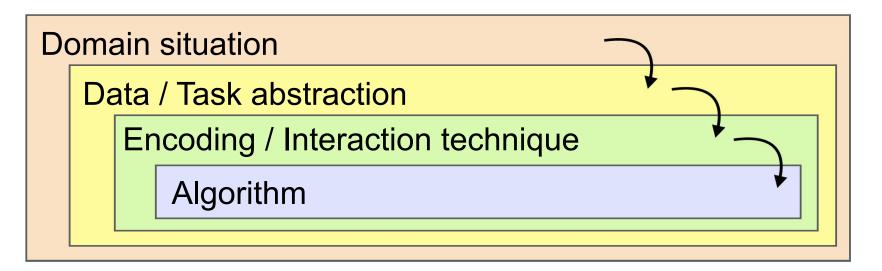
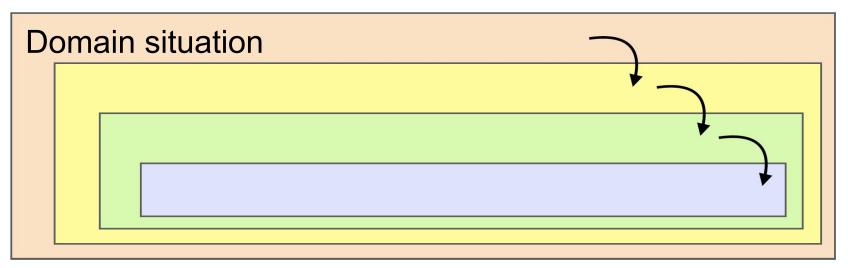
# **Visualization Design**

Chen He

#### Nested Model for Visualization Design



#### **Domain situation**



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

#### Data / Task abstraction

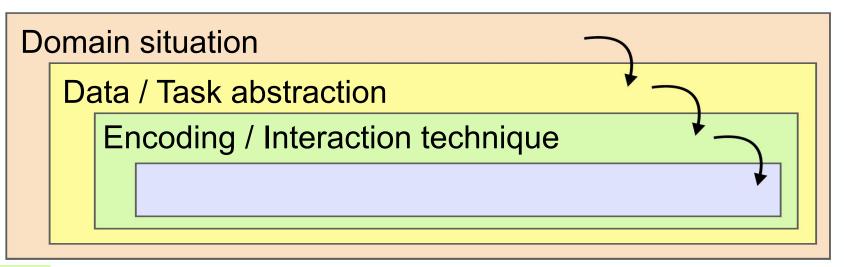
Do	omain situation
	Data / Task abstraction
	*

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.

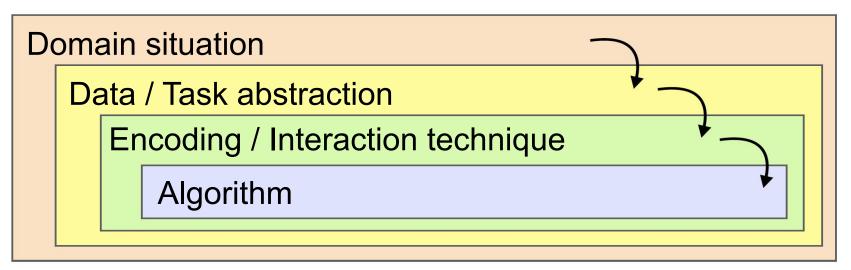
#### Encoding / Interaction technique



How is the data shown?

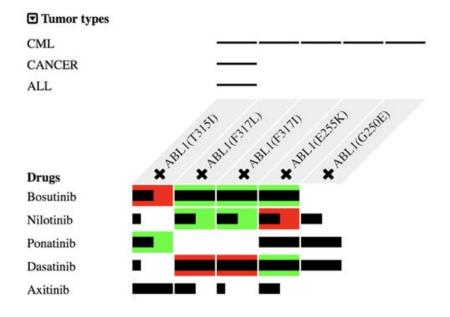
Decide on the specific way to **create** and **manipulate** the **visual representation** of the **abstraction**.

# 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**

O	m	ain	situation -			
	Da	ata	/ Task abstraction	+		
		Er	ncoding / Interaction technique		* ~	
			Algorithm		+	
						_

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

D

	Biomarker 🧿	<b>→</b> Drug	Effect 🧿	Evidence 🧿	Source	Curator	Tumor type
	E255K						
	ABL1 (T315A	Bosutinib (BCR	Responsive	∧ NCCN guidelines	PMID:21562040	RDientsmann	CML
CANCER GENOME	ABL1 (F359V	Dasatinib (BCR	Responsive	∧ NCCN guidelines	PMID:21562040	RDientsmann	CML
INTERPRETER	ABL1 (1242T,	Imatinib (BCR-A	Resistant		PMID:21562040	CRubio-Perez	CML
(CGI)	ABL1 (E255K	Nilotinib (BCR-A	Resistant		PMID:21562040	CRubio-Perez	CML

Comp(~Y	✓ UniprcY	Target Pref Name Y	Gene Name Y	Wild ty Y	Mutation ir	PubMed ID Y	End Y	EnY	End P.Y	EΥ	
PONATINIB	A Sort As	cending	ABL1	mutated	ABL1(E255K)	23301703	IC50	=	1.08	nM	
NILOTINIB	A Sort De	scending	ABL1	mutated	ABL1(E255K)	23301703	IC50	=	27.8	nM	
IMATINIB	2× Remove	e Sort	ABL1	mutated	ABL1(E255K)	23301703	IC50	=	485.8	nM	
DASATINIB	Show rows where:		ABL1	mutated	ABL1(E255K)	23301703	IC50	=	0.21	nM	
			ABL1, ABL2	mutated	ABL1(E255K)	23414803	ACTI	=	100	%	
	And 👻			ABL1, ABL2	mutated	ABL1(E255K)	23414803	ACTI	=	88	%
			ABL1	mutated	ABL1(E255K)	23301703	IC50	=	5.4	nM	
	contains		ABL1	mutated	ABL1(E255K)	23301703	IC50	=	2.9	nM	
			ABL1	mutated	ABL1(E255K)	23301703	KD	=	0.28	nM	
	Filter	Clear	ABL1	mutated	ABL1(E255K)	23301703	IC50	=	0.27	nM	



#### **Task abstraction**

om	ain situation
D	ata / Task abstraction
	Encoding / Interaction technique
	Algorithm

D

**Integrate** drug-target relations from two different sources. Allow user **exploration** of drug-target relations.

#### **Data abstraction**

D	oma	ain	situation	$\overline{}$	
	Da	ata	/ Task abstraction	١	
		Er	ncoding / Interaction technique		* ~
			Algorithm		¥

	<b>Discrete</b> (no between values)	<b>Continuous</b> (values between)
<b>Ordered</b> (values are comparable)	<b>Ordinal,</b> e.g. size: S,M,L,XL, <b>Quantitative,</b> e.g. counts: 1,2,3,	<b>Fields,</b> e.g. altitude, temperature
<b>Unordered</b> (values not comparable)	<b>Nominal,</b> e.g. shape: □O∆ <b>Categories,</b> e.g. nationality	<b>Cyclic values,</b> e.g. directions, hues

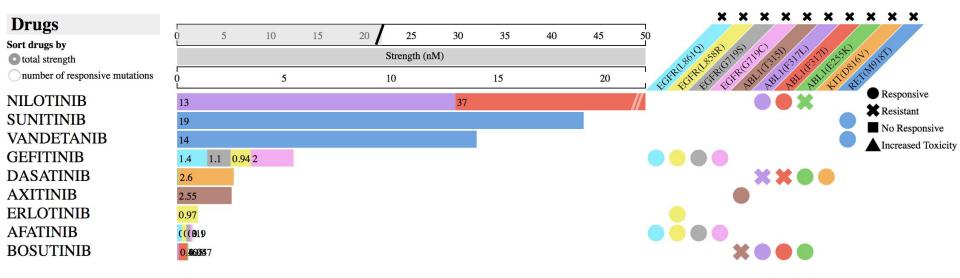
D	oma	ain	situation	$\overline{}$	
	Da	ata	/ Task abstraction	١	
		Er	coding / Interaction technique		+ ~
			Algorithm		+

#### **Data abstraction**

Data	Data type													
Drug	Nominal	_												
Mutation	Nominal	_												
		-		r 🕜 ADrug		Effect 📀	Evidence	0	Sourc	ce	Curator		Tumor t	уре
Tumor type	Nominal		E255K	15A Bosutin	ib (BCR	Responsive	⊗ NCCN gui	idelines	PMID:2	1562040	RDientsm	ann	CML	
		CANCER GENOME	ABL1 (F3	59V Dasatin	nib (BCR	Responsive	NCCN gu	idelines	PMID:2	1562040	RDientsm	ann	CML	
Drug-target relation	n from CGI	INTERPRETER		42T, Imatinib		Resistant					CRubio-P			
			ABL1 (E2	255K Nilotinit	b (BCR-A	Resistant		Leuk	PMID:2	1562040	CRubio-P	erez	CML	
Effects	Nominal		Comp(+Y	✓ UniprcY Tar	rget Pref Nam	e Y Gene Na	me Y Wild t	Y Mutatio	on irY	PubMed ID	Y End	Y En	Y End P.Y	EΥ
	Norman		PONATINIB	2 Sort Ascen	ding	ABL1	mutate	ed ABL1(E	255K)	23301703	IC50	=	1.08	nM
		-	NILOTINIB	A Sort Descer		ABL1	mutate	ed ABL1(E	255K)	23301703	IC50	) =	27.8	nM
Evidence level	Ordinal		IMATINIB	2× Remove So Show rows wh		ABL1		ed ABL1(E			IC50		485.8	nM
			DASATINIB	contains	lere.	ABL1		ABL1(E			IC50		0.21	nM %
Drug-target relation	from DTC			[		ABL1, AB		ed ABL1(E			ACTI		88	%
Diug-laiget relation		DRUGTARGET		And 👻		ABL1	1000	ad ABL1(E			IC50		5.4	nM
	-	_		contains		* ABL1		ad ABL1(E			IC50		2.9	nM
Potency	Quantitative					ABL1	mutate	ed ABL1(E	255K)	23301703	KD	=	0.28	nM
				Filter	Clear	ABL1	mutate	ABL1(E	255K)	23301703	IC50	=	0.27	nM

#### Prototype version Zero

Quantita	tive	Ordina	al	Nomin	al
Position	•••	Position	•••	Position	•••
Length	=	Density		Hue	
Angle	4	Saturation		Density	
Slope	11	Hue	•••	Saturation	
Area	••	Length	_	Shape	• • =
Density		Angle	2	Length	—
Saturation		Slope	1-	Angle	2
Hue		Area	••	Slope	1-
Shape		Shape	• • =	Area	••



#### **Task abstraction**

oma	ain situation	$\overline{}$
Da	ata / Task abstraction	* ~
	Encoding / Interaction technique	* ~
	Algorithm	ł

D

**Integrate Compare** drug-target relations from two different sources. Allow user **exploration** of drug-target relations.

D	oma	ain	situation	$\overline{}$			
	Da	ata	/ Task abstraction	١			
		Er	ncoding / Interaction technique		١	5	
			Algorithm			ł	·
							_

#### **Data abstraction**

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

0 DRUGTARGET COMMONS

	Biomarker 🧿	-Drug	Effect 🧿	Evidence 🧿	Source	Curator	Tumor type
	E255K						
••••••	ABL1 (T315A	Bosutinib (BCR	Responsive	∧ NCCN guidelines	PMID:21562040	RDientsmann	CML
CANCER GENOME	ABL1 (F359V	Dasatinib (BCR	Responsive	∧ NCCN guidelines	PMID:21562040	RDientsmann	CML
INTERPRETER	ABL1 (1242T,	Imatinib (BCR-A	Resistant		PMID:21562040	CRubio-Perez	CML
	ABL1 (E255K	Nilotinib (BCR-A	Resistant		PMID:21562040	CRubio-Perez	CML

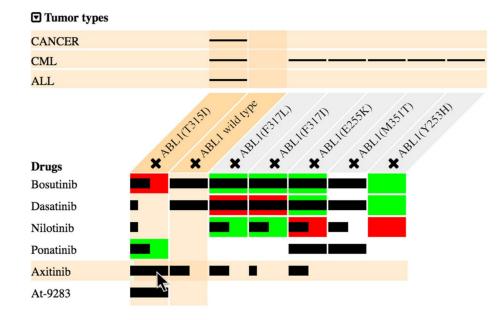
Comp( - Y	✓ Uniprc Y Target Pref Name Y	Gene Name Y	Wild ty Y	Mutation ir	PubMed ID Y	End Y	EnY	End P.Y	EΥ
PONATINIB	2 Sort Ascending	ABL1	mutated	ABL1(E255K)	23301703	IC50	=	1.08	nM
NILOTINIB	ZI Sort Descending	ABL1	mutated	ABL1(E255K)	23301703	IC50	=	27.8	nM
IMATINIB	2× Remove Sort	ABL1	mutated	ABL1(E255K)	23301703	IC50	=	485.8	nM
DASATINIB	Show rows where:	ABL1	mutated	ABL1(E255K)	23301703	IC50	=	0.21	nM
	contains v	ABL1, ABL2	mutated	ABL1(E255K)	23414803	ACTI	=	100	%
	And 👻	ABL1, ABL2	mutated	ABL1(E255K)	23414803	ACTI	=	88	%
		ABL1	mutated	ABL1(E255K)	23301703	IC50	=	5.4	nM
	contains 👻	ABL1	mutated	ABL1(E255K)	23301703	IC50	=	2.9	nM
	Filter Clear	ABL1	mutated	ABL1(E255K)	23301703	KD	=	0.28	nM
		ABL1	mutated	ABL1(E255K)	23301703	IC50	=	0.27	nM

#### Visual encoding - Layout

(	oma	ain situation	$\overline{}$
	Da	ata / Task abstraction	* ~
		Encoding / Interaction technique	* ~
		Algorithm	¥

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

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



D

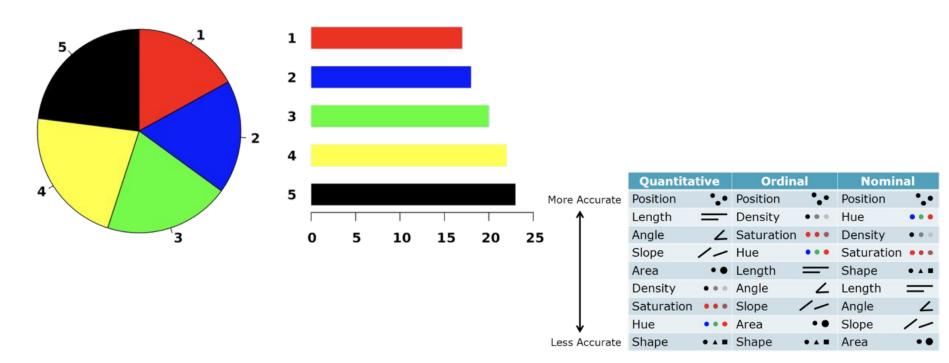
#### Visual encoding

	Quantitative		Ordinal		Nominal	
More Accurate	Position	•••	Position	•••	Position	•••
1	Length	=	Density		Hue	
	Angle	2	Saturation		Density	
	Slope	1-	Hue		Saturation	
	Area	••	Length	=	Shape	• • =
	Density		Angle	2	Length	=
	Saturation		Slope	1-	Angle	2
¥	Hue		Area	••	Slope	1-
Less Accurate	Shape	• • =	Shape	• • =	Area	••

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

#### Visual encoding

Angle & Area < Position & Length



			Domain situation
		Data / Task abstraction	
			Encoding / Interaction technique
Visual enco	aing		Algorithm
Data	Data type	Visual variable	
Drug	Nominal	Position	
Mutation	Nominal	Position	Matrix layout
Tumor type	Nominal	Position	layout
Drug-target relation	n from CGI	·	
Effects	Nominal	Hue	Responsive provide and the second sec
Evidence level	Ordinal	Position, length,	Rest to Truce Guidelines Quantitative Ordinal Nominal
		saturation	Early trials Position •• Position •• Position
			Case report Pre-clinical Angle ∠ Saturation ••• Density •••
Drug-target relation from DTC		Slope / Hue ••• Saturation •••	
	-		Highly potent Area •• Length — Shape ••
Potency level	Ordinal	Position, length	Potent     Density     •••     Angle     ∠     Length       Waskly potent     Saturation     •••     Slope     /     Angle     ∠
		r contori, iongui	Weakly potent Inactive Hue ••• Area •• Slope
			Shape • • Shape • • Area • •

## Visual encoding

Potency level

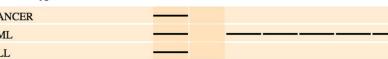
Data	Data type	Visual variable			
Drug	Nominal	Position	🛡 Tumor		
Mutation	Nominal	Position	CANCER CML ALL		
Tumor type	Nominal	Position			
Drug-target relation	n from CGI		Drugs		
Effects	Nominal	Hue	Bosutinib Dasatinib		
Evidence level	Ordinal	Position, length,	Nilotinib Ponatinib		
		saturation	Axitinib At-9283		
Drug-target relation from DTC					

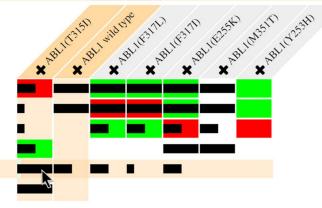
Ordinal

Position, length

domain situation data/task abstraction encoding/interaction technique algorithm

#### **Tumor types**



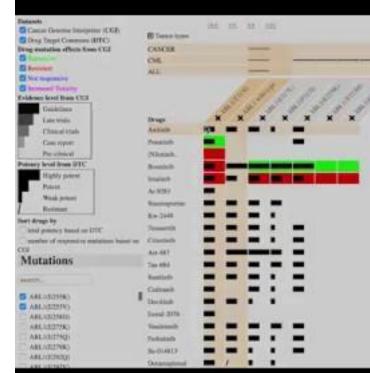


#### **Domain situation**

Data / Task abstraction

Encoding / Interaction technique

#### Interaction technique



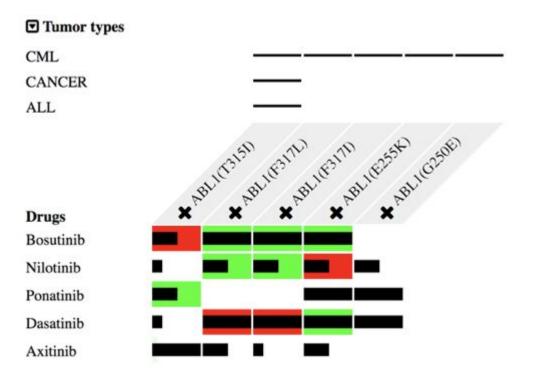
Algorithm

CCE Responsive between Askinds and ARLI(T3121) issue on Pre-effected

Tab Actual effectively address BCR ABC//T2157 with a disease biodiag conferencess. Address

The BCR ABLI ferror gene is a driver storagene in chemic reprint? traincross and 30 SPE of cours of adult acore (prephotosate traincross. introduction of AEL! know adultion (for express), instially for marked); improved patient carting), but acquired along ministance remains a challenge. Point metalasis to the APL I khose dowain weakers anothers backing self. represent the seast containers choical resistence inachations. The BCB ARCI kinss detrare protosper notaine Th/1110 conten wanted a still approval ABCI inhibitors anorys possibility which has accordy limitations." Here we combrane comprehensive drag sensitivity and resistance profiling of potenti cella en irivo with ideactatal analysis is intabilist the VEEPR process. kinner infoltutor aubigilt as a arbeitive and official or inhistory for TSUSI anatase WOR-ARCI driven took across. Available percently tobobilised INCR-ARCUTETENES. at both biochemical and collater levels, by fireding to the active Bonn of AIR STTUND is a mainteen adjustice hitsing much. These Stalings access that the TNDD maturiou old the the crediteleastment significant of the kinetic to lavour of an active (DFG on) A long-contennation, which has none optimal binding interactions with manuals. Deamage of a CDDB chevels: myscheid instances points with assists mached in a signif reductors of T3155 produce culls from hora marton. Taken expellen, our Railings-demonstrate an unorganized opportantly to repurpose axilially, an anti-angiogenic drog. approval for your concer, as an oddition for ARL 1 passionper mature dragrestinant halkannin pataret. This stary shows that wild trys pression do not afterney sample the conductorations are deliver to discuss othersard washed principal and that comproheneitye drug testing of particul durities cells can identify proposition to be a substantiation of the second se 3804 30.025%/satura14115

#### 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

Domain situation
Data / Task abstraction
Encoding / Interaction technique
Algorithm

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

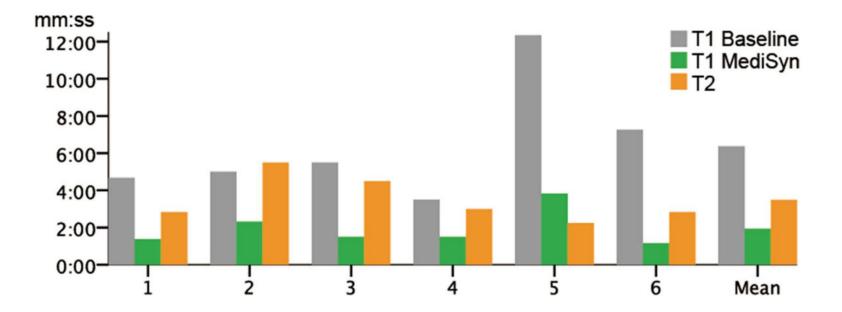
Baseline: two unlinked datasets

Participants: 6 domain experts

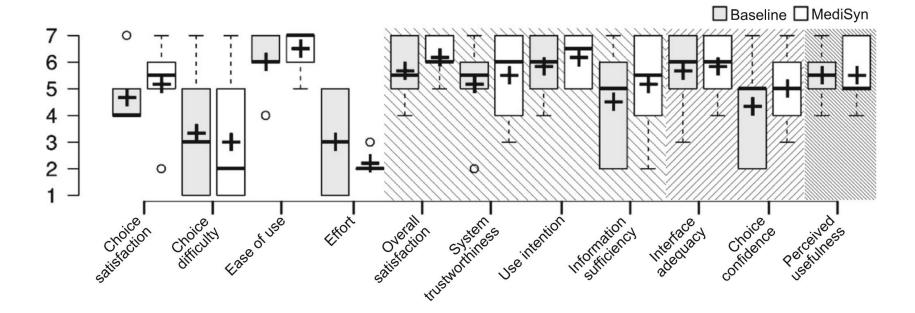
Tasks: 1. drug selection; 2. inconsistency discovery

Measure: task performance, subjective feedback

#### **Results - Lab study**



#### Results - Lab study

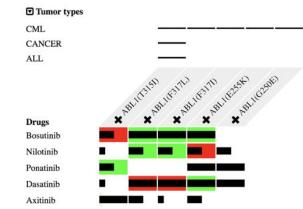


#### Findings - Lab study

Matrix view supports drug comparison and exposes missing data.

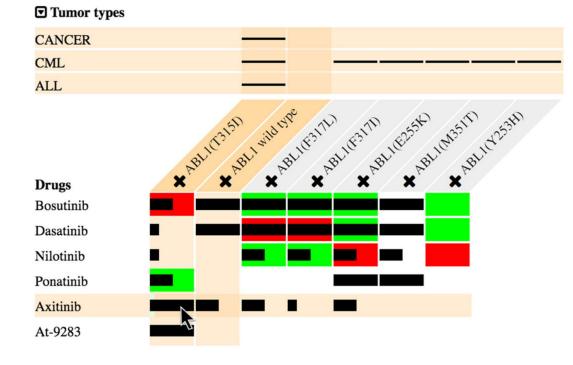
Depiction of datasets in **overlaid layers** facilitates **direct comparison of data from multiple sources**. (Data consistency)

**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					
v	alid	lat	e: observe and interview target users		
Г	th		at: bad data/operation abstraction		
L		tł	nreat: ineffective encoding/interaction technique	Э	
L		v	alidate: justify encoding/interaction design		
L			threat: slow algorithm		
L		П	validate: analyze computational complexity		
L		П	implement system		
L		П	validate: measure system time/memory		
L		v	alidate: qualitative/quantitative result image and	alysis	
L		[t	est 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					
v	alid	at	e: observe adoption rates		

#### **Iteration: Task abstraction**

om	ain situation	$\overline{}$
Da	ata / Task abstraction	* ~
	Encoding / Interaction technique	* ~
	Algorithm	+

D

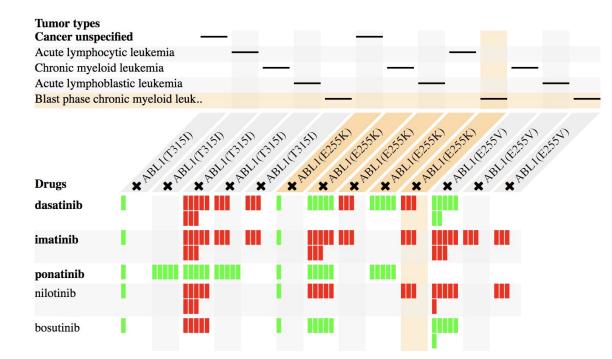
Compare drug-target relations from more than two sources.

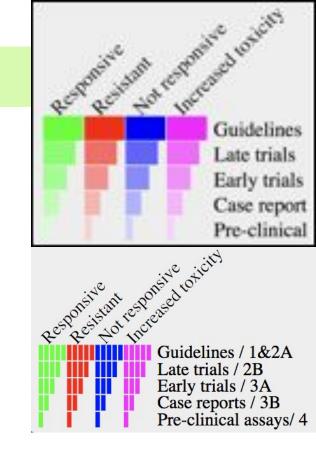
Allow user exploration of drug-target relations.

Support insight recording and sharing.

#### Iteration: Visual encoding

#### Five datasets - Juxtaposed bars





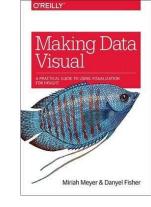
#### Iteration: Interaction technique

<ul> <li>MediSyn Biomedical Da</li> <li>A</li> <li>search</li> <li>Mutations</li> <li>Tumors</li> <li>Acute leukemia of ambiguous lineage</li> <li>Acute lymphoblastic leukemia</li> <li>Acute lymphocytic leukemia</li> <li>Acute myeloid leukemia</li> <li>Acute myeloid leukemia associated with mds</li> <li>Selected entities</li> </ul>	tasets Synthesizer Instruction●         Tumor types         Acute lymphocytic leukemia         Cancer unspecified         Chronic myeloid leukemia         Acute lymphoblastic leukemia         Blast phase chronic myeloid leuk         Drugs         ★ Button Button Button         Acute lymphoblastic leukemia         Blast phase chronic myeloid leuk	CGI: Chronic myeloid leukemia with ABL1(T3151) mutation is Resistant to dasatinib treatment in European LeukemiaNet guidelines. Title BCR-ABL kinase domain mutation analysis in chronic myeloid leukemia patients treated with tyrosine kinase inhibitors: recommendations from an expert panel on behalf of European LeukemiaNet. DOI 10.1182/blood-2010-12-326405	<ul> <li>✓ My private notes</li> <li>✓ Other public notes</li> <li>✓ Other public notes</li> <li>All notes related to: dasatinib</li> <li>★</li> <li>The references given in case of dasatinib responses towards BCR-ABL kinase mutation are inconclusive, primarily due to insufficient information. i.e. mutation in only ABL or BCR or due to the fusion are not mentioned, neither the reference matched.</li> <li>Entity: dasatinib</li> <li>✓ ♥ View provenance</li> </ul>
ABL1(T315I) ★ Datasets of Drug Effects  The Cancer Genome Interpreter (CGI)  OncoKB  The Gene Drug Knowledge database (Synapse)  COSMIC	dasatinib       asstinib         select       asstinib         imatin       view notes         tyrosine kinase       asstinib         nilotinib       asstinib	Dasatinib	
<ul> <li>COSMIC</li> <li>Drug Target Commons (DTC)</li> </ul>	aurk inhibitors + B	Cancel Save&close   Private	D

#### 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

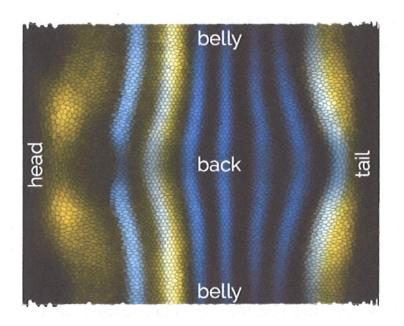
Goal:

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

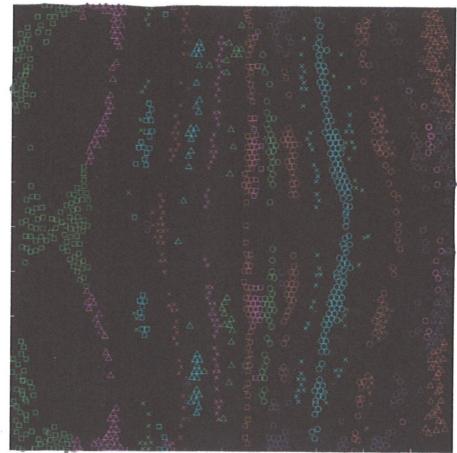
### Existing task and tool

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

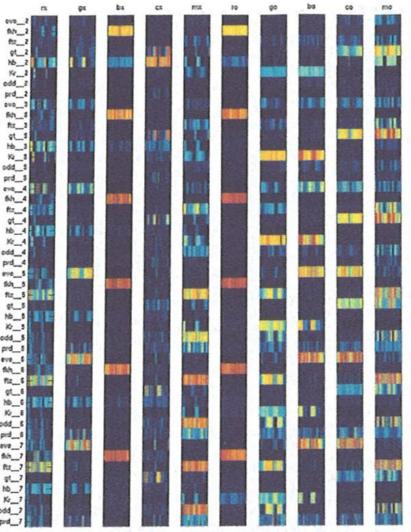
Task 2: Find out which genes were different in the outlier cells.

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.

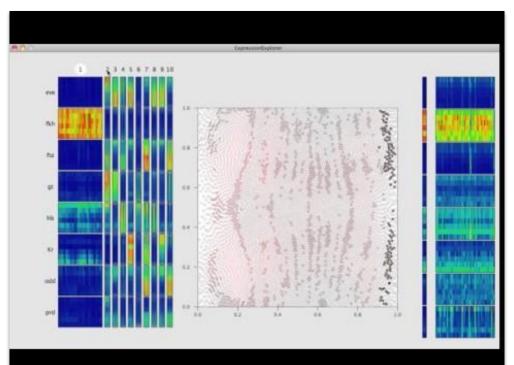
Task 3 (not supported): characterize how this gene expression is different from the corresponding cells in another embryo.

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.

D	oma	ain	situation	$\neg$	
	Da	ata	/ Task abstraction	-	
		Encoding / Interaction technique		*	
			Algorithm		+



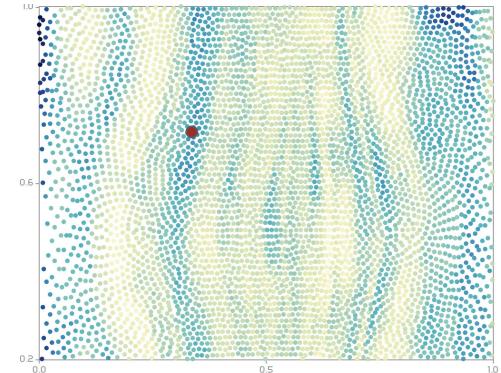
# Second iteration -- Similarity, not outliers

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

om	ain situation	$\supset$		
Da	ata / Task abstraction	*		
	Encoding / Interaction technique		+_	$\overline{}$
	Algorithm			ł

Explore many more cells than the first version.

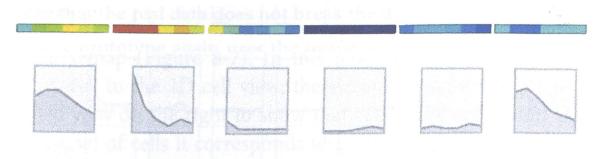
Finding: The experimental measure from one of the species was plagued with low-level noise, causing the biologists to go back and recapture the data.

D

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	•.•
1	Length	=	Density		Hue	
	Angle	2	Saturation		Density	
	Slope	1-	Hue		Saturation	
	Area	••	Length	=	Shape	• • =
	Density	• • •	Angle	2	Length	—
	Saturation		Slope	1-	Angle	2
↓ ↓	Hue		Area	••	Slope	1-
Less Accurate	Shape	• • =	Shape	• • =	Area	••

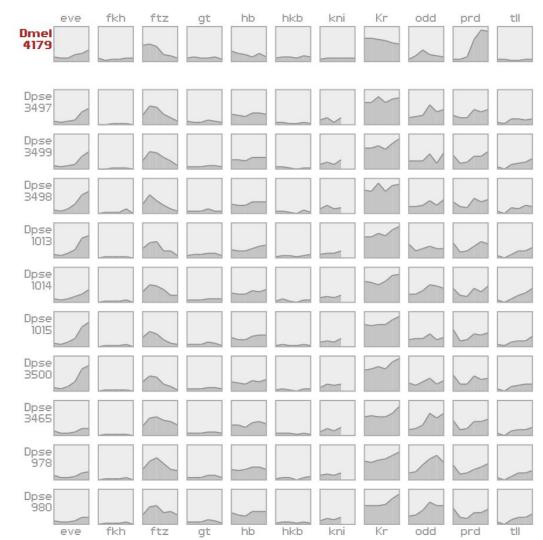
# A final version

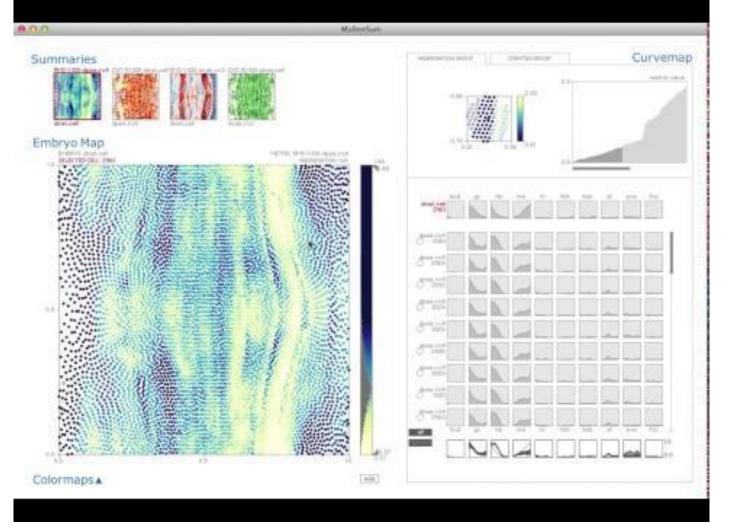
Columns: gene

Top row: a selected cell

Bottom rows: corresponding cells from other species.

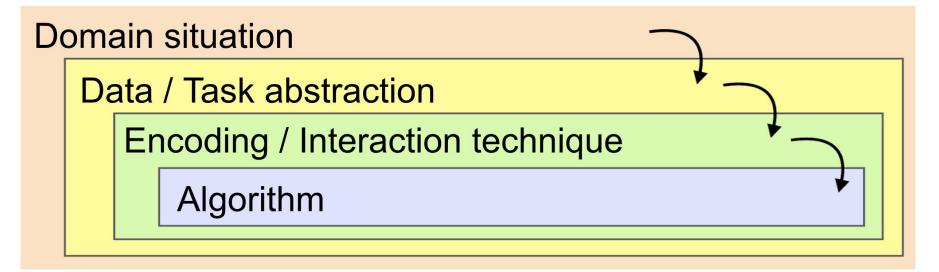
Guideline: It is important to present new ideas to the target users with their own data.





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!