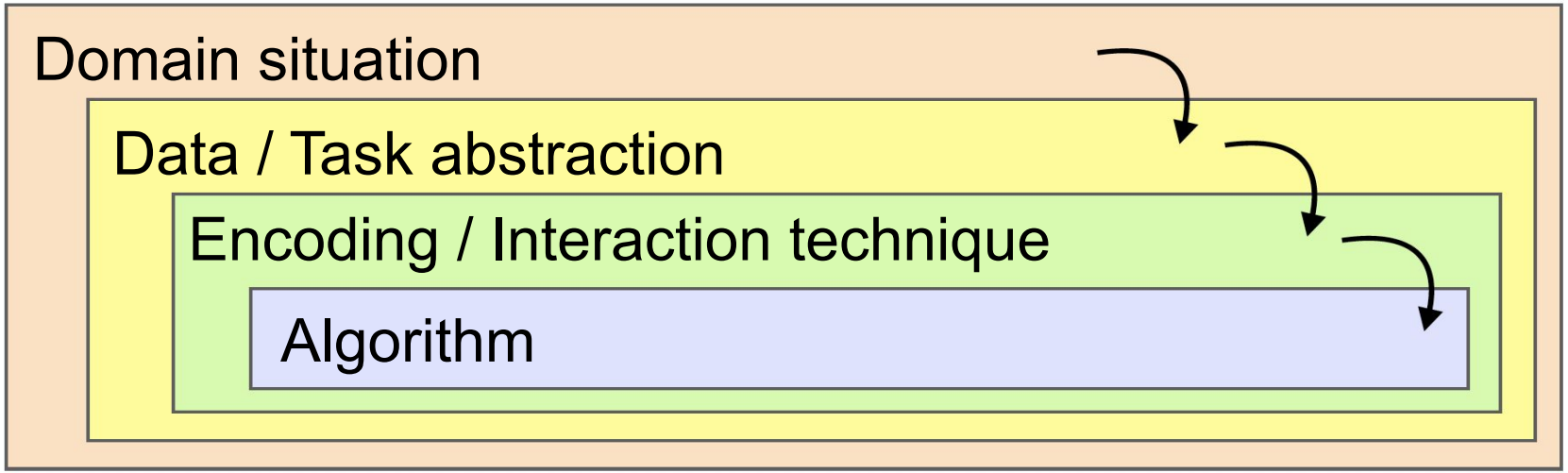
Visualization Design

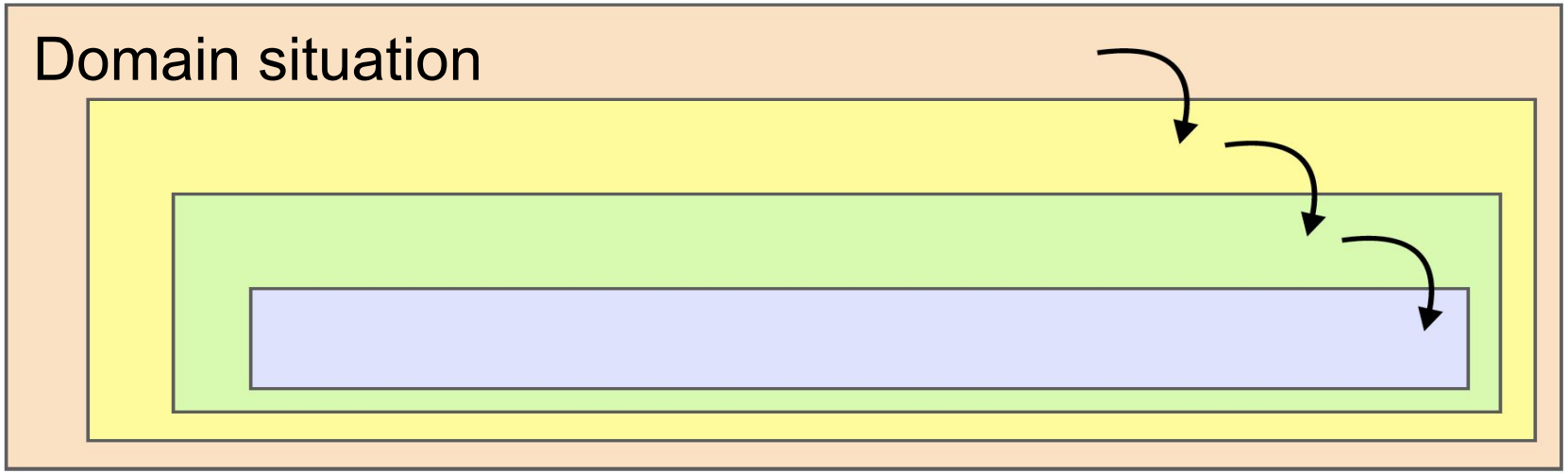
Chen He

Munzner, T. (2009). A nested process model for visualization design and validation. *IEEE TVCG*.

Nested Model for Visualization Design

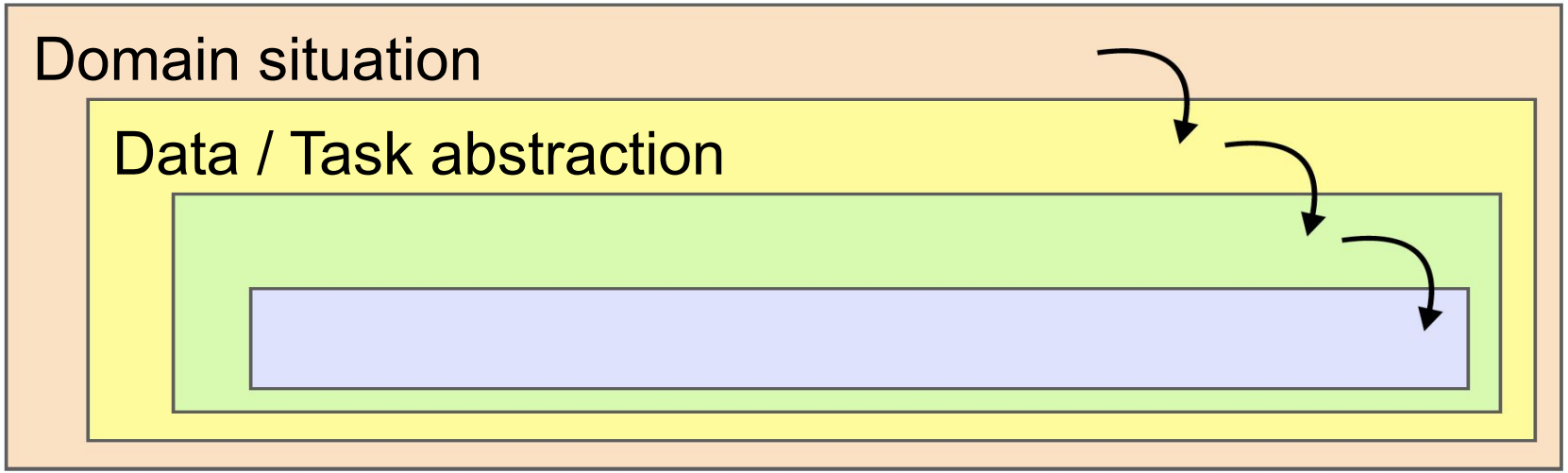


Domain situation



Characterize the **problems** and **data** of **target users** in some particular **target domain.**

Data / Task abstraction

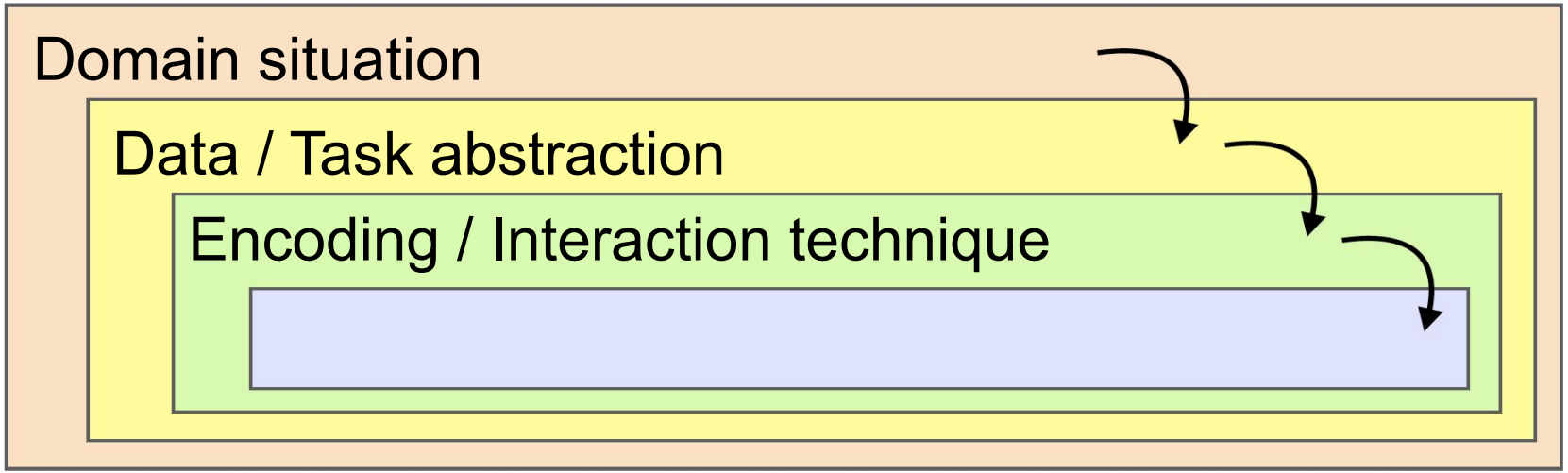


What type of data is shown? **Data abstraction**

Why is the user looking at it? **Abstraction of user tasks**

Abstract domain-specific problems and **data** into a more **generic** description that is in the vocabulary of computer science.

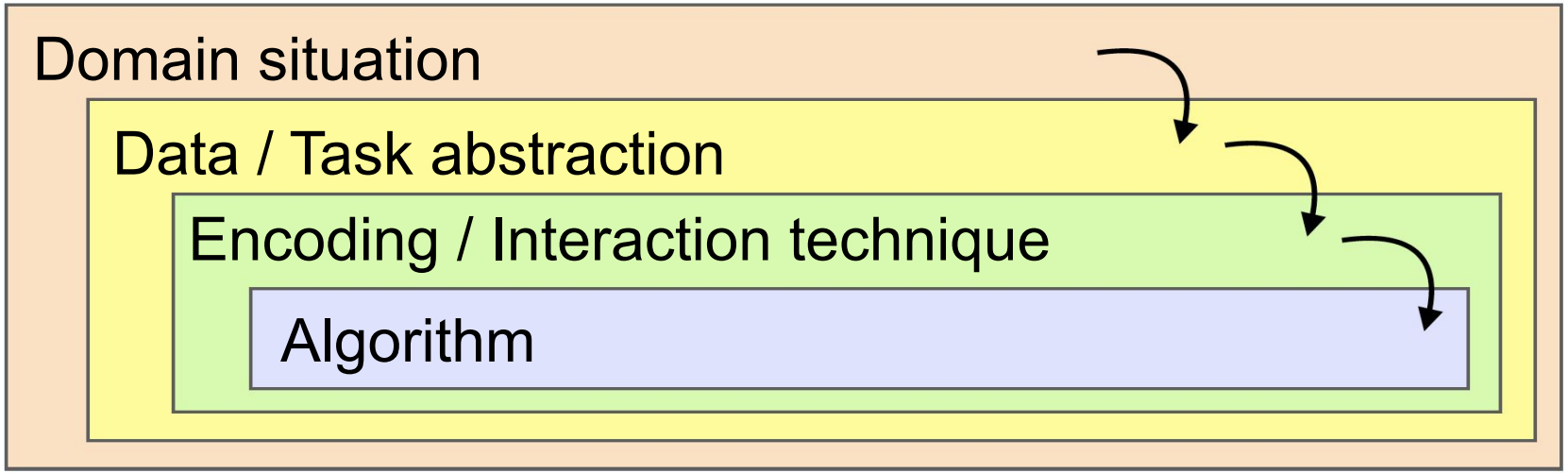
Encoding / Interaction technique



How is the data shown?

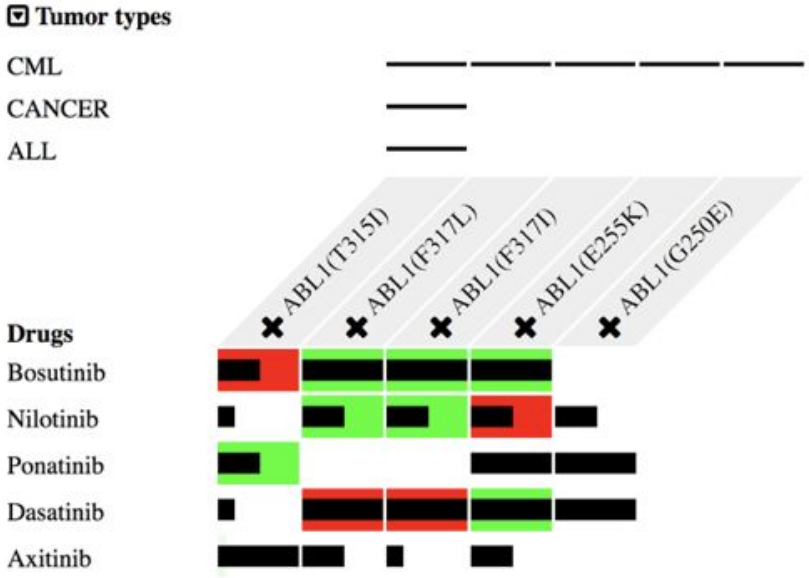
Decide on the specific way to **create** and **manipulate** the **visual representation** of the **abstraction**.

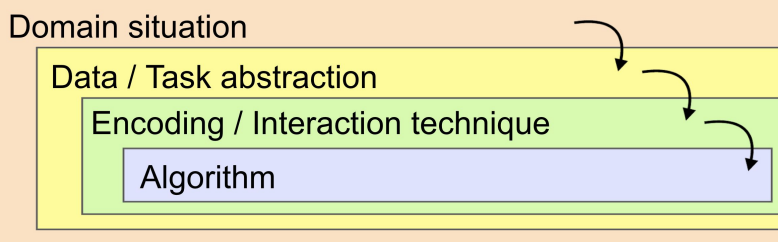
Algorithm



Crafting a detailed procedure that allows a computer to **automatically** and **efficiently** carry out **the desired visualization goal.**

Case study 1: Visualizing drug-target datasets





Domain situation

Biologists: **drug-target datasets** are usually **dispersed** in various sources, which hinders exploration.

CANCER GENOME INTERPRETER
(CGI)

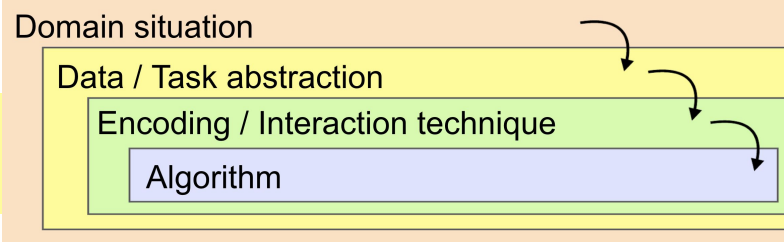
Biomarker	Drug	Effect	Evidence	Source	Curator	Tumor type
E255K						
ABL1 (T315A...	Bosutinib (BCR-...	Responsive	NCCN guidelines	PMID:21562040	RDientsmann	CML
ABL1 (F359V...	Dasatinib (BCR-...	Responsive	NCCN guidelines	PMID:21562040	RDientsmann	CML
ABL1 (I242T,...	Imatinib (BCR-A...	Resistant	European Leuk...	PMID:21562040	CRubio-Perez	CML
ABL1 (E255K...	Nilotinib (BCR-A...	Resistant	European Leuk...	PMID:21562040	CRubio-Perez	CML

DRUGTARGET COMMONS
(DTC)

Compo	Uniprc	Target Pref Name	Gene Name	Wild ty	Mutation ir	PubMed ID	End	En	End P.	E
PONATINIB			ABL1	mutated	ABL1(E255K)	23301703	IC50	=	1.08	nM
NILOTINIB			ABL1	mutated	ABL1(E255K)	23301703	IC50	=	27.8	nM
IMATINIB			ABL1	mutated	ABL1(E255K)	23301703	IC50	=	485.8	nM
DASATINIB			ABL1	mutated	ABL1(E255K)	23301703	IC50	=	0.21	nM
			ABL1, ABL2	mutated	ABL1(E255K)	23414803	ACTI...	=	100	%
			ABL1, ABL2	mutated	ABL1(E255K)	23414803	ACTI...	=	88	%
			ABL1	mutated	ABL1(E255K)	23301703	IC50	=	5.4	nM
			ABL1	mutated	ABL1(E255K)	23301703	IC50	=	2.9	nM
			ABL1	mutated	ABL1(E255K)	23301703	KD	=	0.28	nM
			ABL1	mutated	ABL1(E255K)	23301703	IC50	=	0.27	nM

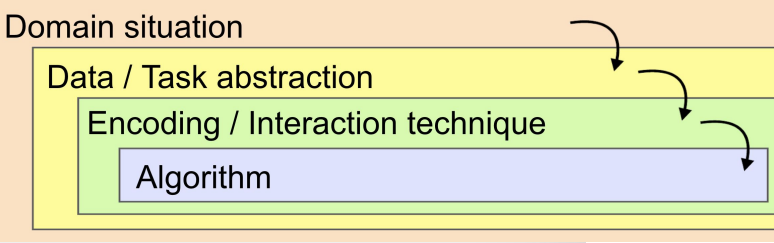
Sort Ascending
Sort Descending
Remove Sort
Show rows where:
contains
And
contains
Filter
Clear

Task abstraction



Integrate drug-target relations from two different sources.

Allow user **exploration** of drug-target relations.

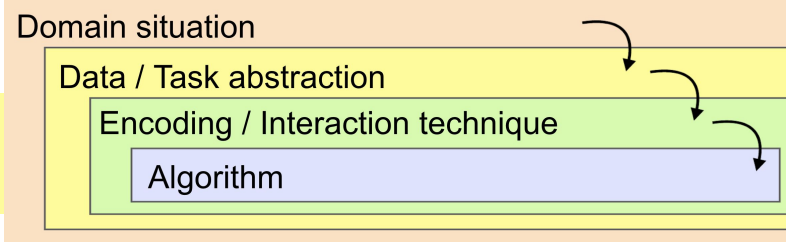


Data abstraction

	Discrete (no between values)	**Continuous** (values between)
Ordered (values are comparable)	**Ordinal,** e.g. size: S,M,L,XL,… **Quantitative,** e.g. counts: 1,2,3,…	**Fields,** e.g. altitude, temperature
Unordered (values not comparable)	**Nominal,** e.g. shape: ☐○△ **Categories,** e.g. nationality	**Cyclic values,** e.g. directions, hues

https://www.coursera.org/learn/datavisualization/lecture/ONoPE/2-1-1-data

Data abstraction



Data	Data type
Drug	Nominal
Mutation	Nominal
Tumor type	Nominal
Drug-target relation from CGI	
Effects	Nominal
Evidence level	Ordinal
Drug-target relation from DTC	
Potency	Quantitative

CANCER GENOME INTERPRETER

Biomarker	Drug	Effect	Evidence	Source	Curator	Tumor type
E255K						
ABL1 (T315A...	Bosutinib (BCR-...	Responsive	NCCN guidelines	PMID:21562040	RDientsmann	CML
ABL1 (F359V...	Dasatinib (BCR-...	Responsive	NCCN guidelines	PMID:21562040	RDientsmann	CML
ABL1 (I242T,...	Imatinib (BCR-A...	Resistant	European Leuk...	PMID:21562040	CRubio-Perez	CML
ABL1 (E255K...	Nilotinib (BCR-A...	Resistant	European Leuk...	PMID:21562040	CRubio-Perez	CML

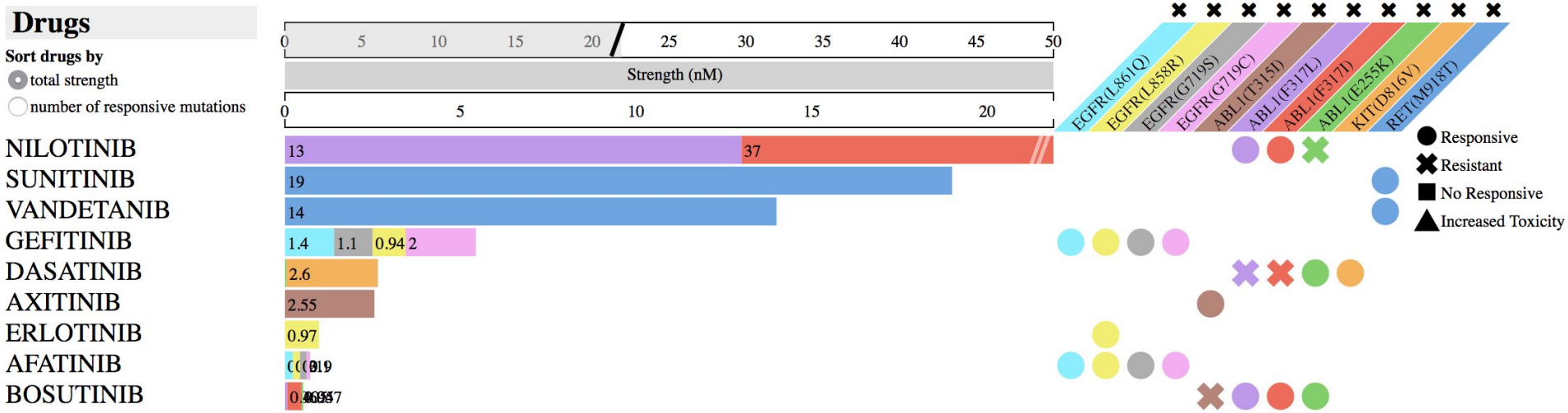
DRUGTARGET COMMONS

Compo	Unipro	Target Pref Name	Gene Name	Wild ty	Mutation in	PubMed ID	End	En	End P.	E
PONATINIB			ABL1	mutated	ABL1(E255K)	23301703	IC50	=	1.08	nM
NILOTINIB			ABL1	mutated	ABL1(E255K)	23301703	IC50	=	27.8	nM
IMATINIB			ABL1	mutated	ABL1(E255K)	23301703	IC50	=	485.8	nM
DASATINIB			ABL1	mutated	ABL1(E255K)	23301703	IC50	=	0.21	nM
			ABL1, ABL2	mutated	ABL1(E255K)	23414803	ACTI...	=	100	%
			ABL1, ABL2	mutated	ABL1(E255K)	23414803	ACTI...	=	88	%
			ABL1	mutated	ABL1(E255K)	23301703	IC50	=	5.4	nM
			ABL1	mutated	ABL1(E255K)	23301703	IC50	=	2.9	nM
			ABL1	mutated	ABL1(E255K)	23301703	KD	=	0.28	nM
			ABL1	mutated	ABL1(E255K)	23301703	IC50	=	0.27	nM

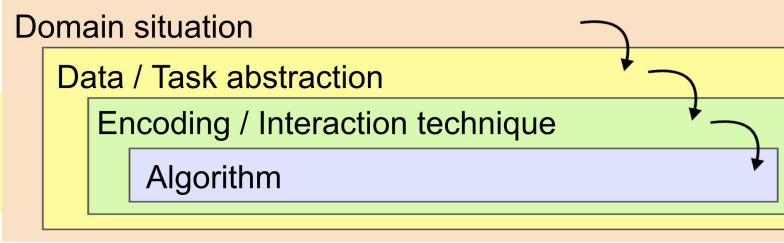
Sort Ascending
Sort Descending
Remove Sort
Show rows where:
contains
And
contains
Filter
Clear

Prototype version Zero

Quantitative	Ordinal	Nominal
Position	Position	Position
Length	Density	Hue
Angle	Saturation	Density
Slope	Hue	Saturation
Area	Length	Shape
Density	Angle	Length
Saturation	Slope	Angle
Hue	Area	Slope
Shape	Shape	Area



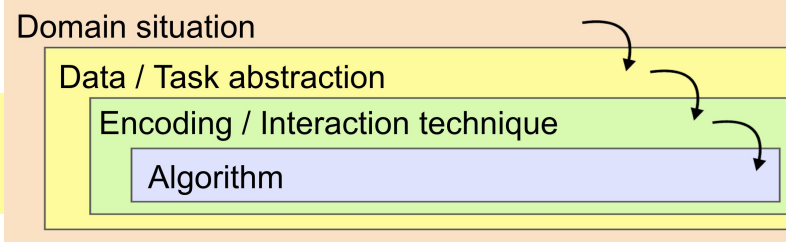
Task abstraction



~~**Integrate**~~ **Compare** drug-target relations from two different sources.

Allow user **exploration** of drug-target relations.

Data abstraction



Data	Data type
Drug	Nominal
Mutation	Nominal
Tumor type	Nominal
Drug-target relation from CGI	
Effects	Nominal
Evidence level	Ordinal
Drug-target relation from DTC	
Potency level	Ordinal

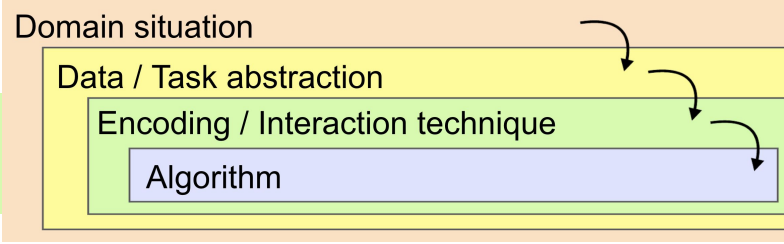
CANCER GENOME INTERPRETER

Biomarker	Drug	Effect	Evidence	Source	Curator	Tumor type
E255K						
ABL1 (T315A...	Bosutinib (BCR-...	Responsive	NCCN guidelines	PMID:21562040	RDientsmann	CML
ABL1 (F359V...	Dasatinib (BCR-...	Responsive	NCCN guidelines	PMID:21562040	RDientsmann	CML
ABL1 (I242T,...	Imatinib (BCR-A...	Resistant	European Leuk...	PMID:21562040	CRubio-Perez	CML
ABL1 (E255K...	Nilotinib (BCR-A...	Resistant	European Leuk...	PMID:21562040	CRubio-Perez	CML

DRUGTARGET COMMONS

Compo	Unipro	Target Pref Name	Gene Name	Wild ty	Mutation in	PubMed ID	End	En	End P.	E
PONATINIB			ABL1	mutated	ABL1(E255K)	23301703	IC50	=	1.08	nM
NILOTINIB			ABL1	mutated	ABL1(E255K)	23301703	IC50	=	27.8	nM
IMATINIB			ABL1	mutated	ABL1(E255K)	23301703	IC50	=	485.8	nM
DASATINIB			ABL1	mutated	ABL1(E255K)	23301703	IC50	=	0.21	nM
			ABL1, ABL2	mutated	ABL1(E255K)	23414803	ACTI...	=	100	%
			ABL1, ABL2	mutated	ABL1(E255K)	23414803	ACTI...	=	88	%
			ABL1	mutated	ABL1(E255K)	23301703	IC50	=	5.4	nM
			ABL1	mutated	ABL1(E255K)	23301703	IC50	=	2.9	nM
			ABL1	mutated	ABL1(E255K)	23301703	KD	=	0.28	nM
			ABL1	mutated	ABL1(E255K)	23301703	IC50	=	0.27	nM

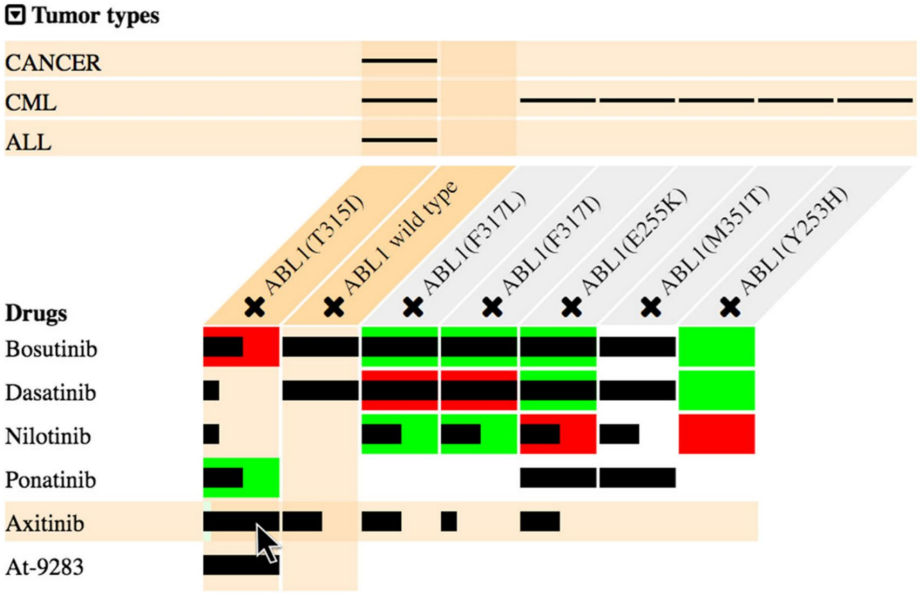
Sort Ascending
Sort Descending
Remove Sort
Show rows where:
contains
And
contains
Filter
Clear



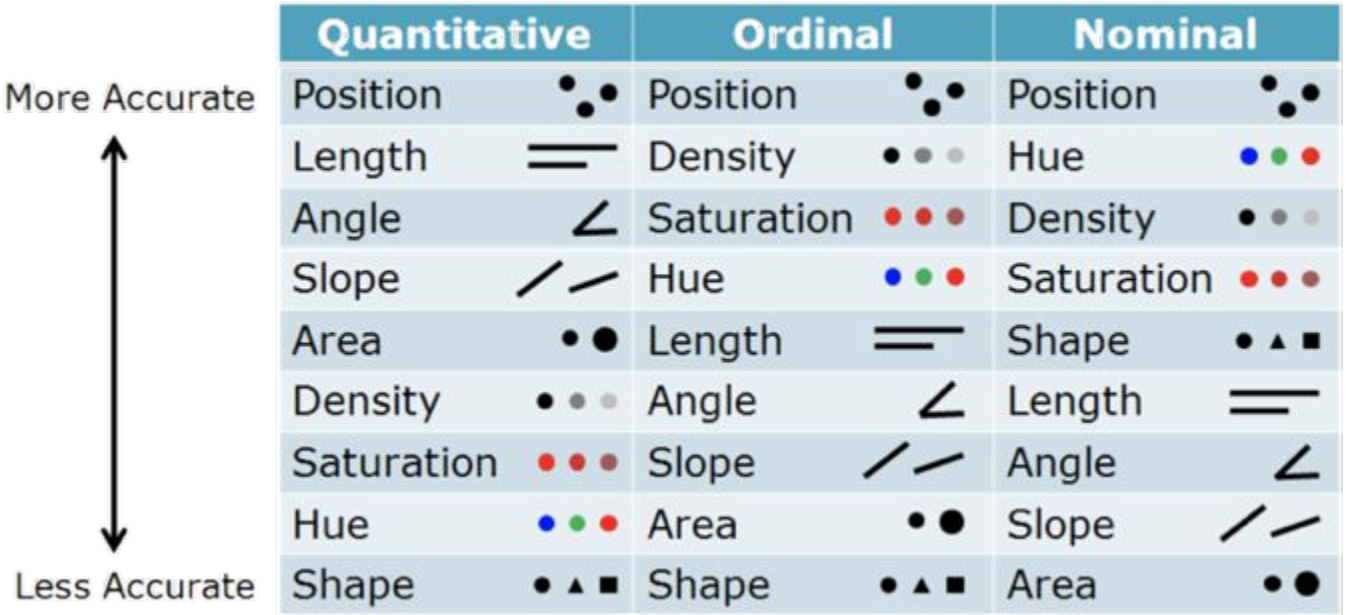
Visual encoding - Layout

Matrix-based layout: **scalable** numbers of drugs and targets;

Overlaid layers: facilitate the **comparison** of data from multiple sources.



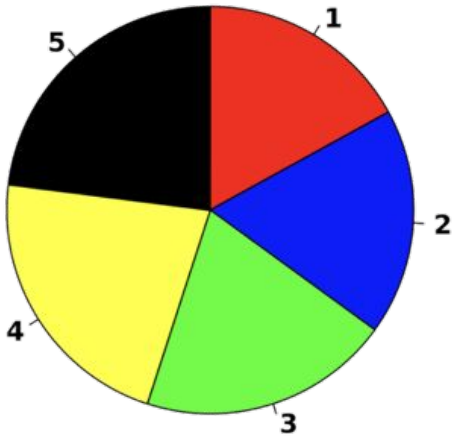
Visual encoding

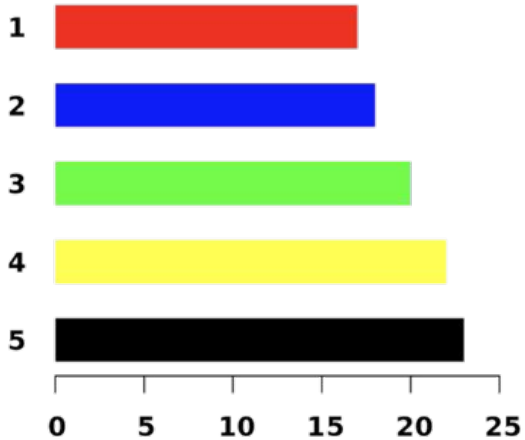


J. Mackinlay, Automating the Design of Graphical Presentations of Relational Information, ACM Transactions on Graphics 5(2), 1986.

Visual encoding

Angle & Area < Position & Length

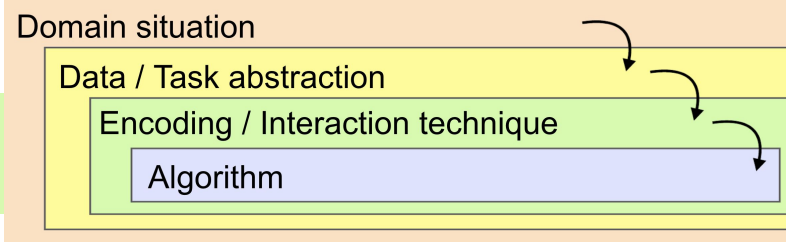




More Accurate

Less Accurate

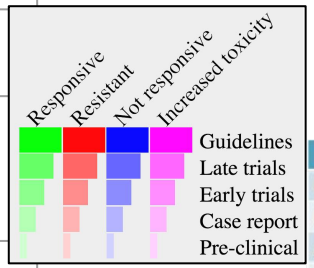
Quantitative	Ordinal	Nominal
Position	Position	Position
Length	Density	Hue
Angle	Saturation	Density
Slope	Hue	Saturation
Area	Length	Shape
Density	Angle	Length
Saturation	Slope	Angle
Hue	Area	Slope
Shape	Shape	Area

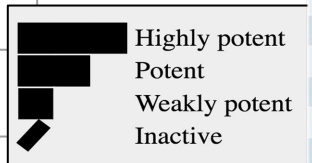


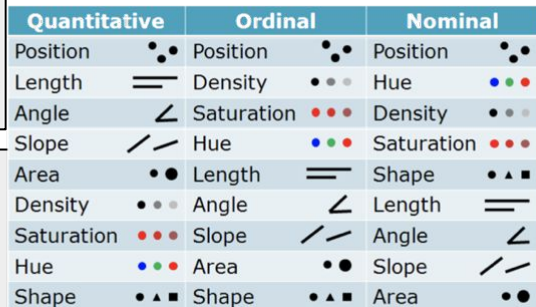
Visual encoding

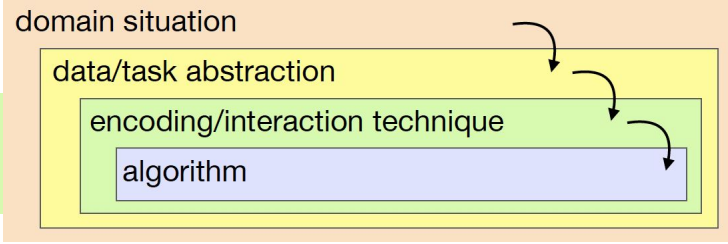
Data	Data type	Visual variable
Drug	Nominal	Position
Mutation	Nominal	Position
Tumor type	Nominal	Position
Drug-target relation from CGI		
Effects	Nominal	Hue
Evidence level	Ordinal	Position, length, saturation
Drug-target relation from DTC		
Potency level	Ordinal	Position, length

Matrix layout



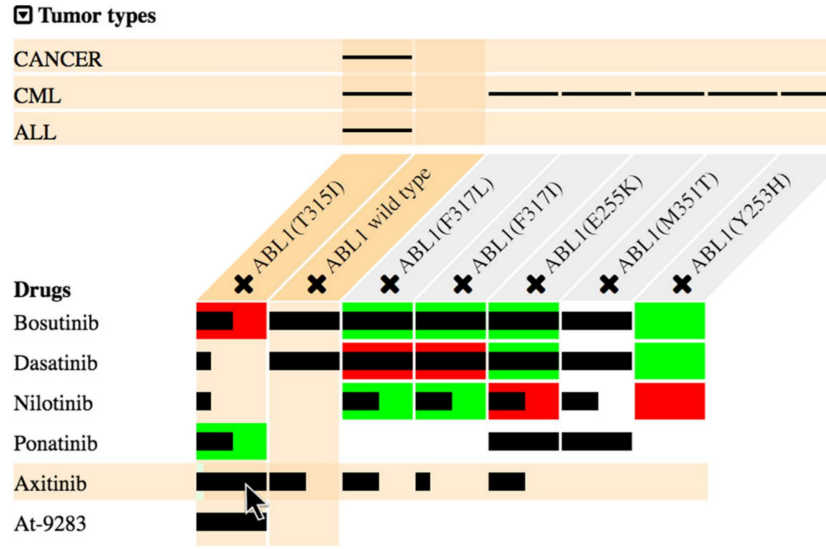






Visual encoding

Data	Data type	Visual variable
Drug	Nominal	Position
Mutation	Nominal	Position
Tumor type	Nominal	Position
Drug-target relation from CGI		
Effects	Nominal	Hue
Evidence level	Ordinal	Position, length, saturation
Drug-target relation from DTC		
Potency level	Ordinal	Position, length



https://www.youtube.com/watch?v=Bg_YvhBs1sg

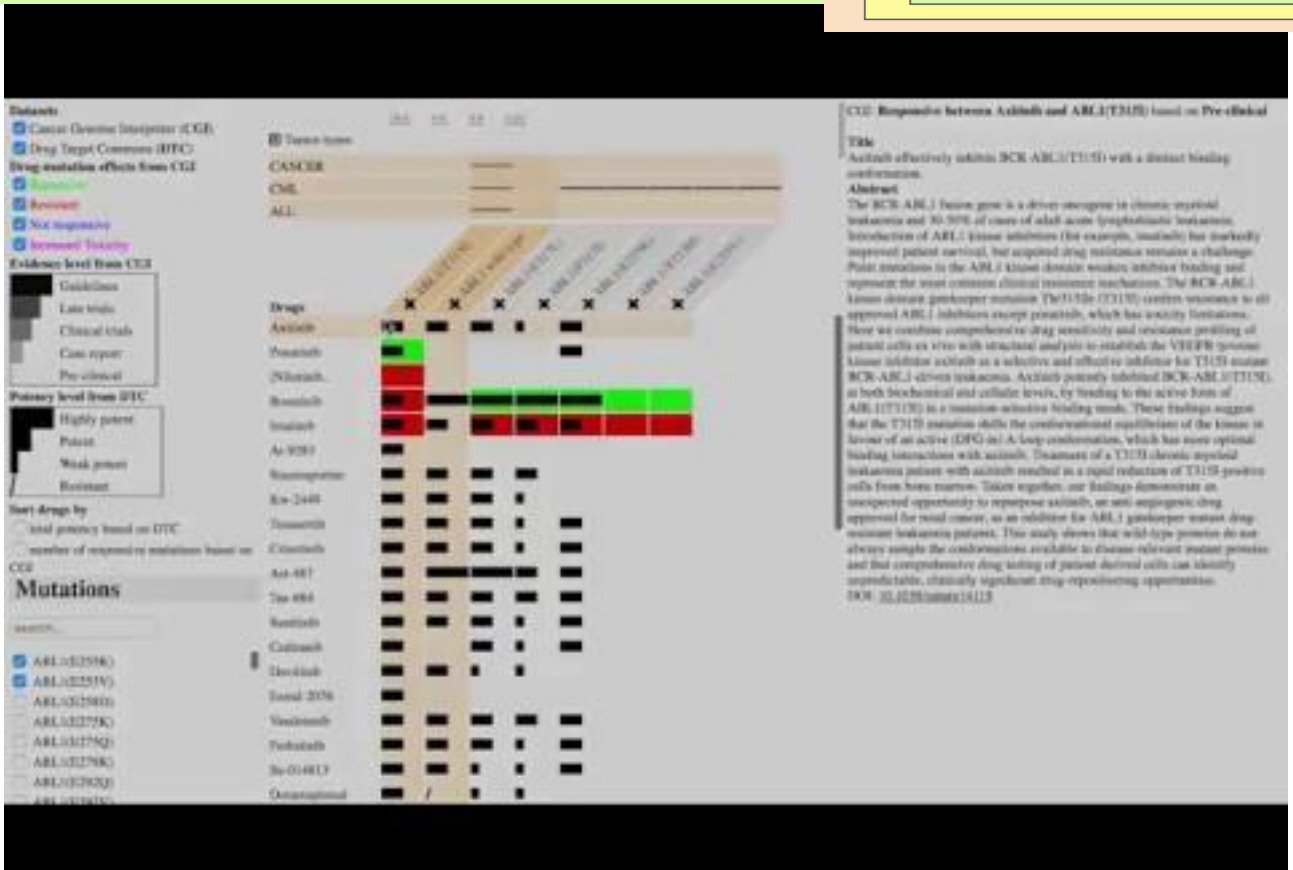
Domain situation

Data / Task abstraction

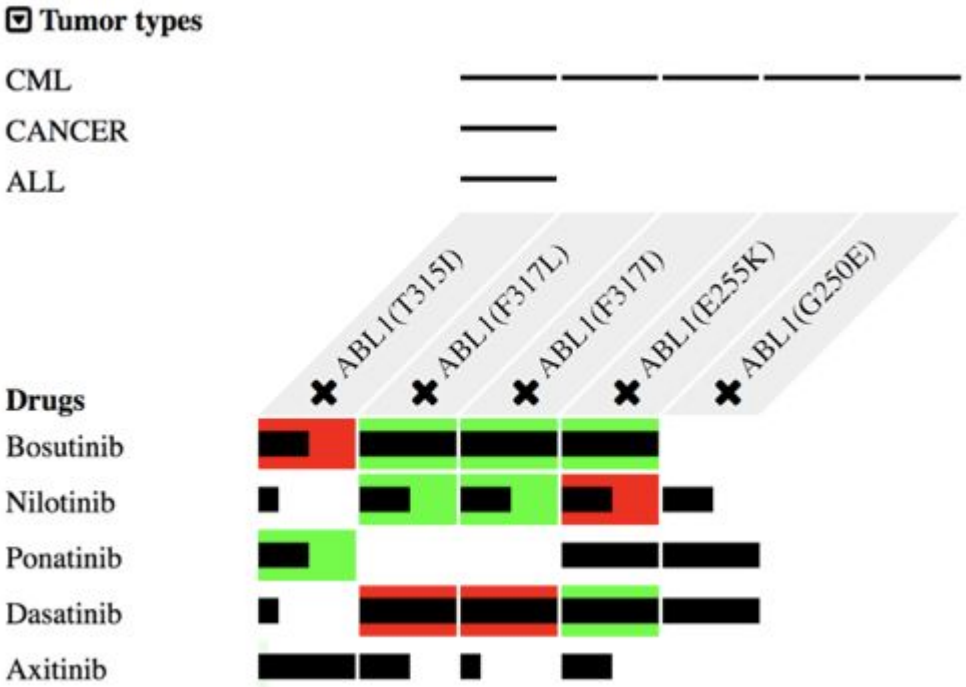
Encoding / Interaction technique

Algorithm

Interaction technique

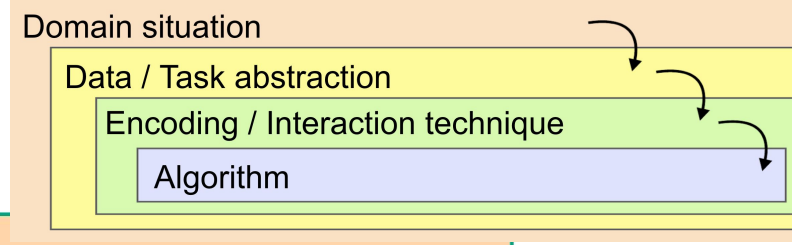


New discovery - Inconsistency Exposure



He, C., Micallef, L., Kaski, S., Aittokallio, T. and Jacucci, G., 2017. MediSyn: uncertainty-aware visualization of multiple biomedical datasets to support drug treatment selection. *BMC bioinformatics*.

Nested Model for Validation



threat: wrong problem
validate: observe and interview target users
threat: bad data/operation abstraction
threat: ineffective encoding/interaction technique
validate: justify encoding/interaction design
threat: slow algorithm
validate: analyze computational complexity
implement system
validate: measure system time/memory
validate: qualitative/quantitative result image analysis
[test on any users, informal usability study]
validate: lab study, measure human time/errors for operation
validate: test on target users, collect anecdotal evidence of utility
validate: field study, document human usage of deployed system
validate: observe adoption rates

Validation - Lab study

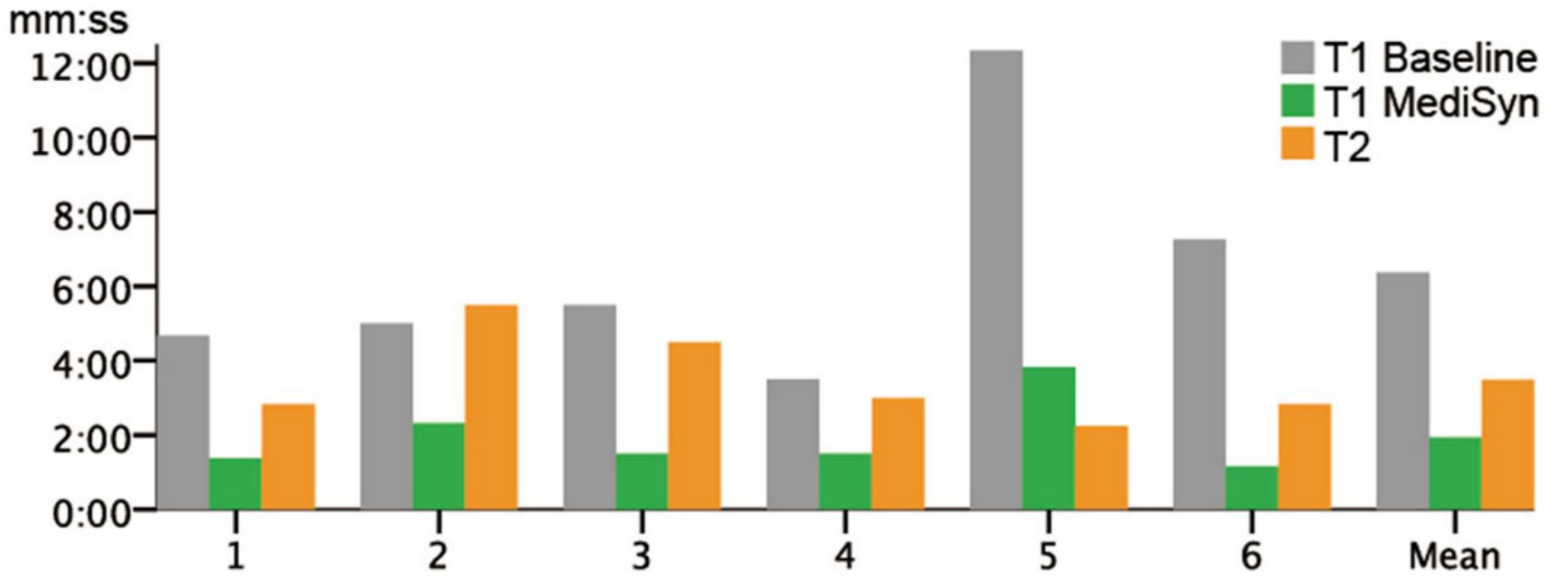
Baseline: two unlinked datasets

Participants: 6 domain experts

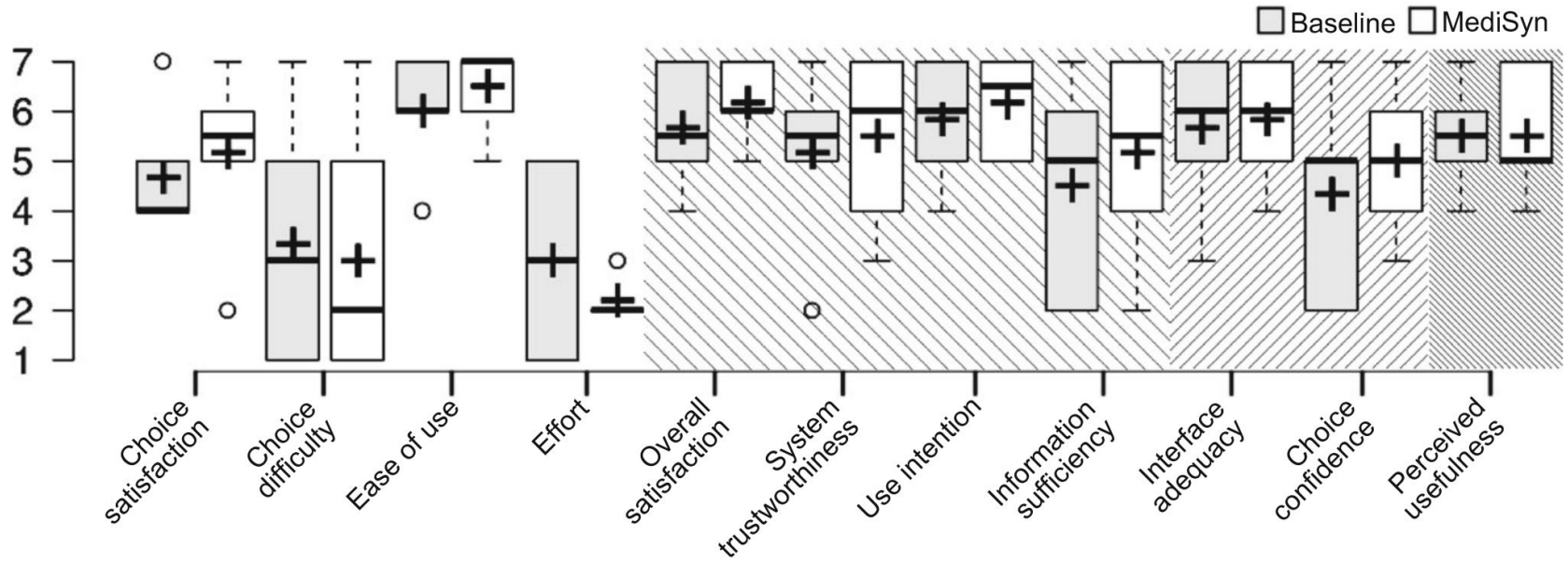
Tasks: 1. drug selection; 2. inconsistency discovery

Measure: task performance, subjective feedback

Results - Lab study



Results - Lab study

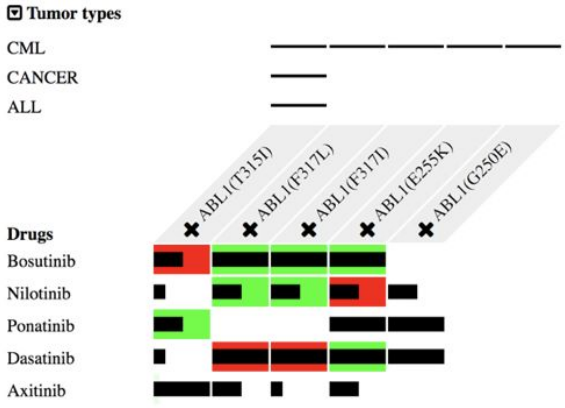


Findings - Lab study

Matrix view supports **drug comparison** and exposes missing data.

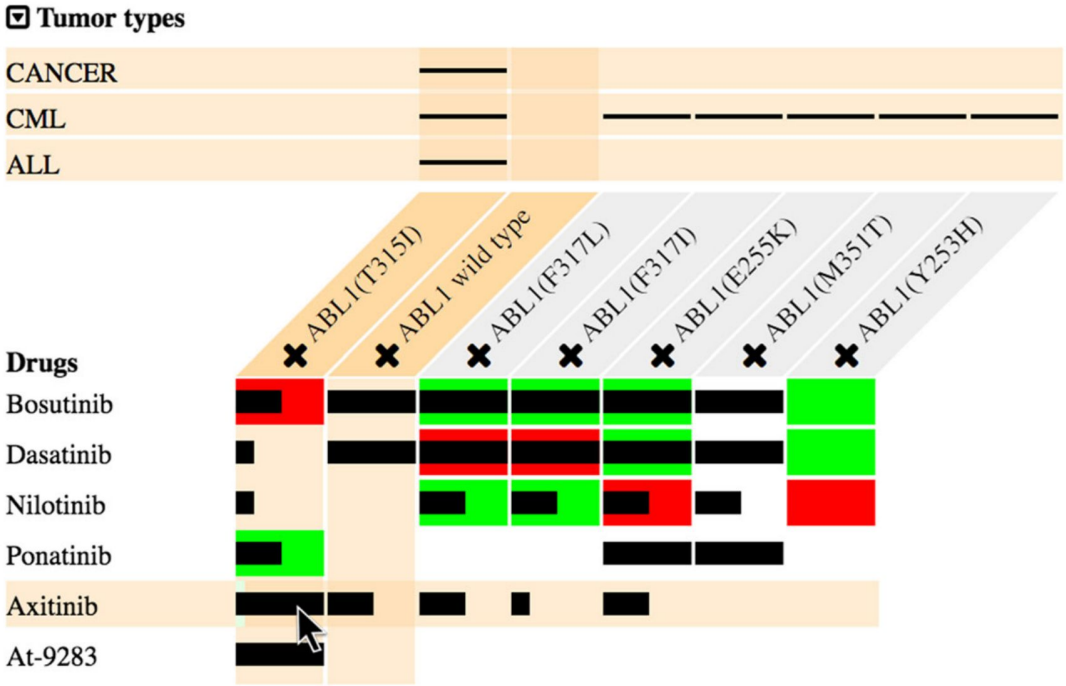
Depiction of datasets in **overlaid layers** facilitates **direct comparison of data from multiple sources**. (Data consistency)

Exposed data conflicts tend to **lower user trust** in MediSyn but do not have observable effects of user trust in data.



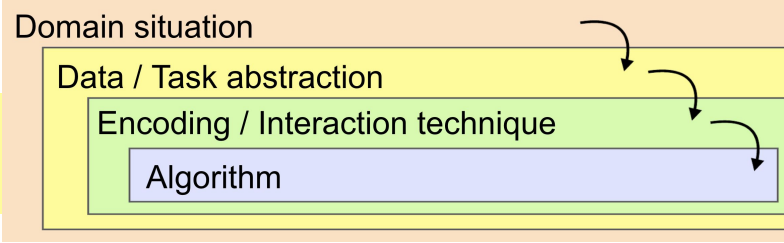
Validation - Test on target users, collect anecdotal evidence of utility

Discovery of drug repurposing opportunity.



- threat: wrong problem
- validate: observe and interview target users
- threat: bad data/operation abstraction
- threat: ineffective encoding/interaction technique
- validate: justify encoding/interaction design
- threat: slow algorithm
- validate: analyze computational complexity
- implement system
- validate: measure system time/memory
- validate: qualitative/quantitative result image analysis
- [test on any users, informal usability study]
- validate: lab study, measure human time/errors for operation
- validate: test on target users, collect anecdotal evidence of utility
- validate: field study, document human usage of deployed system
- validate: observe adoption rates

Iteration: Task abstraction



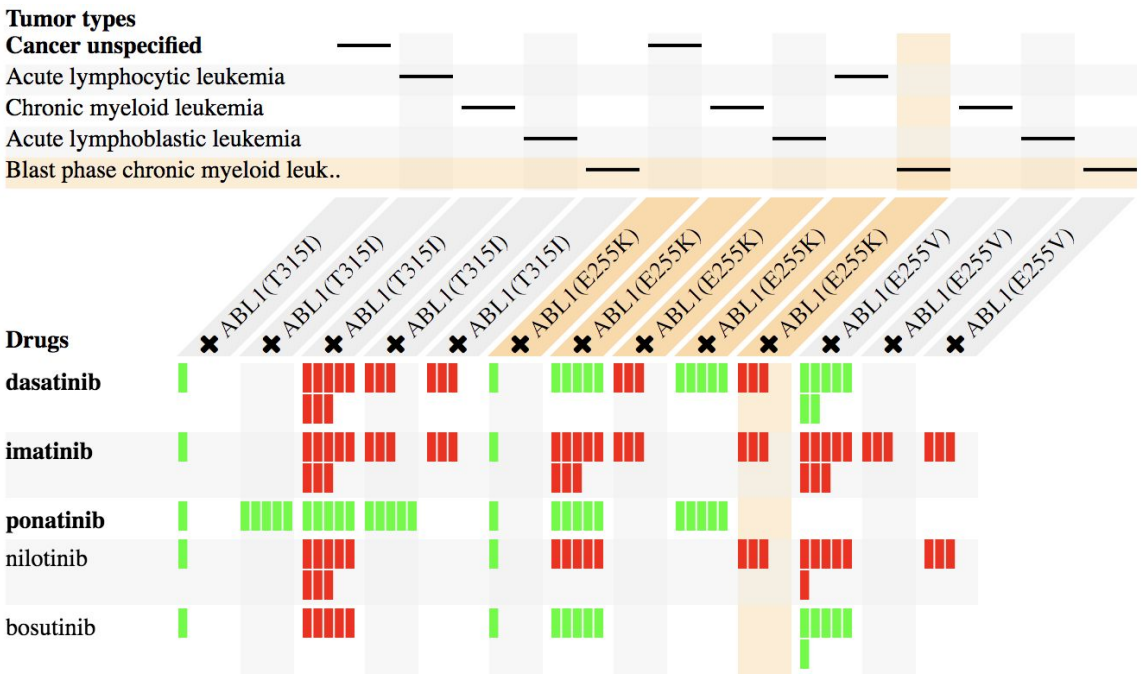
Compare drug-target relations from **more than two sources**.

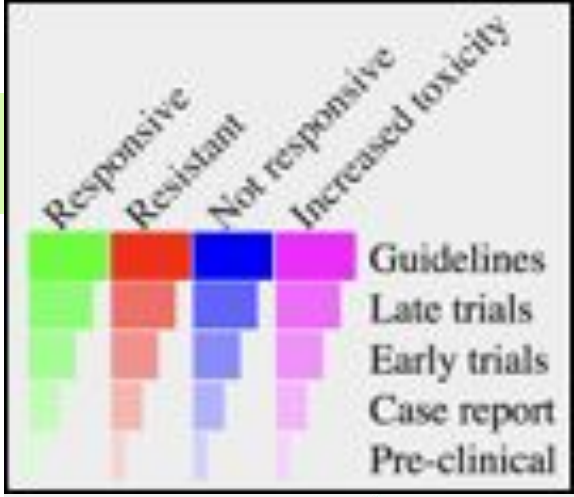
Allow user exploration of drug-target relations.

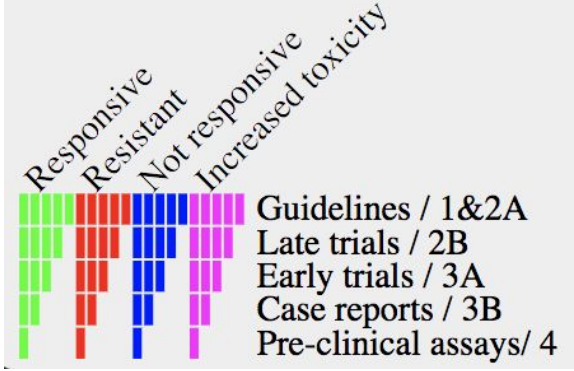
Support **insight recording and sharing**.

Iteration: Visual encoding

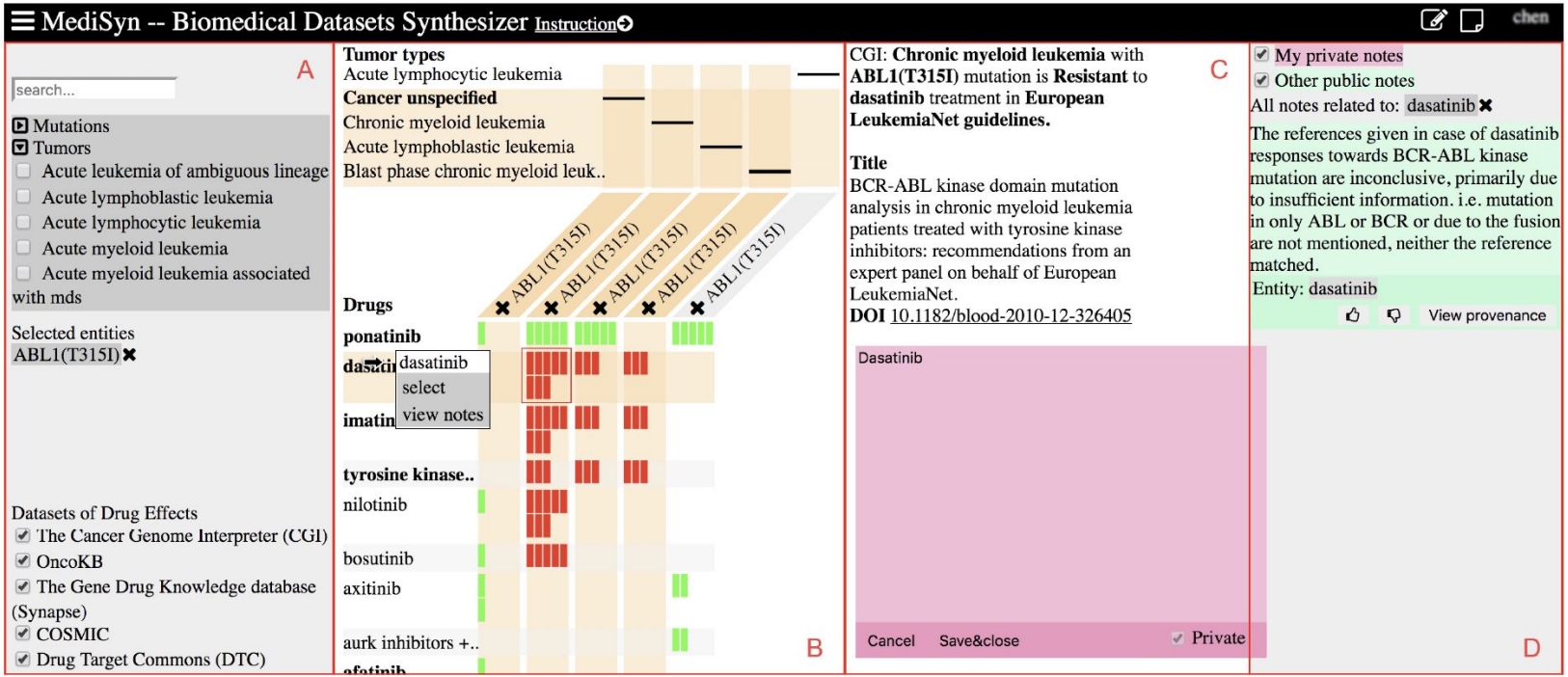
Five datasets - Juxtaposed bars





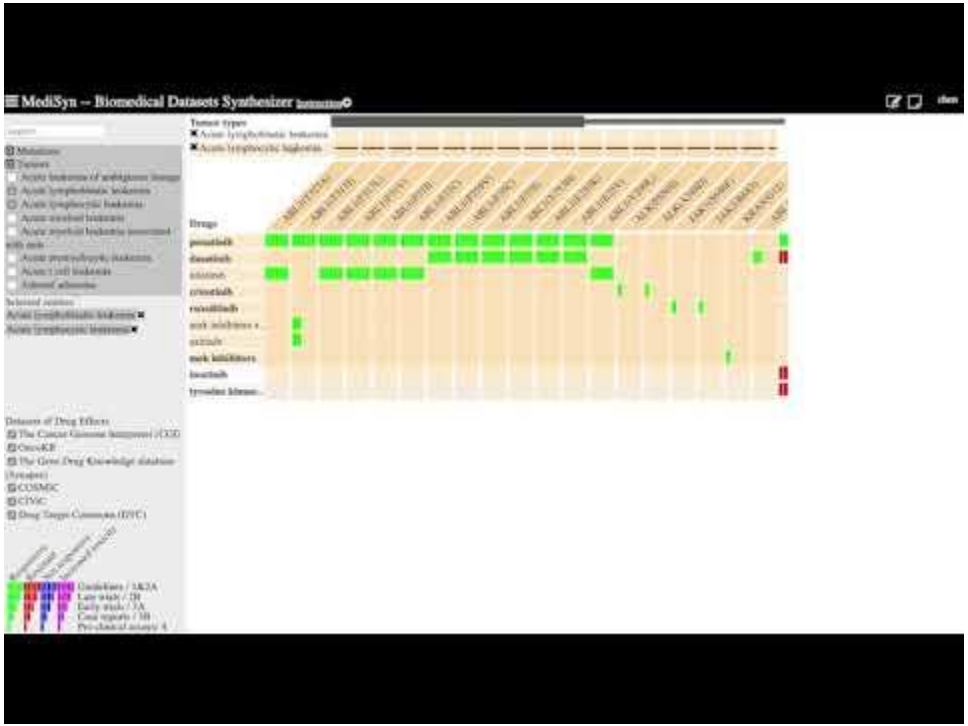


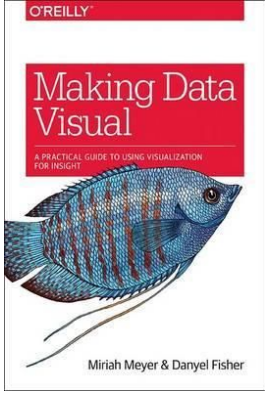
Iteration: Interaction technique



Interaction technique -- Entity-Based Interaction

Select
Connect
Elaborate
Explore
Insight-sharing





Case Study 2: Visualizing Biological Data

From Miriah Meyer, Danyel Fisher. **Making Data Visual: A Practical Guide to Using Visualization for Insight**. O'Reilly Media, 2018.

Domain situation

How genes influence physical features of animals?

Biologists study a set of **fundamental genes** that are shared across many species, and control the development of body parts in developing embryos.

They are nearly the same in many species, and yet these species are physically very different.

Domain situation

What is known: Differences between species are related to **when and where** (in which cells) these genes are turned on and off in developing embryos. -- Gene expression

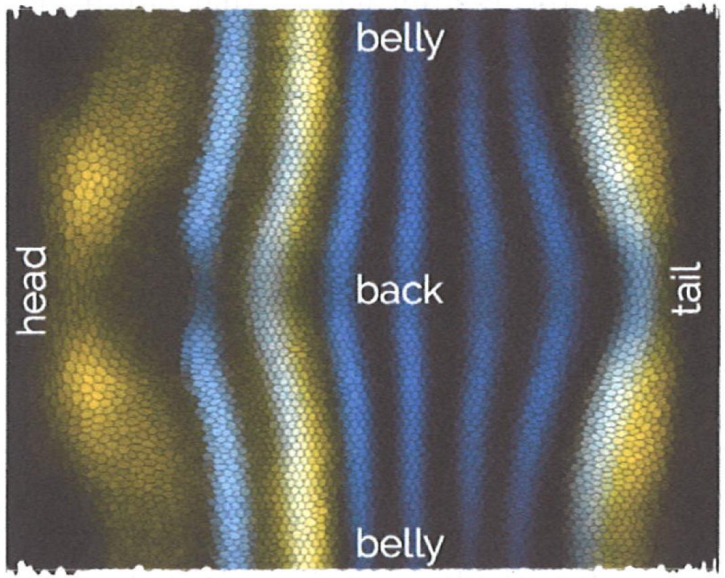
Goal:

Link the differences in gene expression to the differences in physical traits.

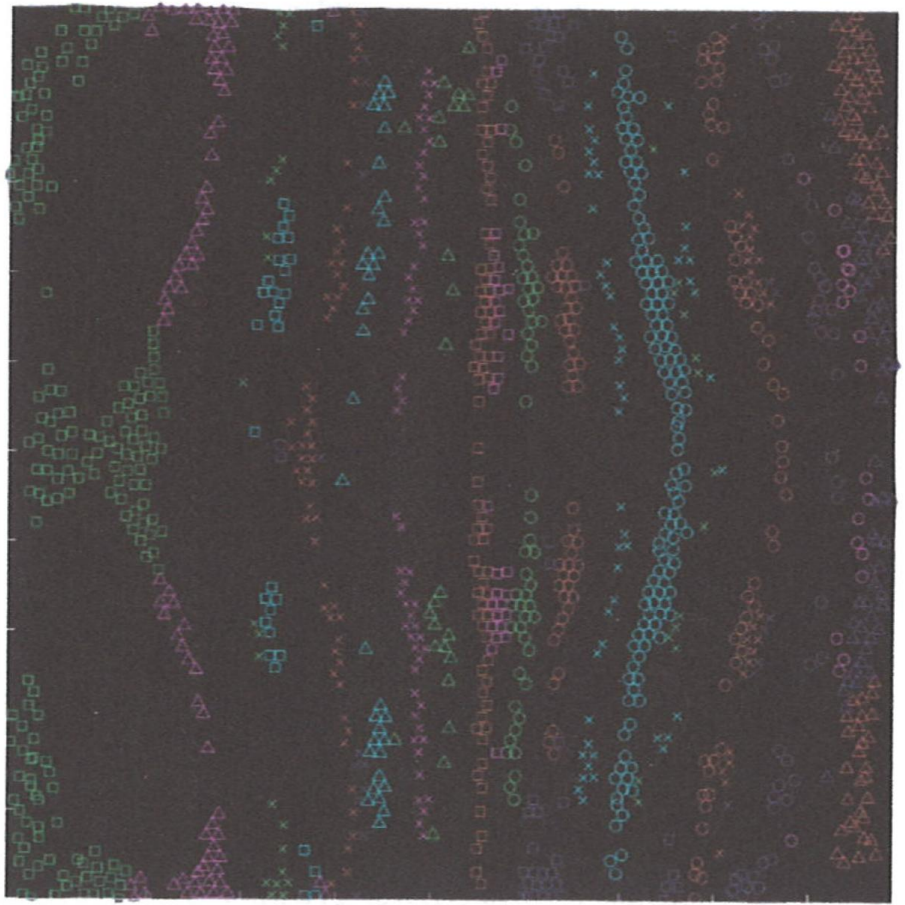
Existing task and tool

Task 1: Find cells in one embryo that had significantly different gene expression from cells in another embryo. -- Outlier cells

2D representation of a fruit fly embryo.



Outlier cells are clustered by color and shape.



Existing task and tool

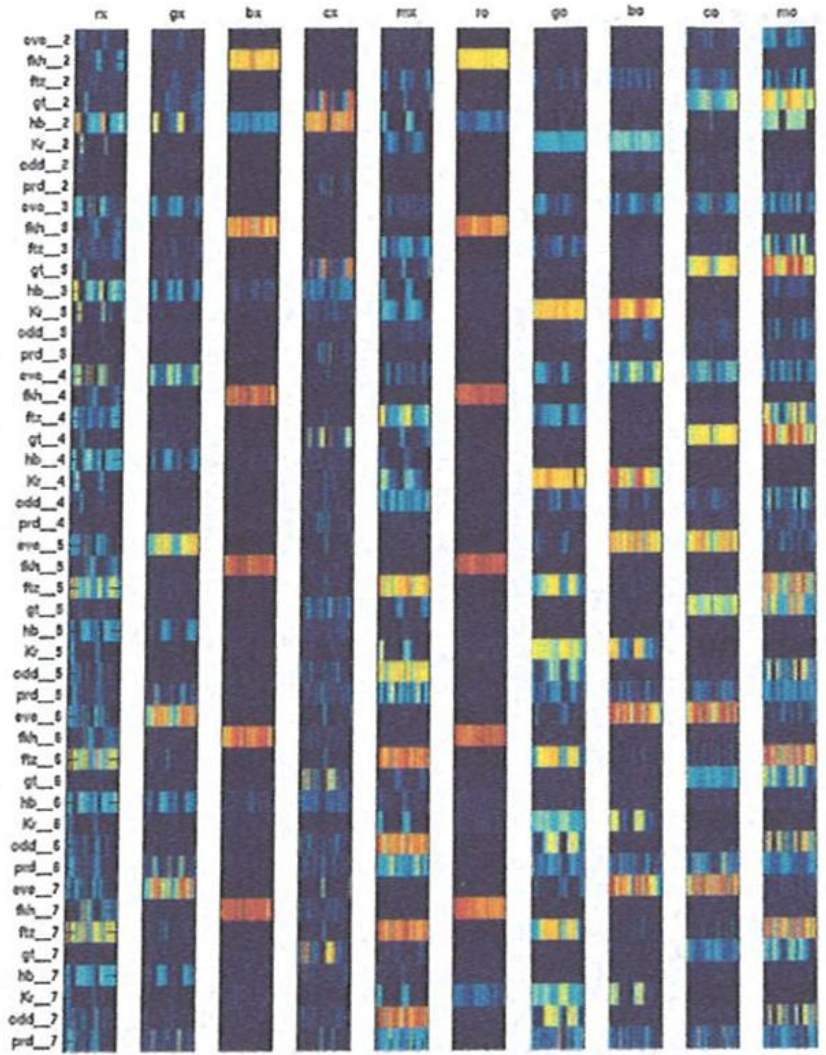
Task 2: Find out which genes were different in the outlier cells.

Column: a cell.

Grouped columns: clusters of cells in the outlier cell plot.

Rows: genes and 6 time points of each gene.

Heatmap: encode gene expression values using color.



Limitation of the existing tool

Manual look-ups between multiple views.

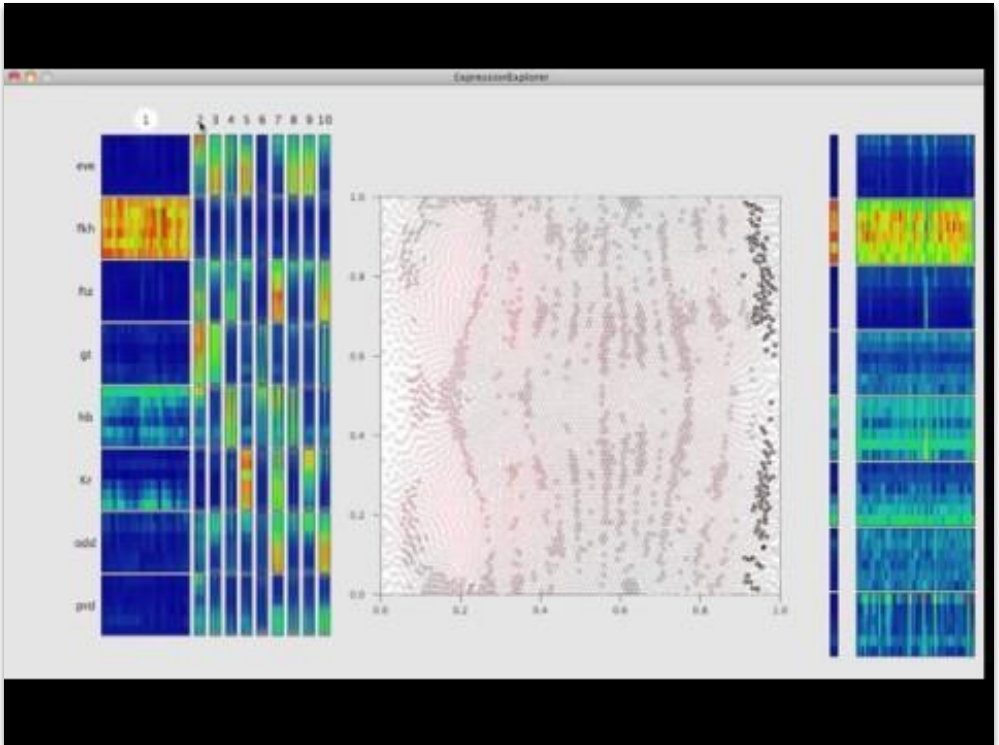
Task 3 (not supported): characterize how this gene expression is different from the corresponding cells in another embryo.

Task 3 requires comparison of numerous numbers of heatmaps.

First iteration

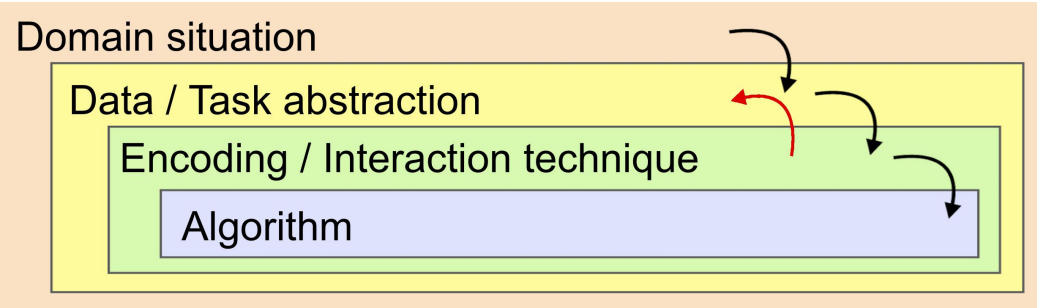
Link two views together via user interaction.

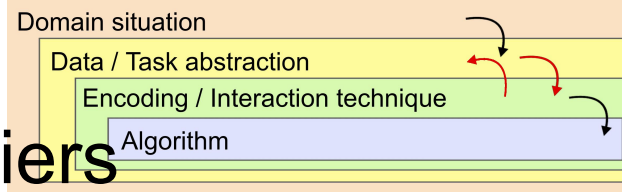
Details on demand.



Deploy and interview

Problem: The outlier detection algorithm was too restrict, resulting in a rethinking of biologists' computational approach.





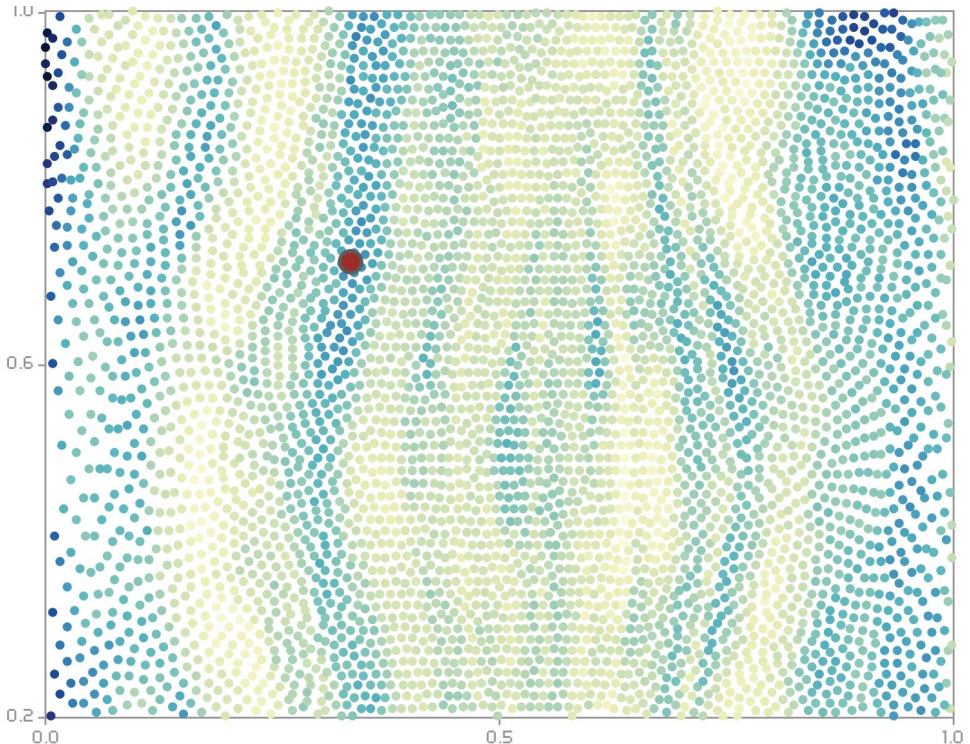
Second iteration -- Similarity, not outliers

How similar each cell in one embryo was compared to corresponding cells in the other embryo.

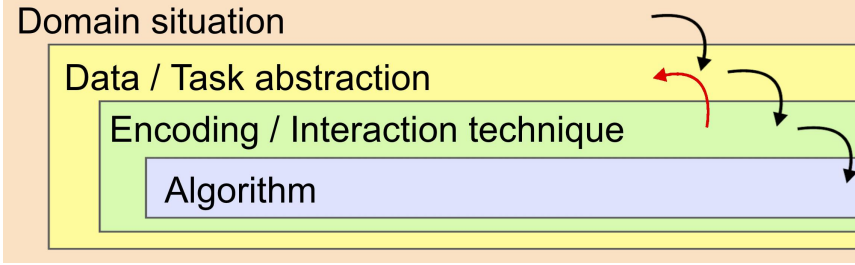
Task 1:

From “Find outlier cells” to

“Find cells with low similarity.”



Second iteration -- Results



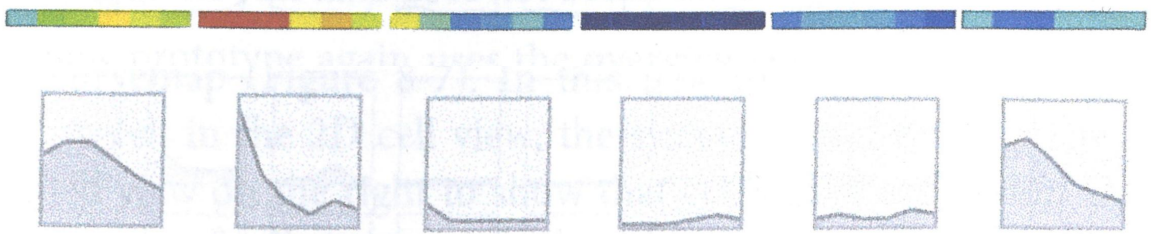
Explore many more cells than the first version.

Finding: The experimental measure from one of the species was plagued with low-level noise, causing the biologists to go back and recapture the data.

Emerging question: What would a different similarity metric reveal?

A final version -- Apply good design principles

Measure a gene in 6 time points of its expression values.



	Quantitative	Ordinal	Nominal
More Accurate	Position	Position	Position
	Length	Density	Hue
	Angle	Saturation	Density
	Slope	Hue	Saturation
	Area	Length	Shape
	Density	Angle	Length
	Saturation	Slope	Angle
	Hue	Area	Slope
Less Accurate	Shape	Shape	Area

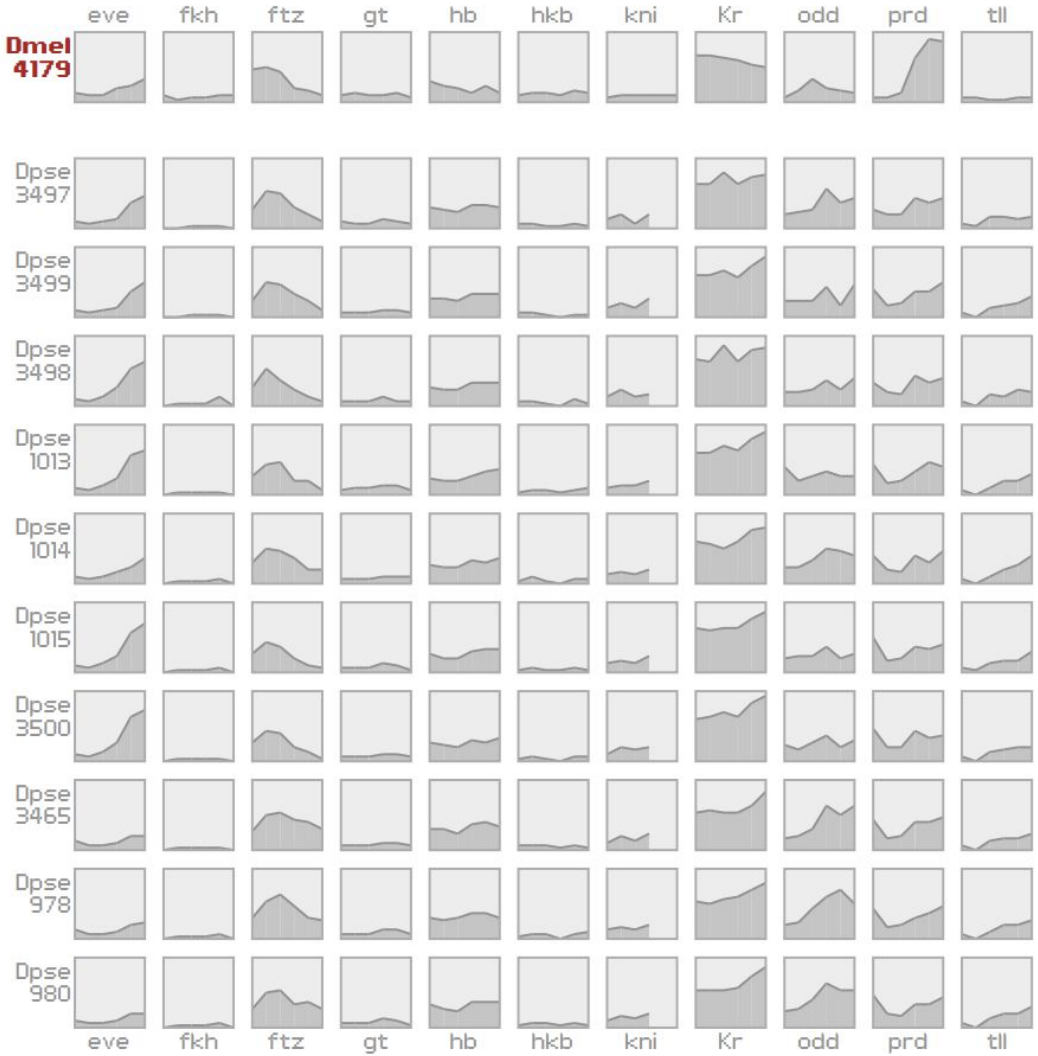
A final version

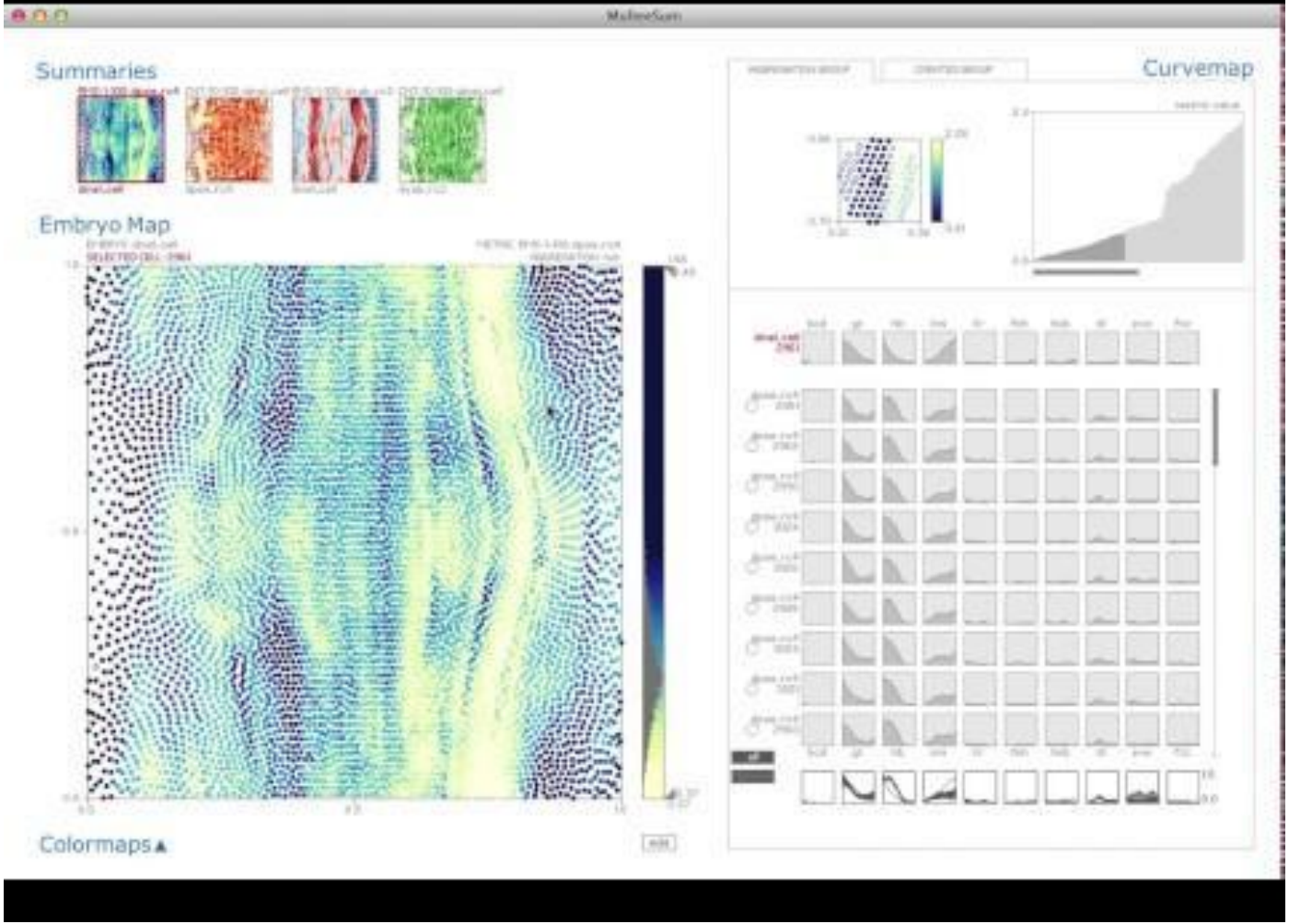
Columns: gene

Top row: a selected cell

Bottom rows: corresponding cells from other species.

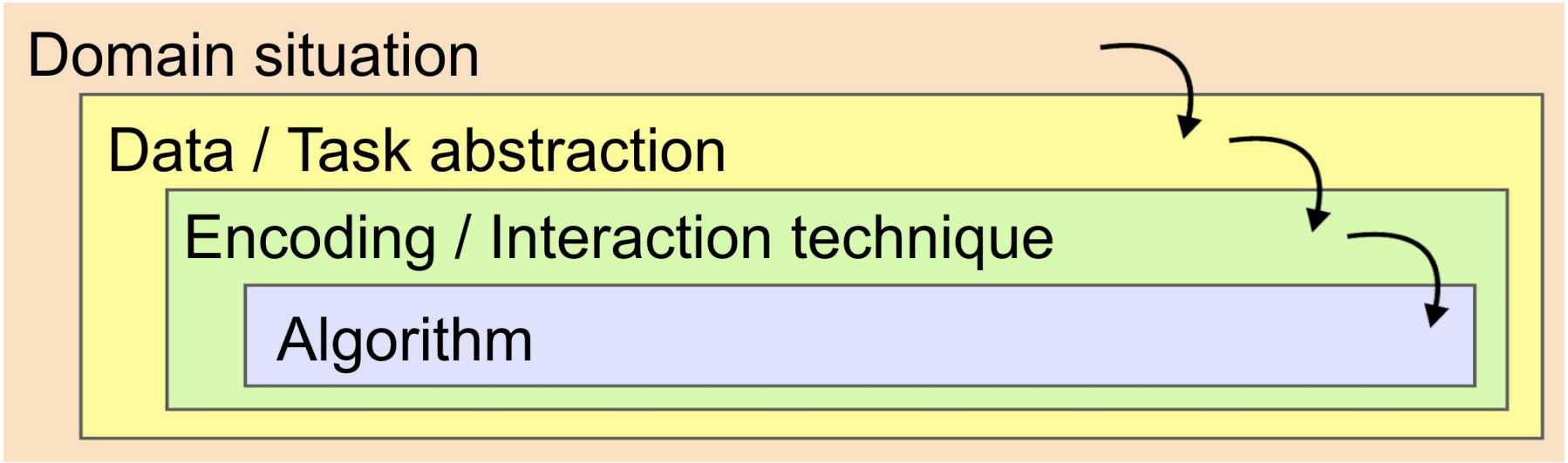
Guideline: It is important to present new ideas to the target users with their own data.





Meyer, M., Munzner, T., DePace, A. and Pfister, H., 2010. MulteeSum: a tool for comparative spatial and temporal gene expression data. *IEEE TVCG*.

Recap -- An iterative process with rapid prototyping



Try out this design process with your project!