

Variant View

Genomics, Big Data, and Patient Privacy Implications

Tamara Munzner

Department of Computer Science
University of British Columbia

THINK Conference

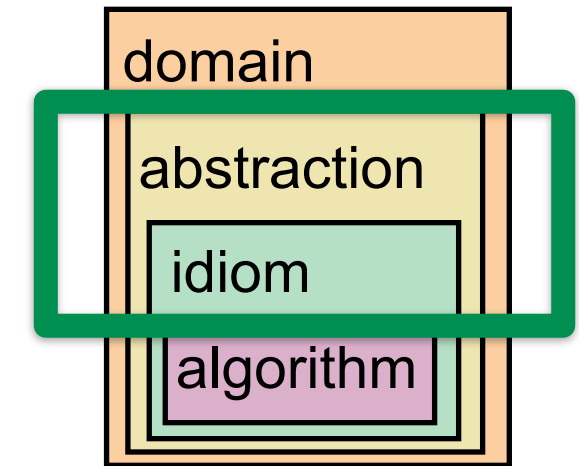
6 Nov 2015, Santa Cruz CA

<http://www.cs.ubc.ca/~tmm/talks.html#think15>

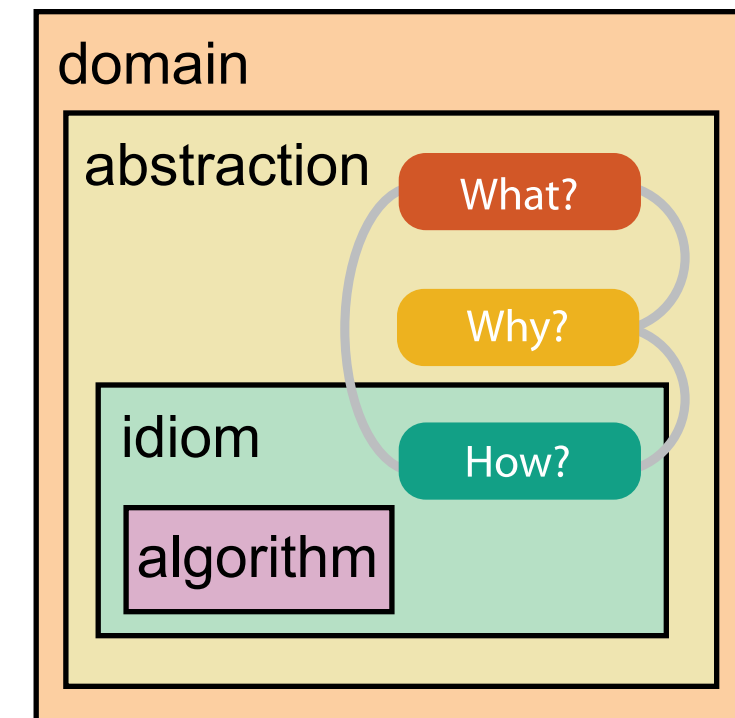
[@tamaramunzner](#)

Visualization analysis framework: Four levels, three questions

- *domain* situation
 - who are the target users?
- *abstraction*
 - translate from specifics of domain to vocabulary of vis
- **what** is shown? **data abstraction**
 - often don't just draw what you're given: transform to new form
- **why** is the user looking at it? **task abstraction**
- *idiom*
 - **how** is it shown?
 - **visual encoding idiom**: how to draw
 - **interaction idiom**: how to manipulate
- *algorithm*
 - efficient computation



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



[A Multi-Level Typology of Abstract Visualization Tasks
Brehmer and Munzner. *IEEE TVCG* 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]

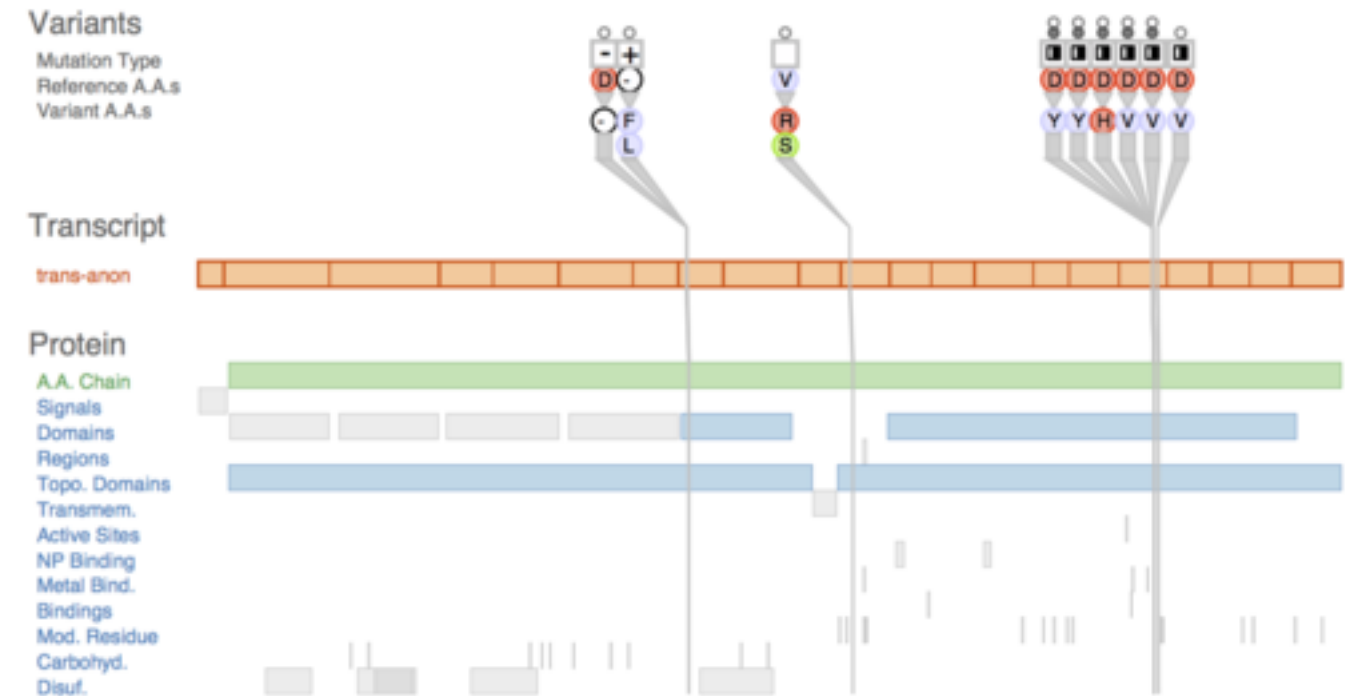
Variant View

Visualizing Sequence Variants in their Gene Context

joint work with:

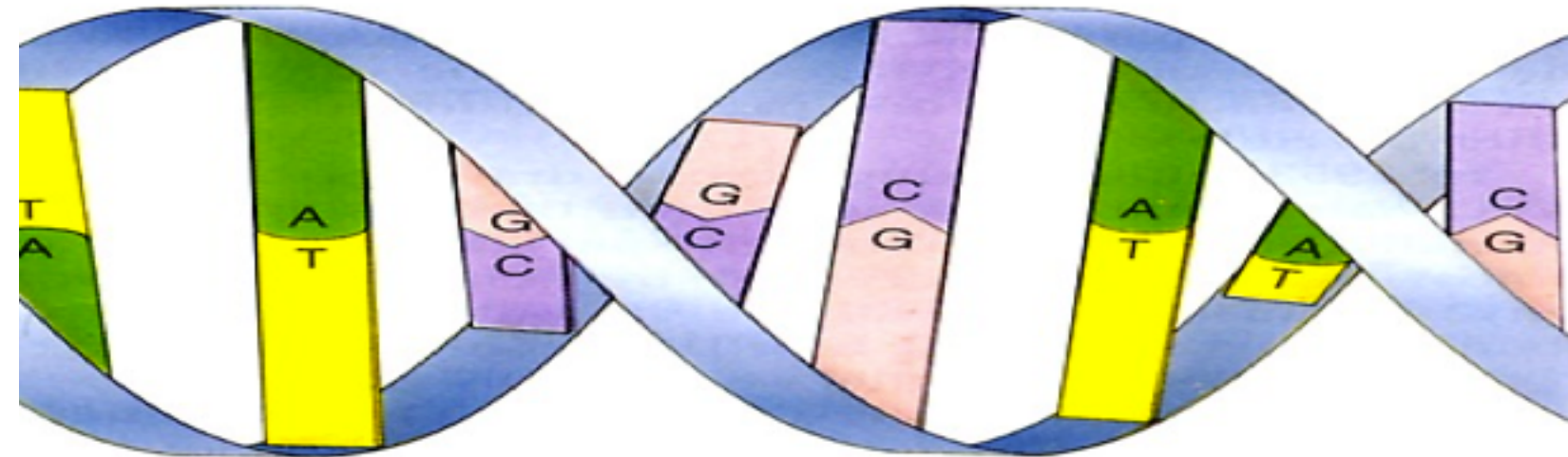
Joel Ferstay, Cydney Nielsen

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



Sequence Variant Definition

- Sequence variants
 - Difference between reference and given person's 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?

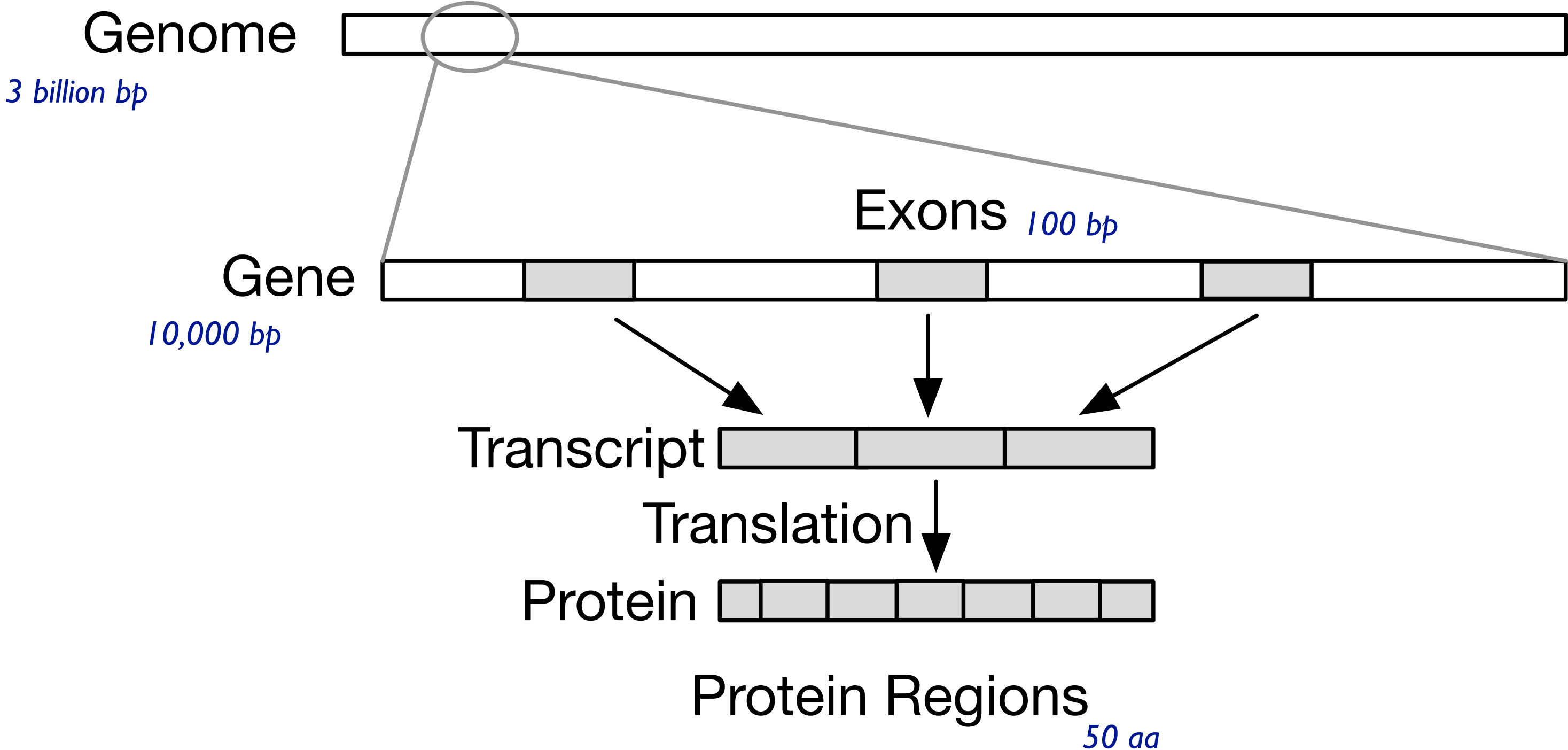
Harmless?

Cancer Research

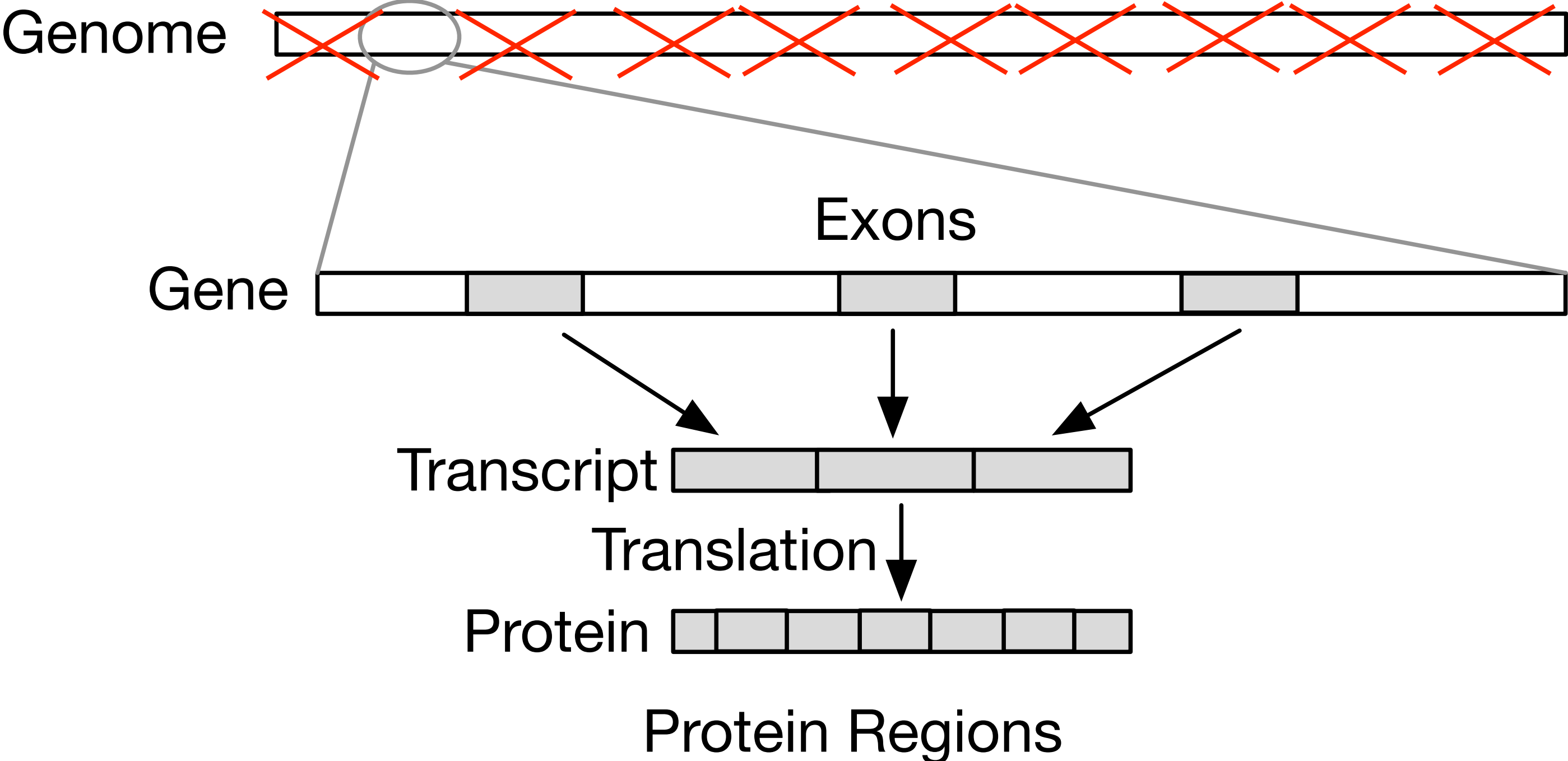
- collaboration with analysts at BC Genome Sciences Center
 - studying genetic basis of leukemia
- driving task
 - discover new candidate genes with harmful variants
- two big questions
 - what to show
 - data abstraction
 - challenge: enormous range of scales in the data
 - how to show it
 - visual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

Abstractions

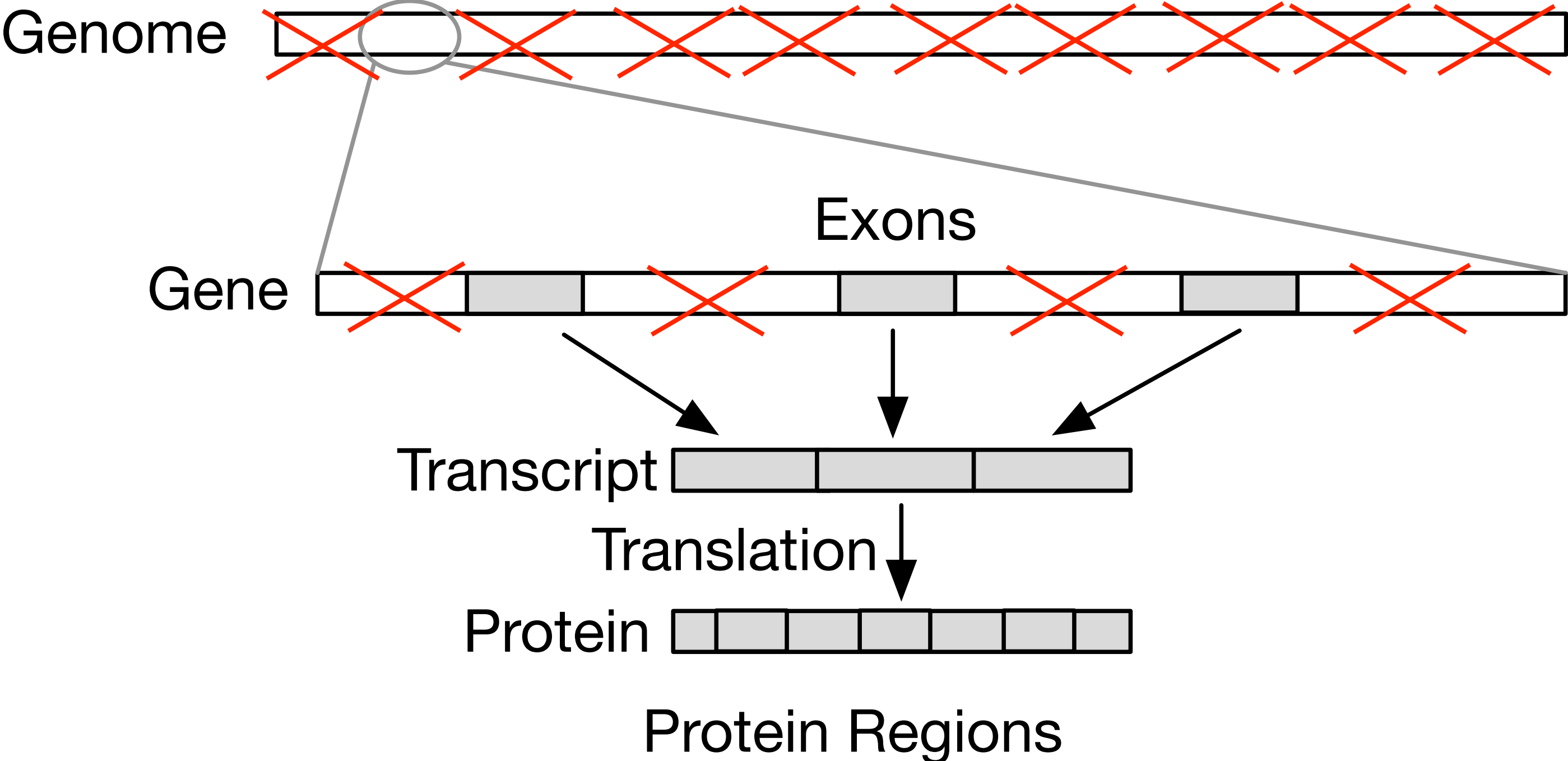
Data: Filtering to relevant biological levels and scales



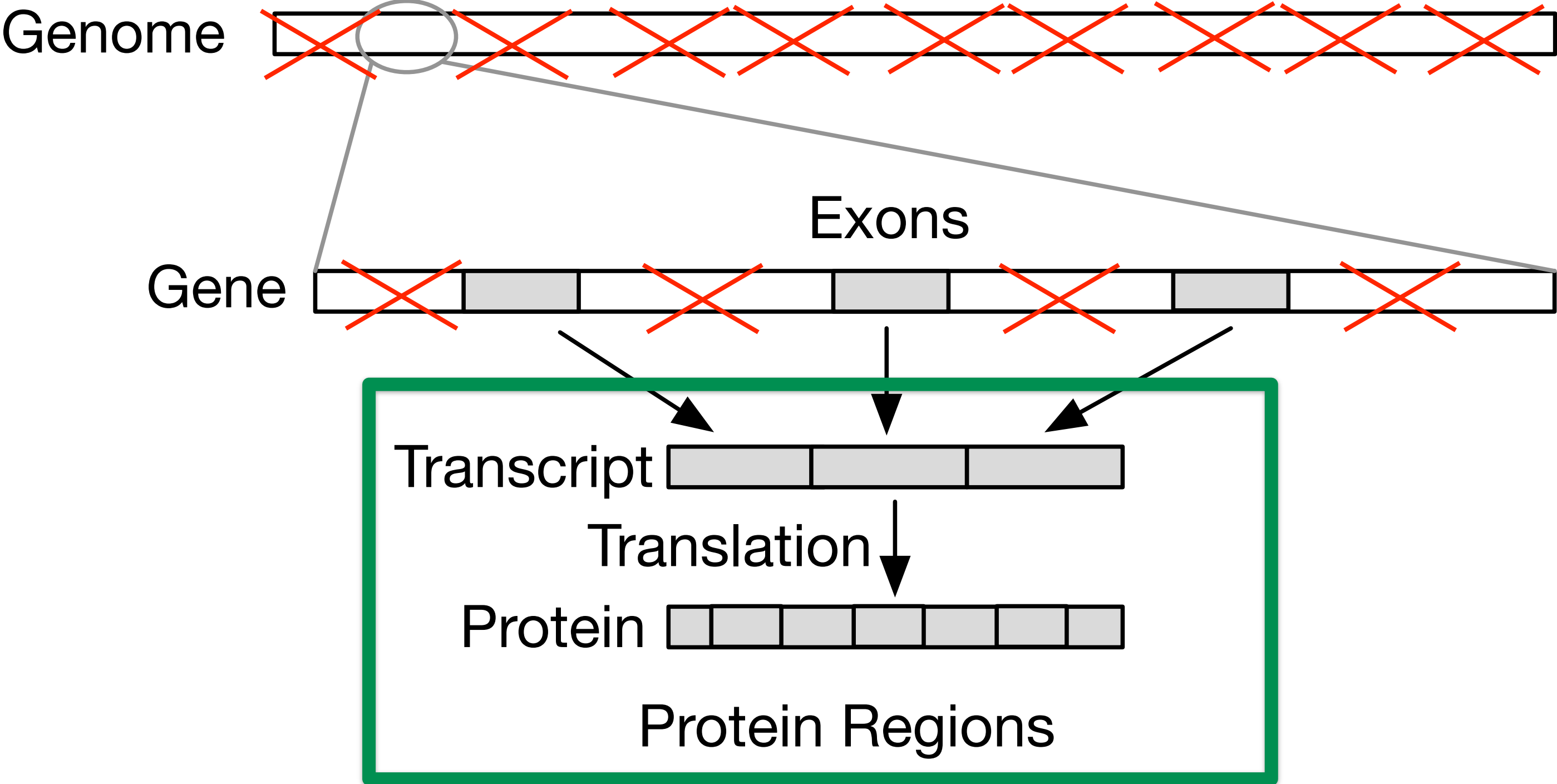
Filter out whole genome; keep genes



Filter out non-exon regions

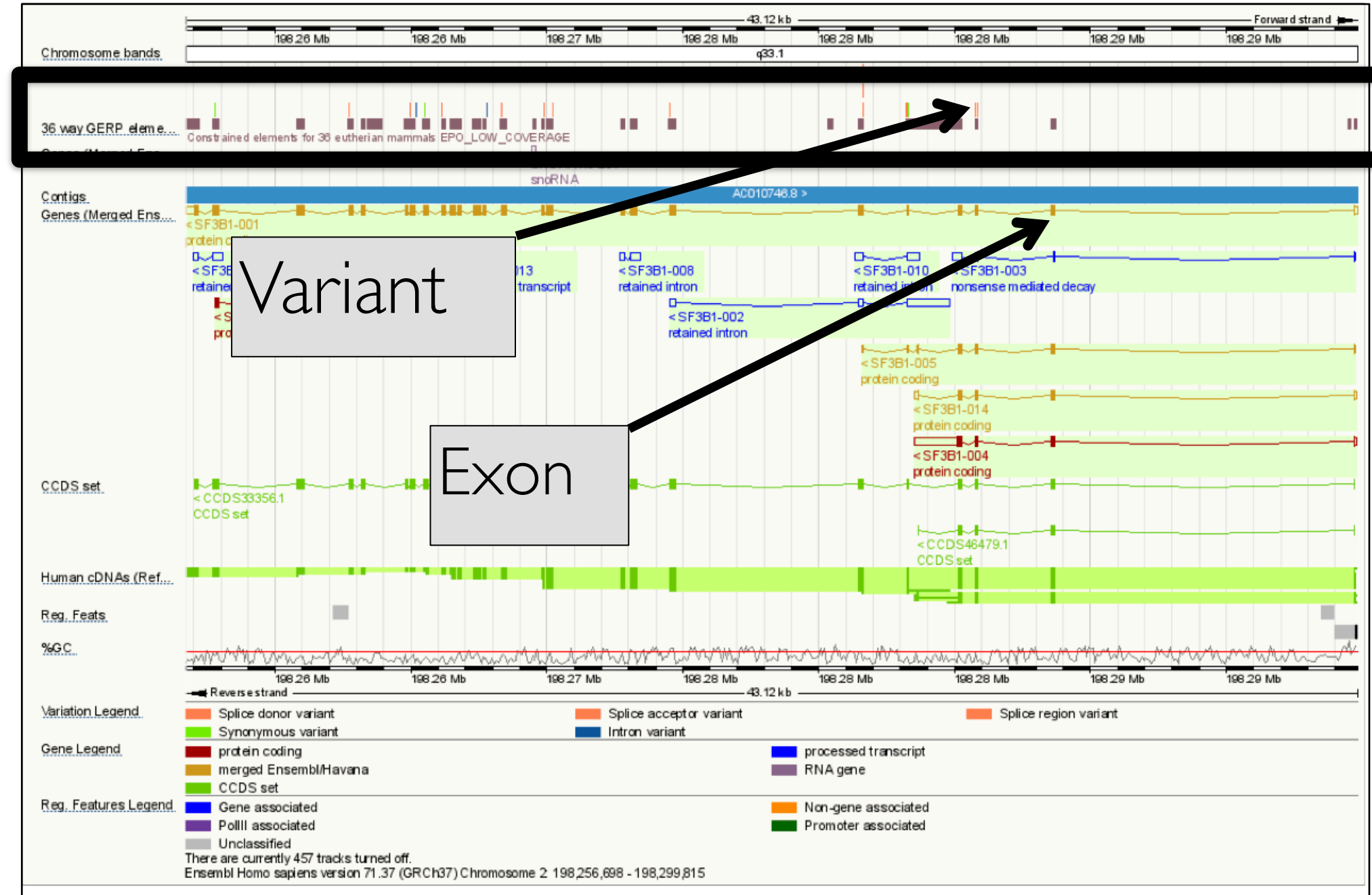


Data abstraction: highly filtered scope



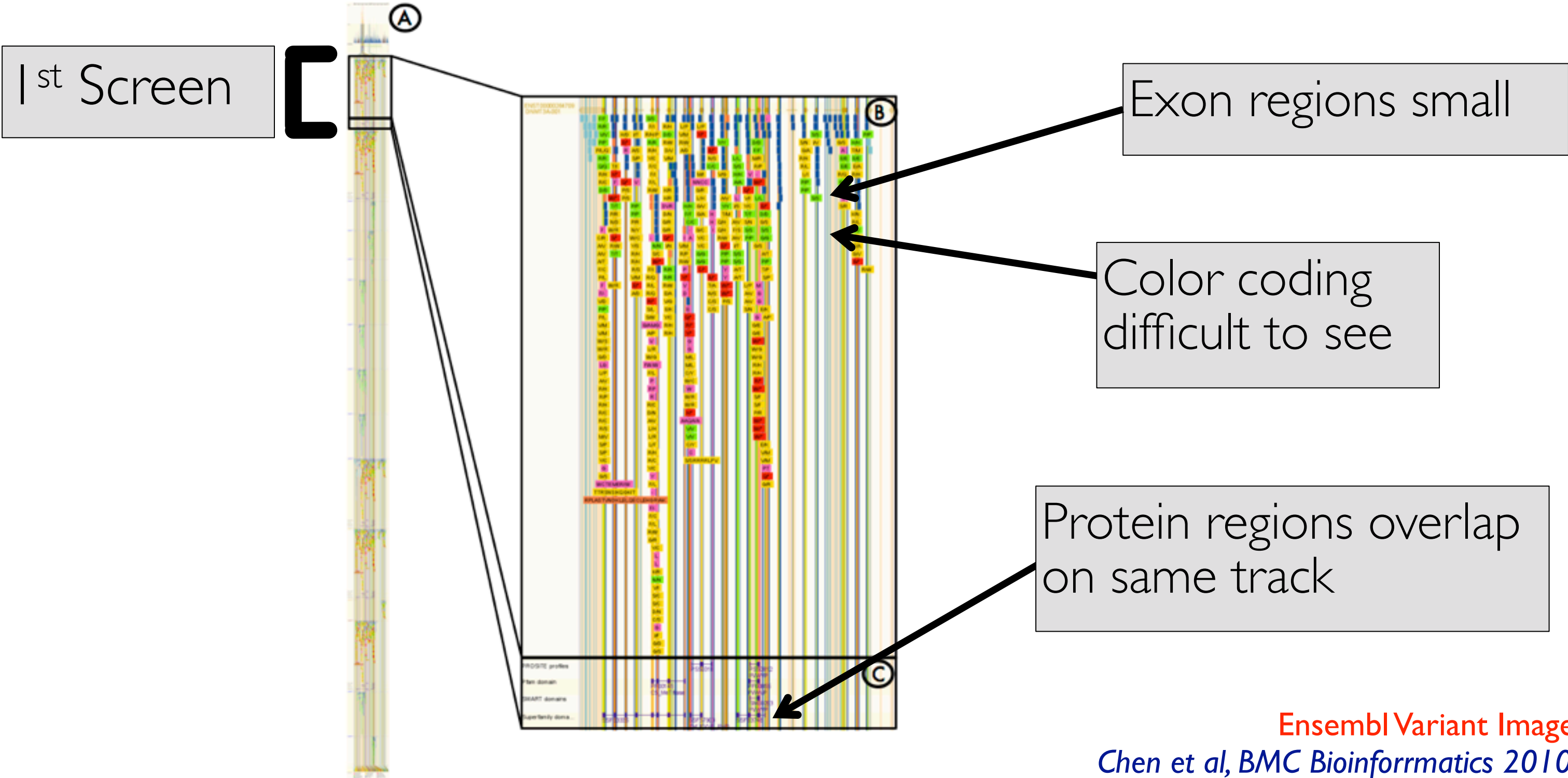
Dominant paradigm: genome browsers

- 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**

- 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 |

Variant View

Previous: Table, one row per variant

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

Sort By Gene:
Alpha Cluster Score Variant Count

Variant Data

| Patient ID | Chr. | Coord. | Ref Base | Var Base | dbSNP129 | dbSNP135 | dbSNP137 | COSMIC | A.A. Ch | Gene | Transcript |
|------------|------|--------|----------|----------|----------|----------|----------|---------|---------|-----------|------------|
| pid-anon | 11 | 288816 | G | T | . | . | . | *13028, | G60V | gene-anon | trans-anon |
| pid-anon | 11 | 288816 | G | T | . | . | . | *13012, | D61Y | gene-anon | trans-anon |
| pid-anon | 11 | 288819 | G | T | . | rs121918 | . | 13014 | A72S | gene-anon | trans-anon |
| pid-anon | 11 | 288819 | C | T | . | . | . | *13035, | A72V | gene-anon | trans-anon |
| pid-anon | 11 | 288821 | G | C | . | . | . | *13016, | E76Q | gene-anon | trans-anon |
| pid-anon | 11 | 288821 | A | G | . | rs121918 | . | *13017, | E76G | gene-anon | trans-anon |
| pid-anon | 11 | 288821 | G | T | . | . | . | . | E76D | gene-anon | trans-anon |
| pid-anon | 11 | 292688 | T | A | . | rs121918 | . | *13020, | S502T | gene-anon | trans-anon |
| pid-anon | 11 | 292688 | T | G | . | . | . | *13020, | S502A | gene-anon | trans-anon |
| pid-anon | 11 | 292688 | C | T | . | . | . | 13023 | S502L | gene-anon | trans-anon |

No need for pan and zoom

The screenshot displays a web-based variant view for a single gene. At the top, there is a 'Gene Search' field with a 'Submit' button. Below this, the 'Alternative Transcripts' section shows 'gene-anon (trans-anon)'. The 'Variants' section includes a legend for Mutation Type, Reference A.A.s, and Variant A.A.s, with a diagram showing amino acid changes: G to D, A to A, E to E, E to E, V to Y, S to V, Q to G, and D to D. The 'Transcript' section shows an orange bar representing the 'trans-anon' transcript. The 'Protein' section shows a green bar representing the 'A.A. Chain' and a blue bar representing 'Domains'. A 'Variant Data' table at the bottom lists patient IDs, coordinates, reference and variant bases, dbSNP IDs, COSMIC IDs, and amino acid changes. On the right side, there is a 'Sort By Gene' section with options for 'Alpha', 'Cluster Score', and 'Variant Count' (selected). A list of genes is shown, including DNMT3A, IDH2, FLT3, ANKRD36, ARID1B, STAG2, TNRC18, WT1, ABCA13, CEBPA, TET2, DNAH10, GPSM1, ASXL1, DNAH1, and DNAH9. Annotations with arrows point to the main view area and the variant data table.

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:

- 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 and amino acid changes. The 'Transcript' section shows a bar representing the transcript with variant positions. The 'Protein' section shows a bar representing the protein with domains and regions. The 'Variant Data' table is at the bottom. On the right, a list of genes is shown, sorted by variant count. Callouts A, B, and C point to specific features: A points to the 'Sort By Gene' dropdown, B points to the 'Variant Data' table, and C points to the gene list.

Alternative Transcripts: gene-anon (trans-anon)

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 |
|-----------|------------|----------|---------|------------|-----------|------------|
| gene-anon | trans-anon | rs121918 | *13028, | G60V | gene-anon | trans-anon |
| gene-anon | trans-anon | rs121918 | *13012, | D61Y | gene-anon | trans-anon |
| gene-anon | trans-anon | rs121918 | 13014 | A72S | gene-anon | trans-anon |
| gene-anon | trans-anon | rs121918 | *13035, | A72V | gene-anon | trans-anon |
| gene-anon | trans-anon | rs121918 | *13016, | E76Q | gene-anon | trans-anon |
| gene-anon | trans-anon | rs121918 | *13017, | E76G | gene-anon | trans-anon |
| gene-anon | trans-anon | rs121918 | . | E76D | gene-anon | trans-anon |
| gene-anon | trans-anon | rs121918 | *13020, | S502T | gene-anon | trans-anon |
| gene-anon | trans-anon | rs121918 | *13020, | S502A | gene-anon | trans-anon |
| gene-anon | trans-anon | rs121918 | 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)

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, | E76D | 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 |

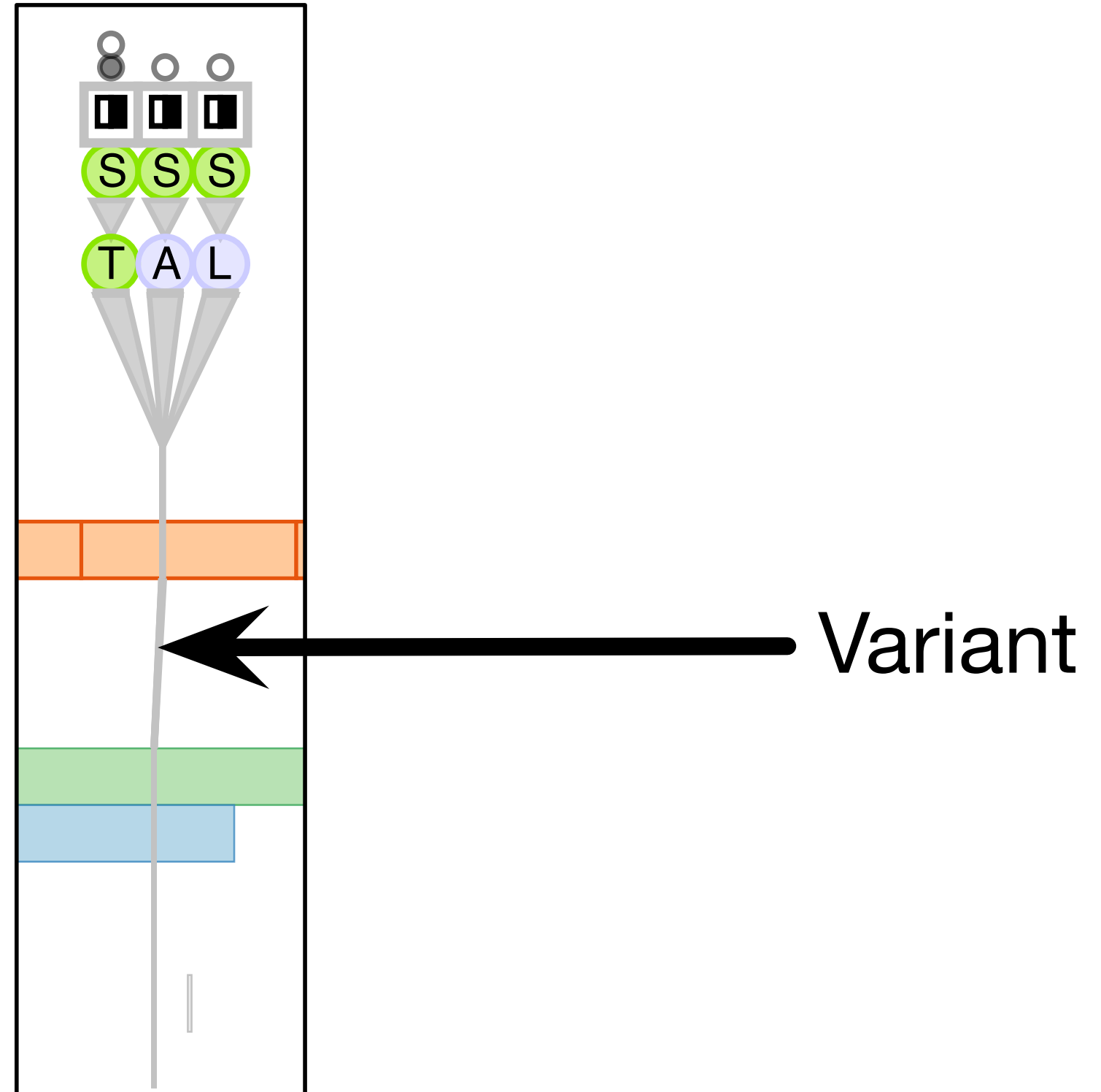
Sort By Gene:
Alpha Cluster Score Variant Count

- 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

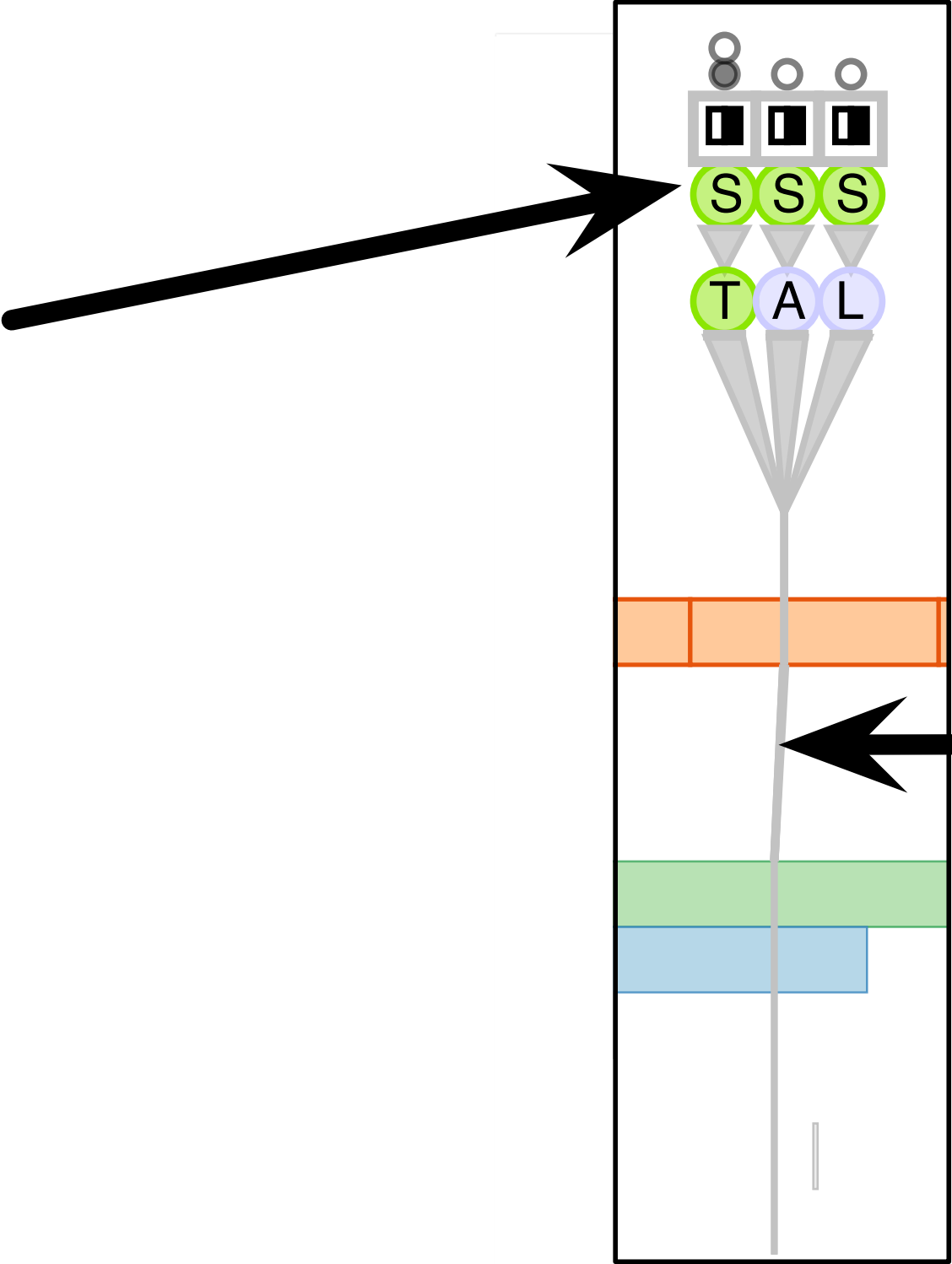
Design information-dense visual encoding

- 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

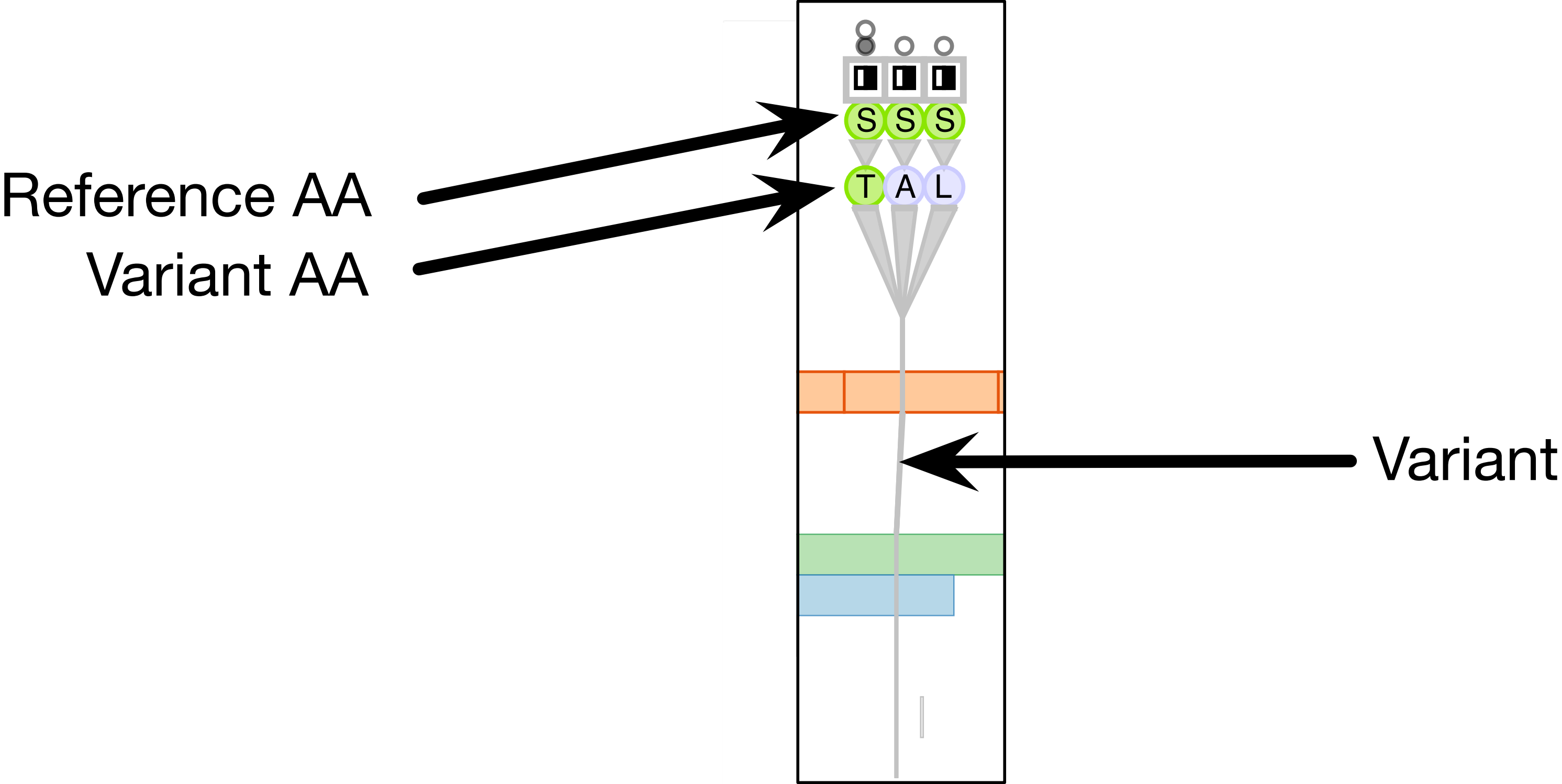
Reference AA



Variant



Design information-dense visual encoding



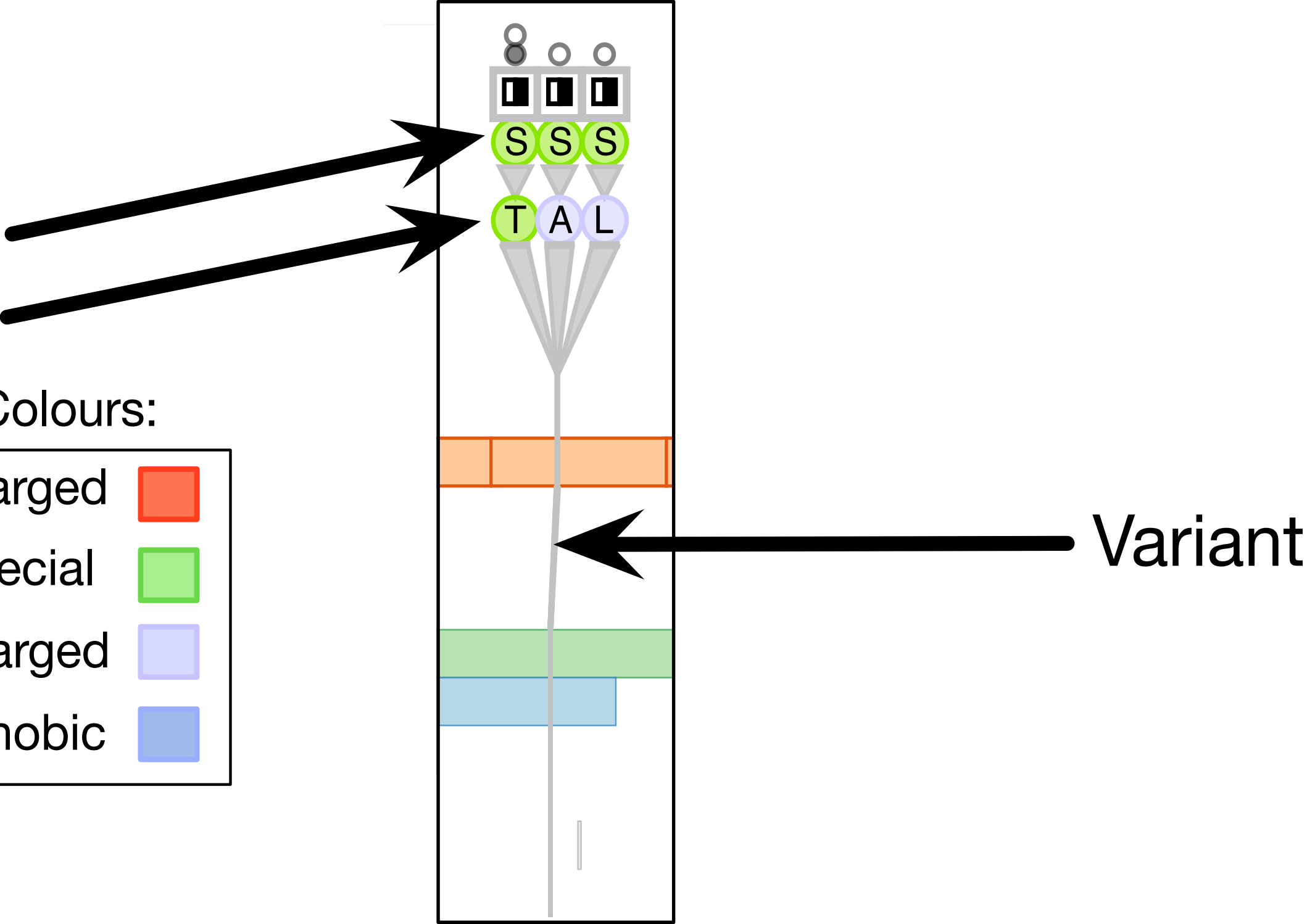
Design information-dense visual encoding

Reference AA

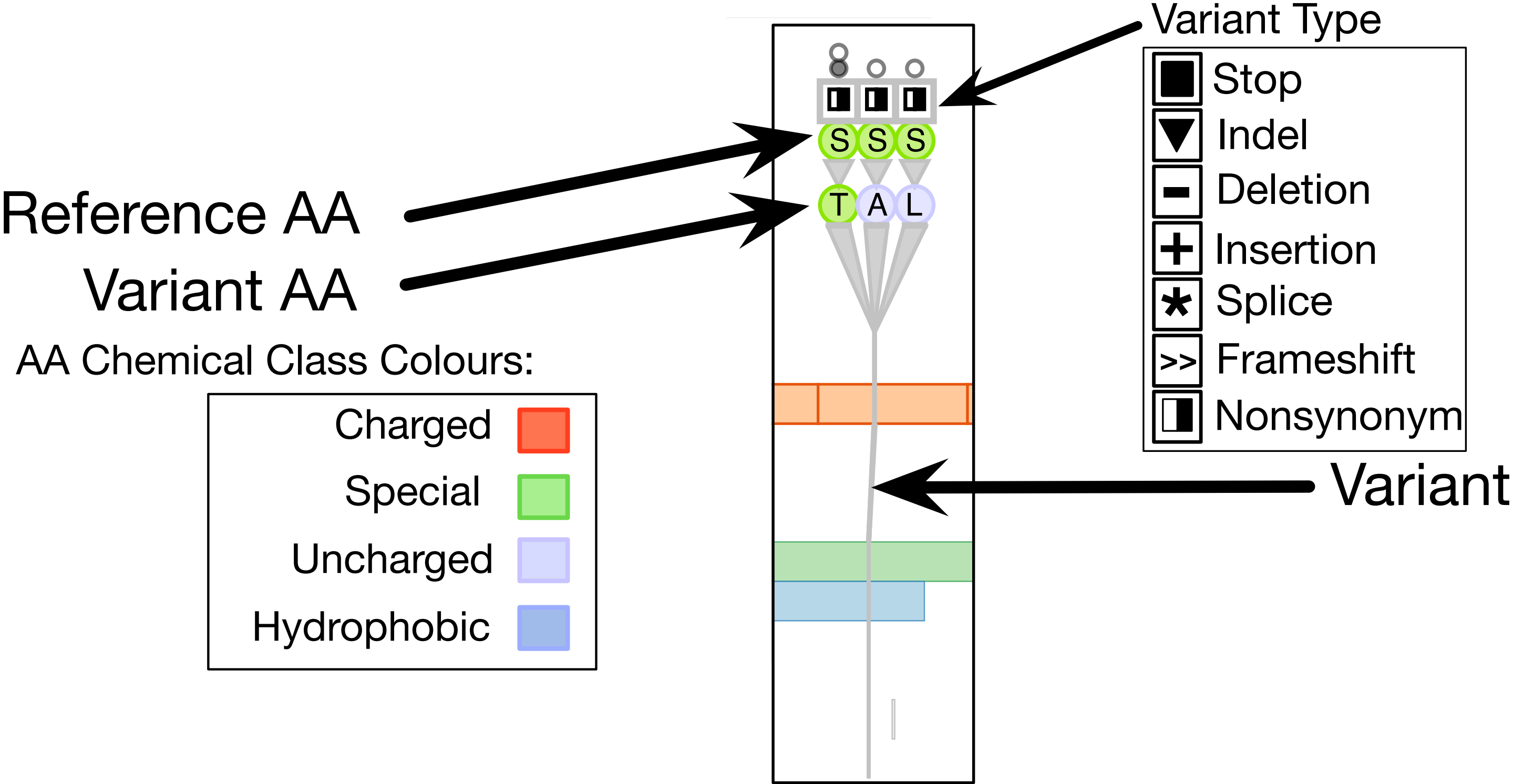
Variant AA

AA Chemical Class Colours:

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



Design information-dense visual encoding



Design information-dense visual encoding

Known Database

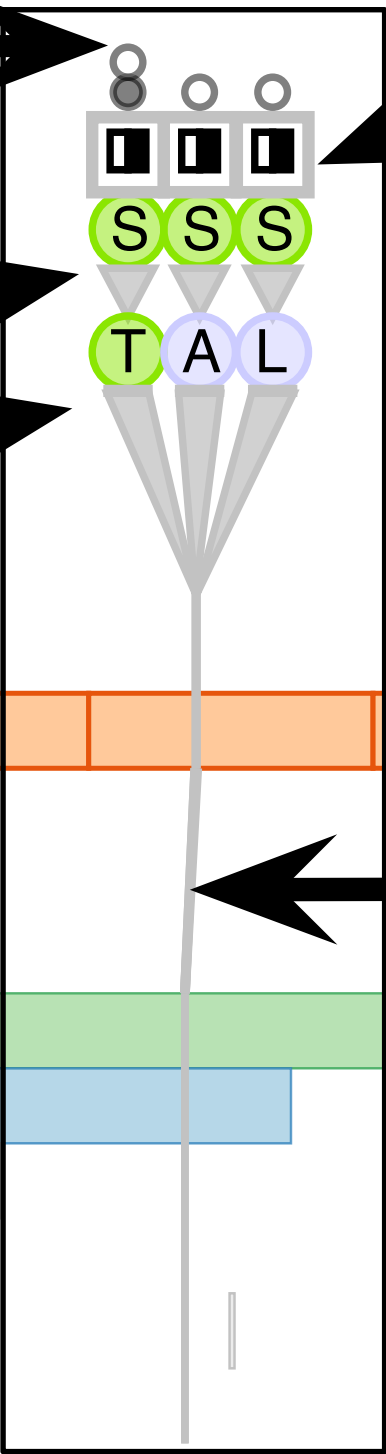
- Known Harmless
- Known Cancer

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 ■

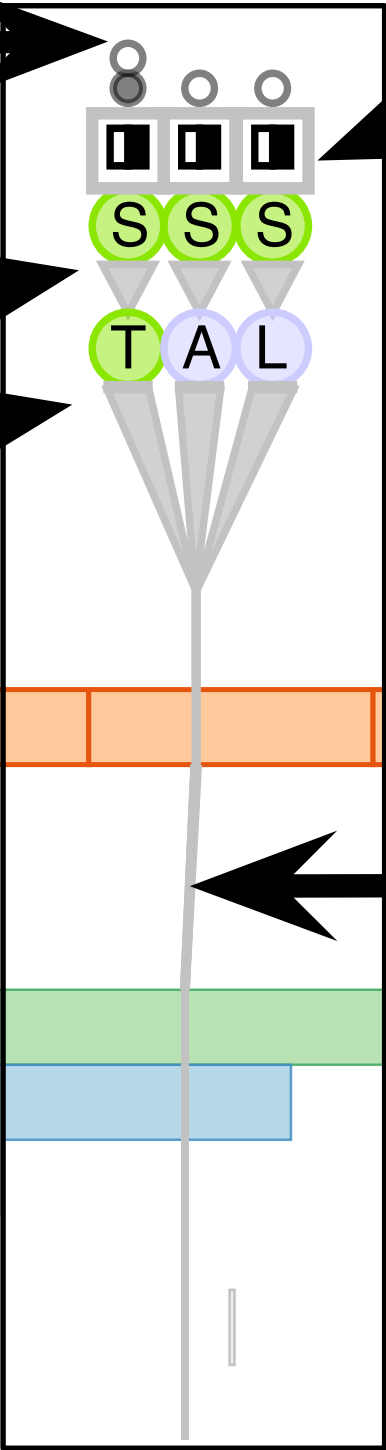
Variant Type

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

Variant

Transcript/Region Colours:

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

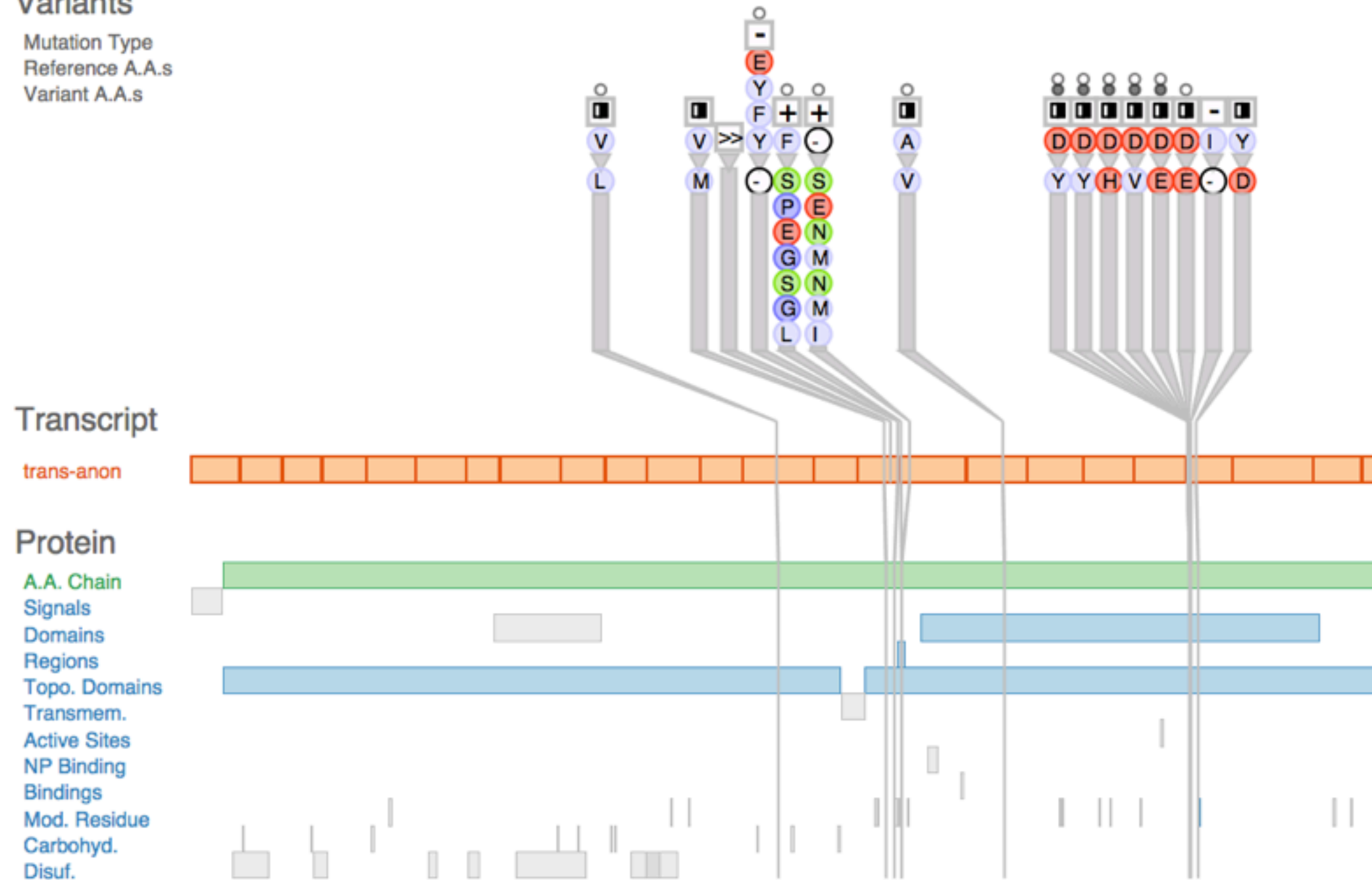


Results

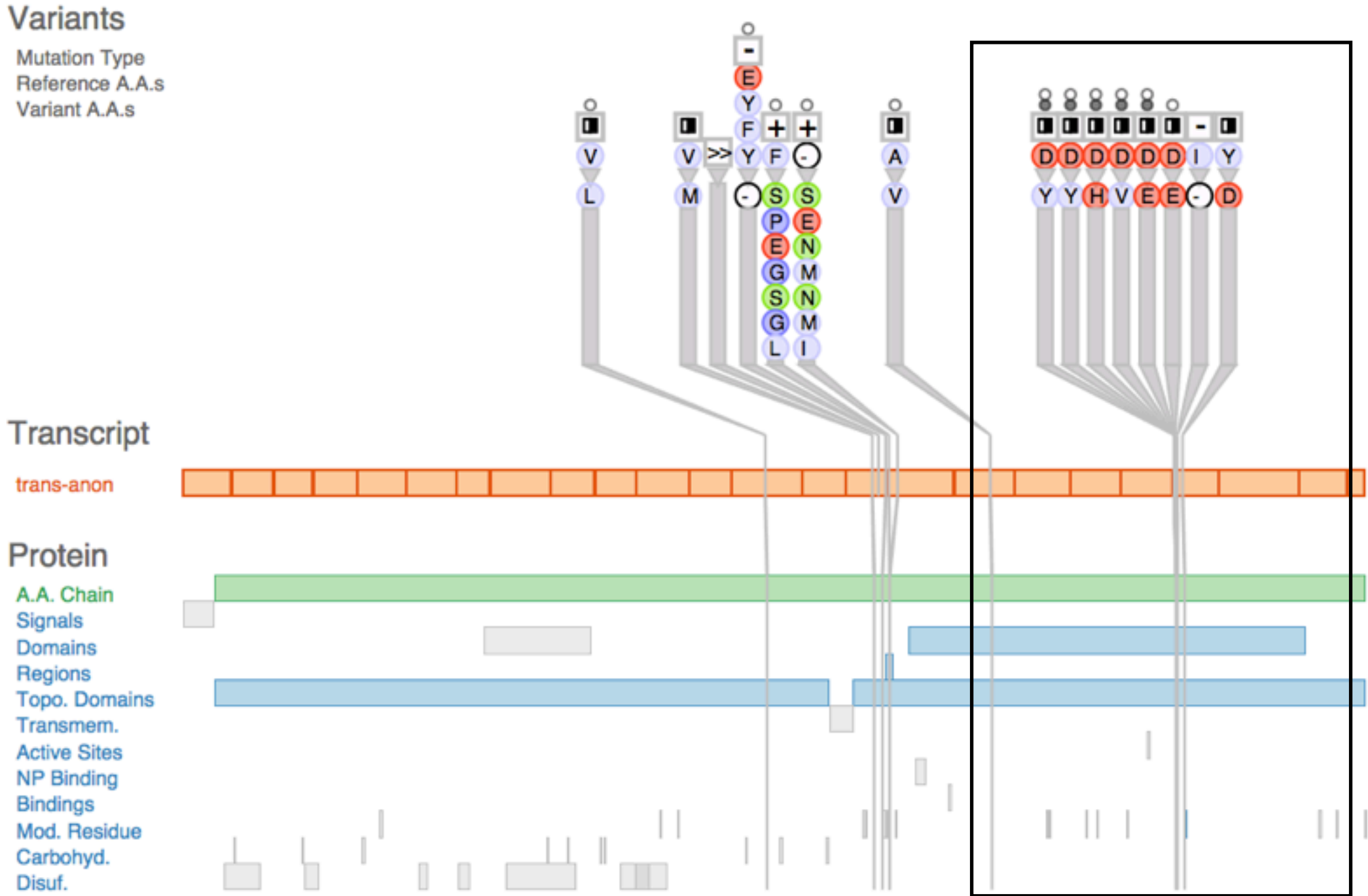
Verify known leukemia gene: Highly scored by sorting metric

Variants

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



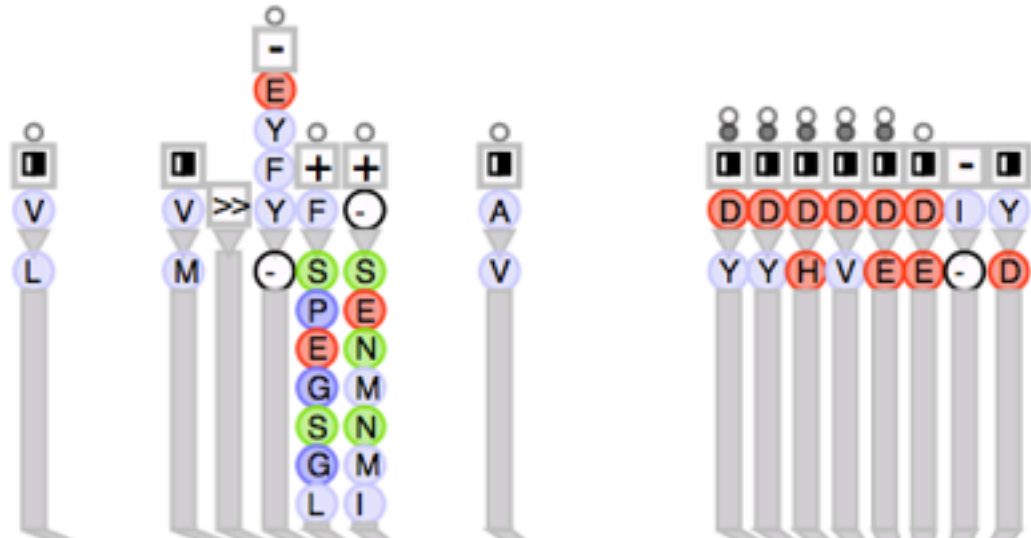
Visual inspection reveals collocation of variants



Several functional protein regions affected

Variants

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



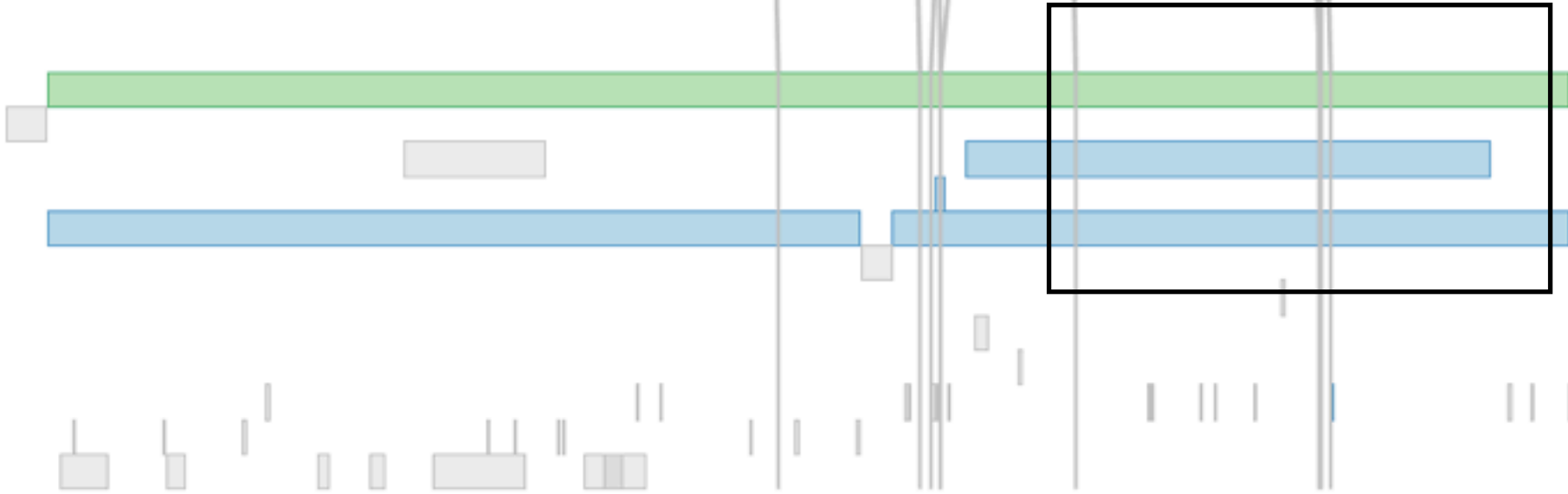
Transcript

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



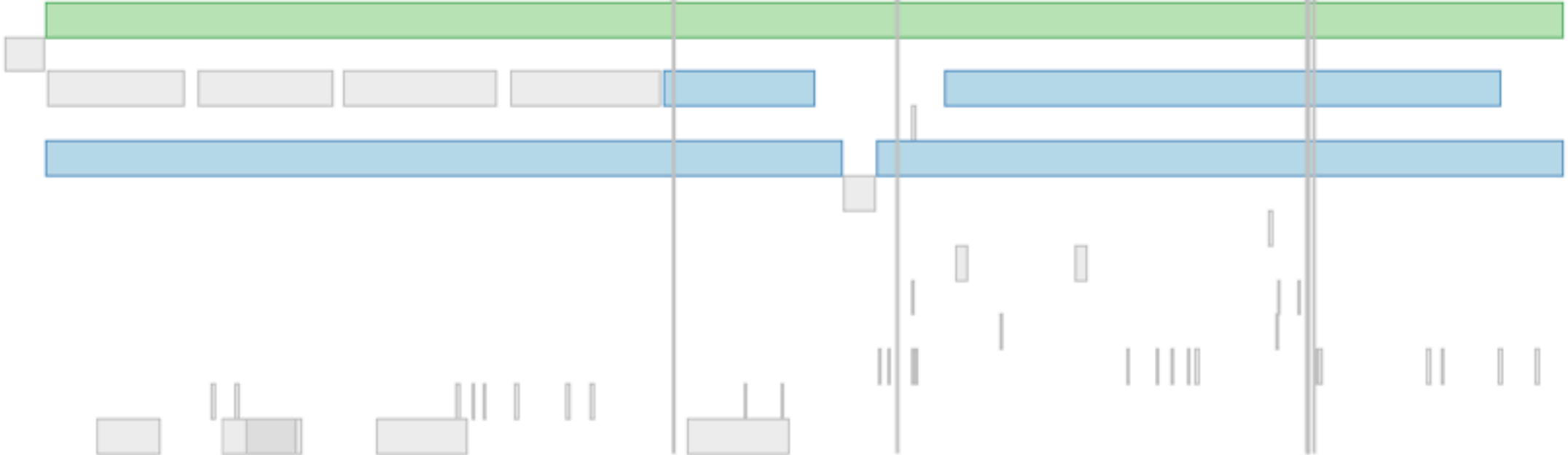
Transcript

trans-anon

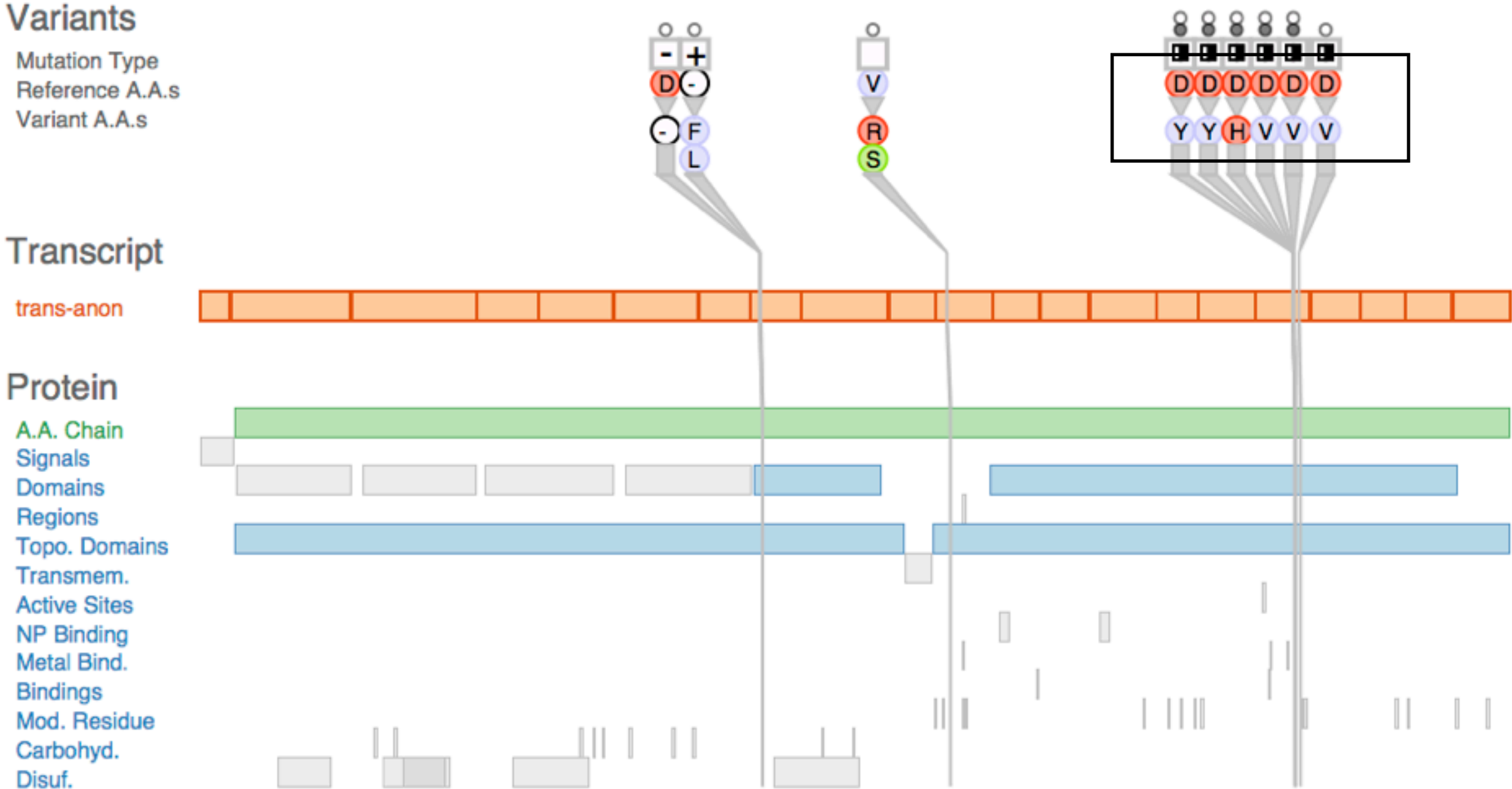


Protein

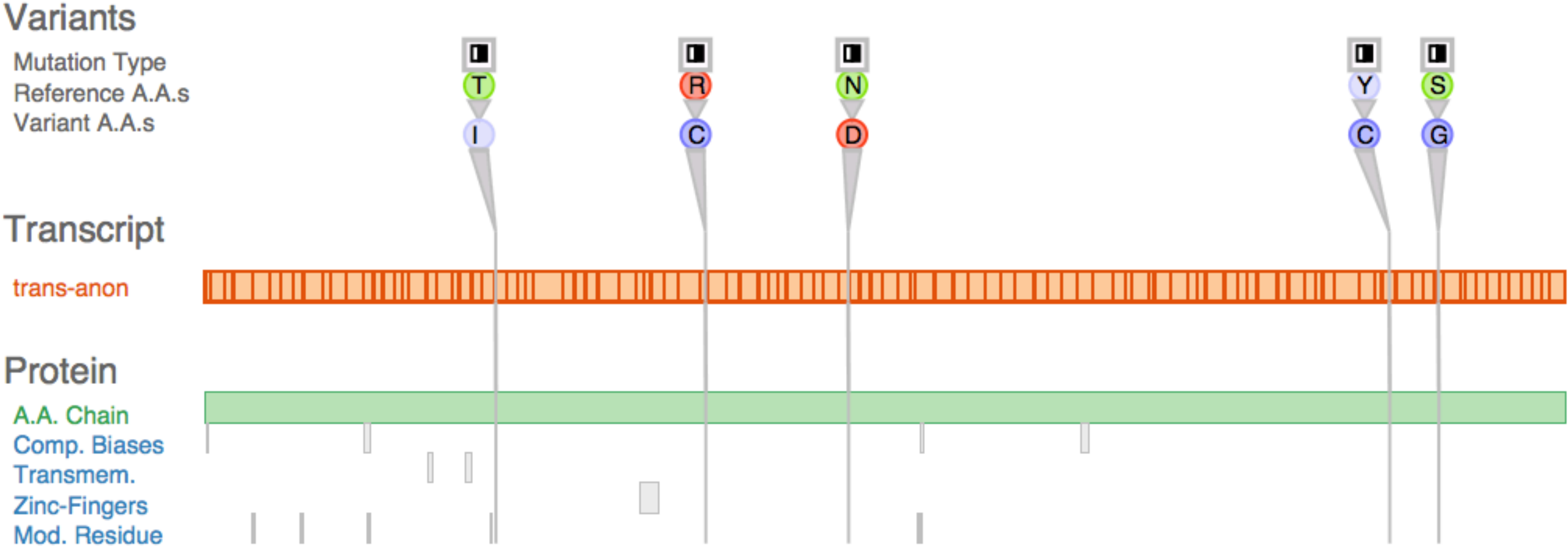
A.A. Chain
Signals
Domains
Regions
Topo. Domains
Transmem.
Active Sites
NP Binding
Metal Bind.
Bindings
Mod. Residue
Carbohyd.
Disuf.



Protein chemical class change evident



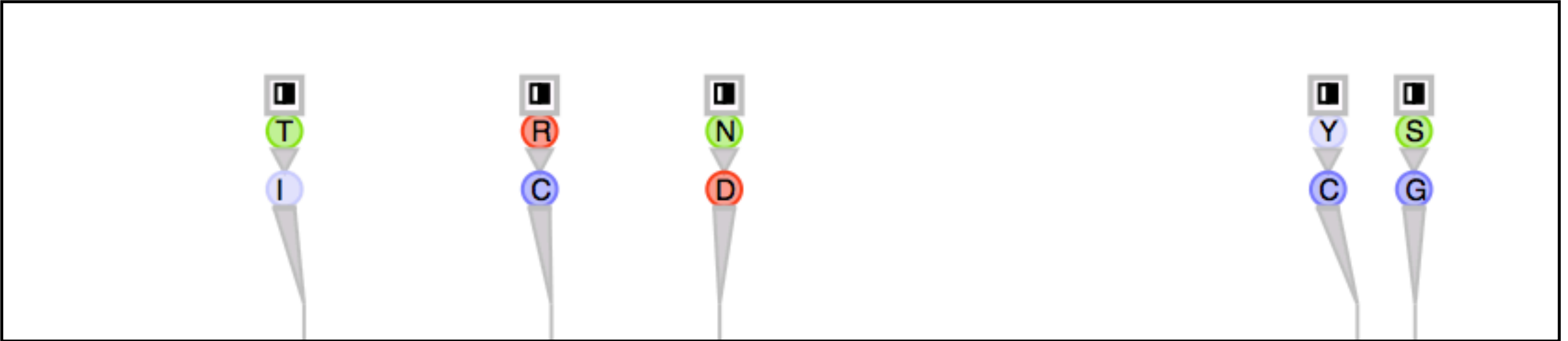
In contrast, low scoring gene



No collocation of variants

Variants

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



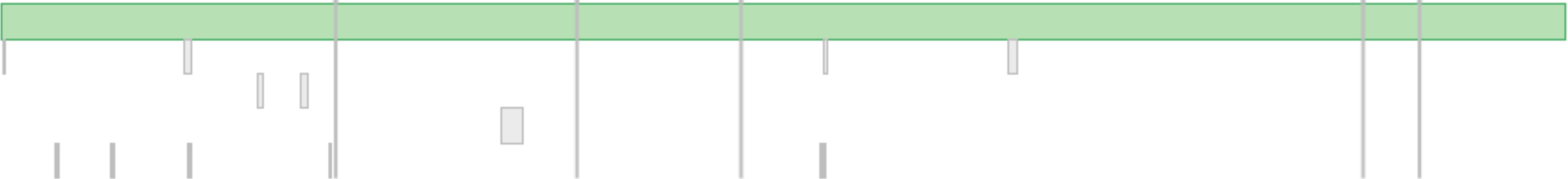
Transcript

trans-anon

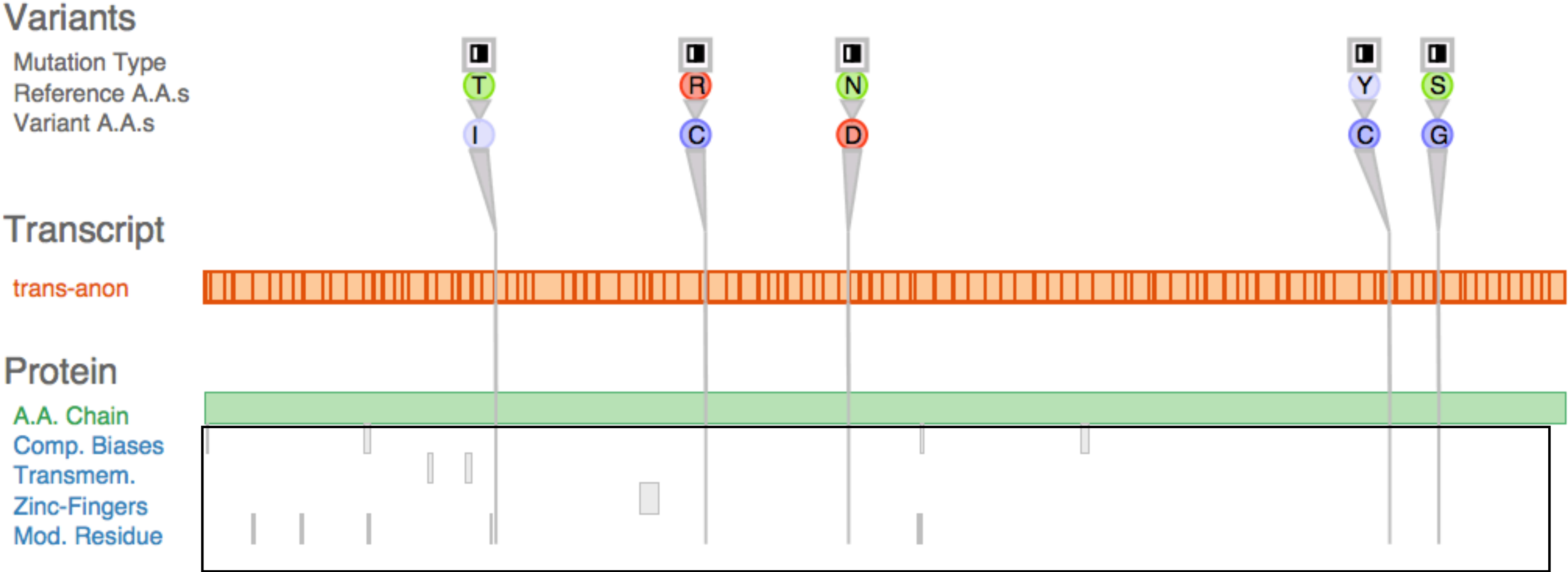


Protein

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



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 Genes:

Alternative Transcripts:

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

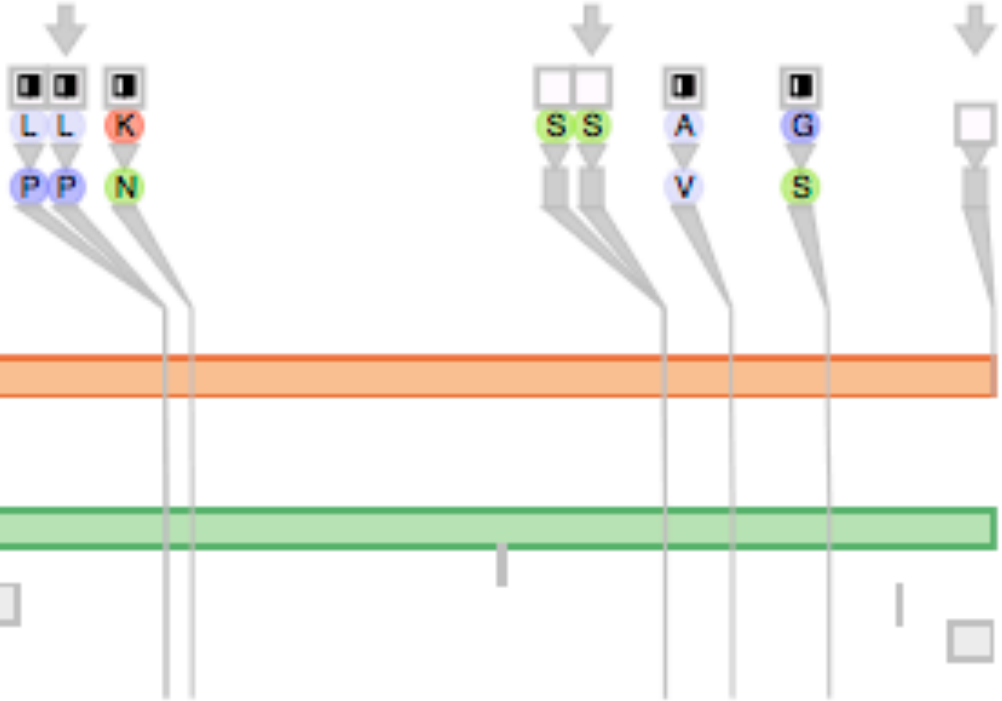
Select Patient:

Patient Genes:

Alternative Transcripts:

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

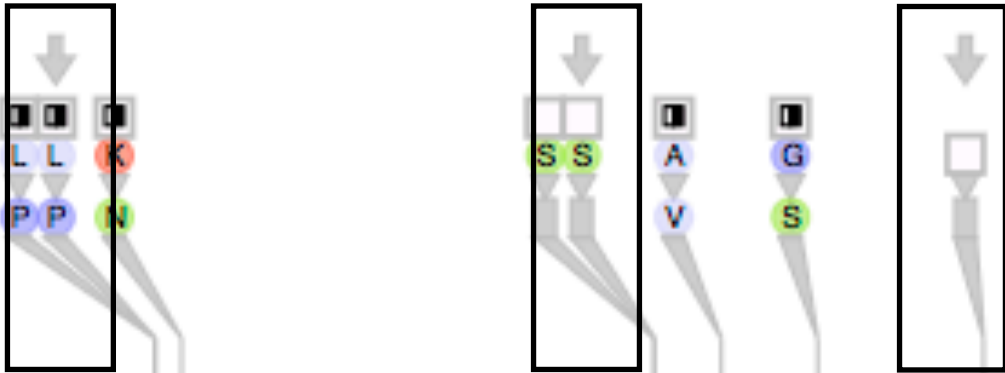
Select Patient:

Patient Genes:

Alternative Transcripts:

Variants

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



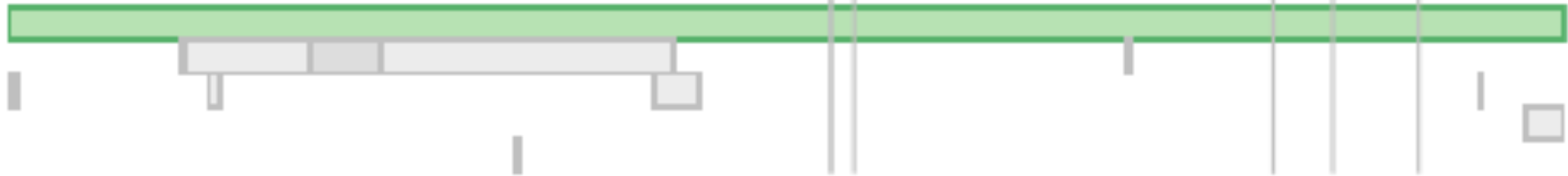
Transcript

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 Genes:

Alternative Transcripts:

Variants

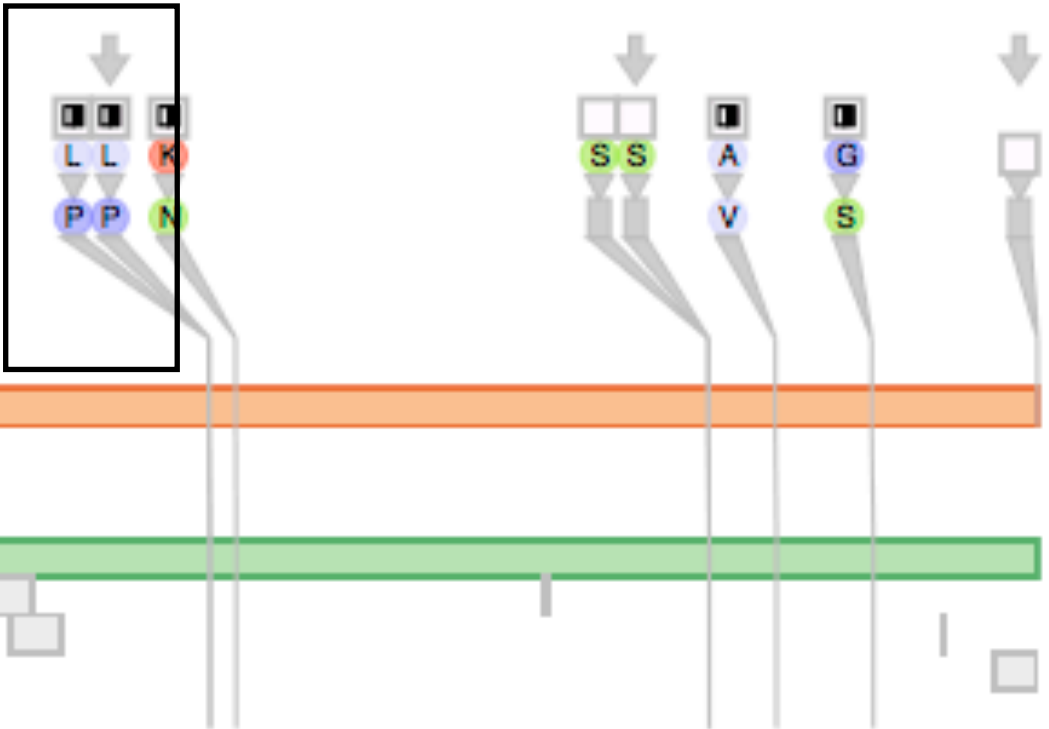
Mutation Type
Reference A.A.s
Variant A.A.s

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

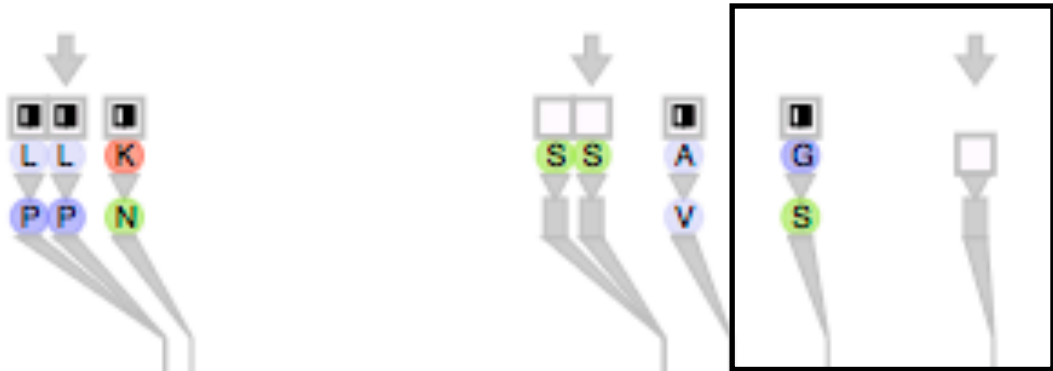
Select Patient:

Patient Genes:

Alternative Transcripts:

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

Variants

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



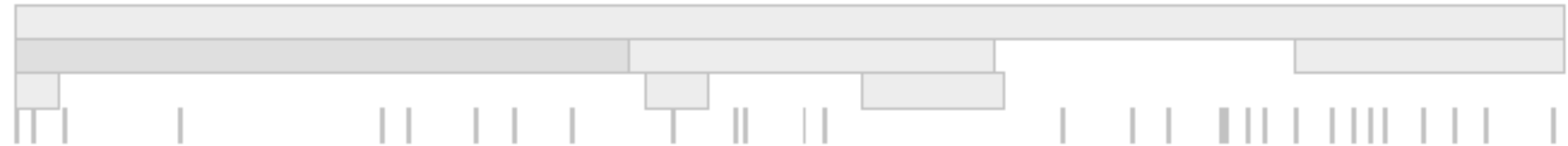
Transcript

trans-anon



Protein

A.A. Chain
Regions
Comp. Biases
Mod. Residue

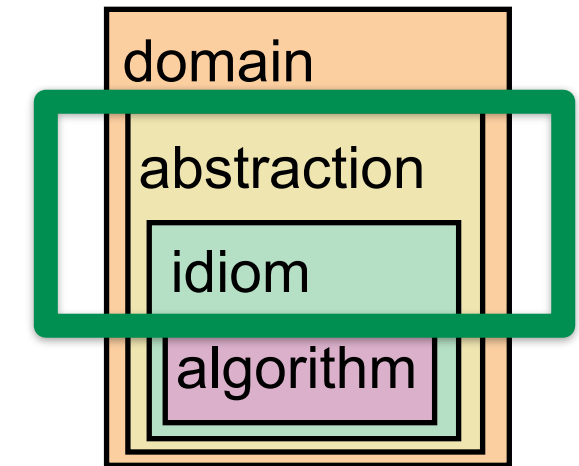


The tool exposed artifacts in the data that slid past at least two rounds of quality metric filtering ... this type of problem would not have been caught by our previous, automated methods.

- Analyst 3

Reduce from big data to manageable data

- at abstraction level, not algorithm level
 - filter away huge amounts of detail
 - transform from variant centric to gene centric



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

Privacy implications

- anonymizing

- protect patients: strip/change patient names
- protect researchers: strip/change gene names
- deanonymization threats: minor in this case

- future projects

- genomics meets clinical meets administrative meets demographic
 - enable data re-use: link individuals across datasets to create research study cohorts
 - health research opportunity: major
 - deanonymization threat: major
- ferociously privacy-protected data silos/vaults
 - can't see data without permission; what to ask for?
 - upcoming project: privacy preserving visual metadata browsing

Select Patient:

Patient Genes:


Alternative Transcripts:

Variants

| Mutation Type | Reference A.A.s | Variant A.A.s |
|---------------|-----------------|---------------|
| | L L | K |
| | P P | N |

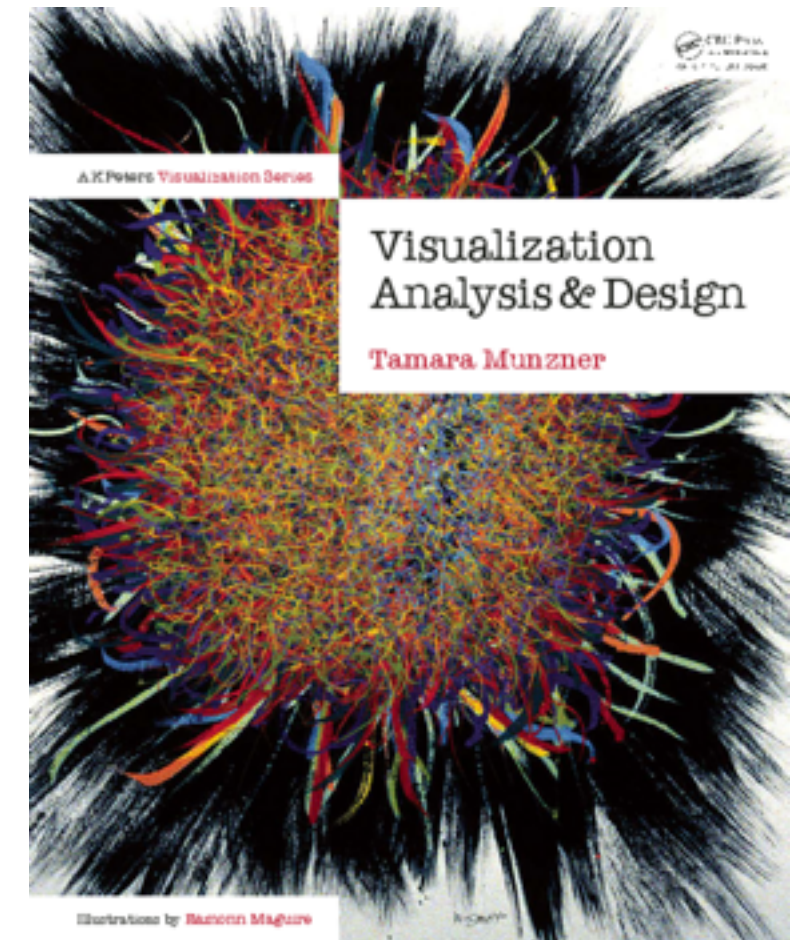
Transcript

trans-anon



More Information

- paper & open source download
<http://www.cs.ubc.ca/labs/imager/tr/2013/VariantView/>
- this talk
<http://www.cs.ubc.ca/~tmm/talks.html#think15>
- papers, videos, software, talks, courses
<http://www.cs.ubc.ca/group/infovis>
<http://www.cs.ubc.ca/~tmm>
- book: Visualization Analysis and Design. CRC Press 2014.
<http://www.cs.ubc.ca/~tmm/vadbook>
 - 20% promo code for book+ebook combo: HVN17
 - <http://www.crcpress.com/product/isbn/9781466508910>
- acknowledgements
 - funding: VIVA, Aeroinfo/Boeing, MITACS



[@tamaramunzner](https://twitter.com/tamaramunzner)