

Information Visualization

Advanced Topics

Tamara Munzner

Department of Computer Science
University of British Columbia

Week 14 sync class, 4 Dec 2025

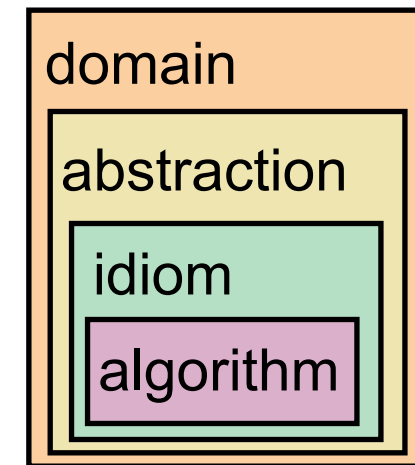
<https://www.cs.ubc.ca/~tmm/courses/547-25/>

Today: Week 14

- UBC InfoVis group research 1: problem-driven
- break
- UBC InfoVis group research 2: Vis & ML intersections
- next steps

Applying visualization to real-world problems

- today: research highlights from my own group
- nested model: going down, starting from top
 - map from domain to abstraction level
 - crucial & difficult, iterative process
 - select appropriate idioms
 - or create new ones if necessary
- many case studies
 - different domains
 - different methods



Four case studies of problem-driven work

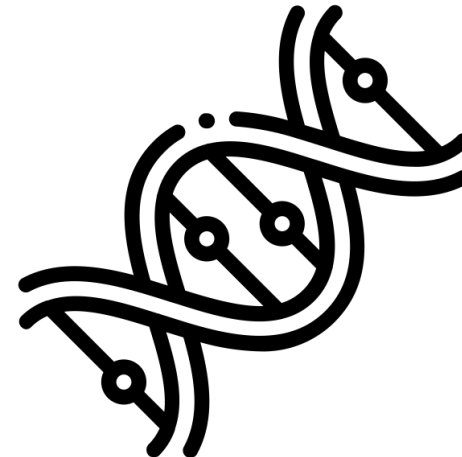
- in-car networks



- facilities management



- biology (x2)



Four case studies of problem-driven work

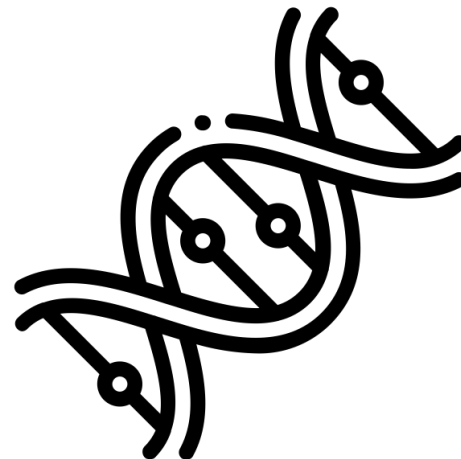
- in-car networks



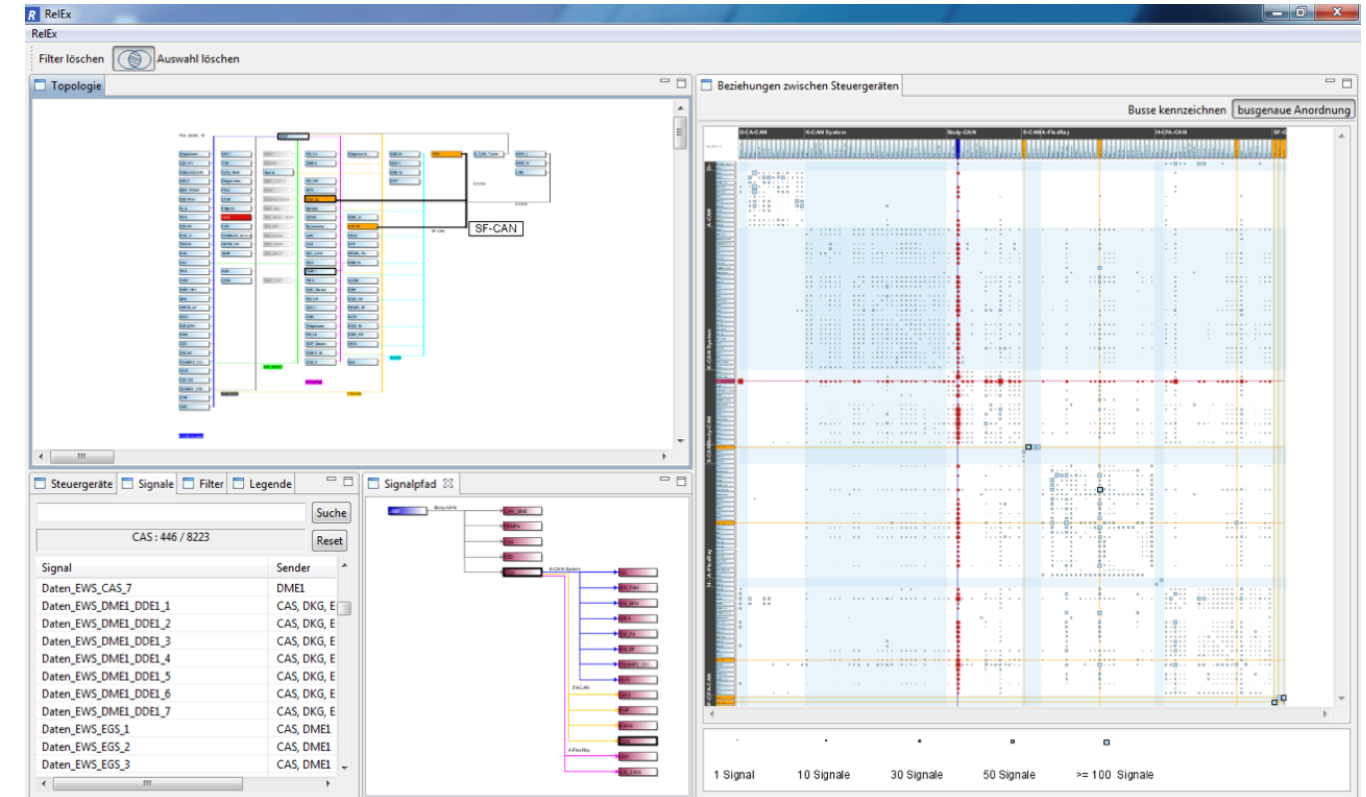
- facilities management



- biology



RelEx



Visualization for Actively Changing Overlay Network Specifications

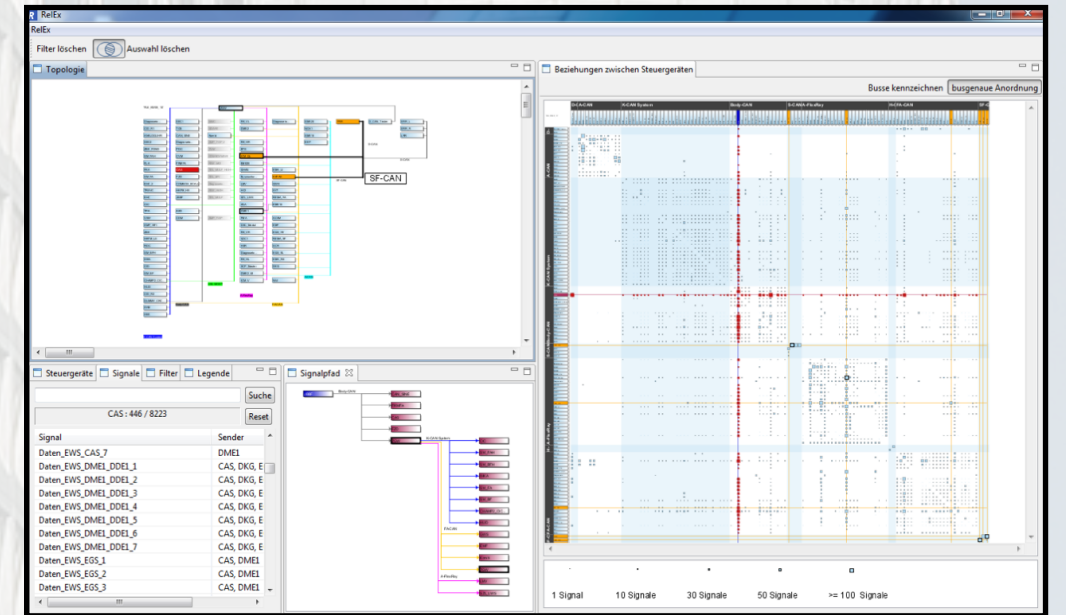
joint work with:

Michael Sedlmair, Annika Frank, Andreas Butz

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

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

Domain: **In-car network engineering**



Abstractions

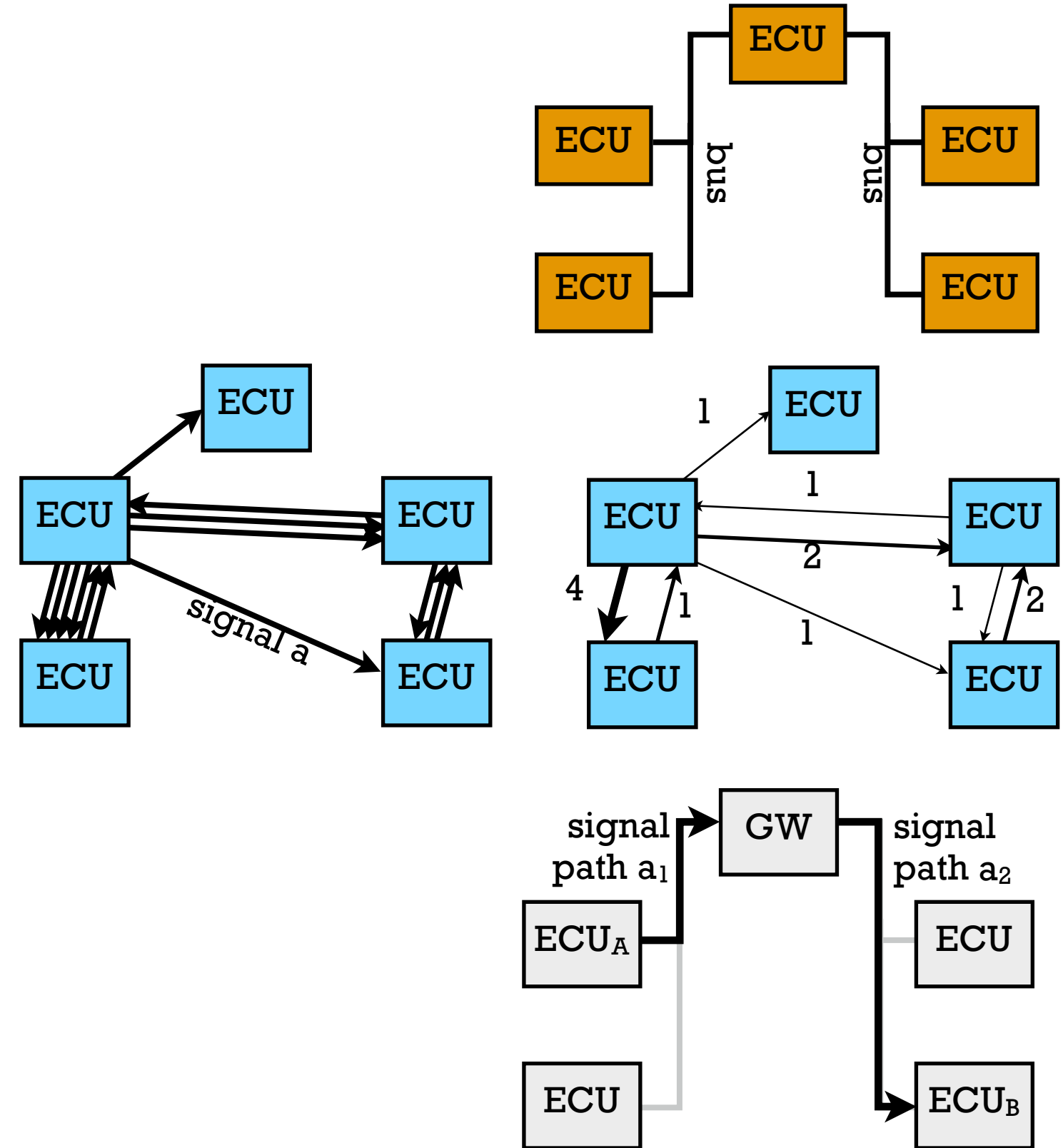
DATA

In-car Electronics



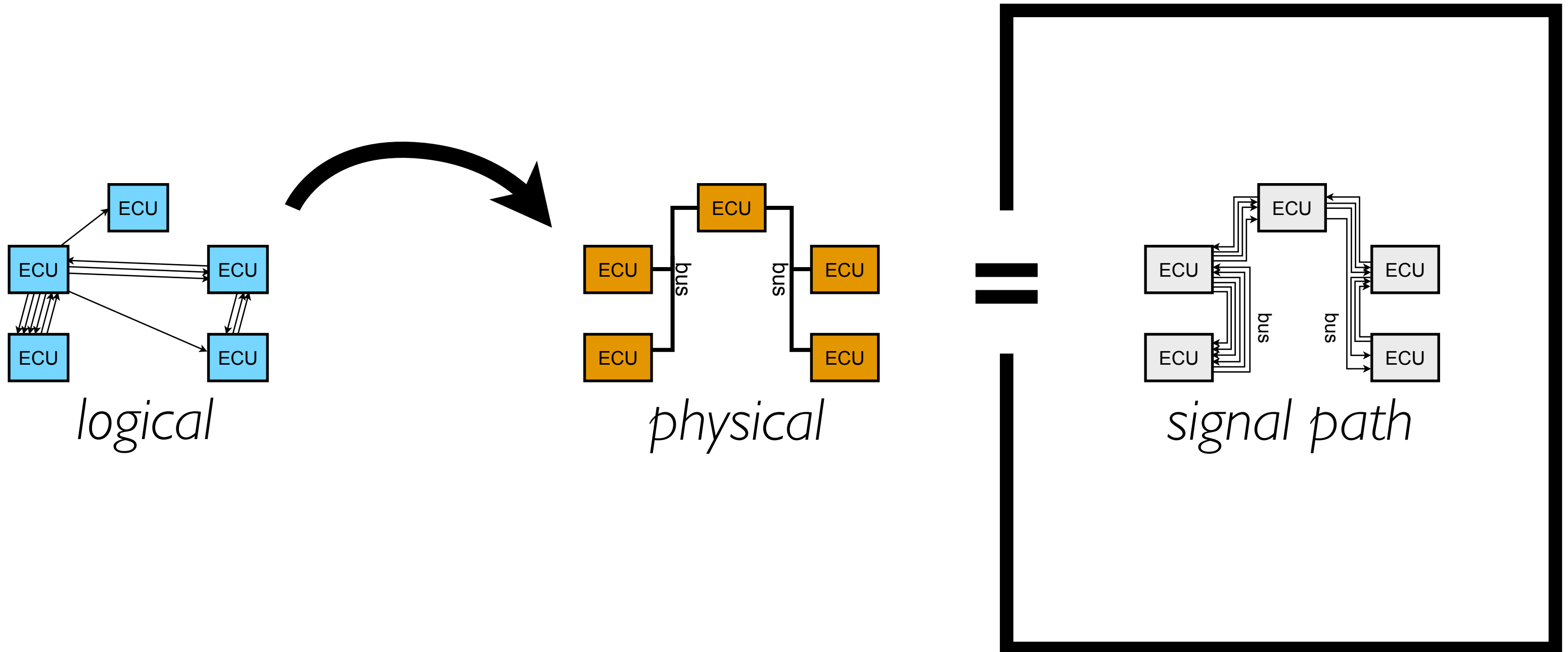
Data Abstraction: 3 Networks

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



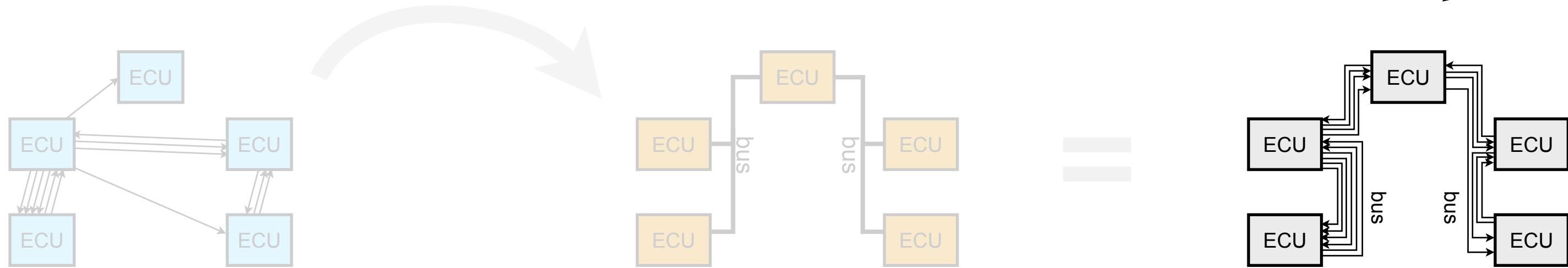
Task Abstraction: Mapping

- specify overlay network that maps logical onto physical



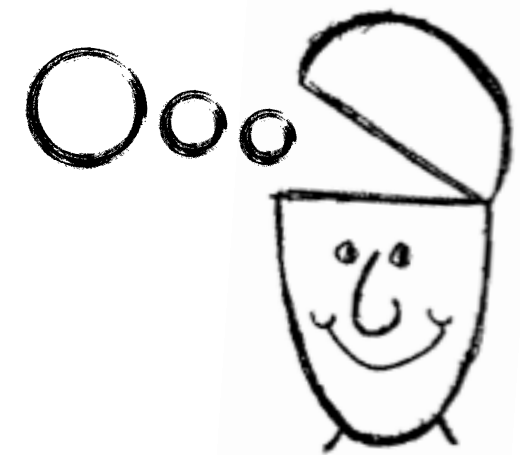
Task Abstraction: Optimizing

- traffic optimization



Many constraints

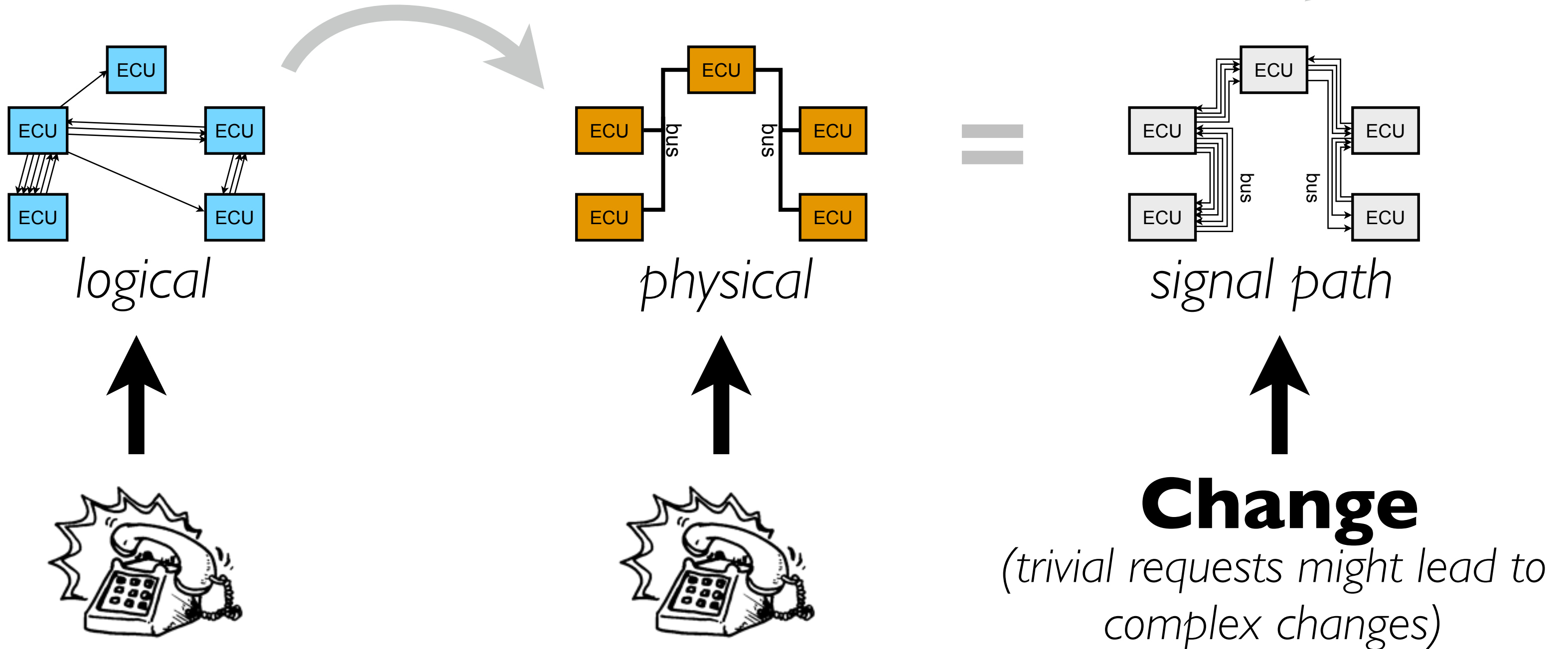
bandwidth ... delay/real time ...
path length ... load balance ...
reliability ... money ...



-- engineer, BMW --

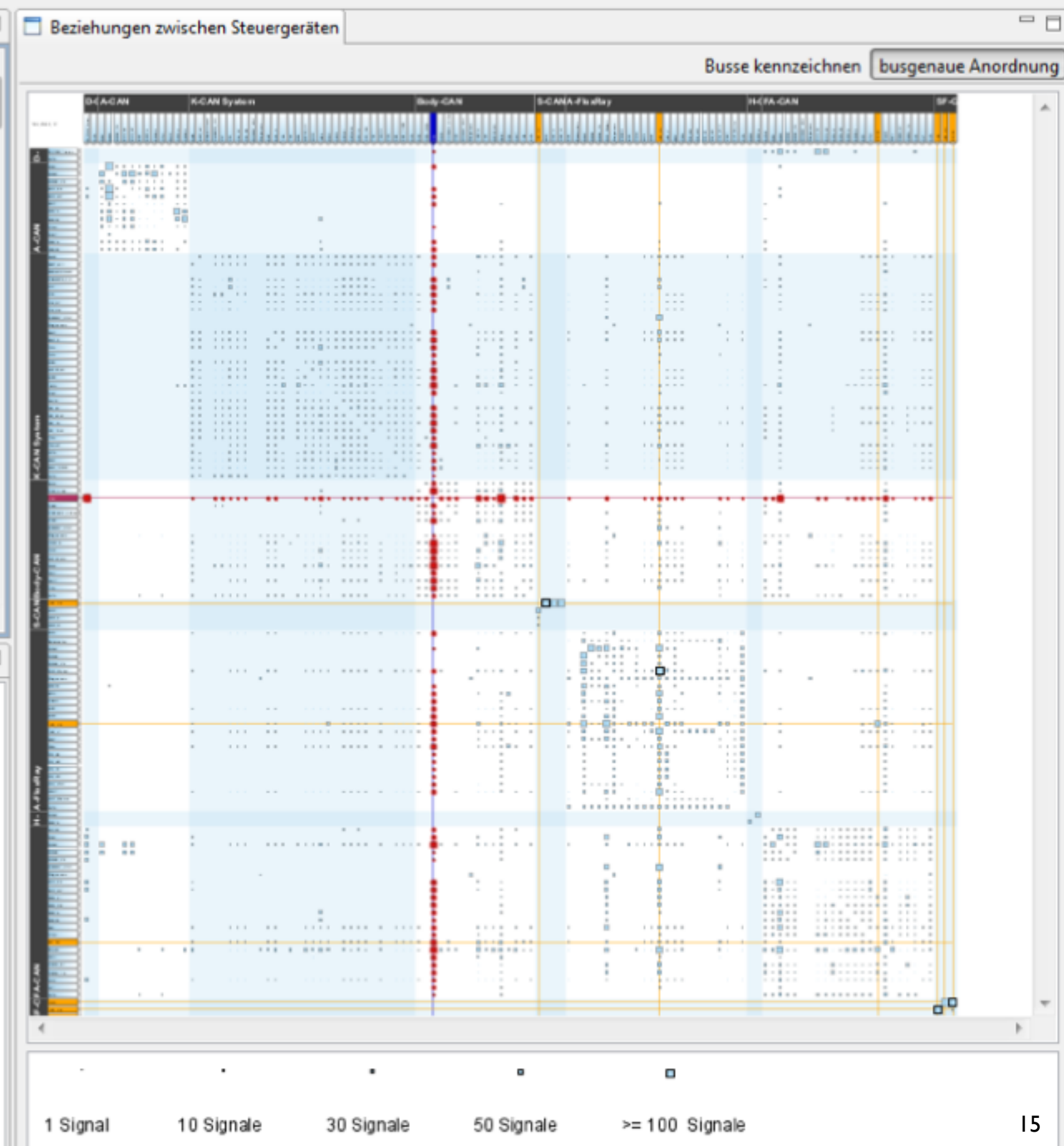
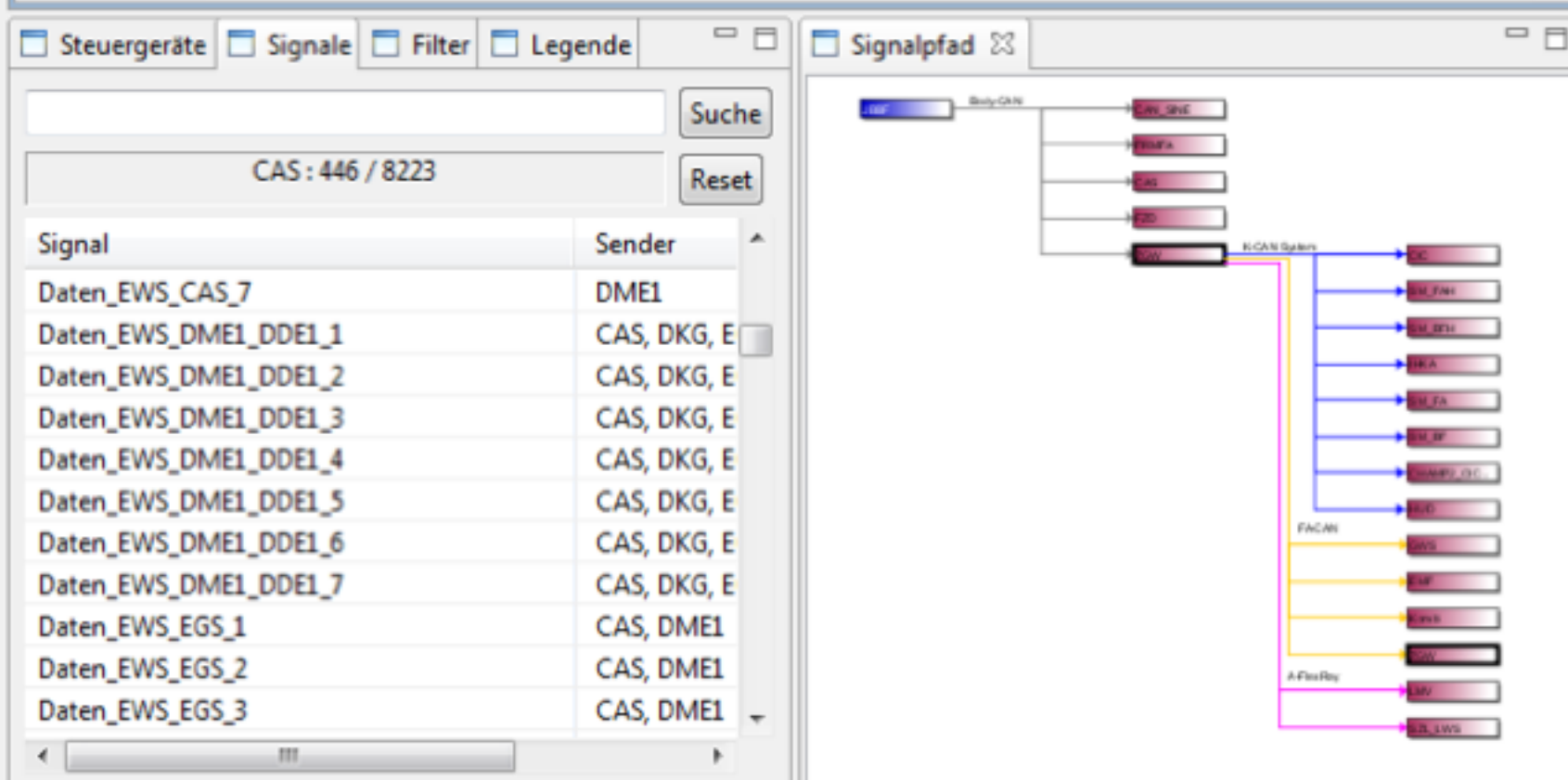
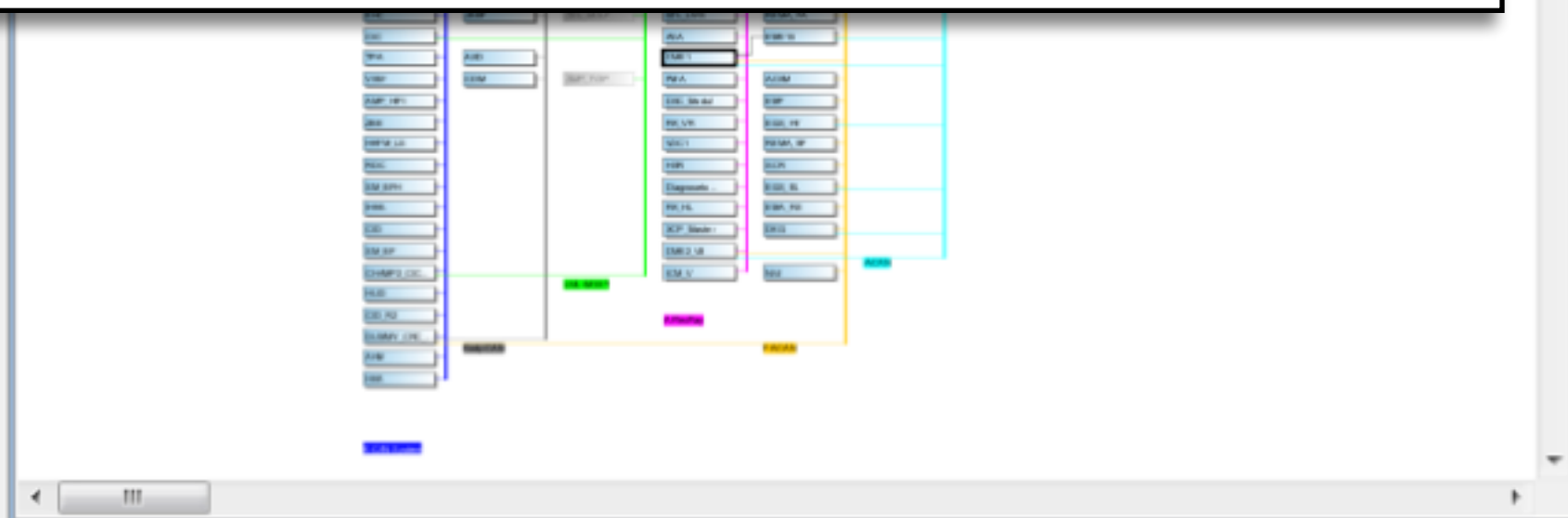
Task Abstraction: Changing

- external change requests



Idioms

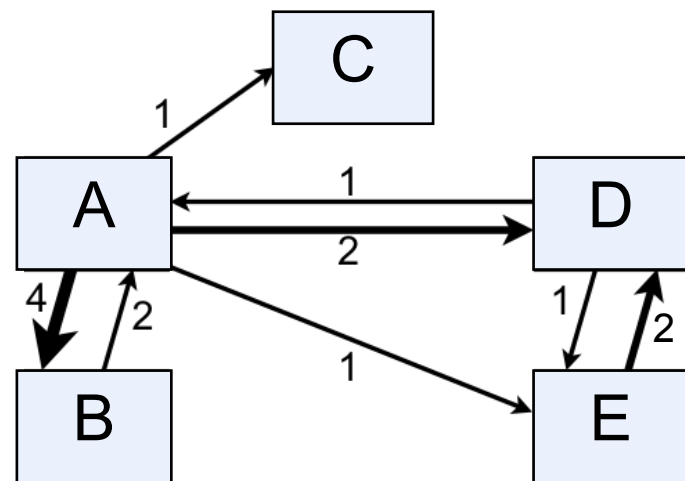
RELEX: Relation Explorer



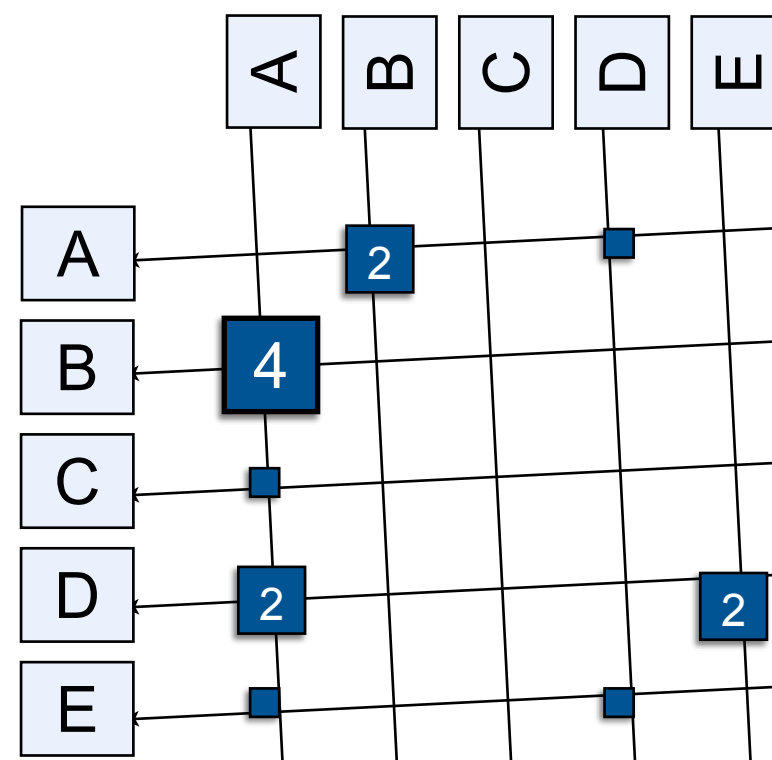
Vis Guideline [Ghoniem 2005]

Matrix for dense network data

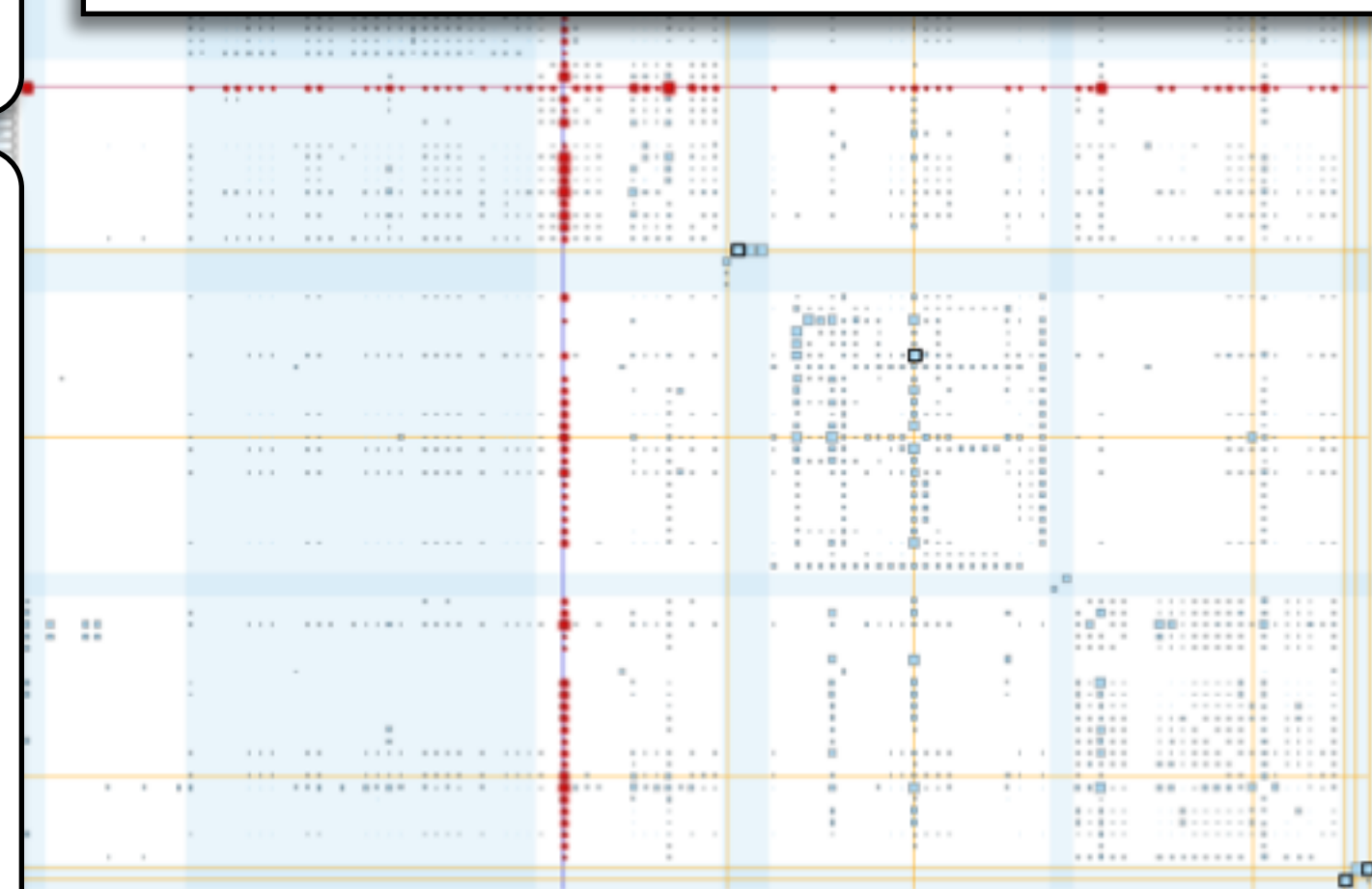
SIGNAL COUNT NETWORK



visual encoding:
size-coded matrix



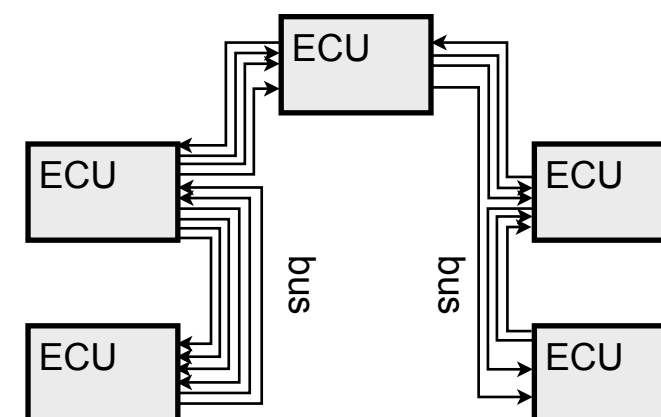
Logical Network View: Overview



Node-link for path following tasks

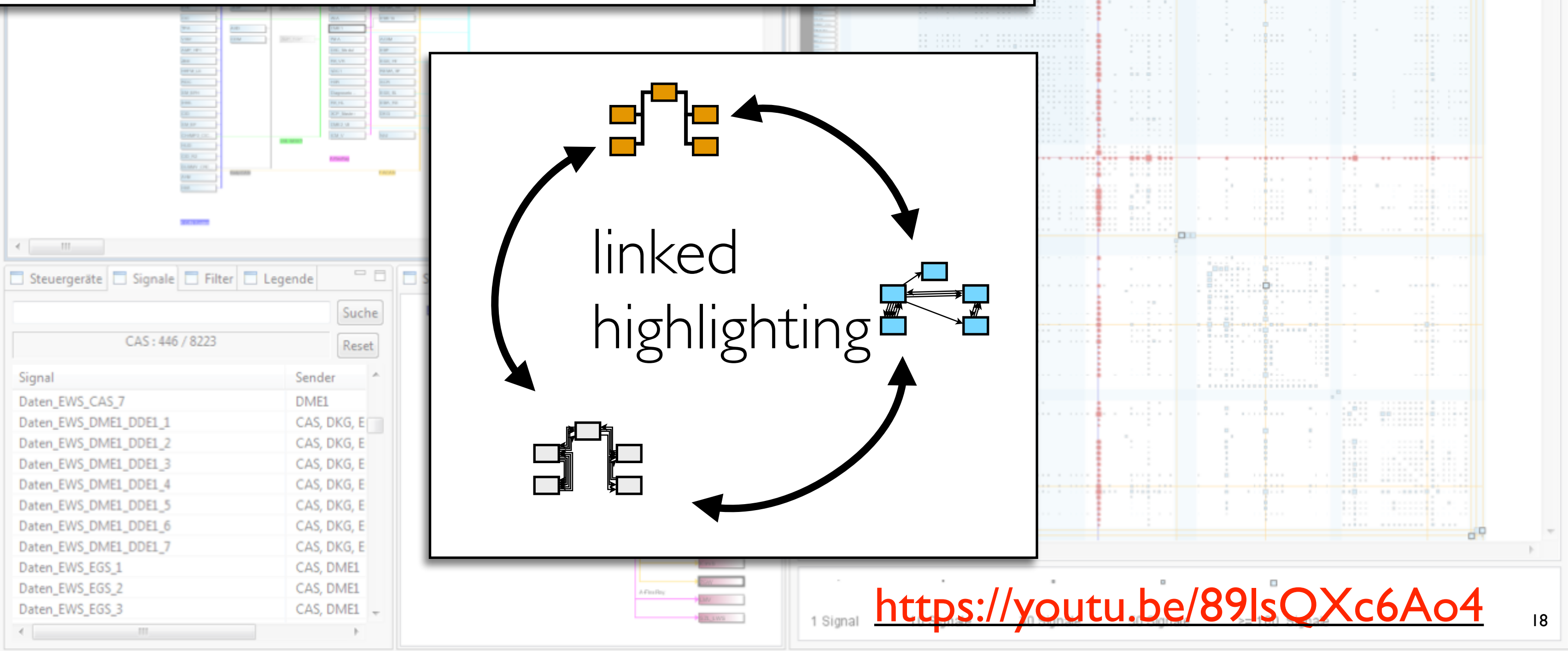
SIGNAL PATH NETWORK

Signal Path View: Selected Signal



filtered by signal

INTERACTION IDIOM: Cross-Network Relations

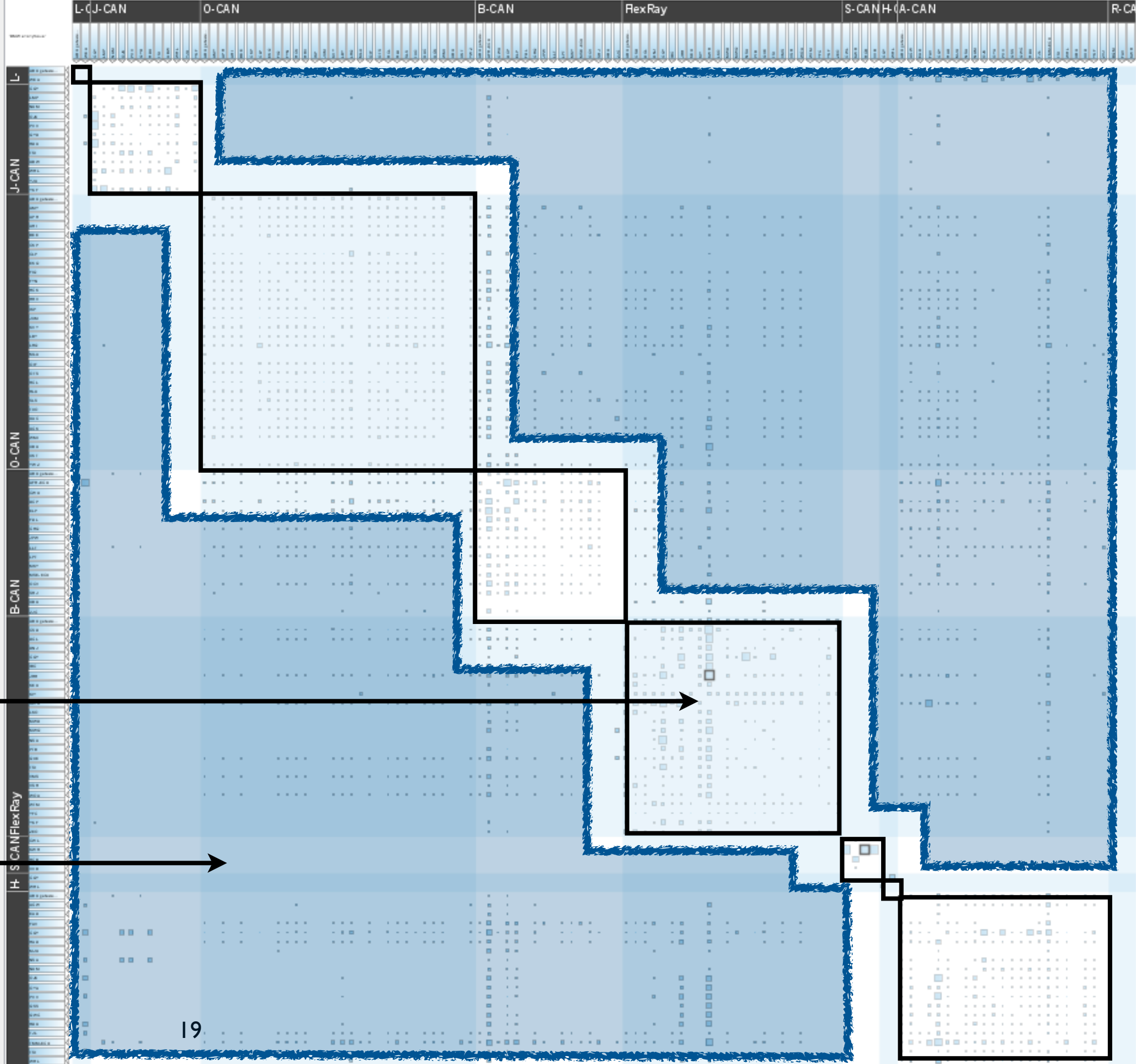


INTERESTS

Bus communication patterns

Within-bus

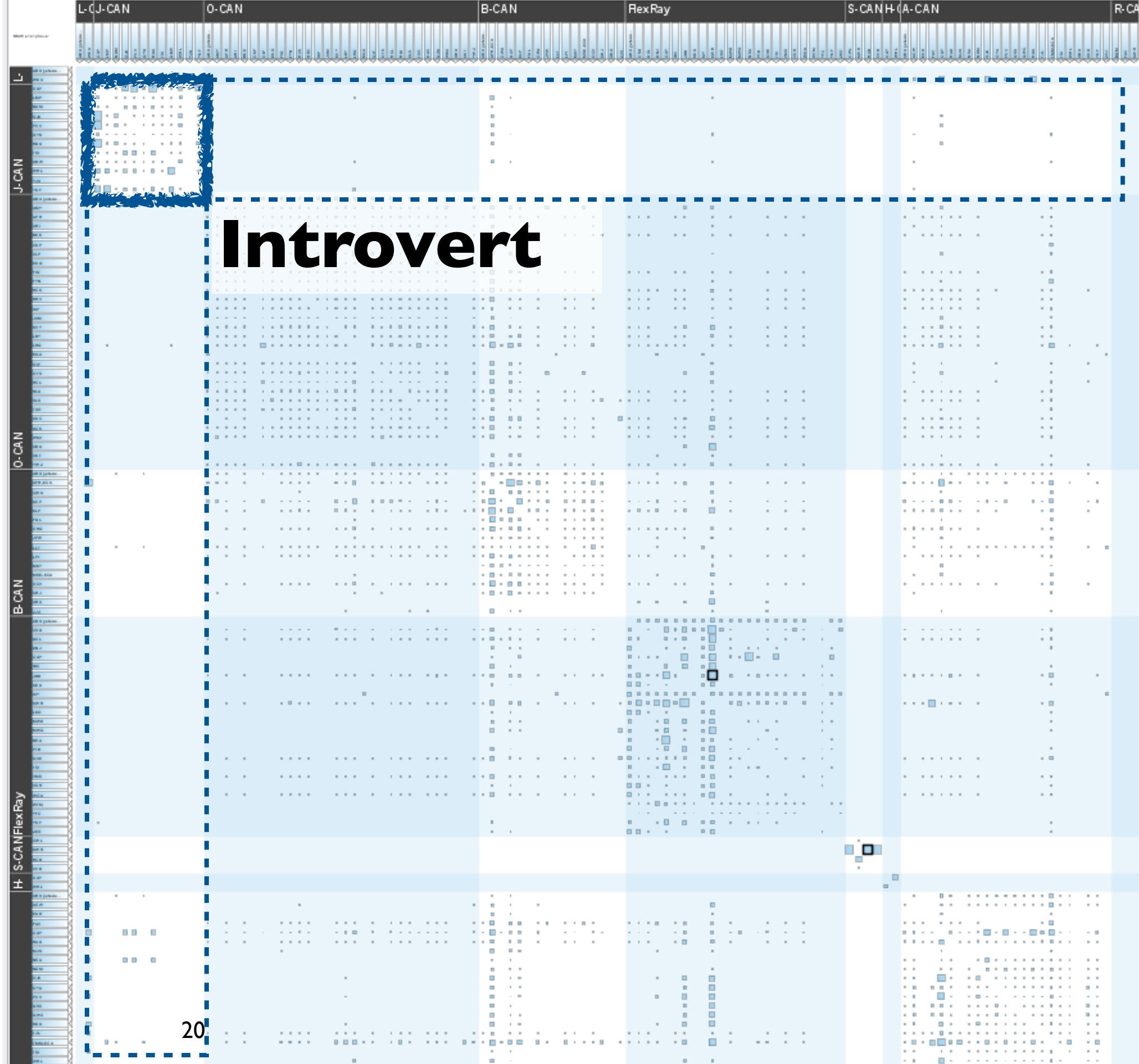
Between-bus



INTERESTS

Bus communication patterns

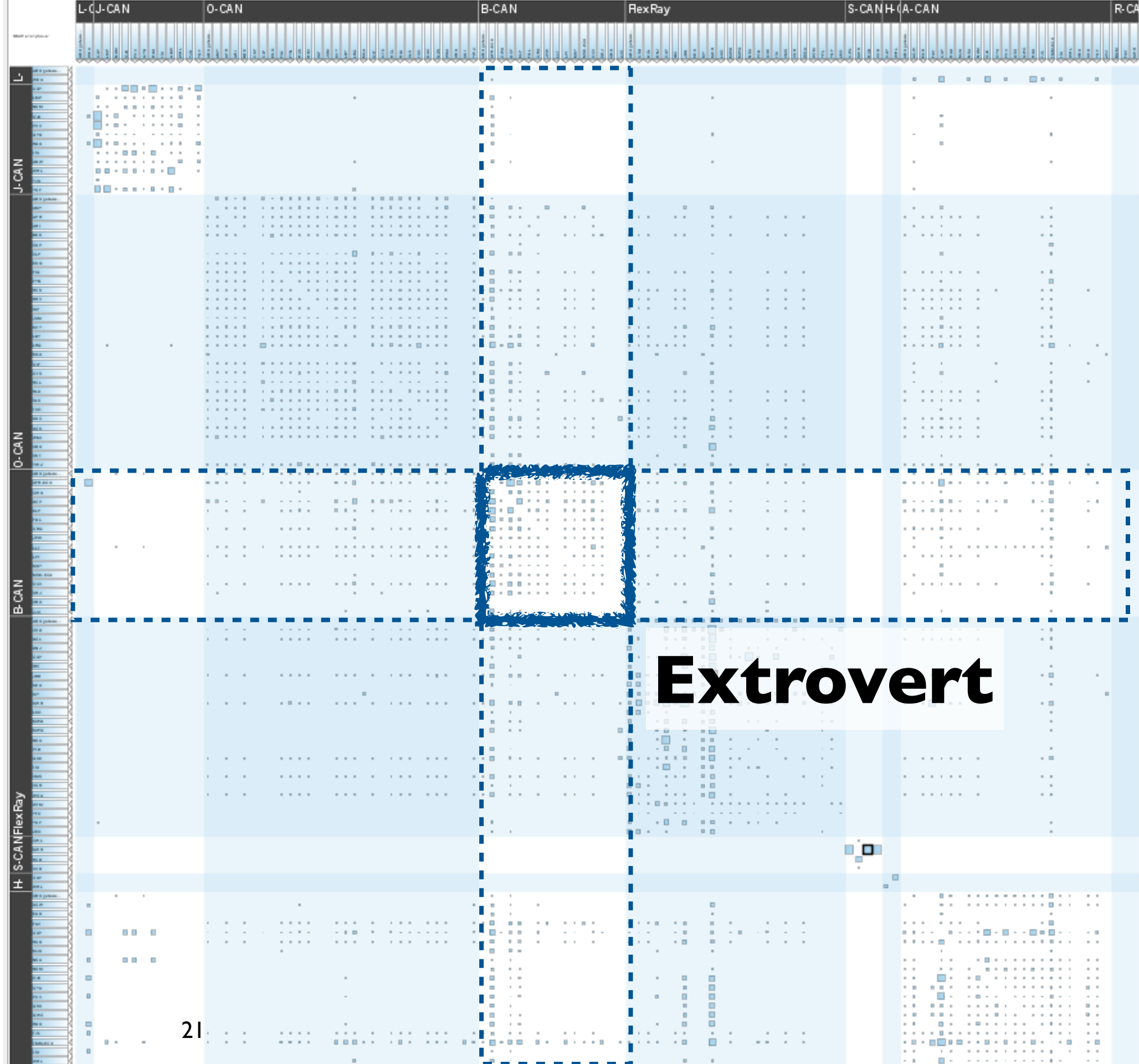
introvert
vs.
extrovert



INTERESTS

Bus communication patterns

introvert
vs.
extrovert



Abstraction Innovation

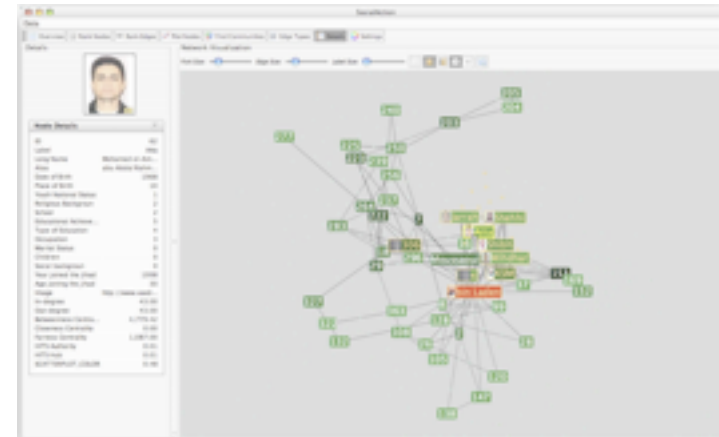
Previous Work

Focus on social network analysis

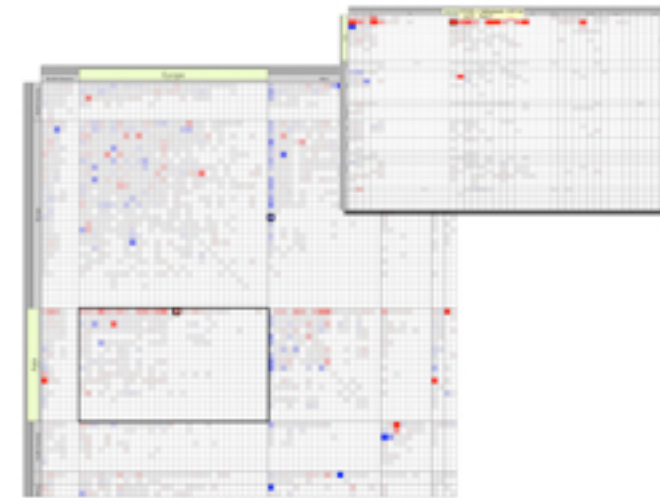
- radically different task and data abstractions



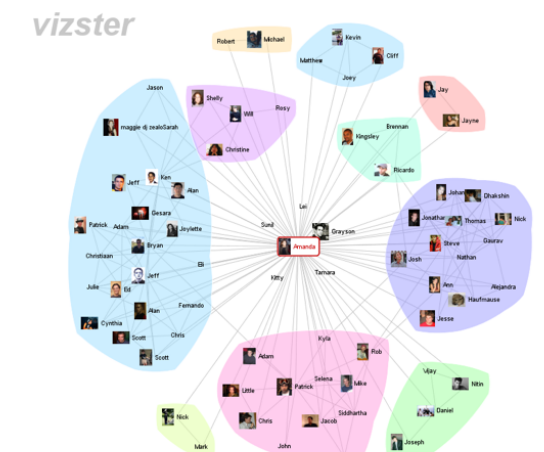
MatrixExplorer



SocialAction



Honeycomb



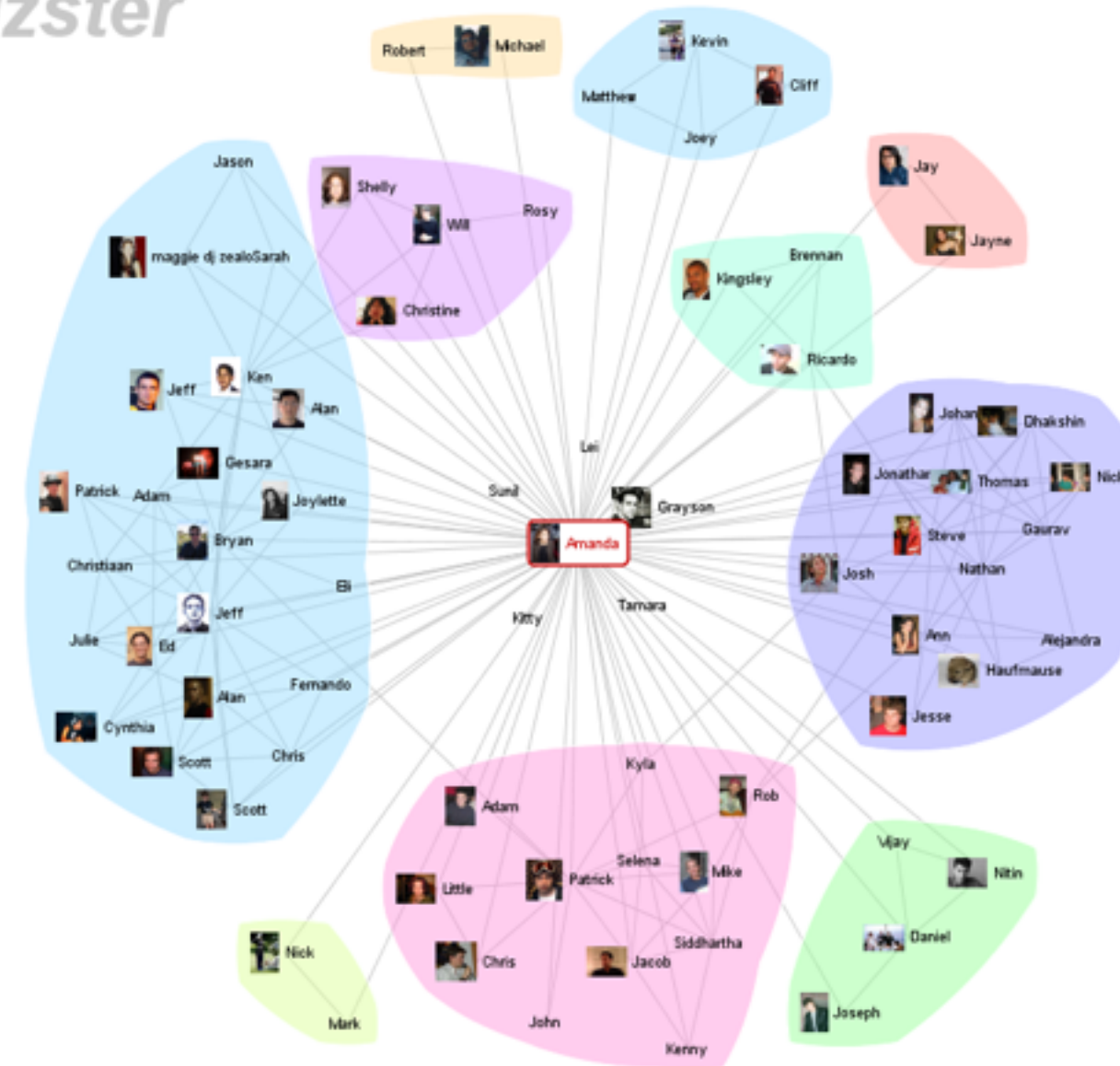
vizster

Task Abstraction

Social Network Analysis Domain

- find clusters

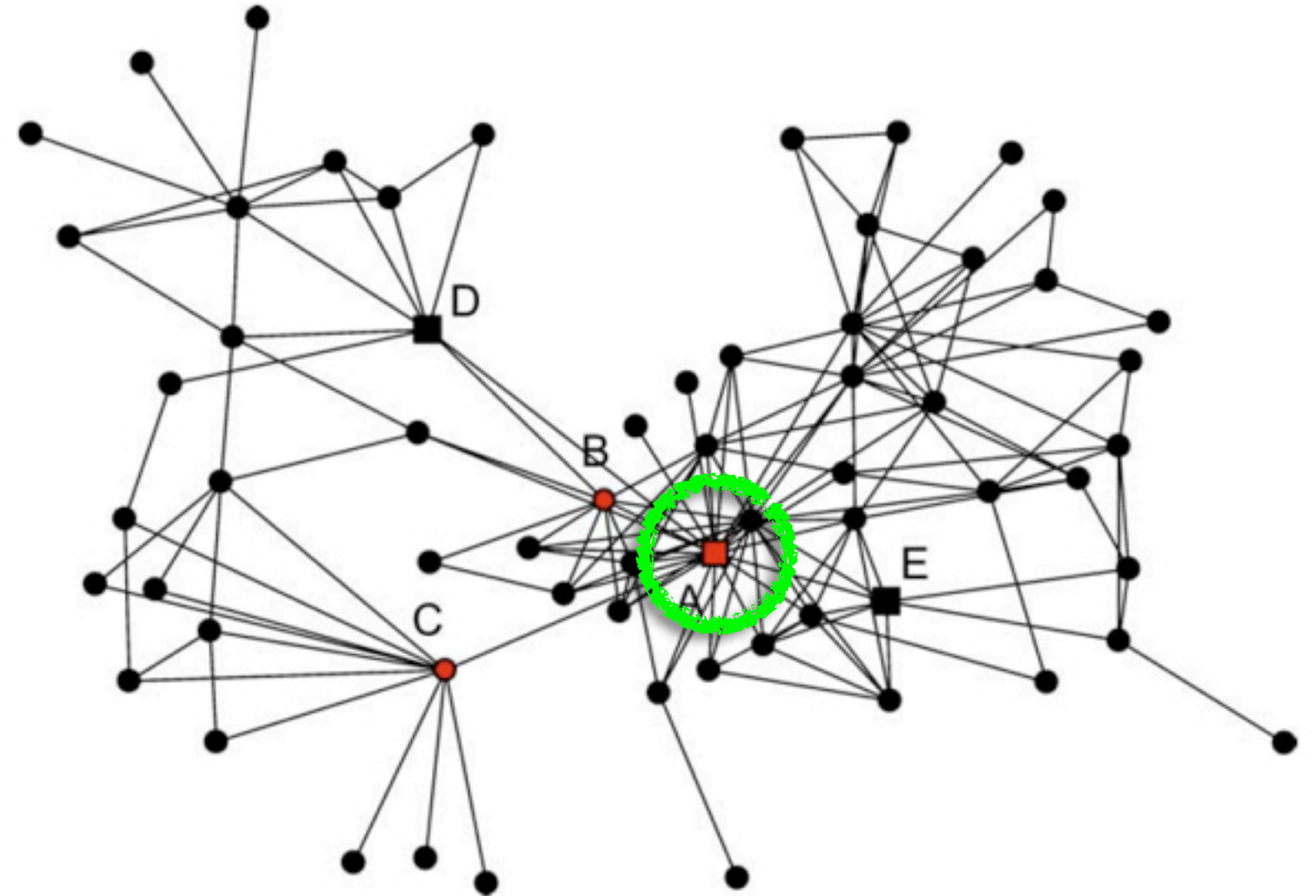
vizster



Task Abstraction

Social Network Analysis Domain

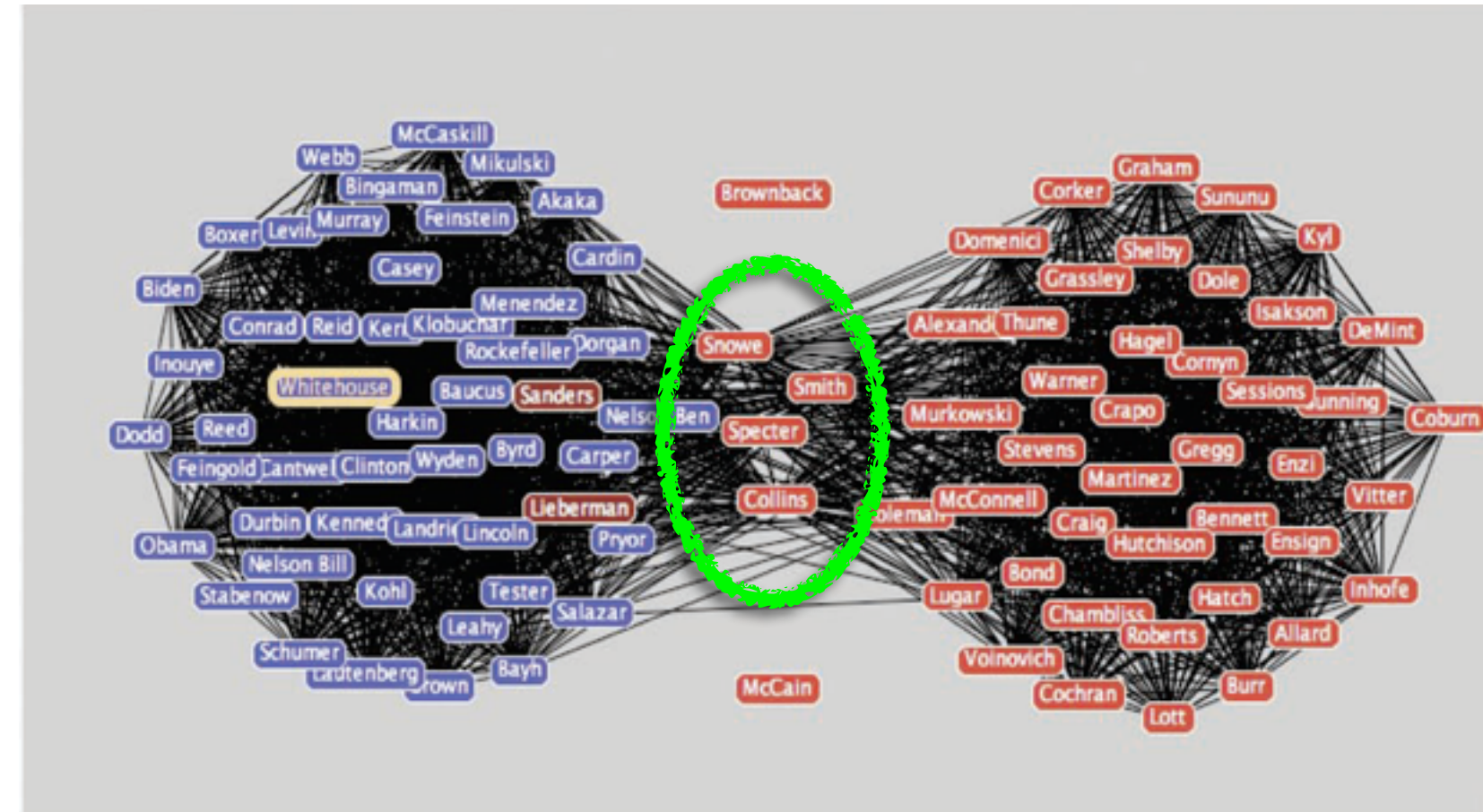
- find clusters
- find high-degree nodes



Task Abstraction

Social Network Analysis Domain

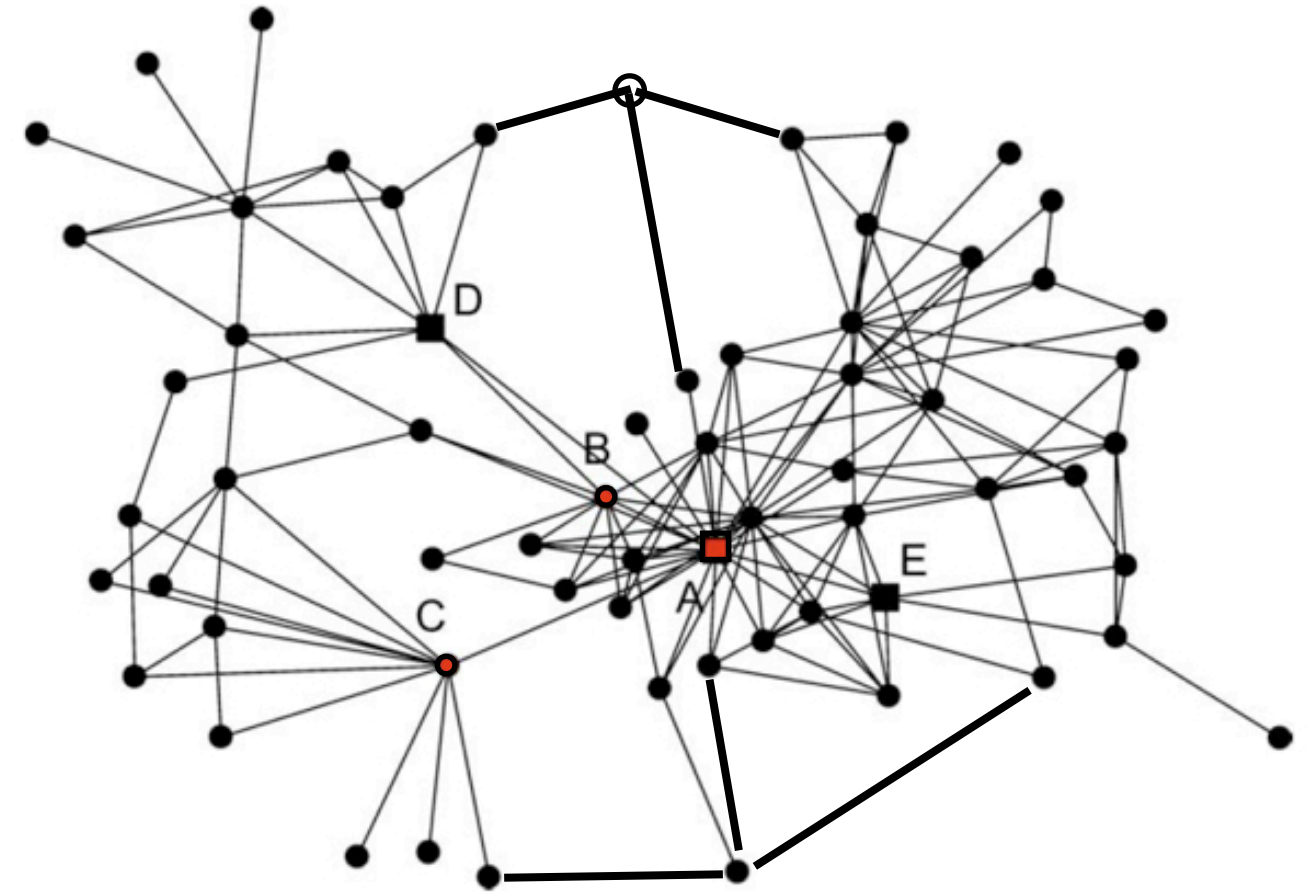
- find clusters
- find high-degree nodes
- find bridge nodes



Task Abstraction

Social Network Analysis Domain

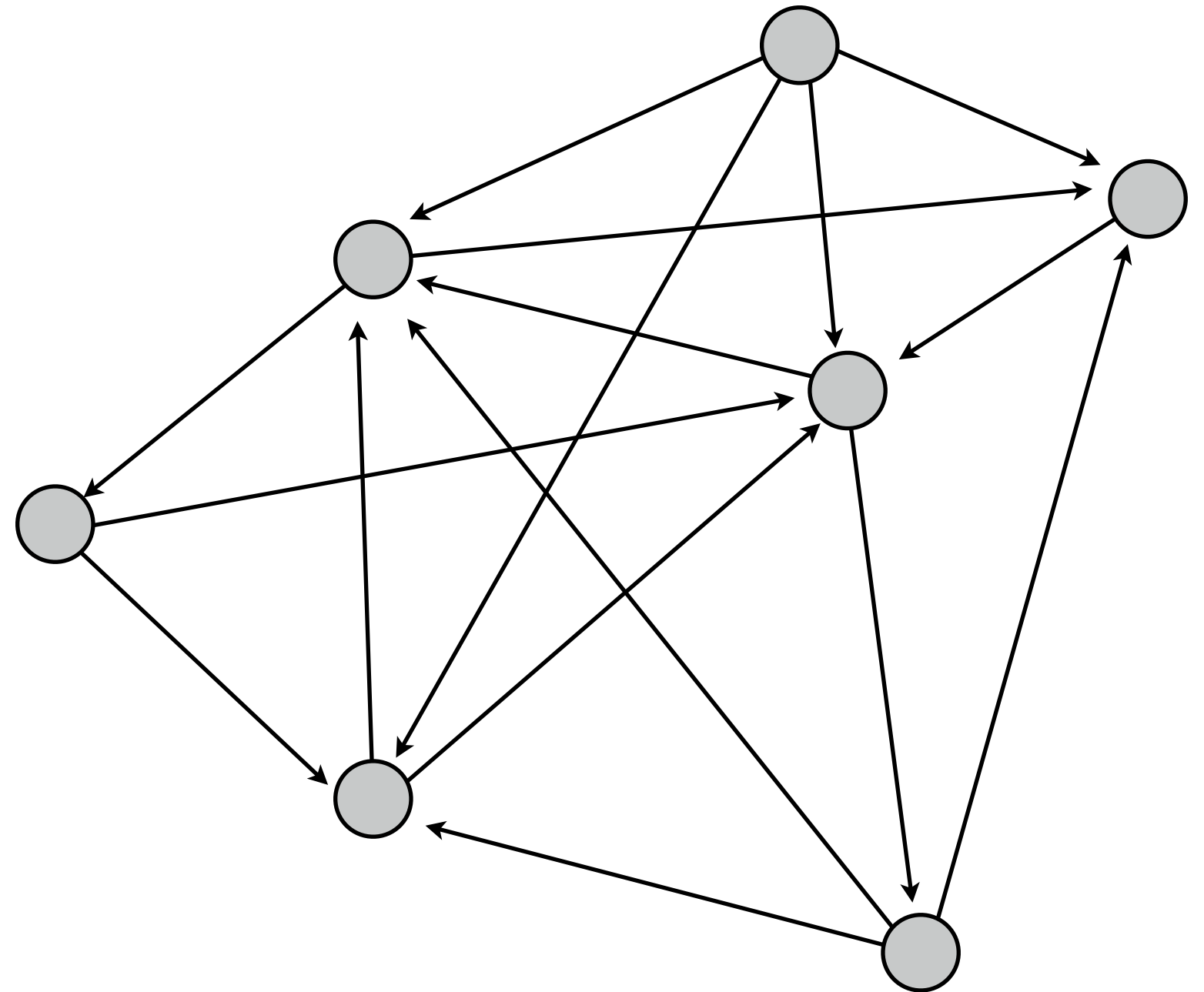
- find clusters
- find high-degree nodes
- find bridge nodes
- understand temporal dynamics
 - passively notice changes



Data Abstraction

Social Network Analysis Domain

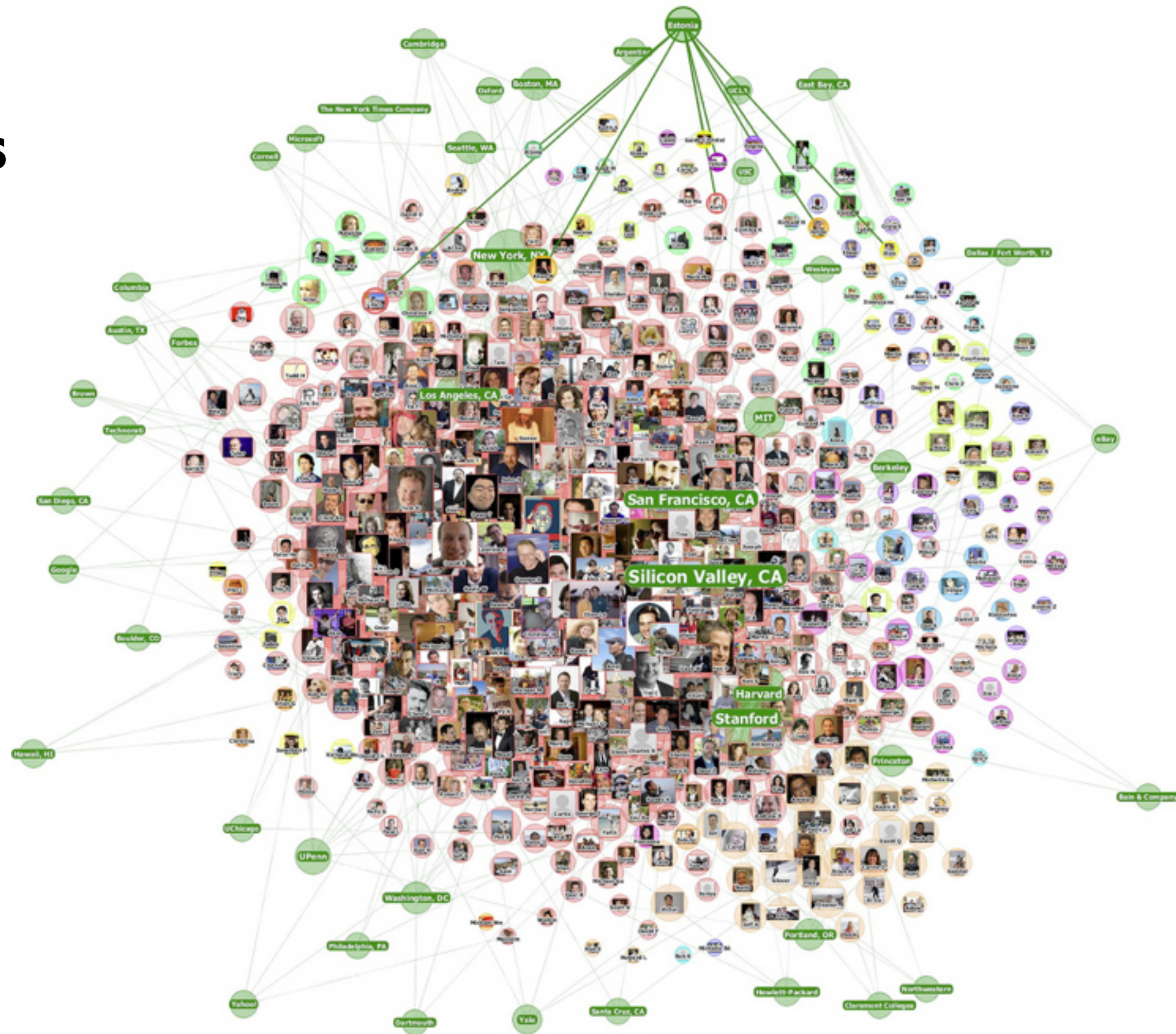
- single graph



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 

Q&A

Four case studies of problem-driven work

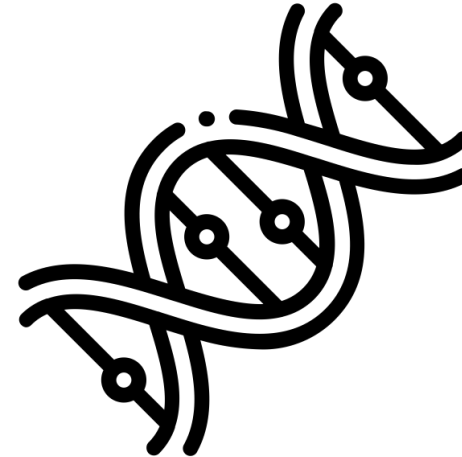
- in-car networks



- facilities management



- biology





Michael
Oppermann



Ocupado

Visualizing Location-Based Counts Over Time Across Buildings

<http://www.cs.ubc.ca/labs/imager/tr/2020/ocupado/>

Ocupado: Visualizing Location-Based Counts Over Time Across Buildings.

Oppermann and Munzner. *Computer Graphics Forum (Proc. EuroVis 2020)* 39(3):127-138 2020.

Location-Based Counts





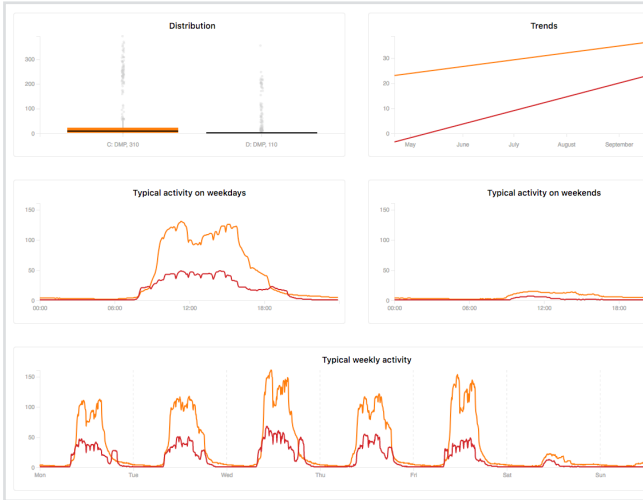
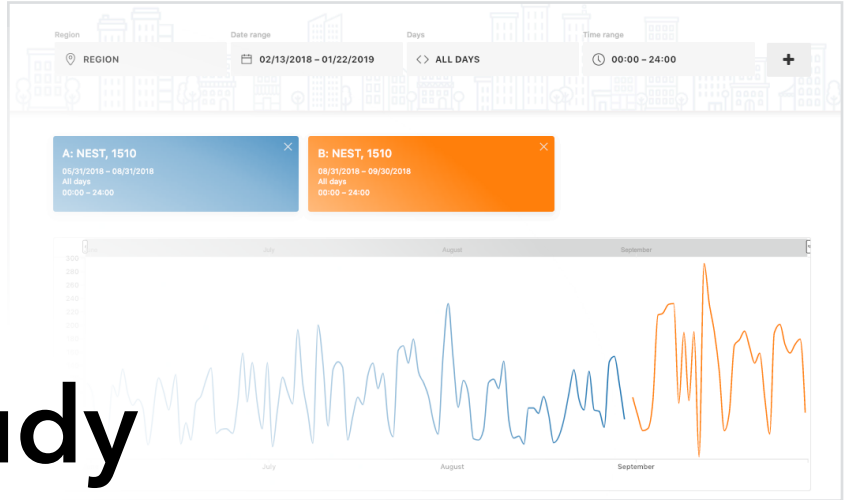
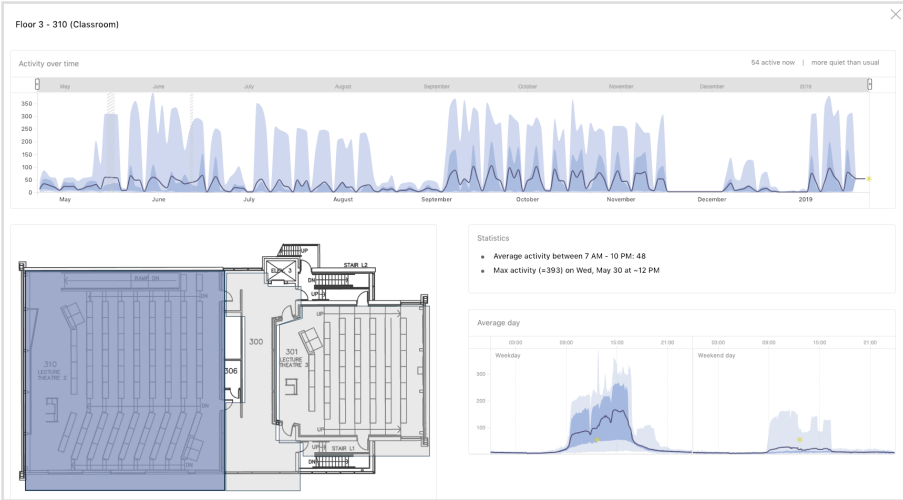
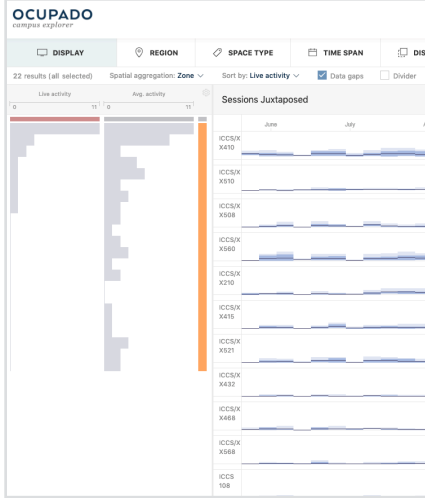
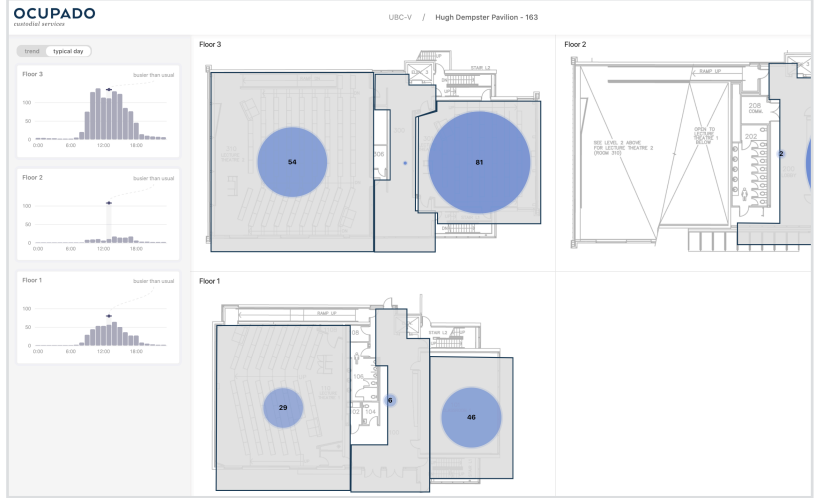
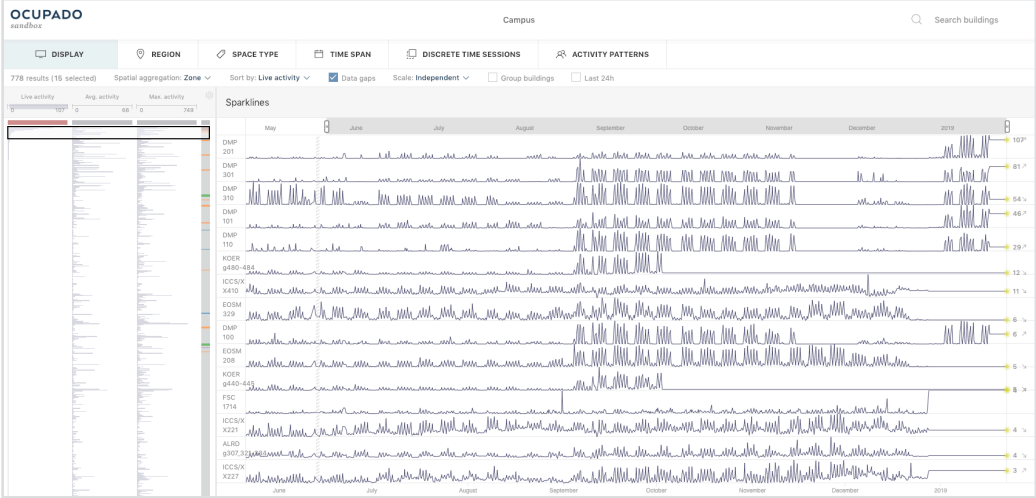
**Previous measurement required
physical counting or installation
of additional hardware.**



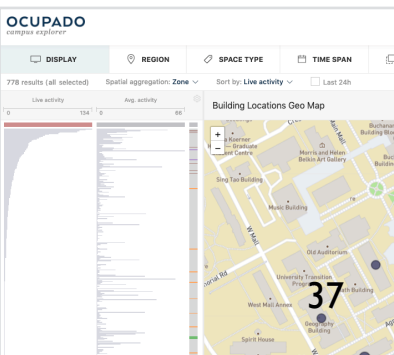
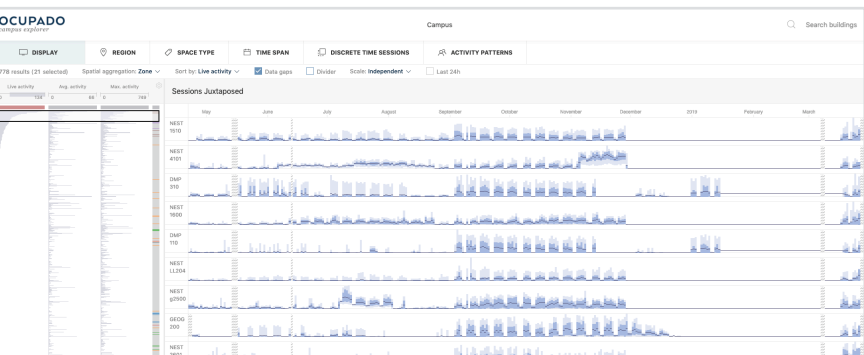
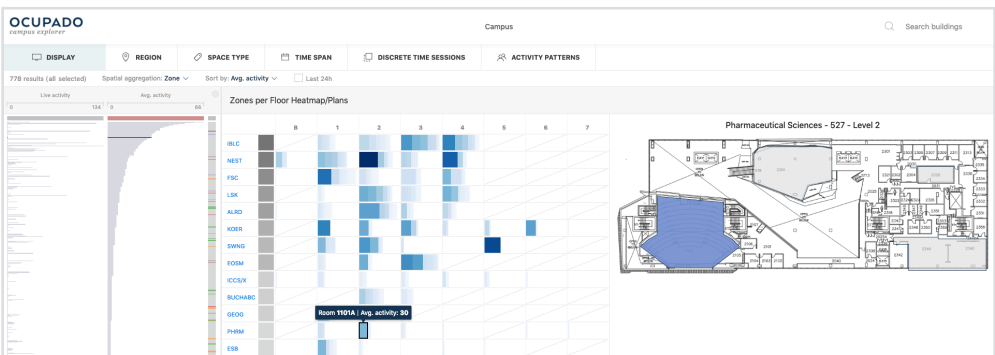
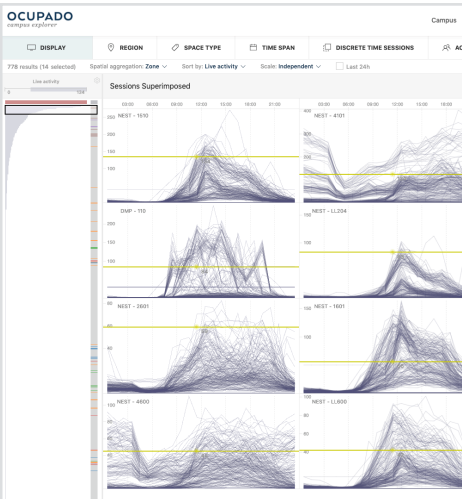
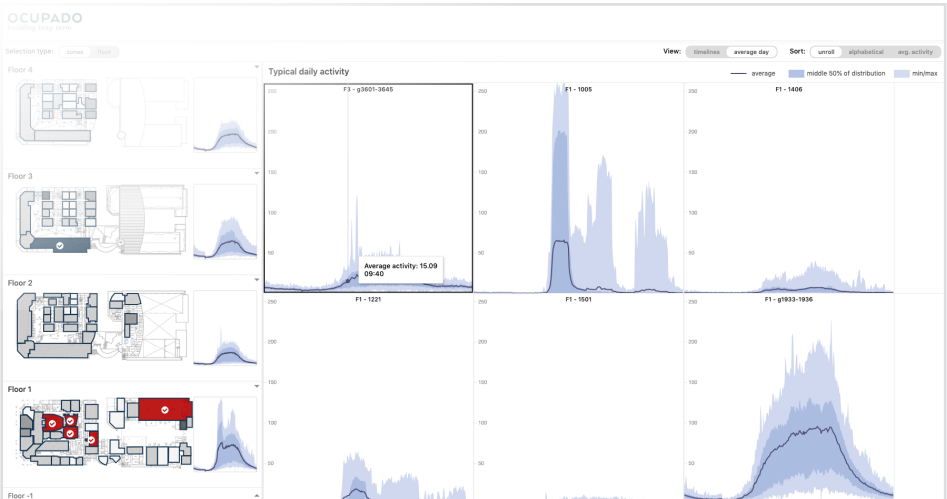
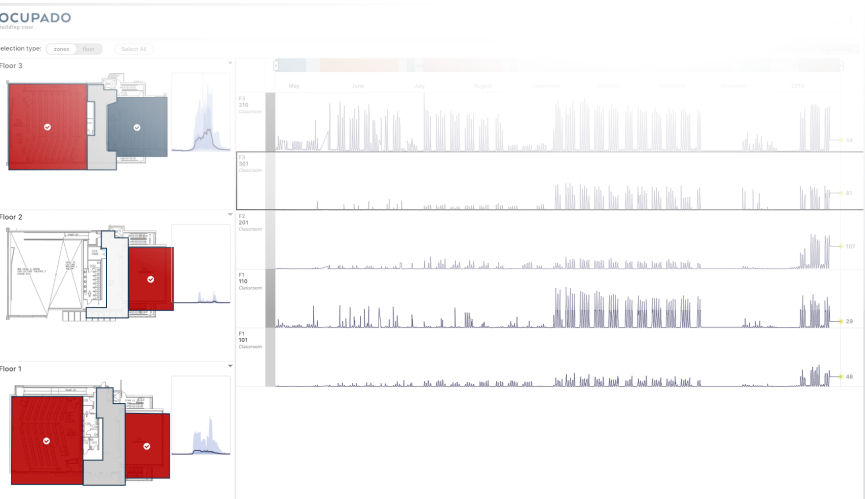
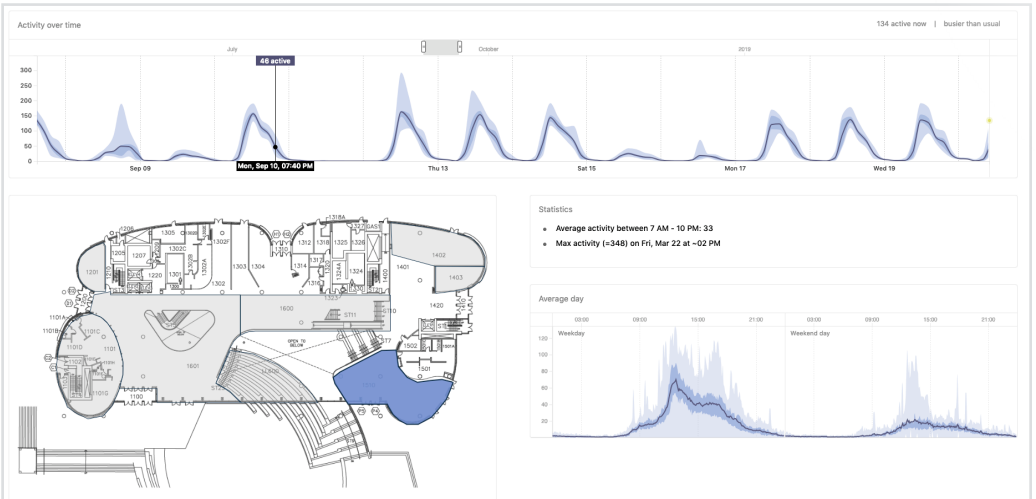
Previous measurement required physical counting or installation of additional hardware.



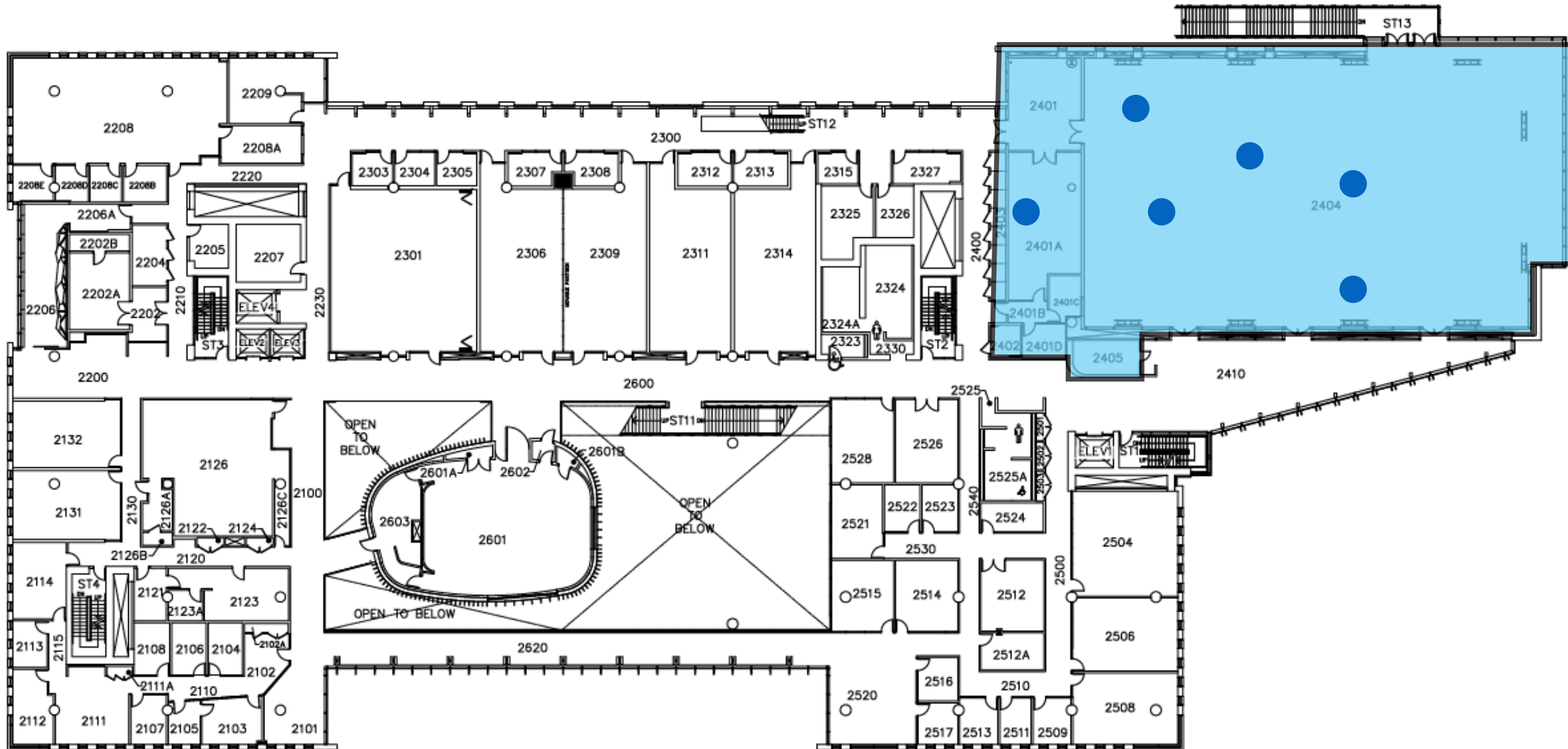
Previous visualization attempts were limited in space and time.



Design Study



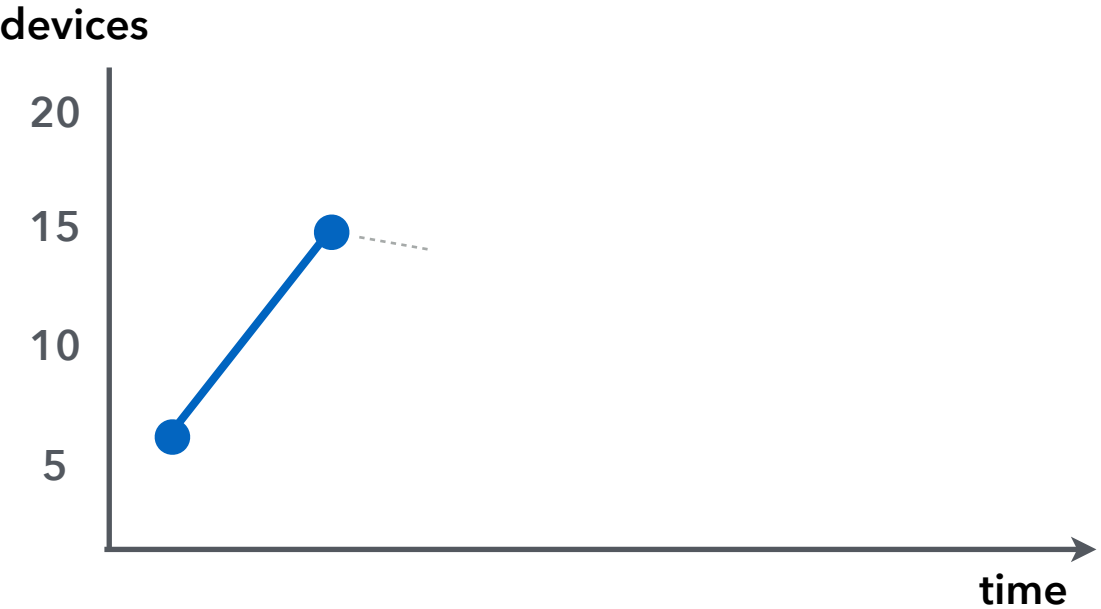
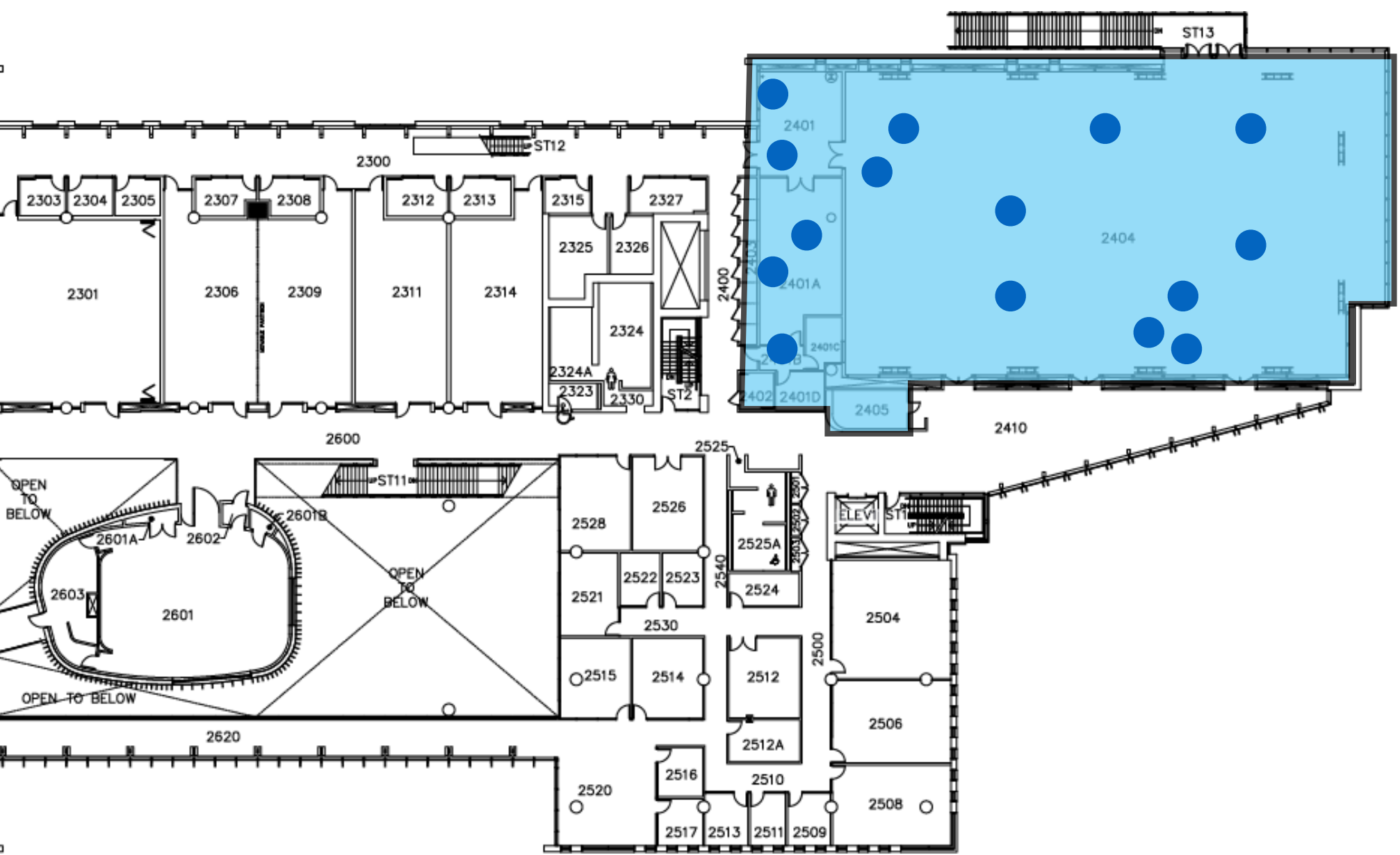
WiFi Connections: Location-Based Counts



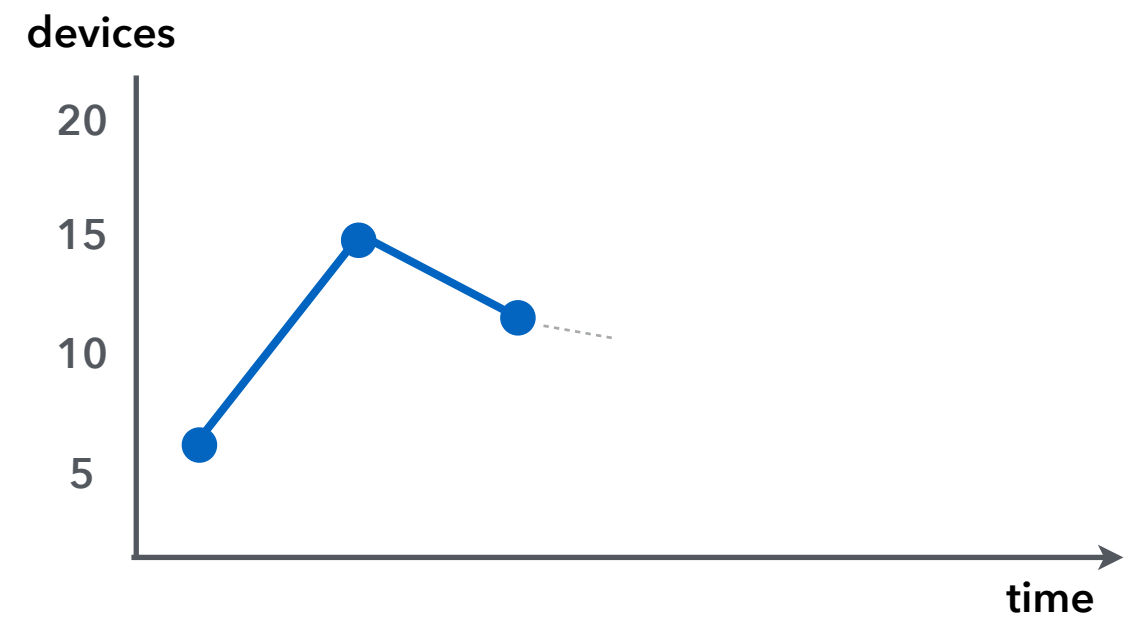
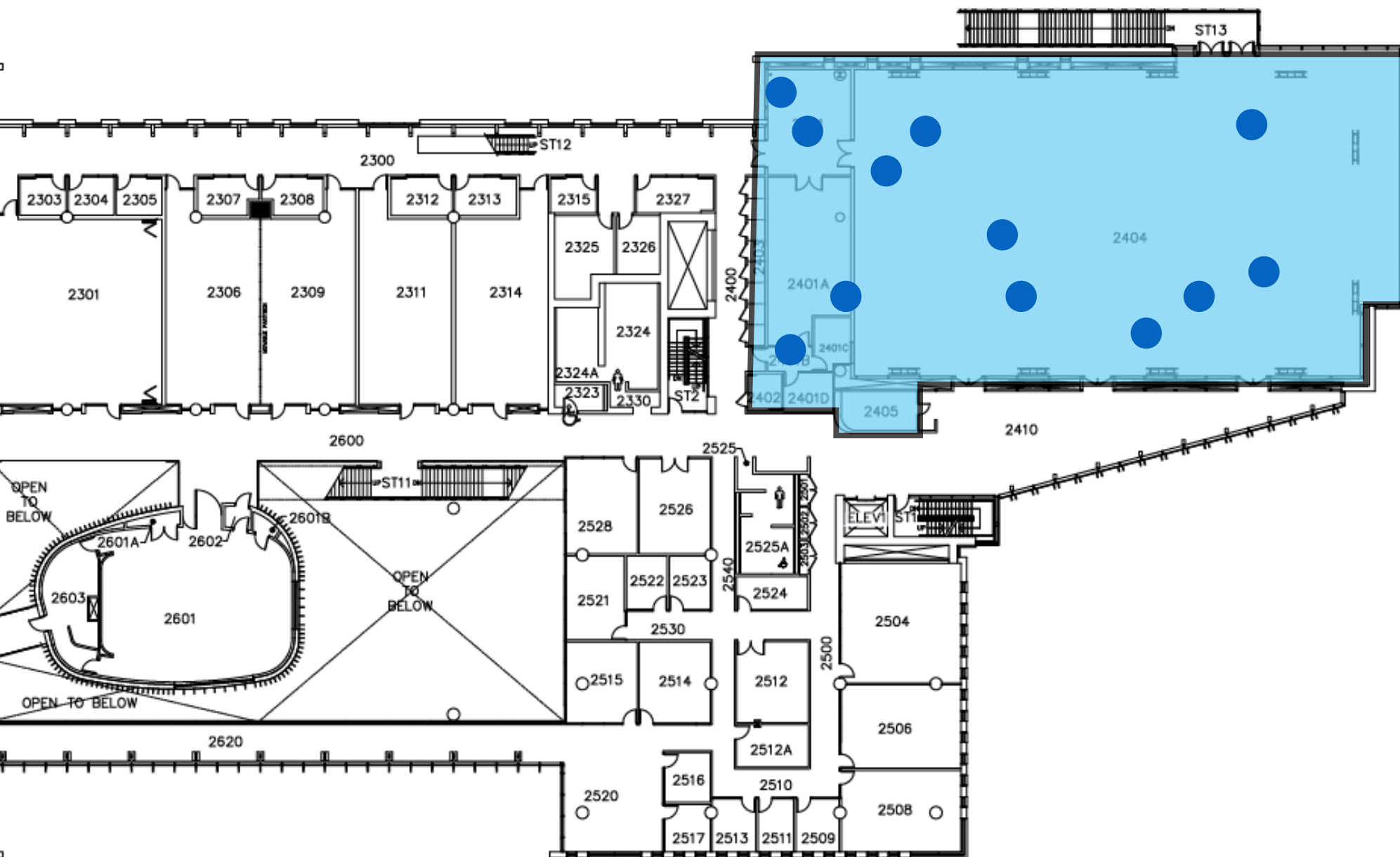
WiFi Connections: Location-Based Counts



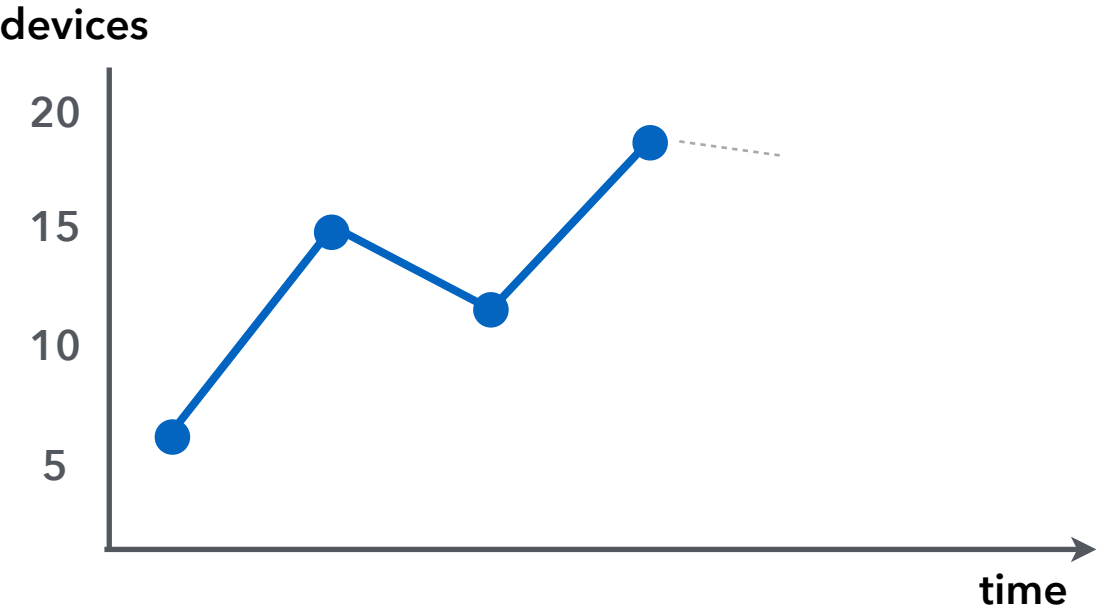
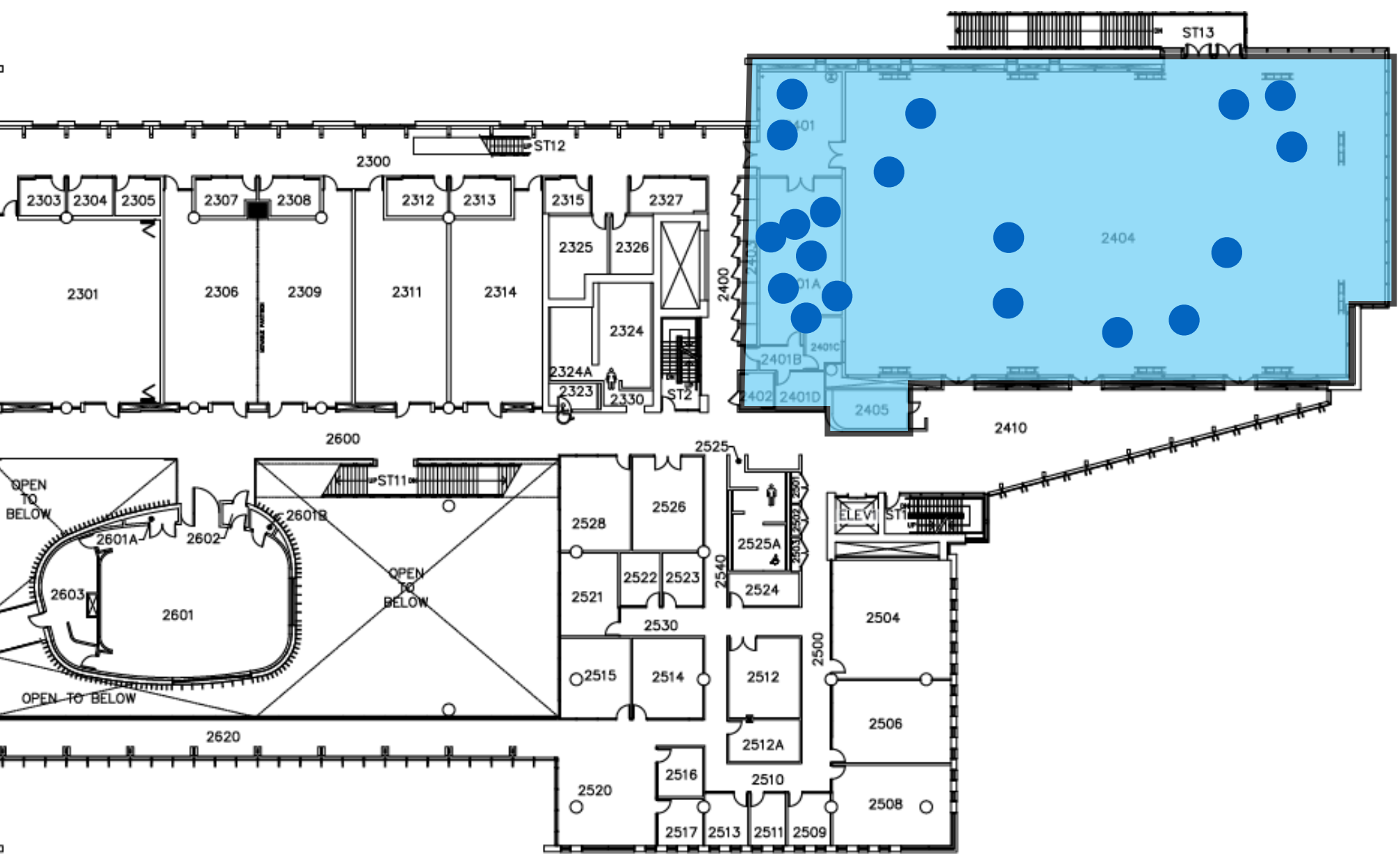
WiFi Connections: Location-Based Counts



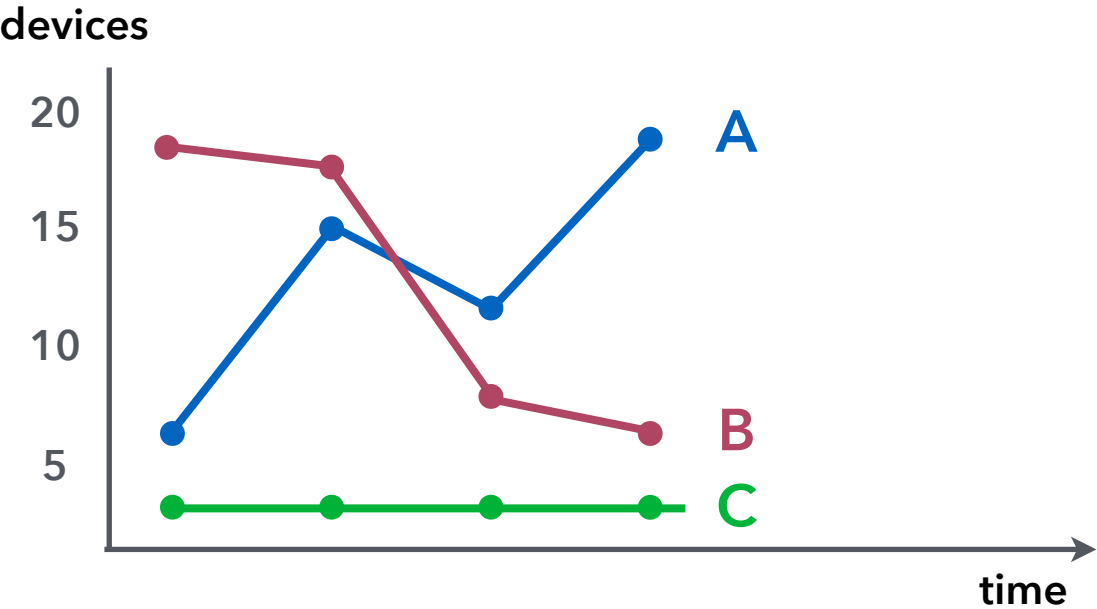
WiFi Connections: Location-Based Counts



WiFi Connections: Location-Based Counts



WiFi Connections: Location-Based Counts



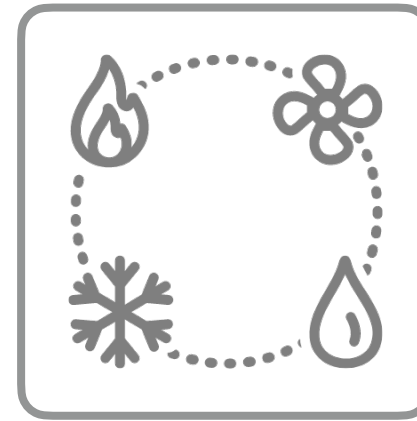
WiFi Connections: Location-Based Counts



Location-Based Counts

- Regular intervals (e.g., every 5 minutes)
- Spatial hierarchy (Zone → Floor → Building → Campus)
- No trajectories or device identifiers are recorded
- Intrinsic privacy advantages

Data



**Automated
HVAC control**

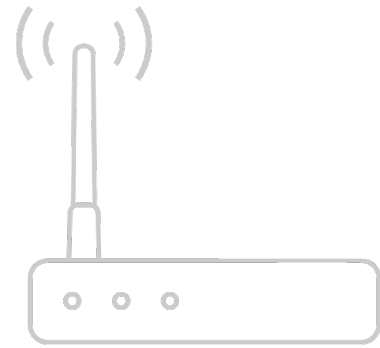
Data



Data



**Decision
making**

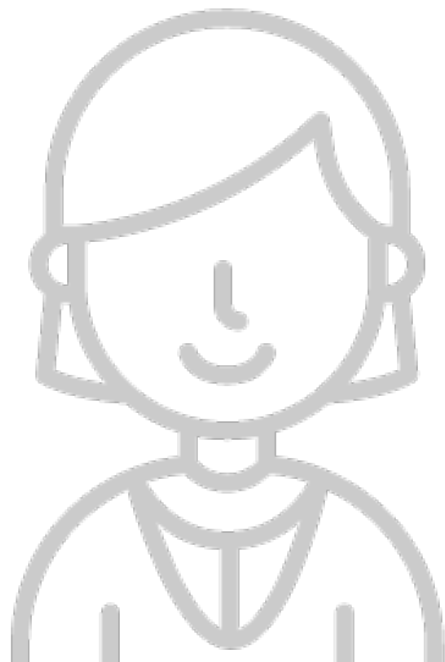
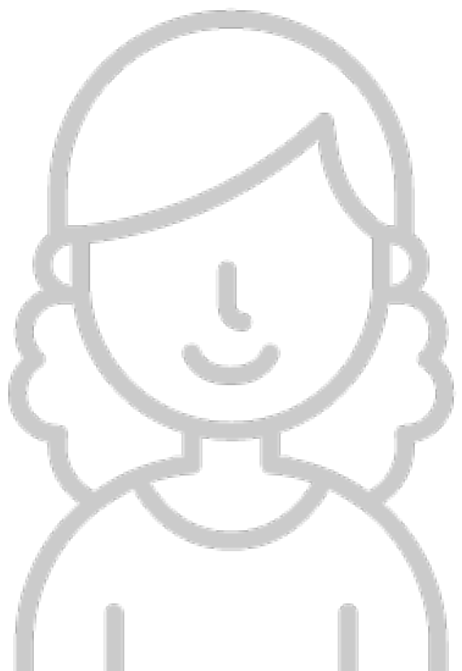


WiFi connections as a proxy for occupancy



WiFi connections as a proxy for occupancy

Interviews with potential stakeholders



Focus Domains

- **Space planning**
- **Building management**
- **Custodial services**
- **Classroom management**
- **Data quality control**

Focus Domains

- Space planning
- Building management
- Custodial services
- Classroom management
- Data quality control



**Semi-structured discussions
and live demos**

Tasks



Confirm assumptions or previous observations.

Do students occupy room x in evenings or on weekends?

Tasks

☒ **Confirm assumptions or previous observations.**

☒ **Monitor the current/recent utilization rate.**

Which rooms are empty/busy?

Tasks

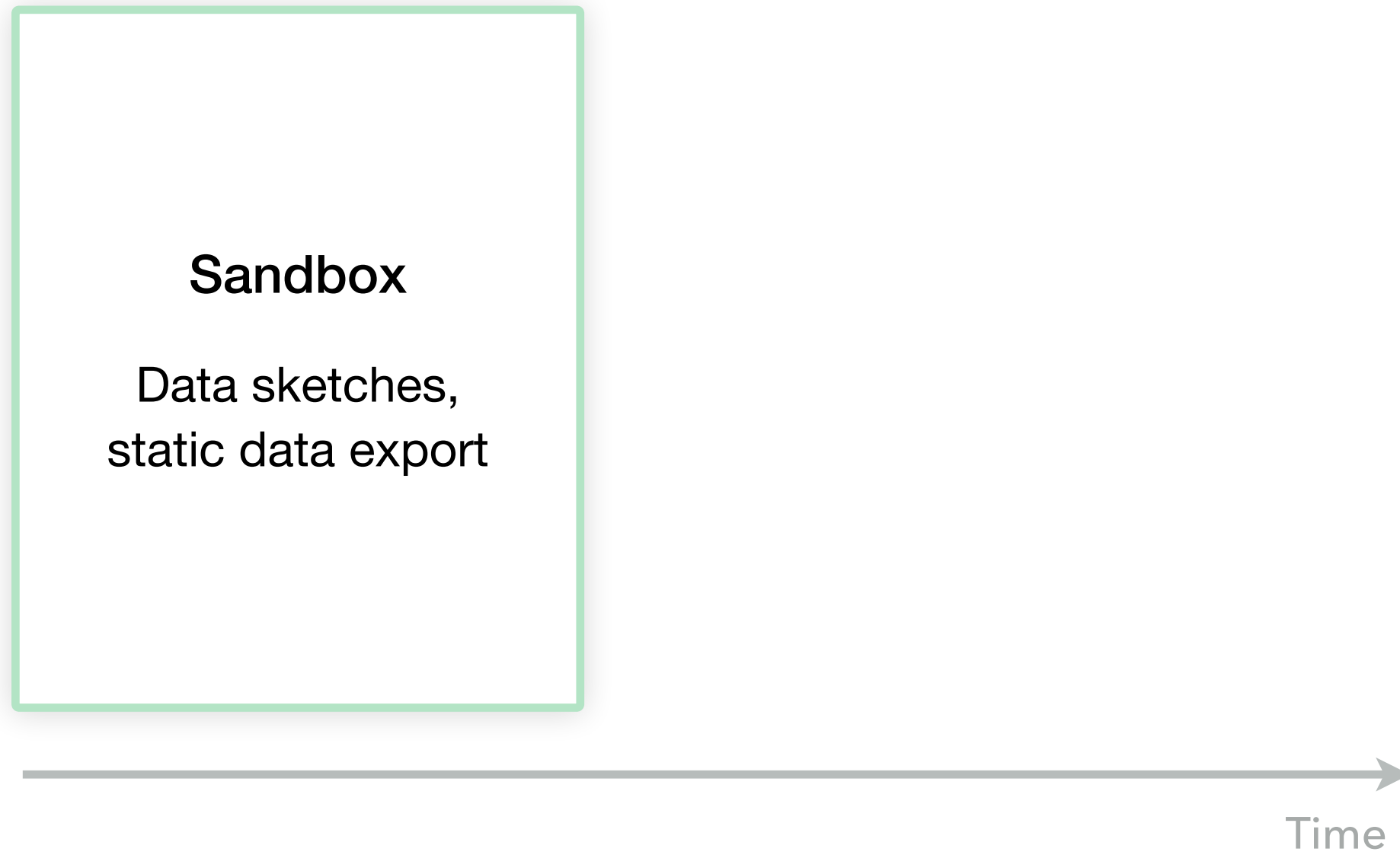
- ☒ **Confirm** assumptions or previous observations.
- ☒ **Monitor** the current/recent utilization rate.
- ☒ **Communicate** space usage and justify decisions.
Space usage improved after renovation.

Tasks

- ☒ **Confirm assumptions or previous observations.**
- ☒ **Monitor the current/recent utilization rate.**
- ☒ **Communicate space usage and justify decisions.**
- ☒ **Validate the data (quality control).**
Check minimum size of a room that can be captured.

Spatial and Temporal Data Granularities

Visualization Prototypes



Visualization Prototypes

Sandbox

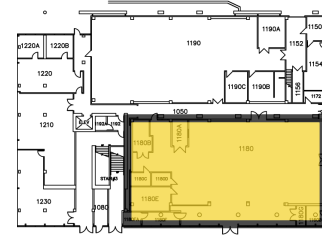
Data sketches,
static data export

- **original plan: different interface for each stakeholder**
- **realization: task & data abstractions match multiple stakeholders**
 - **if slice by space & time granularity**

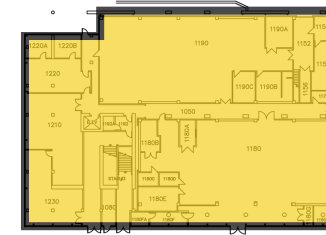
Spatial and Temporal Data Granularities

Regions of interest

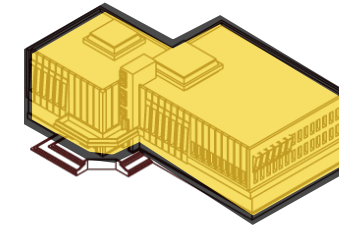
Zone



Floor



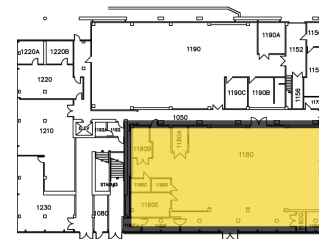
Building



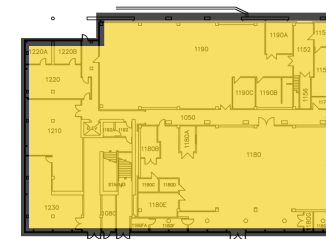
Spatial and Temporal Data Granularities

Regions of interest

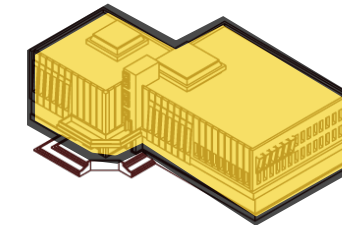
Zone



Floor



Building



Periods of interest

Mondays

Weekdays

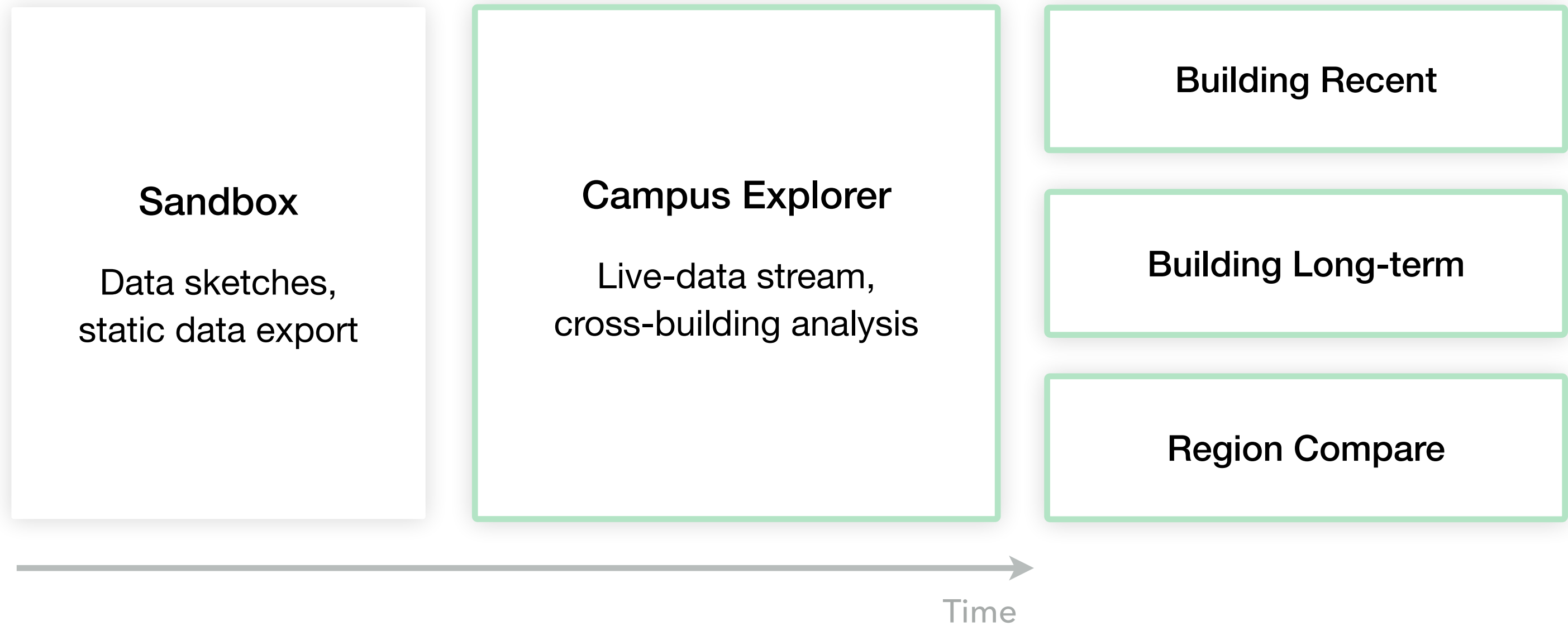
last 12 hours

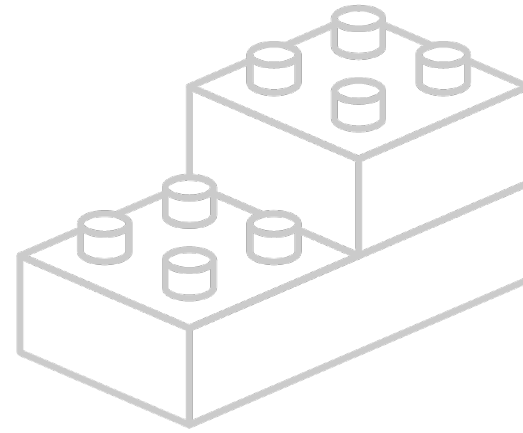
Summer term

Fr 8-10am

Weekends

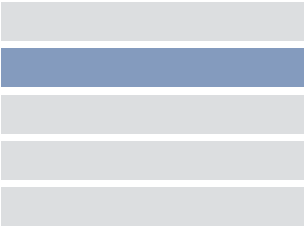
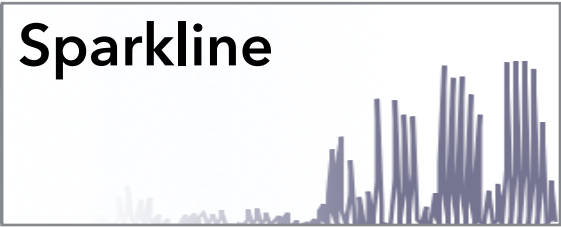
Visualization Prototypes



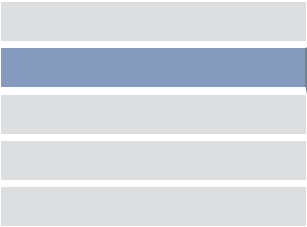
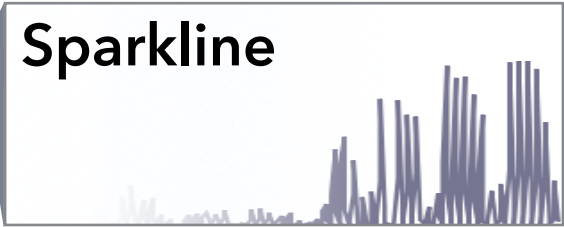
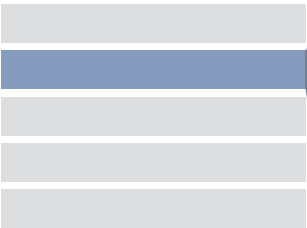
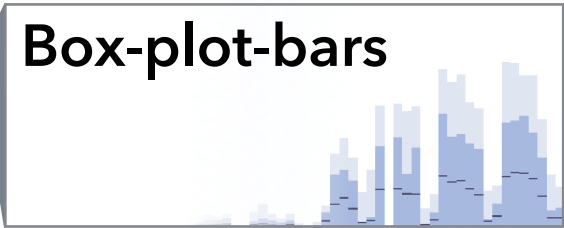


Reusable Visualization Components

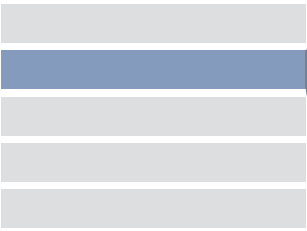
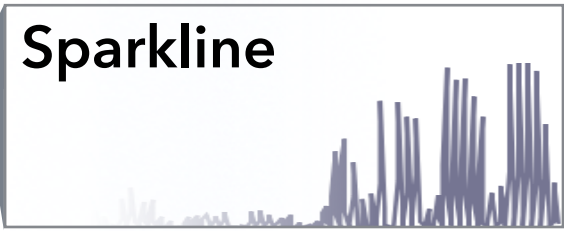
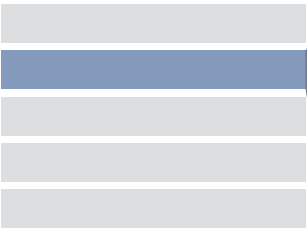
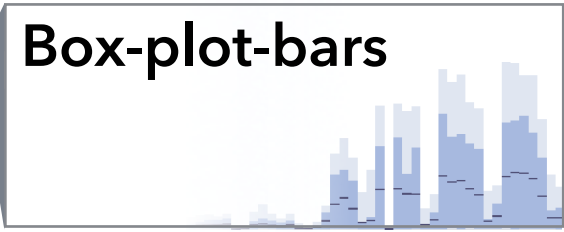
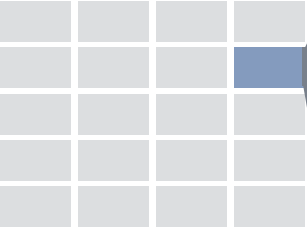
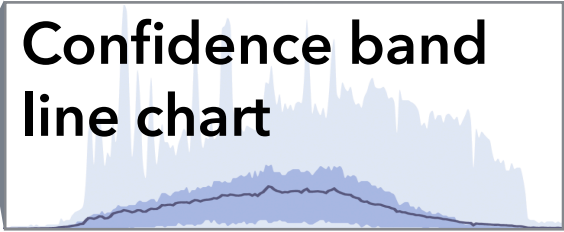
Reusable Visualization Components

Layout	Visual Encoding	Facet	Comparisons
	Sparkline 	Juxtaposition	Repeating patterns, trends, outliers (contiguous)

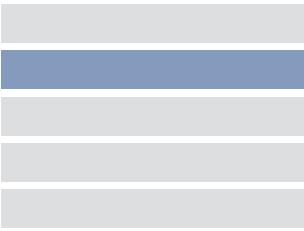
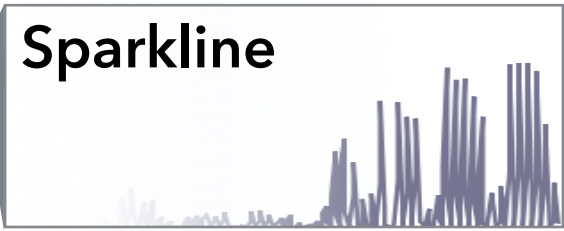
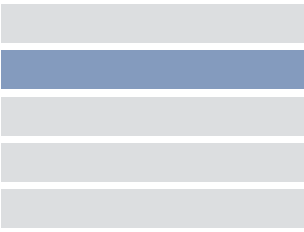
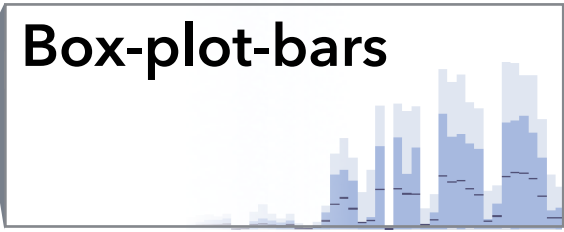
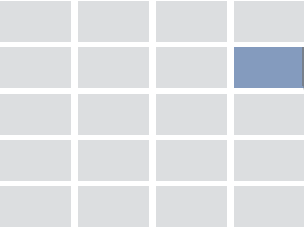
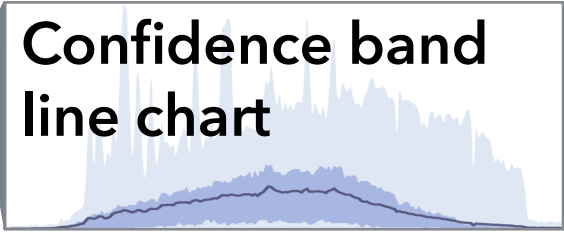
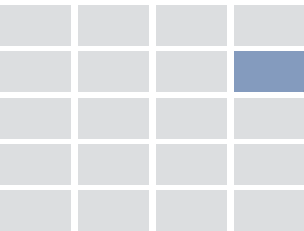
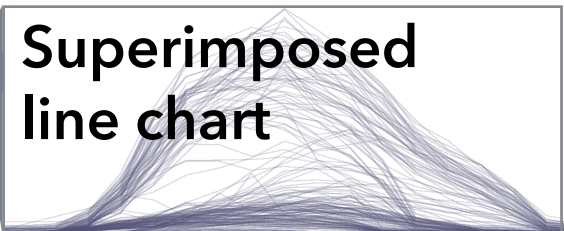
Reusable Visualization Components

Layout	Visual Encoding	Facet	Comparisons
	Sparkline 	Juxtaposition	Repeating patterns, trends, outliers (contiguous)
	Box-plot-bars 	Juxtaposition	Repeating patterns, trends, outliers (<i>non-contiguous</i>)

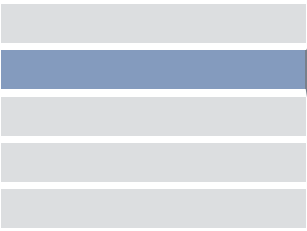
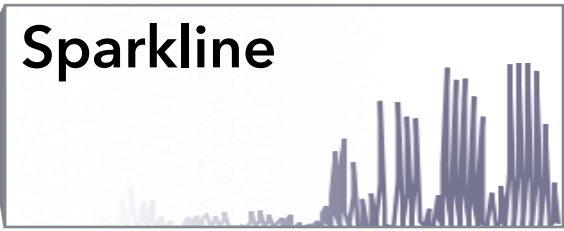
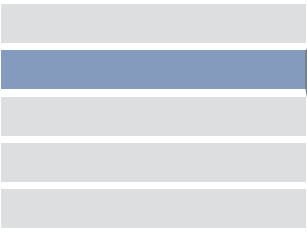
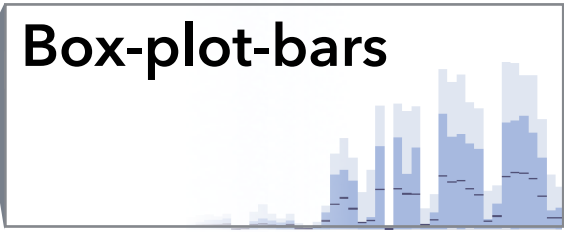
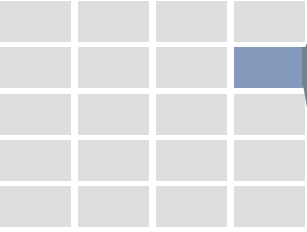
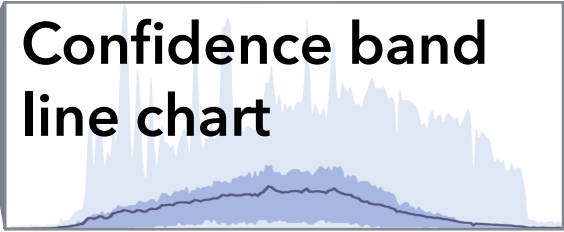
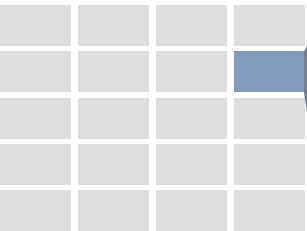
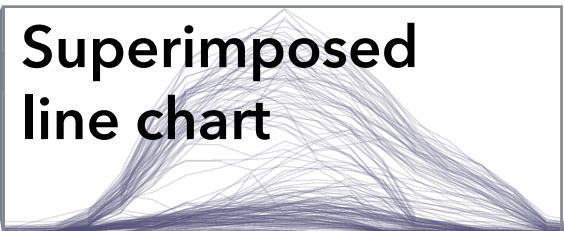
Reusable Visualization Components

Layout	Visual Encoding	Facet	Comparisons
	Sparkline 	Juxtaposition	Repeating patterns, trends, outliers (contiguous)
	Box-plot-bars 	Juxtaposition	Repeating patterns, trends, outliers (<i>non-contiguous</i>)
	Confidence band line chart 	Aggregation	Typical utilization profiles

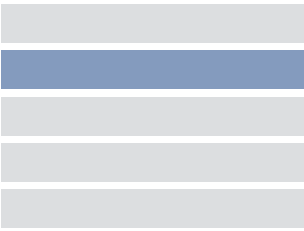
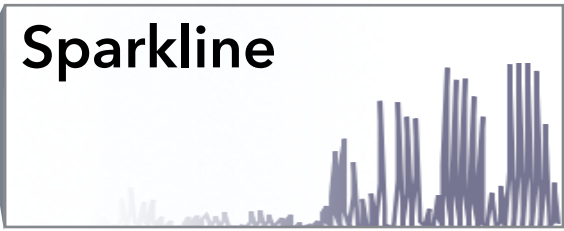
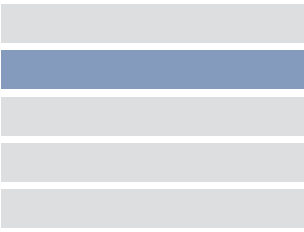
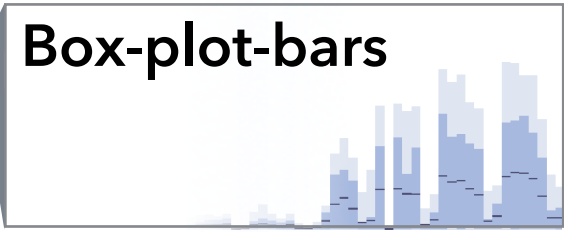
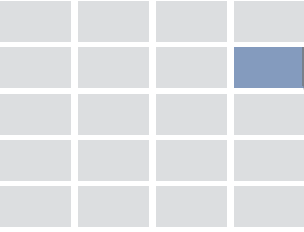
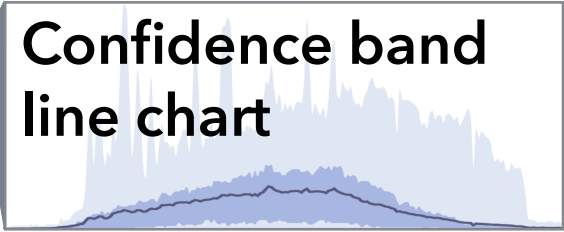
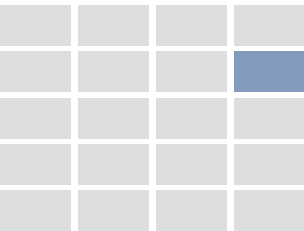
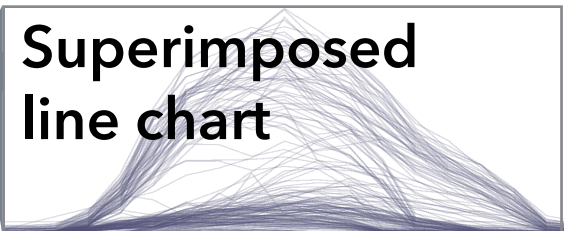
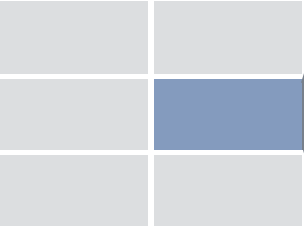
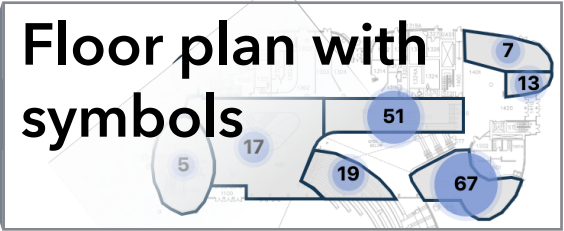
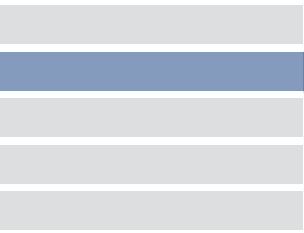
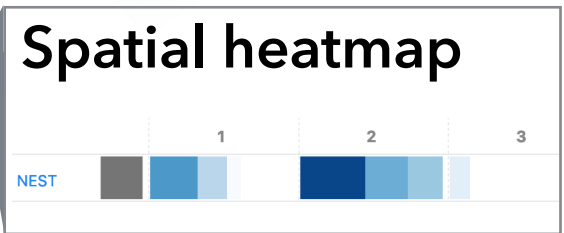
Reusable Visualization Components

Layout	Visual Encoding	Facet	Comparisons
	Sparkline 	Juxtaposition	Repeating patterns, trends, outliers (contiguous)
	Box-plot-bars 	Juxtaposition	Repeating patterns, trends, outliers (<i>non-contiguous</i>)
	Confidence band line chart 	Aggregation	Typical utilization profiles
	Superimposed line chart 	Superposition	Within-session patterns, outliers

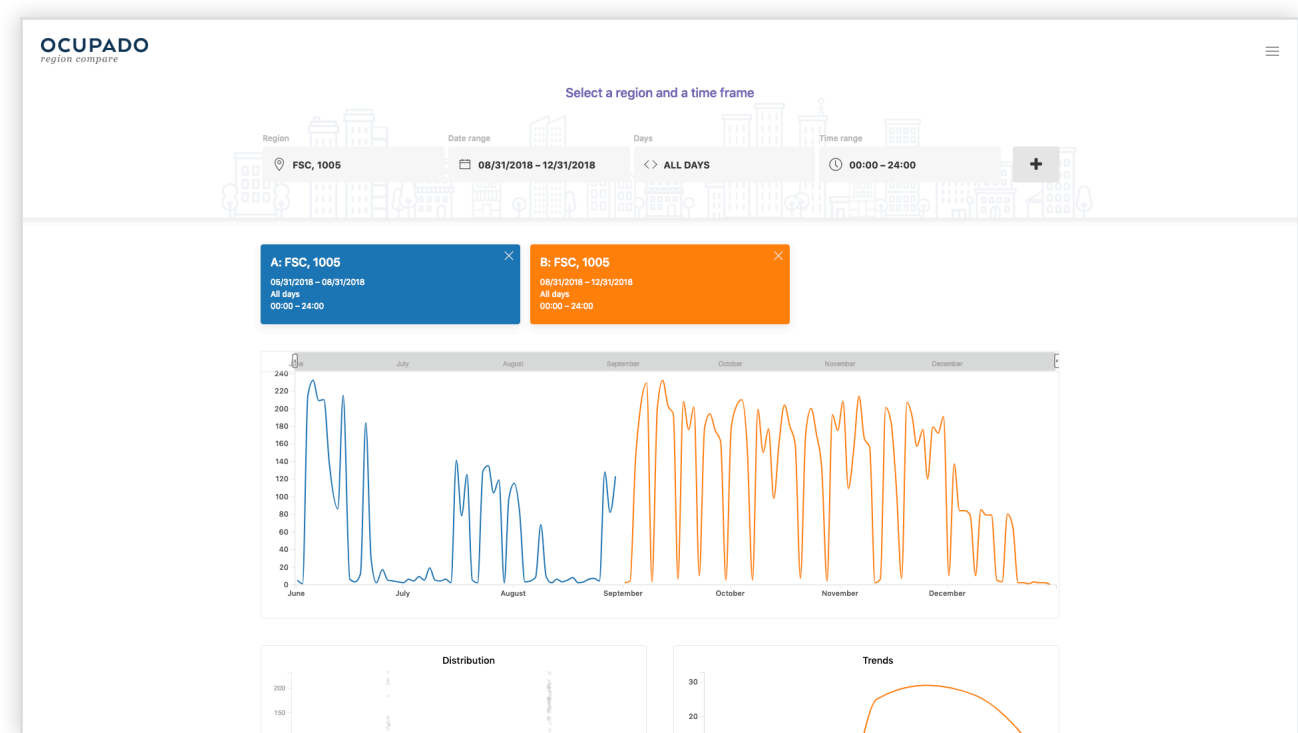
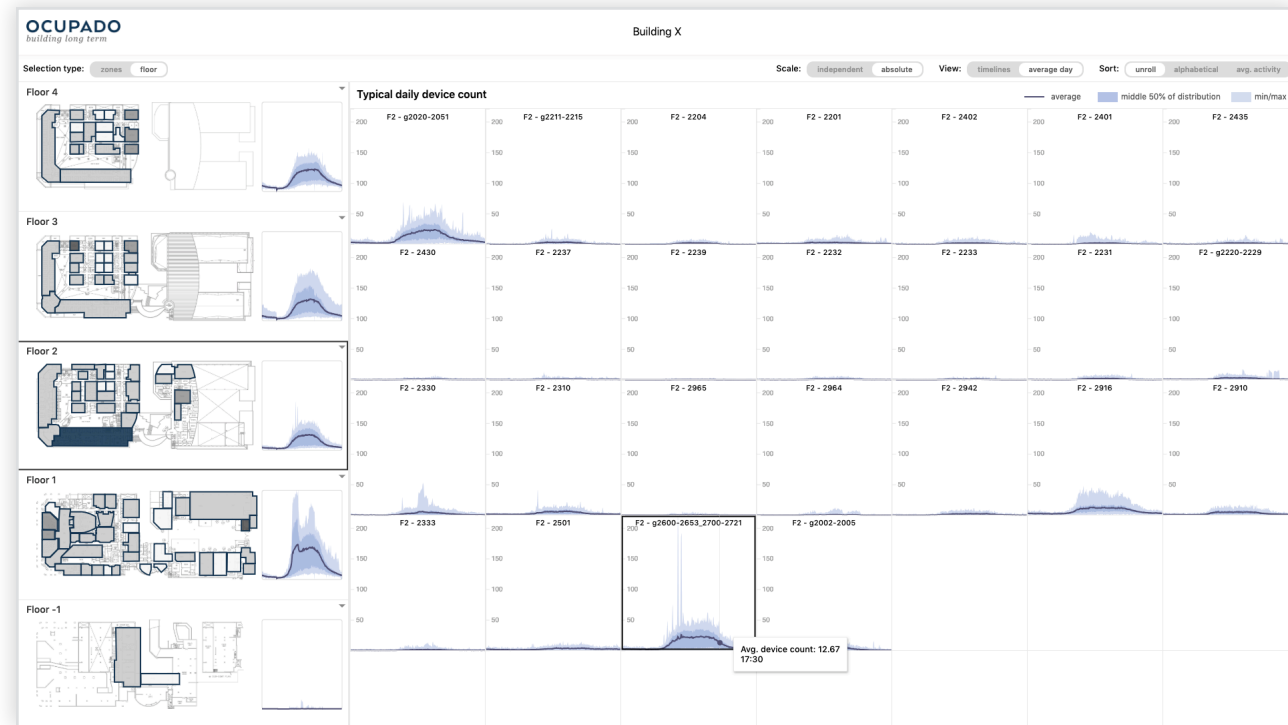
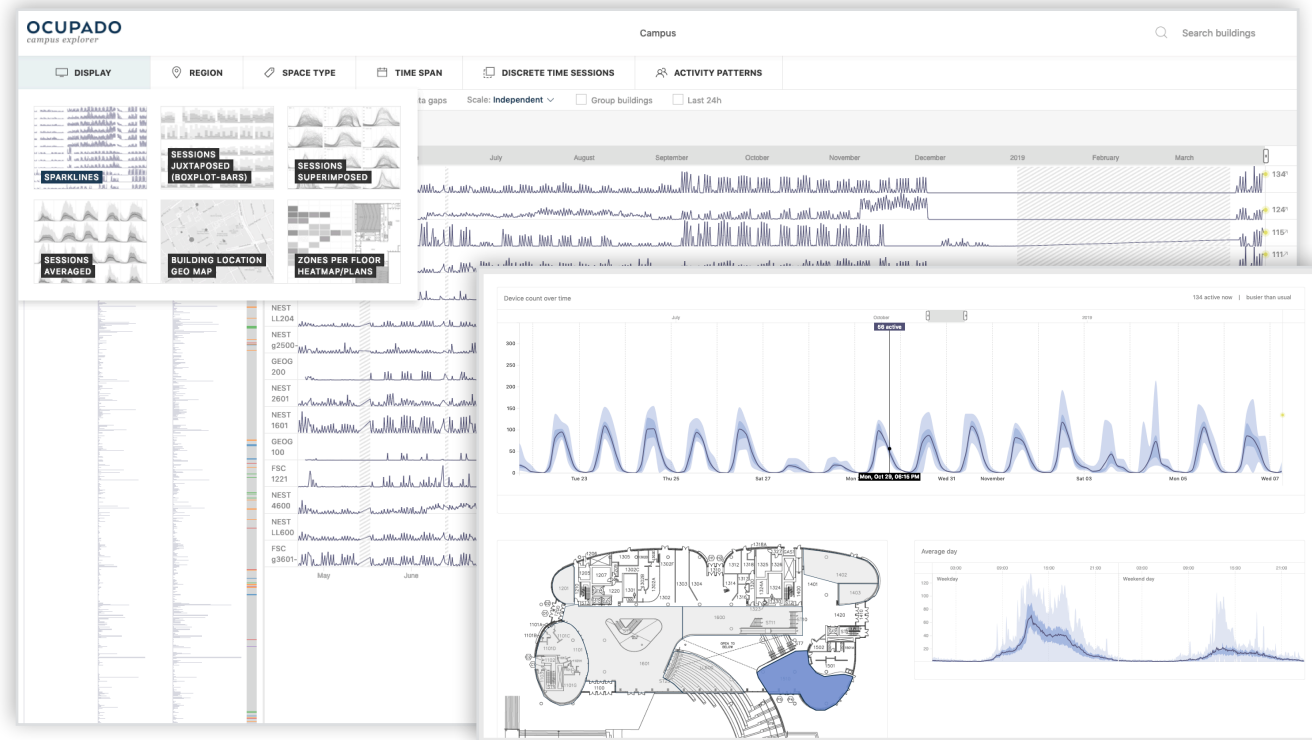
Reusable Visualization Components

	Layout	Visual Encoding	Facet	Comparisons
Temporal			Juxtaposition	Repeating patterns, trends, outliers (contiguous)
			Juxtaposition	Repeating patterns, trends, outliers (<i>non-contiguous</i>)
			Aggregation	Typical utilization profiles
			Superposition	Within-session patterns, outliers

Reusable Visualization Components

	Layout	Visual Encoding	Facet	Comparisons
Temporal			Juxtaposition	Repeating patterns, trends, outliers (contiguous)
			Juxtaposition	Repeating patterns, trends, outliers (non-contiguous)
			Aggregation	Typical utilization profiles
			Superposition	Within-session patterns, outliers
Spatial			Superposition	Within local spatial neighborhood
			Containment (nested)	Across distributed regions

Ocupado Interfaces



Ocupado: Visualizing Location-Based Counts Over Time Across Buildings

Michael Oppermann

Tamara Munzner



THE UNIVERSITY
OF BRITISH COLUMBIA



Project partner:



<https://youtu.be/KcwjVK8eUdw?t=83>

Ocupado Contributions

- Analysis and abstraction of data and tasks for studying space utilization
- Ocupado, a set of visual decision support tools
- Generalizable design choices for visualizing non-trajectory spatiotemporal data relating to large-scale indoor environments

Spinoff toolkit paper: TimeElide

<https://youtu.be/a7vJxTRWbml>

TimeElide

Select data

bike_rides.csv | [Other data source](#)

Title for variable of interest:

Select slicing method

Automatically detect slices

Number of total slices: 20

Median within-slice distance5 minutes

Median between-slice distance4 days

Threshold: 8 hours

See details

Reset

Select visualization type







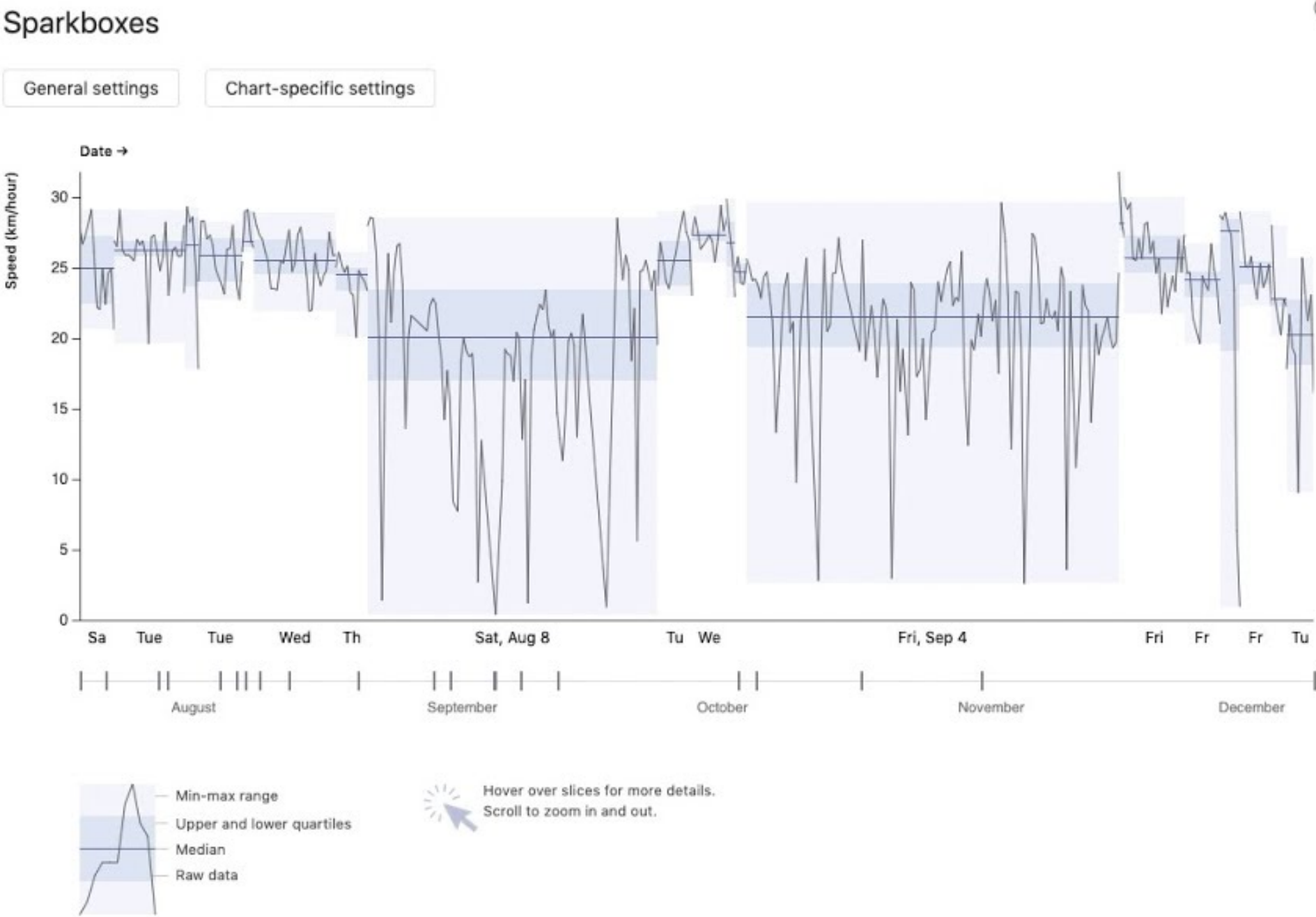






UBC InfoVis

Open Source



Q&A

Four case studies of problem-driven work

- in-car networks



- facilities management



- biology





Zipeng
Liu



Shing Hei
Zhan



Aggregated Dendrograms

for Visual Comparison Between Many Phylogenetic Trees

<http://www.cs.ubc.ca/labs/imager/tr/2019/adview>

Aggregated Dendrograms for Visual Comparison Between Many Phylogenetic Trees.

Liu, Zhan, Munzner. *IEEE Trans. Visualization and Computer Graphics (TVCG)* 26(9):2732-2747, 2019.

Phylogenetic tree

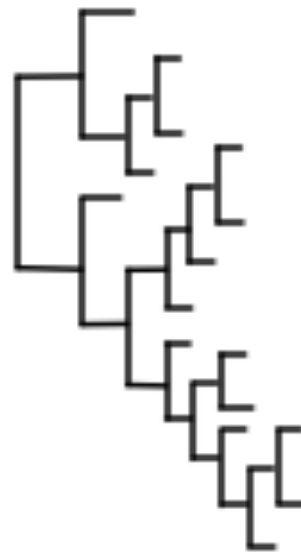
Evolutionary relationships of organisms

Human	A	T	G	G	A	C	A
Chimpanzee	A	T	G	G	A	C	A
Macaque	A	C	G	G	A	C	A

Genetic information

Computational workflow

Phylogenetic tree



Many phylogenetic trees

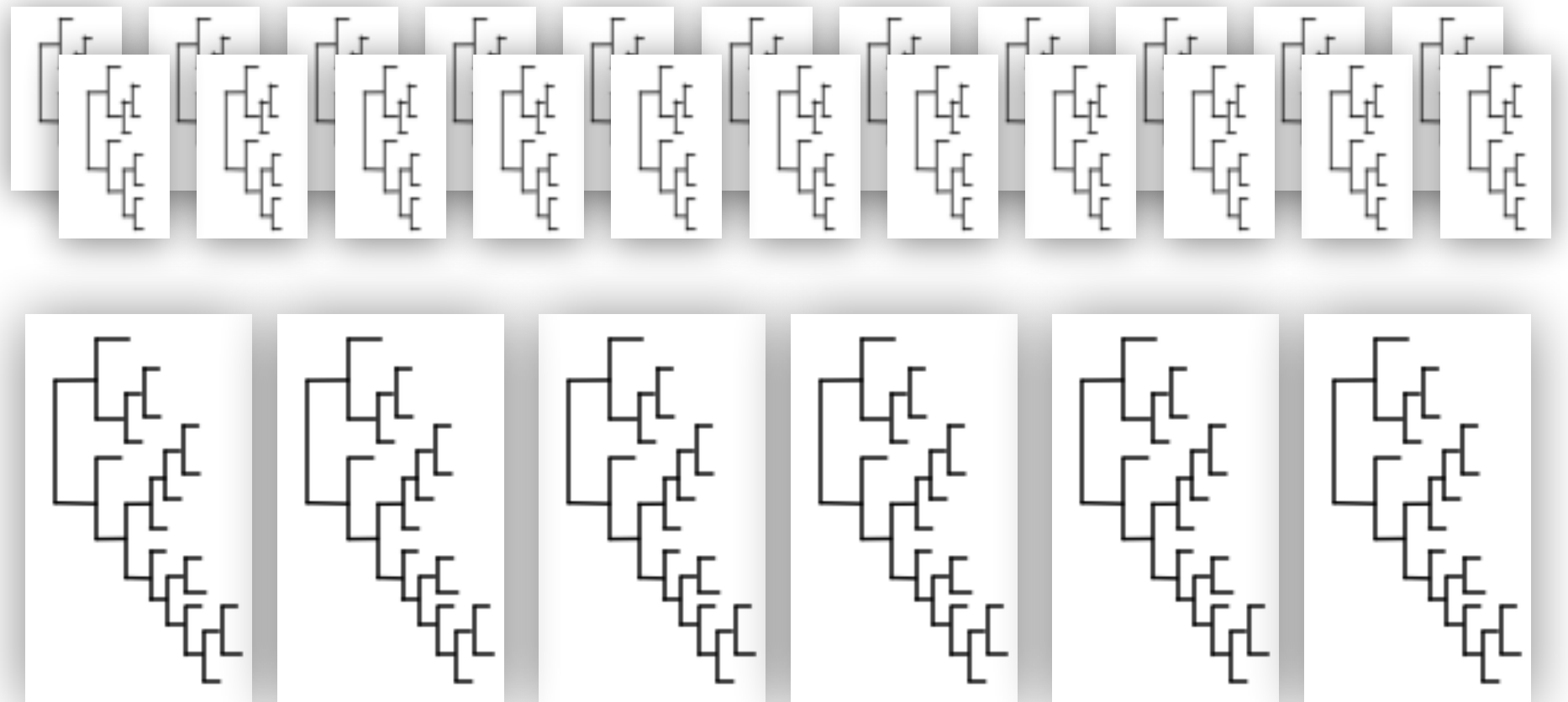
Human	A	T	G	G	A	C	A
Chimpanzee	A	T	G	G	A	C	A
Macaque	A	C	G	G	A	C	A

Genetic information

Computational workflow

Phylogenetic tree

- Understand relationships between genes and species trees
- Explore trees generated with different methods and data



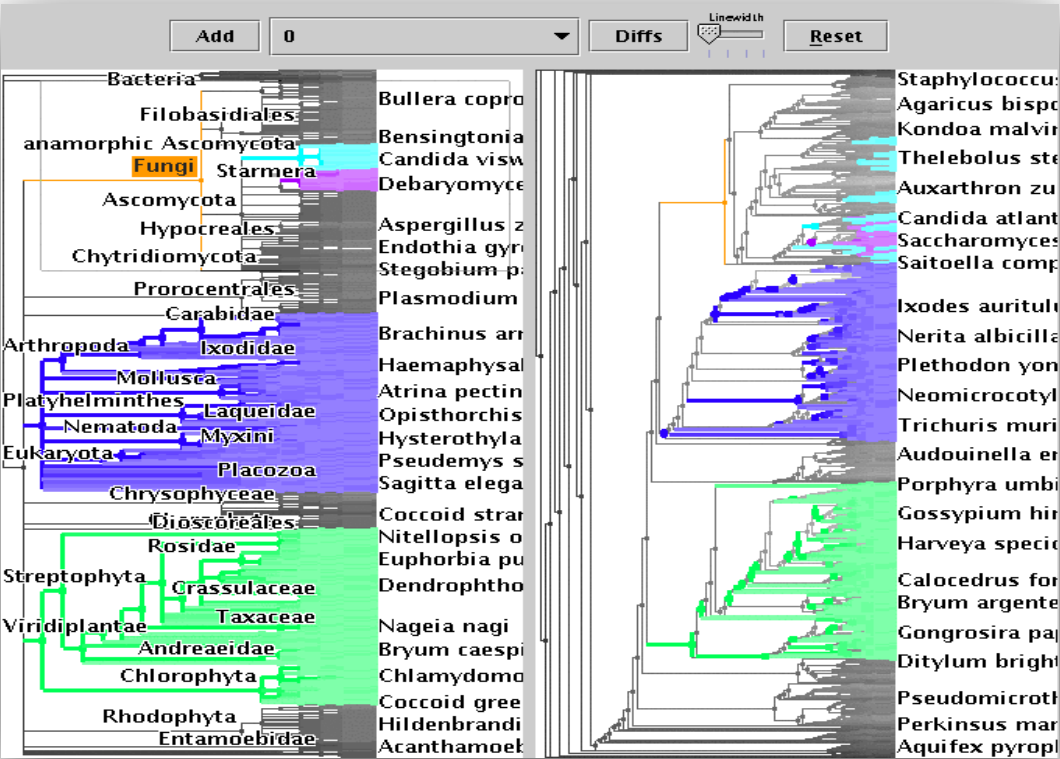
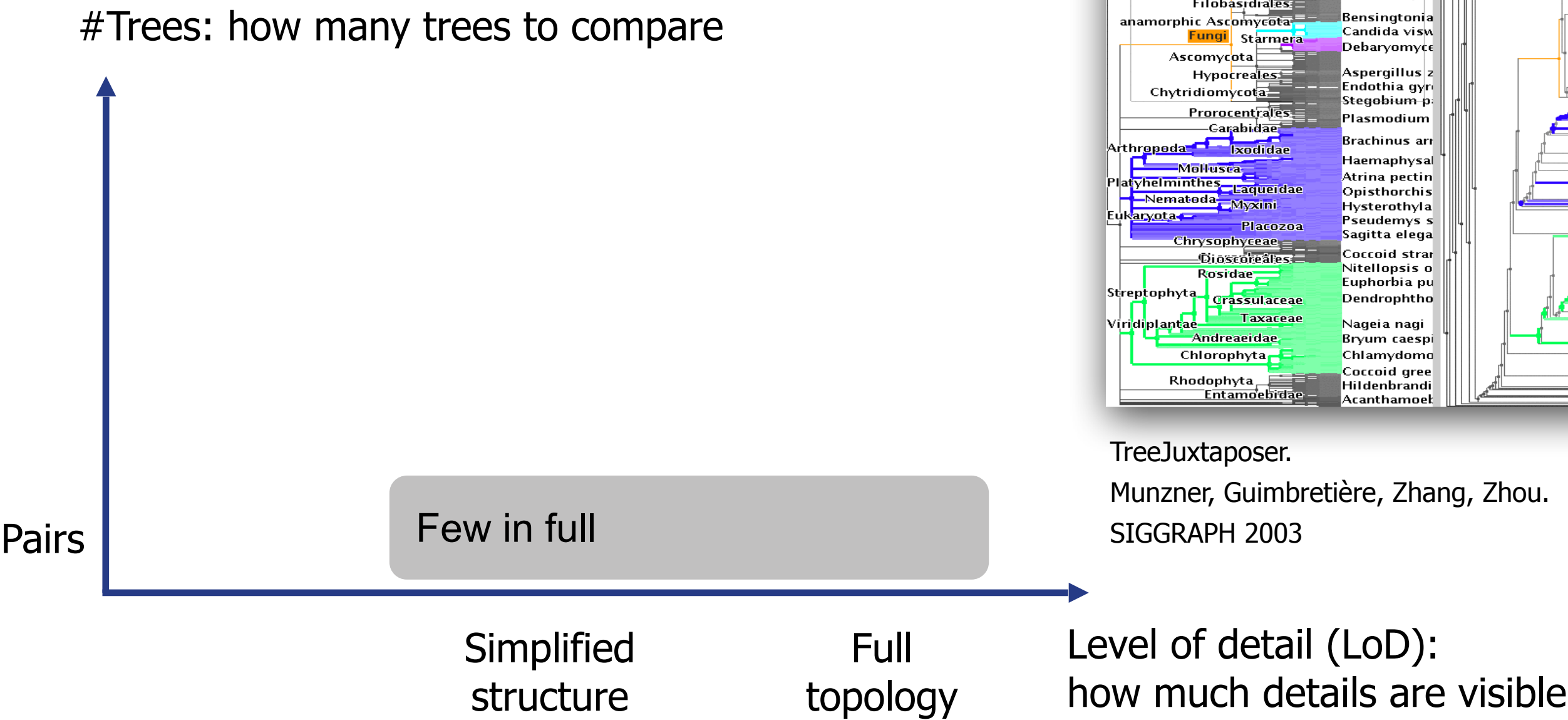
Scalability of Existing Tree Comparison Systems

#Trees: how many trees to compare



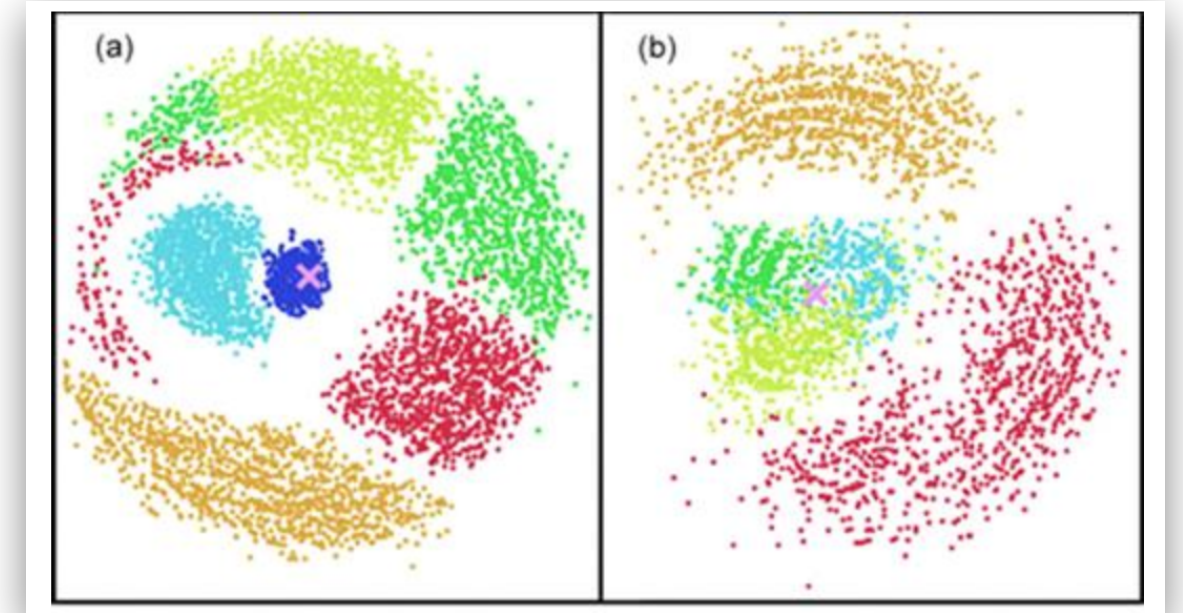
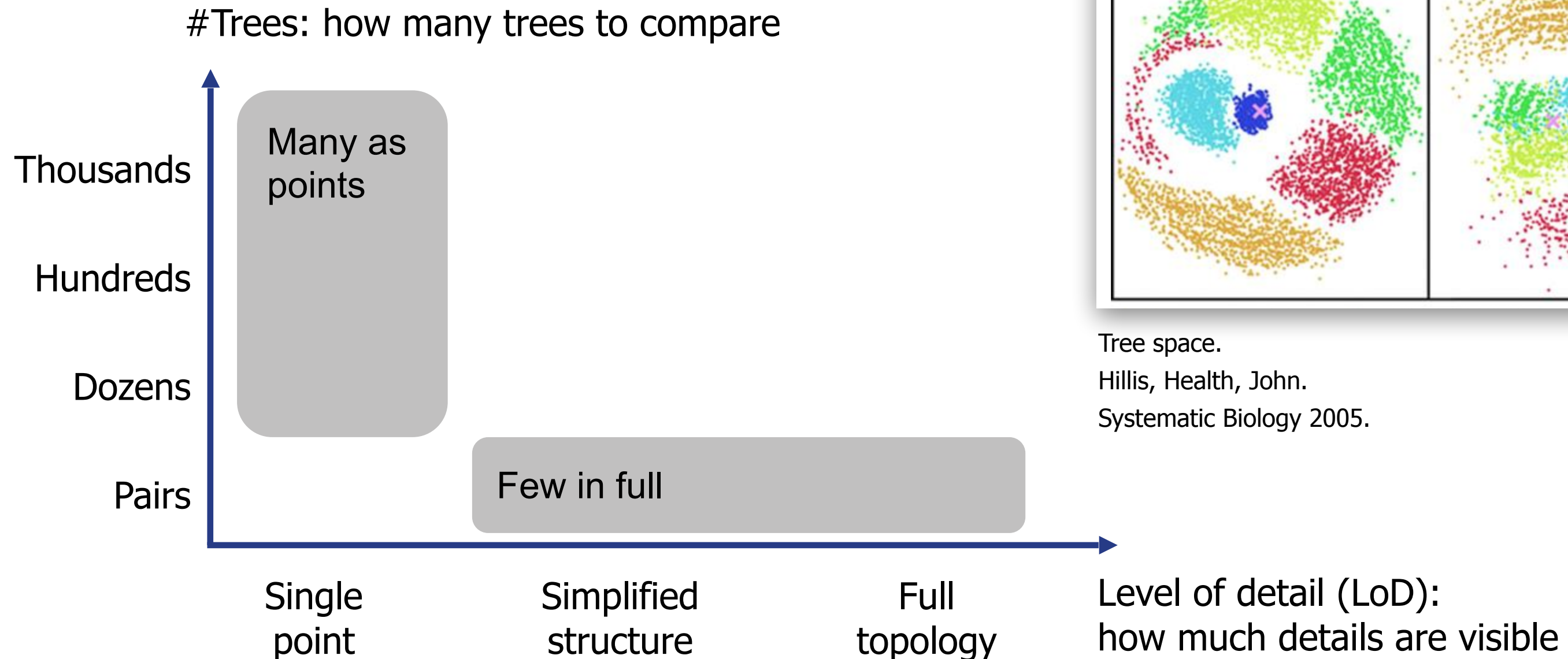
Level of detail (LoD):
how much details are visible

Scalability of Existing Tree Comparison Systems



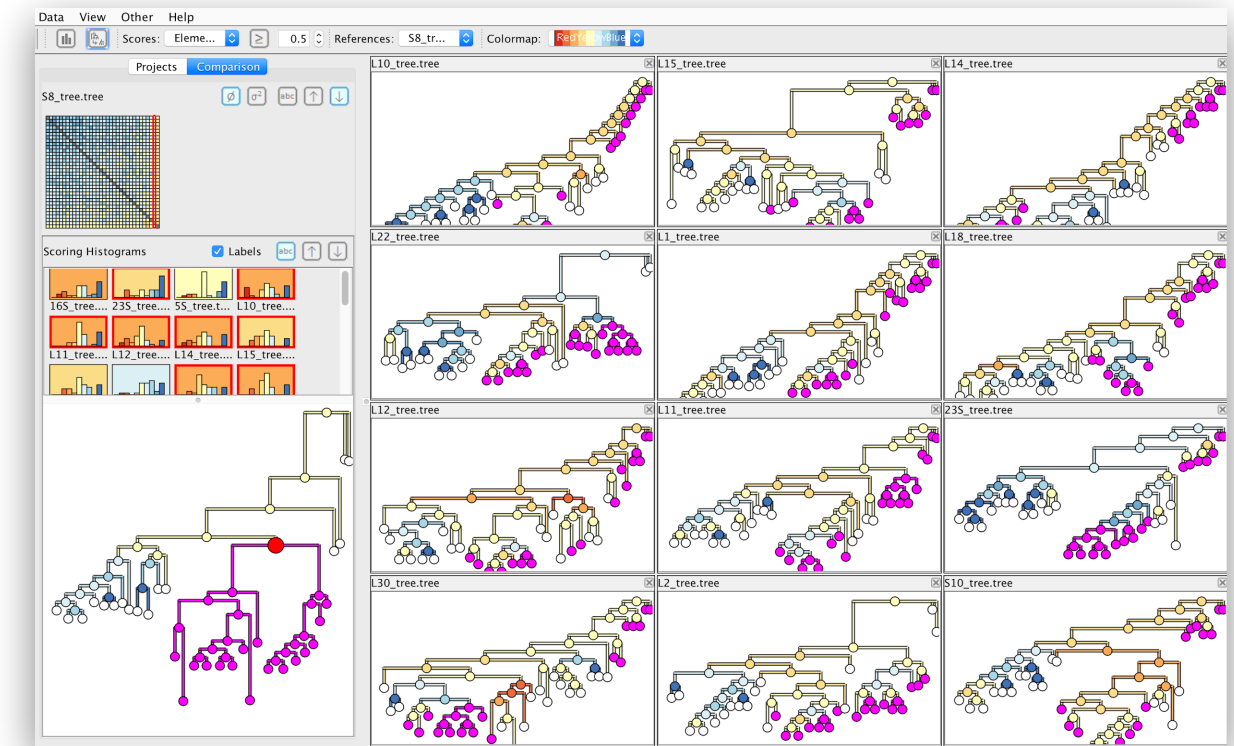
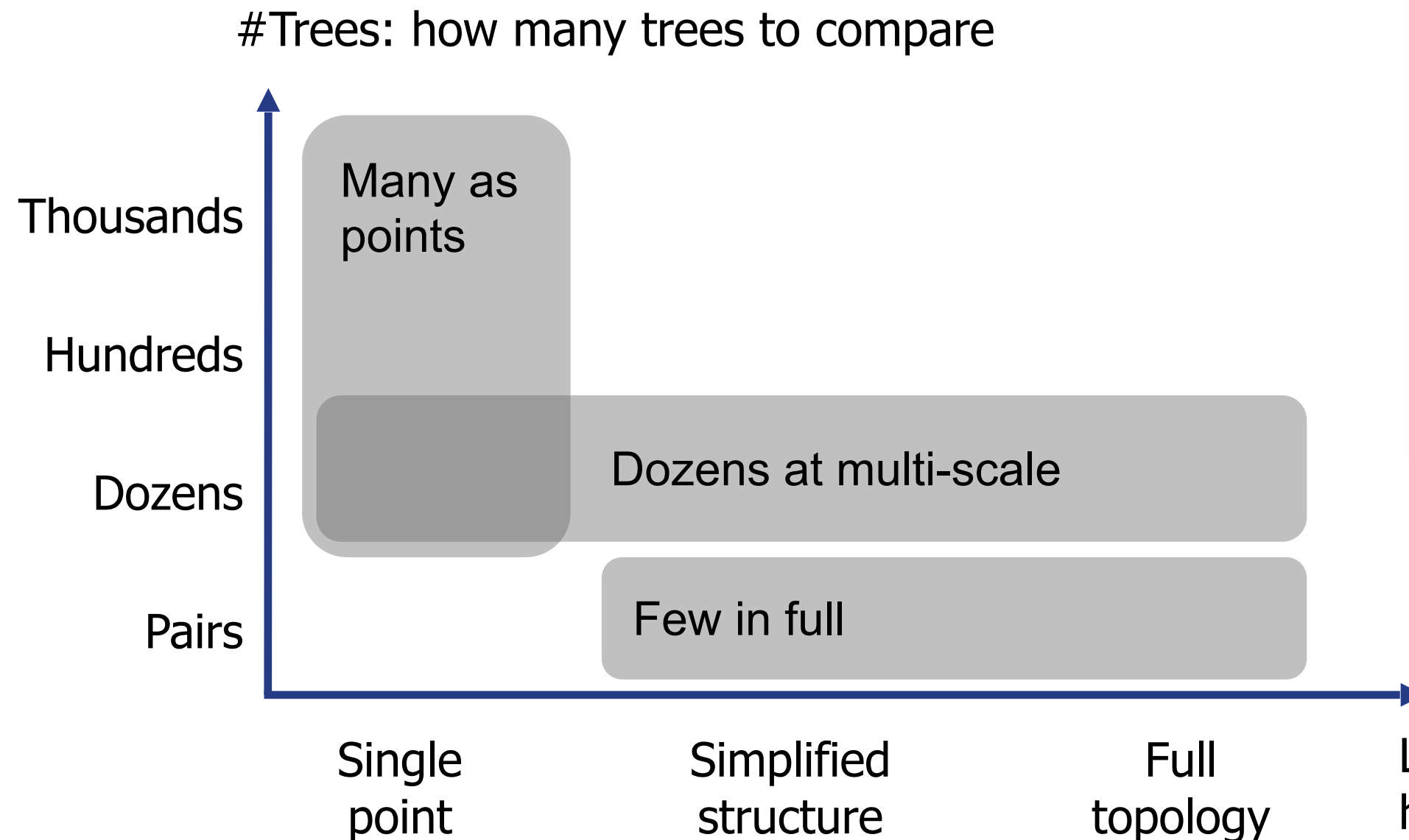
TreeJuxtaposer.
Munzner, Guimbretière, Zhang, Zhou.
SIGGRAPH 2003

Scalability of Existing Tree Comparison Systems



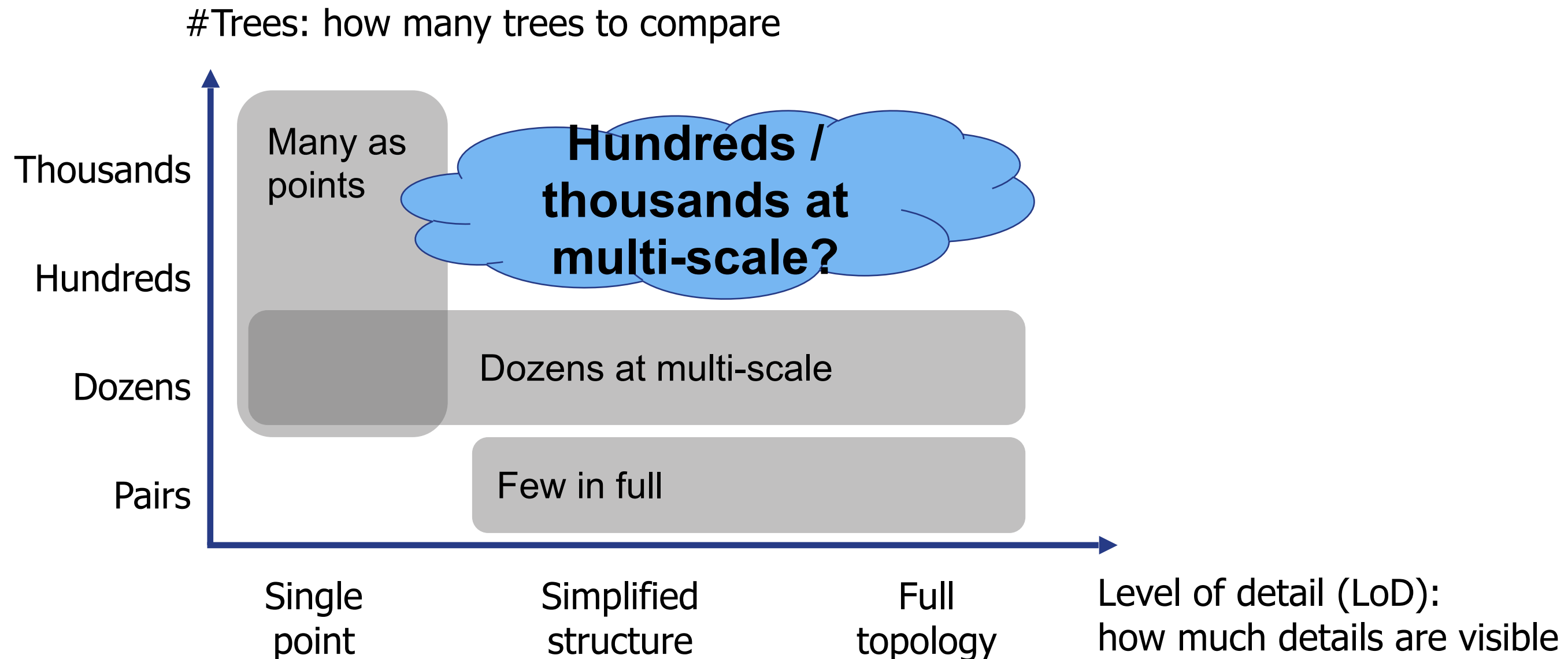
Tree space.
Hillis, Heath, John.
Systematic Biology 2005.

Scalability of Existing Tree Comparison Systems



Interactive visual comparison of multiple trees.
Bremm, Landesberger, Heß, Schreck, Weil, Hamacher.
VAST 2011.

Comparing many phylogenetic trees

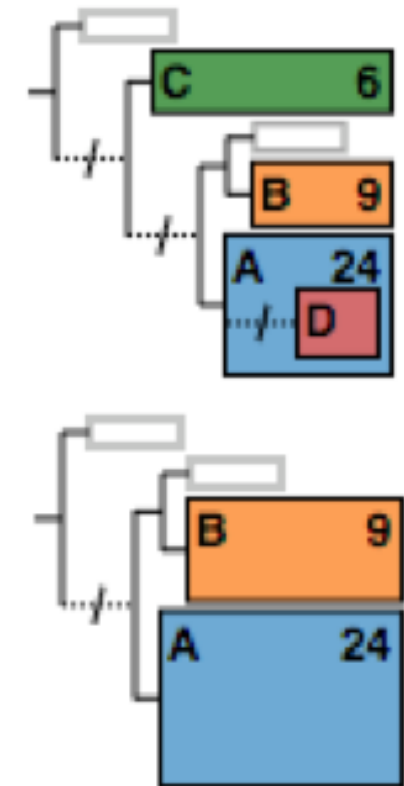


Contributions include idiom & algorithm levels

- Data and task abstractions for comparison of phylogenetic trees

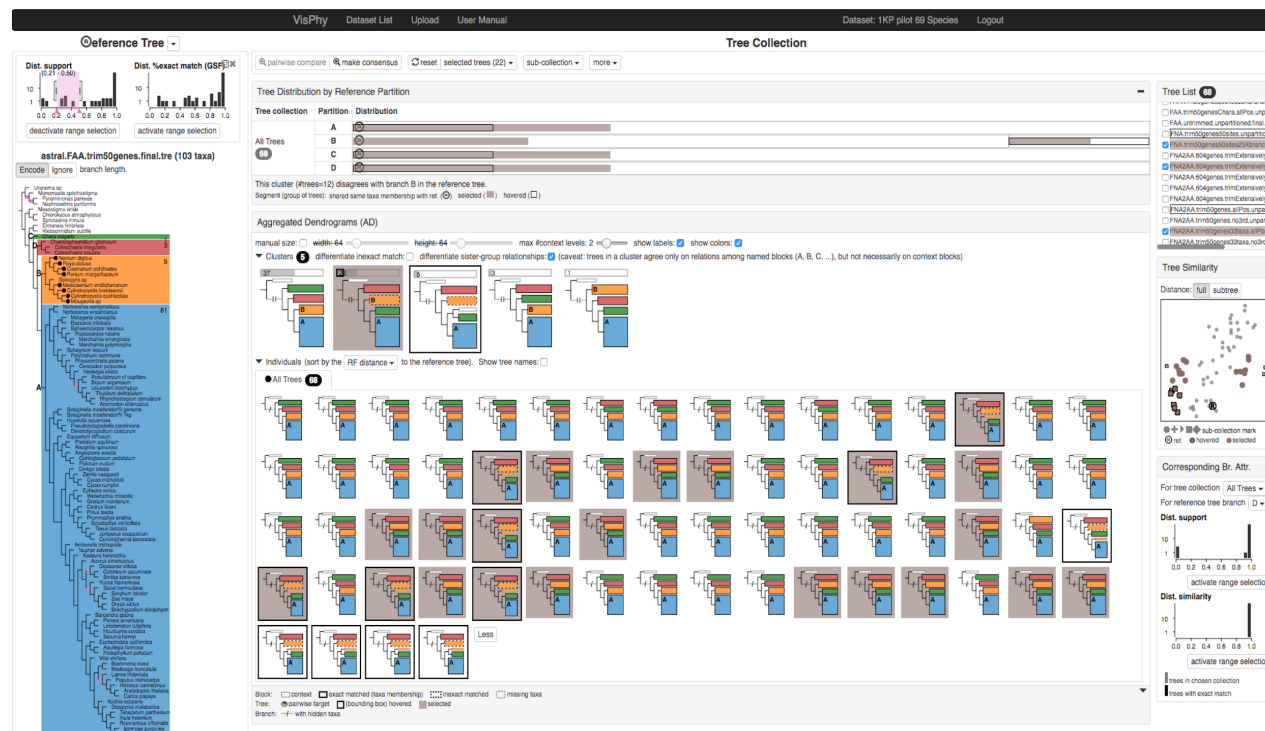
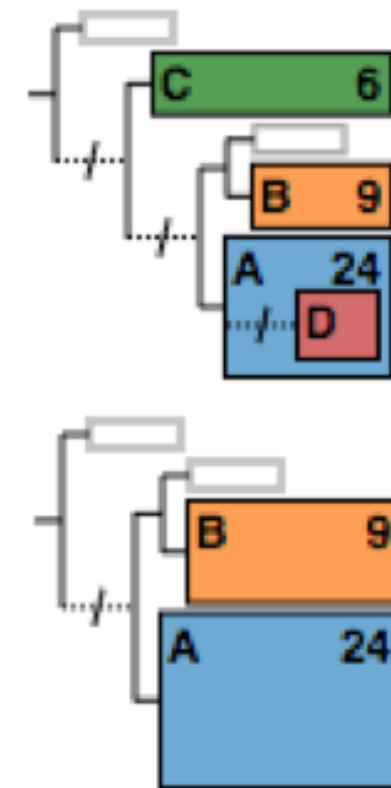
Contributions include idiom & algorithm levels

- Data and task abstractions for comparison of phylogenetic trees
- A new visual encoding: **Aggregated Dendrogram**
 - Compact tree representation that focuses on selected subtrees
 - Adapts to available screen space



Contributions include idiom & algorithm levels

- Data and task abstractions for comparison of phylogenetic trees
- A new visual encoding: **Aggregated Dendrogram**
 - Compact tree representation that focuses on selected subtrees
 - Adapts to available screen space
- A multi-view interactive tool: **ADView**
 - Covers multiple levels of details for tree comparison



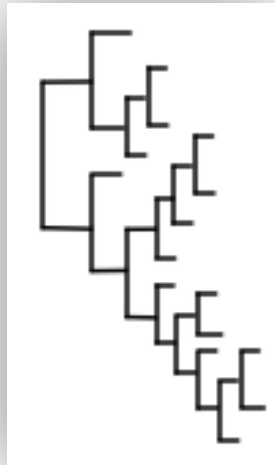


Data & Tasks

- Tree data
- Two crucial tasks

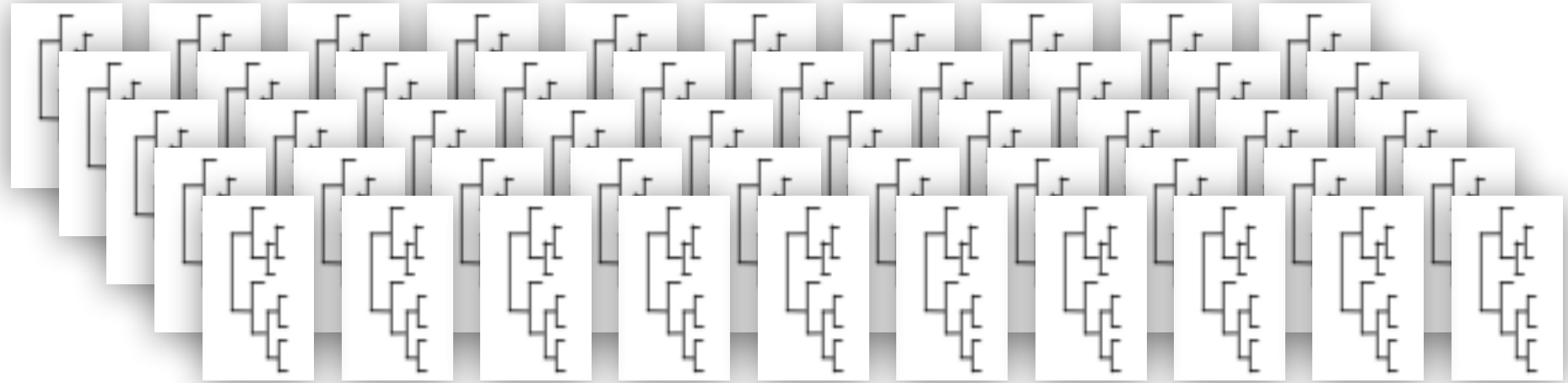
Tree data

Reference tree



vs.

Tree collection

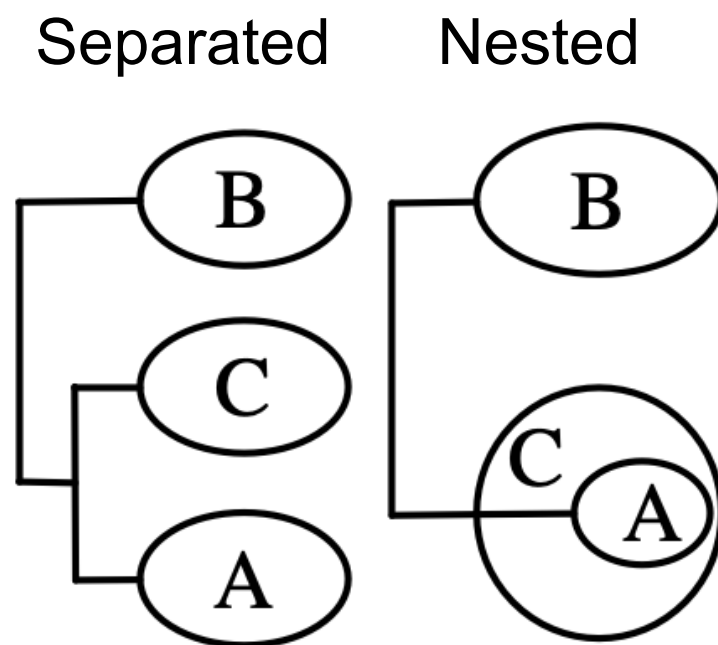


Two crucial tasks

Topological relationships between
subtrees / leaf nodes

Two crucial tasks

Topological relationships between subtrees / leaf nodes

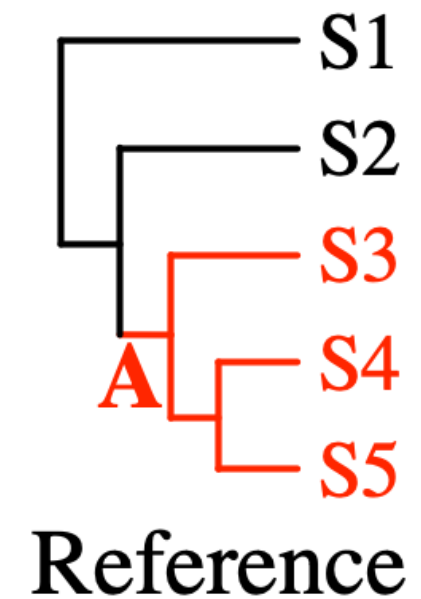
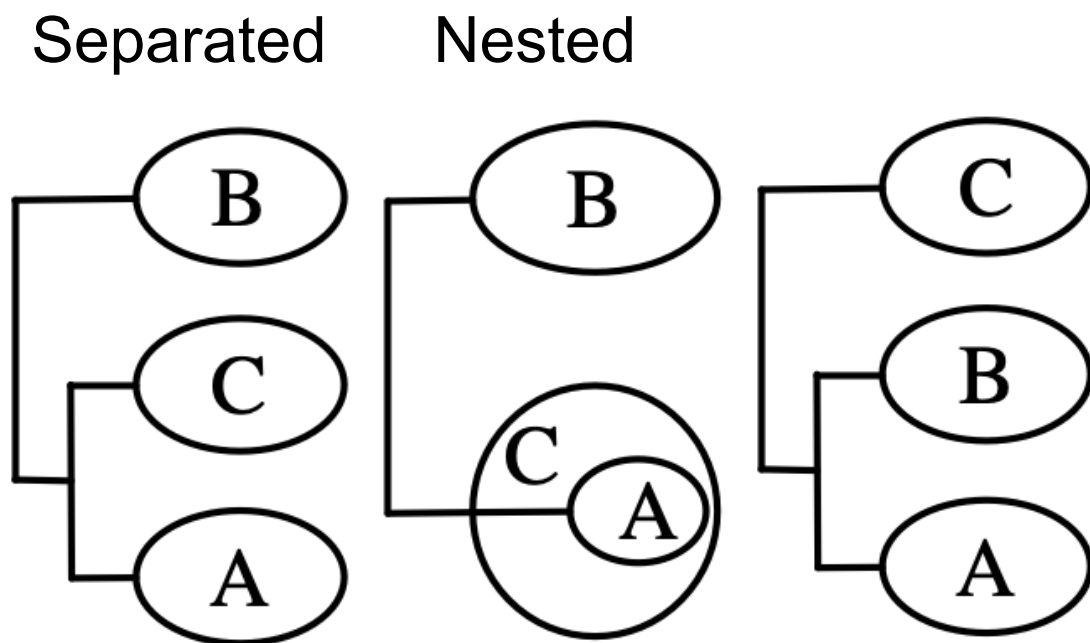


Two crucial tasks

Topological relationships between subtrees / leaf nodes

- Topological distance

Leaf node memberships compared to reference tree

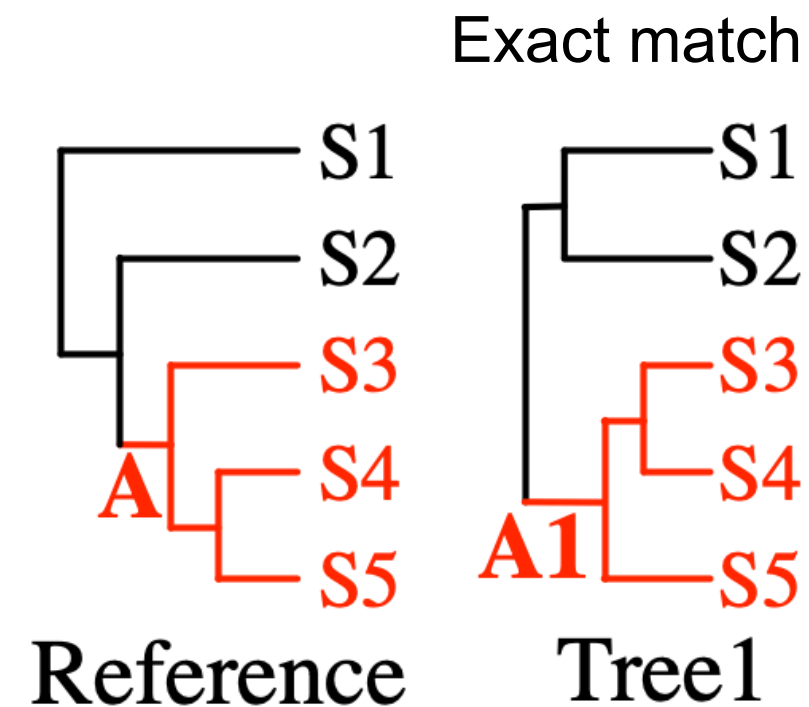
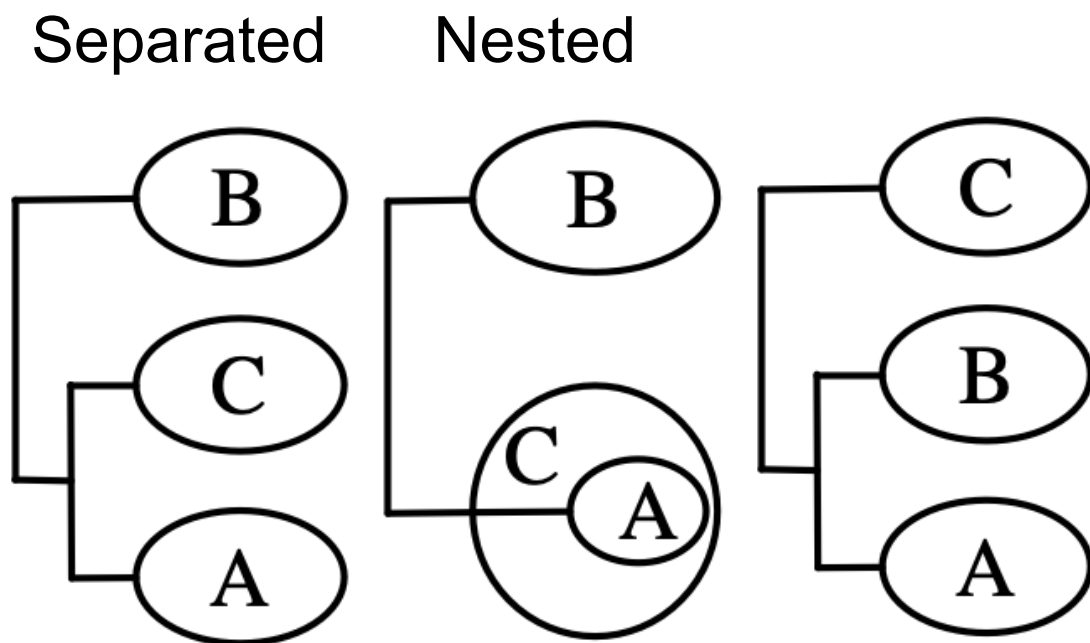


Two crucial tasks

Topological relationships between subtrees / leaf nodes

- Topological distance

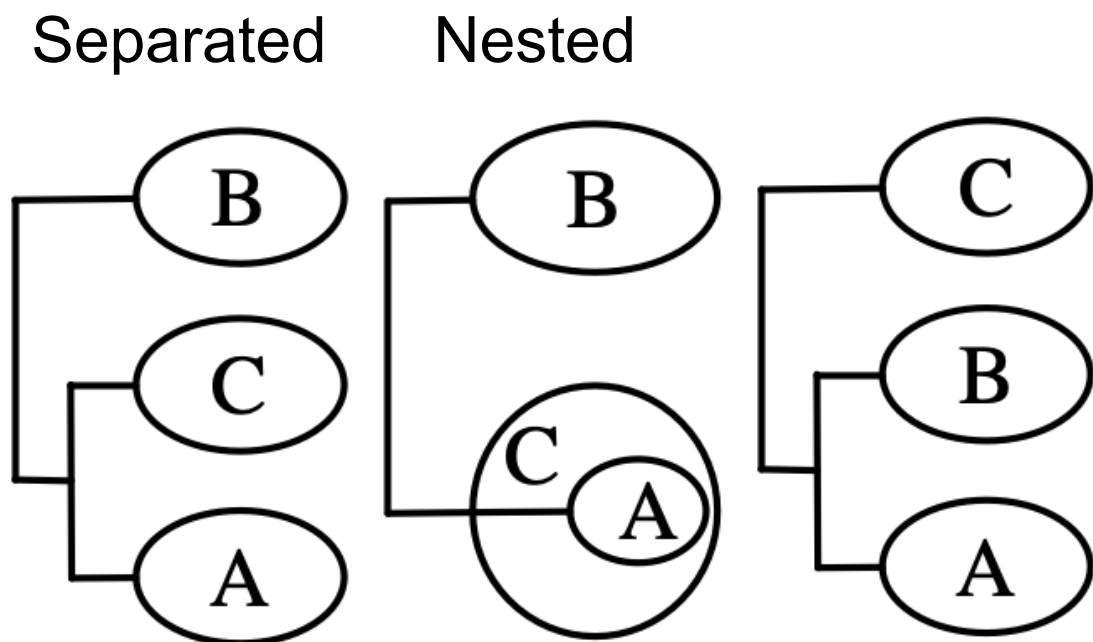
Leaf node memberships compared to reference tree



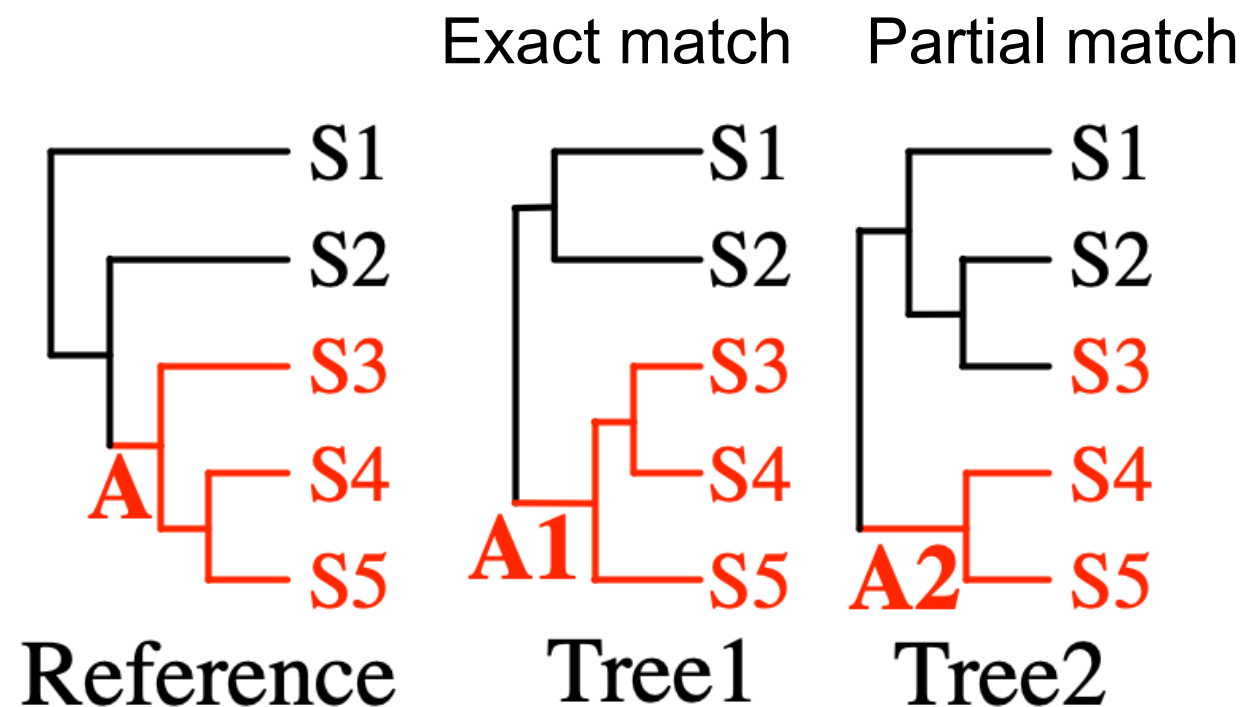
Two crucial tasks

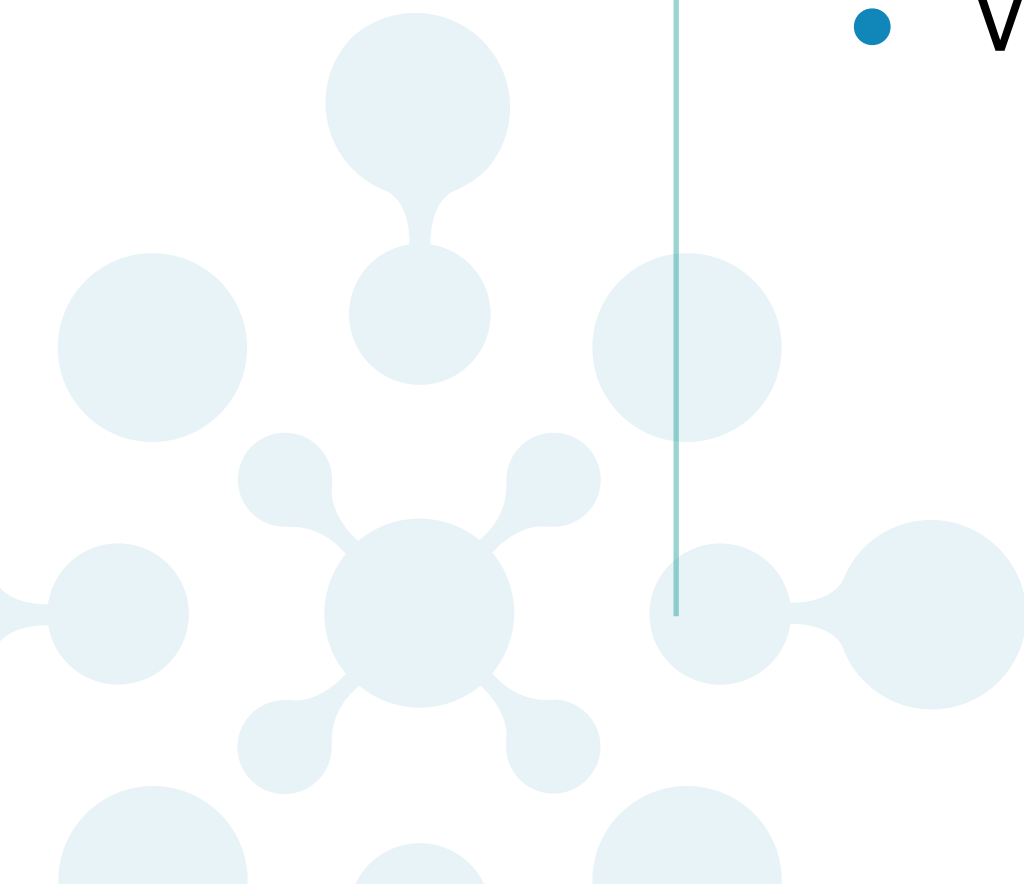
Topological relationships between subtrees / leaf nodes

- Topological distance



Leaf node memberships compared to reference tree



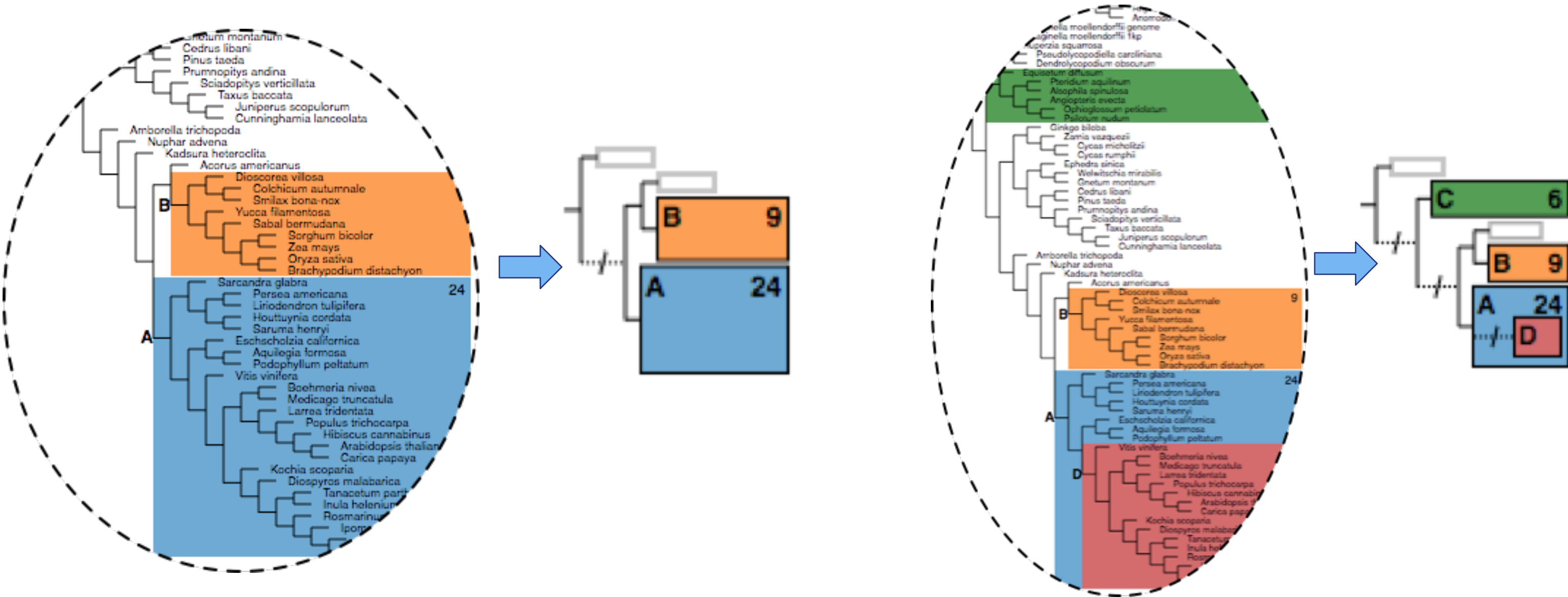


Aggregated Dendrogram (AD)

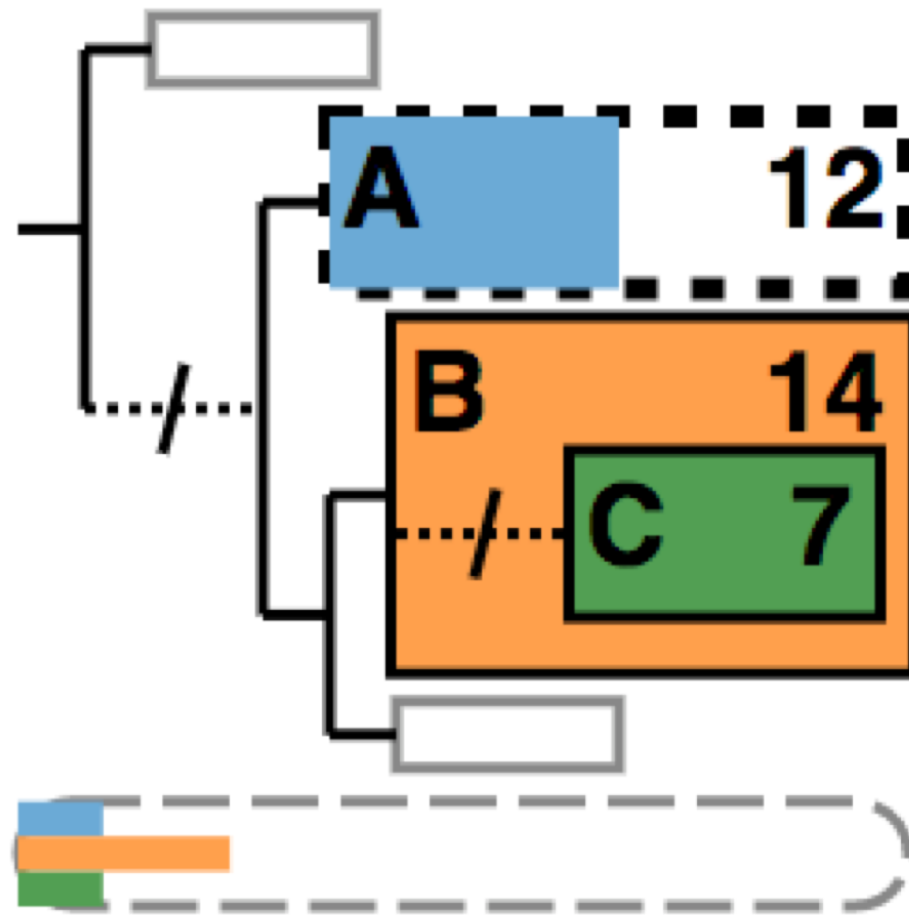
- Intuition
- Visual design

Intuition

Use glyphs to compress a tree according to user selections

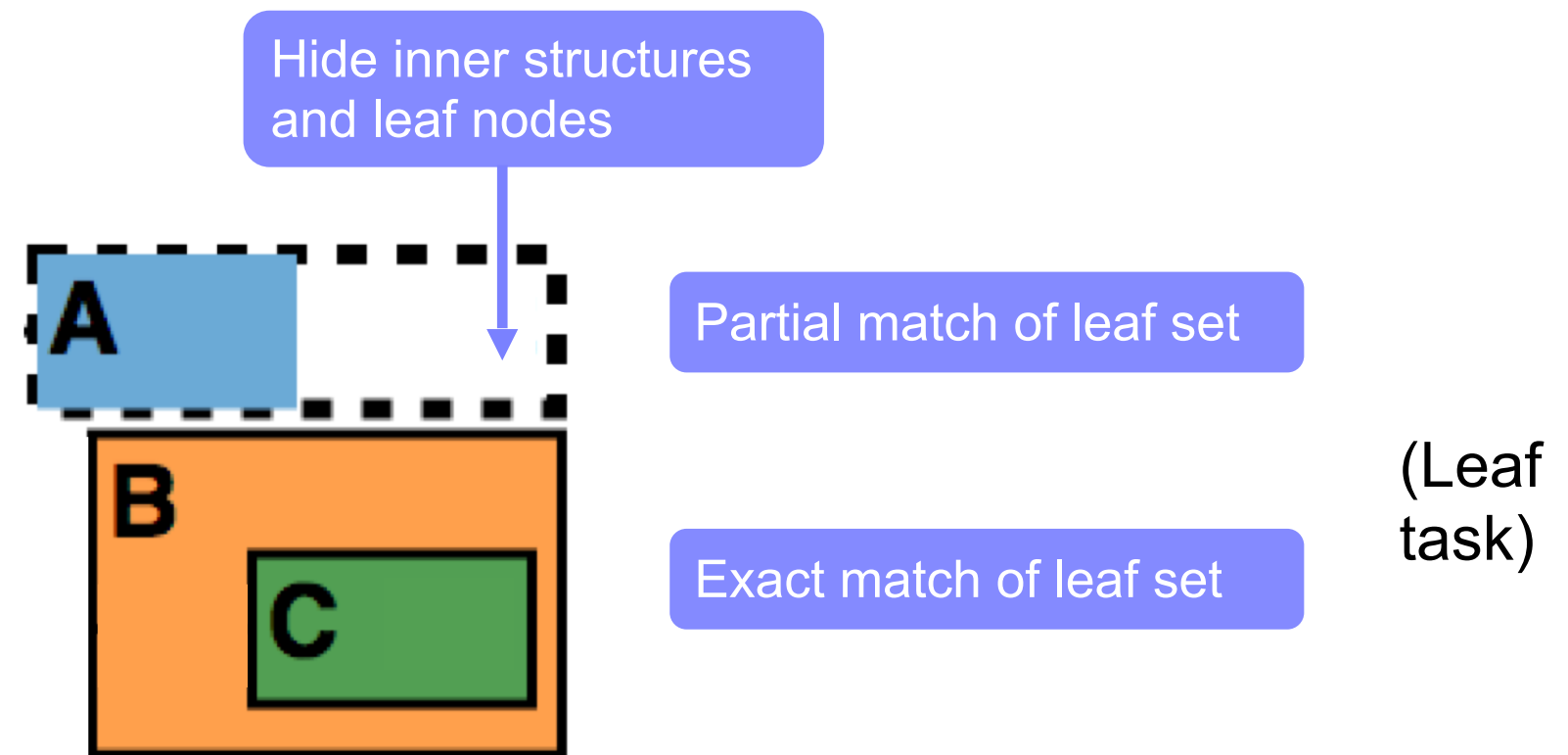


Visual design: focus + context



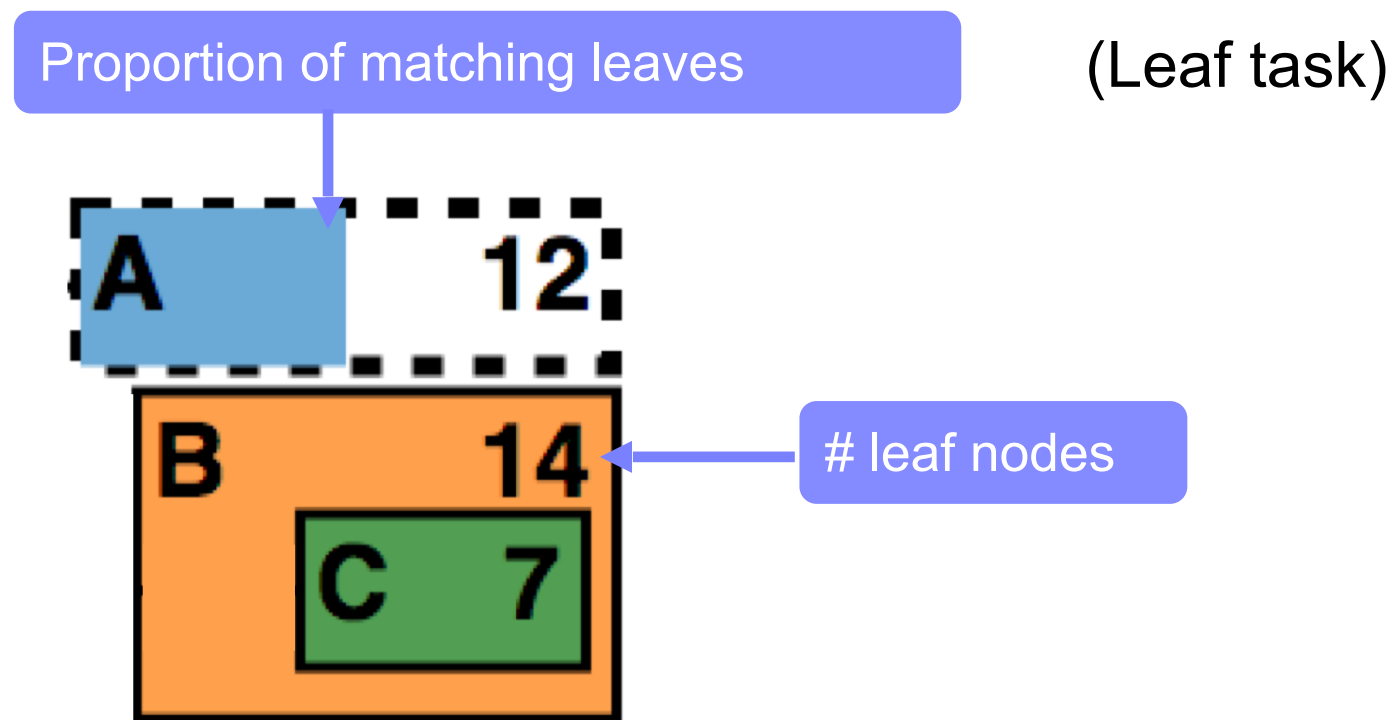
Visual design: focus + context

- Focus
 - Selected subtrees



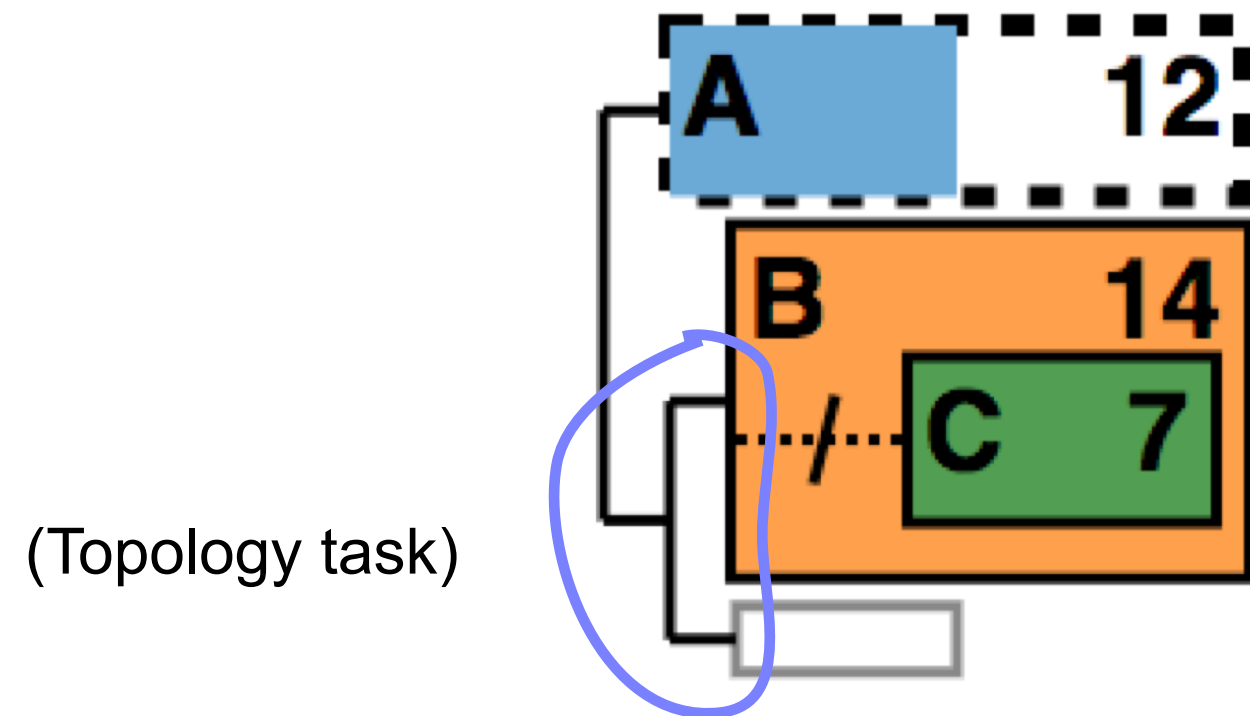
Visual design: focus + context

- Focus
 - Selected subtrees



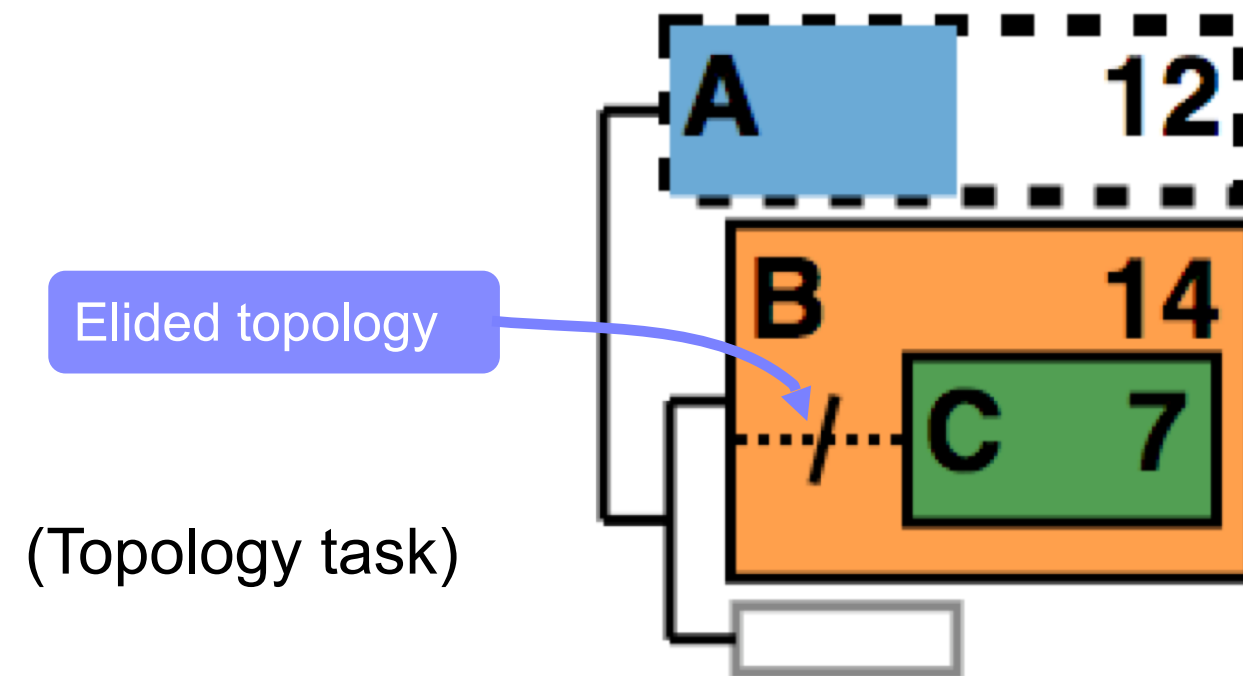
Visual design: focus + context

- Focus
 - Selected subtrees
 - Topological relationships between them



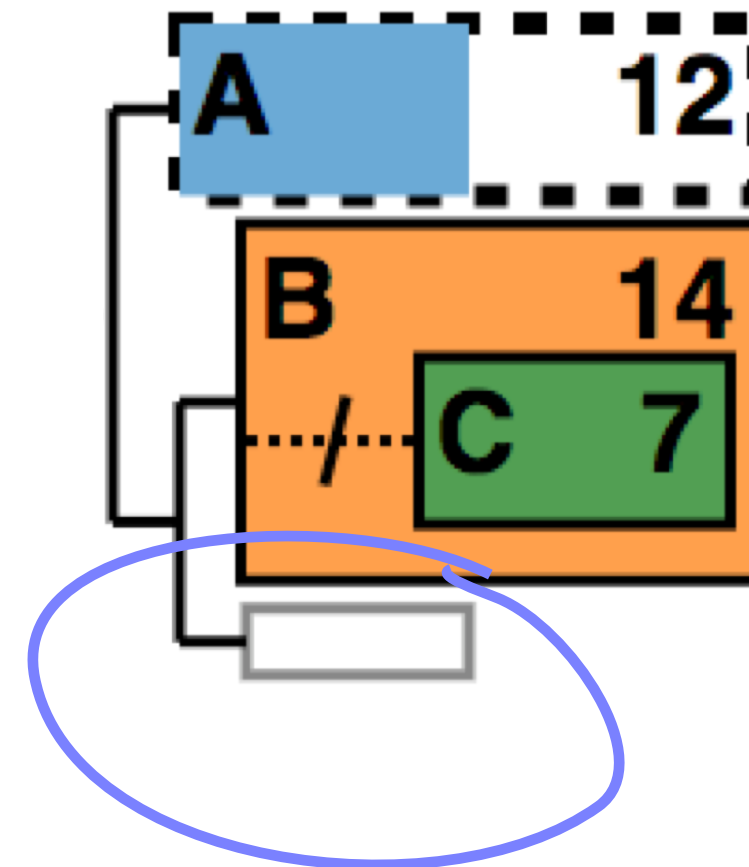
Visual design: focus + context

- Focus
 - Selected subtrees
 - Topological relationships between them



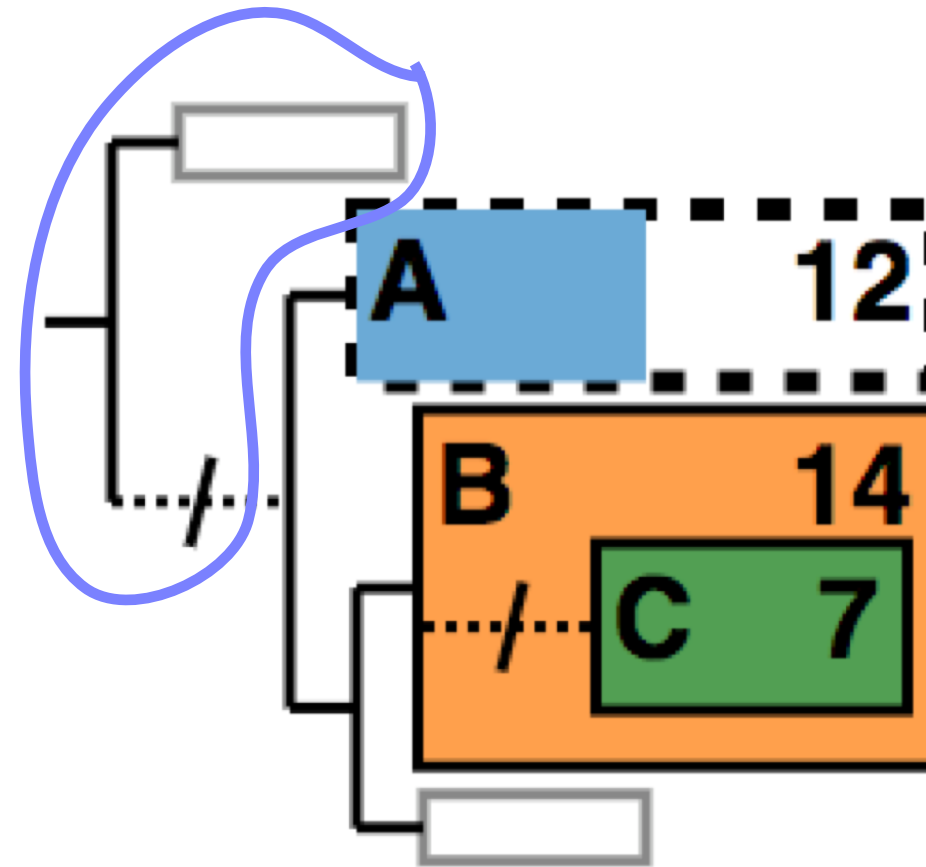
Visual design: focus + context

- Focus
 - Selected subtrees
 - Topological relationships between them
- Context
 - Neighboring subtrees



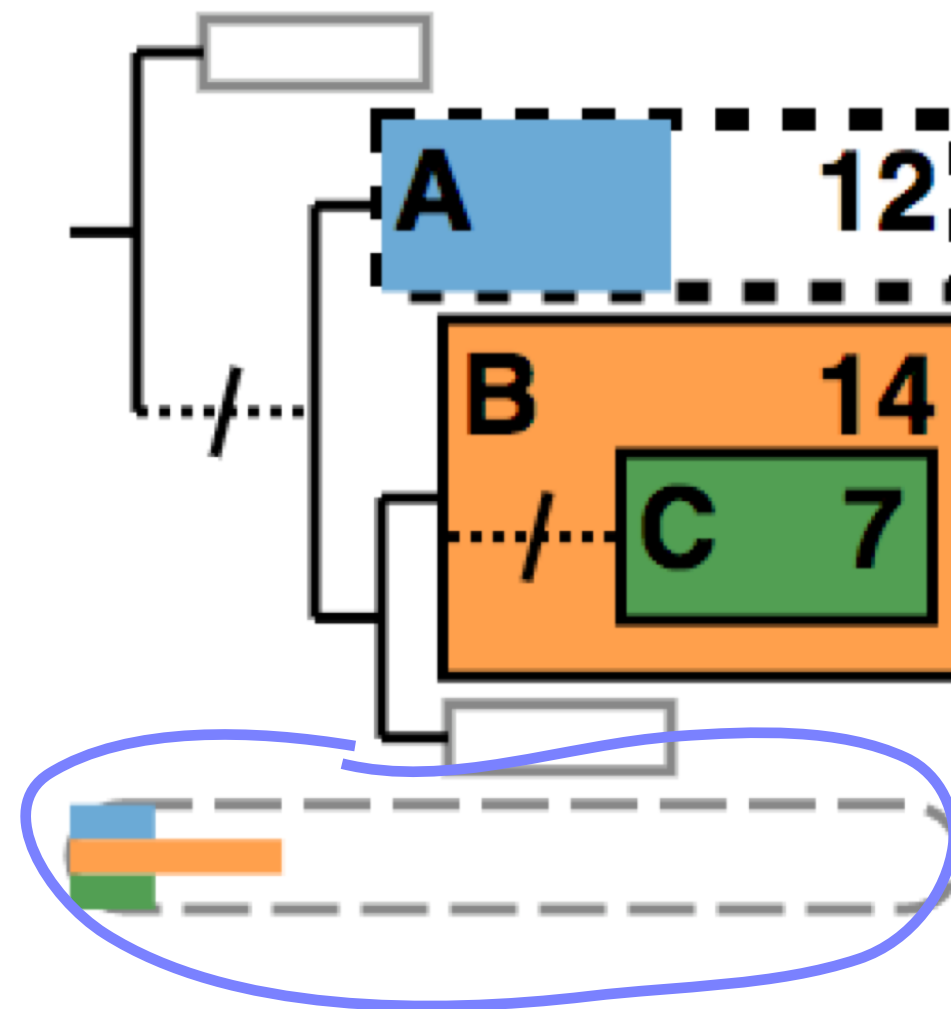
Visual design: focus + context

- Focus
 - Selected subtrees
 - Topological relationships between them
- Context
 - Neighboring subtrees
 - Upstream topology and root



Visual design: focus + context

- Focus
 - Selected subtrees
 - Topological relationships between them
- Context
 - Neighboring subtrees
 - Upstream topology and root
 - Missing leaf nodes



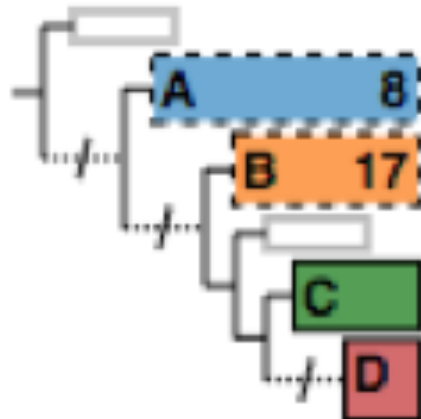
Visual design: algorithm adapts to space

- Show more info when space permitted
 - Labels
 - #leaf nodes
 - Neighboring blocks

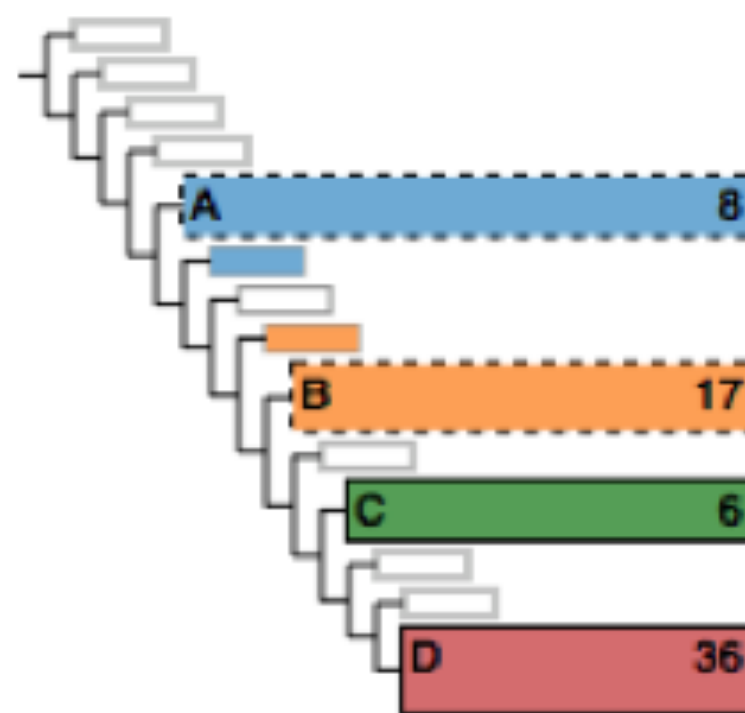
40x40 px



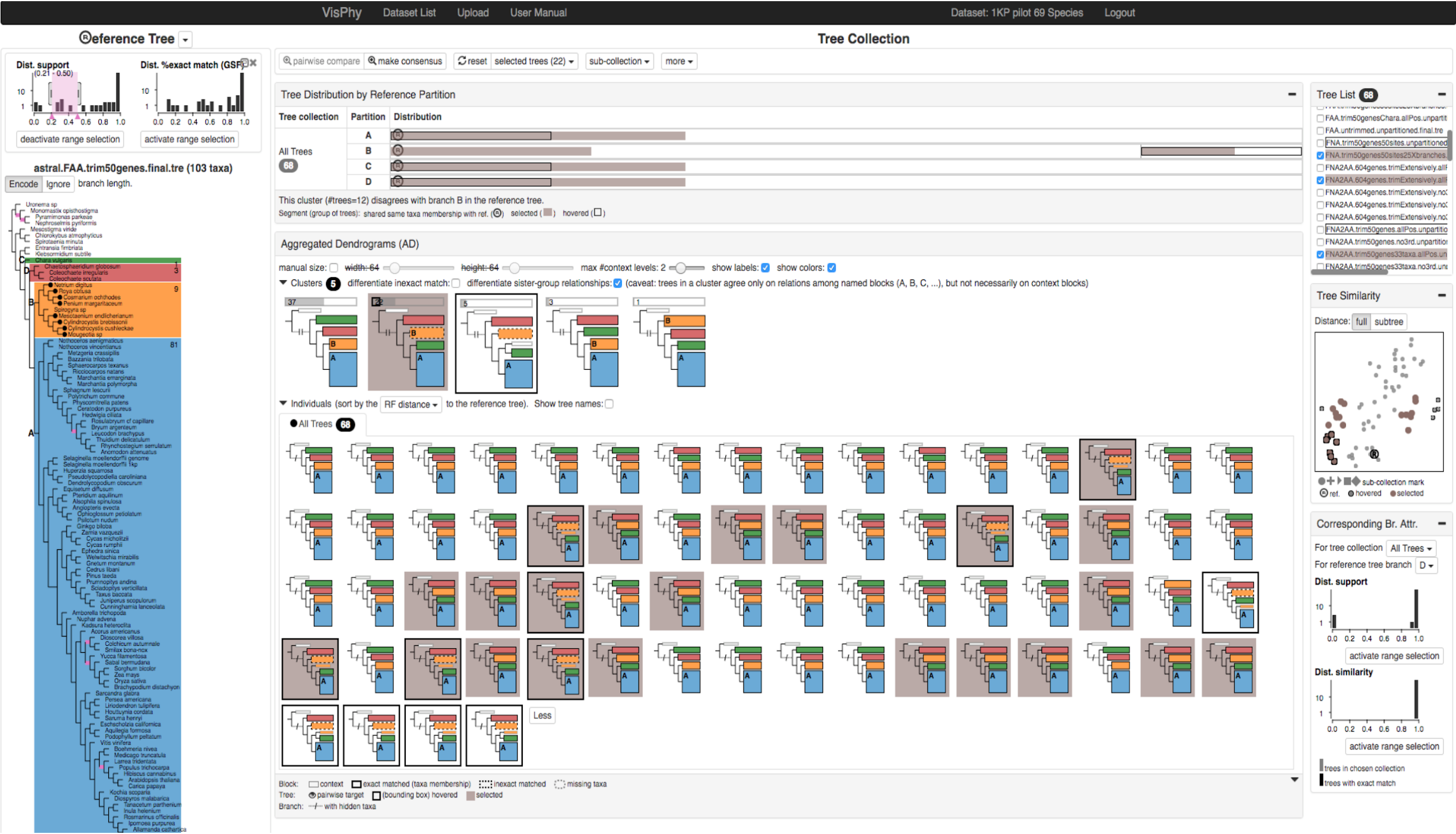
80x80 px



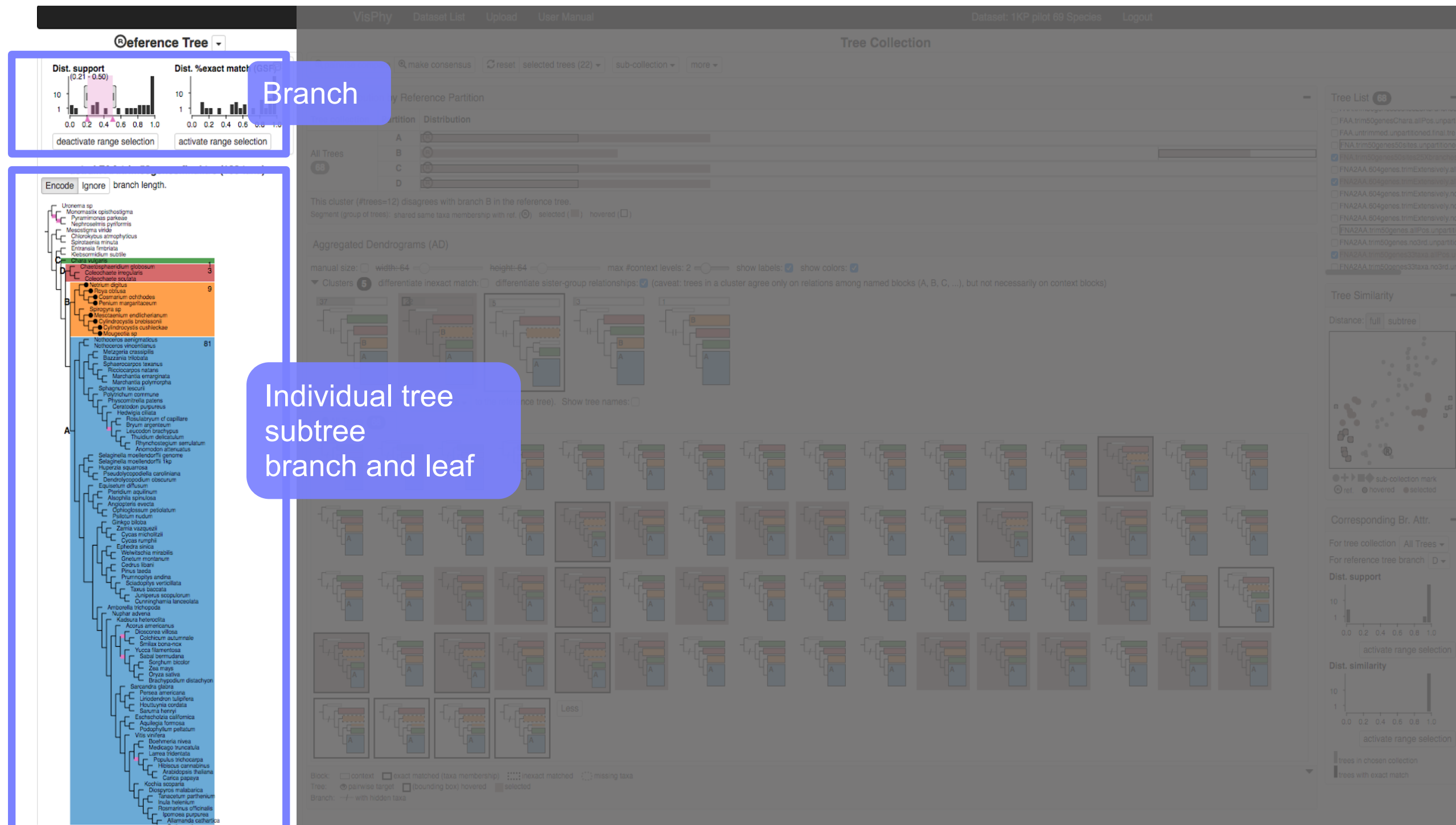
160x160 px



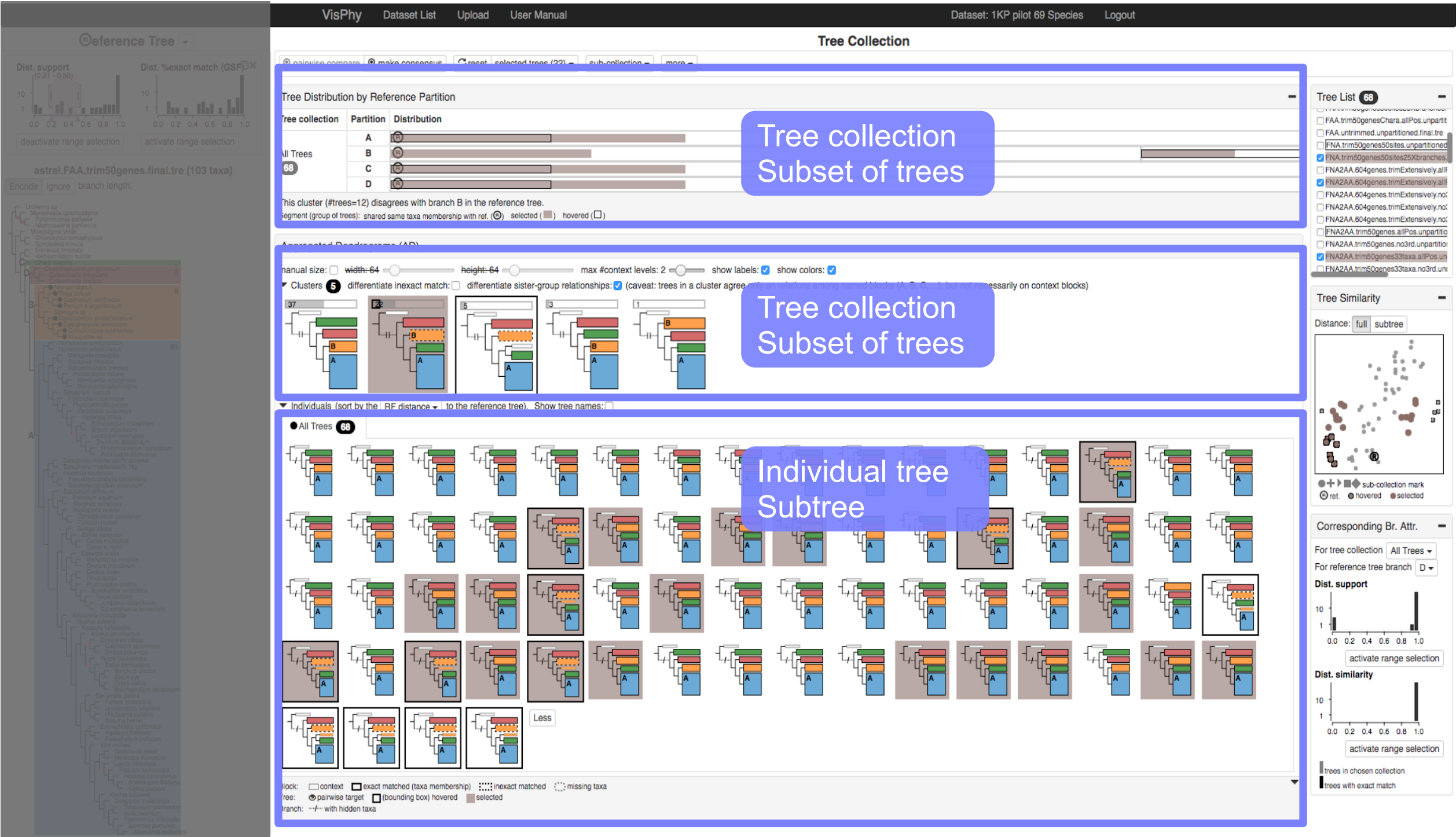
ADView Interface: Multi-level structure across views



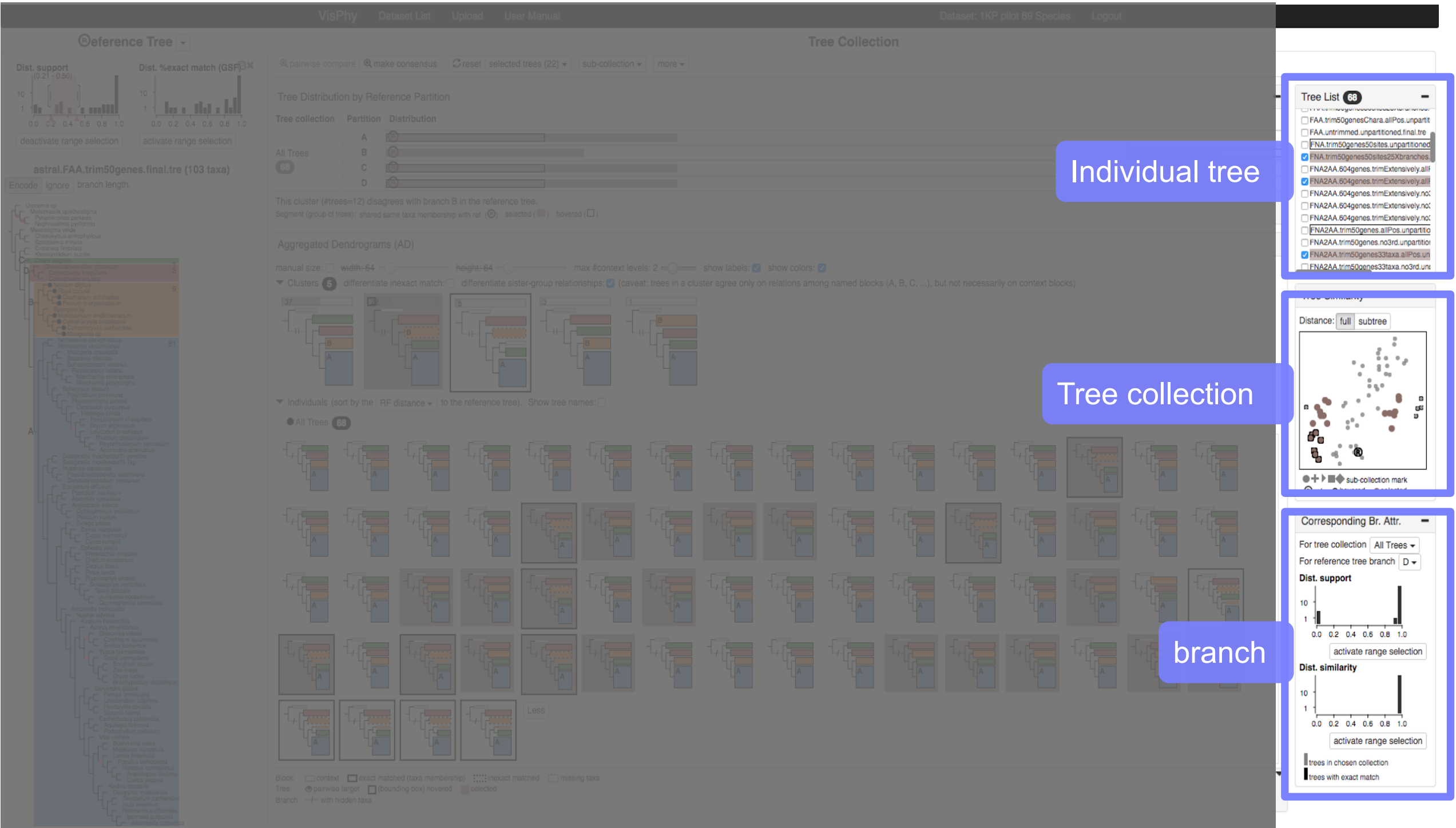
Multi-level structure across views



Interface walkthrough: tree collection main views



Interface walkthrough: tree collection aux. views

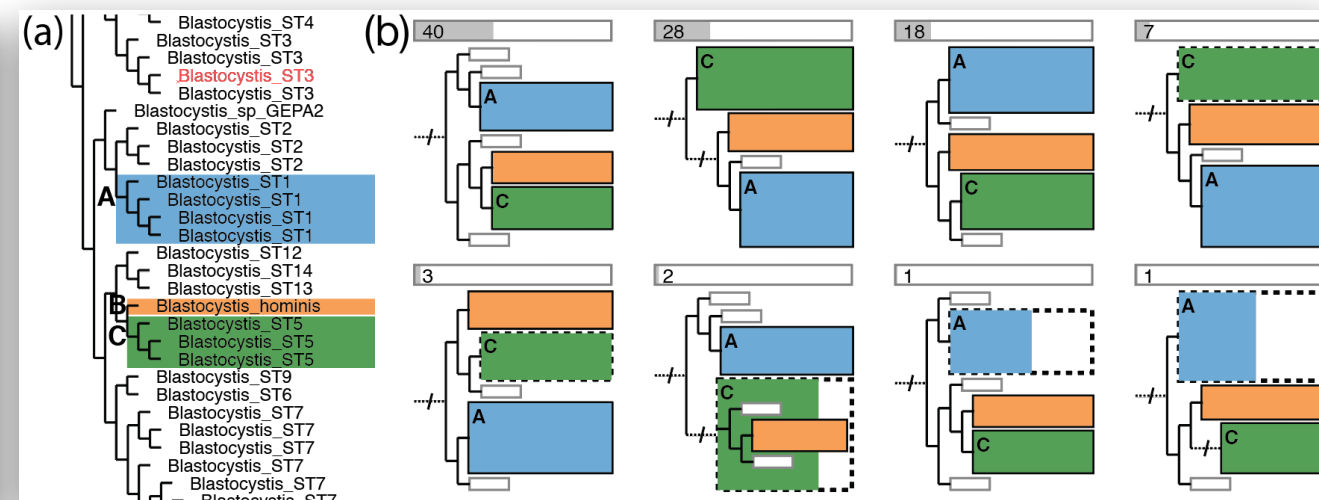
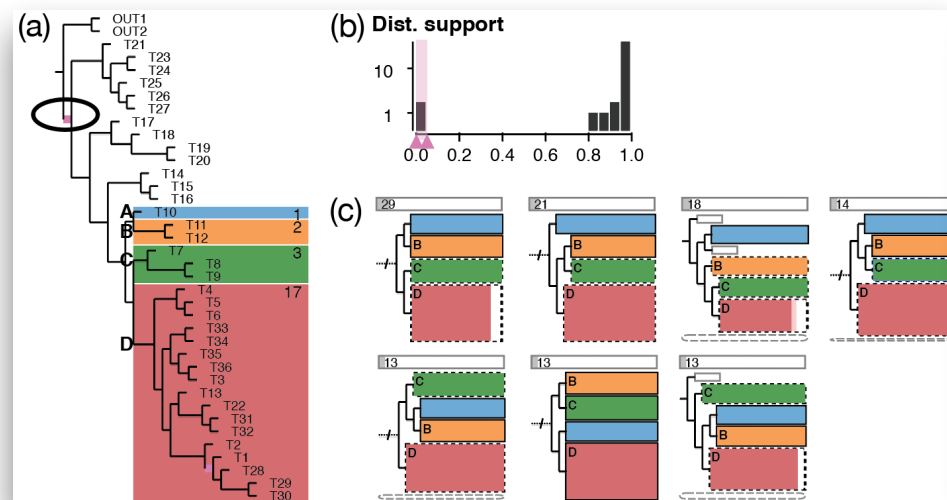
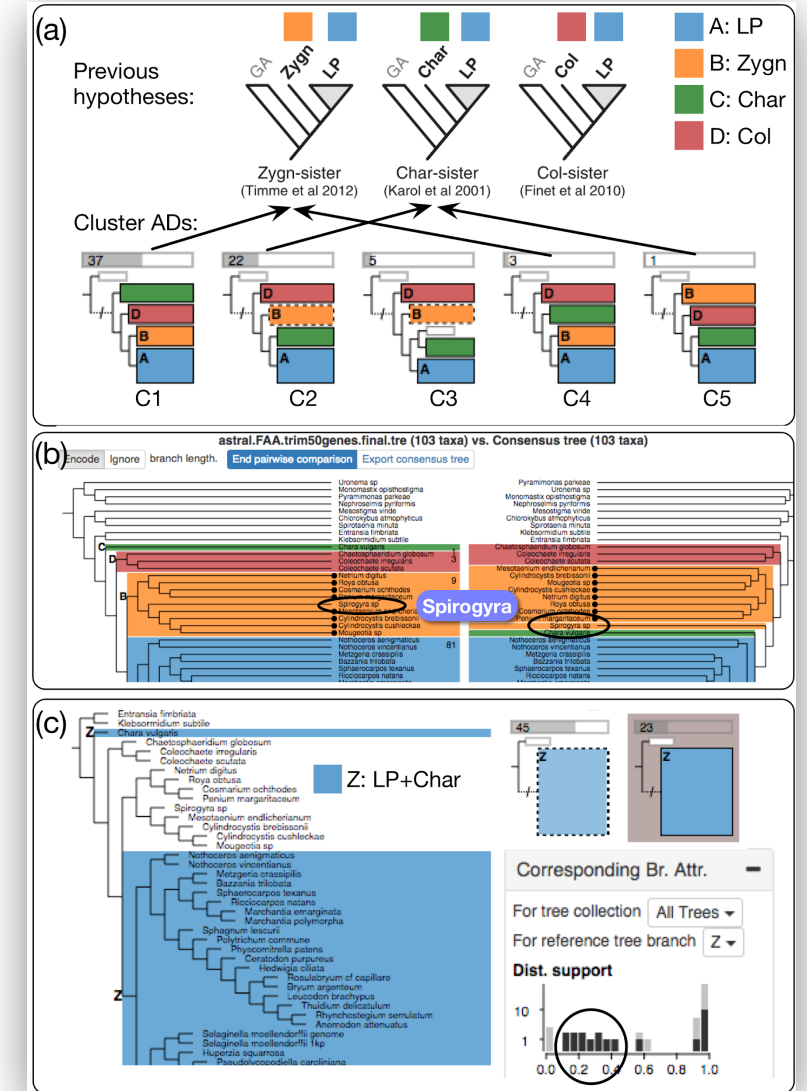


Validation with many biologists

- Work closely with a biology PhD student (second author)
- Demos, interviews and discussions
 - 10 biologists at different times throughout project

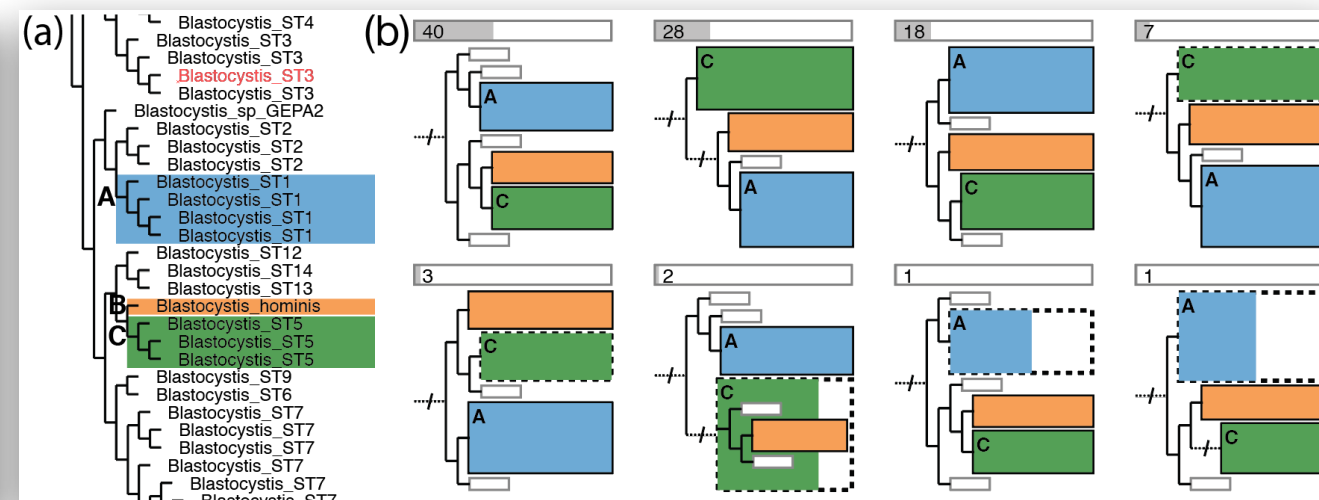
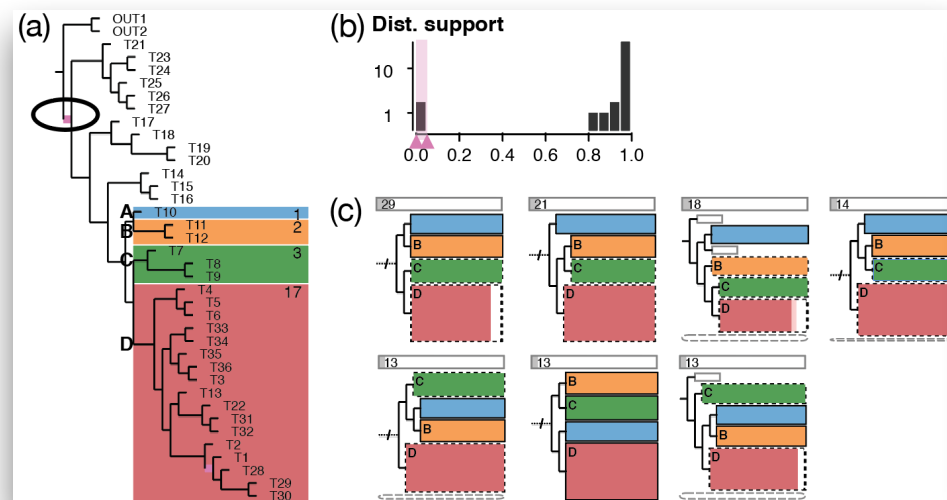
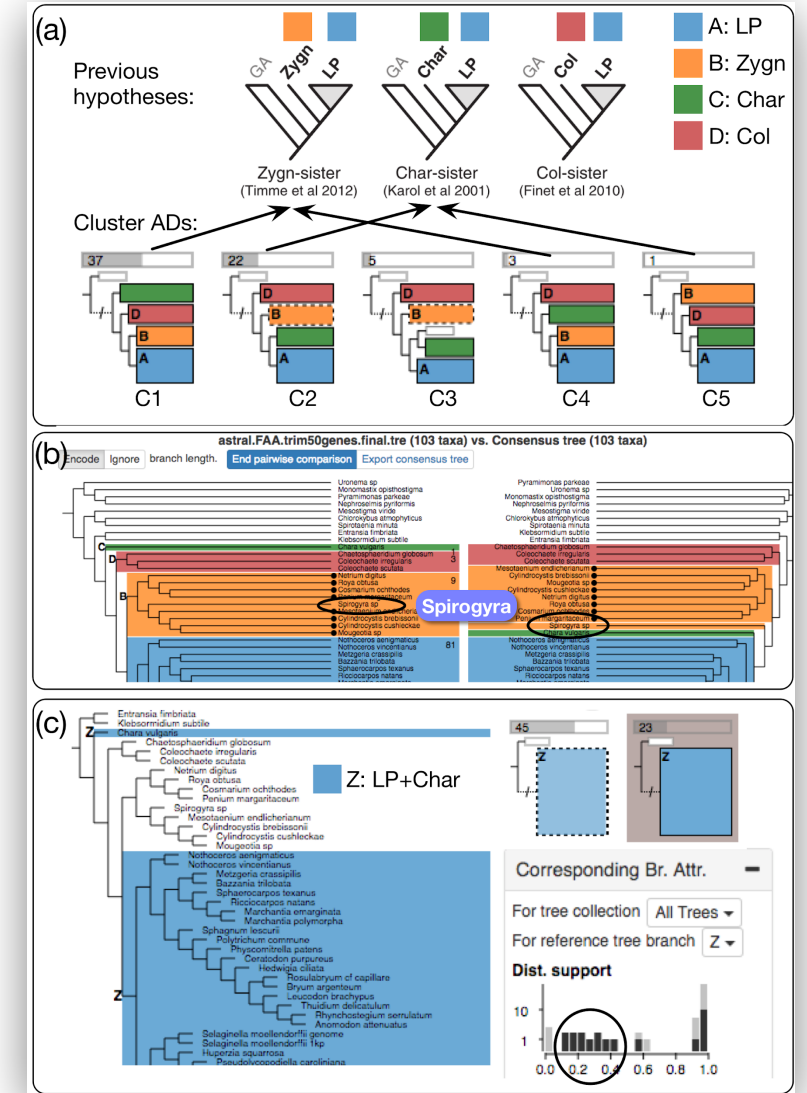
Validation with many biologists

- Work closely with a biology PhD student (second author)
- Demos, interviews and discussions
 - 10 biologists at different times throughout project
- User study sessions
 - 5 biologists
 - Using their own datasets



Validation with many biologists

- Work closely with a biology PhD student (second author)
- Demos, interviews and discussions
 - 10 biologists at different times throughout project
- User study sessions
 - 5 biologists
 - Using their own datasets
- Biologists confirmed
 - Validity of data and task abstractions
 - Utility of ADView



Q&A

Break: 3:40-3:50

Four case studies of problem-driven work

- in-car networks



- facilities management



- biology



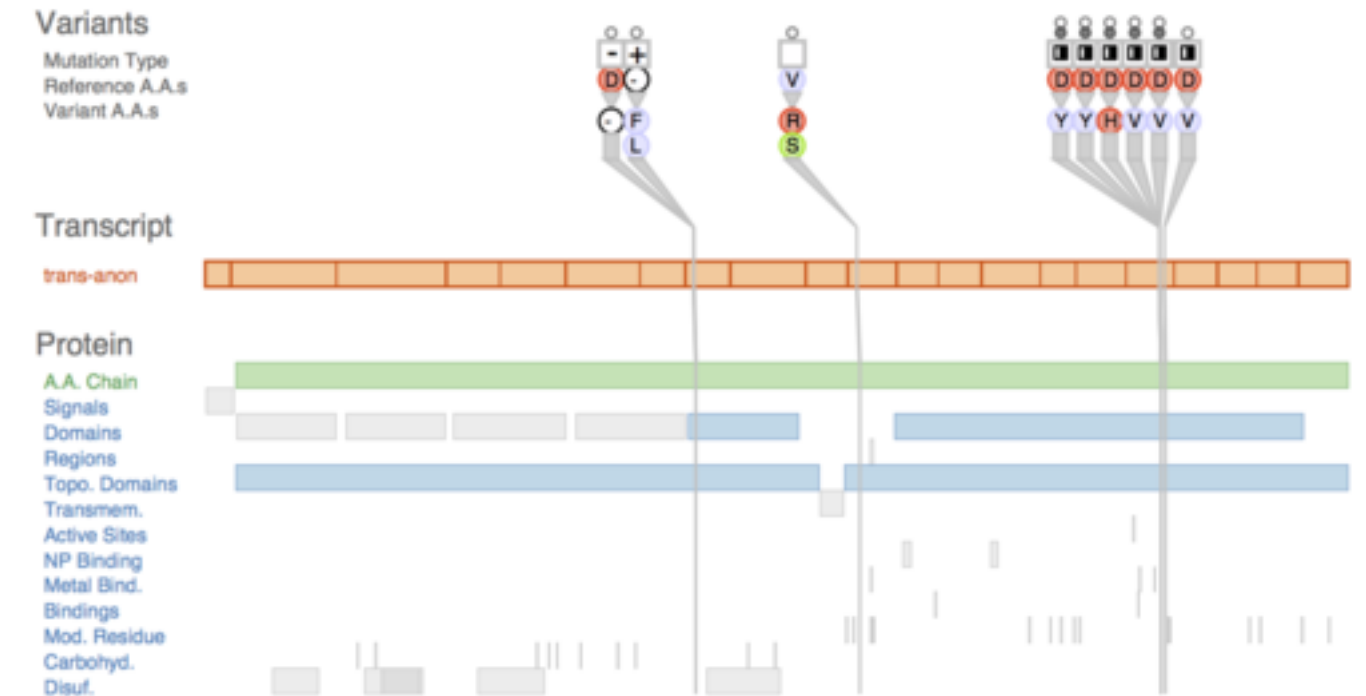
Variant View

Visualizing Sequence Variants in their Gene Context

joint work with:

Joel Ferstay, Cydney Nielsen

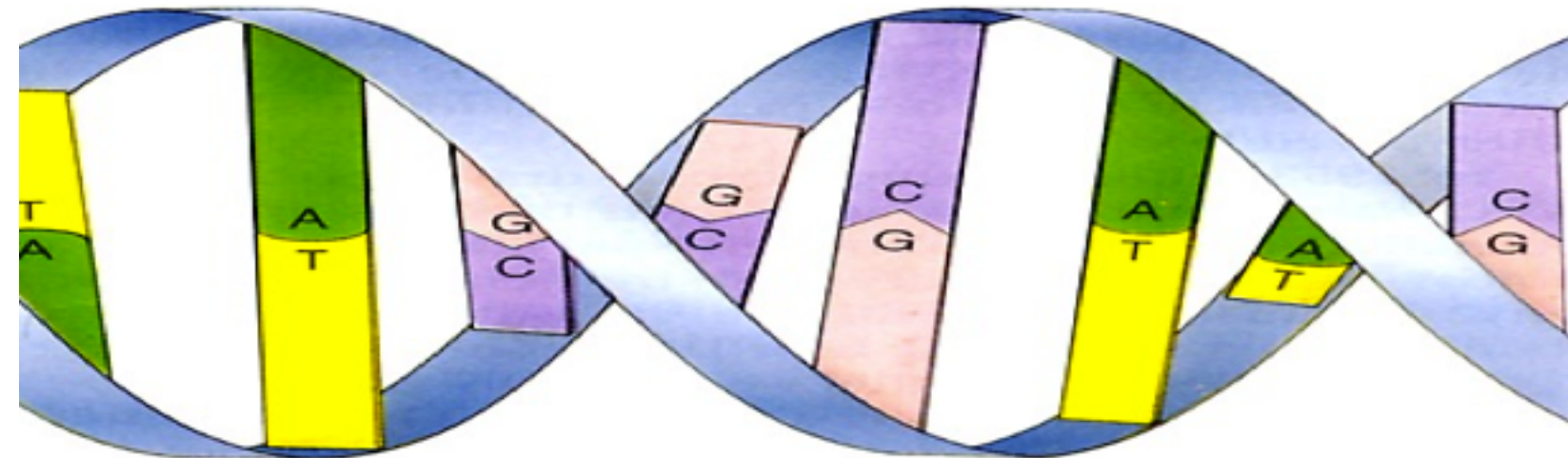
<http://www.cs.ubc.ca/labs/imager/tr/2012/VariantView/>



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 TG**G** TCA **ATA** CTT

Sample 2 Genome DNA: ATA TGA T**G**A ACA C**CT**

Harmful?

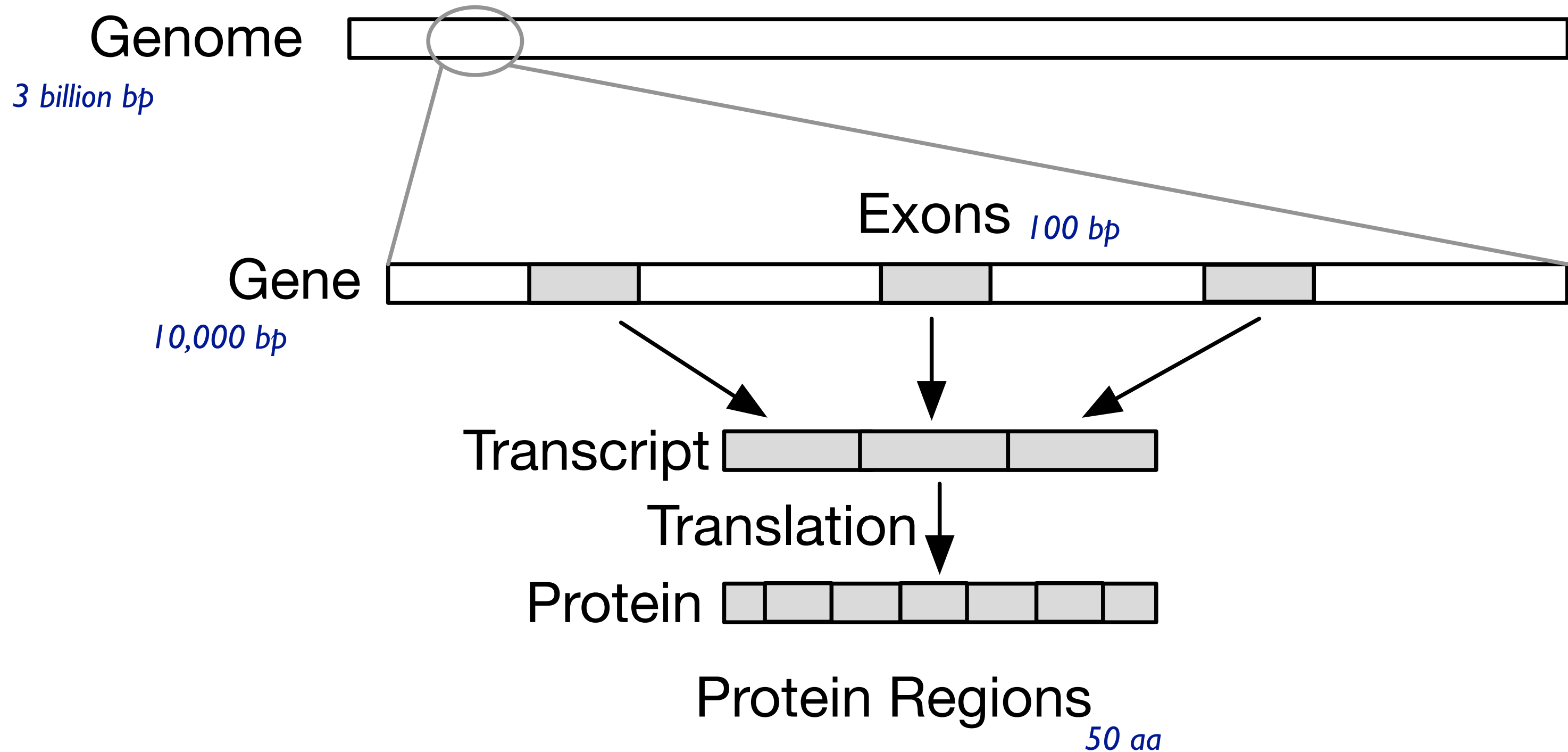
Harmless?

Cancer Research

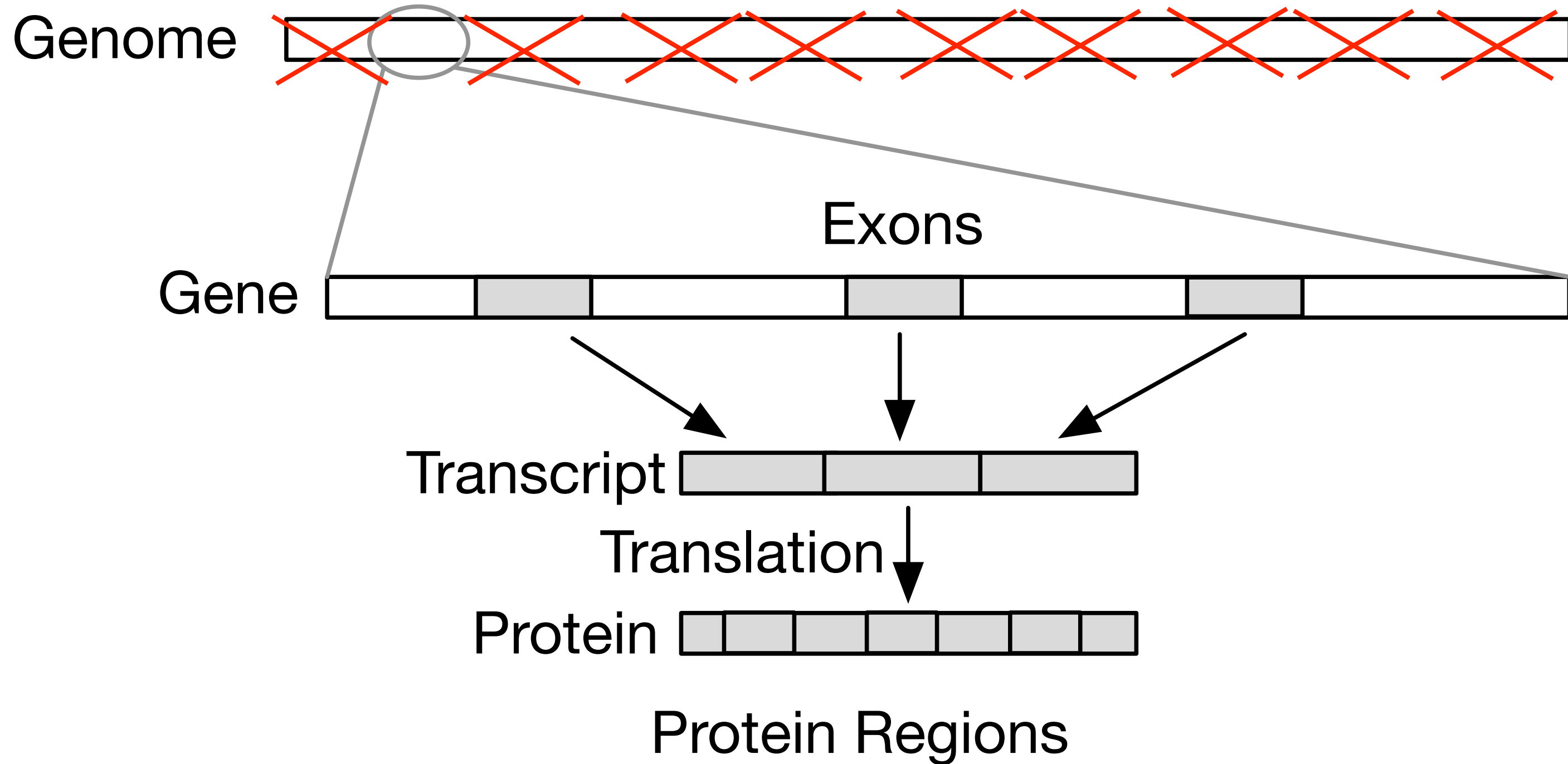
- collaboration with analysts at BC Genome Sciences Center
 - studying genetic basis of leukemia
- driving task
 - discover new candidate genes with harmful variants
- two big questions
 - what to show
 - data abstraction
 - challenge: enormous range of scales in the data
 - how to show it
 - visual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

Abstractions

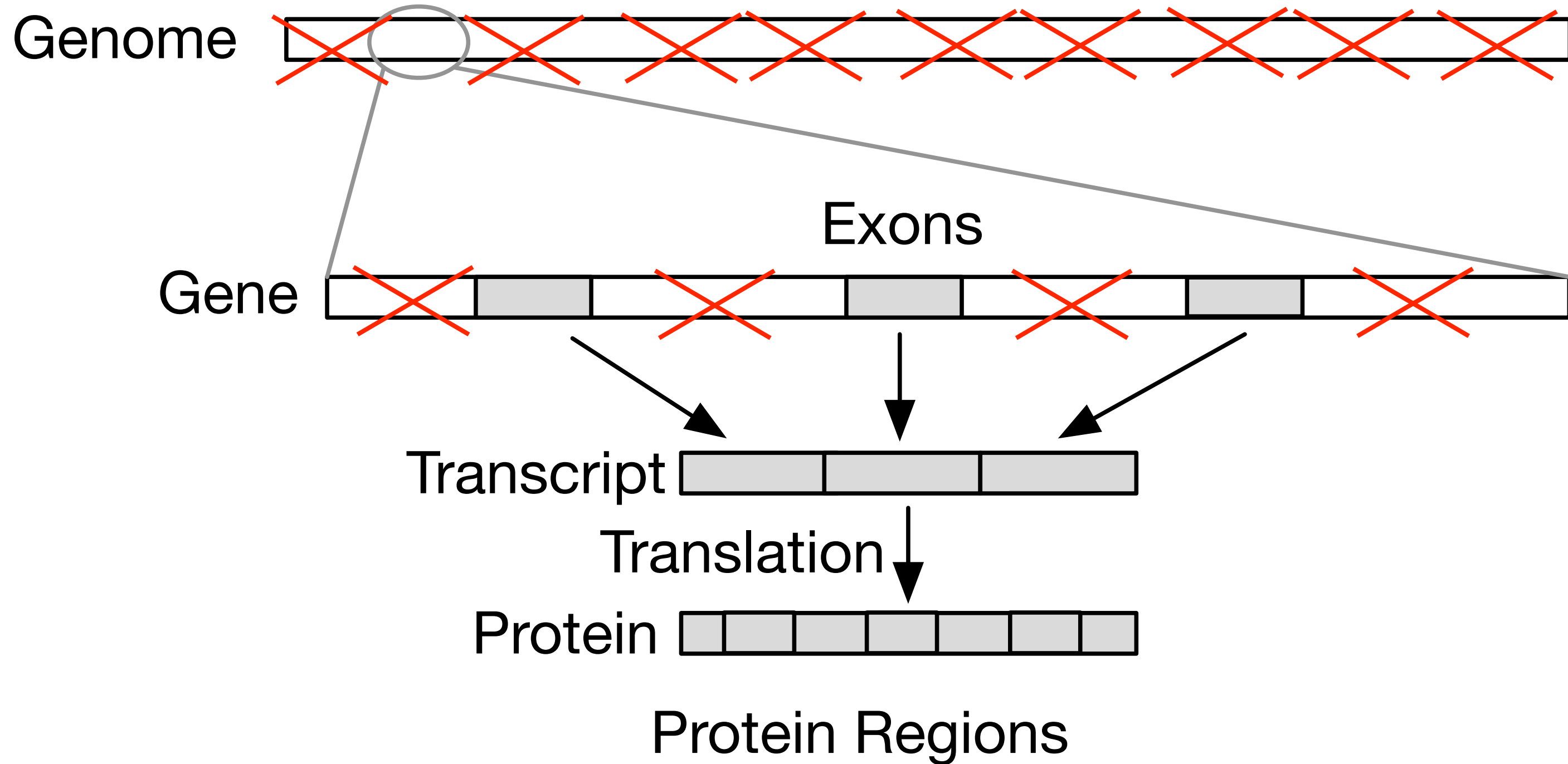
Data: Filtering to relevant biological levels and scales



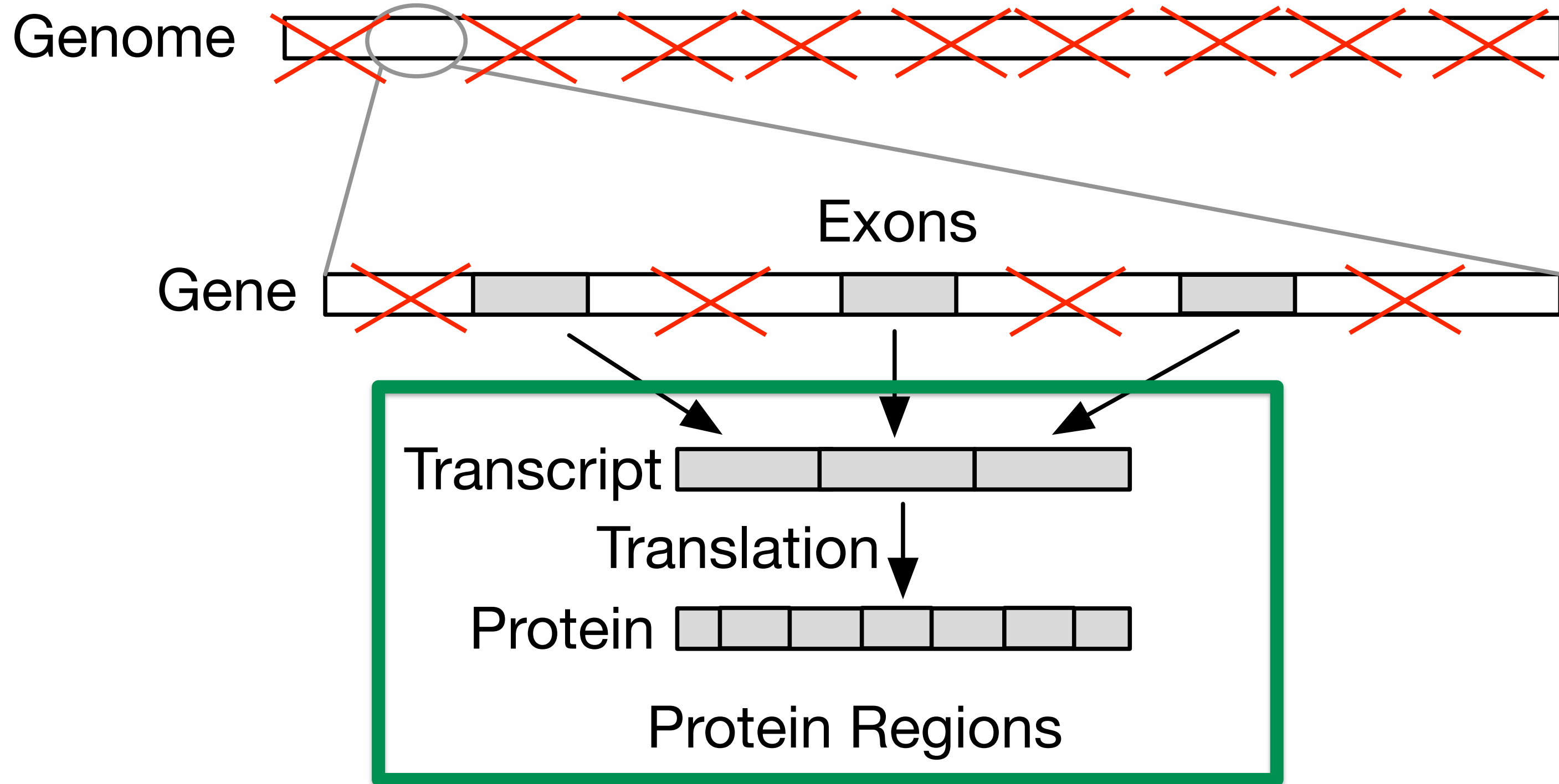
Filter out whole genome; keep genes



Filter out non-exon regions

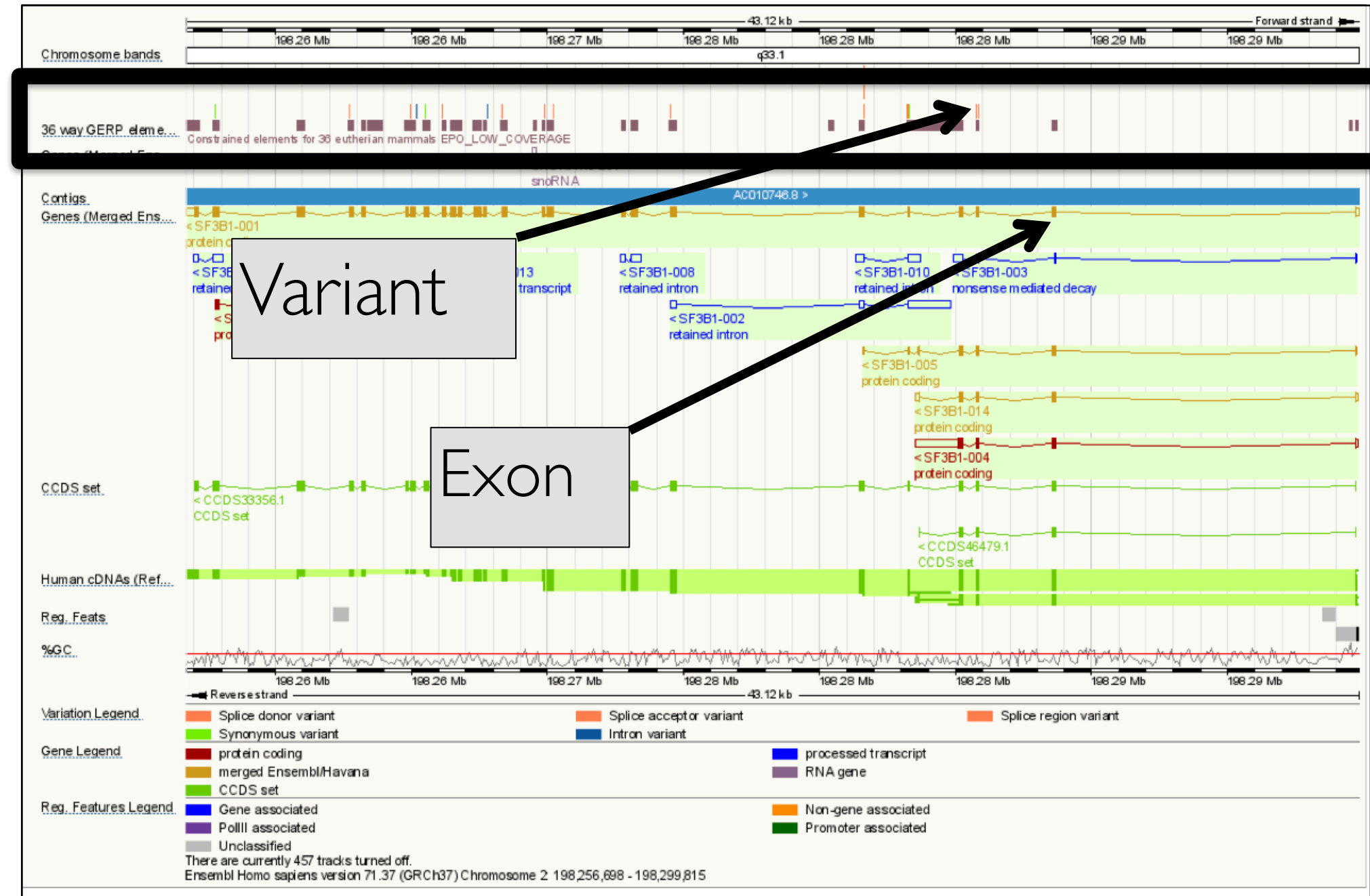


Data abstraction: highly filtered scope



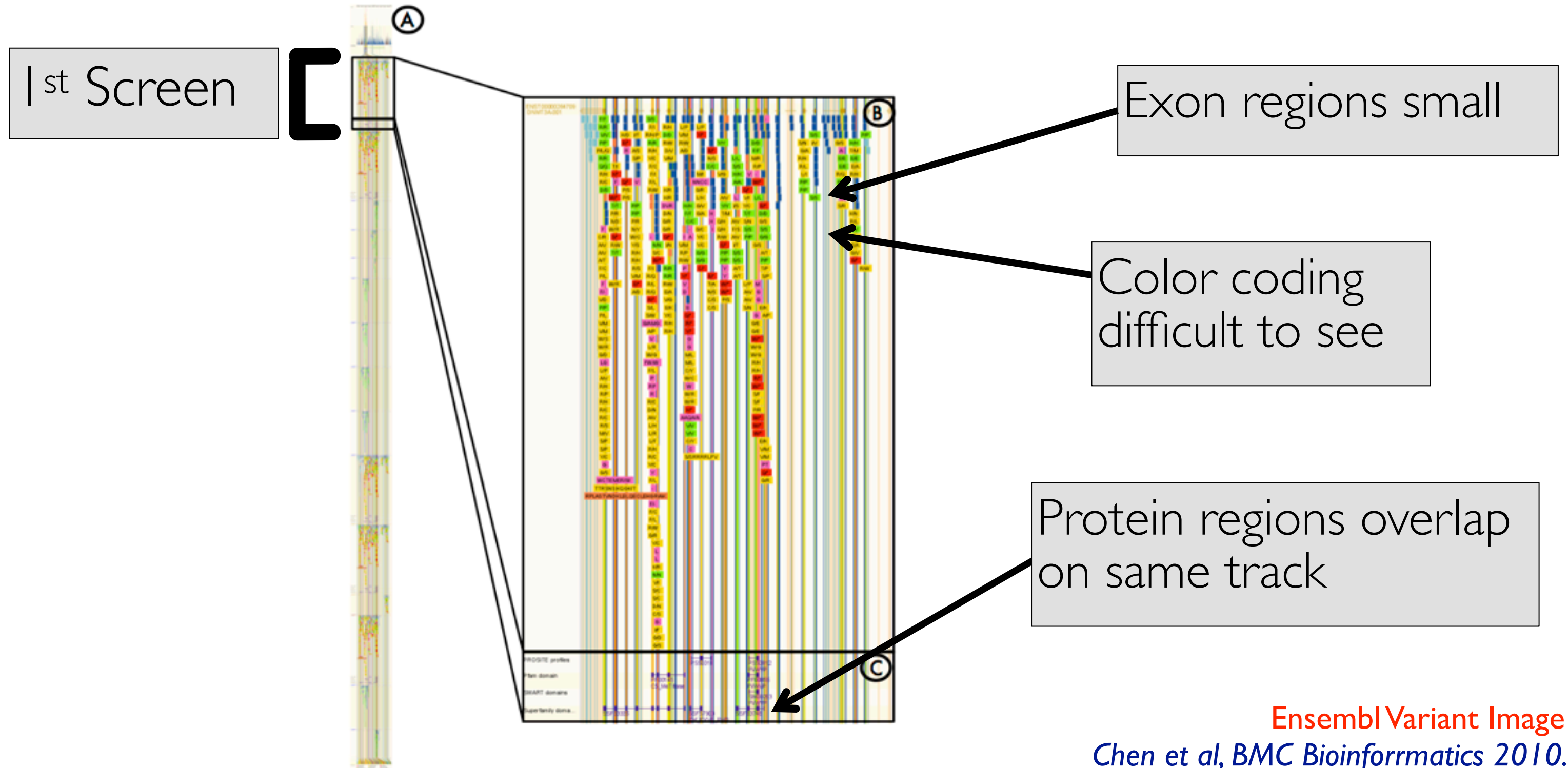
Dominant paradigm: genome browsers

- 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:

Submit

Alternative Transcripts:

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)

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. Seq.
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

Variant

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. Seq.
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

Variant View

Information-dense single gene view

Gene Search: Submit

Alternative Transcripts:

Variants

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

G

D

A

A

E

E

E

E

V

Y

S

V

Q

G

D

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. Seq.
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

Information-dense single gene view

Gene Search: Submit

Alternative Transcripts:

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)

Transcript
trans-anon

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

Variant Data

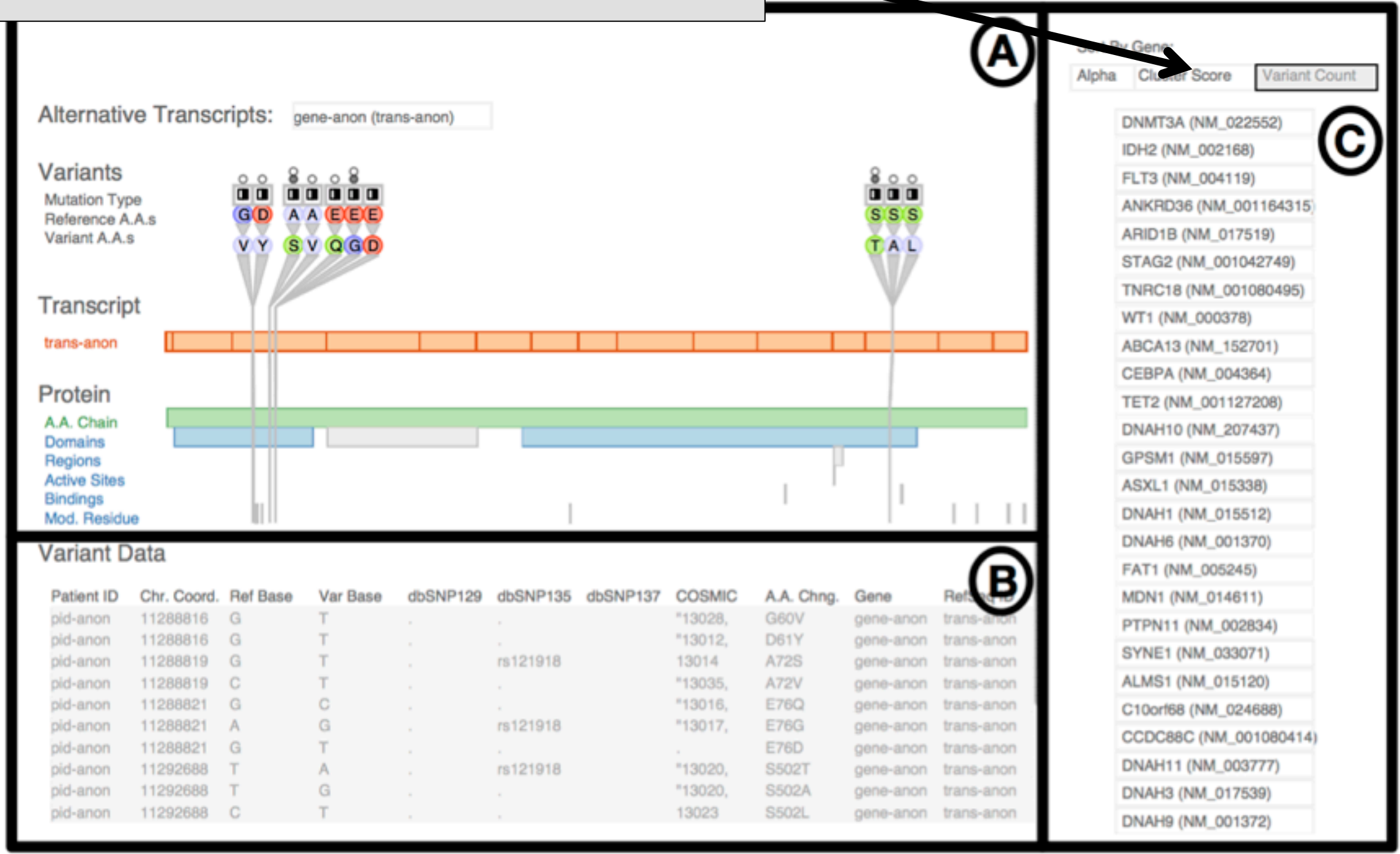
Patient ID	Chr. Coord.	Ref Base	Var Base	dbSNP129	dbSNP135	dbSNP137	COSMIC	A.A. Ch		
pid-anon	11288816	G	T	.	.	.	"13028,	G60V		
pid-anon	11288816	G	T	.	.	.	"13012,	D61Y		
pid-anon	11288819	G	T	.	rs121918	.	13014	A72S	gene-anon	trans-anon
pid-anon	11288819	C	T	.	.	.	"13035,	A72V	gene-anon	trans-anon
pid-anon	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

No need for pan and zoom

129

Variant View

Sorting metrics guide gene navigation



Variant View

Sorting metrics guide gene navigation

Control what shows up here

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

Gene	RefSeq ID	dbSNP	COSMIC	A.A. Chng.	Gene	RefSeq ID
pid-anon	11288819	C	T	.	11288819	11288819
pid-anon	11288821	G	C	.	11288821	11288821
pid-anon	11288821	A	G	.	rs121918	11288821
pid-anon	11288821	G	T	.	11288821	11288821
pid-anon	11292688	T	A	.	rs121918	11292688
pid-anon	11292688	T	G	.	11292688	11292688
pid-anon	11292688	C	T	.	11292688	11292688

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:

Variants

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

G

D

A

A

E

E

E

E

V

Y

S

V

Q

G

D

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. Seq.
pid-anon	11	28816	G	T	.	.	.	"13028,	G60V	gene-anon	trans-anon
pid-anon	11	28816	G	T	.	.	.	"13012,	D61Y	gene-anon	trans-anon
pid-anon	11	28819	G	T	.	rs121918	.	"13014,	A72S	gene-anon	trans-anon
pid-anon	11	28819	C	T	.	.	.	"13035,	A72S	gene-anon	trans-anon
pid-anon	11	28821	G	C	.	.	.	"13016,	E76Q	gene-anon	trans-anon
pid-anon	11	28821	A	G	.	rs121918	.	"13017,	E76G	gene-anon	trans-anon
pid-anon	11	28821	G	T	.	.	.	.	E76D	gene-anon	trans-anon
pid-anon	11	29288	T	A	.	rs121918	.	"13020,	S502T	gene-anon	trans-anon
pid-anon	11	29288	T	G	.	.	.	"13020,	S502A	gene-anon	trans-anon
pid-anon	11	29288	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)

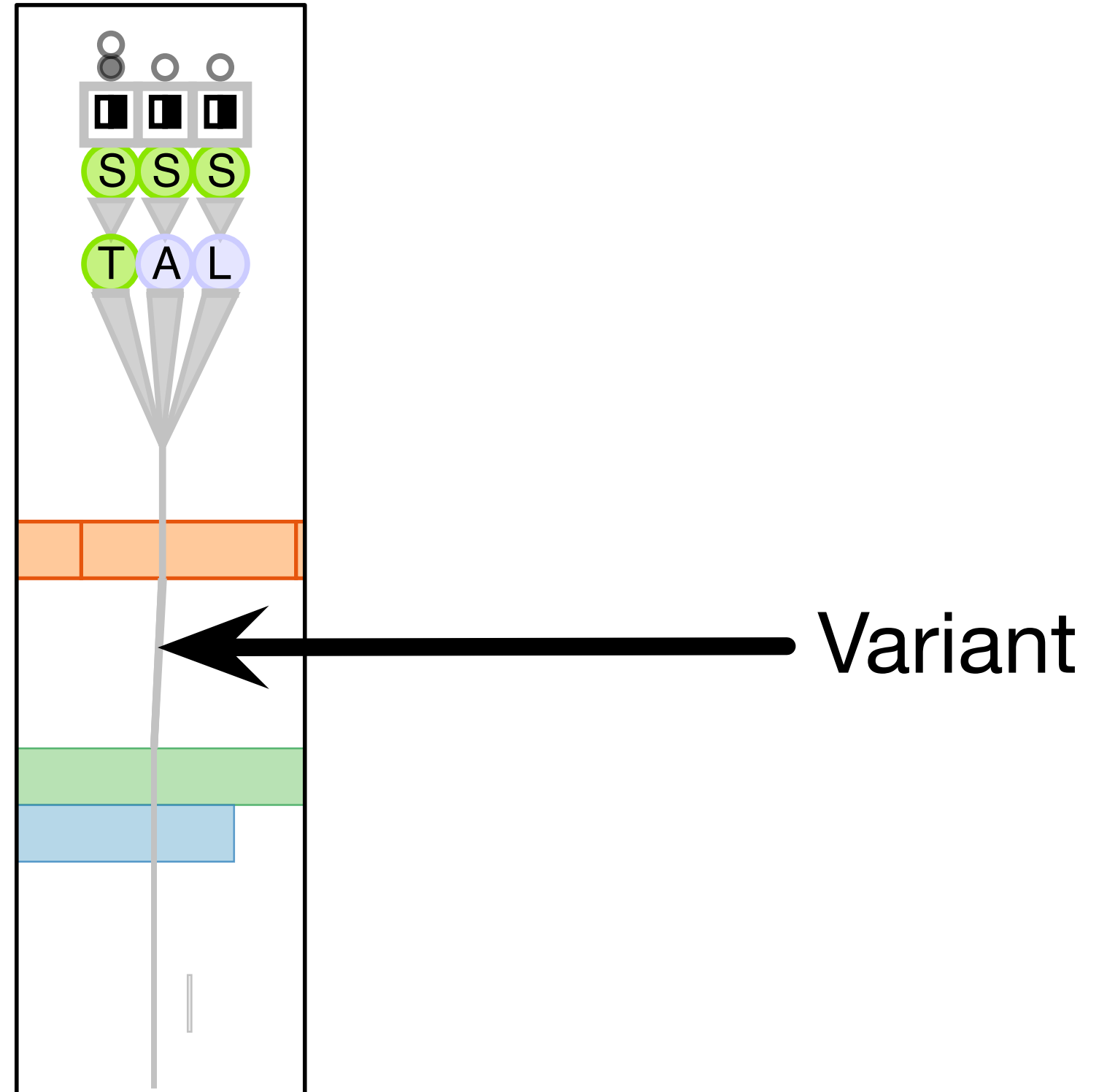
DNAH9 (NM_001372)

Peripheral supporting data

132

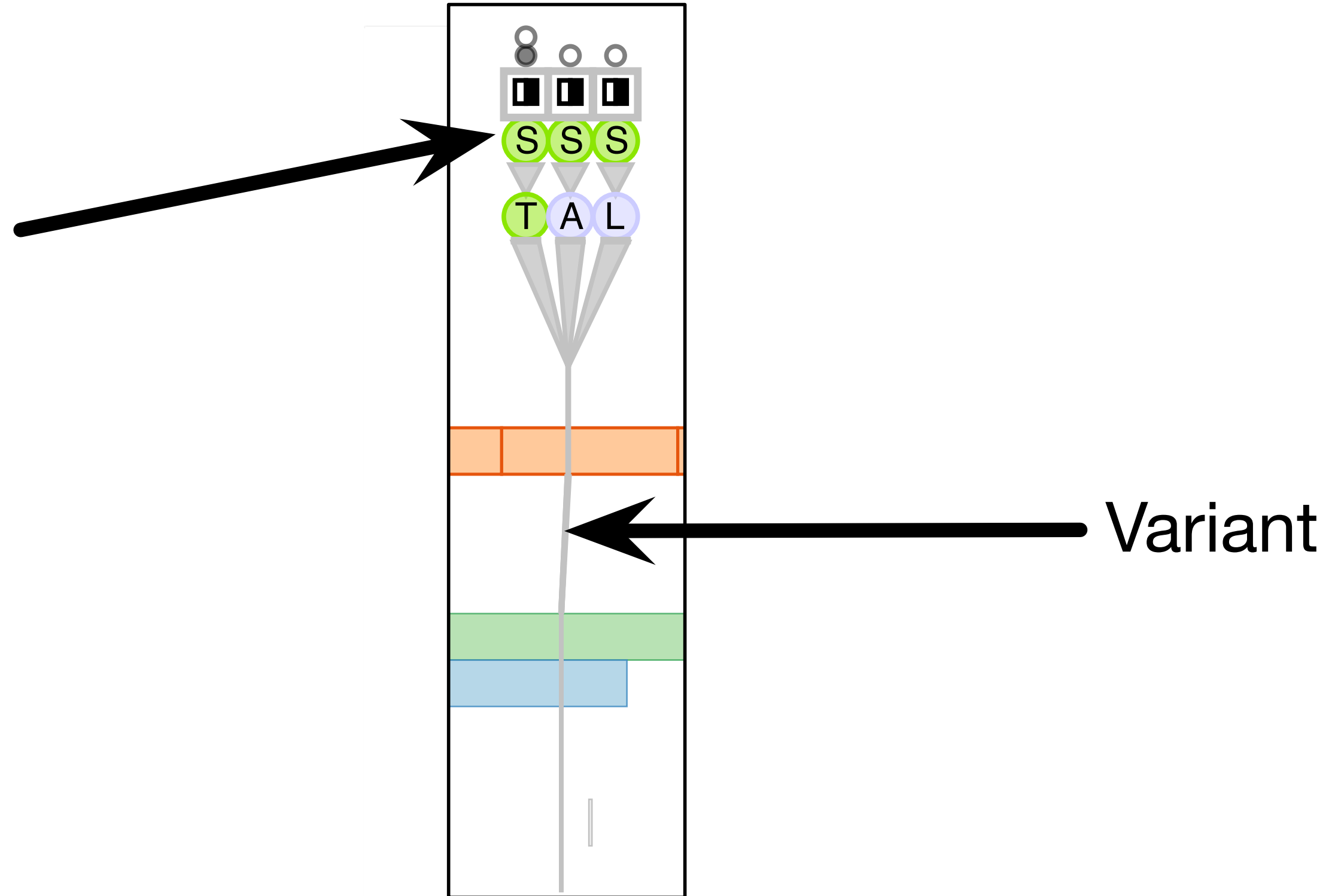
Design information-dense visual encoding

- 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

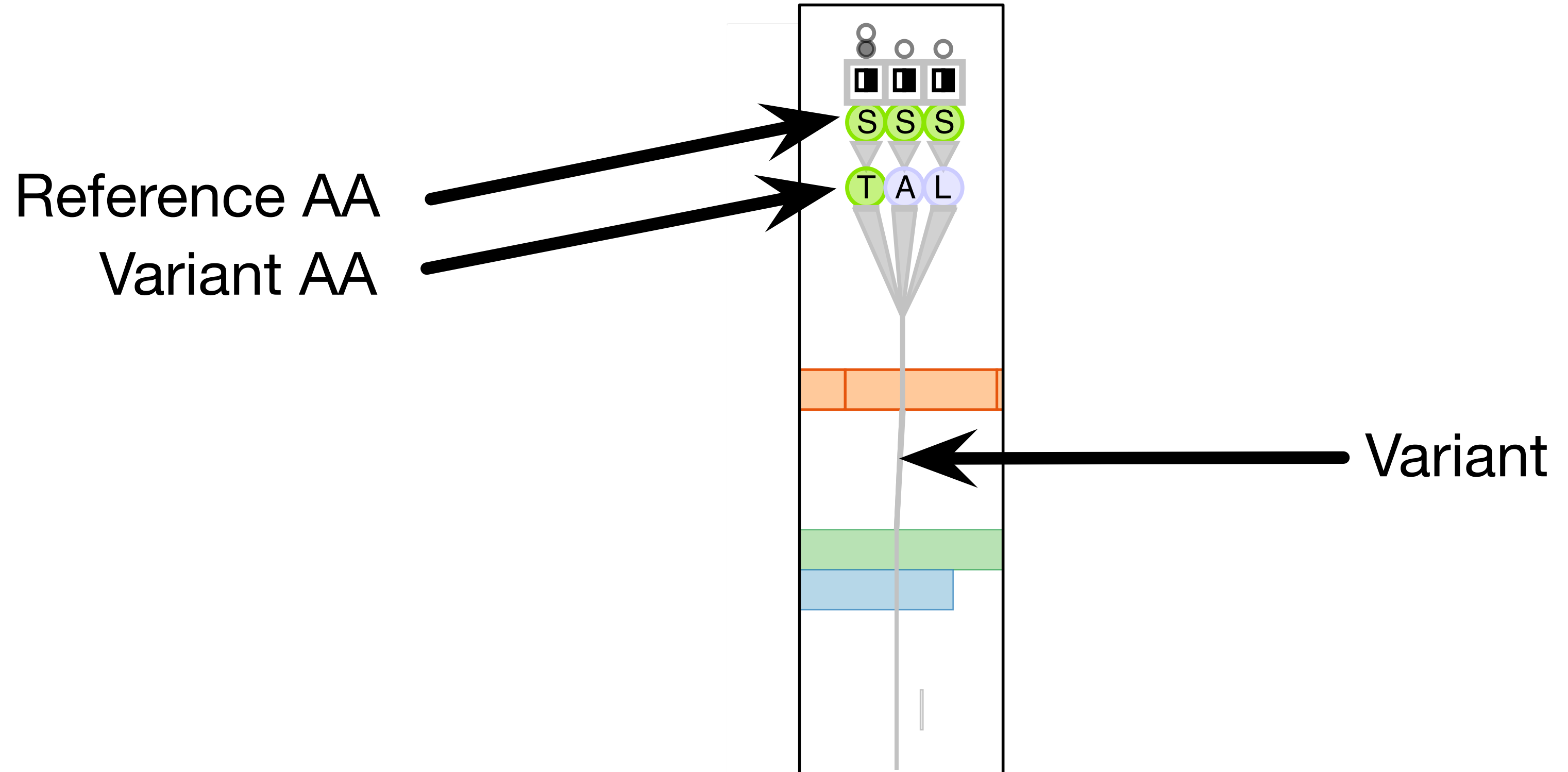


Design information-dense visual encoding

Reference AA



Design information-dense visual encoding



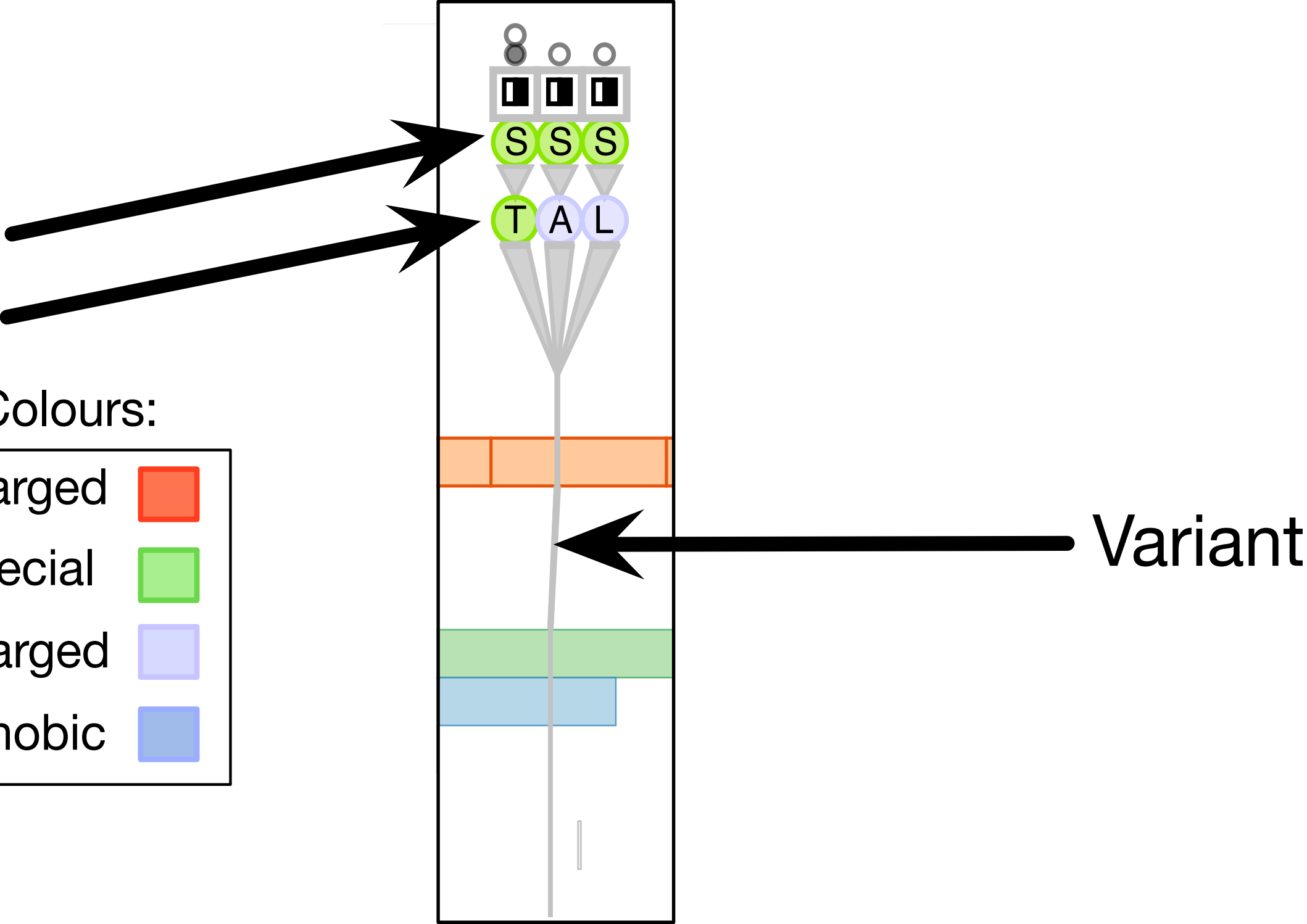
Design information-dense visual encoding

Reference AA

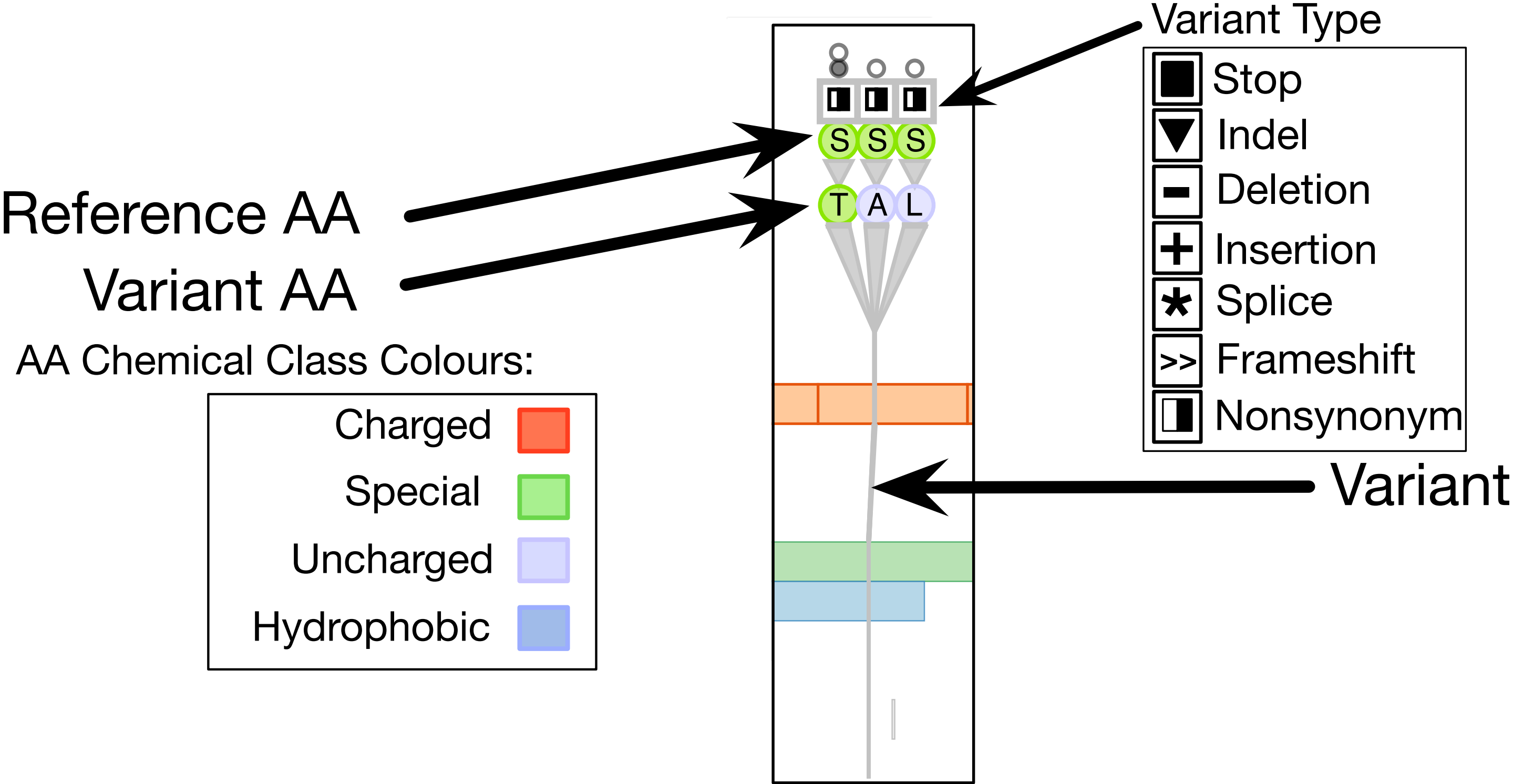
Variant AA

AA Chemical Class Colours:

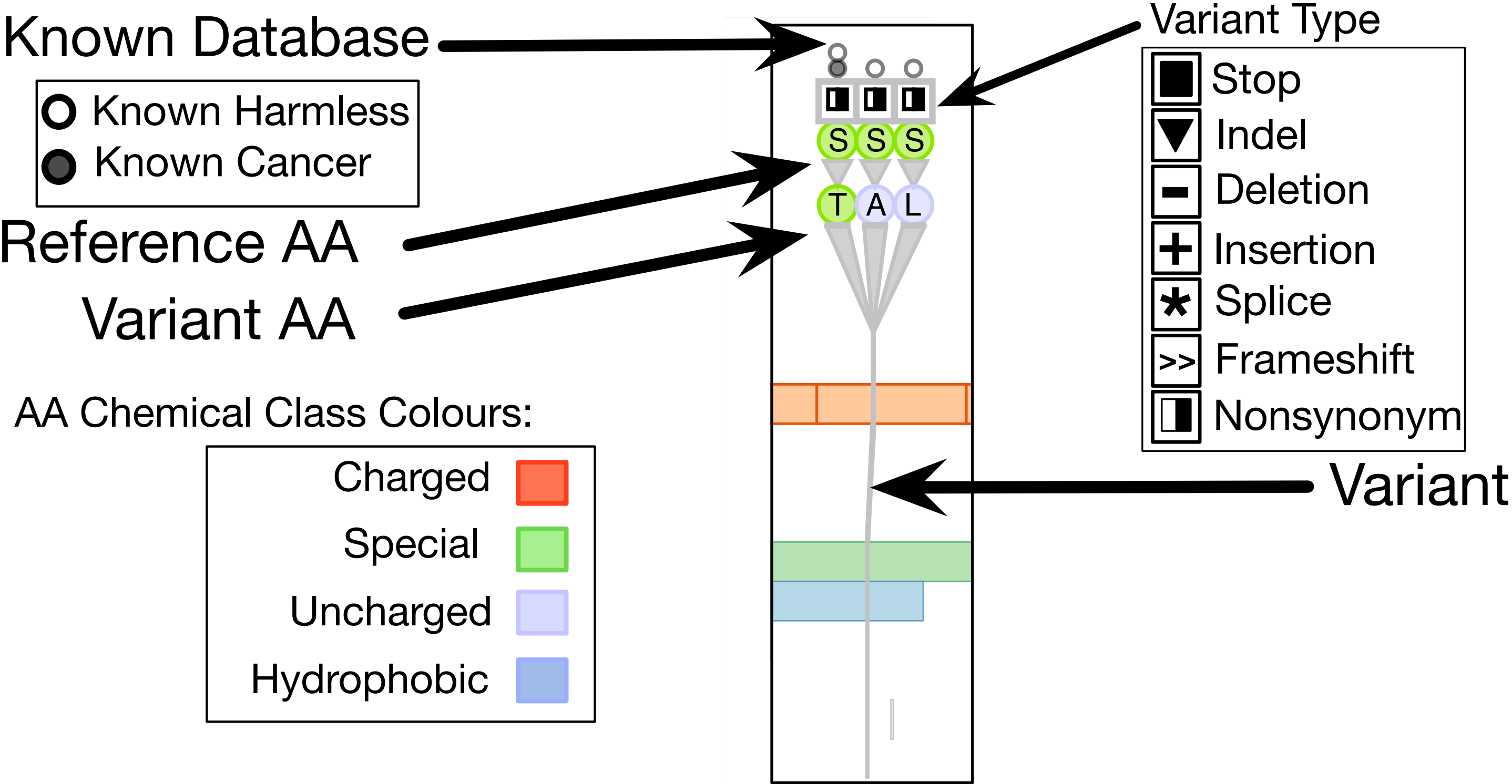
Charged	
Special	
Uncharged	
Hydrophobic	



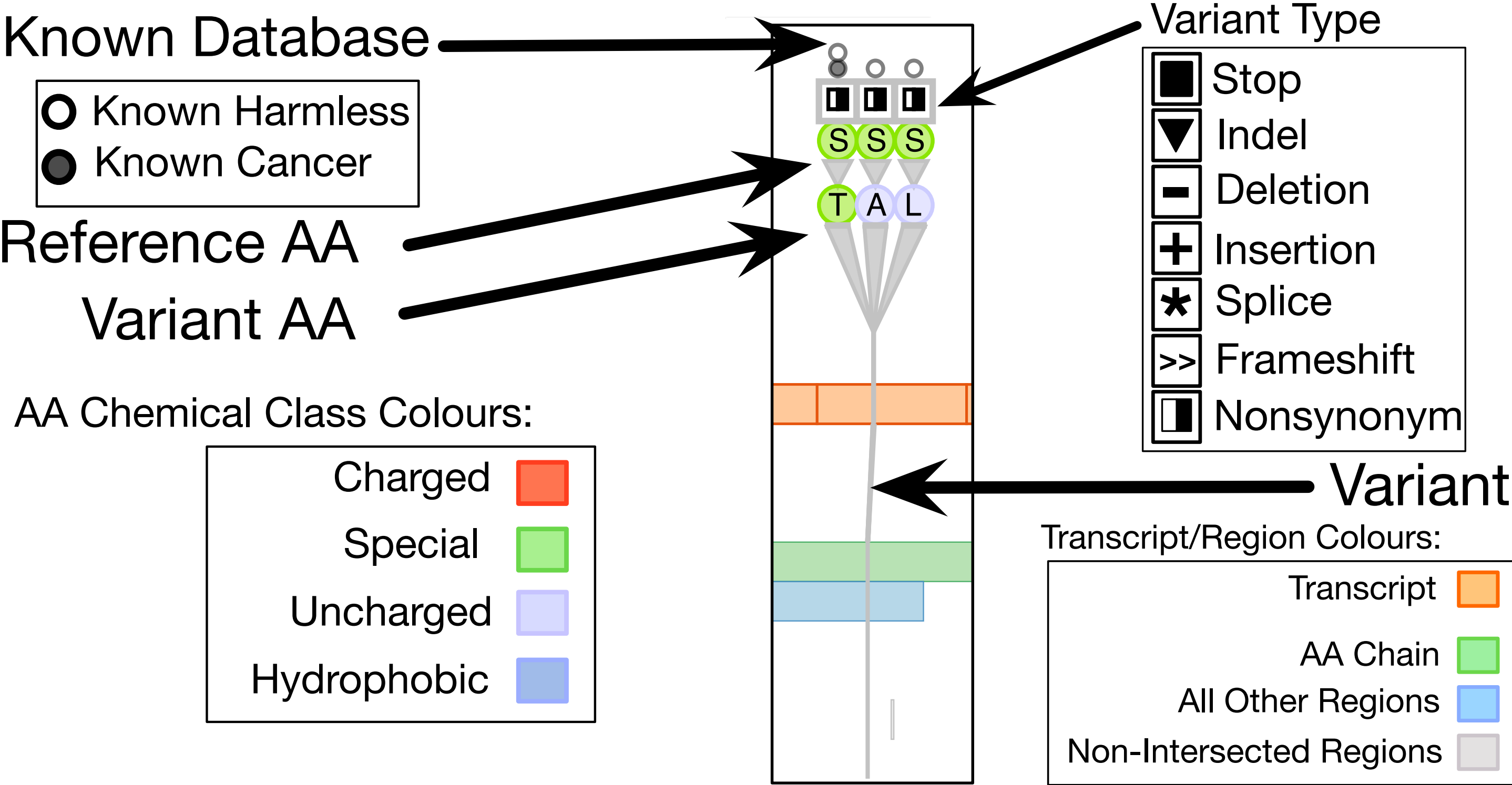
Design information-dense visual encoding



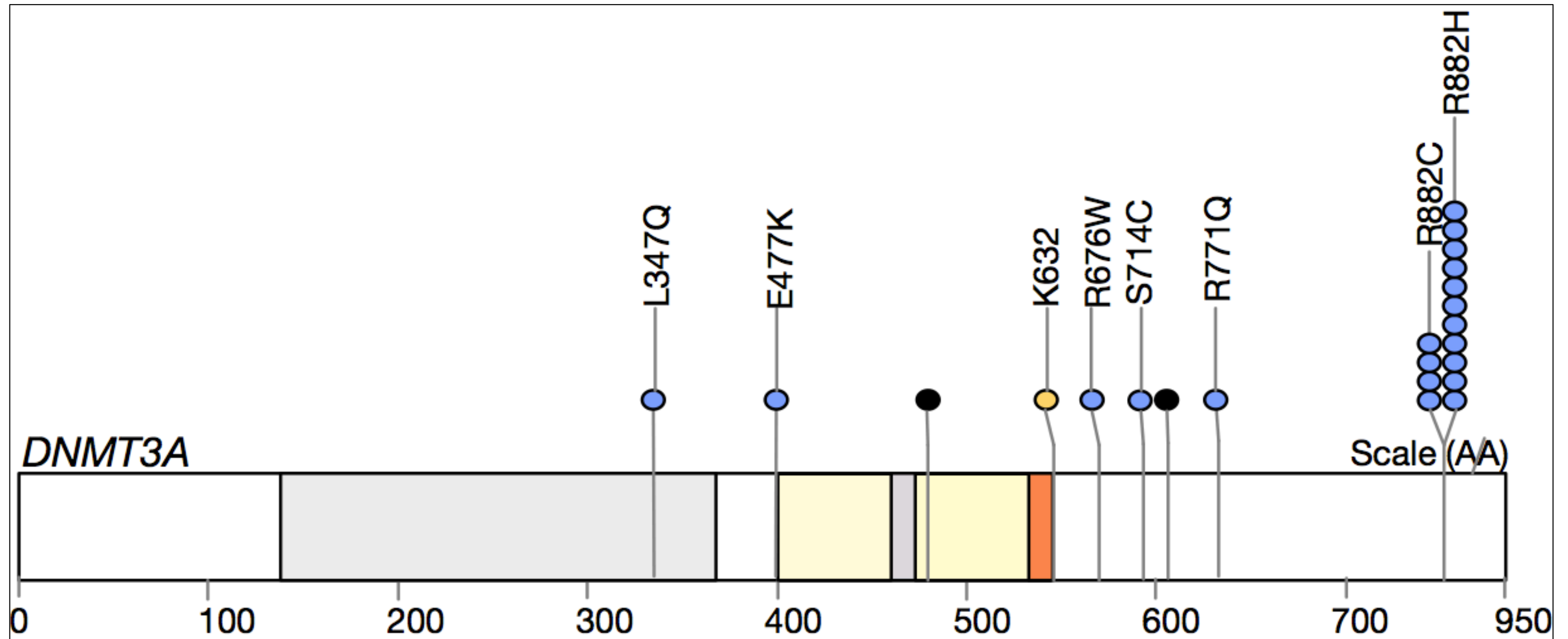
Design information-dense visual encoding



Design information-dense visual encoding



Previous work targeted at variant analysis: MuSiC

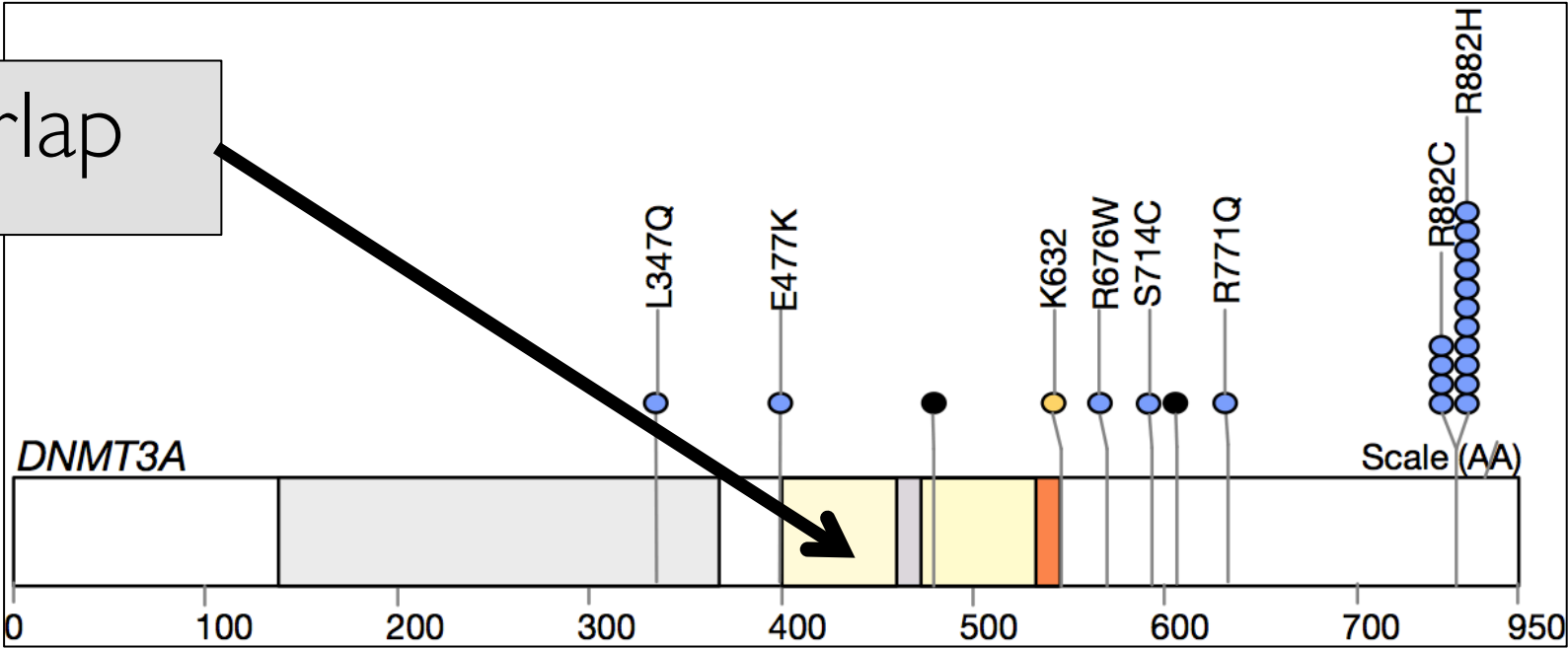


MuSiC

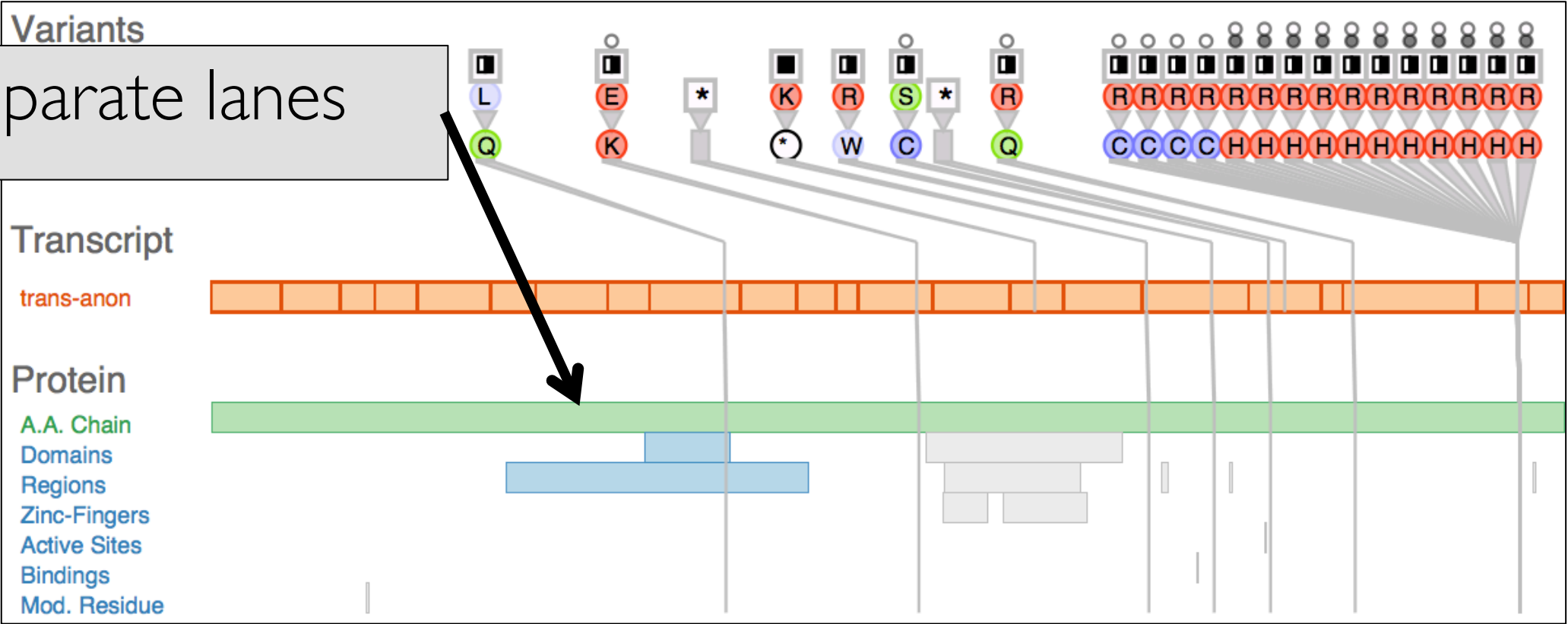
[Dees et al., Genome Research 2012]

Side-by-side comparison: MuSiC vs Variant View

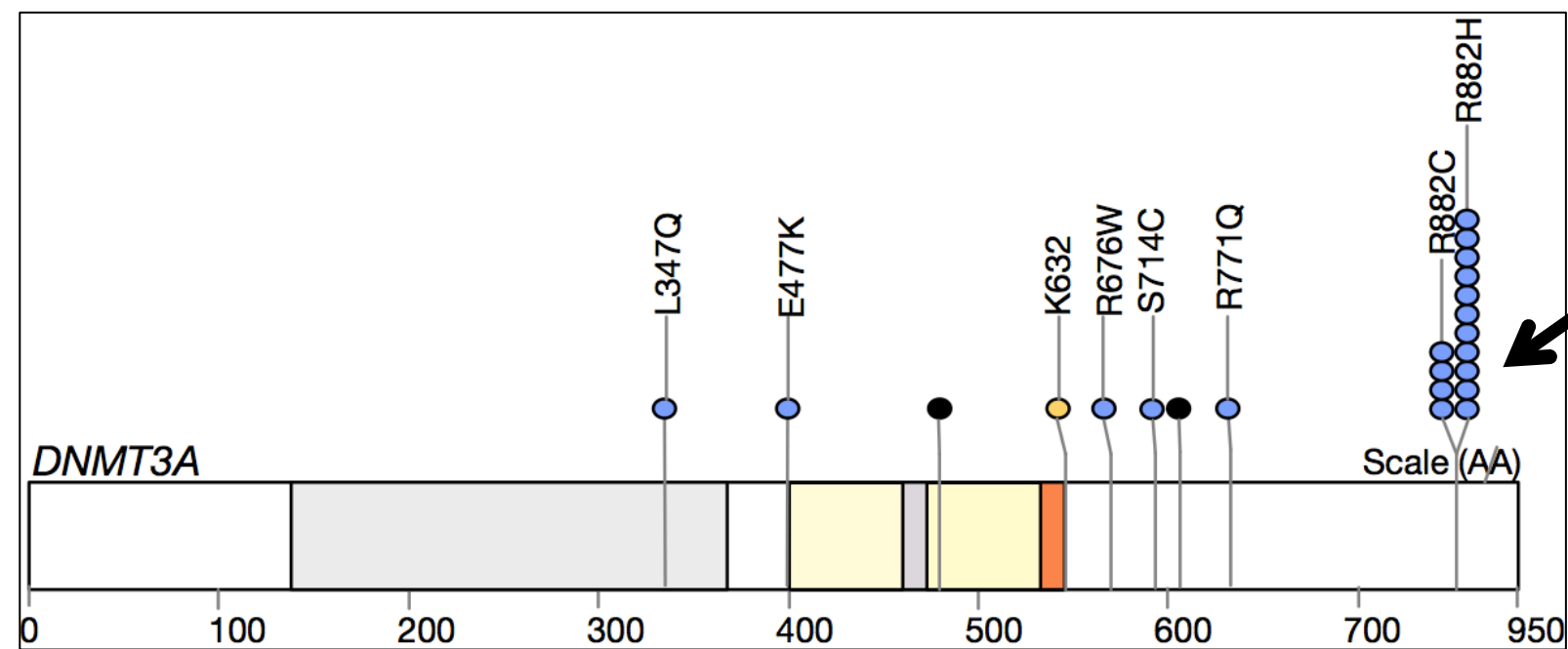
Protein regions can overlap



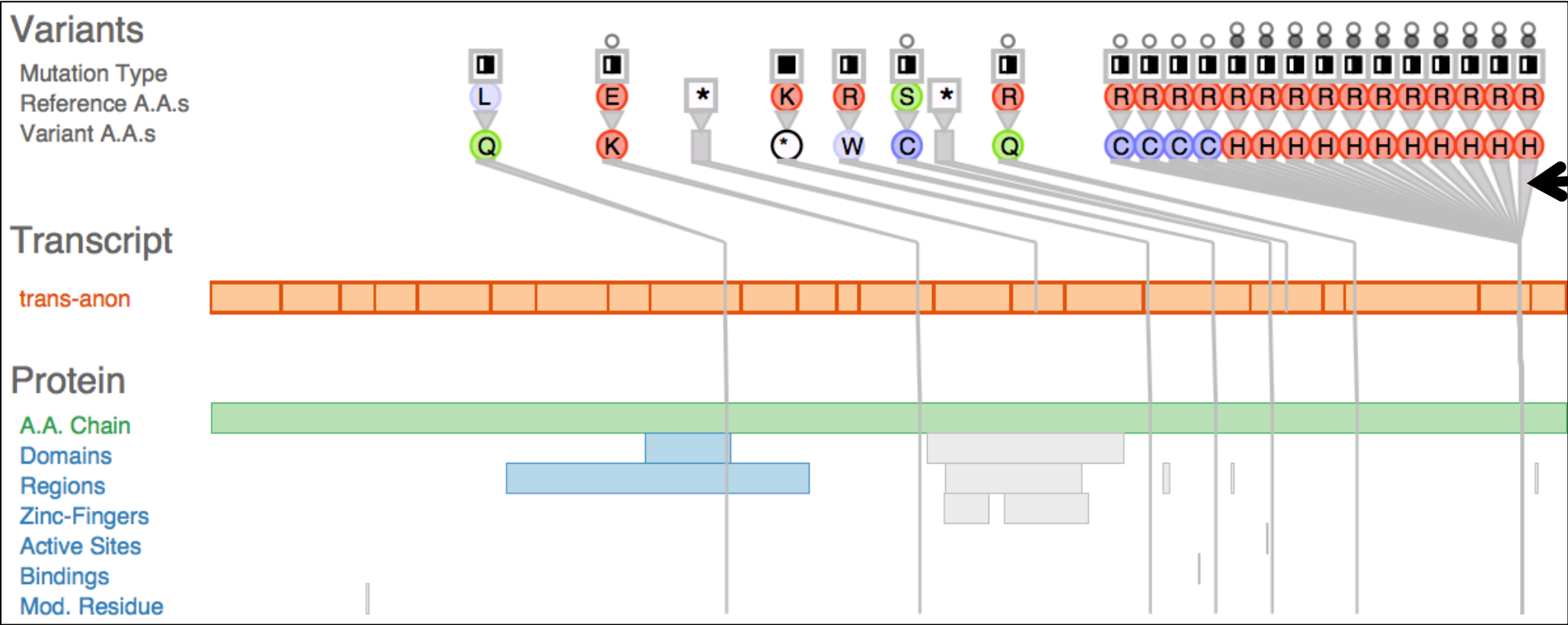
Regions get separate lanes



Side-by-side comparison: MuSiC vs Variant View



Many collocated variants



Large bloom of repeated elements: more salient

Results

Verify known leukemia gene: Highly scored by sorting metric

Variants

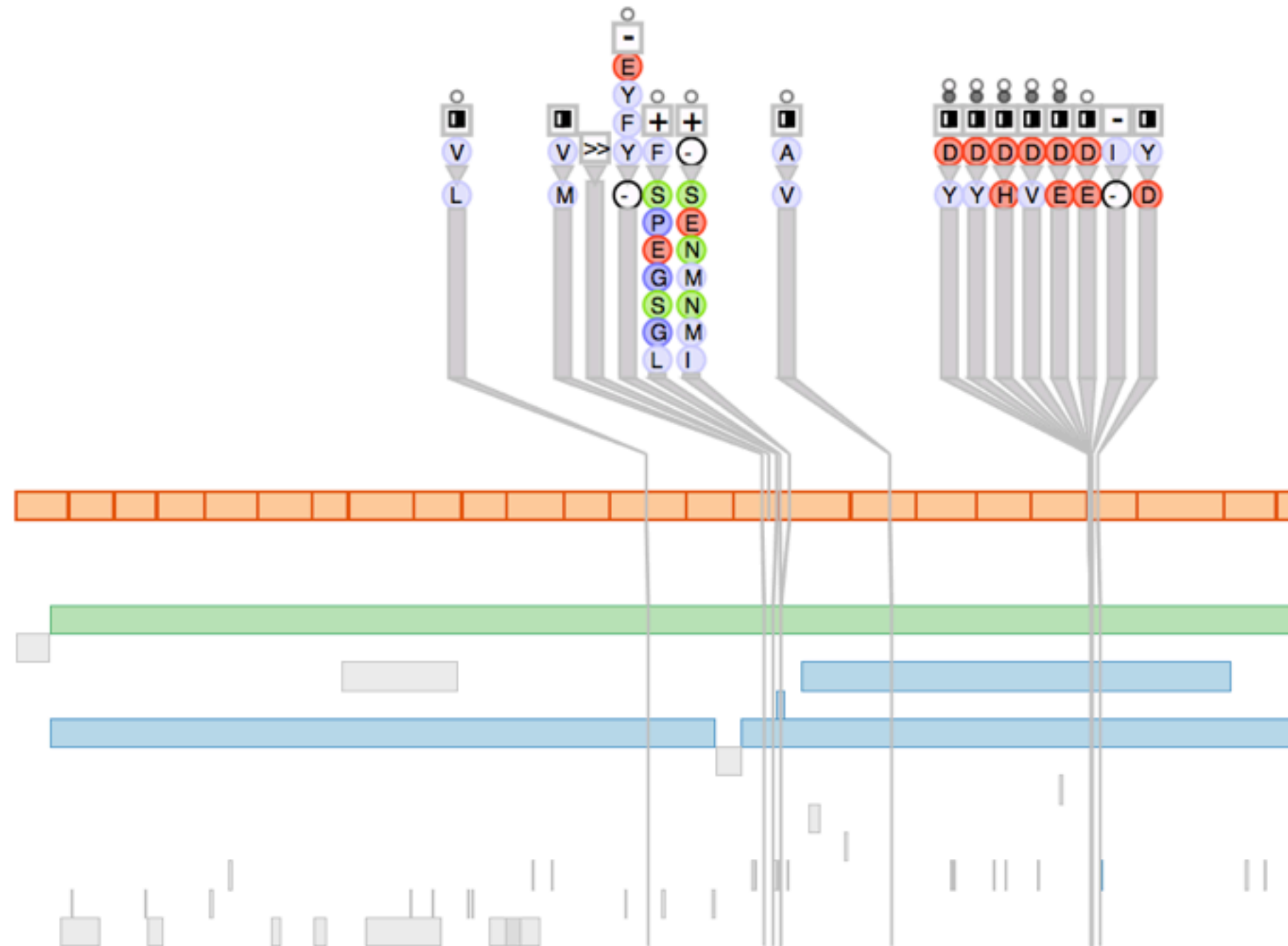
Mutation Type
Reference A.A.s
Variant A.A.s

Transcript

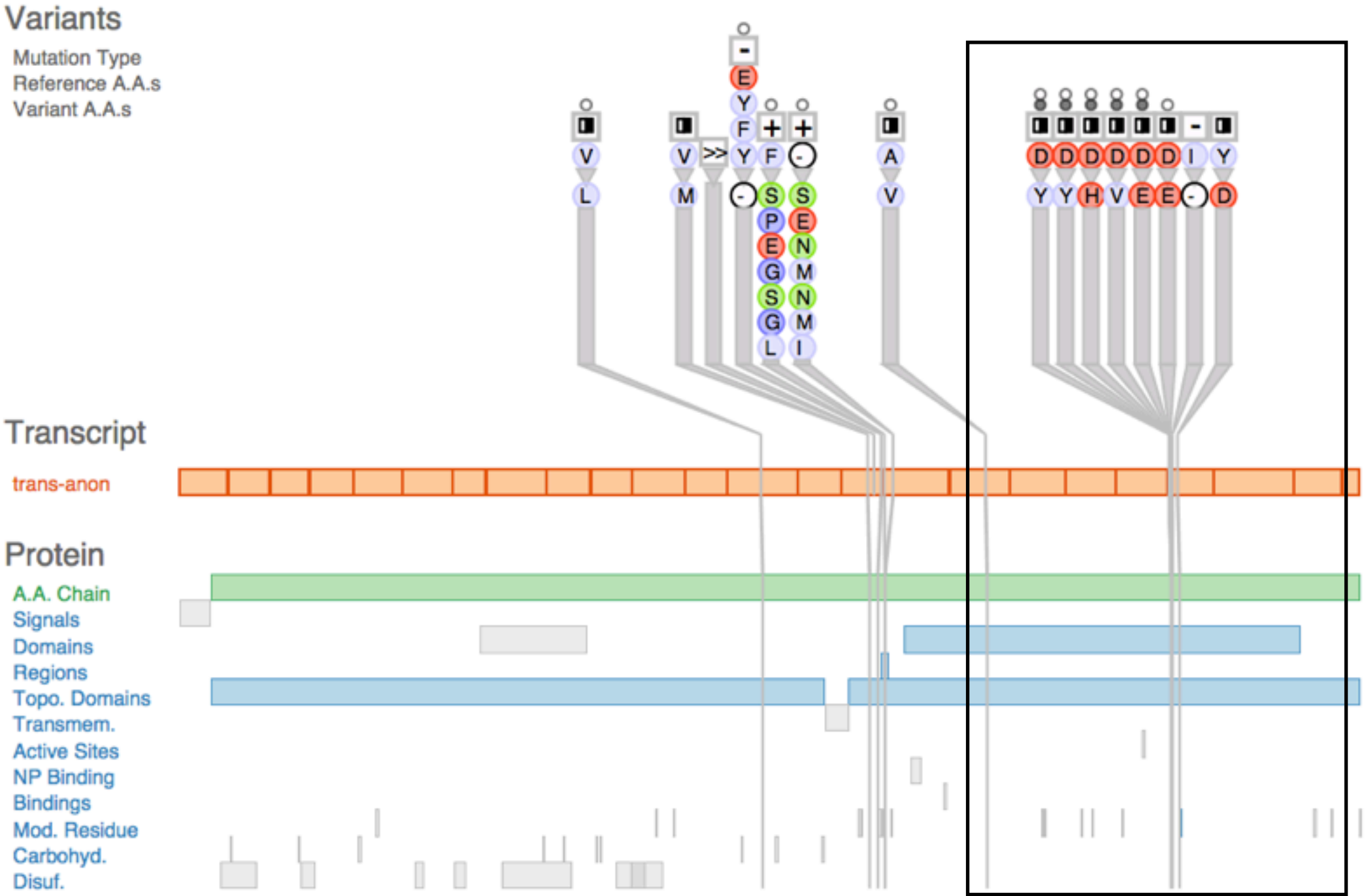
trans-anon

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

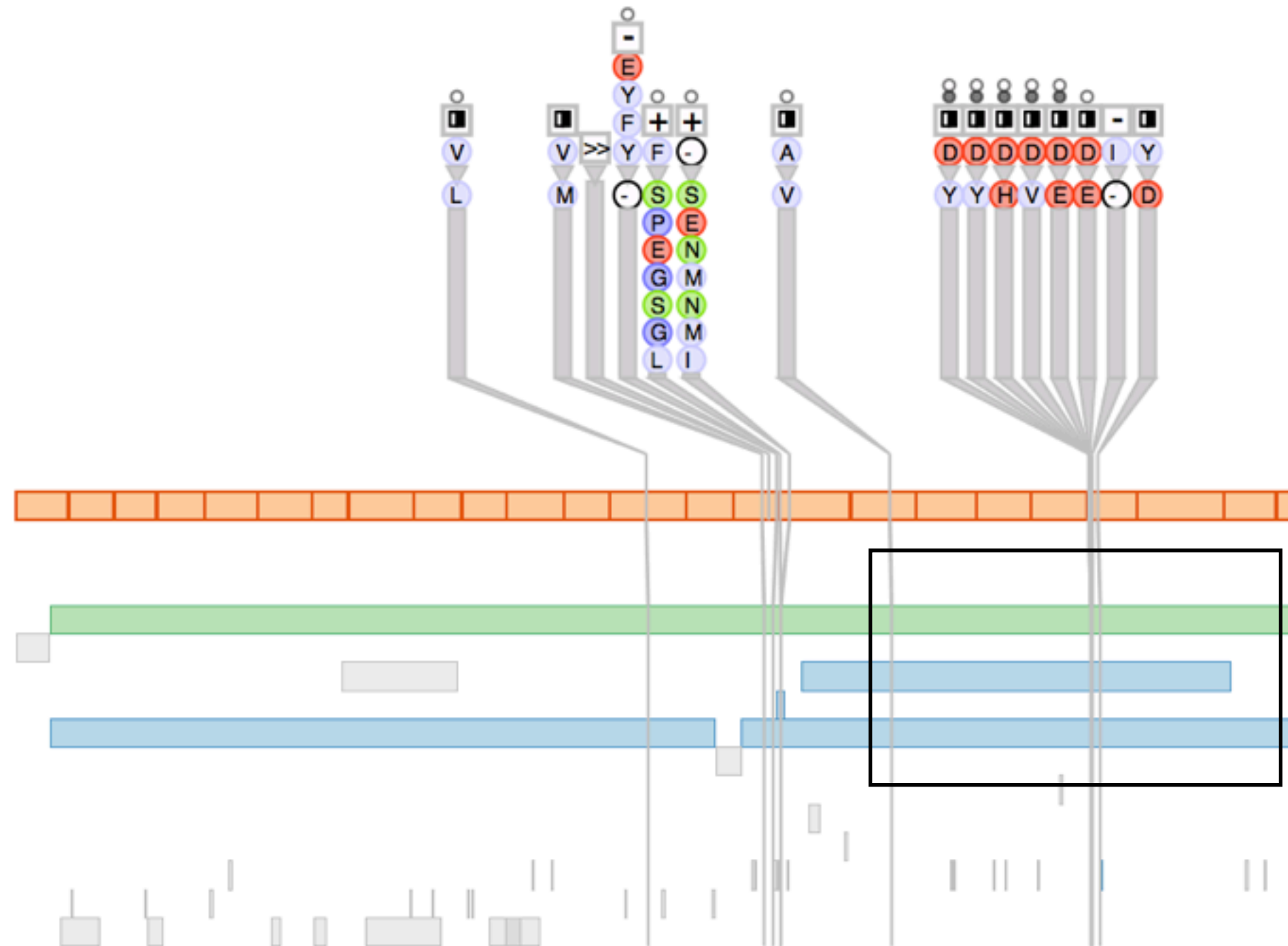
Mutation Type
Reference A.A.s
Variant A.A.s

Transcript

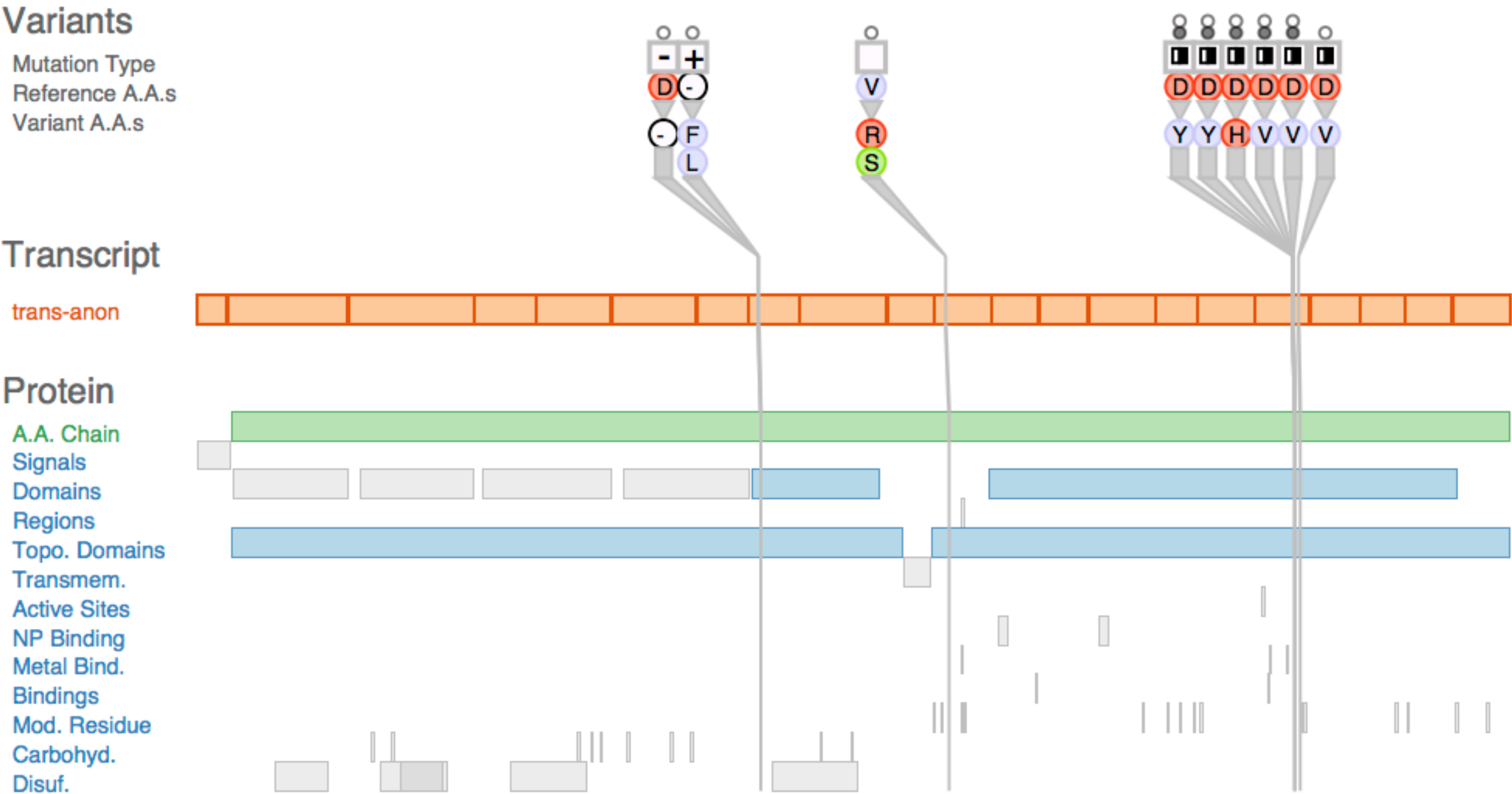
trans-anon

Protein

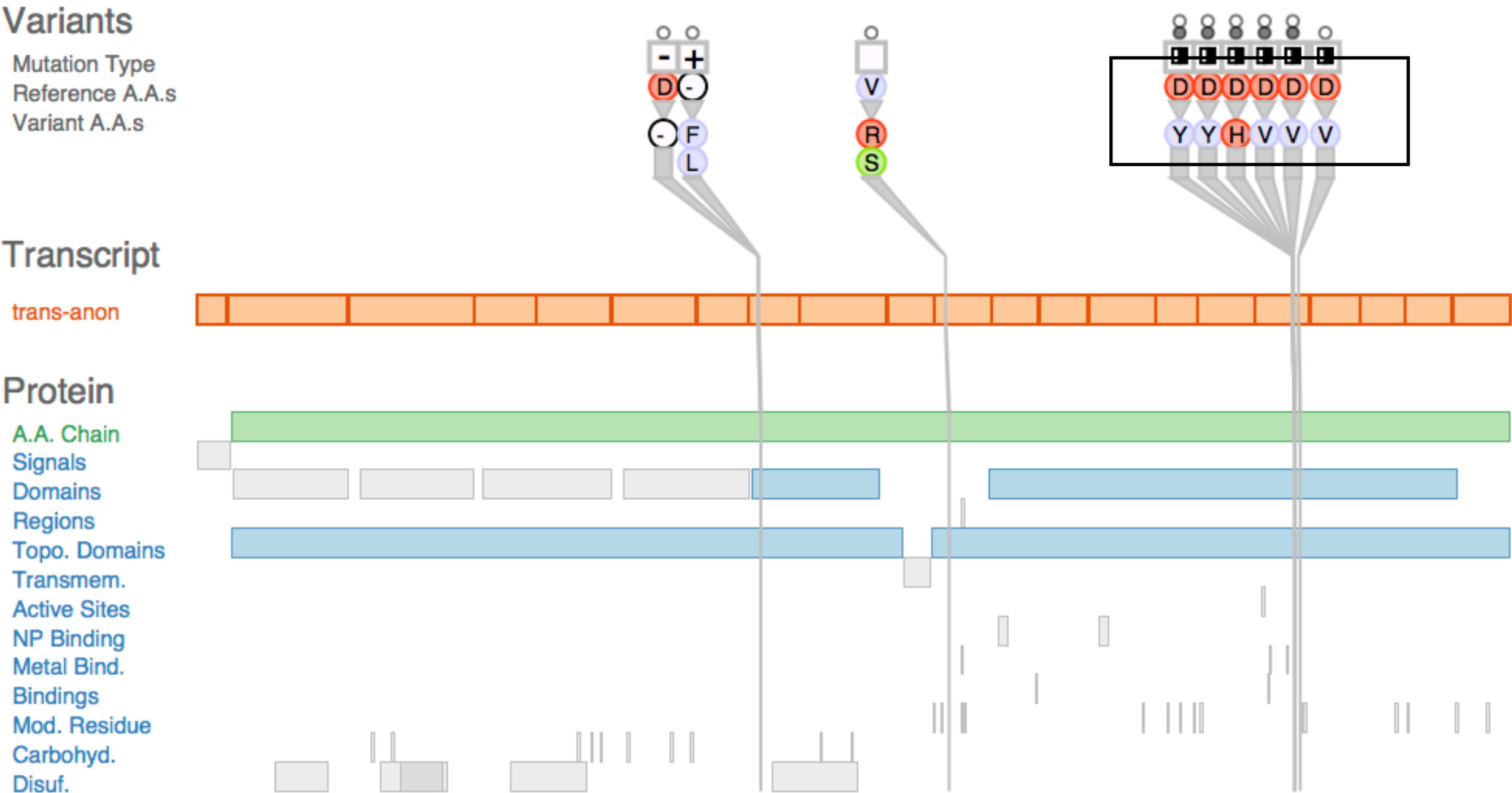
A.A. Chain
Signals
Domains
Regions
Topo. Domains
Transmem.
Active Sites
NP Binding
Bindings
Mod. Residue
Carbohyd.
Disuf.



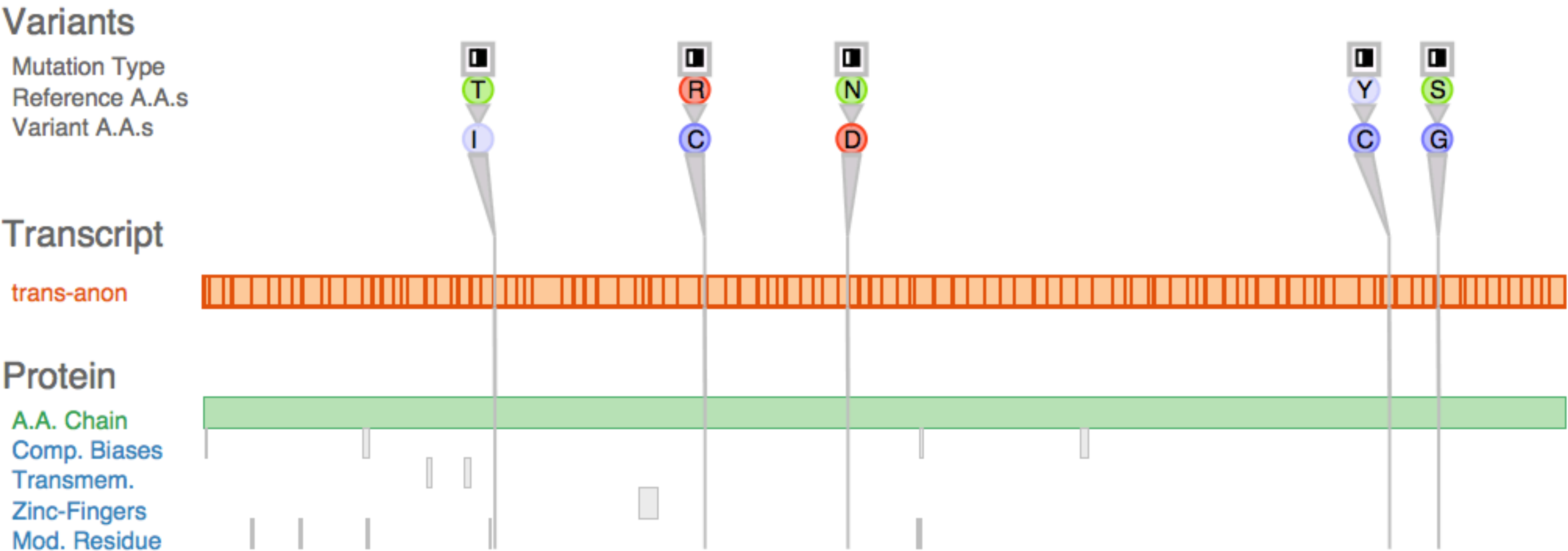
Highly scored by metric: not previously known, good candidate



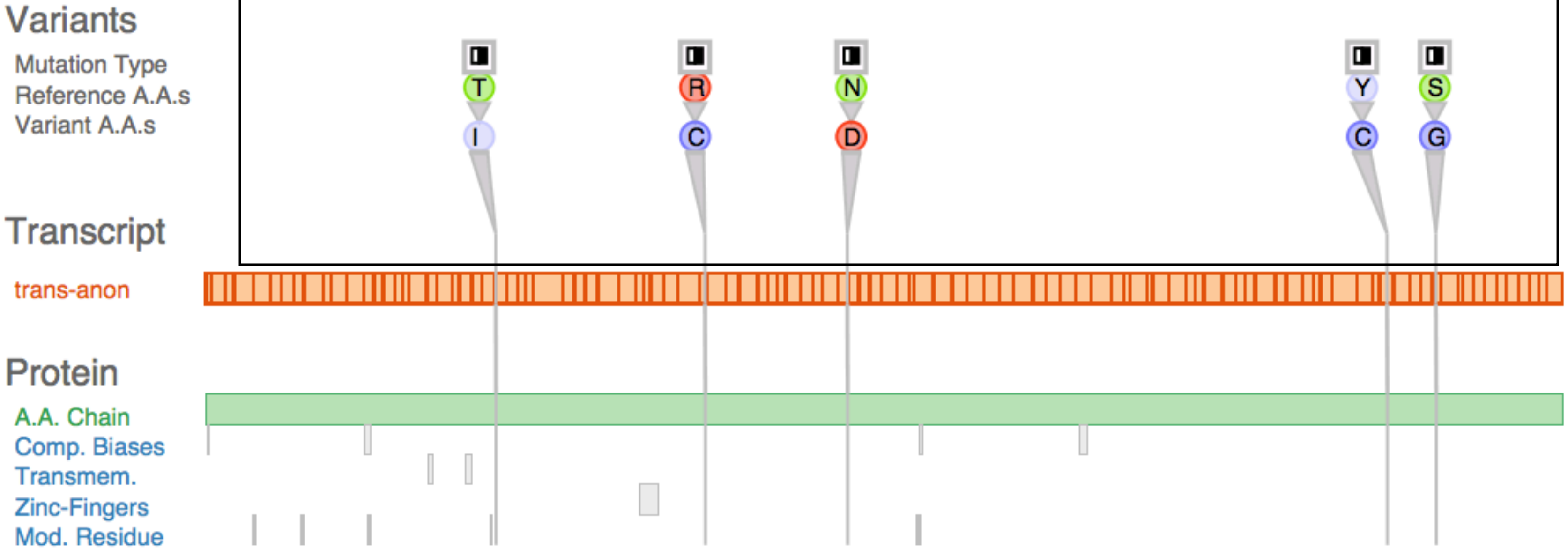
Protein chemical class change evident



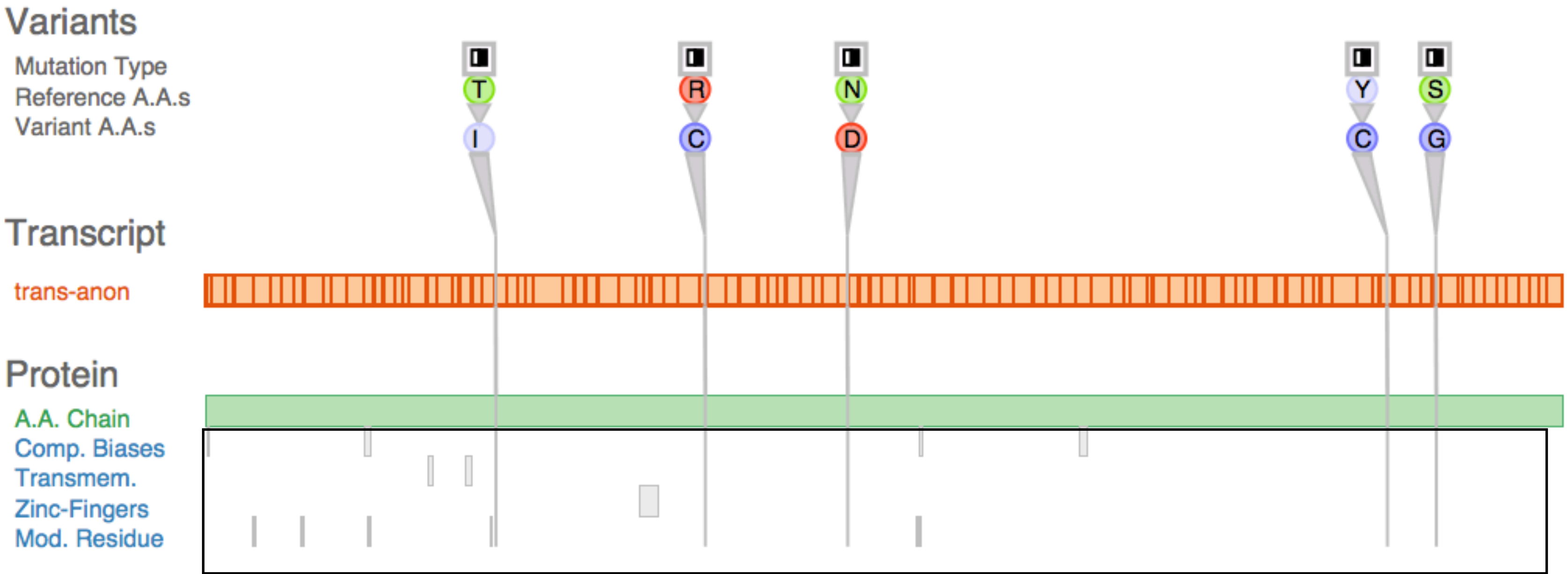
In contrast, low scoring gene



No collocation of variants



Mostly unaffected protein regions



Additional tasks

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

Adapted Variant View with minimal changes

Select Patient: Patient 1 Submit

Patient Genes: gene-anon Submit

Alternative Transcripts: gene-anon (trans-anon) gene-anon (trans-anon)

Variants

Mutation Type

Reference A.A.s

Variant A.A.s

Transcript

trans-anon

Protein

A.A. Chain

Regions

Comp. Biases

Zinc-Fingers

Mod. Residue

Variant Details

Variant ID	Chr. Coord.	Ref Base	Var Base	Effect Level	Effect Type	Gene Name	Trans. Name	Prot. Coord.
pid-anon	31022959	T	C	MODERATE	NON_SYNONY	gene-anon	trans-anon	L815P
pid-anon	31022959	T	C		NON_SYNONY	gene-anon	trans-anon	L815P
pid-anon	31023029	G	T		NON_SYNONY	gene-anon	trans-anon	K838N
pid-anon	31024274	T	C	LOW	SYNONYMOUS	gene-anon	trans-anon	S1253
pid-anon	31024274	T	C		SYNONYMOUS	gene-anon	trans-anon	S1253
pid-anon	31024450	C	T		NON_SYNONY	gene-anon	trans-anon	A1312V
pid-anon	31024704	G	A		NON_SYNONY	gene-anon	trans-anon	G1397S
pid-anon	31025163	A	G	MODIFIER	UTR_3_PRIM	gene-anon	trans-anon	-

Comparison Modes

☐ Show Patient Data Only

☒ Show Patient + Neighborhood

Navigate through patient data with list

Select Patient: Patient 1 Submit

Patient Genes: gene-anon Submit

Alternative Transcripts: gene-anon (trans-anon) gene-anon (trans-anon)

Variants

Mutation Type

Reference A.A.s

Variant A.A.s

Transcript

trans-anon

Protein

A.A. Chain

Regions

Comp. Biases

Zinc-Fingers

Mod. Residue

Variant Details

Variant ID	Chr. Coord.	Ref Base	Var Base	Effect Level	Effect Type	Gene Name	Trans. Name	Prot. Coord.
pid-anon	31022959	T	C	MODERATE	NON_SYNONY	gene-anon	trans-anon	L815P
pid-anon	31022959	T	C		NON_SYNONY	gene-anon	trans-anon	L815P
pid-anon	31023029	G	T		NON_SYNONY	gene-anon	trans-anon	K838N
pid-anon	31024274	T	C	LOW	SYNONYMOUS	gene-anon	trans-anon	S1253
pid-anon	31024274	T	C		SYNONYMOUS	gene-anon	trans-anon	S1253
pid-anon	31024450	C	T		NON_SYNONY	gene-anon	trans-anon	A1312V
pid-anon	31024704	G	A		NON_SYNONY	gene-anon	trans-anon	G1397S
pid-anon	31025163	A	G	MODIFIER	UTR_3_PRIM	gene-anon	trans-anon	-

Comparison Modes

☐ Show Patient Data Only

☒ Show Patient + Neighborhood

Patient data emphasized with arrows

Select Patient: Patient 1 Submit

Patient Genes: gene-anon Submit

Alternative Transcripts: gene-anon (trans-anon) gene-anon (trans-anon)

Variants

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

Transcript

trans-anon

Protein

A.A. Chain
Regions
Comp. Biases
Zinc-Fingers
Mod. Residue

Variant Details

Variant ID	Chr. Coord.	Ref Base	Var Base	Effect Level	Effect Type	Gene Name	Trans. Name	Prot. Coord.
pid-anon	31022959	T	C	MODERATE	NON_SYNONY	gene-anon	trans-anon	L815P
pid-anon	31022959	T	C		NON_SYNONY	gene-anon	trans-anon	L815P
pid-anon	31023029	G	T		NON_SYNONY	gene-anon	trans-anon	K838N
pid-anon	31024274	T	C	LOW	SYNONYMOUS	gene-anon	trans-anon	S1253
pid-anon	31024274	T	C		SYNONYMOUS	gene-anon	trans-anon	S1253
pid-anon	31024450	C	T		NON_SYNONY	gene-anon	trans-anon	A1312V
pid-anon	31024704	G	A		NON_SYNONY	gene-anon	trans-anon	G1397S
pid-anon	31025163	A	G	MODIFIER	UTR_3_PRIM	gene-anon	trans-anon	-

Comparison Modes

- ☐ Show Patient Data Only
- ☒ Show Patient + Neighborhood

Patient has same harmful L to P mutation

Select Patient: Patient 1 Submit

Patient Genes: gene-anon Submit

Alternative Transcripts: gene-anon (trans-anon) gene-anon (trans-anon)

Variants

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

Transcript

trans-anon

Protein

A.A. Chain
Regions
Comp. Biases
Zinc-Fingers
Mod. Residue

Variant Details

Variant ID	Chr. Coord.	Ref Base	Var Base	Effect Level	Effect Type	Gene Name	Trans. Name	Prot. Coord.
pid-anon	31022959	T	C	MODERATE	NON_SYNONY	gene-anon	trans-anon	L815P
pid-anon	31022959	T	C		NON_SYNONY	gene-anon	trans-anon	L815P
pid-anon	31023029	G	T		NON_SYNONY	gene-anon	trans-anon	K838N
pid-anon	31024274	T	C	LOW	SYNONYMOUS	gene-anon	trans-anon	S1253
pid-anon	31024274	T	C		SYNONYMOUS	gene-anon	trans-anon	S1253
pid-anon	31024450	C	T		NON_SYNONY	gene-anon	trans-anon	A1312V
pid-anon	31024704	G	A		NON_SYNONY	gene-anon	trans-anon	G1397S
pid-anon	31025163	A	G	MODIFIER	UTR_3_PRIM	gene-anon	trans-anon	-

Comparison Modes

- ☐ Show Patient Data Only
- ☒ Show Patient + Neighborhood

Nonmatching variants

Select Patient: Patient 1 Submit

Patient Genes: gene-anon Submit

Alternative Transcripts: gene-anon (trans-anon) gene-anon (trans-anon)

Variants

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

Transcript

trans-anon

Protein

A.A. Chain
Regions
Comp. Biases
Zinc-Fingers
Mod. Residue

Variant Details

Variant ID	Chr. Coord.	Ref Base	Var Base	Effect Level	Effect Type	Gene Name	Trans. Name	Prot. Coord.
pid-anon	31022959	T	C	MODERATE	NON_SYNONY	gene-anon	trans-anon	L815P
pid-anon	31022959	T	C		NON_SYNONY	gene-anon	trans-anon	L815P
pid-anon	31023029	G	T		NON_SYNONY	gene-anon	trans-anon	K838N
pid-anon	31024274	T	C	LOW	SYNONYMOUS	gene-anon	trans-anon	S1253
pid-anon	31024274	T	C		SYNONYMOUS	gene-anon	trans-anon	S1253
pid-anon	31024450	C	T		NON_SYNONY	gene-anon	trans-anon	A1312V
pid-anon	31024704	G	A		NON_SYNONY	gene-anon	trans-anon	G1397S
pid-anon	31025163	A	G	MODIFIER	UTR_3_PRIM	gene-anon	trans-anon	-

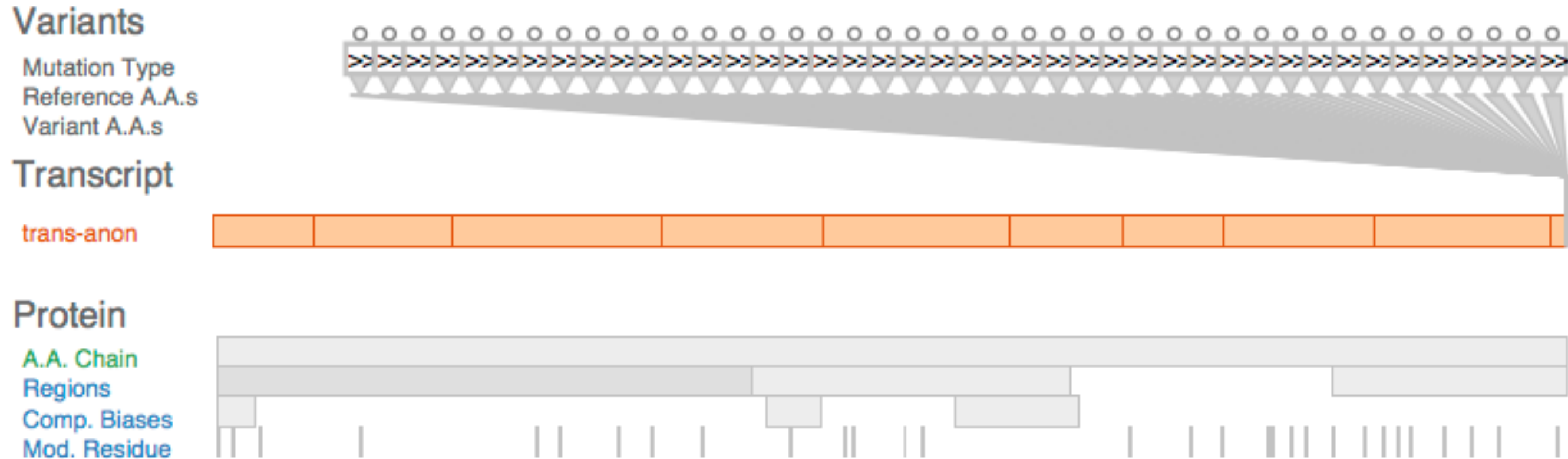
Comparison Modes

- ☐ Show Patient Data Only
- ☒ Show Patient + Neighborhood

Additional tasks

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

Tool revealed errors in the data



The tool exposed artifacts in the data that slid past at least two rounds of quality metric filtering ... this type of problem would not have been caught by our previous, automated methods.

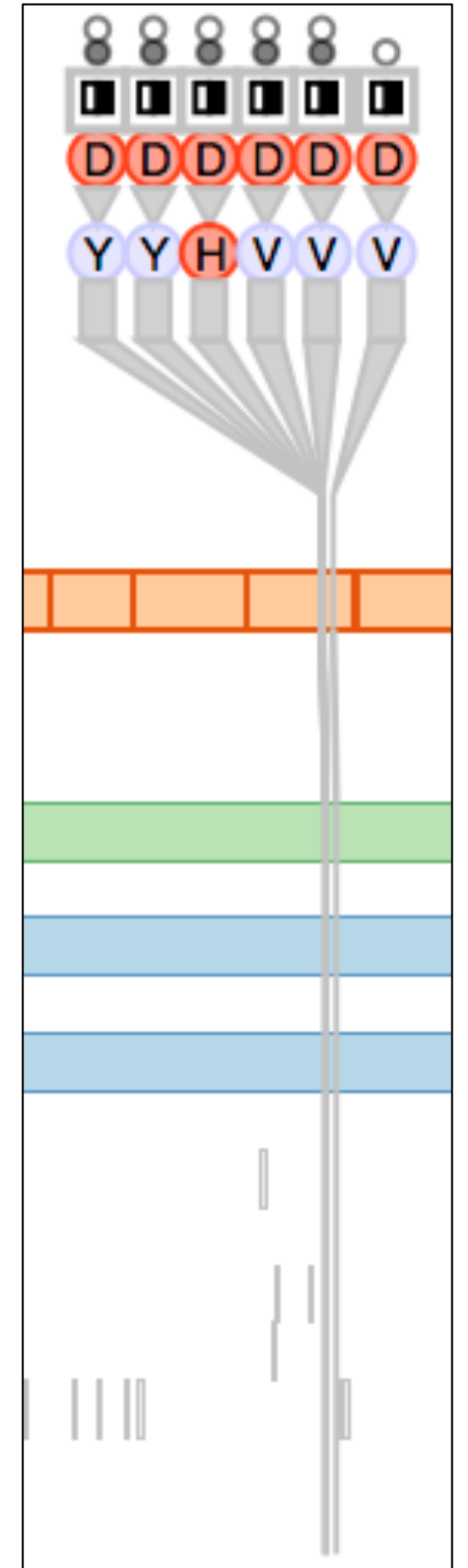
- Analyst 3

Reflections: vis design guidelines

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

Conclusions

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



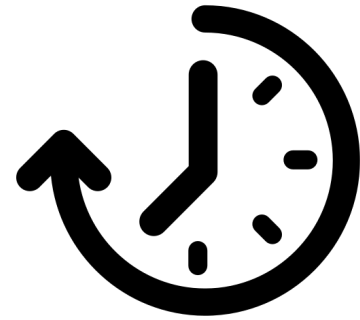
Q&A

Mixed-initiative: Human in the loop with ML

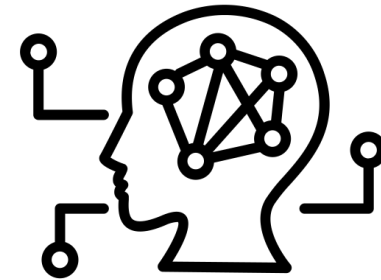
- visualizing imperfect models

Two case studies of visualizing imperfect models

- NLP for temporal data



- ML for graph data

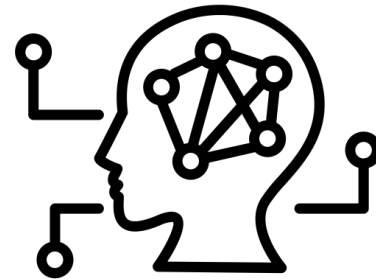


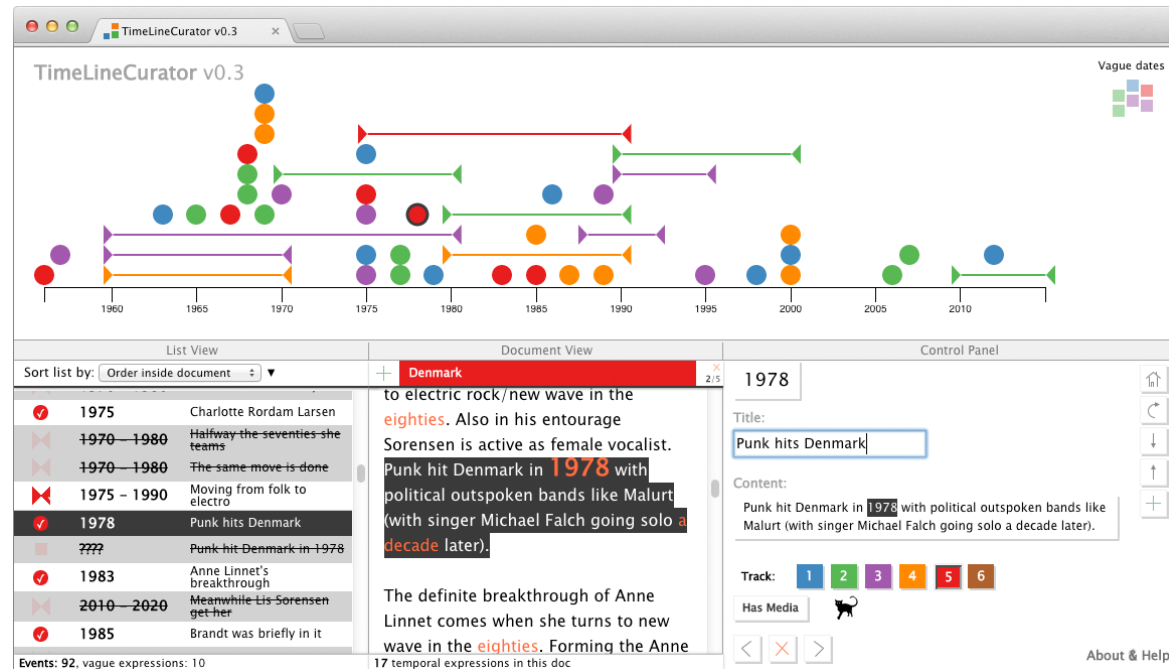
Two case studies of visualizing imperfect models

- NLP for temporal data



- ML for graph data





Johanna Fulda
@jofu_



Matthew Brehmer
@mattbrehmer



TimeLineCurator

Interactive Authoring of Visual Timelines from Unstructured Text

<http://about.timelinecurator.org>

<http://timelinecurator.org>

Tamara Munzner
@tamaramunzner



TimeLineCurator: Interactive Authoring of Visual Timelines from Unstructured Text.

Fulda, Brehmer, Munzner. *IEEE Trans. Visualization and Computer Graphics (Proc IEEE VAST 2015)* 22(1):300-309, 2015.

TimeLineCurator

visual & browser-based

<https://vimeo.com/jofu/tlc>

Manual creation process



1868 The Typewriter
Invented by Christopher Sholes, typewriters quickly became indispensable tools for practically all writing other than personal correspondence. They were widely used by professional writers, in offices, and for business correspondence in private homes

1897 The Stylus
a small pen-shaped instrument that is used to input commands to a computer screen, mobile device or graphics tablet

1907 Multi Touch
With the start of iPhones Multi-touch became a thing

1986 The Mouse
Some additional information here

2012 Speech Recognition

Christopher Sholes
1868
Type writer
"QUERY - Sholes"

Douglas C. Engelbart
'68
Mouse
+ One Button Computer

Stylus
1997
Multi touch
2007

2012
Speech Recogn.

we only have 2 columns for this piece

Why is there such a big gap that wastes whitespace and we only have 2 columns

can we not have so much whitespace here?

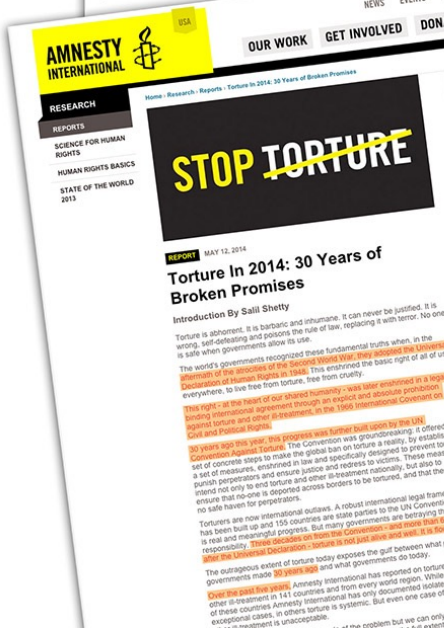
was there nothing else happening in the time between Mouse and Stylus?!

has to be mentioned here, since it is the new shit

3D Touch

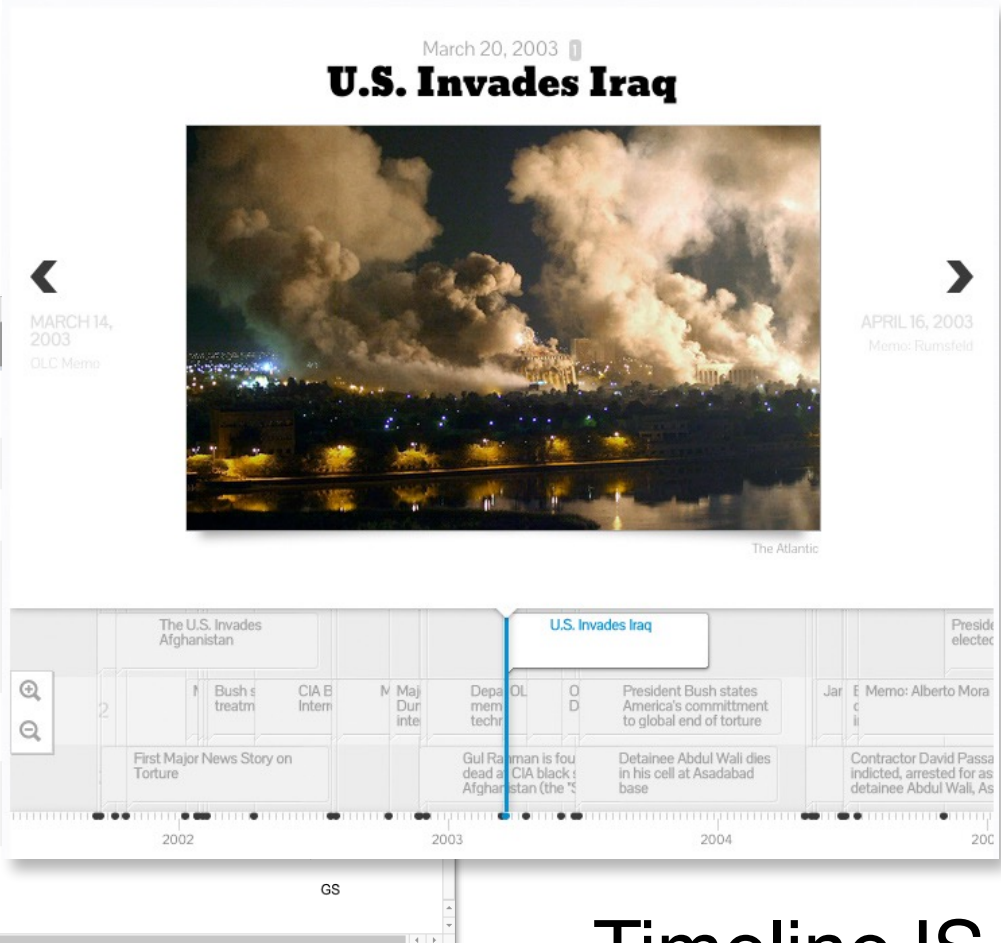
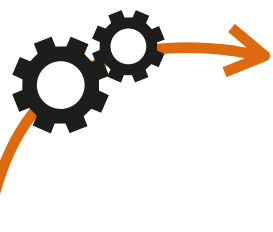
Structured creation process







	A	B	C	D	E	F
	Start Date	End Date	Headline	Text	Media	Media Credit
1				Zuckerberg wrote a program called Facemash on October 28, 2003 while attending Harvard as a sophomore	http://dubindigital.ie/wp-content/uploads/2011/02/facemash-700x299.jpg	Dublin Digital
2	10/28/2003		Facemash	The following semester, Zuckerberg began writing code for a new website in January 2004		
3	1/1/2004	1/31/2004	new website	On February 4, 2004, Zuckerberg launched "thefacebook", originally located at thefacebook.com		
4	2/4/2004		thefacebook.com	Six days after the site launched, three Harvard seniors (Cameron Winklevoss, Tyler Winklevoss, and Divya Narendra) accused Zuckerberg of intentionally misleading them into believing he would help them build a social network called HarvardConnection.com	http://www.capitalberg.com/wp-content/uploads/2015/05/Harvard-wider.jpg http://4.bp.blogspot.com/-KM6s4_O3yys/VV5335_YFsl/AAAAAAAAATB/b15xoCQ3TA/s1600/5.jpeg	Capital Berg Easymese
5	2/10/2004		Harvard Connection	They later filed a lawsuit against Zuckerberg, subsequently settling in 2008[17] for 1.2 million shares (worth \$300 million at Facebook's IPO)		
6	1/1/2008	12/31/2008	Harvard Connection settles	Membership was initially restricted to students of Harvard College; within the first month, more than half the undergraduates at Harvard were registered on the service		
7	2/4/2004	3/4/2004	thefacebook at Harvard only	In March 2004, Facebook expanded to the universities of Columbia, Stanford, and Yale.[20]		
8	3/1/2004	3/31/2004	thefacebook expands to other universities	In mid-2004, entrepreneur Sean Parker (an informal advisor to		



TimelineJS

timeline.knightlab.com/

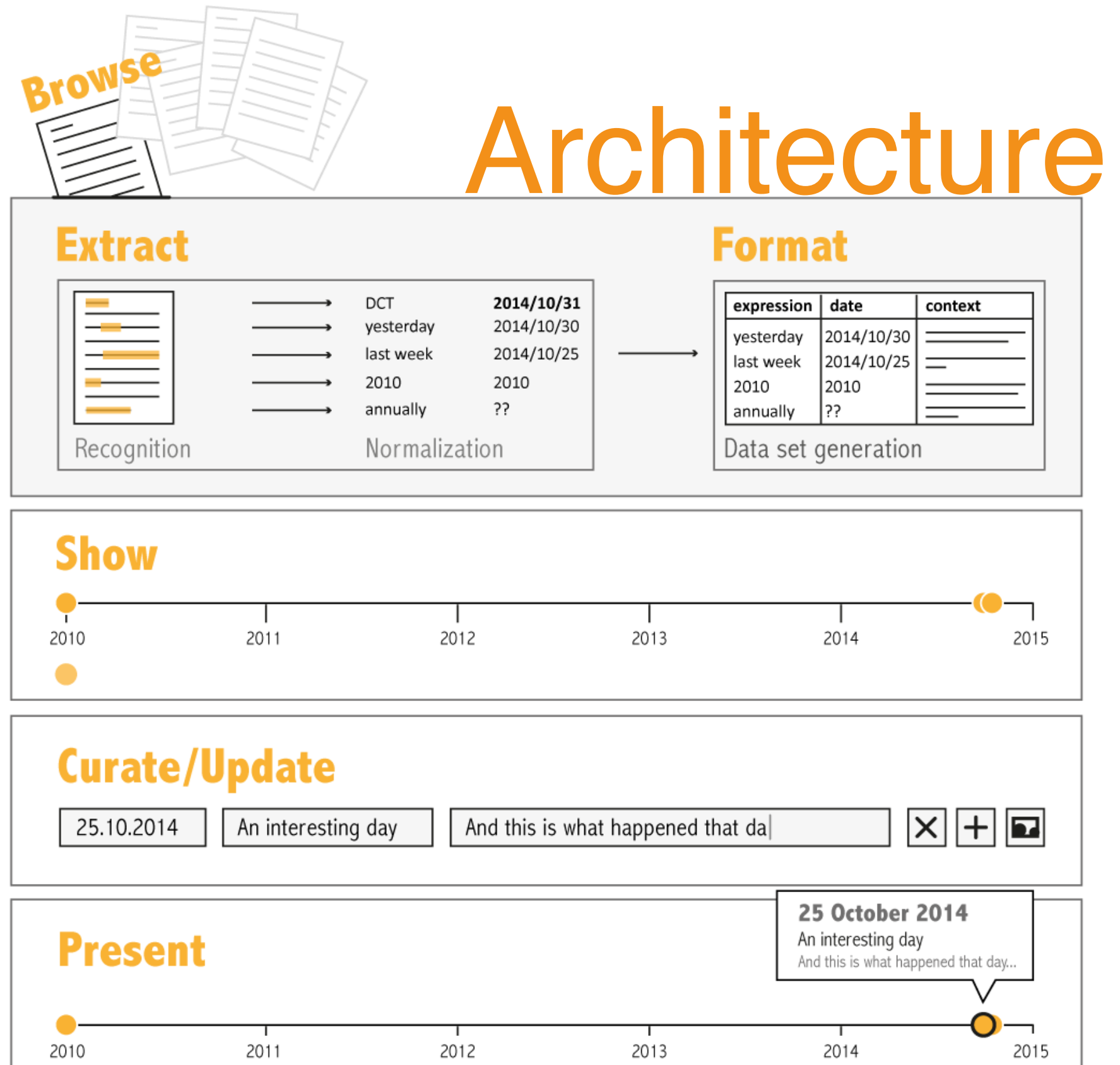
Timeline authoring model

- time required for each task

	Browse	Extract	Format	Show	Update
Manual Drawing	 slow	 slow		 slow	 slow
Structured Creation	 slow	 slow	 slow	 automated	 fast
TimeLine Curator	 fast	 automated	 automated	 fast	 fast

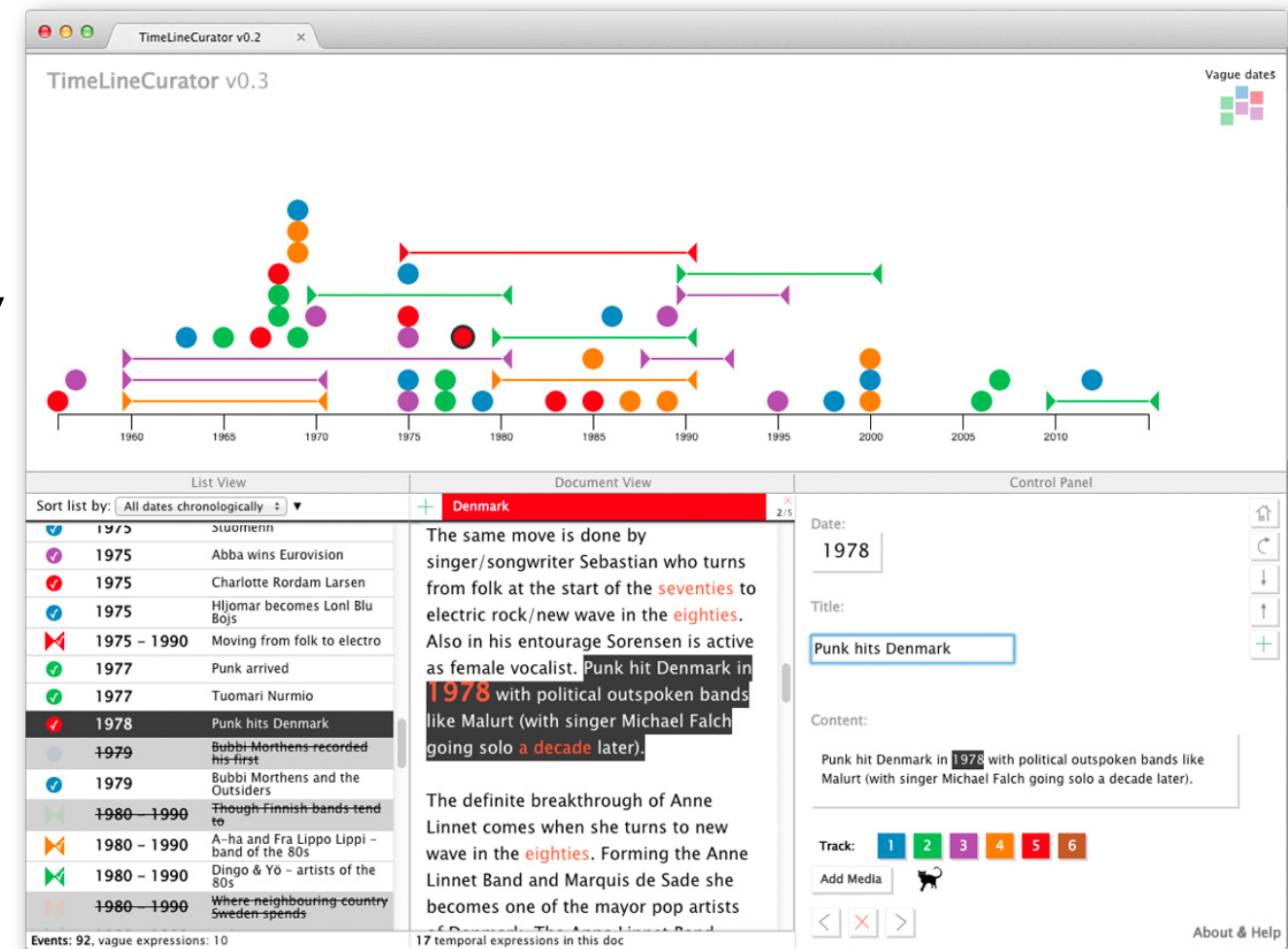
The general case for curation

- build for human in the loop as continuing need
 - automatic processing to accelerate
 - not replace
- **assume computational results good but not perfect**
 - for the indefinite future!
- visual feedback to accelerate



The importance of being brisk

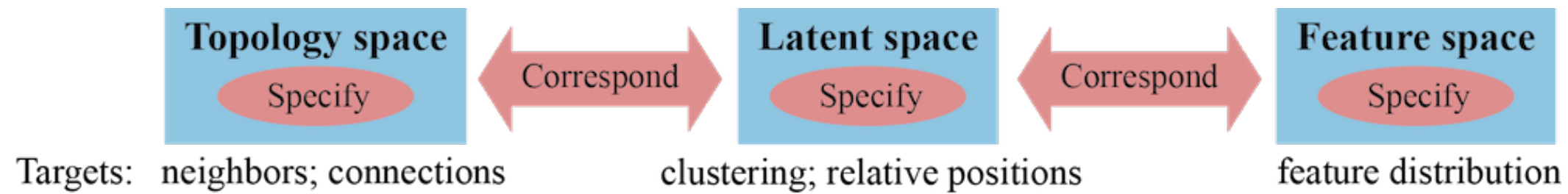
- cool use case: eureka moment
 - success: enable what was impossible before
 - vis tools for new insights & discoveries
- workhorse use case: workflow speedup
 - success: vis tools accelerate your prior workflow
 - sometimes enables the previously infeasible
- TLC use cases
 - started with speedup use case, for presentation
 - make this doc into a timeline now!
 - two other use cases nudge towards exploration
 - comparison between multiple timelines
 - speculative browsing



TimeLineCurator: Speculative Browsing

s p e c u l a t i v e b r o w s i n g

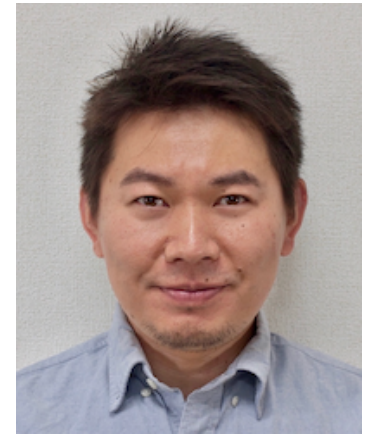
<https://vimeo.com/jofu/tlc>



Zipeng Liu
UBC/Beihang



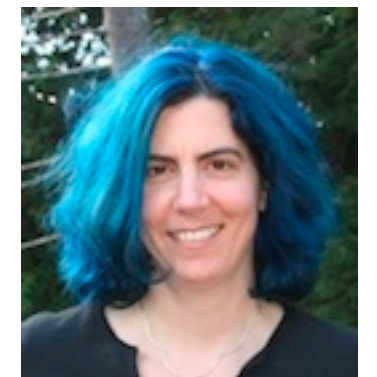
Yang Wang
Uber/Facebook



Jürgen Bernard
UBC/Zurich



Tamara Munzner
UBC



Visualizing Graph Neural Networks with CorGIE:

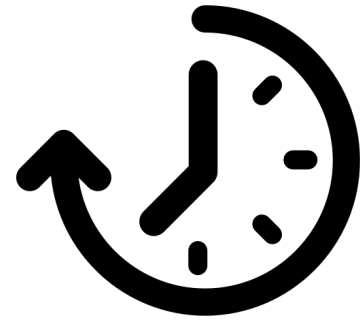
Corresponding a Graph to Its Embedding

<http://www.cs.ubc.ca/group/infovis/pubs/2021/corgie/>

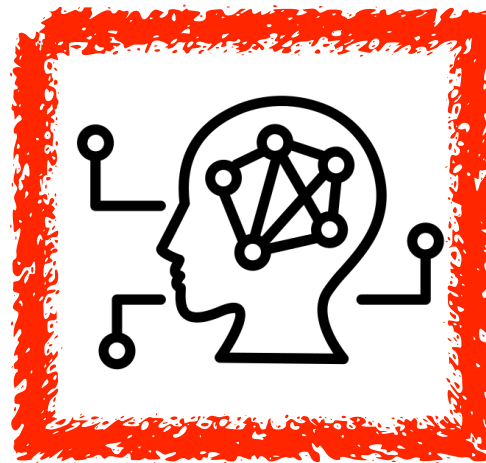
Visualizing Graph Neural Networks with CorGIE: Corresponding a Graph to Its Embedding.
Liu, Wang, Bernard, Munzner. *IEEE TVCG* 28(6):2500-2516 2002.

Two case studies of visualizing imperfect models

- NLP for temporal data

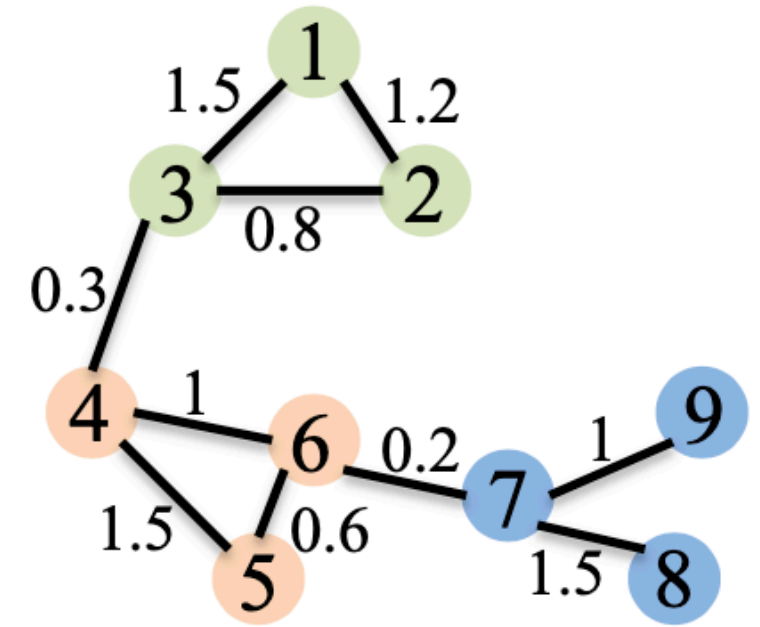


- ML for graph data

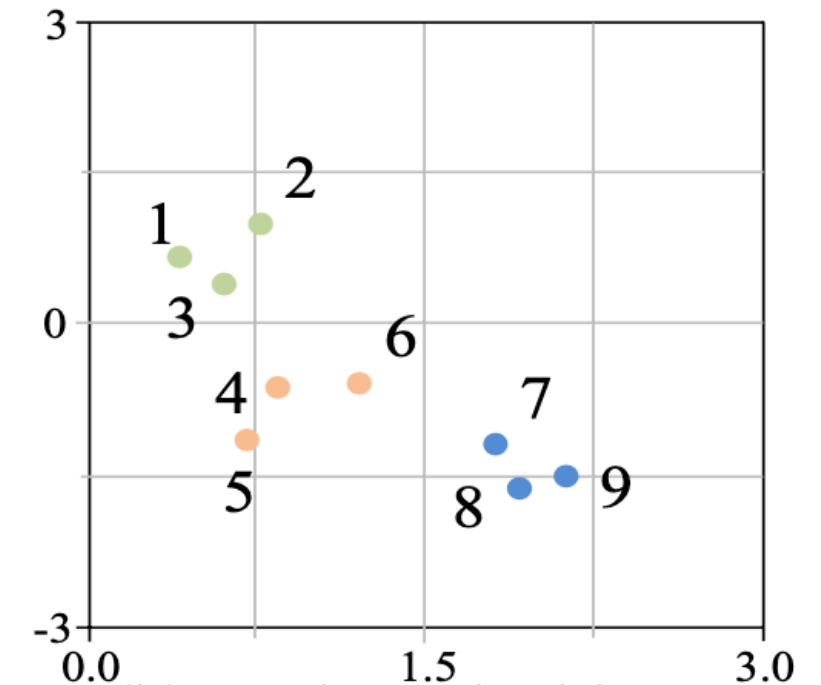
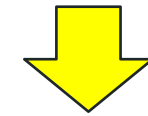


Graph neural network (GNN)

- machine learning (ML) models for graphs
 - like CNN for images
 - like Transformer for text
- many real-world graph-related applications
 - node classification
 - examples: fraud detection, disease classification
 - link prediction
 - examples: product recommendation, protein interactions



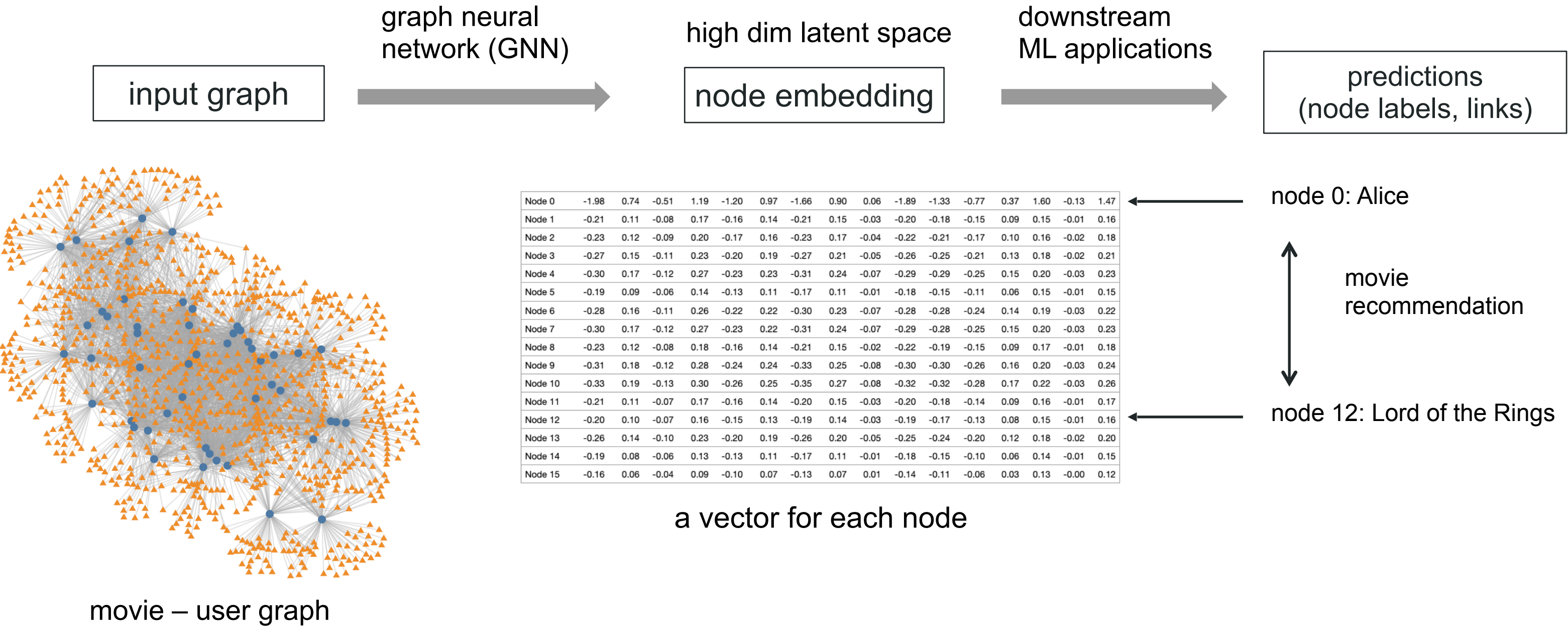
(a) Input Graph G_1



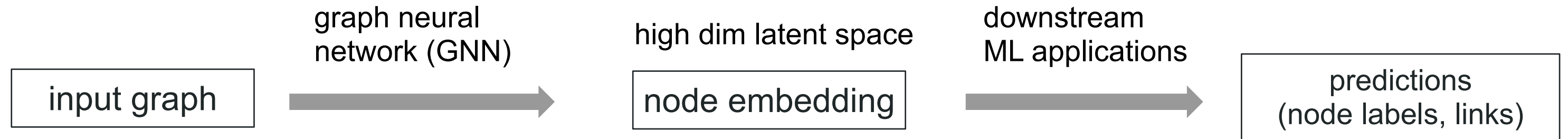
(b) Node Embedding

[Cai et al. TKDE'18]

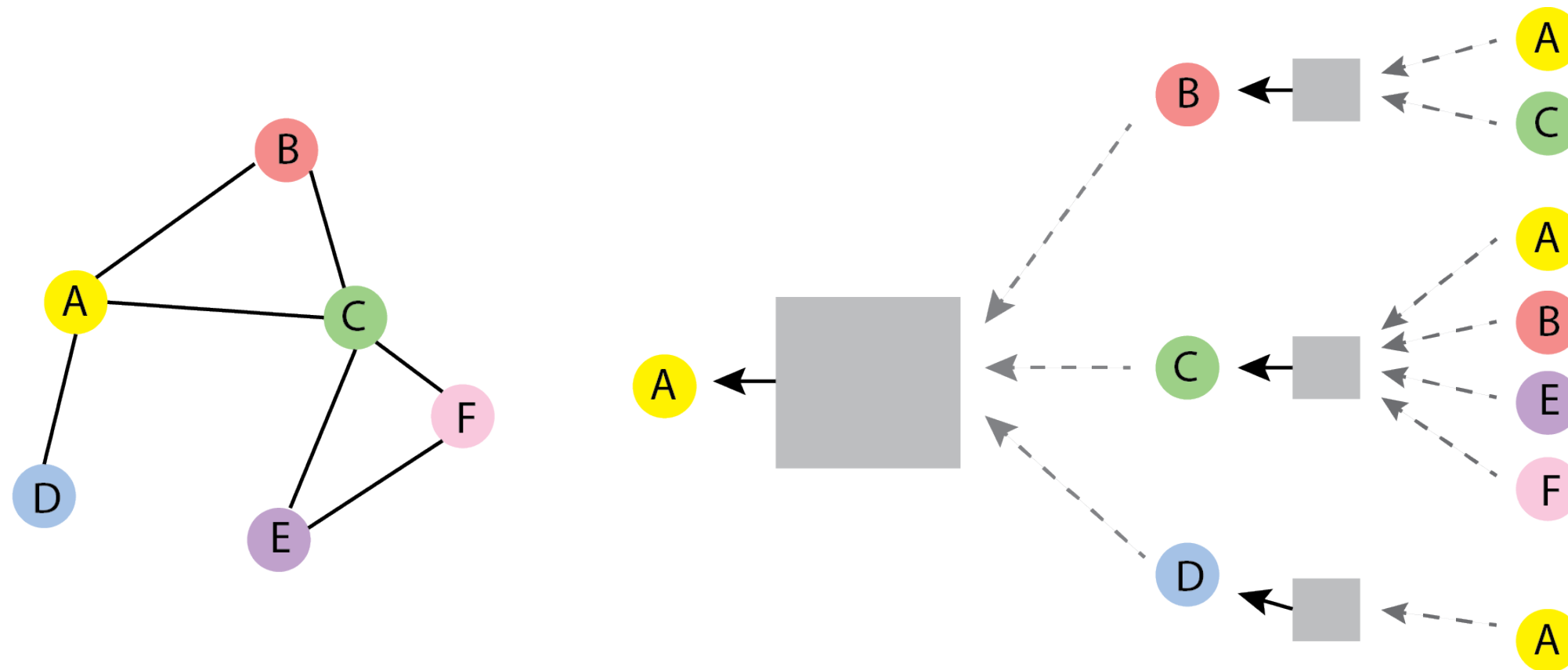
Graph neural network (GNN)



Graph neural network (GNN)



node features are aggregated / passed through **topological neighborhood**



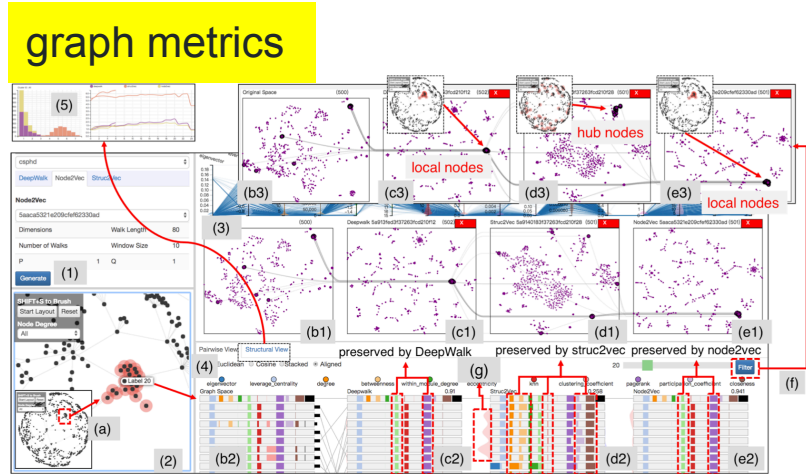
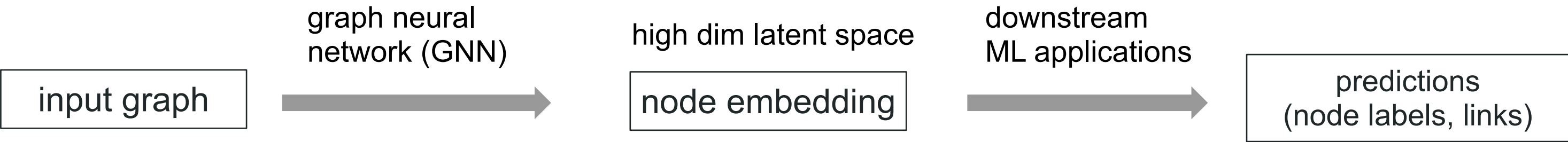
Evaluate GNN

Two big-picture questions

- “Are we there yet?”: should we train / tune more?
- “Are we lost?”: does it behave as we expect?



Evaluate GNN: Previous approaches



Li et al. EmbeddingVis. VAST'18.

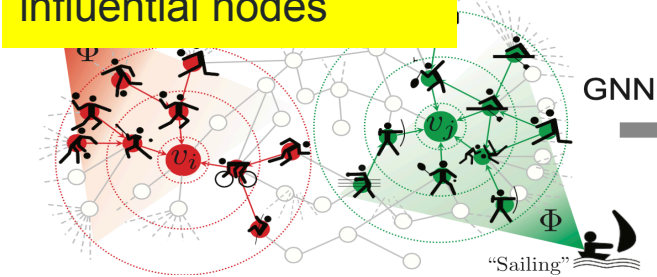
dimensional reduction (simple inspection)



quant metrics

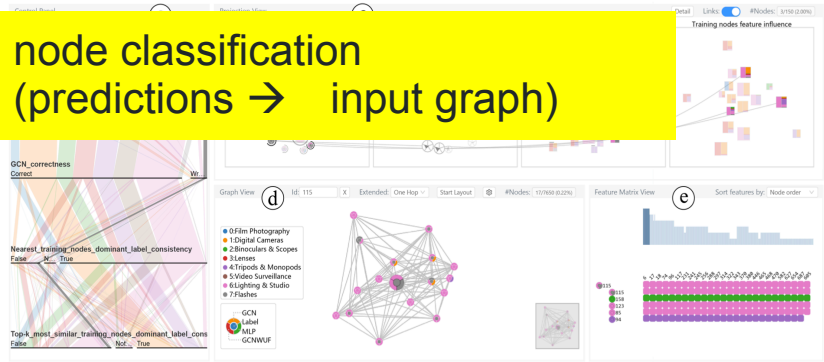
e.g. accuracy, hit rate

influential nodes



Ying et al. GNNExplainer.

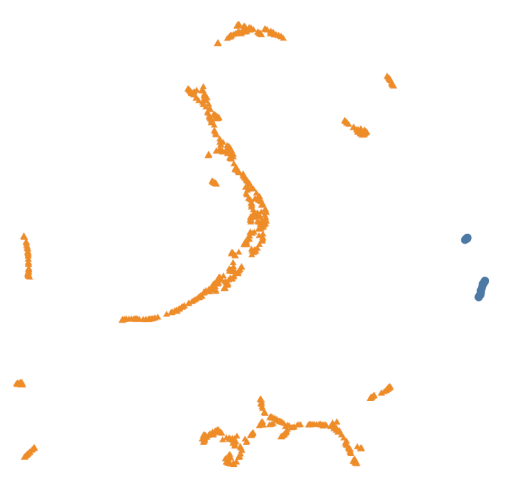
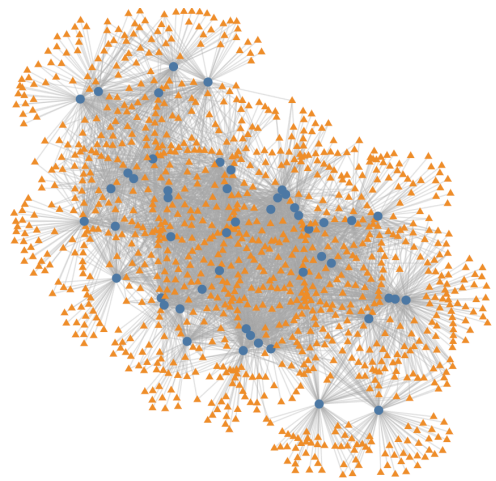
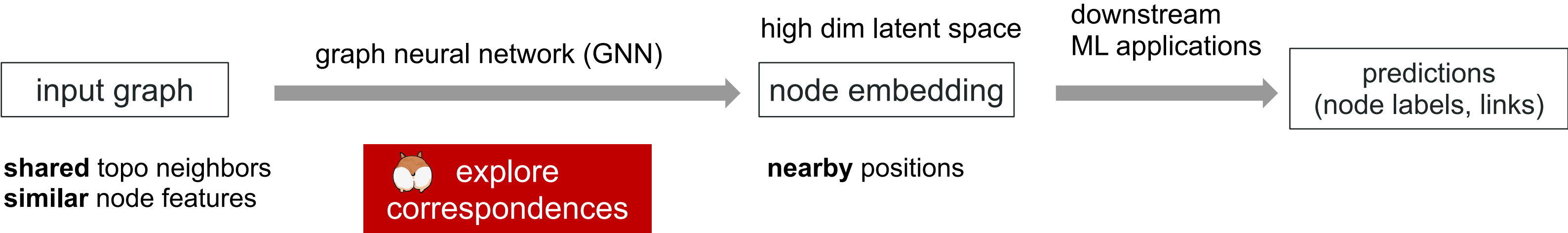
node classification (predictions → input graph)



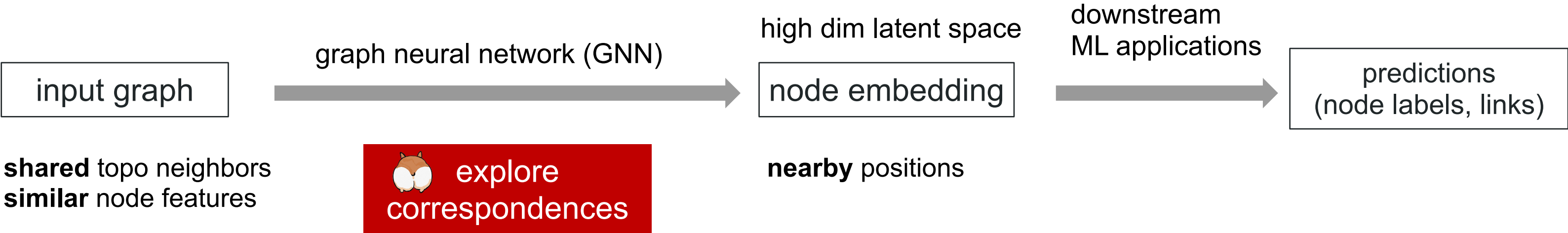
Jin et al. GNNVis. Arxiv'20.

input graph & node embedding under-used!

Evaluate GNN: 🦊 CorGLE idea

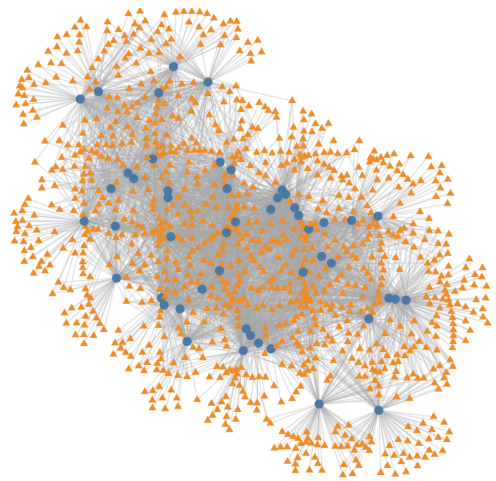


Evaluate GNN: 🦊 CorGLE idea

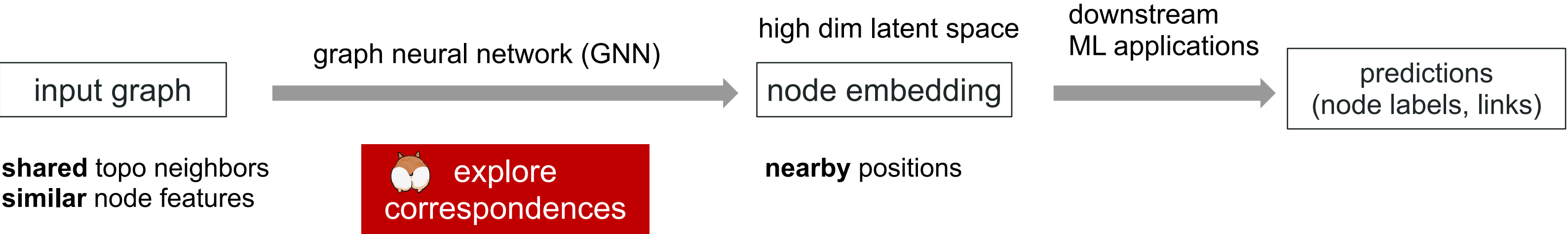


Examples of correspondences:

Check [similar topology? Similar node features?] ← — — — — Pick [a cluster]

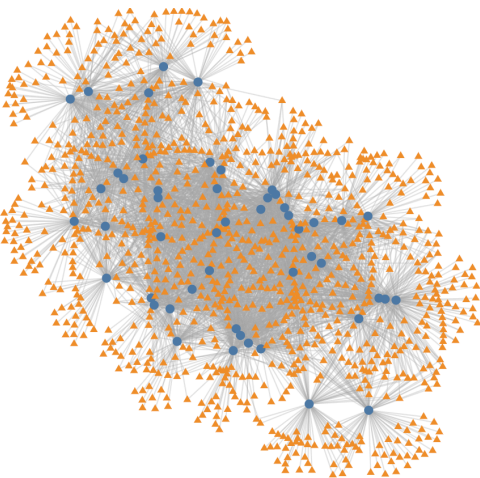


Evaluate GNN: 🦊 CorGLE idea

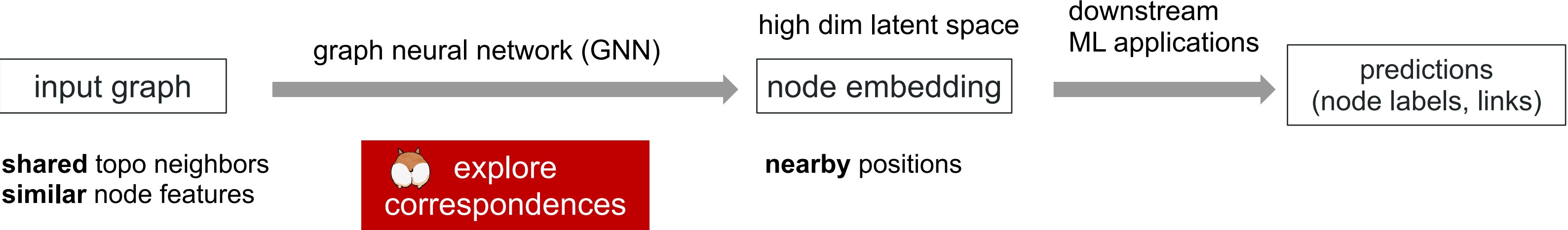


Examples of correspondences:

- Check [similar topology? Similar node features?] ← — — — — Pick [a cluster]
- Check [different topology? Different node features?] ← — — — — Pick [two far-away clusters]

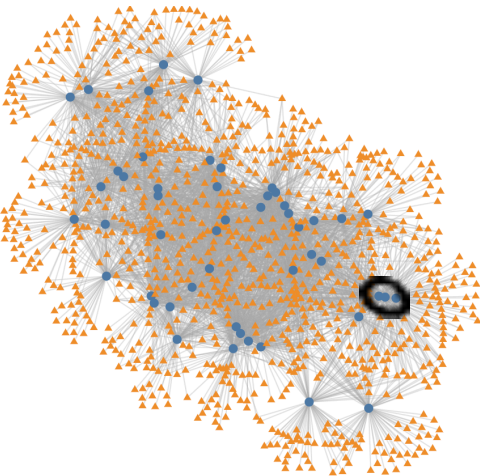


Evaluate GNN: 🦊 CorGLE idea



Examples of correspondences:

- Check [similar topology? Similar node features?] ← - - - - Pick [a cluster]
- Check [different topology? Different node features?] ← - - - - Pick [two far-away clusters]
- Pick [two nodes sharing many topo neighbors] - - - - → Check [how close the nodes are compared to others?]



Data and task abstraction

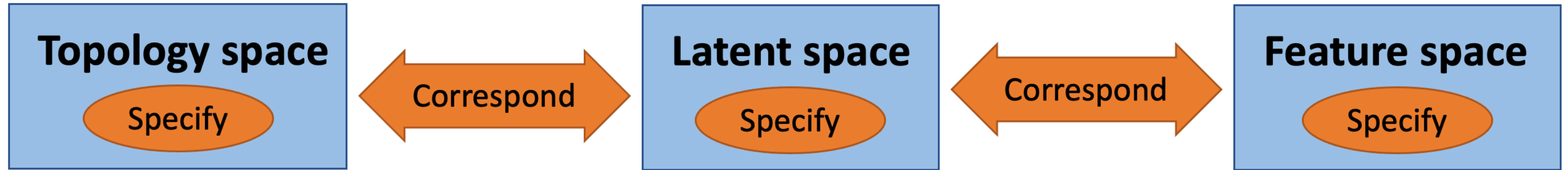
Topology space

Latent space

Feature space

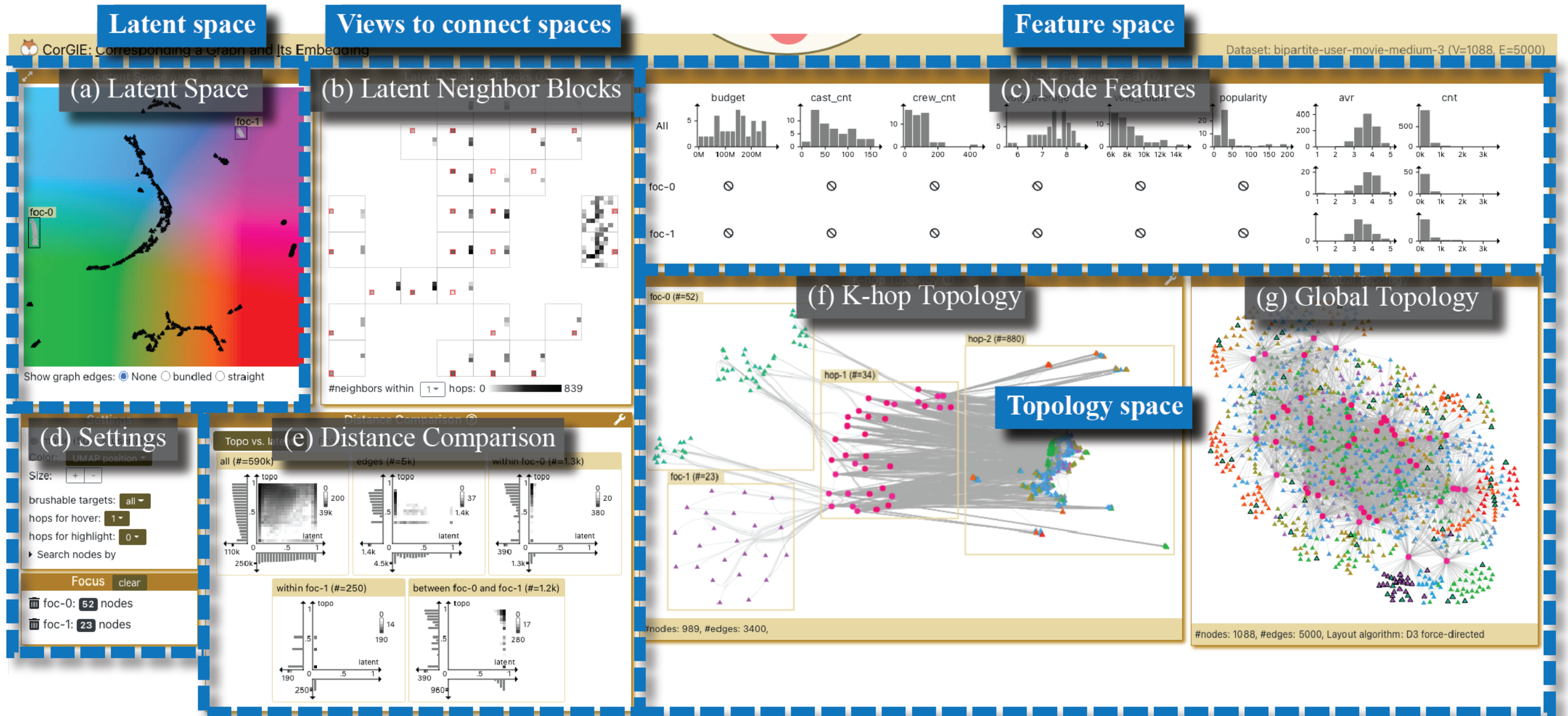
- data spaces

Data and task abstraction



- data spaces
- tasks
 - specify
 - correspond
- task iteration
 - levels in grouping structure of nodes

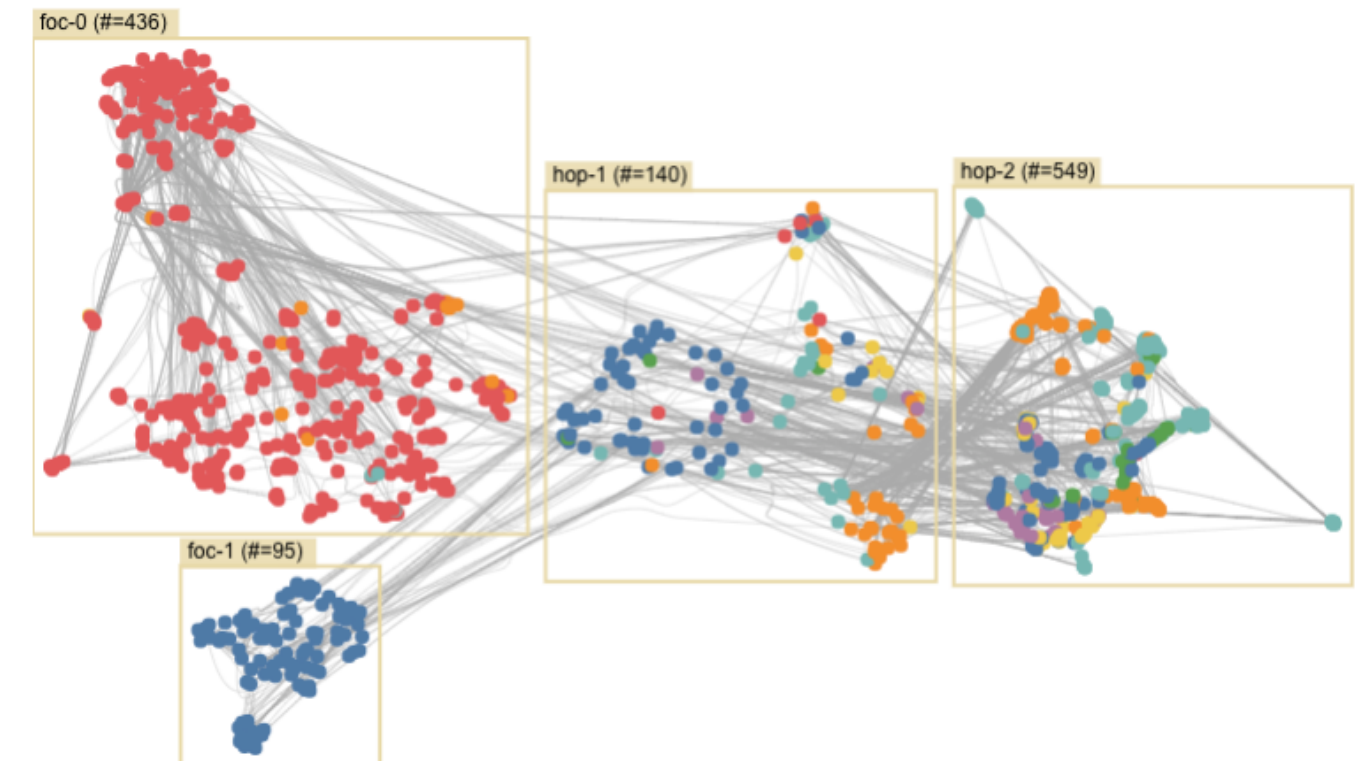
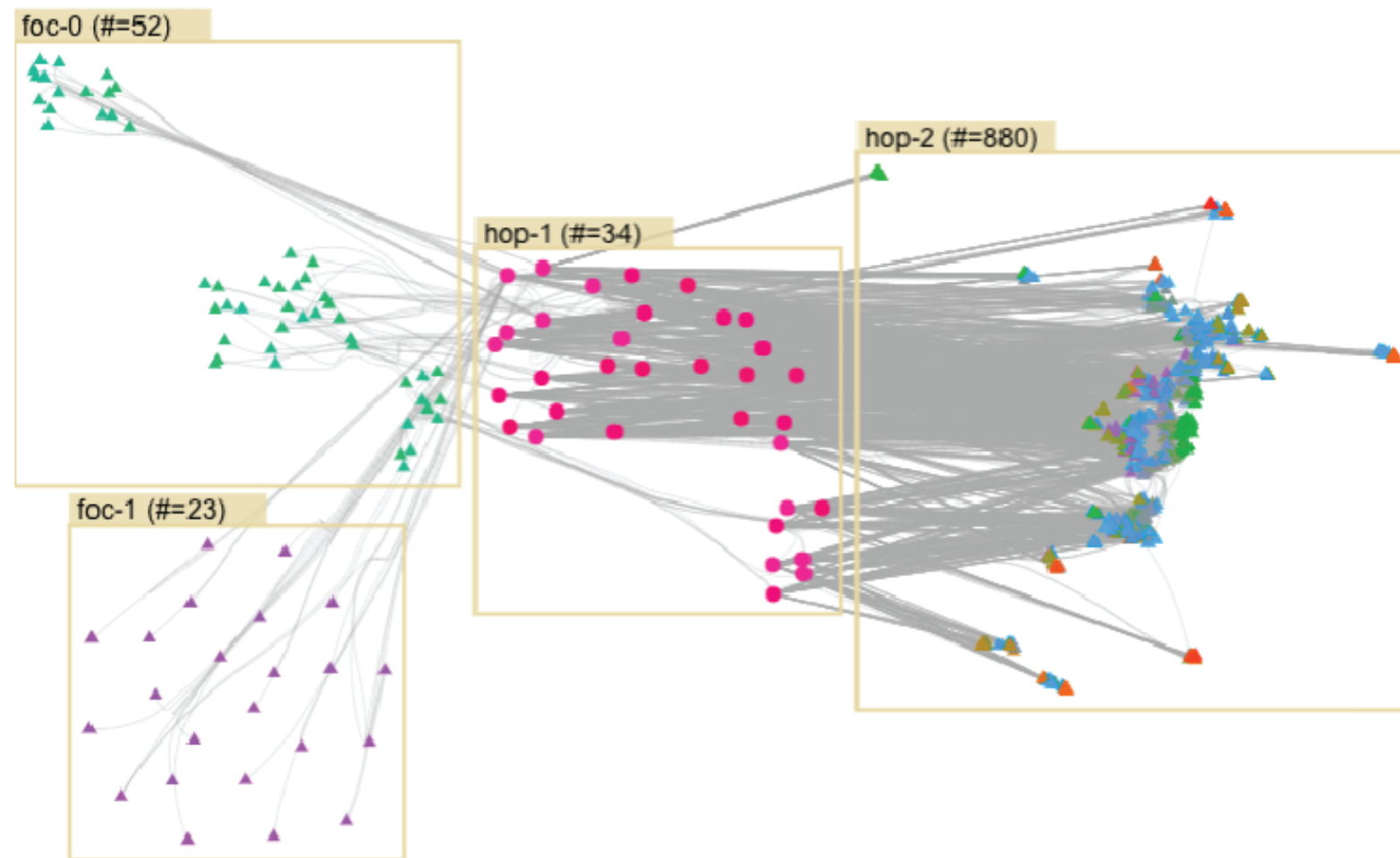
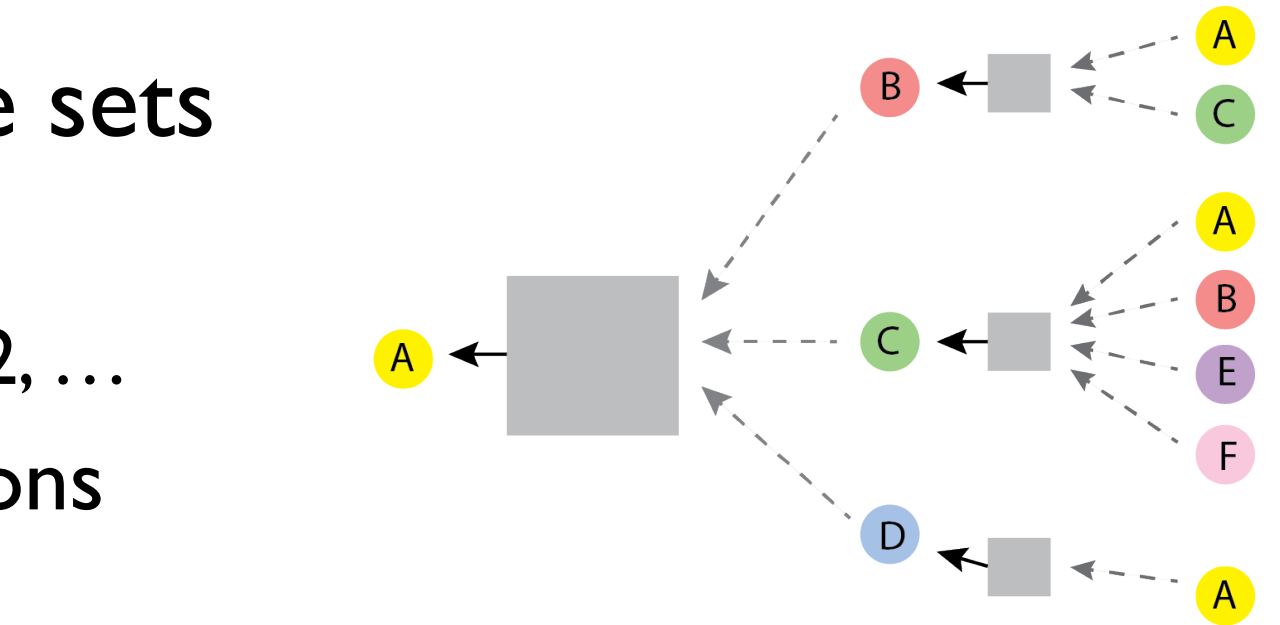
Contribution: Multi-view interactive interface



Video: <https://osf.io/j56hu/>

Contribution: K-hop layout

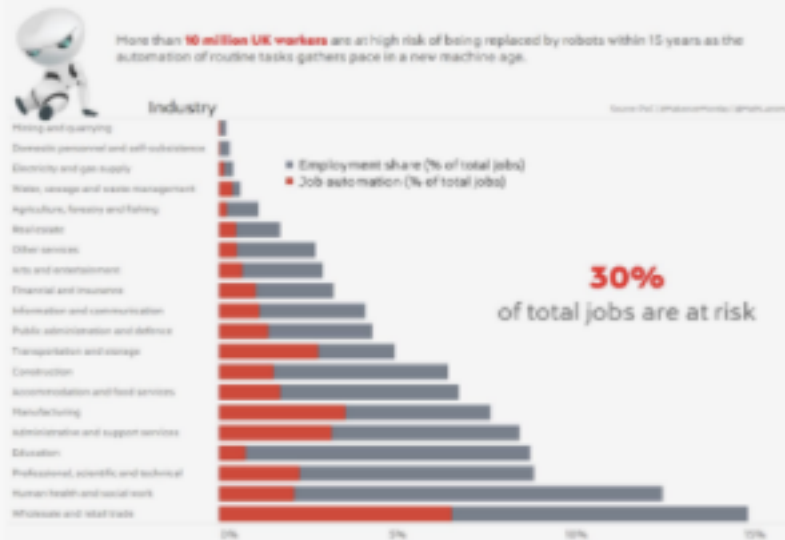
- show topo neighbors of user-specified node sets
 - mimic how info is aggregated in GNN
 - boxes from left to right: specified nodes, hop-1, hop-2, ...
 - within box, cluster nodes by their topo connections



Next Steps

Learning through Redesign En Masse: **Makeover Mondays**

- easy entry point (Tableau focus)



Week 14 – Millions of UK workers at risk of being replaced by robots

Apr 7, 2017

During week 14 we looked at job automation and the potential impact of robots and AI on the UK employment market.



Week 13 – The Secret of Success

Mar 31, 2017

Week 13 took a look at a Russian survey about the secret of success. Dot plot, bump charts, bar charts, radar charts. This week had it all! Plus seven lessons to take on board.



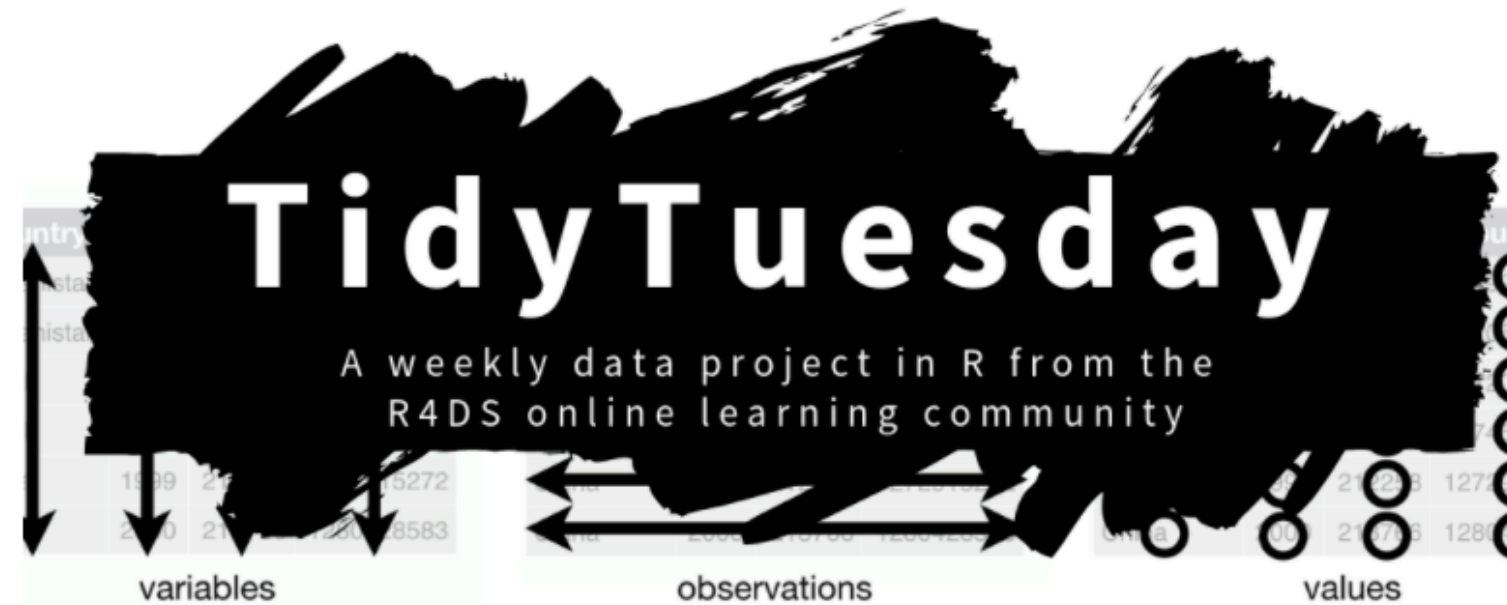
Week 12 – March Madness

Mar 24, 2017

We looked at March Madness data for week 12, highlighting the phenomenon that is US college basketball. Quite a few vizzes showed the passion that

Learning through Redesign En Masse: **Tidy Tuesdays**

- easy entry point (R focus)



A weekly social data project in R

A weekly data project aimed at the R ecosystem. As this project was borne out of the R4DS Online Learning Community and the R for Data Science textbook, an emphasis was placed on understanding how to summarize and arrange data to make meaningful charts with `ggplot2`, `tidyr`, `dplyr`, and other tools in the tidyverse ecosystem. However, any code-based methodology is welcome - just please remember to share the code used to generate the results.

<https://github.com/rfordatascience/tidytuesday>

Visual Design Process In Depth: **Dear Data**

- inspiring celebration of data humanism



Giorgia Lupi and
Stefanie Posavec

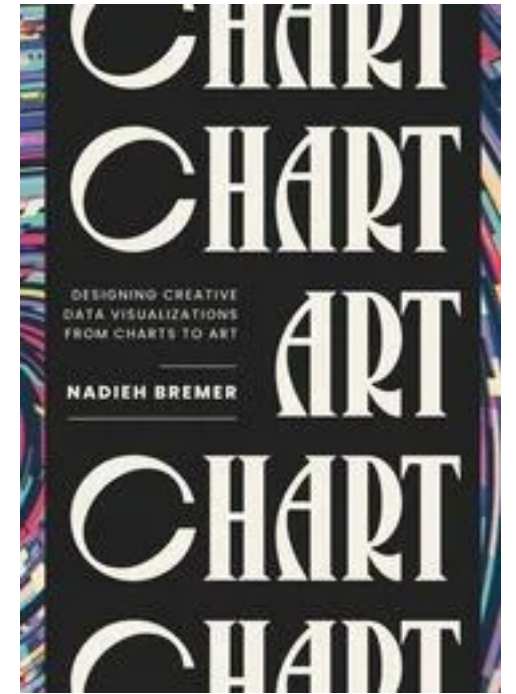
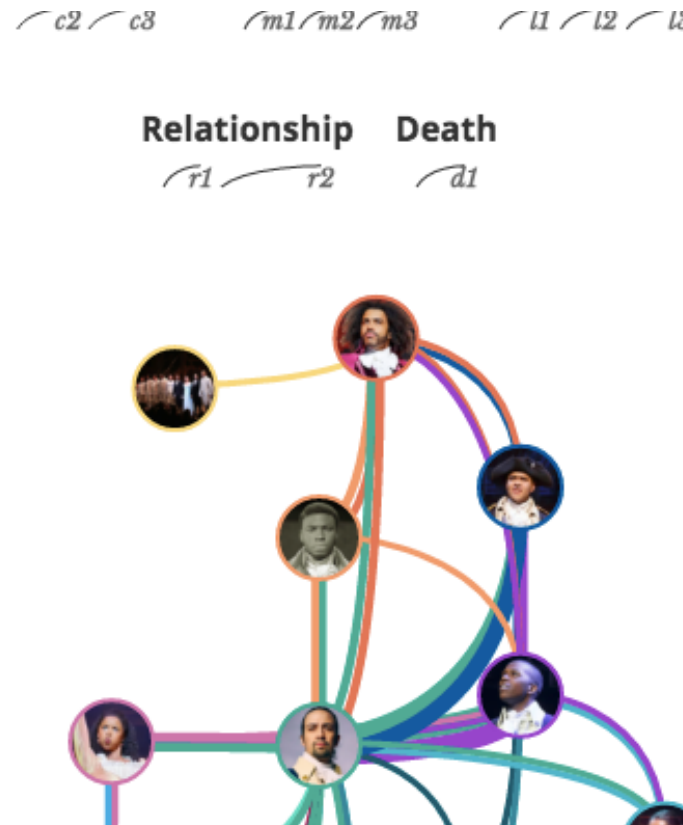
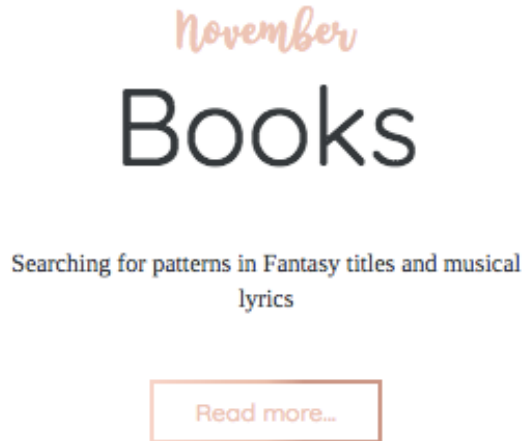
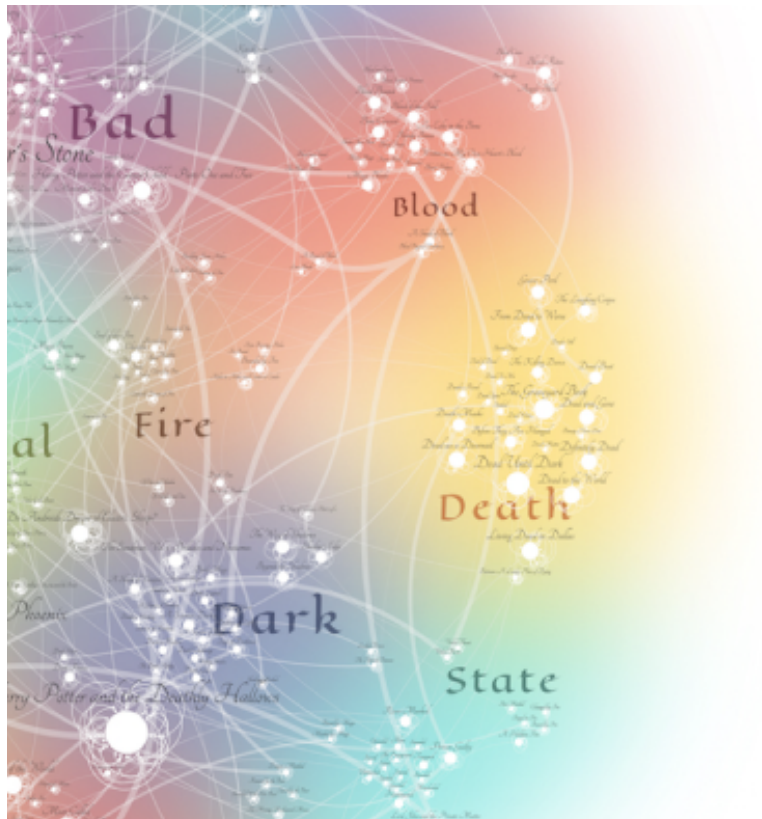
<http://www.dear-data.com/by-week/>

<https://giorgialupi.com/dear-data>

<https://www.stefanieposavec.com/dear-data>

Visual Design Process In Depth: **Data Sketches**

- detailed process notes, from sketching through coding
- Shirley Wu & Nadieh Bremer



new book out
from Bremer:
CHART
Designing Creative
Data Visualizations
from Charts to Art

<http://www.datasketch.es/>

Pathways for more participation: organizations

- join Vancouver Visualization meetup (4K+ members)
 - <https://www.meetup.com/Vancouver-Data-Visualization/>
- join Data Visualization Society
 - <https://www.datavisualizationsociety.org>
 - four years old, 23K+ members around the world
 - **jobs board:** full-time, part-time and contract positions worldwide
<https://jobs.datavisualizationsociety.org/>
 - many other resources, super-active Slack incl local groups, challenges, ...
 - articles: [Nightingale](#)
 - conferences: [Outlier](#)
 - awards: [Information Is Beautiful](#)

Visualization jobs

- spectrum
 - visualization as main/core focus
 - visualization as occasional task
 - visualization skills add strength to your portfolio even if no immediate duties
- local companies
 - Tableau Vancouver is largest company focused on visualization
 - many smaller ones have visualization / data science needs

Upcoming

- see you next week for final presentations
- enjoy visualization, for those who keep going down this path!
 - now or later...