(Designing) Interactive Visualisations to Solve Analytical Problems (in biology)

CAGATAY TURKAY,
giCentre, City University London

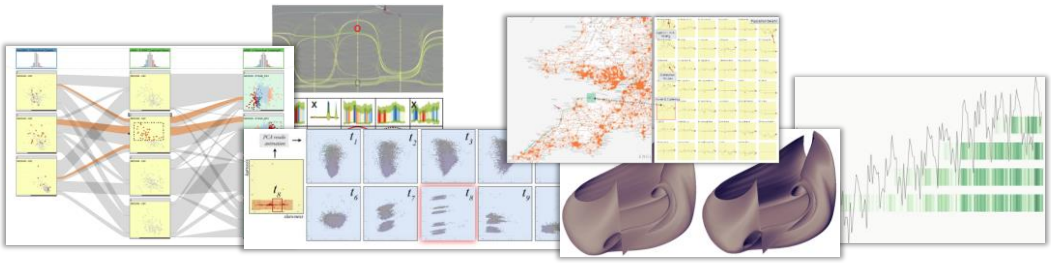
CITY UNIVERSITY LONDON

giCentre
www.gicentre.net

Who?

CITY UNIVERSITY LONDON

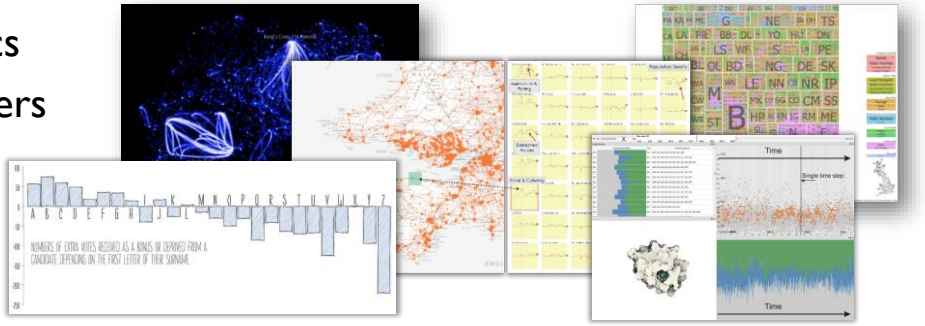
- Lecturer in **Applied Data Science** @ the giCentre, CUL
- PhD @ VisGroup at **Univ. of Bergen, Norway**
- Research interests:
  - Integrating **Computational** Tools in **Interactive Visual** Analysis Methods
  - **Perceptually** Optimized **Visualization**
- Methods for several domains:
  - Biology, transport, intelligence, neuroscience



giCentre *(www.giCentre.net)*

CITY UNIVERSITY LONDON

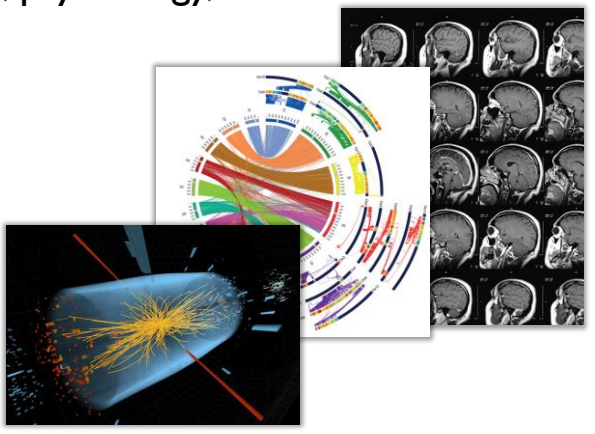
- 6 academics
- 2 researchers
- 5 PhDs



Data **supported** science

CITY UNIVERSITY LONDON

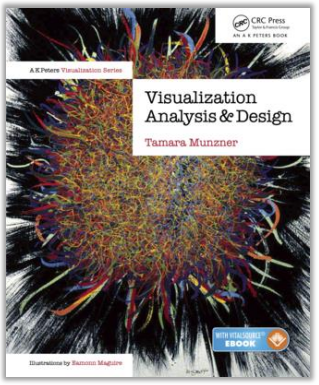
- **Data analysis** in almost all scientific fields
  - Biology, medicine, astronomy, psychology,…
- Data ~~**driven**~~ science
- Research in **several fields**
  - Visualization
  - Data Mining
  - Machine Learning
  - Statistics



Visualization ?

CITY UNIVERSITY LONDON

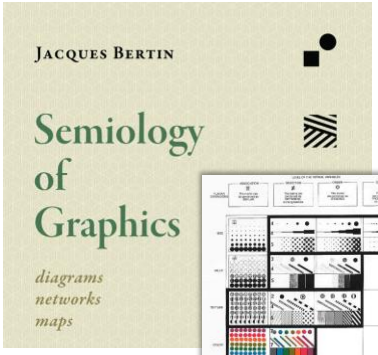
"Computer-based visualization systems provide visual representations of ***datasets*** *designed to help* ***people*** *carry out tasks more effectively."* *[Tamara Munzner, 2014]*

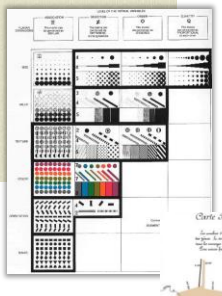


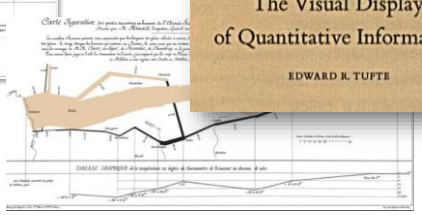
"The use of computer-generated, ***interactive****, visual representations of data to amplify* ***cognition****"* *[Card, Mackinlay, & Shneiderman 1999]*

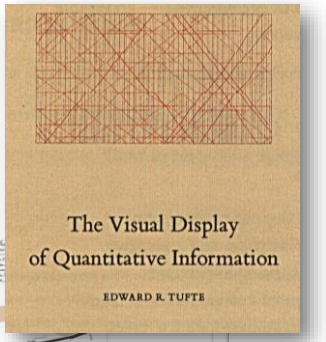
VIS -- a mature field already

CITY UNIVERSITY LONDON









VAST INFOVIS SCIVIS
VIS 2014
9-14 NOVEMBER 2014
PARIS, FRANCE

EuroVis 2014
The Eurographics Conference on Visualization

biovis

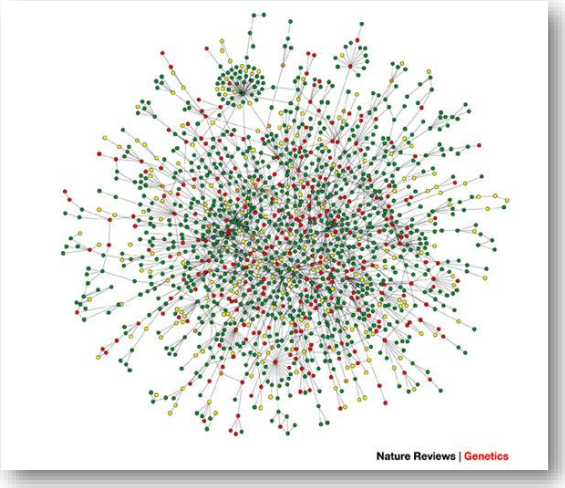
VISUALIZING BIOLOGICAL DATA
VIZBI

Biological data + VIS: A good synergy .. but why?

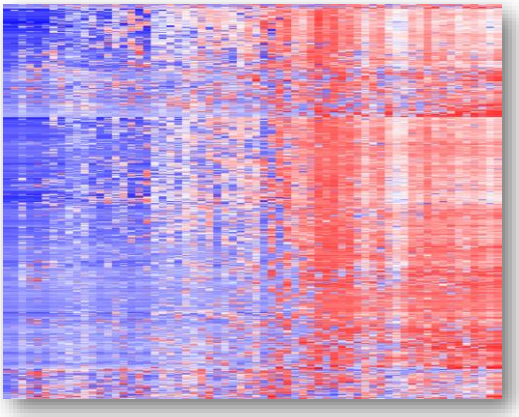
Why biology is interesting for VIS?

CITY UNIVERSITY LONDON

Datasets are **large** & **heterogeneous**



Yeast Protein interaction network, Barabási & Oltvai, 2004



Clustering miR expressions
http://gdac.broadinstitute.org/

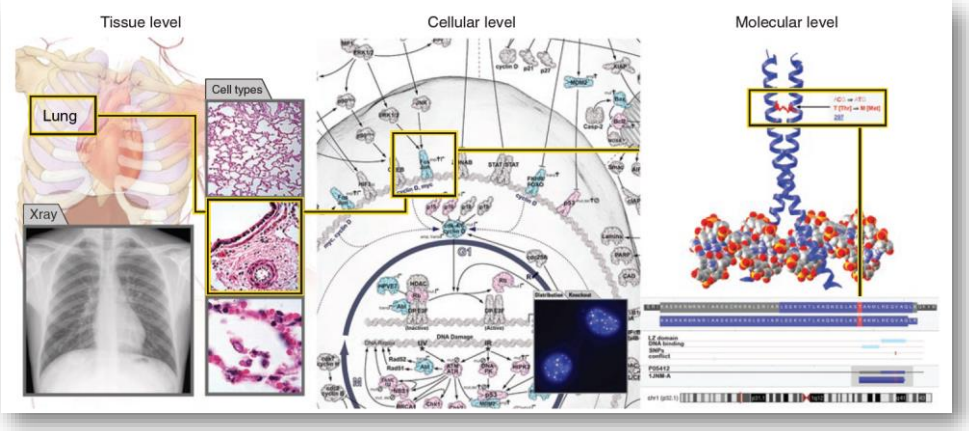
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
ID	BirthDate	Gender	Education	cvlt_tri_tc	cvlt_tri_tc	cvlt_sd_fr	cvlt_sd_fr	cvlt_ld_fr	cvlt_ld_fr	cvlt_rec_h	cvlt_rec_h	cvlt_rec_f	cvlt_rec_f	cvlt_tot_d	cvlt_tot_d	cwi_1_col	cwi_1_col	cwi_2_wo
501	1951	2	14	70	73	15	1.5	16	1.5	16	0.5	0	-1	4	1.5	30	10	22
507	1948	2	15	50	51	11	0	9	-1	15	0	0	-1	3.7	1	33	9	20
508	1947	1	15	52	64	11	1	10	0.5	15	0.5	1	-1	3.4	1	27	12	21
510	1956	1	11	56	62	12	1	12	1	13	-0.5	4	0.5	2	-0.5	37	7	28
512	1935	2	17	61	68	15	1.5	16	2	16	0.5	0	-1	4	1.5	34	10	26
514	1930	1	18	40	54	9	0.5	10	1	15	0.5	1	-1	3.4	1.5	37	9	20
517	1949	1	13	54	60	13	1.5	12	1	15	0.5	5	0.5	2.5	0	24	13	20
518	1942	1	19	45	56	9	0	11	1	14	0	1	-1	3	0.5	38	8	20
519	1944	2	14	61	65	13	1	14	1	16	0.5	1	-0.5	3.7	1	38	8	24
520	1946	1	18	33	41	6	-0.5	7	-0.5	11	-2	3	0	1.8	-1	29	12	24
523	1946	1	19	75	87	16	2.5	16	2.5	16	1	0	-1	4	2	29	12	20
524	1950	2	16	63	66	15	1.5	14	1	16	0.5	0	-1	4	1.5	26	12	21
526	1945	1	18	59	71	12	1	11	1	15	0.5	3	0	2.9	0.5	28	12	22
527	1945	2	12	63	66	15	2	14	1	14	-1	1	-0.5	3	0	27	12	18
529	1942	2	19	65	68	14	1.5	13	0.5	14	-1	2	0	2.7	-0.5	28	12	19
530	1949	1	12	50	56	10	0.5	10	0	15	0.5	0	-1	3.7	1.5	24	13	21
532	1949	1	15													34	8	25
533	1948	2	9	58	62	12	1	12	0.5	15	0	0	-1	3.7	1	26	13	21
537	1951	2	15	61	64	12	0.5	[illegible]	[illegible]	16	0.5	0	-1	4	1.5	27	11	20
538	1956	2	16	66	69	15	1.5	[illegible]	[illegible]	16	0.5	0	-1	4	1.5	28	11	19
539	1950	1	15	62	68	15	2	[illegible]	[illegible]	16	1	0	-1	4	1.5	37	7	26
542	1954	2	15	58	61	11	0	[illegible]	[illegible]	16	0.5	1	-0.5	3.7	1	35	8	21
543	1946	2	18	65	68	15	2	[illegible]	[illegible]	16	0.5	1	-0.5	3.7	1	35	9	23
544	1945	2	12	51	55	10	0	[illegible]	[illegible]	15	0	0	-1	1.7	1	30	11	22
545	1952	2	15	62	65	12	0.5	[illegible]	[illegible]	14	-0.5	0	-1	3.3	0.5	24	13	21
547	1940	2	17	66	70	16	2	[illegible]	[illegible]	15	0	0	-1	3.7	1	24	14	18
549	1951	2	12	75	80	16	2	[illegible]	[illegible]	16	0.5	0	-1	4	1.5	22	13	19
551	1938	2	10	58	62	15	2	15	1.5	15	0	1	-0.5	3.4	0.5	27	12	26
555	1947	1	16	49	60	11	1	14	2	16	1	0	-1	4	2	42	6	29
556	1932	1	15	32	45	6	-0.5	6	-0.5	11	-2	5	0.5	1.5	-1	35	9	23
558	1933	2	18	61	68	14	1.5	14	1	16	0.5	0	-1	4	1.5	32	11	23
559	1944	1	16	50	62	7	-0.5	9	0	14	0	4	0	2.3	-0.5	19	16	17
560	1951	2	20	56	58	10	-0.5	12	0	16	0.5	0	-1	4	1.5	33	9	21
564	1935	2	10	67	74	11	0.5	13	1	15	0	1	-0.5	3.4	0.5			
566	1945	2	10	53	57	13	1	12	0.5	16	0.5	0	-1	4	1.5	38	8	27
567	1949	2	14	69	72	14	1	15	1.5	16	0.5	0	-1	4	1.5	34	8	22
569	1945	2	10	57	61	12	1	13	0.5	16	0.5	8	2	2.5	-1	38	8	30
573	1958	2	16	70	73	15	1.5	15	1.5	16	0.5	0	-1	4	1.5	23	13	19
577	1958	2	16	60	63	16	2	16	1.5	16	0.5	0	-1	4	1.5	26	11	19
579	1957	2	14	65	68	15	1.5	16	1.5	16	0.5	0	-1	4	1.5	27	11	17
581	1935	2	11	56	63	15	1.5	15	1.5	15	0	0	-1	3.7	1	28	12	20
583	1943	1	13	53	65	12	1	13	1.5	16	1	0	-1	4	2	35	9	26
586	1947	2	19	74	79	16	2	16	1.5	16	0.5	0	-1	4	1.5	32	10	19

ShortStatsJoined

Why biology is interesting for VIS?

CITY UNIVERSITY LONDON

Things happen at multiple **scales**



[by O'Donoghue et al., 2010]

[Nye, 2008]

Why biology is interesting for VIS?

CITY UNIVERSITY LONDON

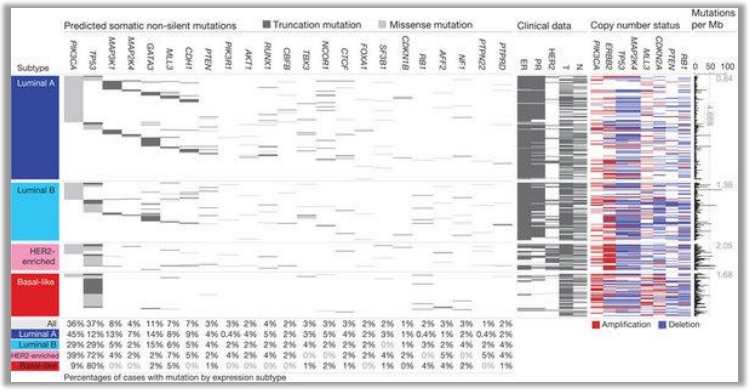
Processes are **dynamic** (spatio-temporal complexity)

Neutrophil chasing a bacteria by David Rogers

Why biology is interesting for VIS?

CITY UNIVERSITY LONDON

- Computational methods are central in analysis
  - **Uncertainties** hinder reliability
  - **Interpretation** is a problem (**black-box** alg., little context)



Comprehensive molecular portraits of human breast tumours, TCGA Network, Nature, 2012

How can visualisation help?

CITY UNIVERSITY LONDON

- Ease of **cognition & communication**
- **Relating** multiple aspects
- **Compare** multiple computational outputs
- Investigate **uncertainties**
- **Seamless integration** of computation

and …

- Enable & foster hypothesis generation

Forms of visualisation support

CITY UNIVERSITY LONDON

VIS as a **presentation medium**

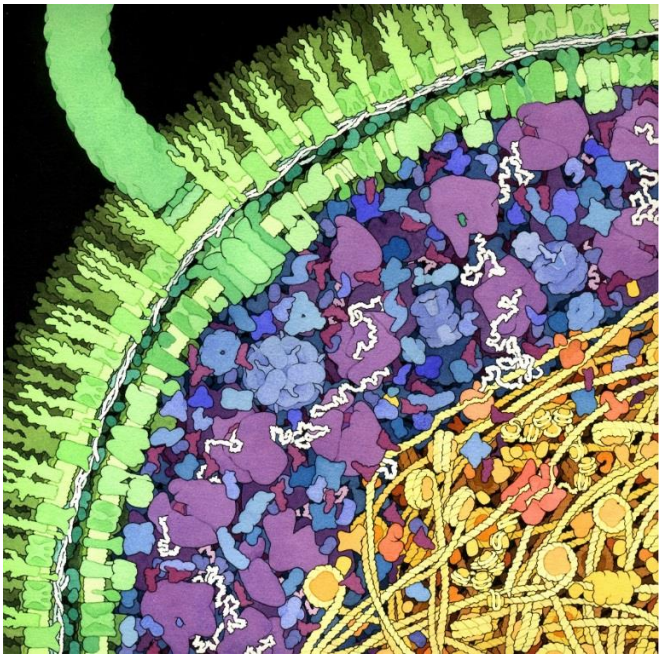
+

VIS with **interaction**

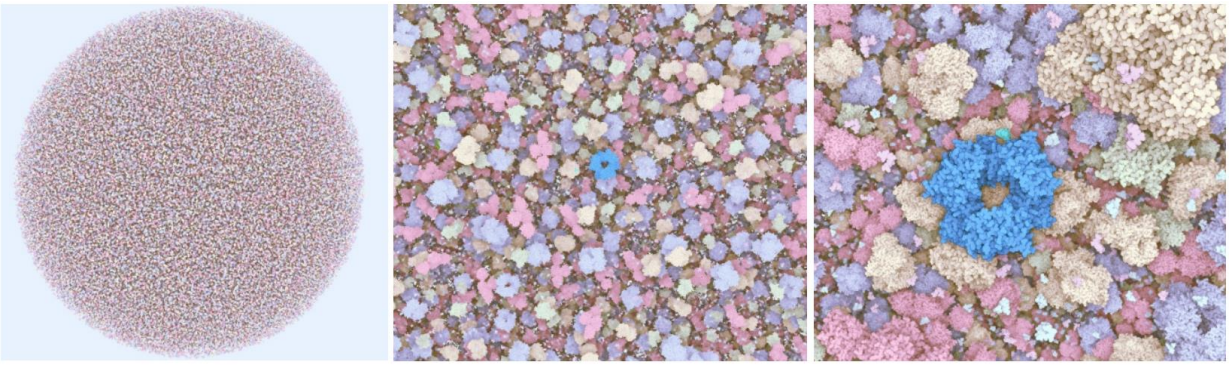
+

VIS with **integrated computations**

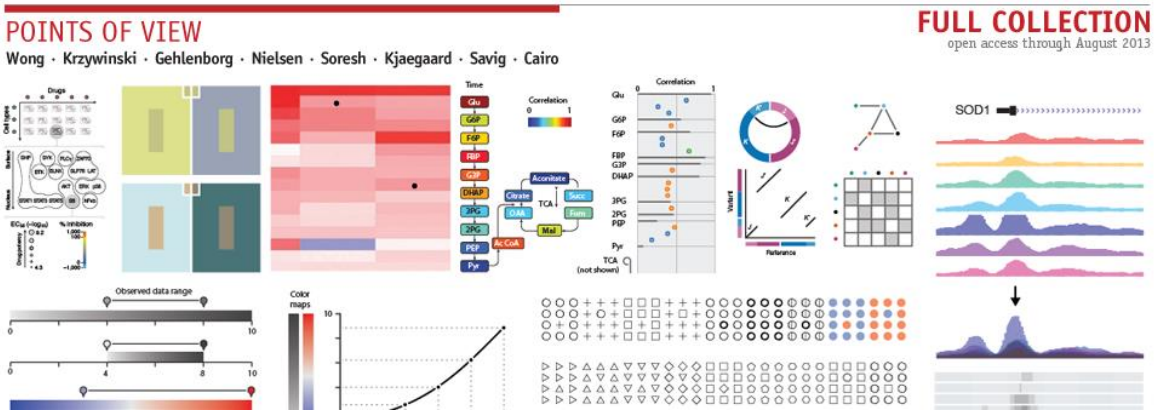
Visualisation as a presentation medium



Cross-section of Escherichia coli cell, Illustration by David S. Goodsell, the Scripps Research Institute



10^6 diffusing and reacting molecules in real-time, Muzic et al., 2014



NATURE METHODS: POINTS OF VIEW, by Wong et al.

http://blogs.nature.nom/methagora/2013/07/data-visualization-points-of-view.html

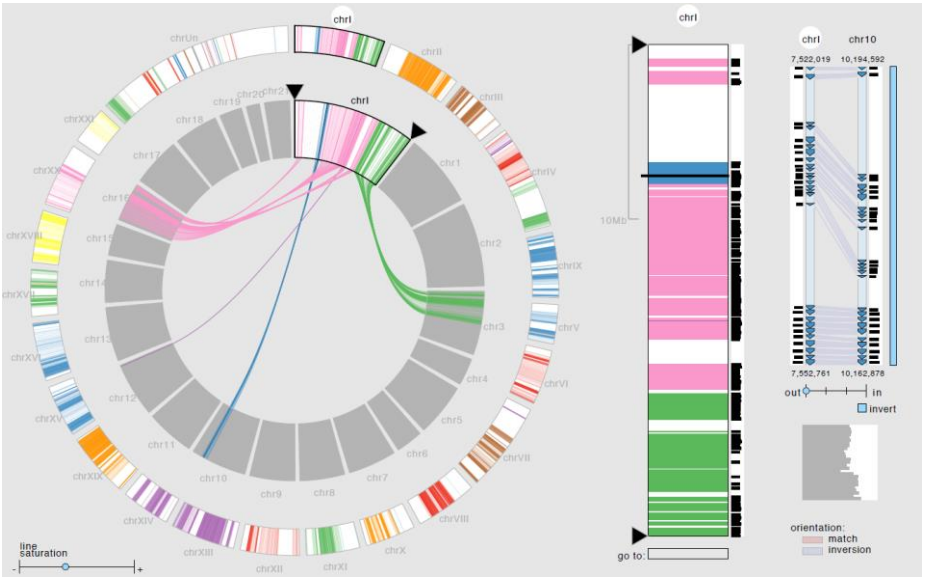
Why is VIS good here?

CITY UNIVERSITY LONDON

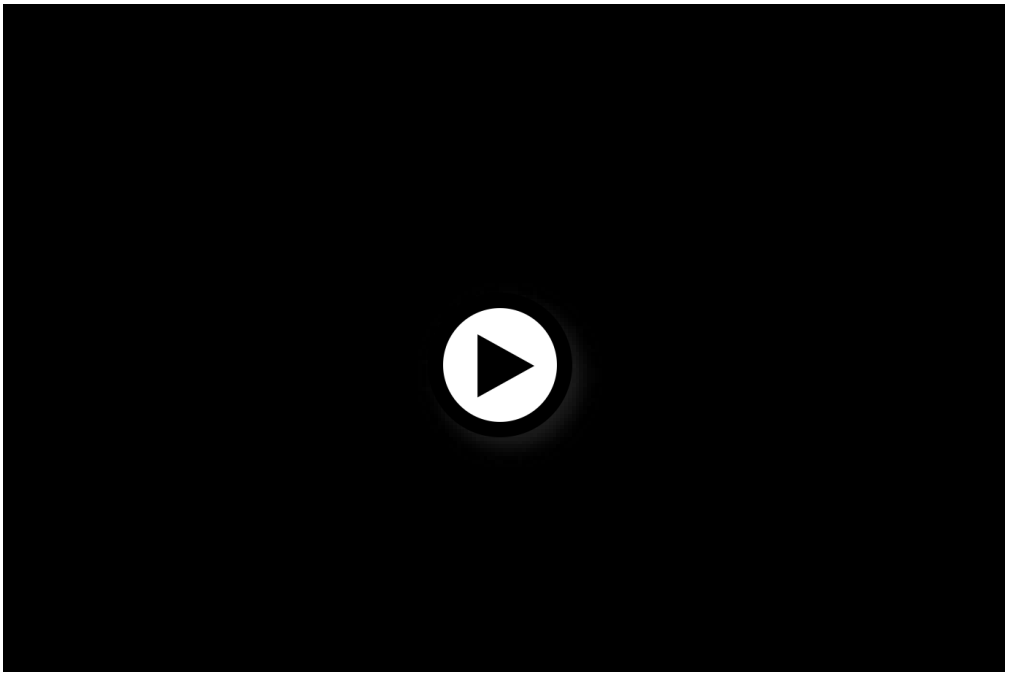
- Analysts' perceptual & cognitive capabilities
- Better interpretation
- Communication

Visualisation with interaction

Example: MizBee - Synteny Browser



Meyer et al., MizBee: A Multiscale Synteny Browser, 2009



Why is VIS good here?

CITY UNIVERSITY LONDON

- **Linking** multiple aspects
- **Interactively varying** the focus
- Display multiple-scales **concurrently**

Visualisation with integrated computations

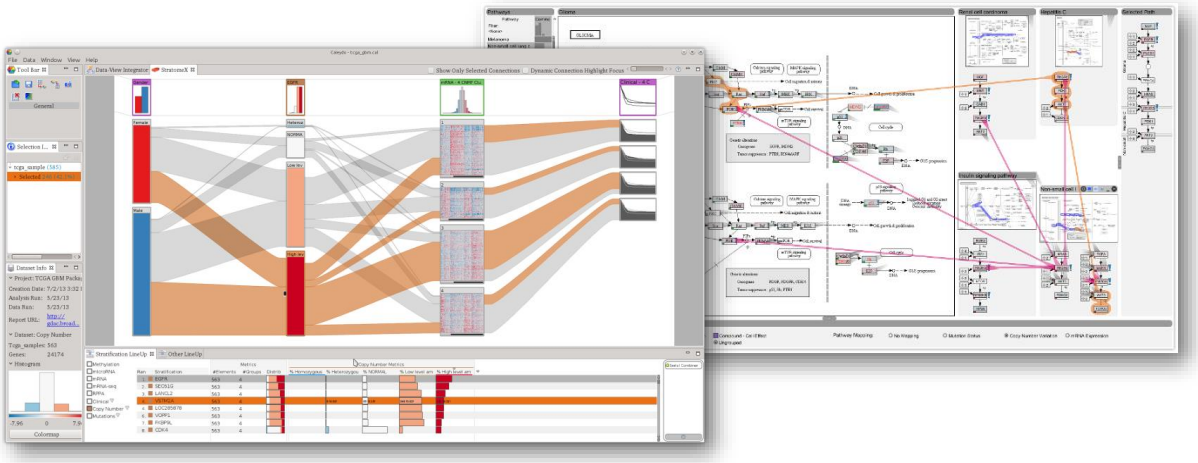
Combine the best of two worlds: **human capabilities** and **computing power**

Facilitate the **informed** use of computation through **interactive visual** methods

(a.k.a. Visual Analytics)

Example: StratomeX, Caleydo

CITY UNIVERSITY LONDON



CALEYDO http://caleydo.org

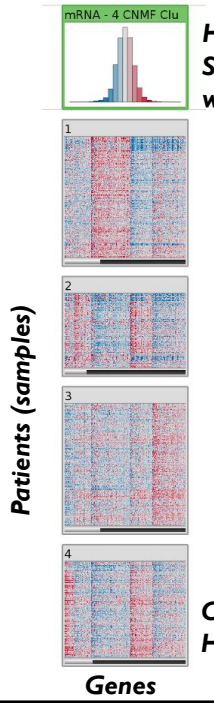
Case: Cancer Subtype Analysis

Cancers have **subtypes**
- different **histology**
- different **molecular alterations**

Subtypes are identified
by **stratifying** datasets, e.g.,
- based on an **expression pattern**
- a **mutation status**
- a **copy number alteration**
- a **combination** of these

The Cancer Genome Atlas — *Understanding genomics to improve cancer care*

FIREHOSE Broad GDAC

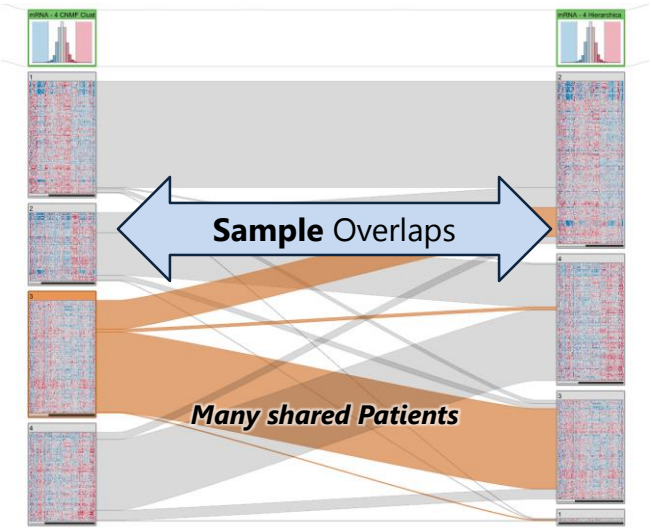


Header / Summary of whole Stratification

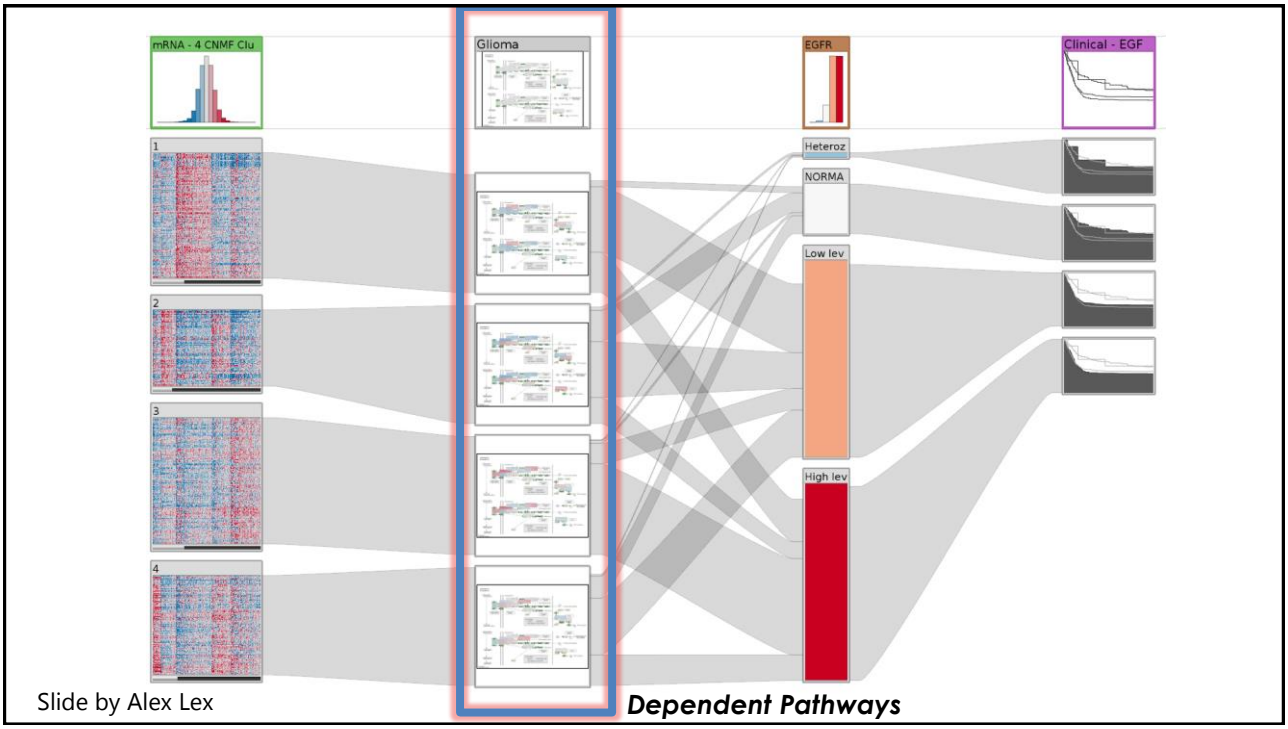
Candidate Subtype / Heat Map

Multiple Stratifications

CITY UNIVERSITY LONDON

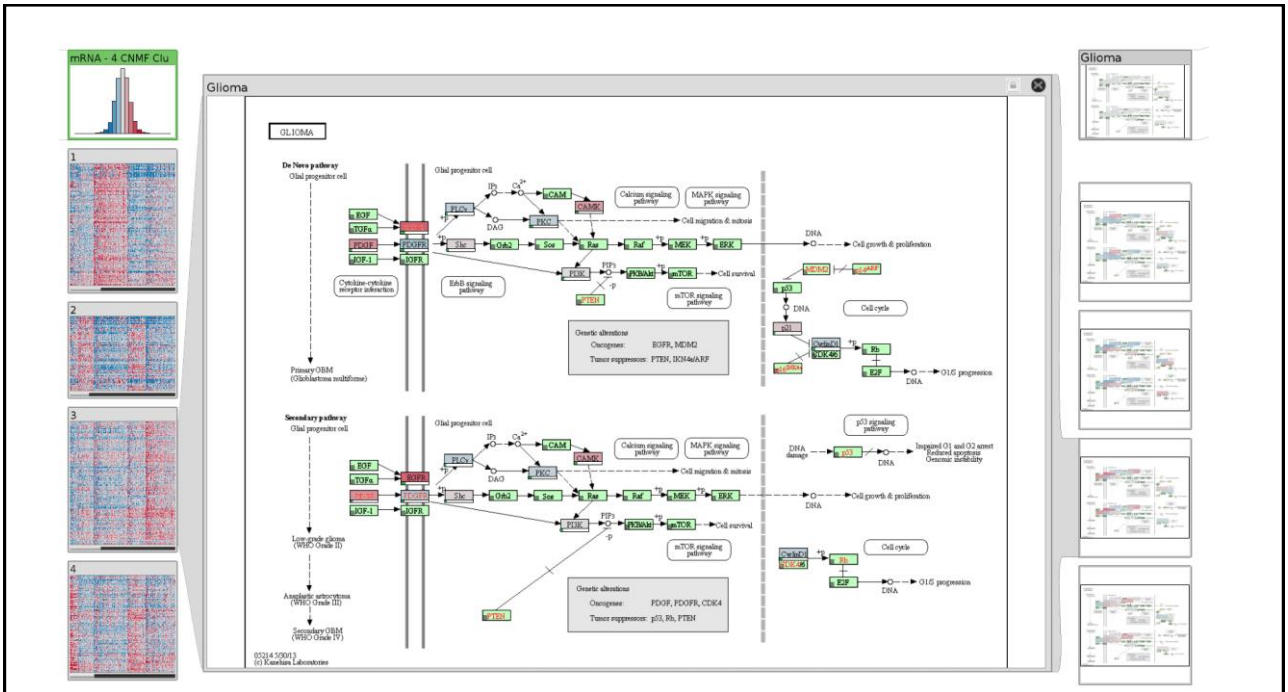


Clustering 1 ***Clustering 2***



Slide by Alex Lex

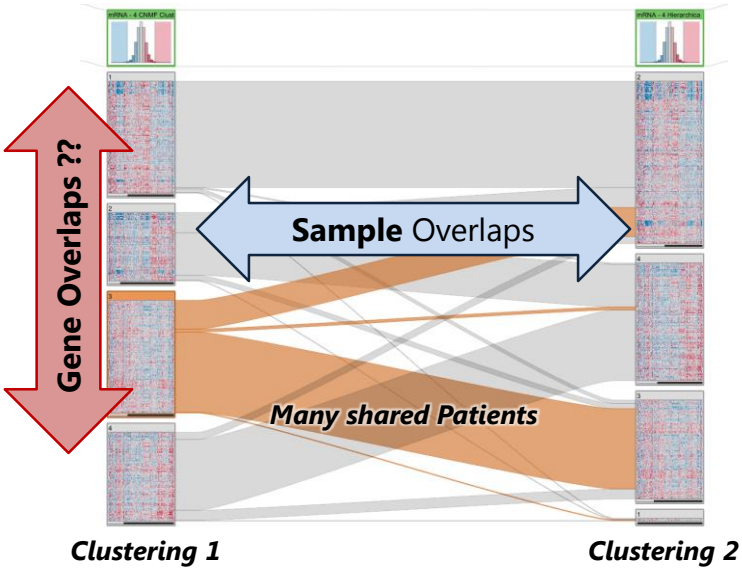
Dependent Pathways



Slide by Alex Lex

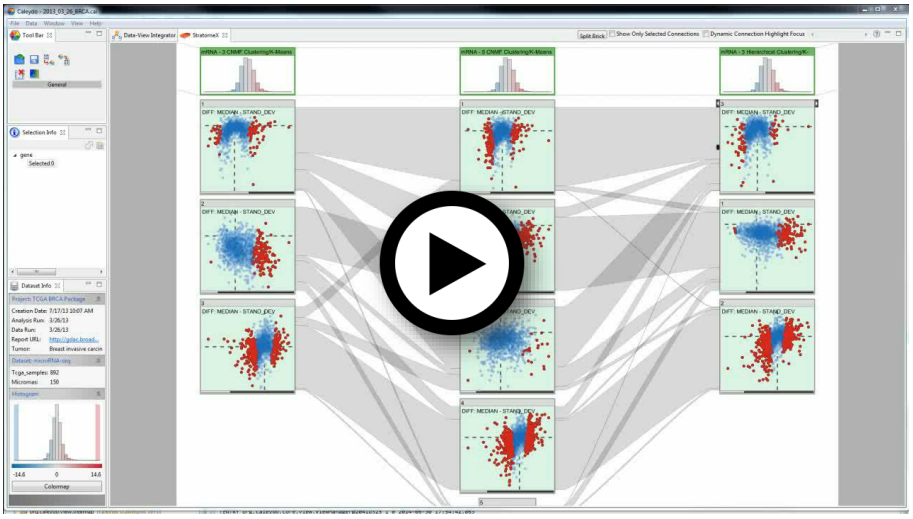
Multiple Stratifications (again)

CITY UNIVERSITY LONDON



Finding distinctive genes

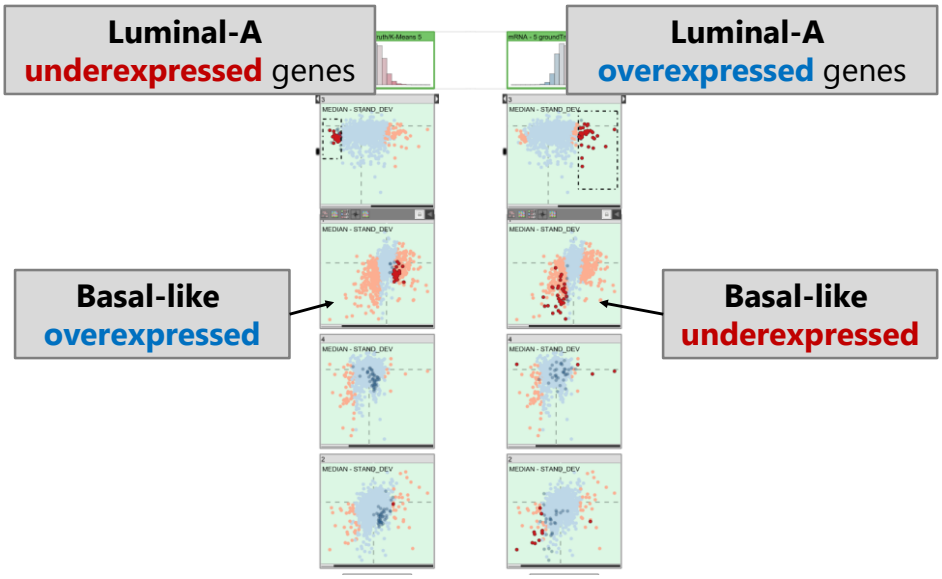
CITY UNIVERSITY LONDON



Characterizing cancer subtypes using dual analysis in Caleydo StratomeX, Turkay et al., IEEE CG&A, 2014

Finding distinctive genes (ex. BRCA types)

CITY UNIVERSITY LONDON



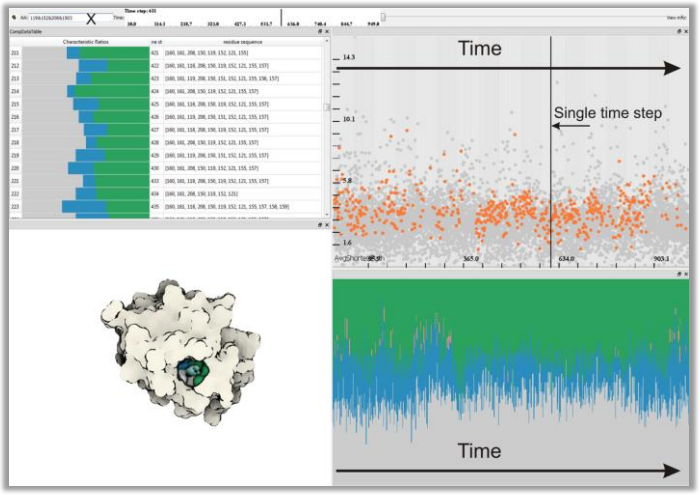
[*] Cancer Genome Atlas Network. (2012). Comprehensive molecular portraits of human breast tumours. *Nature*, *490*(7418), 61-70.

Ex: Cavity analysis in molecular simulations

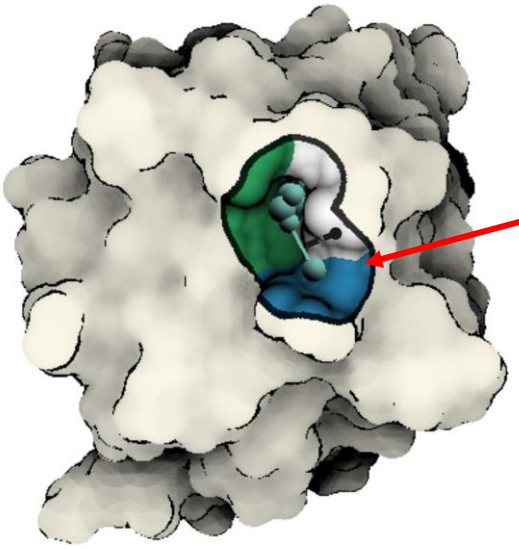
CITY UNIVERSITY LONDON

Cavities on molecular surfaces
- Important in ligand binding
- Drug design, etc.

Long **molecular simulations**
Cavities are **dynamic**, hard to track
Amino-acids to **characterize** the cavity
- hydrophobicity (grey)
- polarity (green)
- positively charged (blue)
- negatively charged (red)

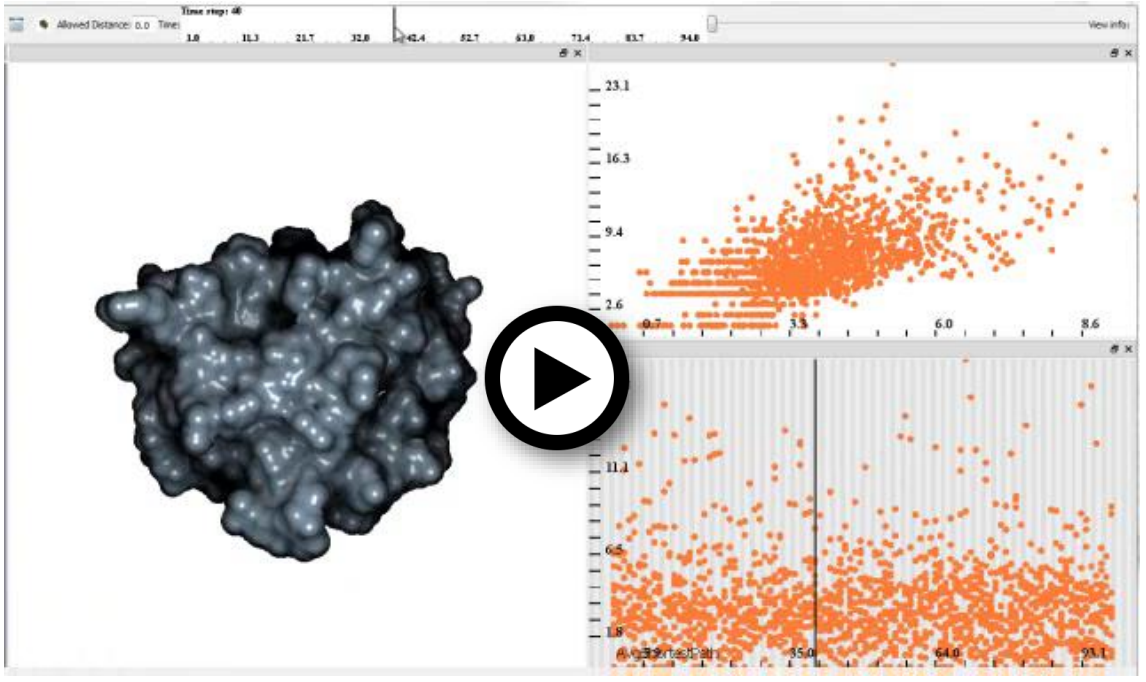


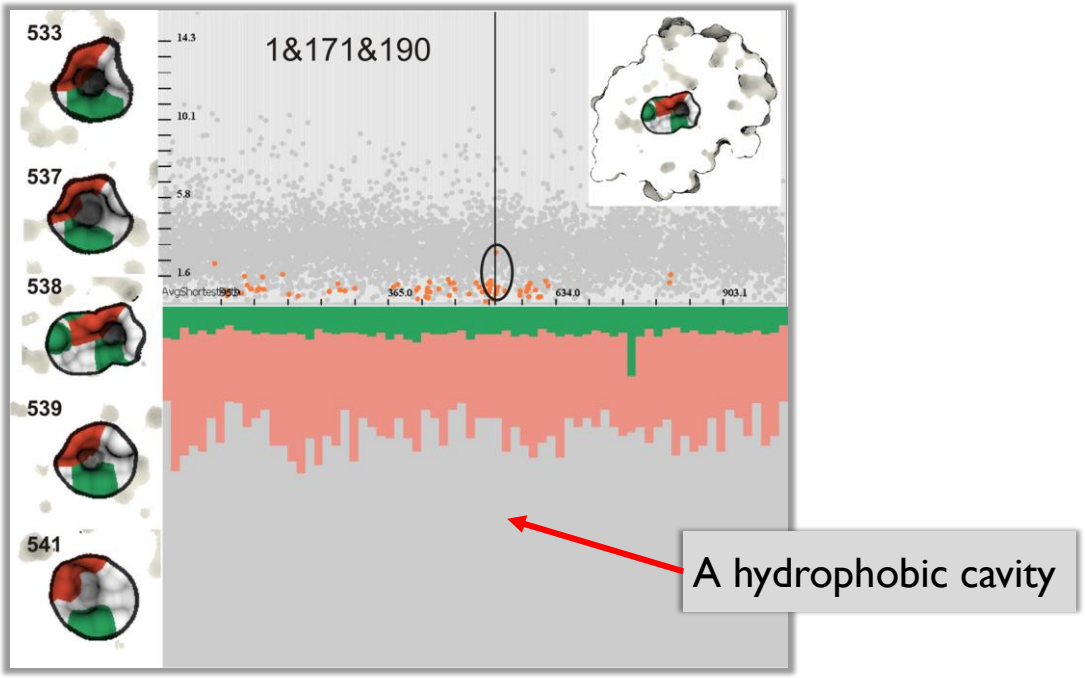
Visual Cavity Analysis in Molecular Simulations
J. Parulek, C. Turkay, N. Reuter, I. Viola. BMC Bioinformatics, 2013.



1. Run the simulation
2. Fit graphs cavities
3. Compute measures
4. Find touching amino-acids
5. Perform visual analysis

Analysis of Proteinase 3





Why is VIS good here?

CITY UNIVERSITY LONDON

- **Multiple linked data sets** – improve **interpretation**
- Multiple computational results – deal with **uncertainty**
- Integrate computation outputs, i.e., clusters, derived data
- Allows a fast-paced **iterative process**
- Quick **idea prototyping**

Wrap up !

CITY UNIVERSITY LONDON

VIS as a **presentation medium**

+

VIS with **interaction**

+

VIS with **integrated computations**

Visualisation is very good to answer

HOW & WHY?

questions ..

- How do these genomes overlap?
- Why is this a cluster?

....

Outlook

CITY UNIVERSITY LONDON

- **Interaction** and **explorative** analysis is key!
- **Seamless support** from **integrated** computation, *i.e., t-tests*
- Visual analysis as an everyday tool for analysts

Thanks ! (& more biovis ?)

- VisGroup (**Helwig Hauser**, **Julius Parulek** & **Ivan Viola**) and **Nathalie Reuter** from University of Bergen
- Caleydo team (**Alex Lex**, **Hanspeter Pfister, Nils Gehlenborg, Marc Streit**)

biovis

5th Symposium on Biological Data Visualization

10- 11 July // Dublin, Ireland @ ISMB/ECCB 2015

http://www.biovis.net
#biovis
Paper deadline: **February 15, 2015**
Data & Design Contests: **May 1, 2015**