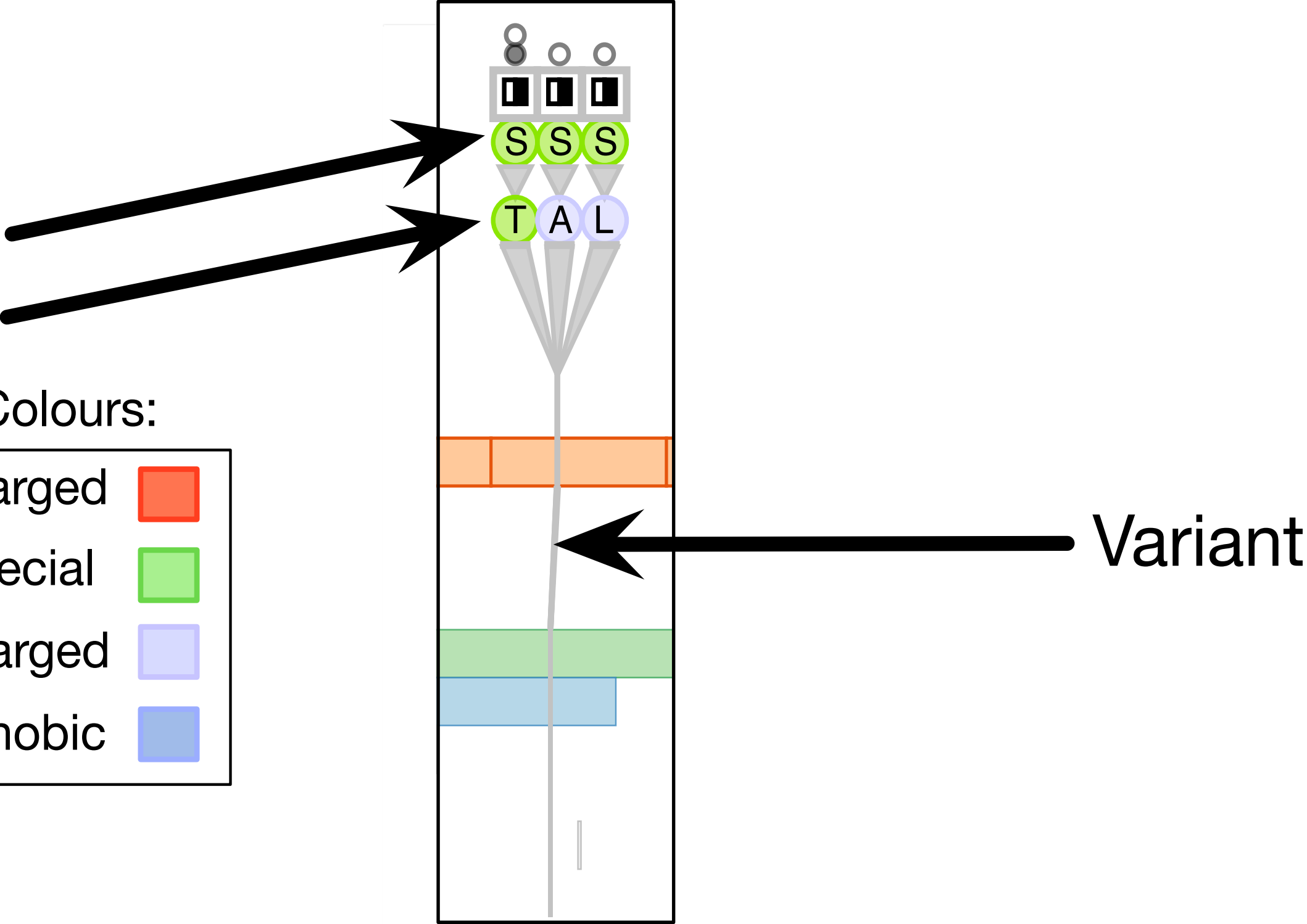
Design information-dense visual encoding

Reference AA

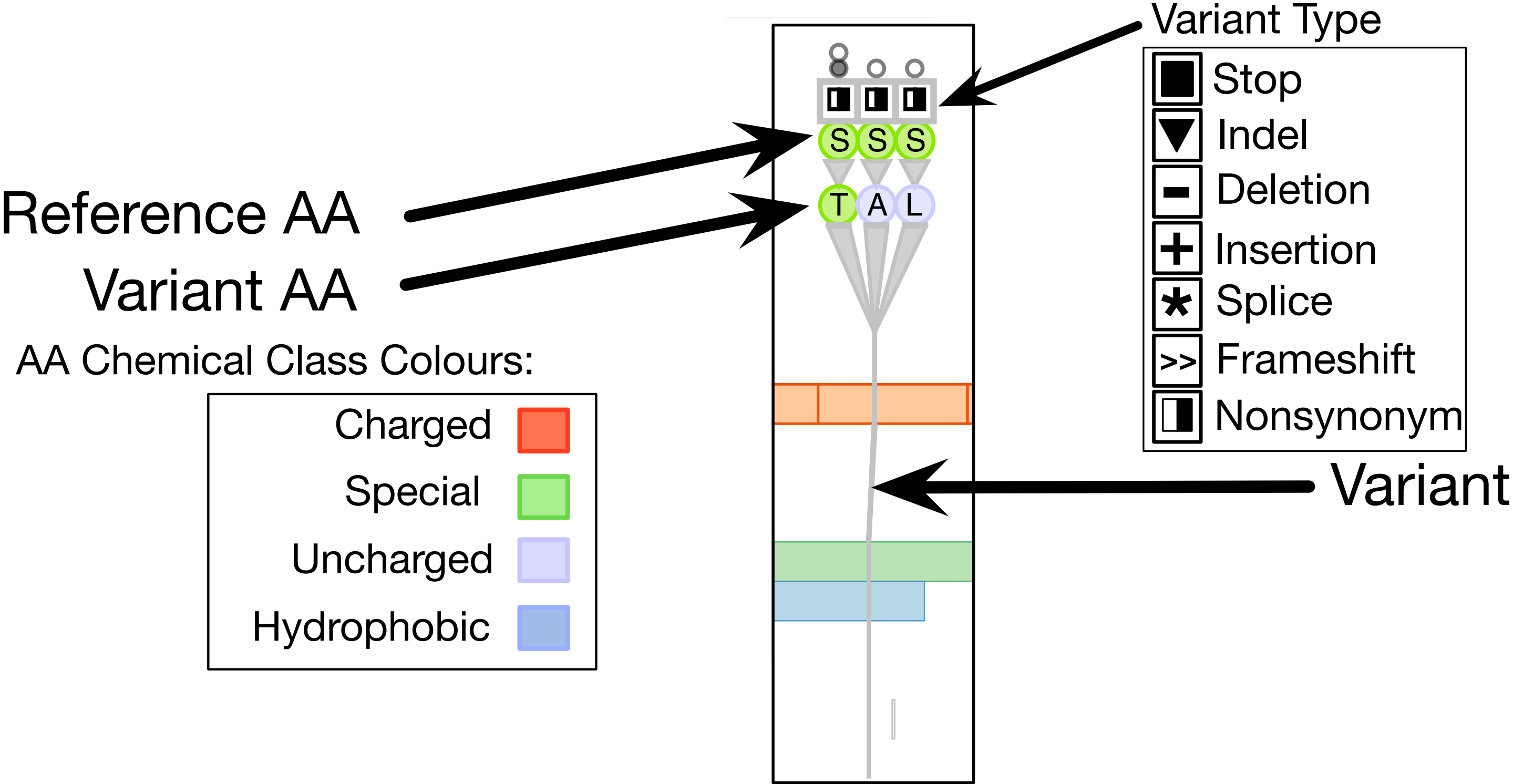
Variant AA

AA Chemical Class Colours:

Charged	■
Special	■
Uncharged	■
Hydrophobic	■



Design information-dense visual encoding



Design information-dense visual encoding

Known Database

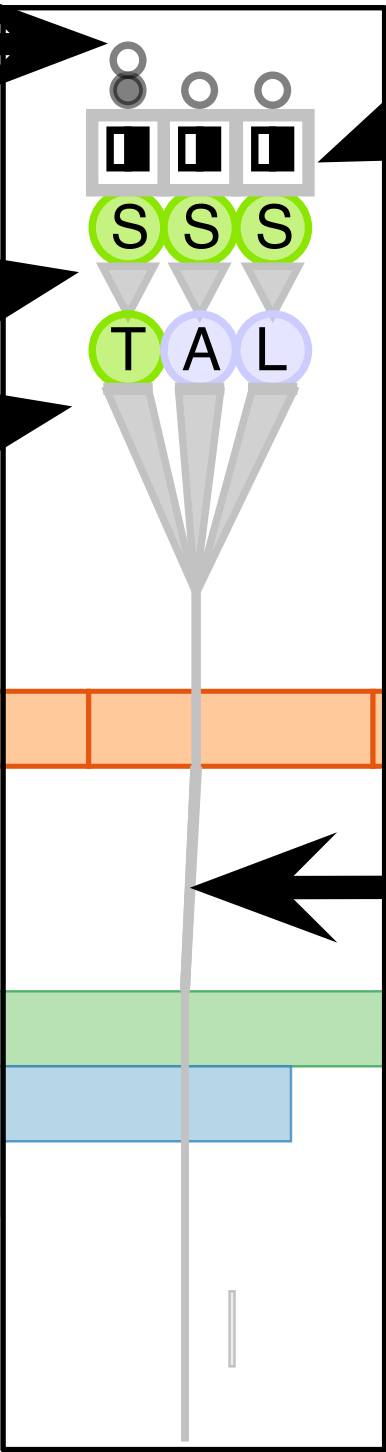
- Known Harmless
- Known Cancer

Reference AA

Variant AA

AA Chemical Class Colours:

- Charged ■
- Special ■
- Uncharged ■
- Hydrophobic ■



Variant Type

- Stop
- ▼ Indel
- ▬ Deletion
- ⊕ Insertion
- * Splice
- >> Frameshift
- ▬ Nonsynonym

Variant

Design information-dense visual encoding

Known Database

- Known Harmless
- Known Cancer

Reference AA

Variant AA

AA Chemical Class Colours:

- Charged ■
- Special ■
- Uncharged ■
- Hydrophobic ■

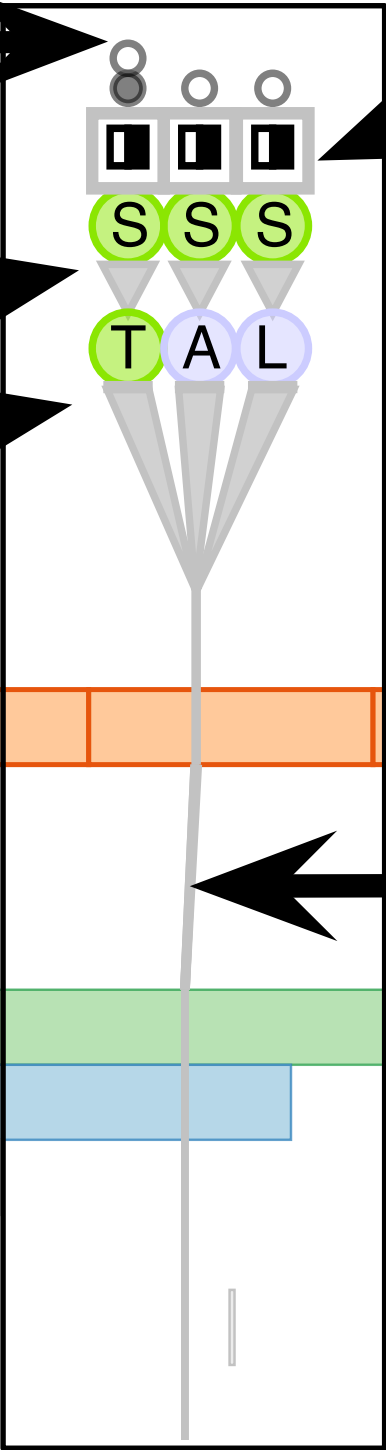
Variant Type

- Stop
- ▼ Indel
- ▬ Deletion
- ⊕ Insertion
- * Splice
- >> Frameshift
- ▣ Nonsynonym

Variant

Transcript/Region Colours:

- Transcript ■
- AA Chain ■
- All Other Regions ■
- Non-Intersected Regions ■

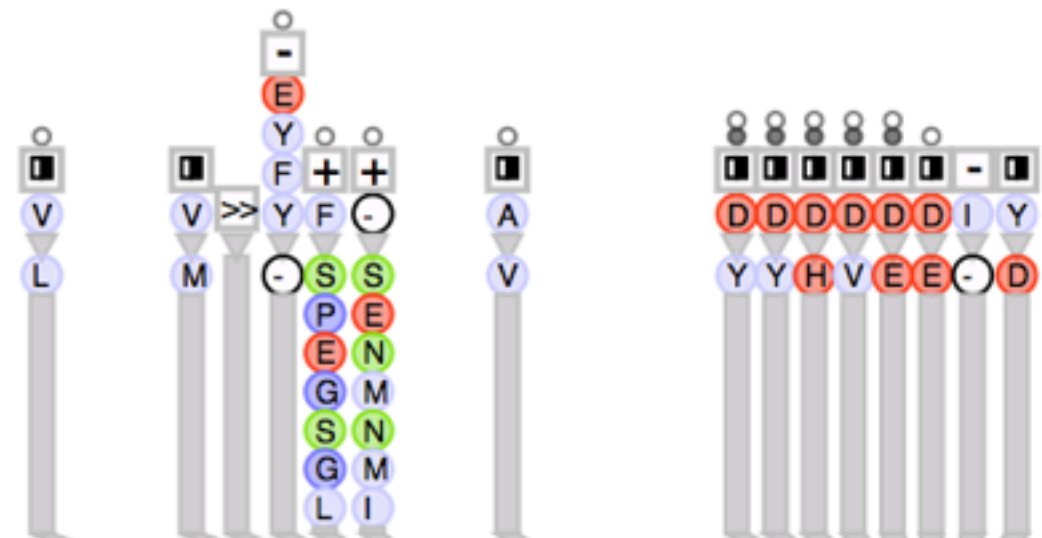


Results

Highly scored gene by sorting metric: known leukemia gene

Variants

Mutation Type
Reference A.A.s
Variant A.A.s



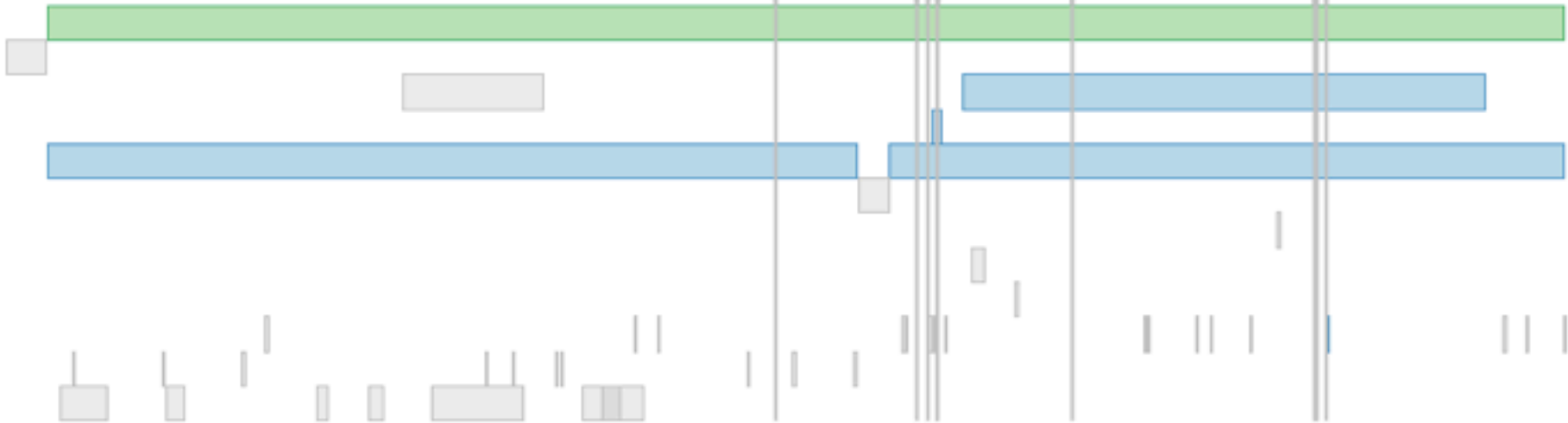
Transcript

trans-anon

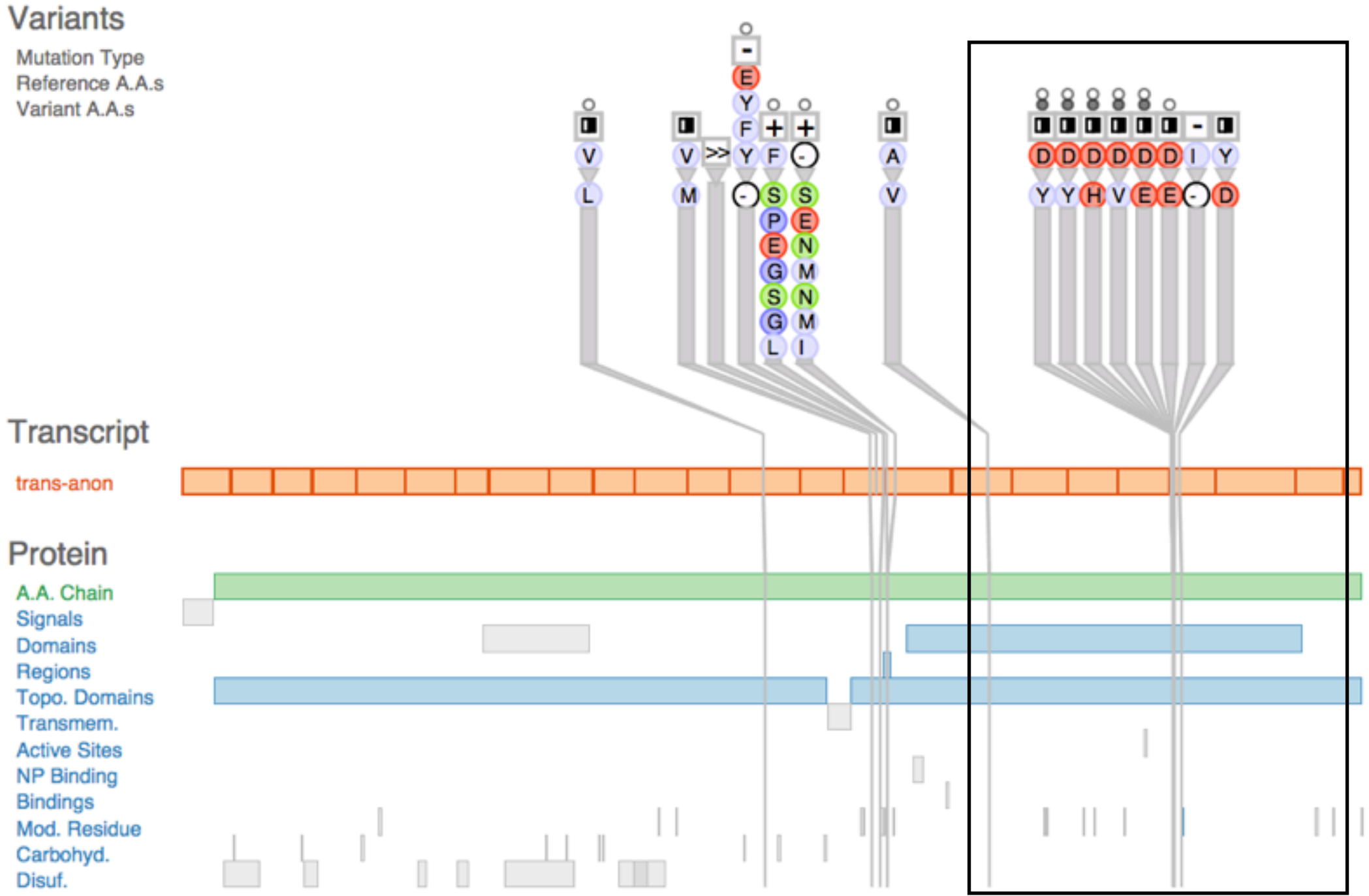


Protein

A.A. Chain
Signals
Domains
Regions
Topo. Domains
Transmem.
Active Sites
NP Binding
Bindings
Mod. Residue
Carbohyd.
Disuf.



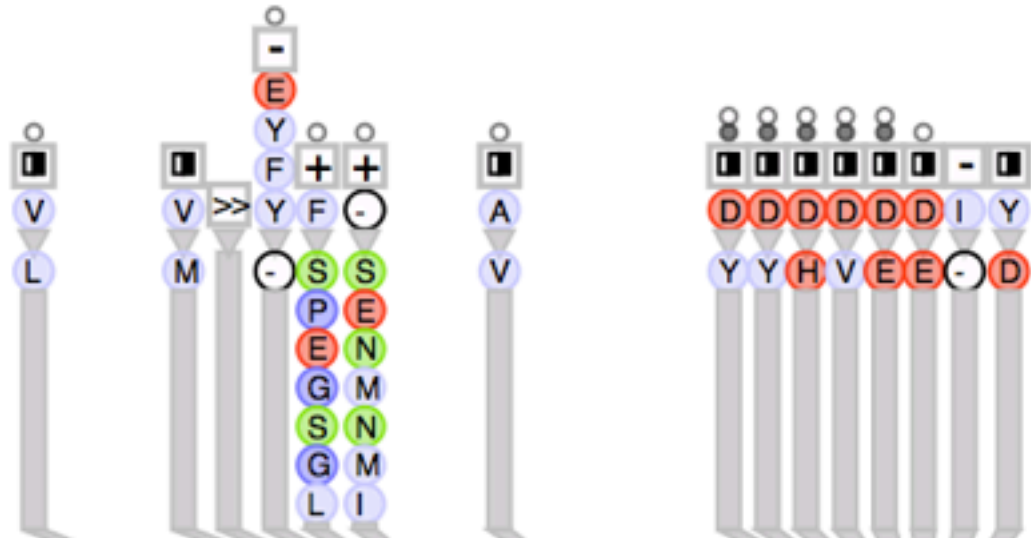
Visual inspection reveals collocation of variants



Several functional protein regions affected

Variants

Mutation Type
Reference A.A.s
Variant A.A.s



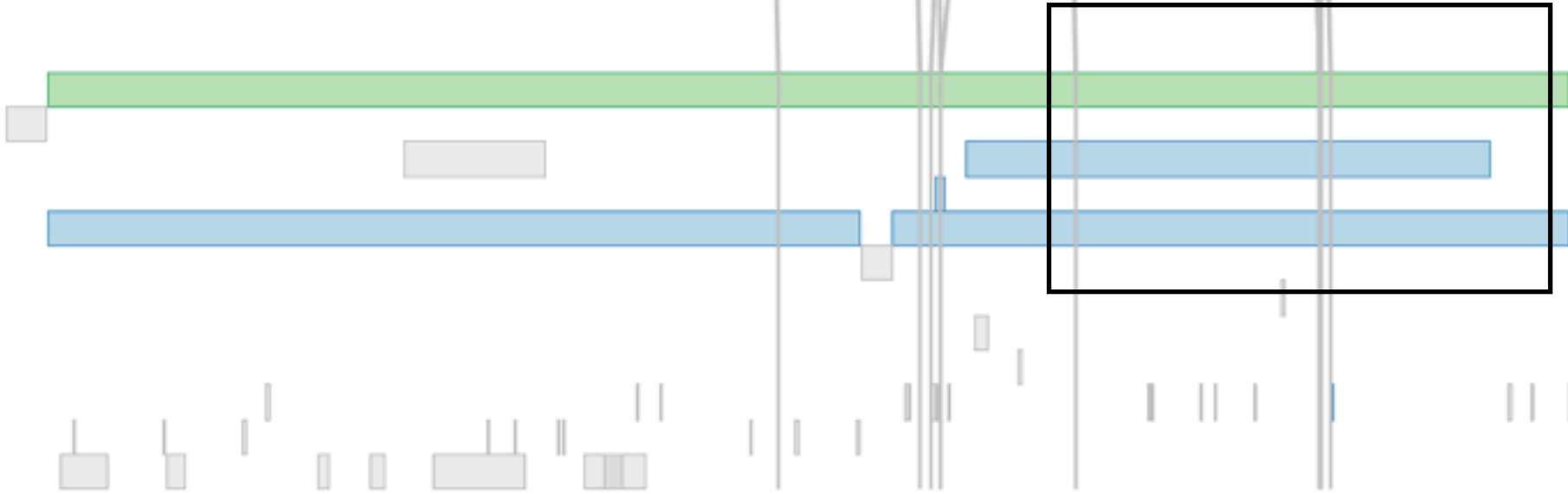
Transcript

trans-anon



Protein

A.A. Chain
Signals
Domains
Regions
Topo. Domains
Transmem.
Active Sites
NP Binding
Bindings
Mod. Residue
Carbohyd.
Disuf.



Highly scored by metric: not previously known, good candidate

Variants

Mutation Type
Reference A.A.s
Variant A.A.s



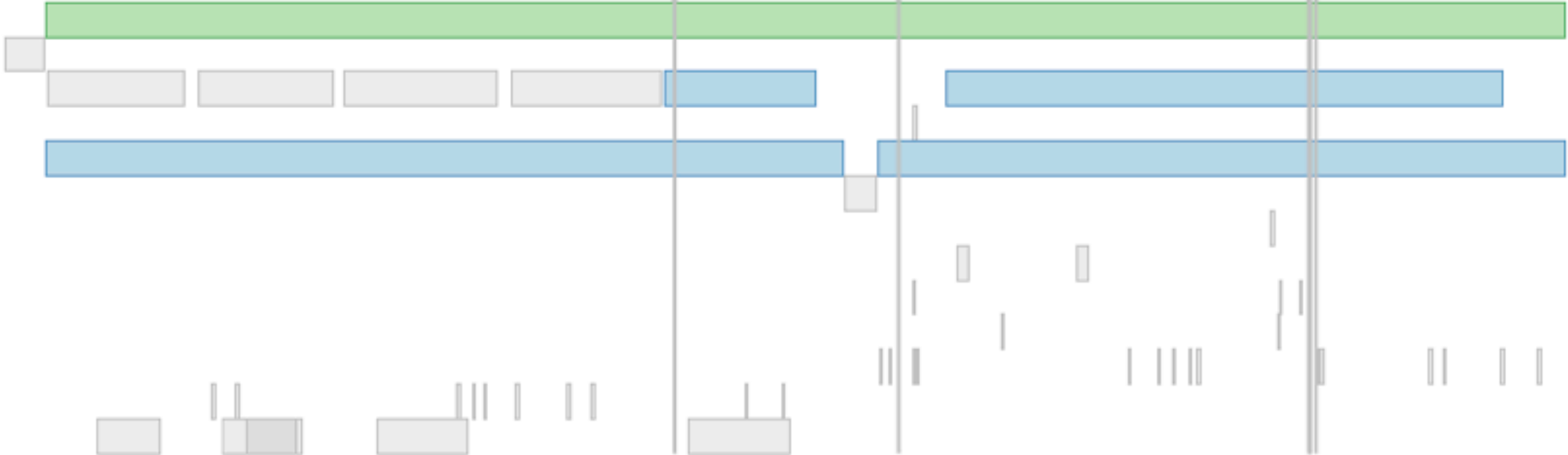
Transcript

trans-anon



Protein

A.A. Chain
Signals
Domains
Regions
Topo. Domains
Transmem.
Active Sites
NP Binding
Metal Bind.
Bindings
Mod. Residue
Carbohyd.
Disuf.



Protein chemical class change evident

Variants

Mutation Type
Reference A.A.s
Variant A.A.s



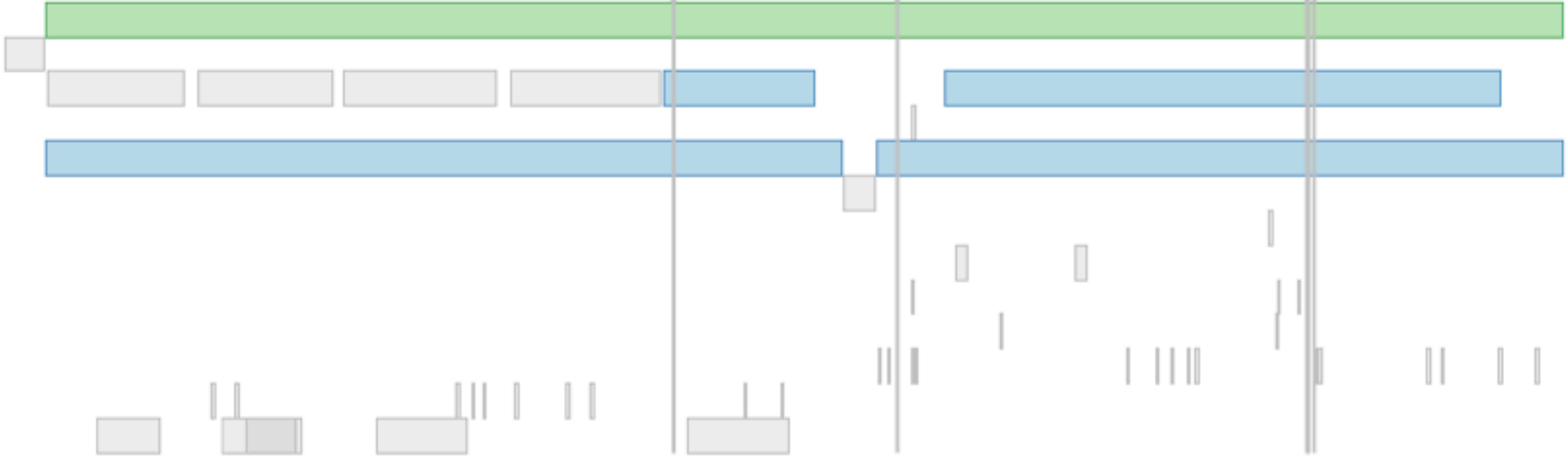
Transcript

trans-anon

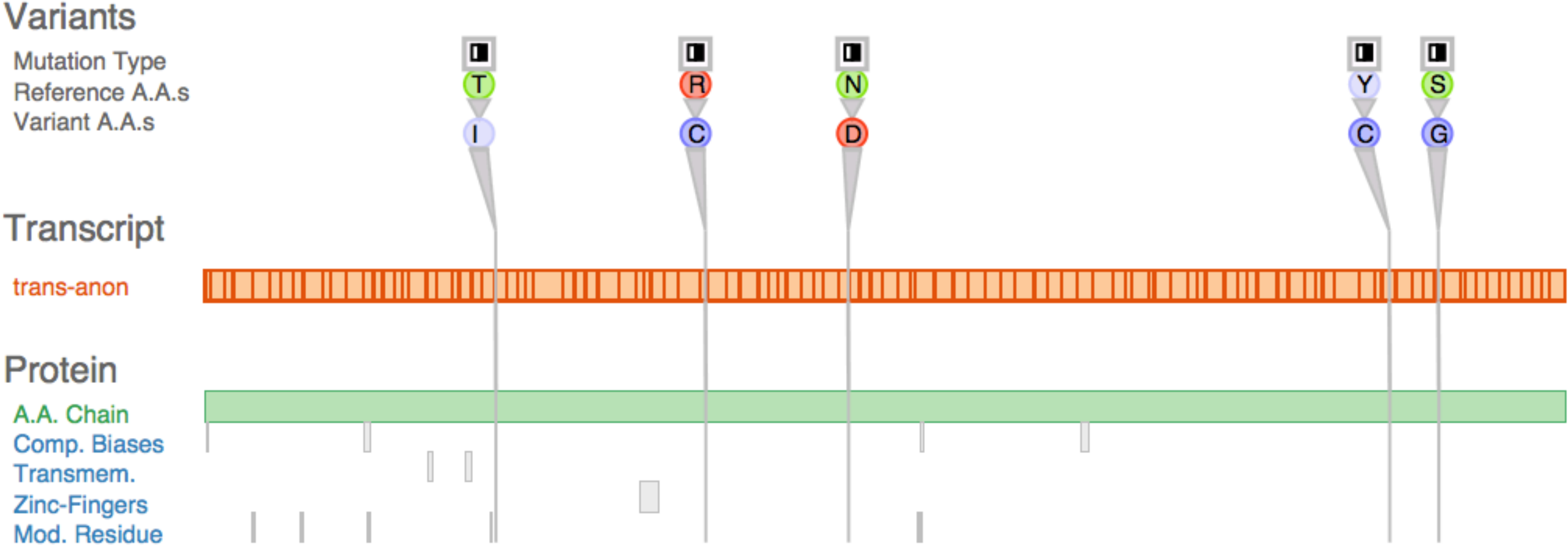


Protein

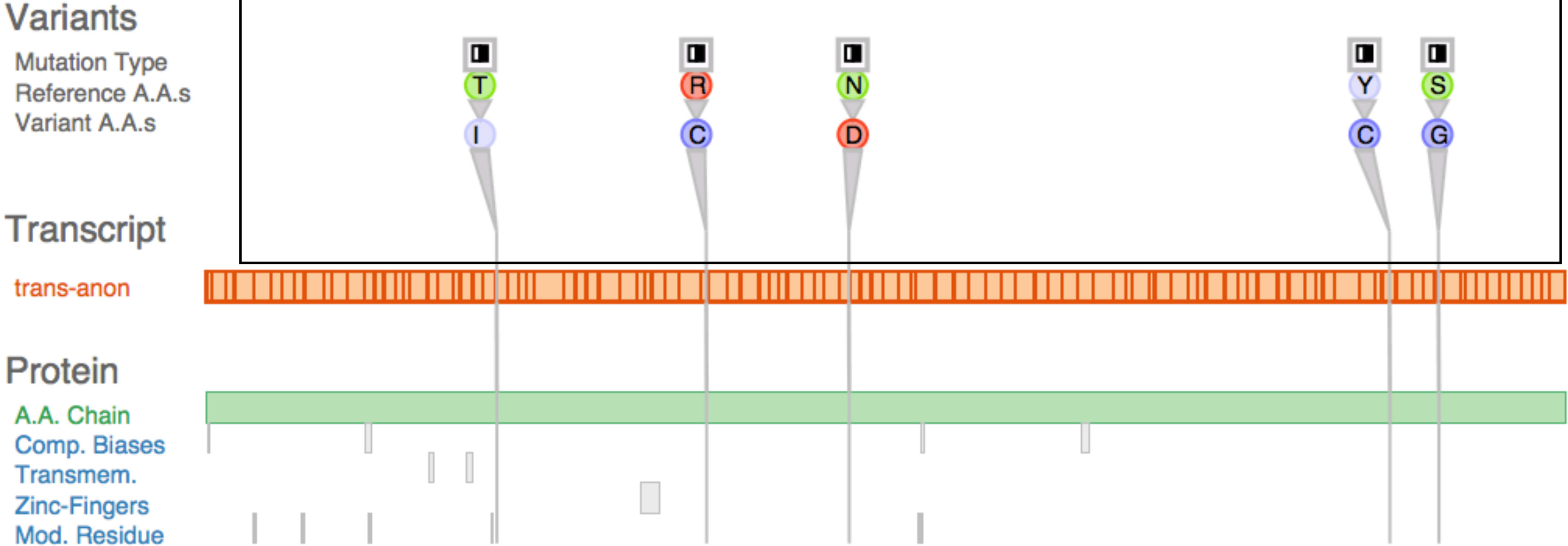
A.A. Chain
Signals
Domains
Regions
Topo. Domains
Transmem.
Active Sites
NP Binding
Metal Bind.
Bindings
Mod. Residue
Carbohyd.
Disuf.



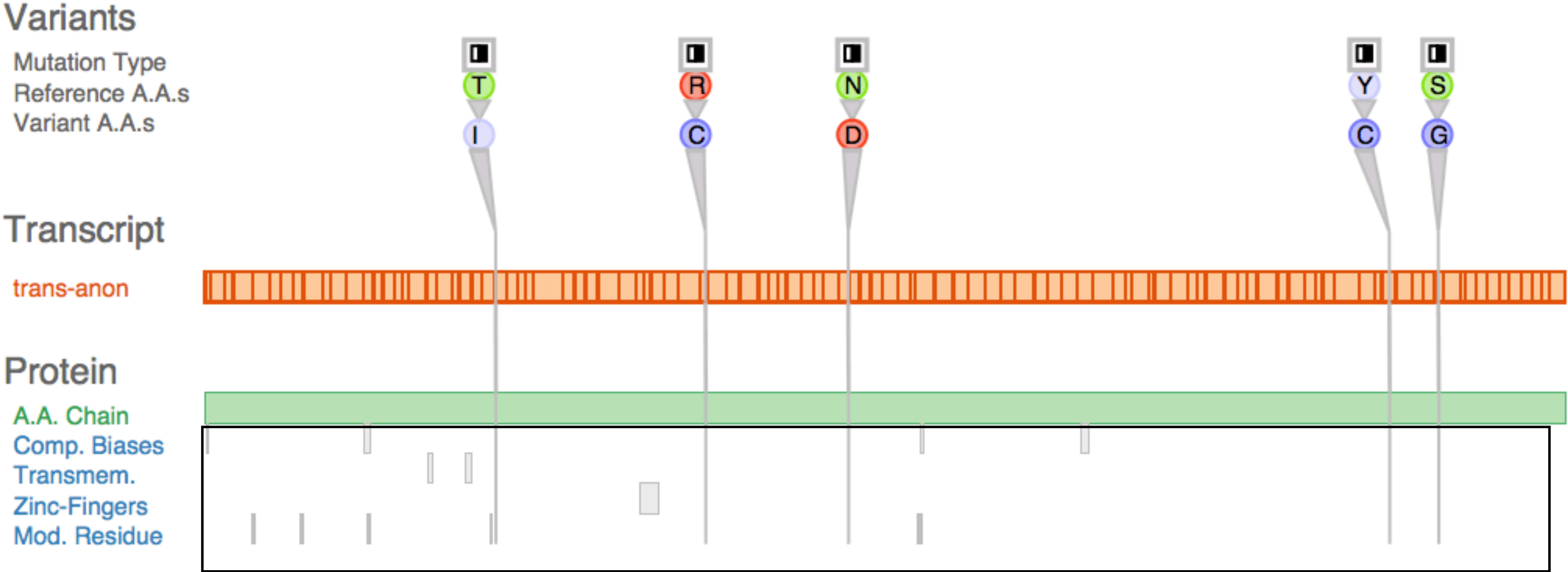
In contrast, low scoring gene



No collocation of variants



Mostly unaffected protein regions



Methods

Phase I: 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



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

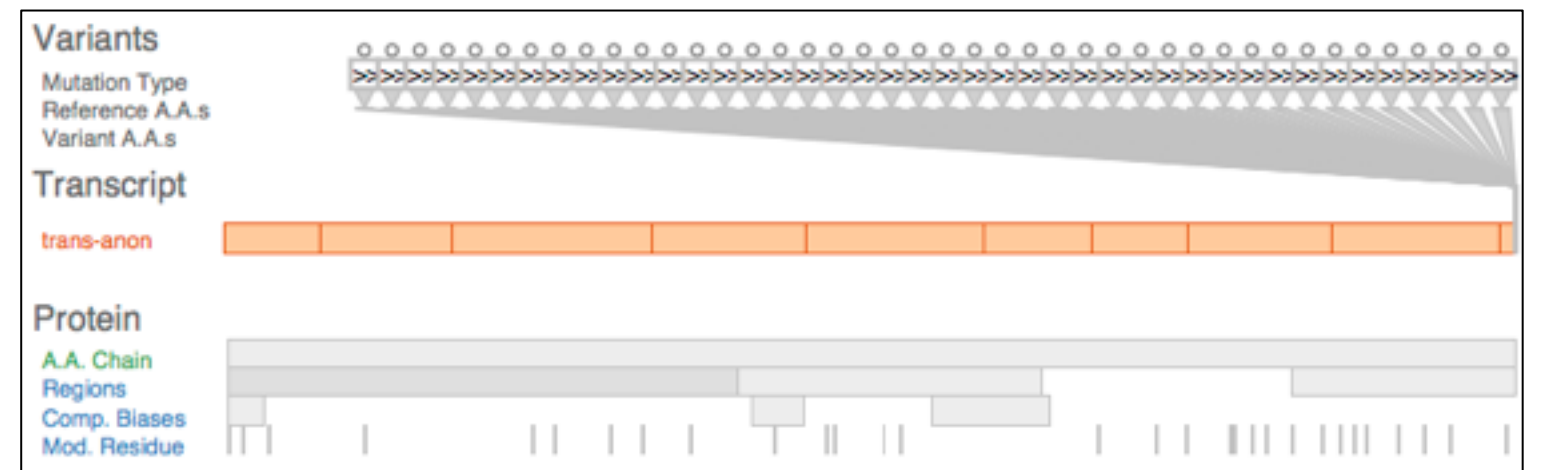
Lloyd and Dykes. *IEEE TVCG (Proc. InfoVis)*, 17(12):2498–2507, 2011.

Phase 3: Two More Tasks

1 month



- 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

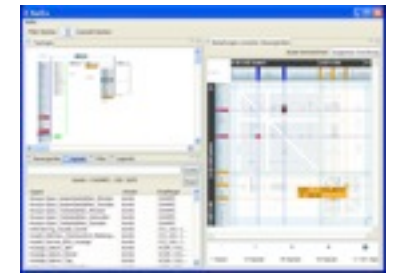
- identifying scales of interest

- what to visually encode directly vs what to support through interaction

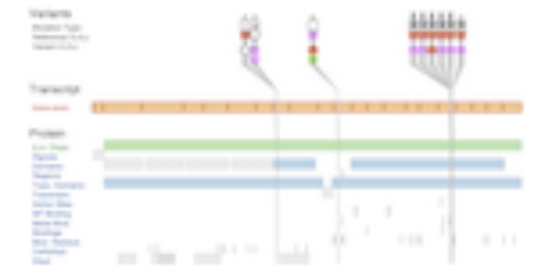
- when (and how) to eliminate navigation

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

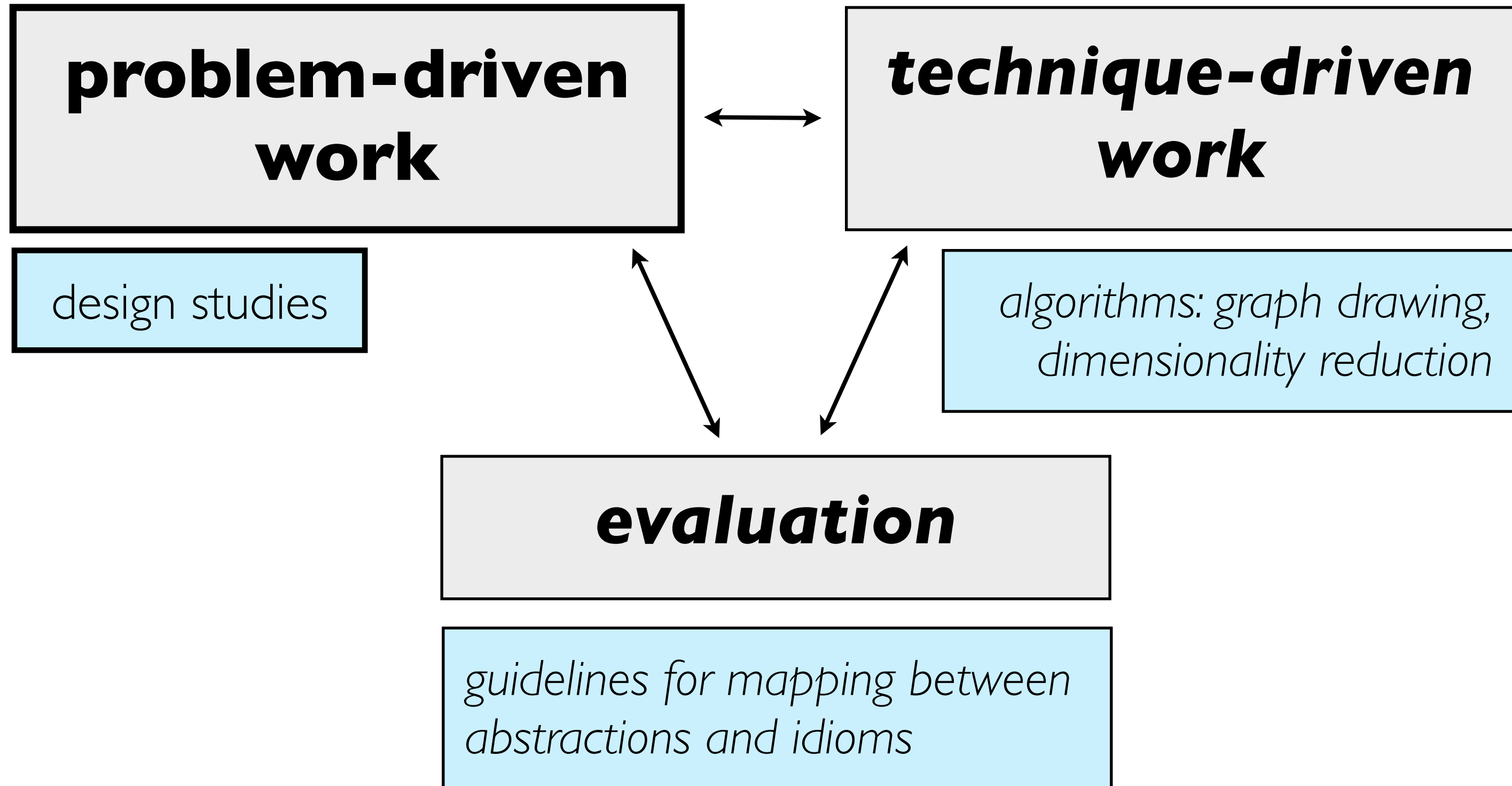


RelEx
in-car networks



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

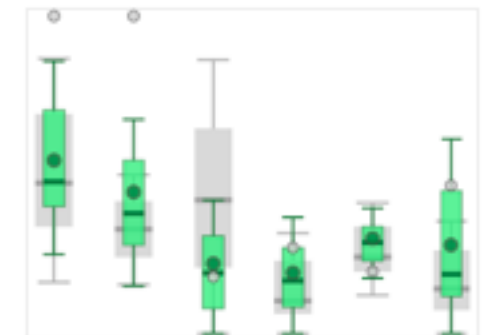
sketchy rendering
IEEE TVCG 2012



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

