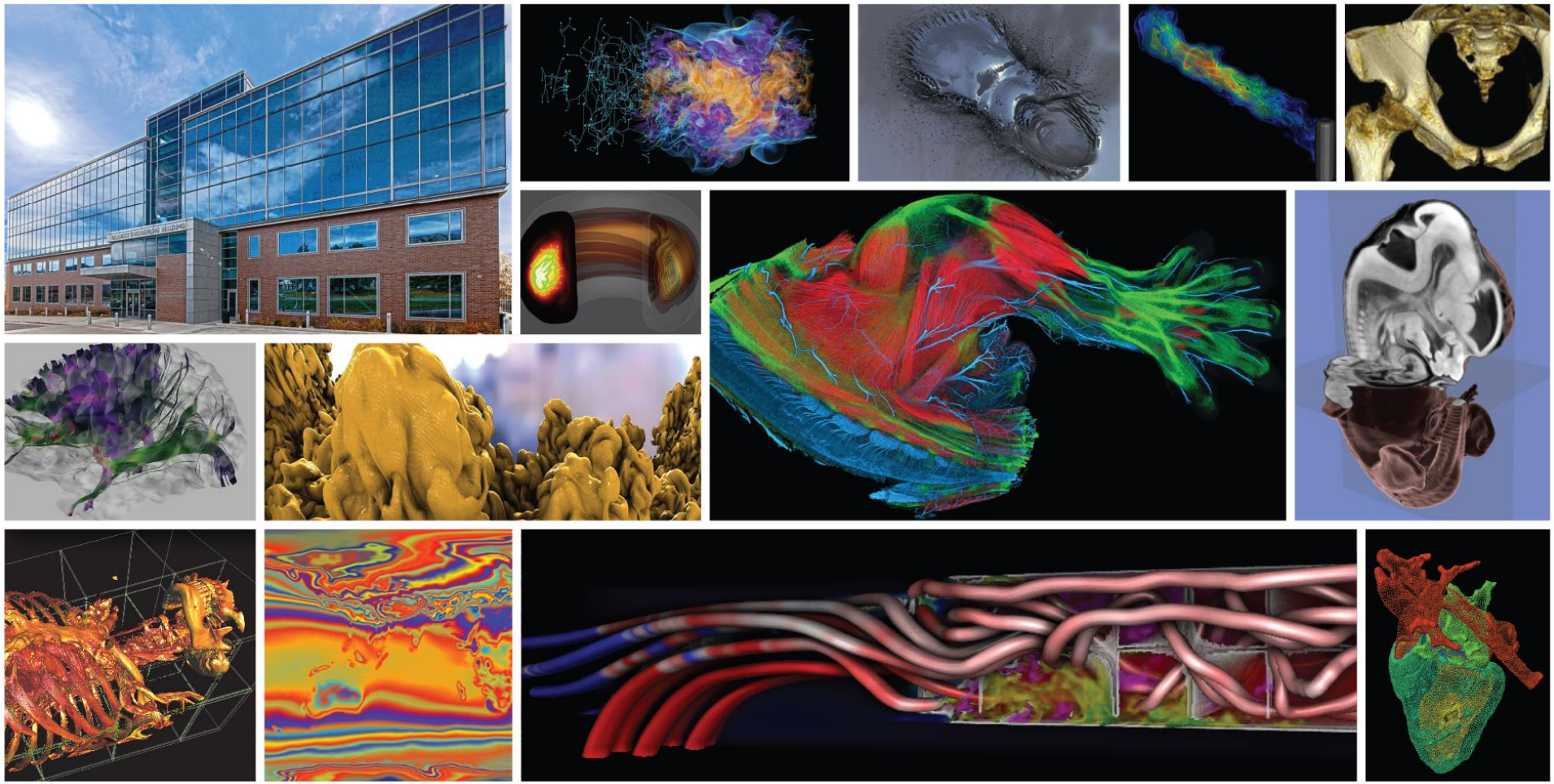
Big Data Visual Analysis



Chris Johnson
Scientific Computing and Imaging Institute
University of Utah

History of Computer Graphics in Utah

1, 2. David Evans /Ivan Sutherland
-Founded CS Dept at the UofU in 1968
-Ivan Sutherland - Turing award
-Founded Evans & Sutherland Company

3. John Warnock
-Worked at Evans & Sutherland
-Founded Adobe
-Hidden Line Removal Algorithm
-Helped invent Postscript @ Adobe

4. Tom Stockham
-Known for work in Signal Processing
-Helped to invent the CD Player

5. Ed Catmull
-Worked at Lucas Film
-Co-Founded Pixar
-President of Disney Animation Studios
-Chair of CoE External Advisory Board

6. Alan Kay
- Personal Computer
-Turing Award Winner
-Object Oriented Languages

7. Jim Kajyia
-VP Research at Microsoft

8. Jim Clark
-Founded SGI, Netscape, Healtheon
- Work in Geometry Pipelines

9. Jim Blinn
-Invented Blinn-Phong Shading Model

10. Nolan Bushnell
-Invented Pong
-Founded Atari

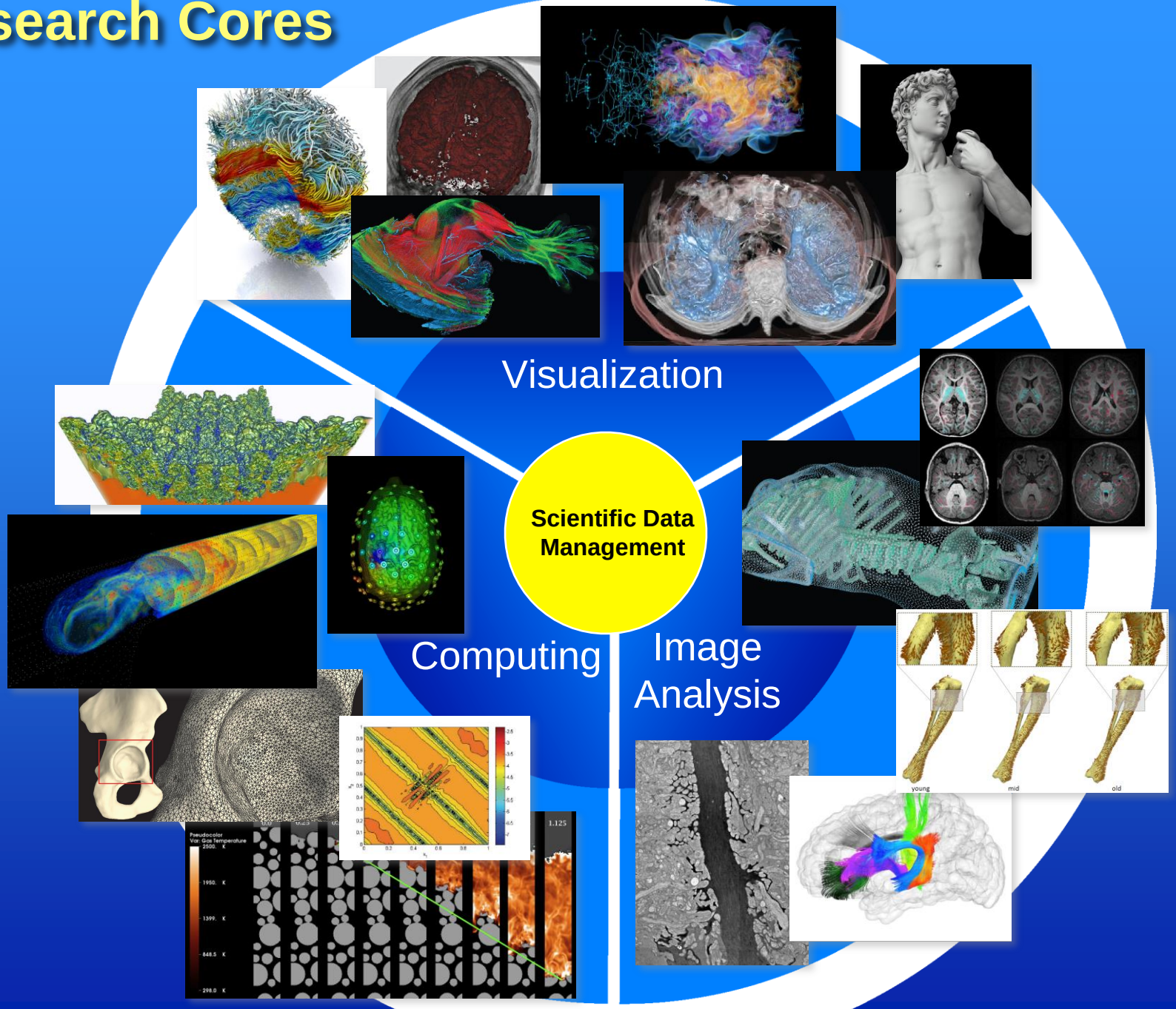
11. Henri Gouraud
-Invented Gouraud Shading Model

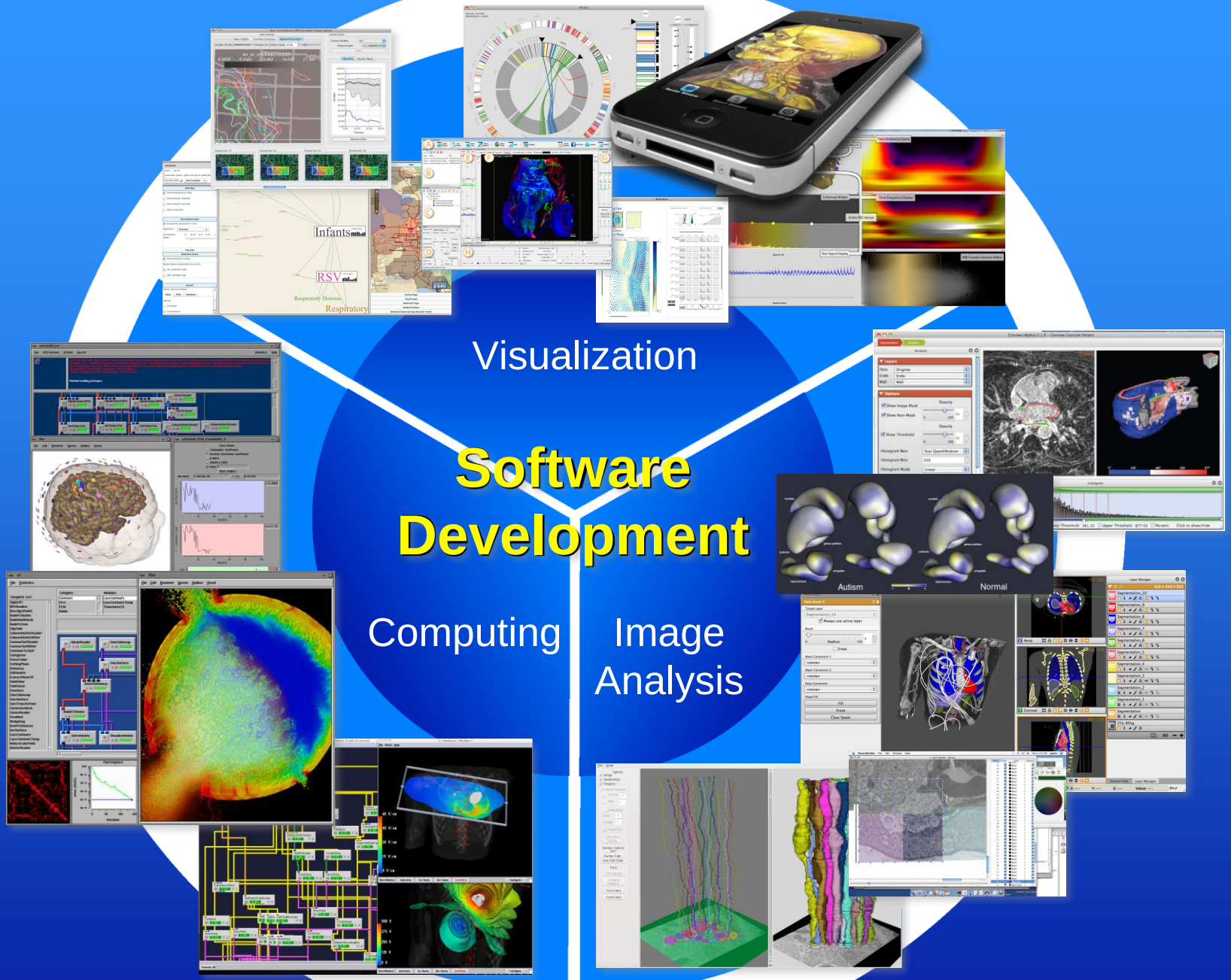
12. Allen Ashton
-Word Perfect
-My CFO Founder

13. Bui Tuong Phong
-Invented Phong Reflection and Shading Models

SCI Institute Faculty

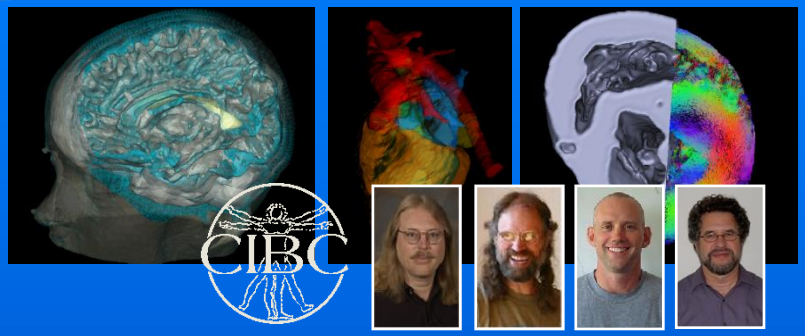
Research Cores




Visualization
Software Development
Computing
Image Analysis
Infants
RSV
Respiratory Distress
Respiratory
Autism
Normal

Centers We Direct

NIH/NIGMS Center for Integrative Biomedical Computing



Utah Center for Neuroimage Analysis

UTAH Center for Computational Earth Sciences

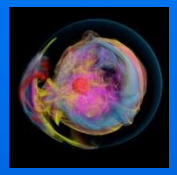
Center for Extreme Data Management, Analysis, and Visualization

CEDMAV

National Centers We are Affiliated With

SDAV

Scalable Data Management, Analysis and Visualization


NIH NAMIC

NIH NAMIC

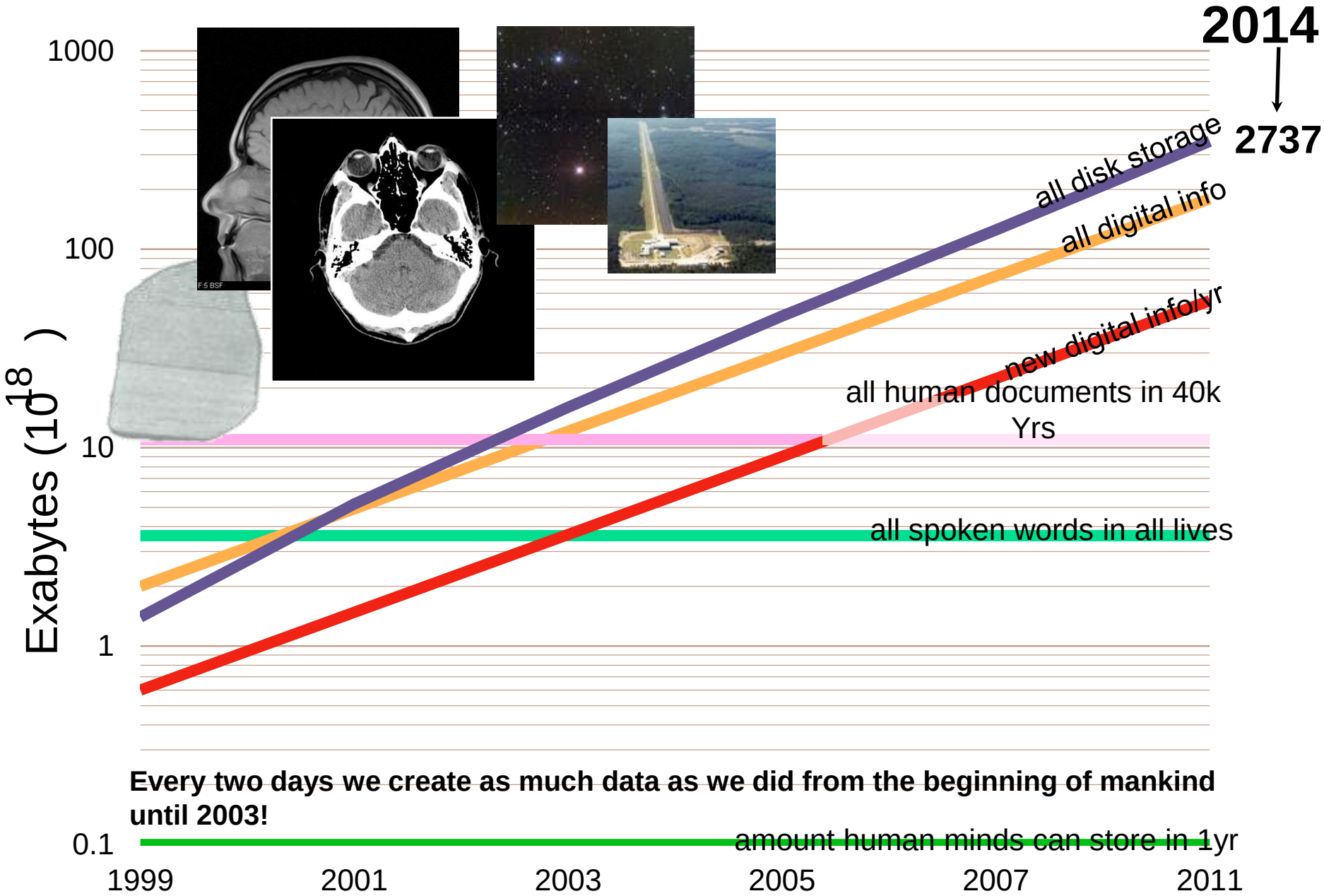
KAUST

IAMCS
Institute for Applied Mathematics and Computational Science

CDC Decision-Support for Infectious Disease Epidemiology

ExaCT

Center for Exascale Simulation of Combustion in Turbulence



Sources: Lesk, Berkeley SIMS, Landauer, EMC, TechCrunch, Smart Planet

How Much is an Exabyte?



How many trees does it take to print out an Exabyte?

1 Exabyte = 1000 Petabytes = could hold approximately 500,000,000,000,000 pages of standard printed text

It takes one tree to produce **94,200** pages of a book

Thus it will take **530,785,562,327** trees to store an Exabyte of data

In 2005, there were **400,246,300,201** trees on Earth

We can store **.75** Exabytes of data using all the trees on the entire planet.

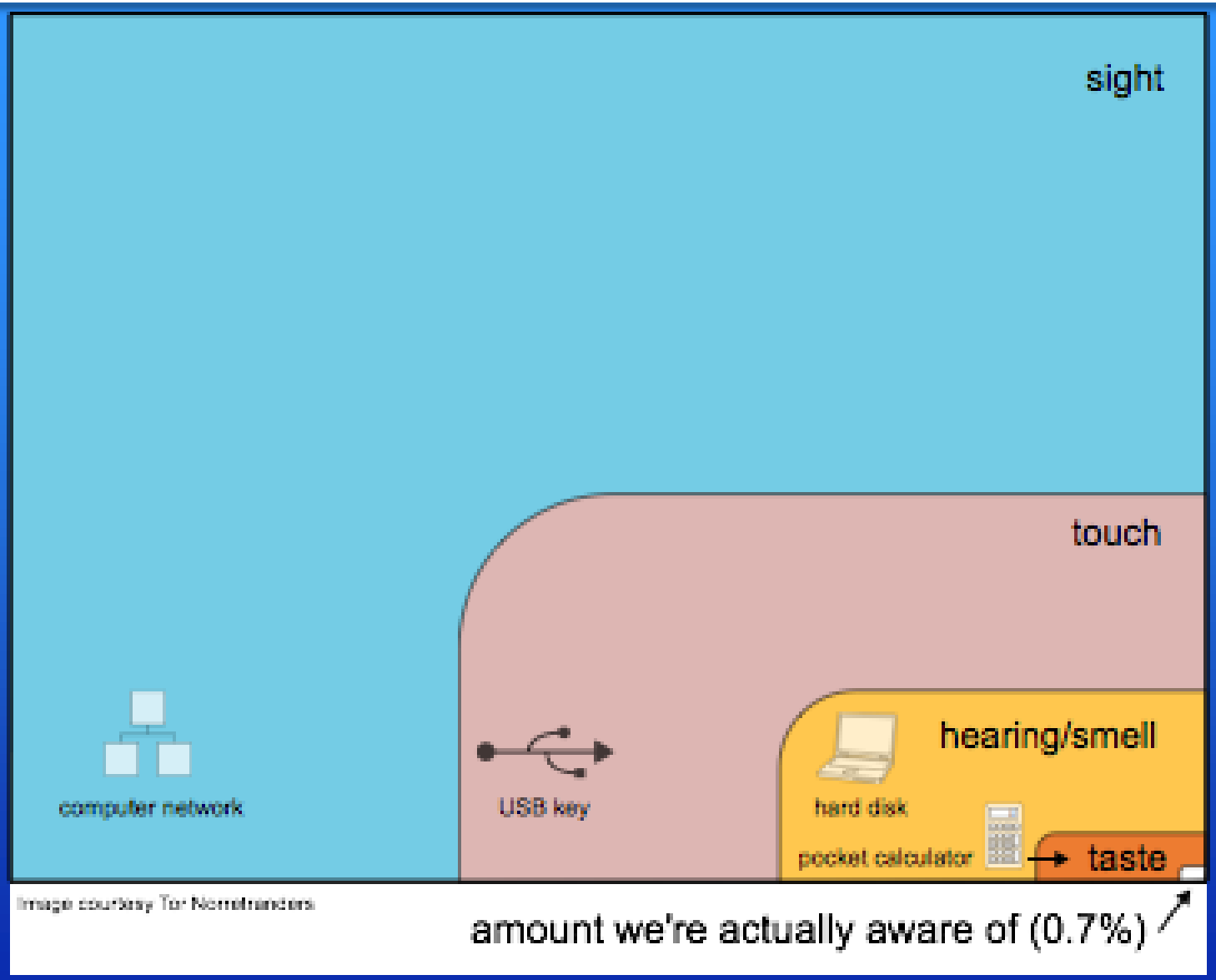
Sources: http://www.whatsabyte.com/ and http://wiki.answers.com

Big Data

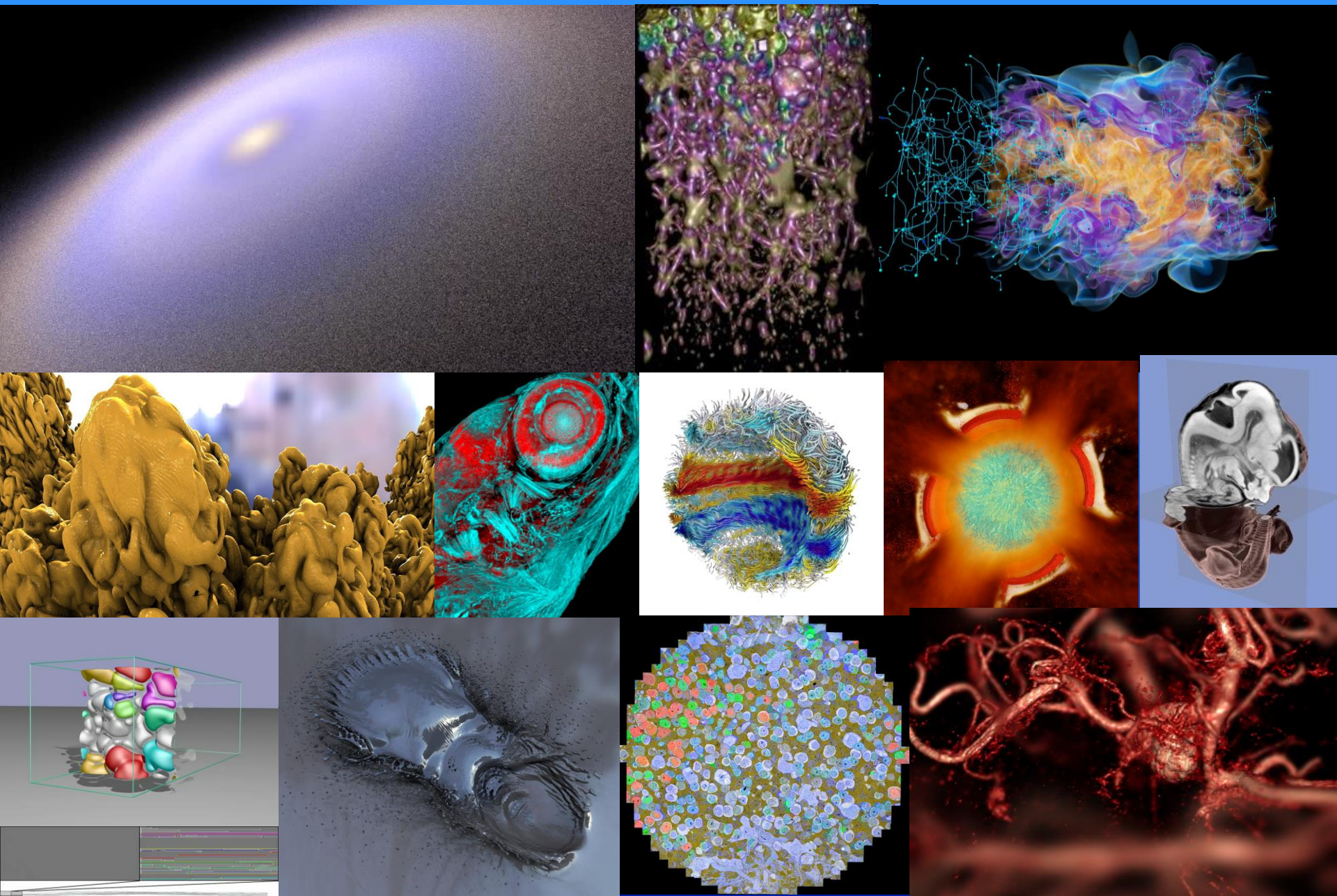
Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it...

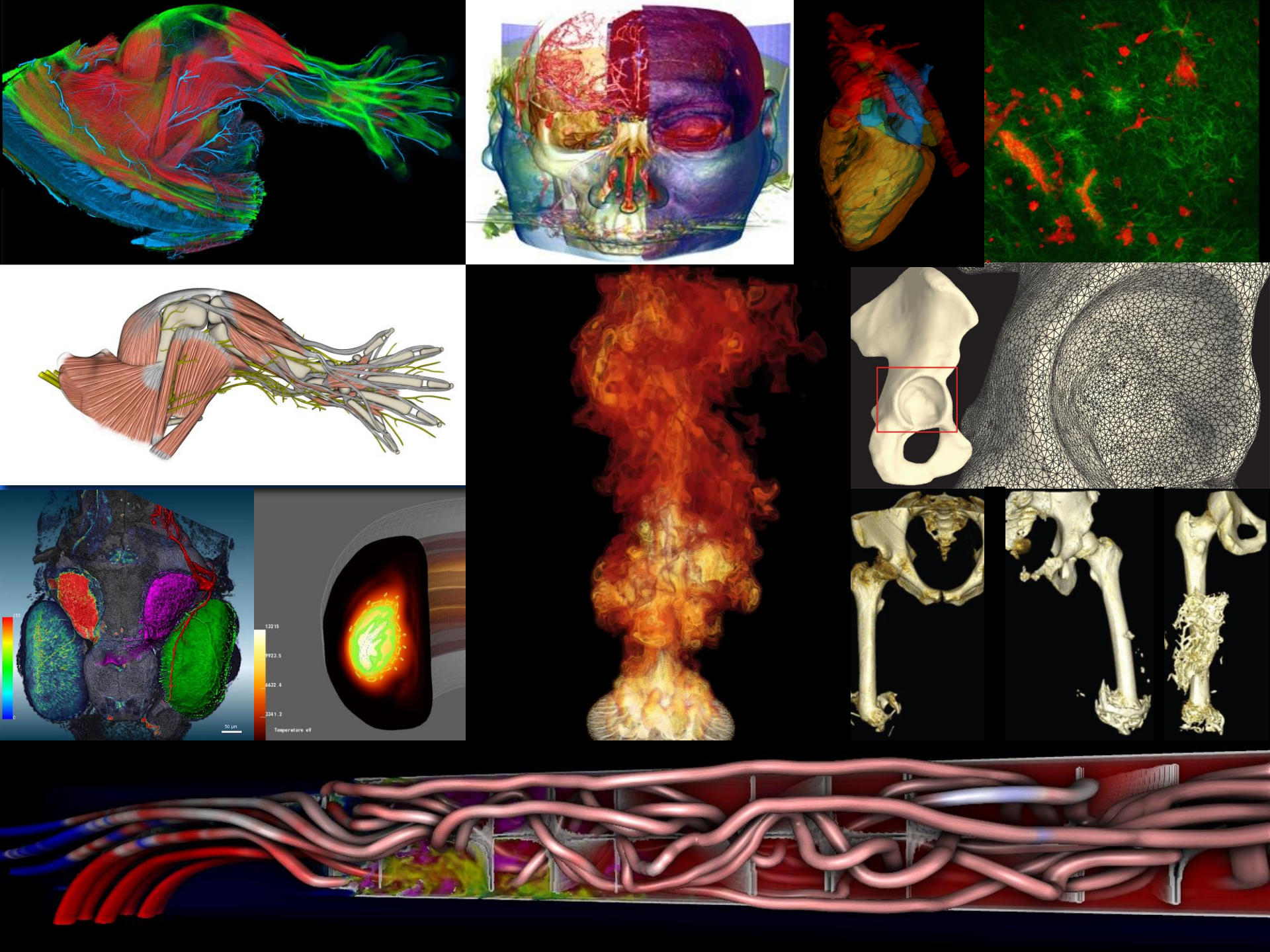
Dan Ariely

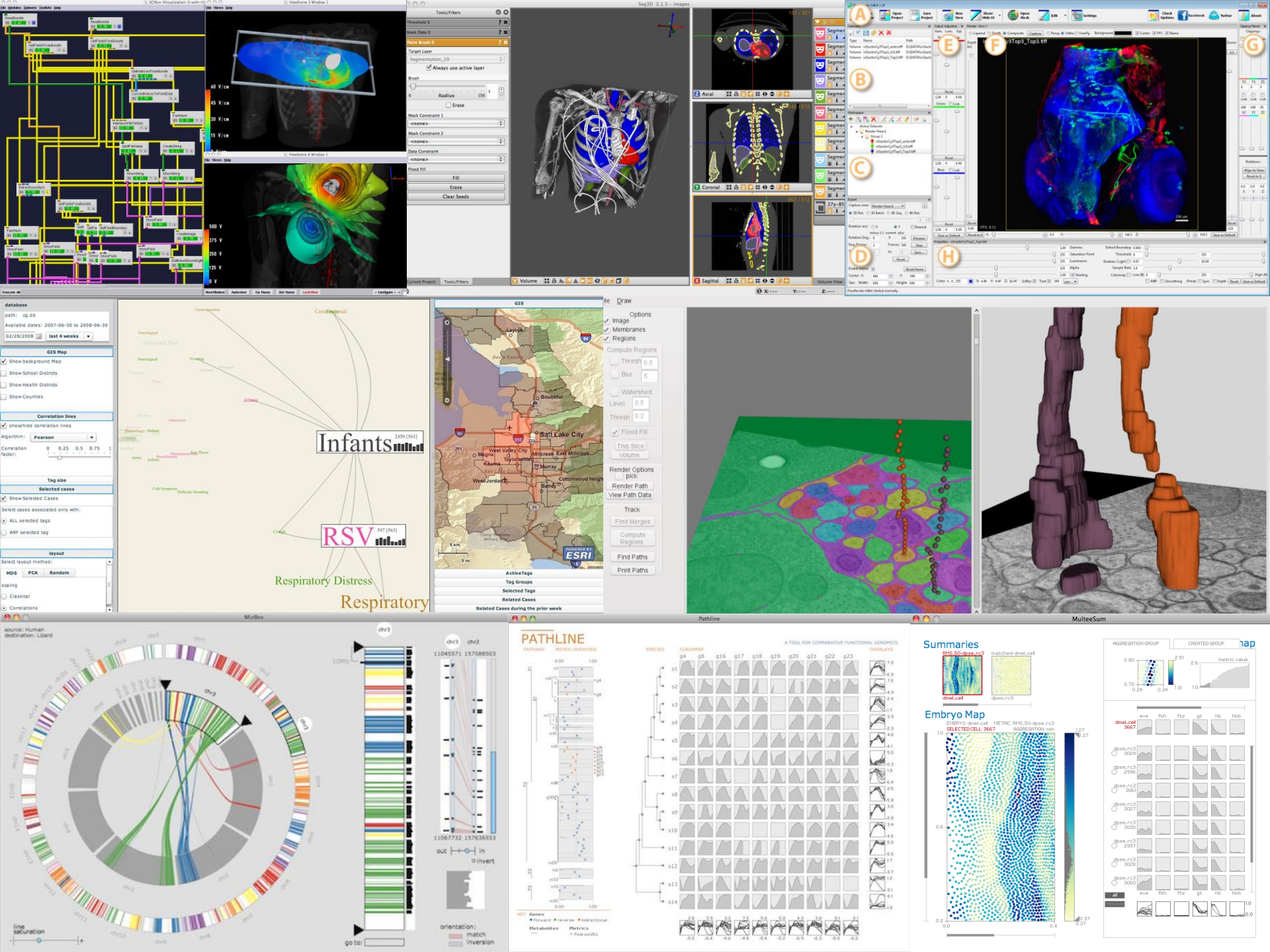
Brain Information Bandwidth



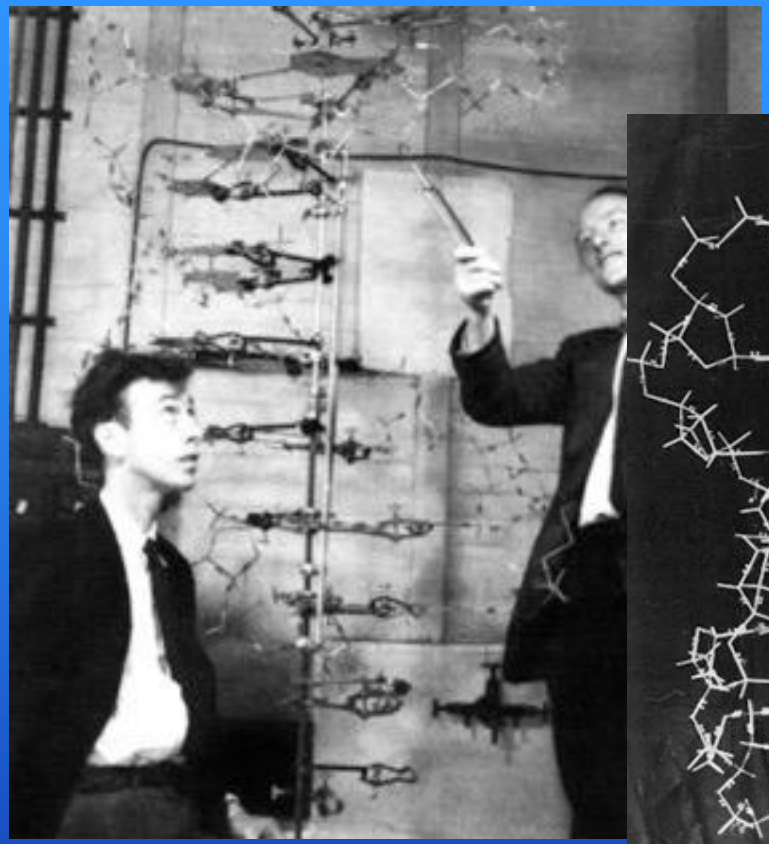
New Visual Analysis Techniques

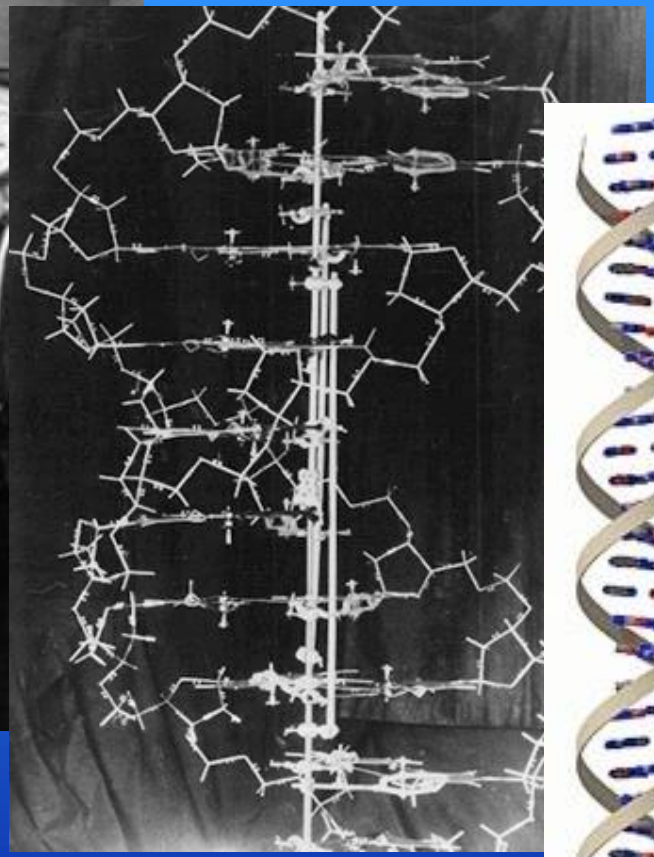


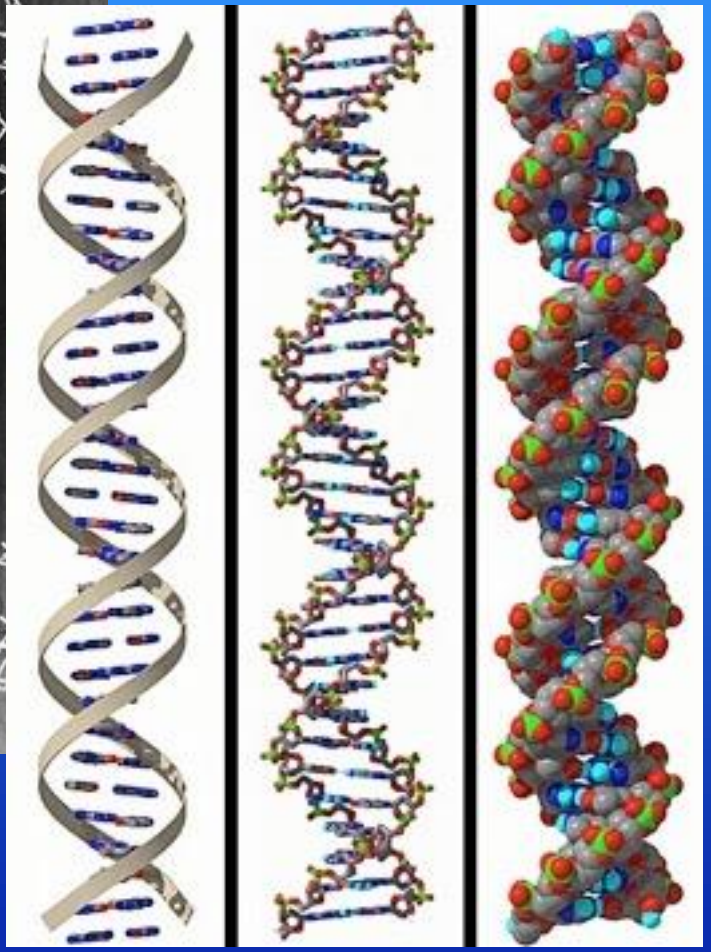

50 µm
13215
9923.5
6632.4
3341.2
Temperature eV


Infants
RSV
Respiratory Distress
Respiratory
GIS
Salt Lake City
PATHLINE
Summaries
Embryo Map
MulteeSum

Watson and Crick - DNA

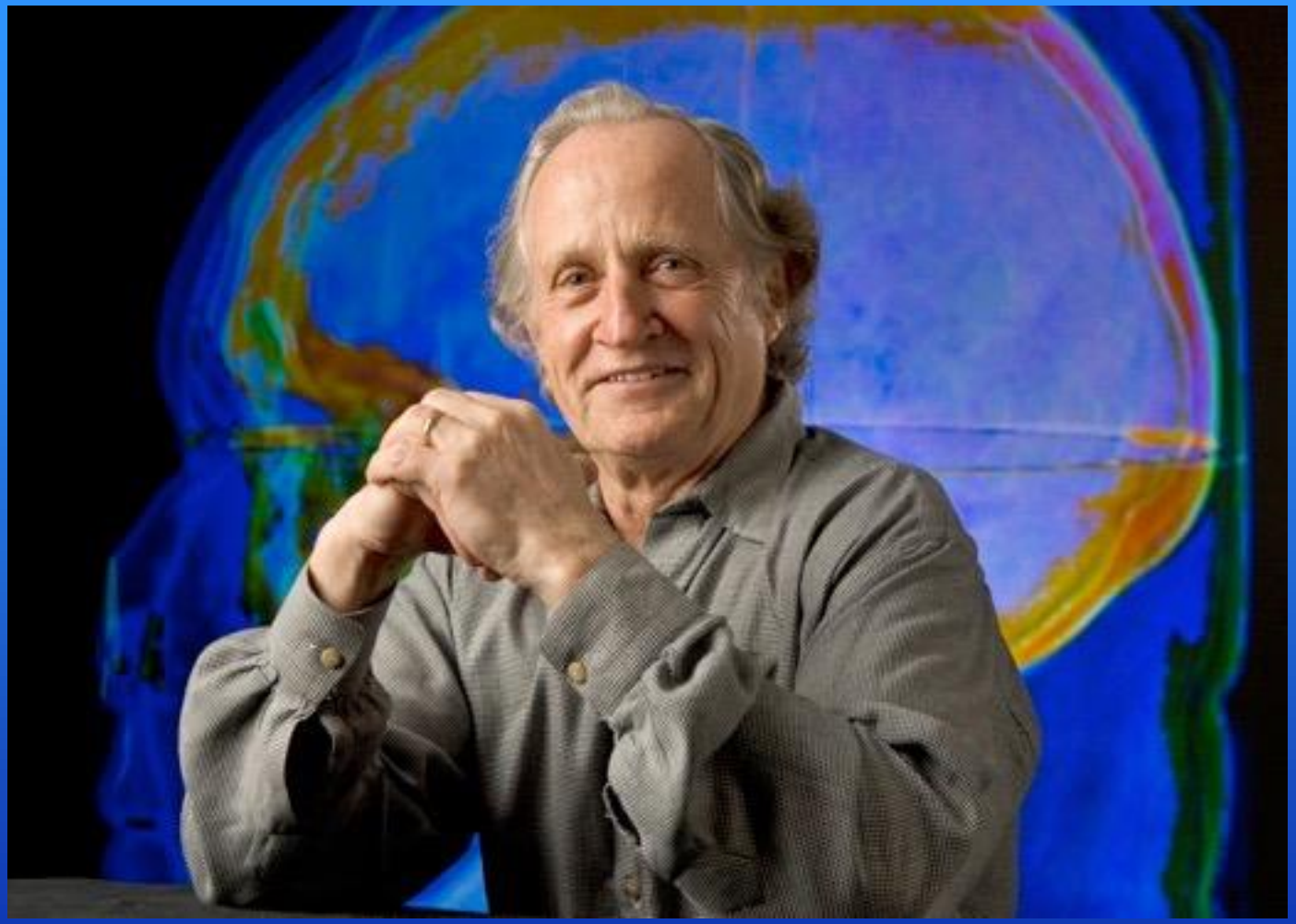






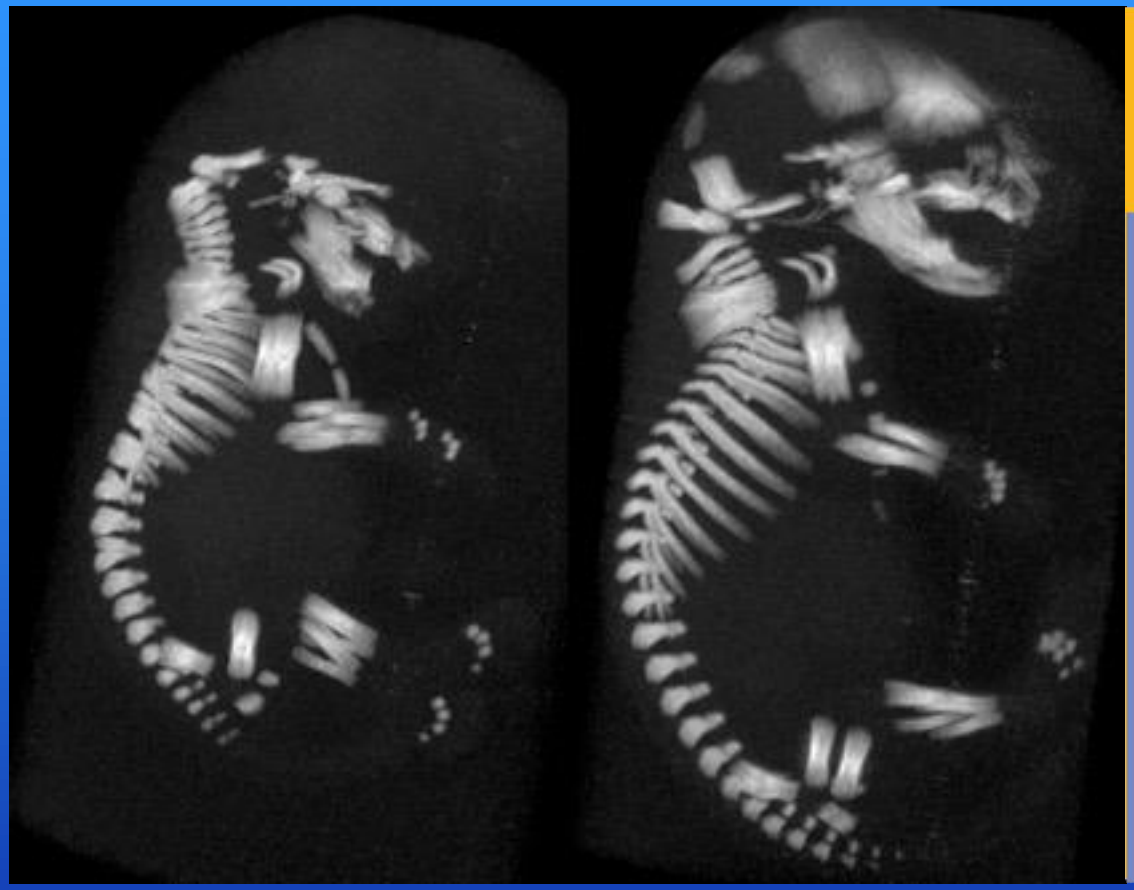
James Watson and Francis Crick - 1953
Nobel Prize - 1962
X-ray diffraction data from Maurice Wilkins and Rosalind Franklin

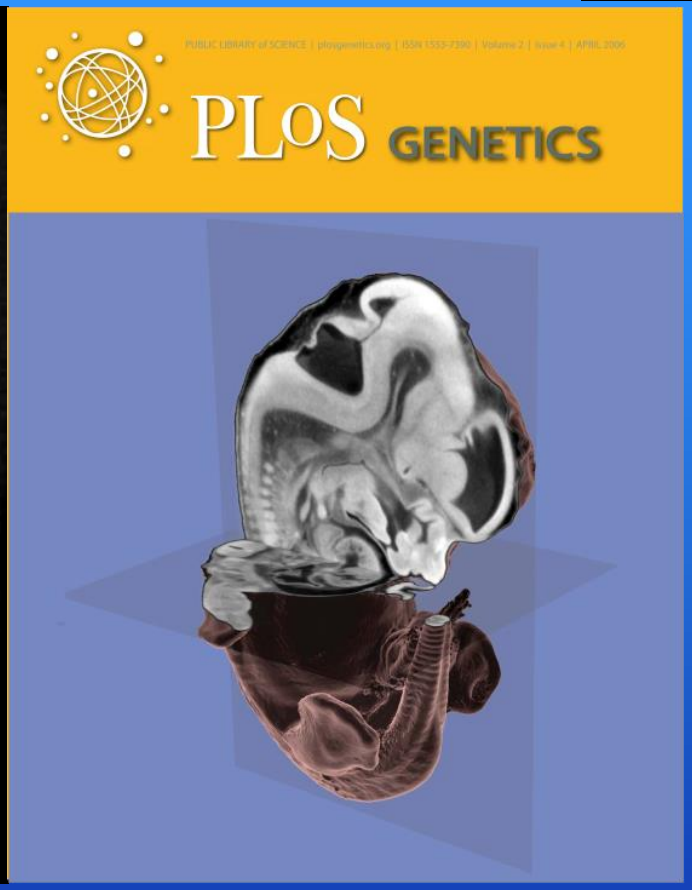
Mario Capecchi - Nobel Prize 2007



Mario R. Capecchi, Ph.D., distinguished professor of human genetics and biology at the University of Utah's Eccles Institute of Human Genetics has won the 2007 Nobel Prize in Physiology or Medicine.

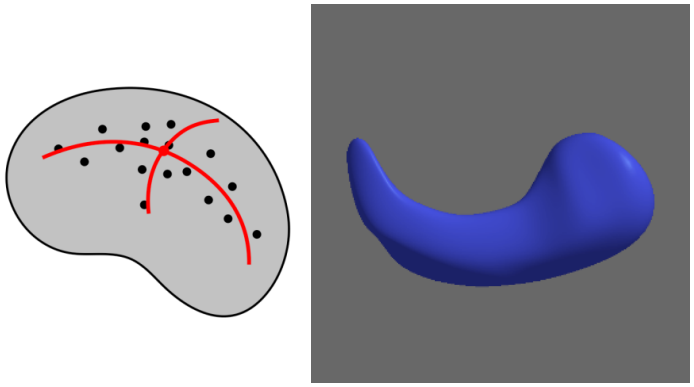
Image Based Phenotyping



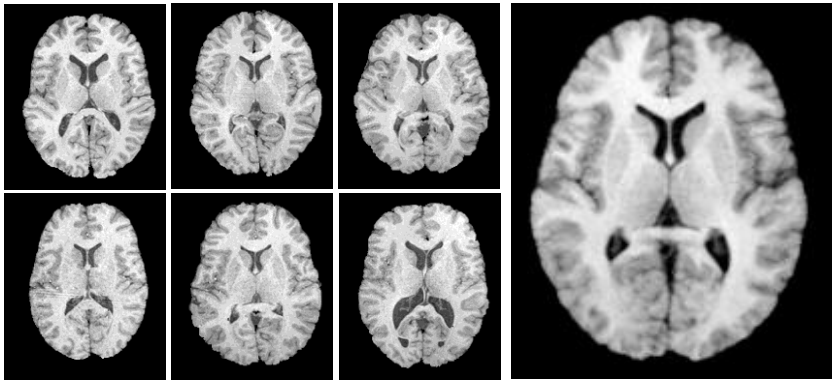


J.T. Johnson III, M.S. Hansen, I. Wu, L.J. Healy, C.R. Johnson, G.M. Jones, M.R. Capecchi, C. Keller.
"Virtual Histology of Transgenic Mouse Embryos for High-Throughput Phenotyping,"
In PLoS Genetics, Vol. 2, No. 1, pp. 471-477. April, 2006.

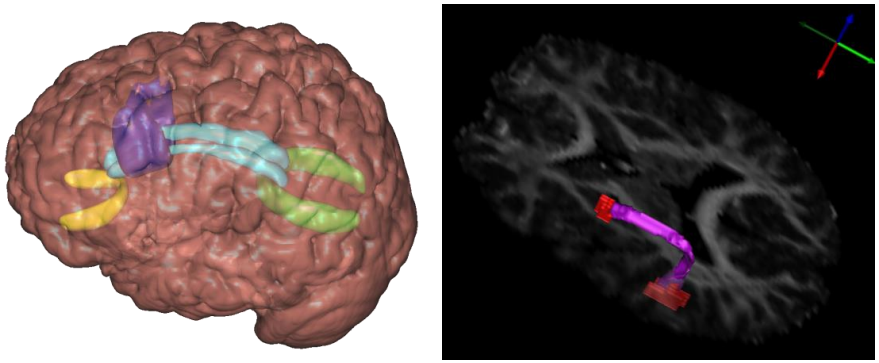
Statistics of Shape, Connectivity, and Function



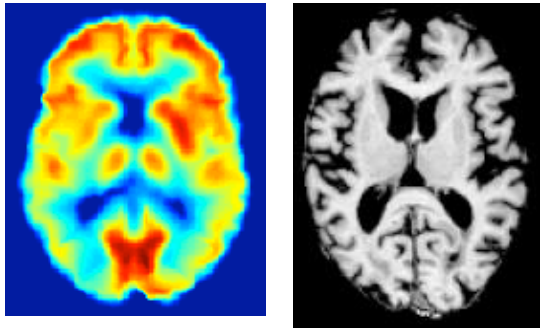
Computational Statistics in Nonlinear Spaces



Anatomical shape averaging and variability



Diffusion Tensor Image Analysis Autism project

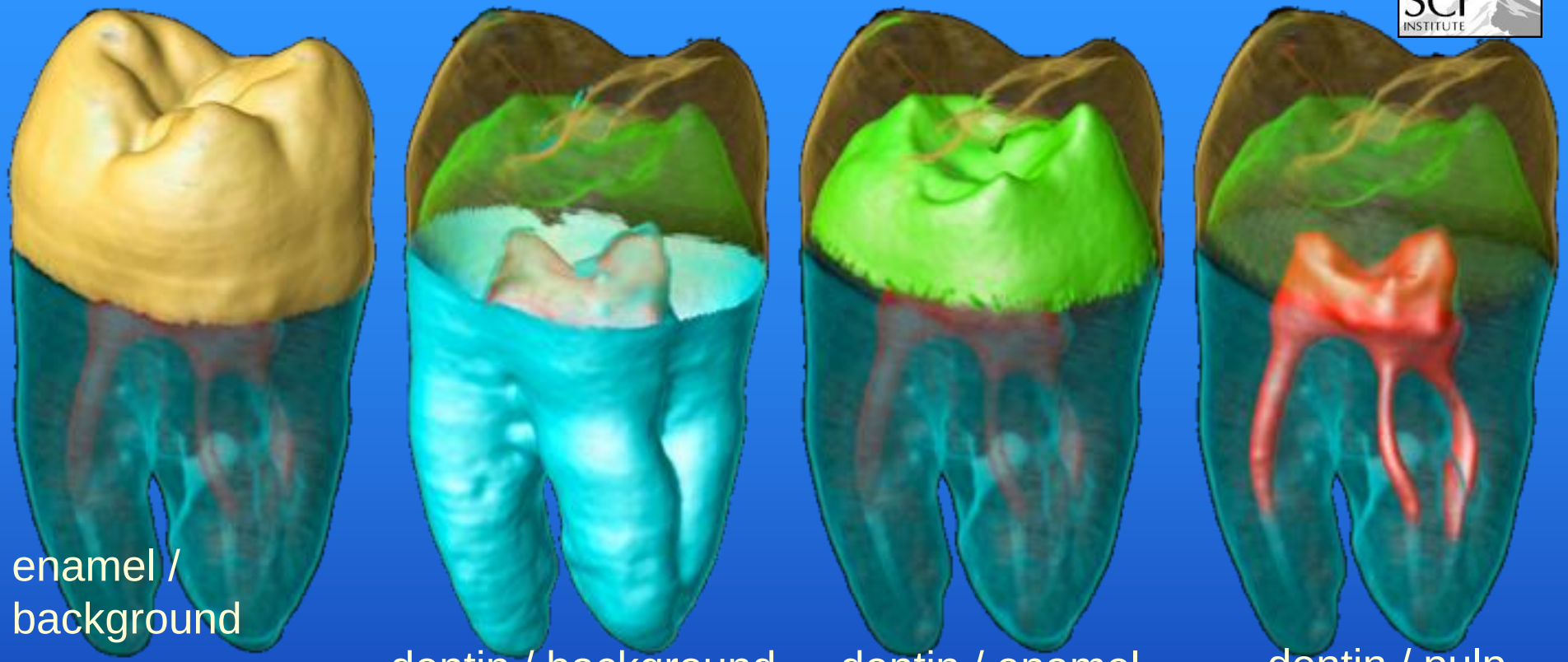


Combined PET + MRI analysis Alzheimer's disease project

SCI INSTITUTE

THE BRAIN INSTITUTE
THE UNIVERSITY OF UTAH

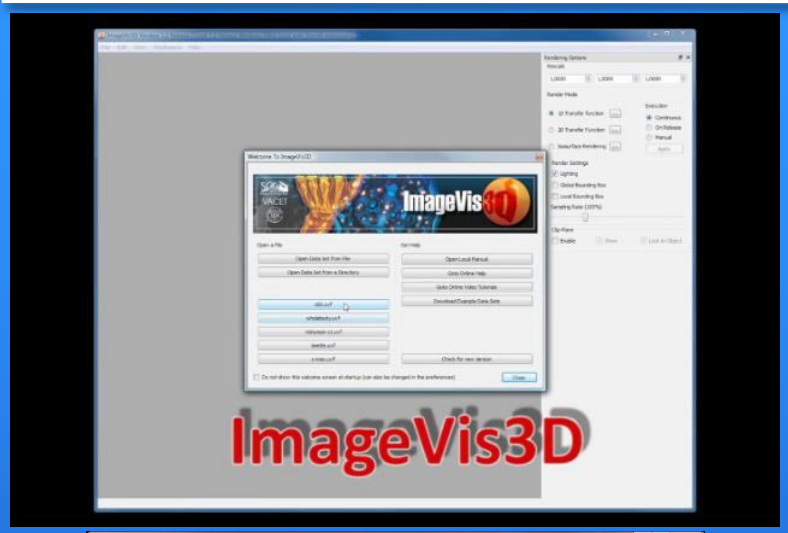
Volume Rendering

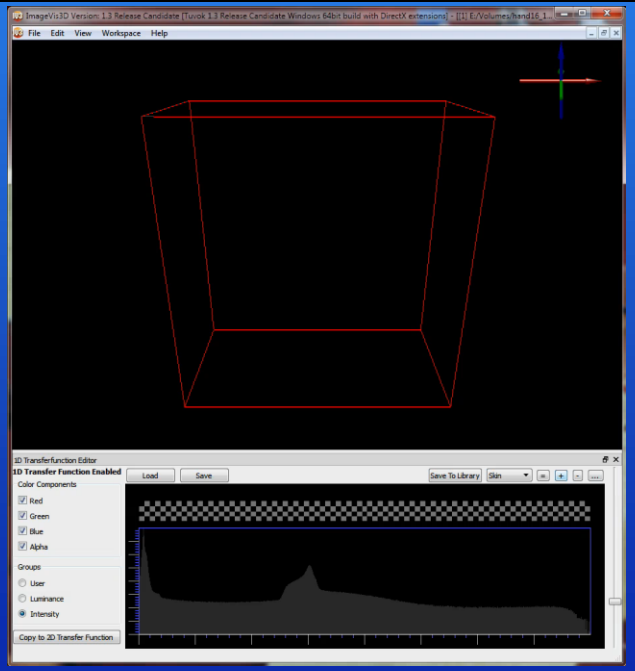


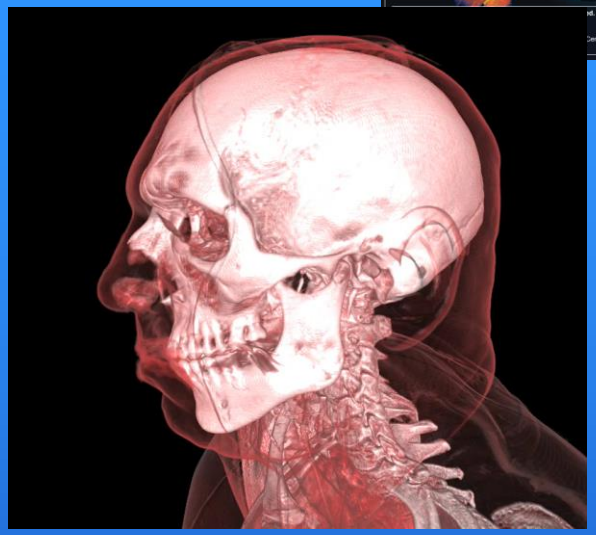
1D: not possible

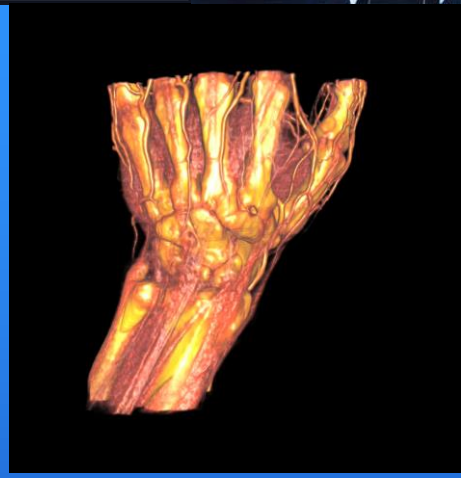
2D: specificity not as good

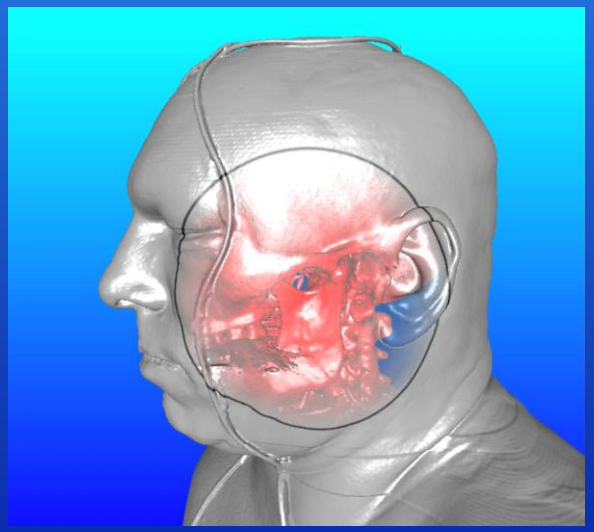
Volume Visualization

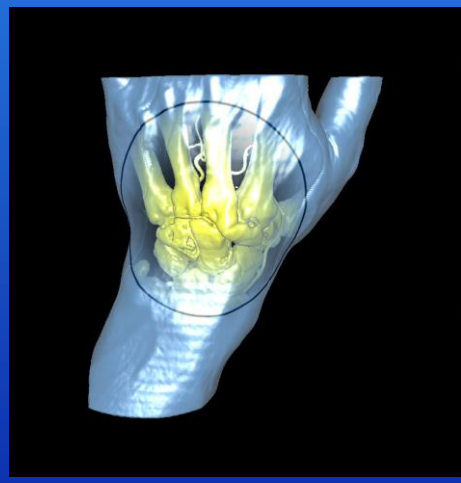




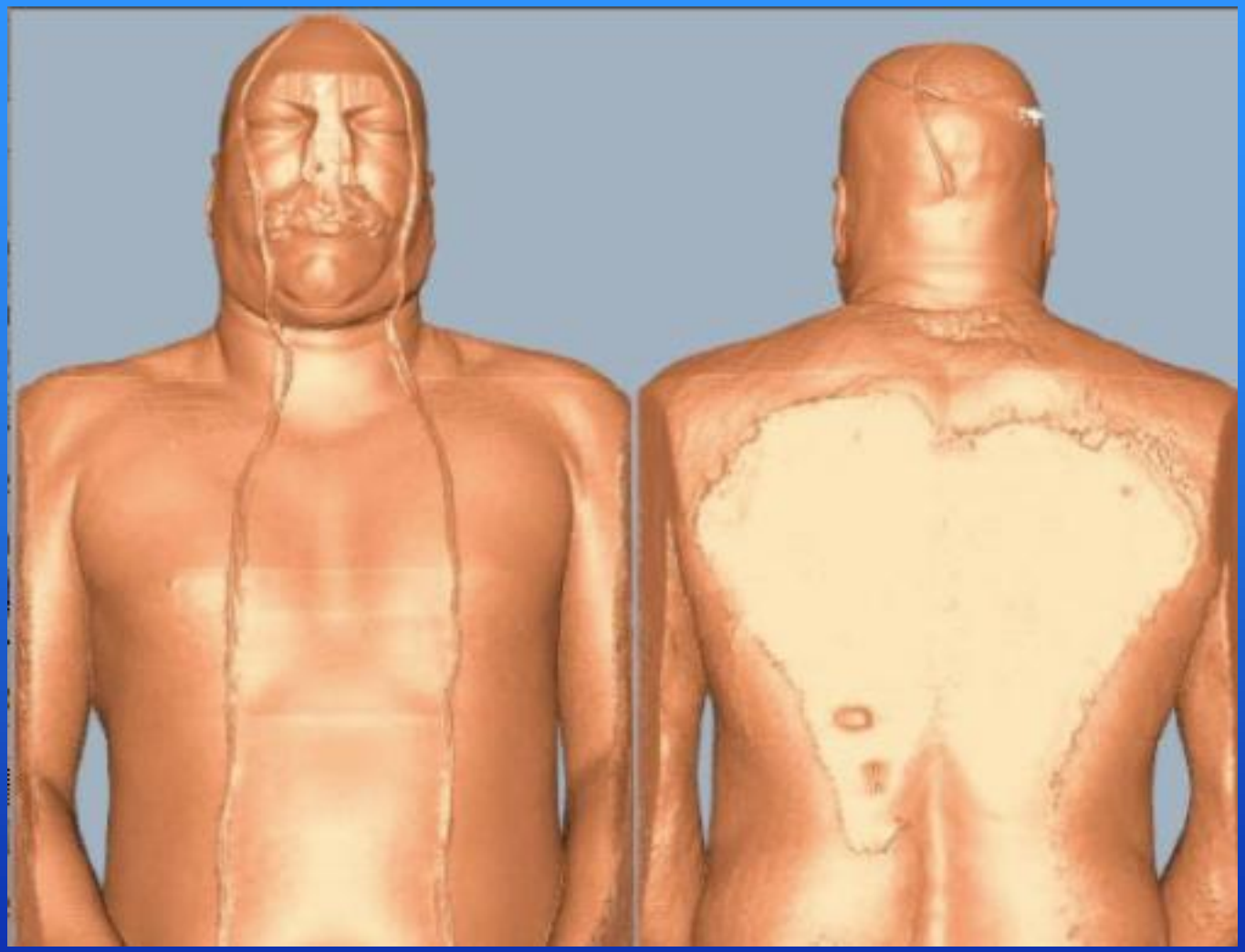








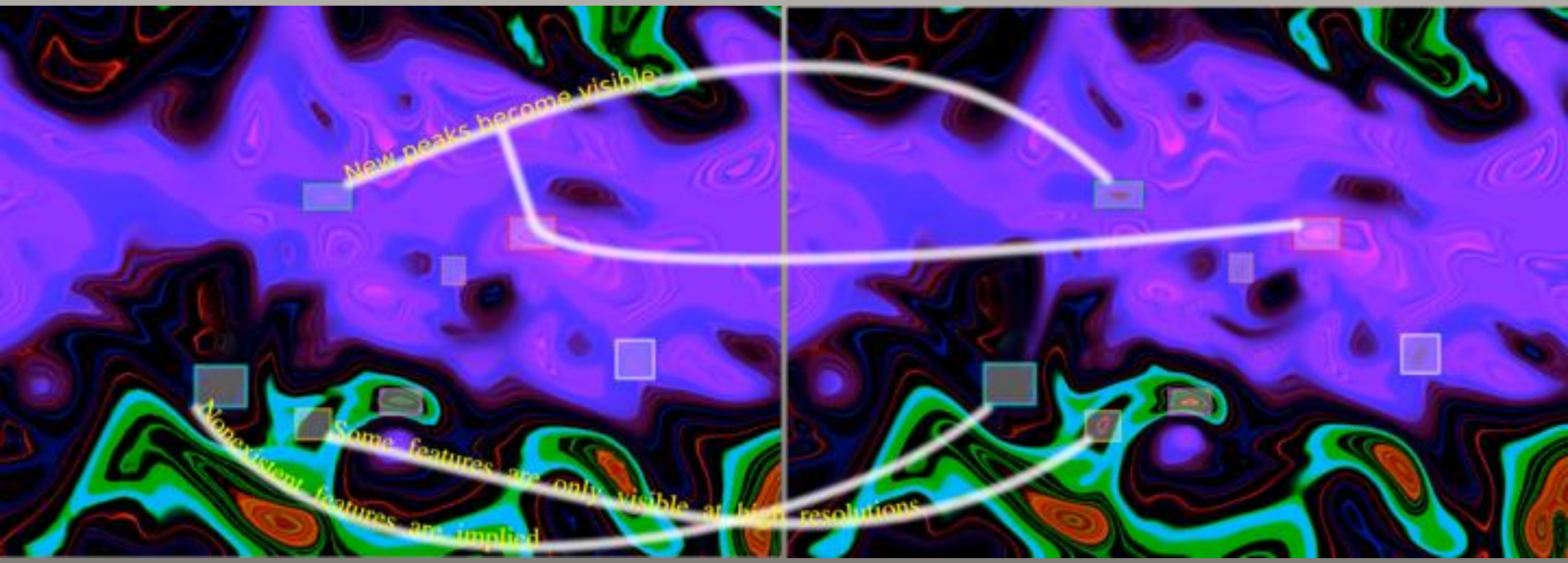
NIH Visible Male



Visible Human - High Resolution

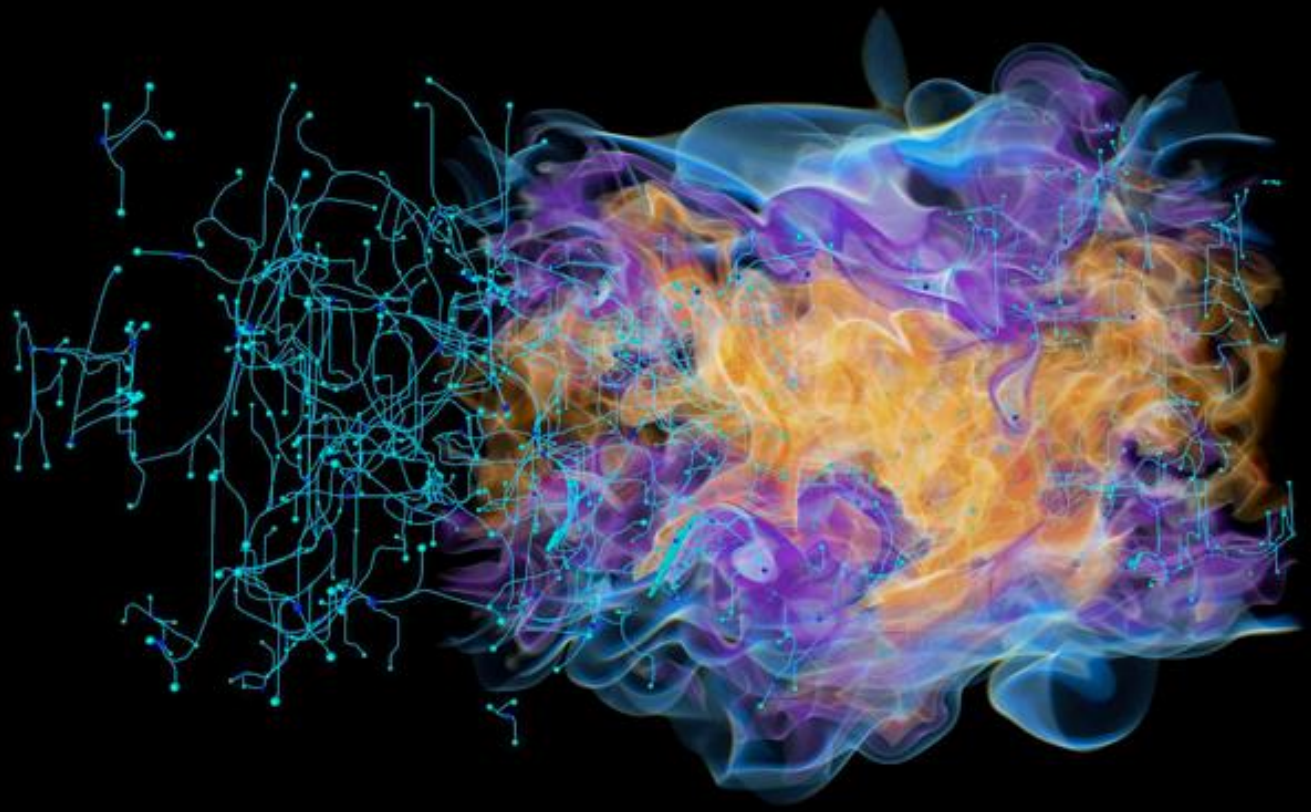
The Need for High Resolution Visualization

"…the data show for the first time how detailed transport and chemistry effects can influence the mixing of reactive scalars. It may be advantageous to incorporate these effects within molecular mixing models. It is worth noting that at present it is impossible to obtain this type of information any other way than by using the type of highly resolved simulation performed here."
Jacqueline Chen, Sandia National Laboratories



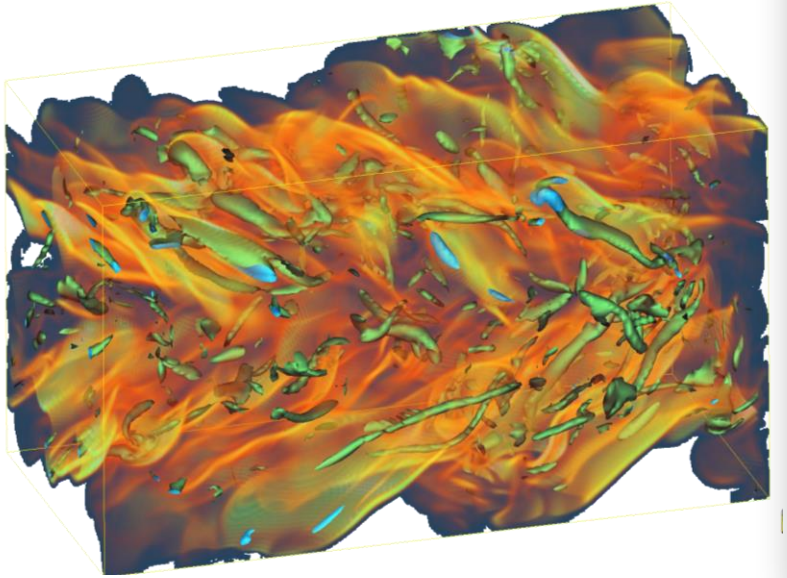
Lower Resolution

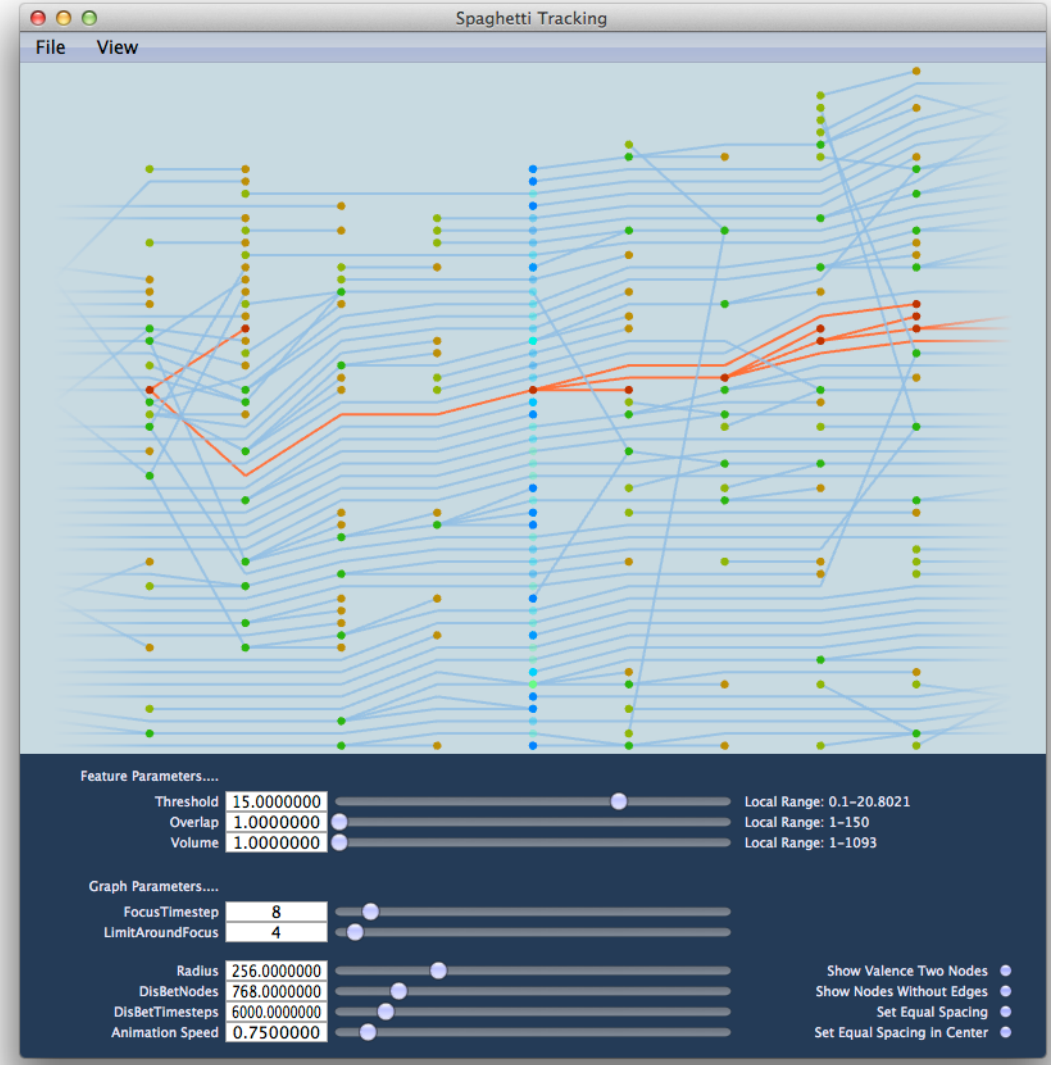
High Resolution



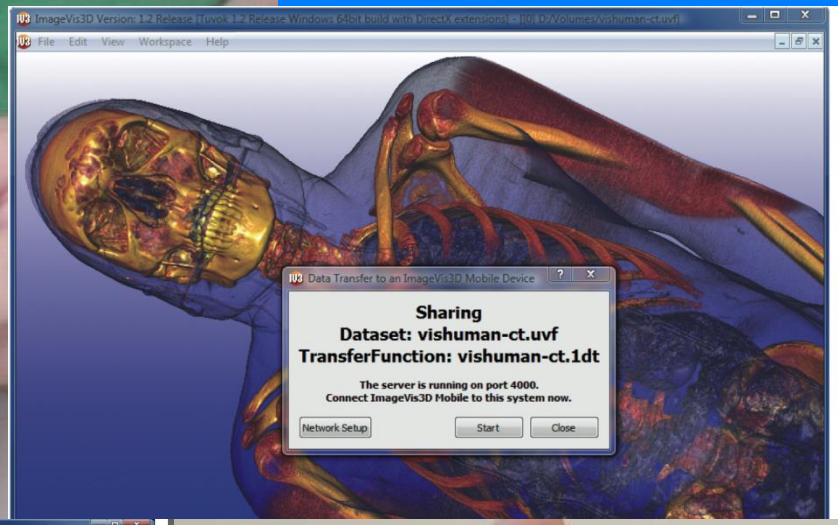
Topological Analysis of Massive Combustion Simulations

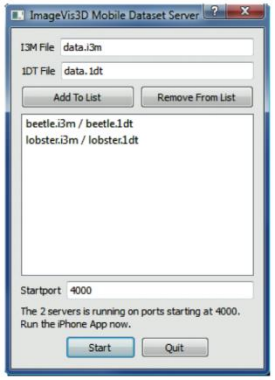
- Non-premixed DNS combustion (J. Chen, SNL): Analysis of the time evolution of extinction and reignition regions for the design of better fuels

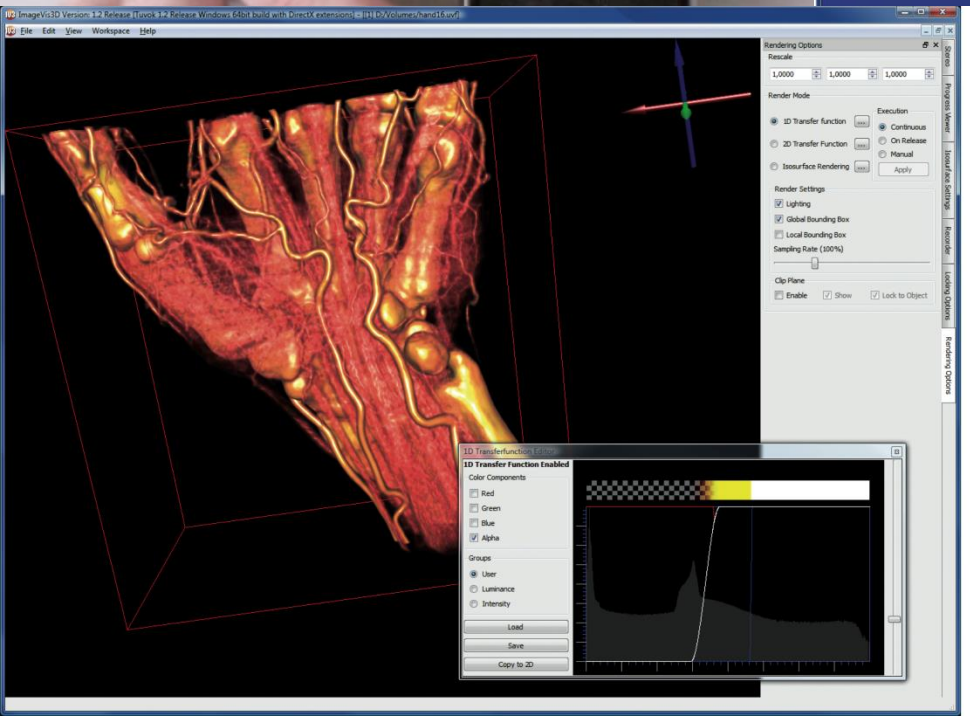




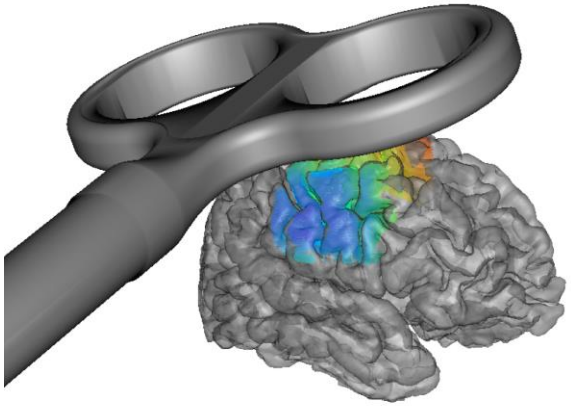
ImageVis3D - Mobile



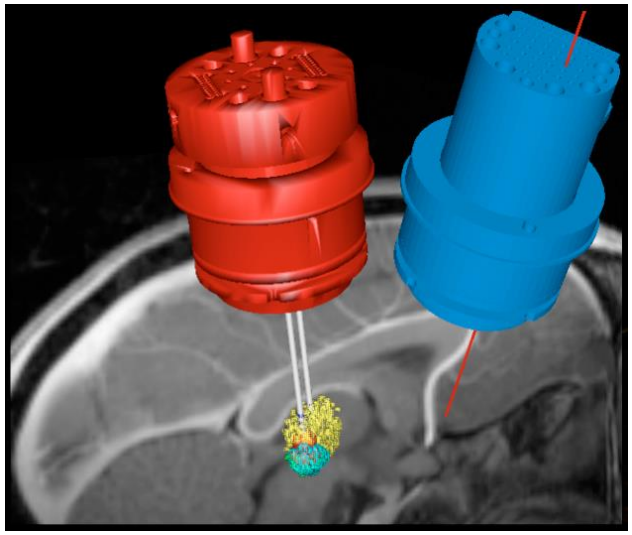


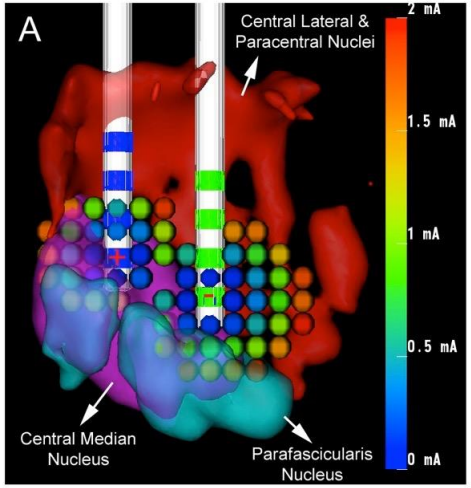


Transcranial Magnetic Stimulation for Depression

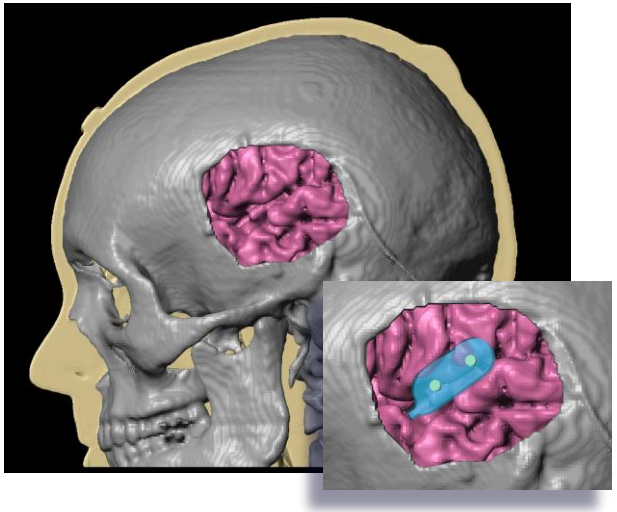


Deep Brain Stimulation for Traumatic Brain Injury: Primate, Human & Mouse

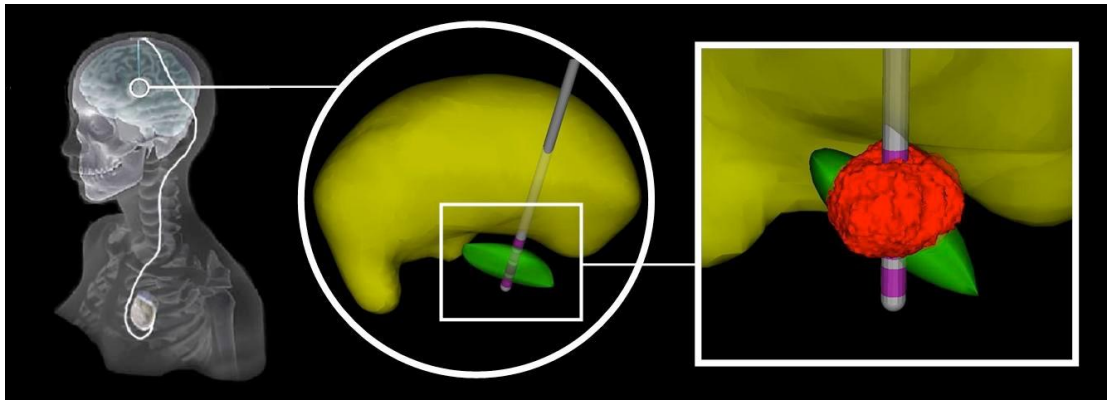




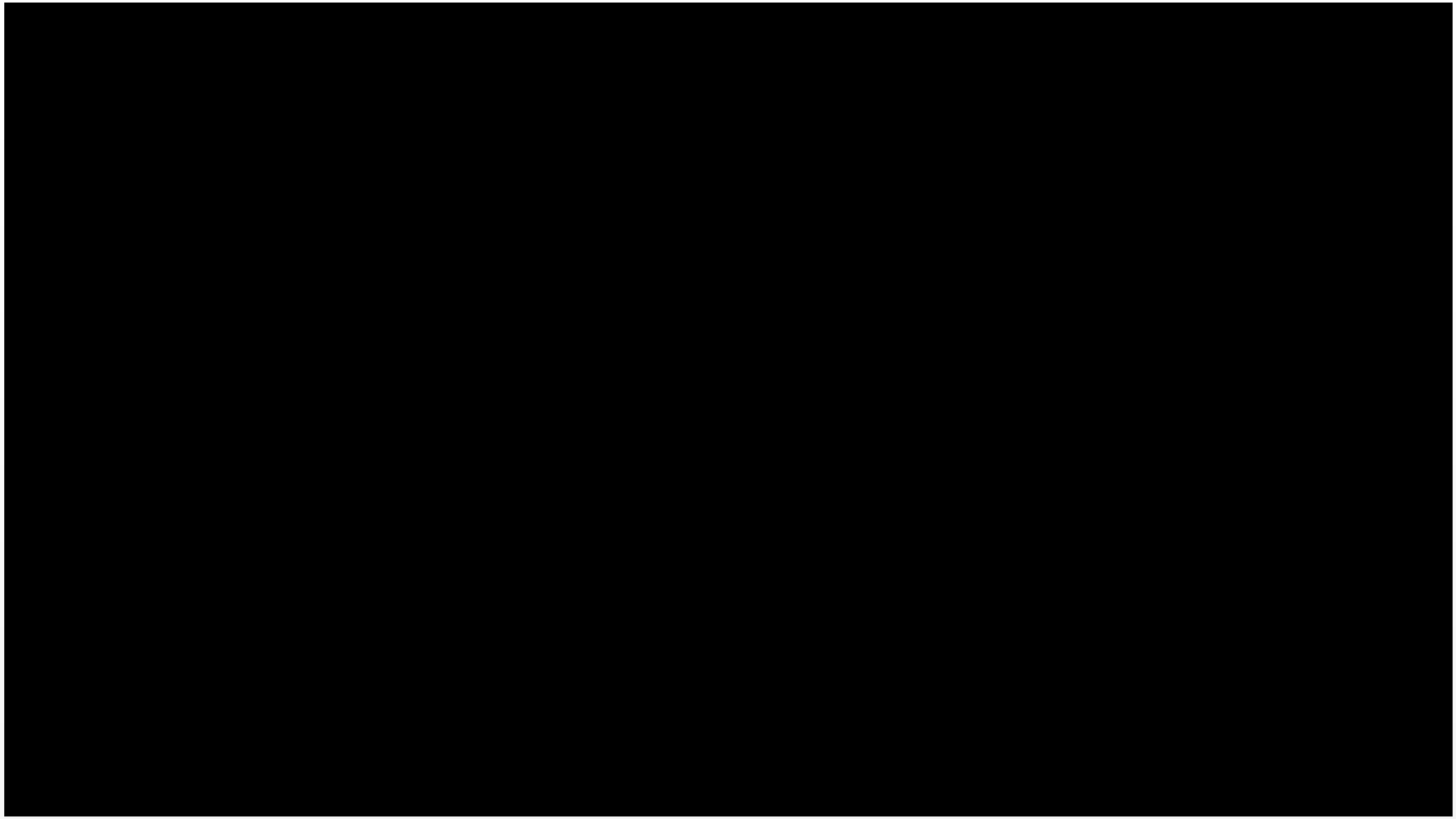
Cortical Stimulation for Depression



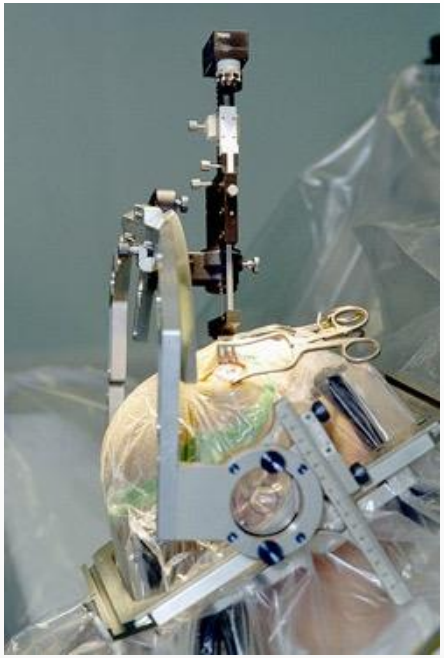
Deep Brain Stimulation (DBS) for Movement Disorders



Deep Brain Stimulation



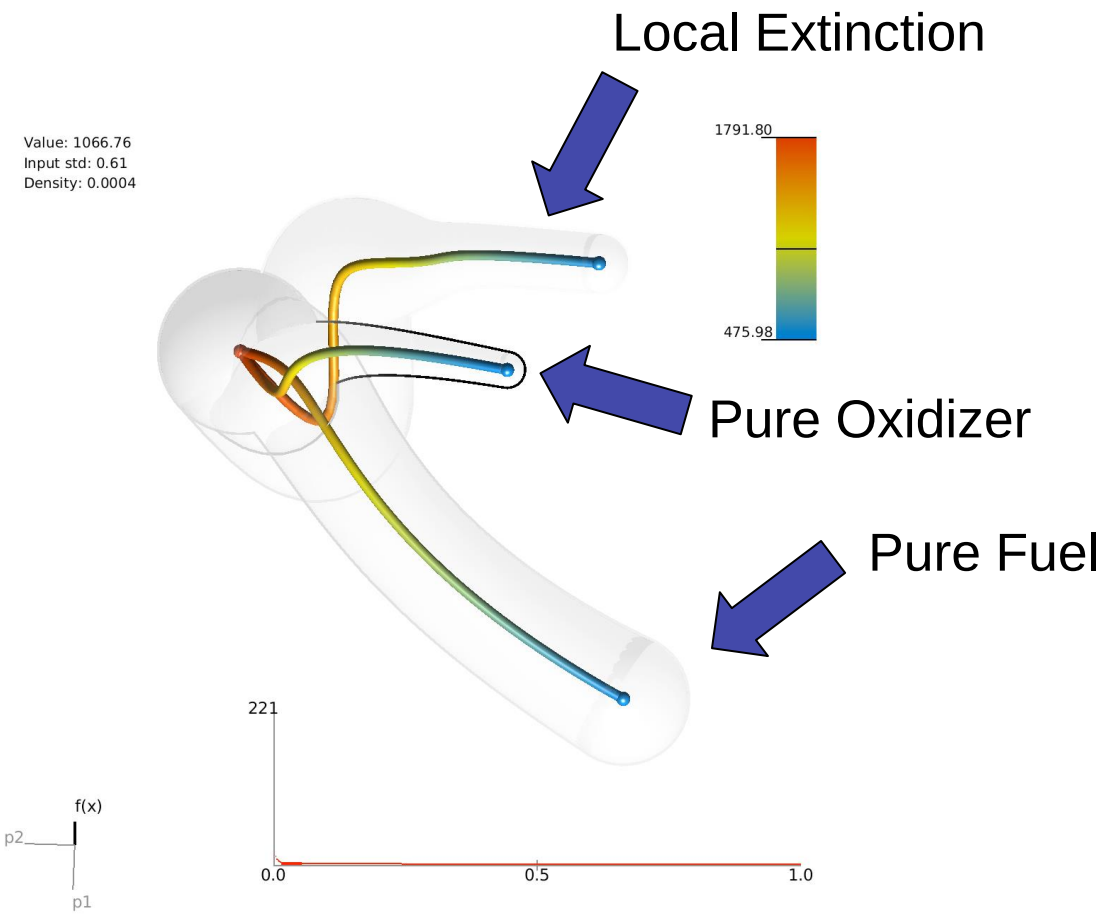
ImageVis3D Mobile DBS App

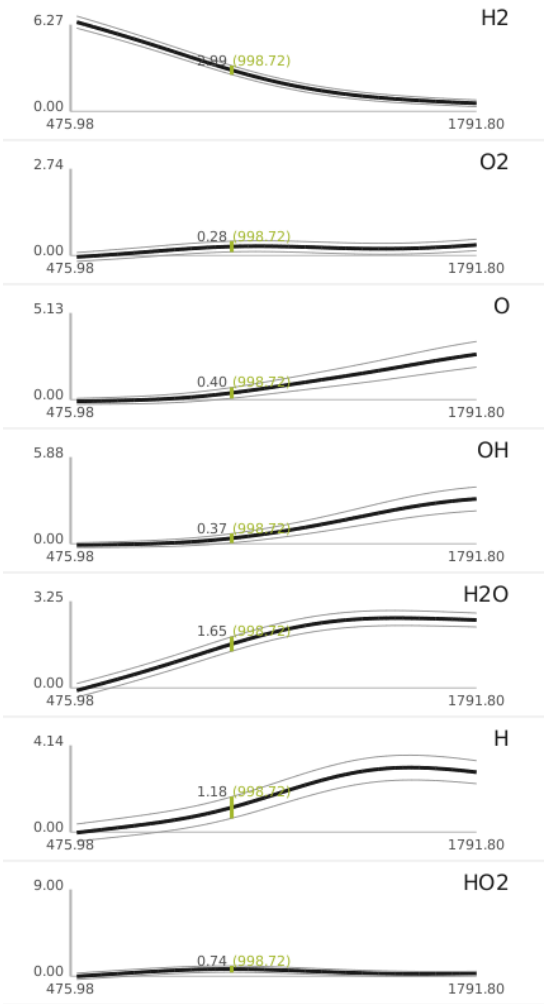


Deep Brain Stimulation
DBP: Chris Butson

C. Butson, G. Tamm, S. Jain, T. Fogal and J. Krüger "Evaluation of Interactive Visualization on Mobile Computing Platforms for Selection of Deep Brain Stimulation Parameters" *IEEE Transactions on Visualization and Computer Graphics, 2012 (in press).*

Visualization of 10D Combustion Simulation of Jet CO/H2-Air Flames





10 dimensional data set describing the heat release wrt. to various chemical species in a combustion simulation

Analysis of Combustion Simulations

Combustion Simulation of Jet CO/H2-Air Flames

Input: Composition of 10 chemical species

Output: Temperature

Interactive Streamline Exploration and Manipulation using Deformation

Xin Tong[1], John Edwards[2], Chun-Ming Chen[1],
Han-Wei Shen[1], Chris R. Johnson[2], Pak Chung Wong[3]
[1]The Ohio State University
[2]Scientific Computing and Imaging Institute, University of Utah
[3]Pacific Northwest National Laboratory

Michelangelos David

Michelangelos David - Part 2

Gigapixel David - iTunes App

crjsci@mac.com | Music | Movies | TV Shows | App Store | Books | Podcasts | iTunes U

App Store › Education › Valerio Pascucci

GigapixelDavid

Valerio Pascucci ›

Details | Ratings and Reviews | Related

Free

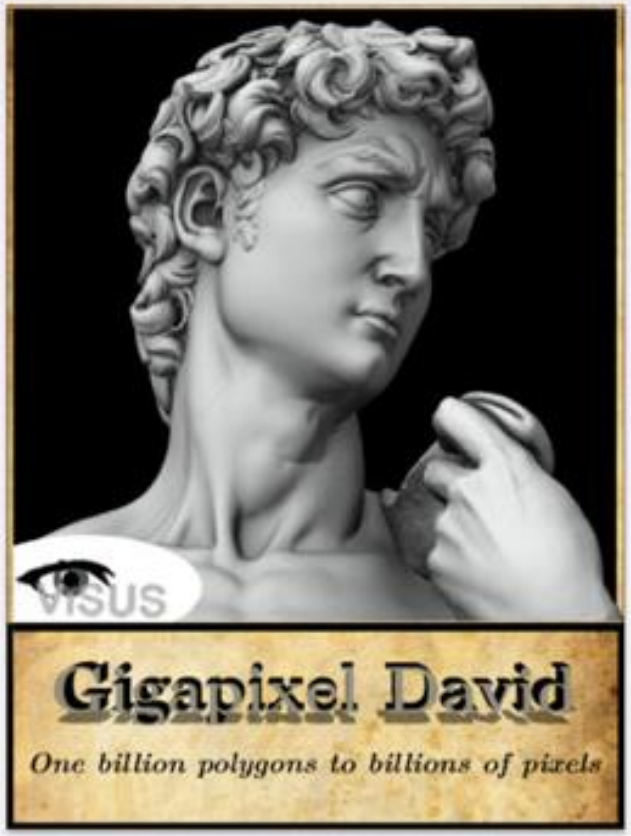
This app is designed for both iPhone and iPad

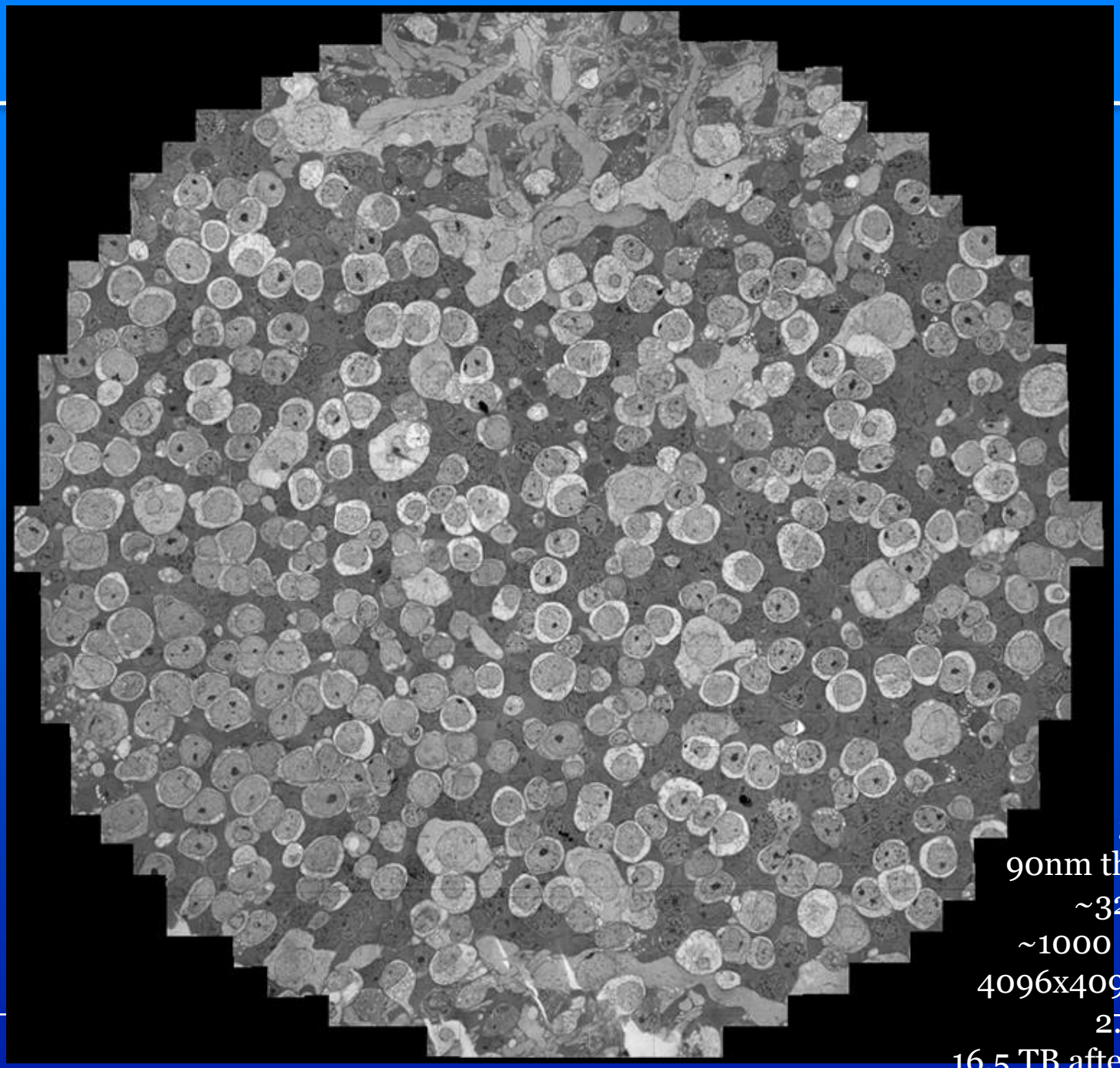
No Ratings
Rated 17+

LINKS

Privacy Policy
Developer Website

Screenshots | iPhone | iPad



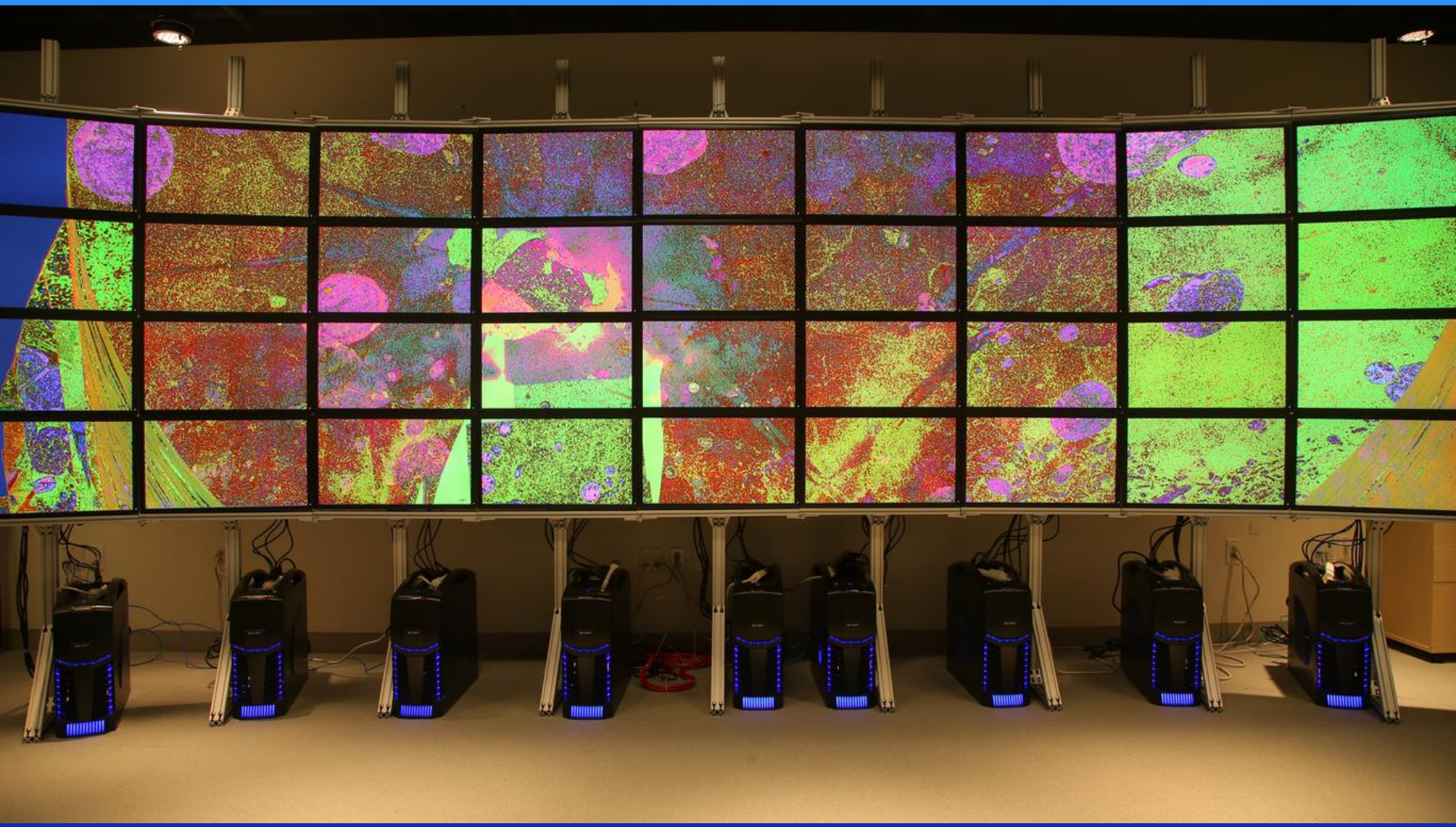


341 Sections
90nm thick sections
~32GB/Section
~1000 tiles/section
4096x4096 pixels/tile
2.18 nm/Pixel
16.5 TB after processing

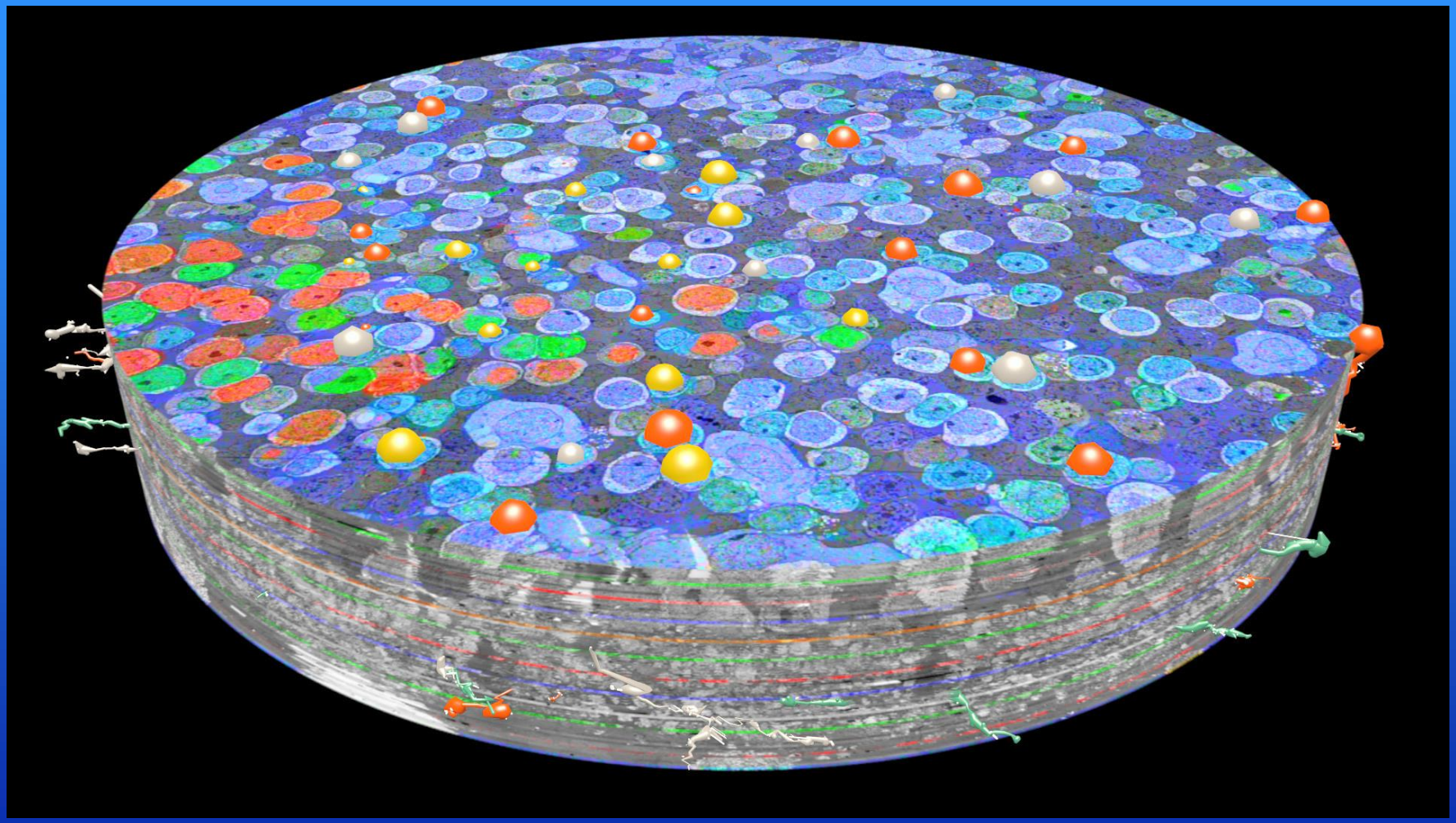
Antony van Leeuwenhoek (1632-1723)

. . . my work, which I've done for a long time, was not pursued in order to gain the praise I now enjoy, but chiefly from a craving after knowledge, which I notice resides in me more than in most other men. And therewithal, whenever I found out anything remarkable, I have thought it my duty to put down my discovery on paper, so that all ingenious people might be informed thereof.

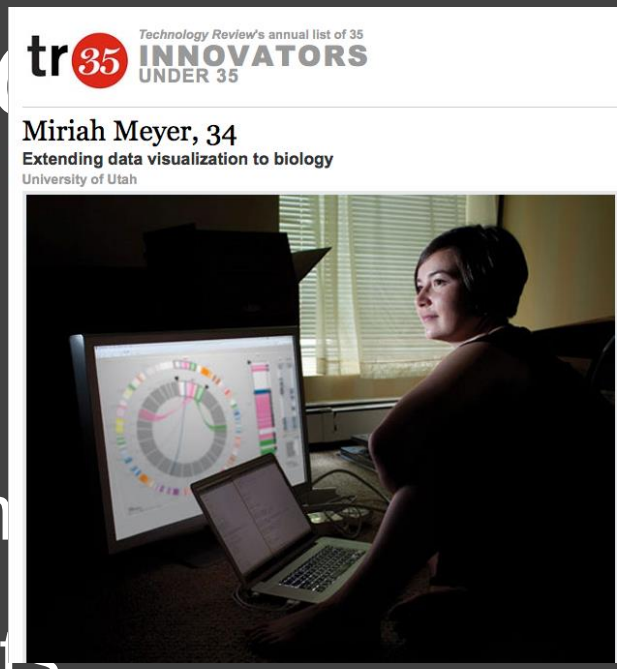
Antony van Leeuwenhoek. Letter of June 12, 1716



Connectome

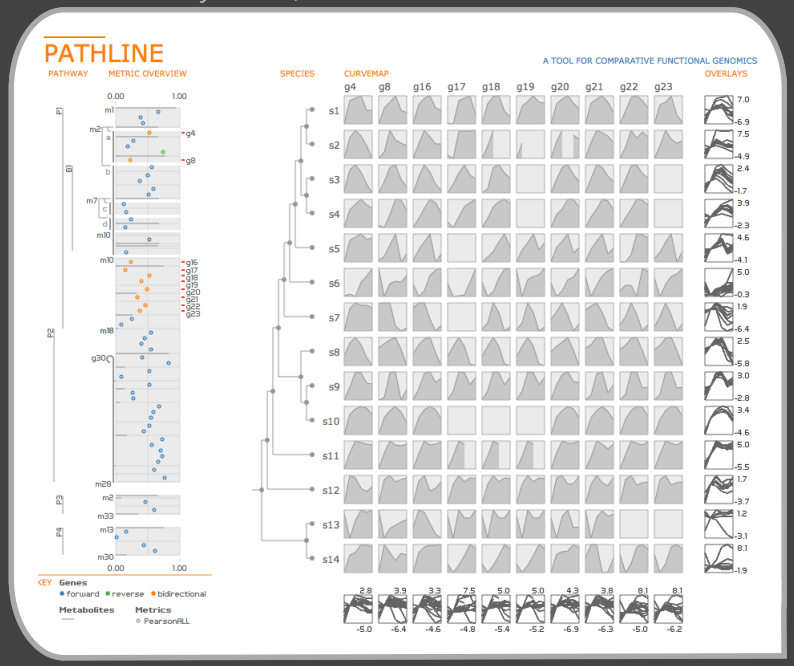


PROBLEM-DRIVEN VISUALIZATION RESEARCH *for biol data*

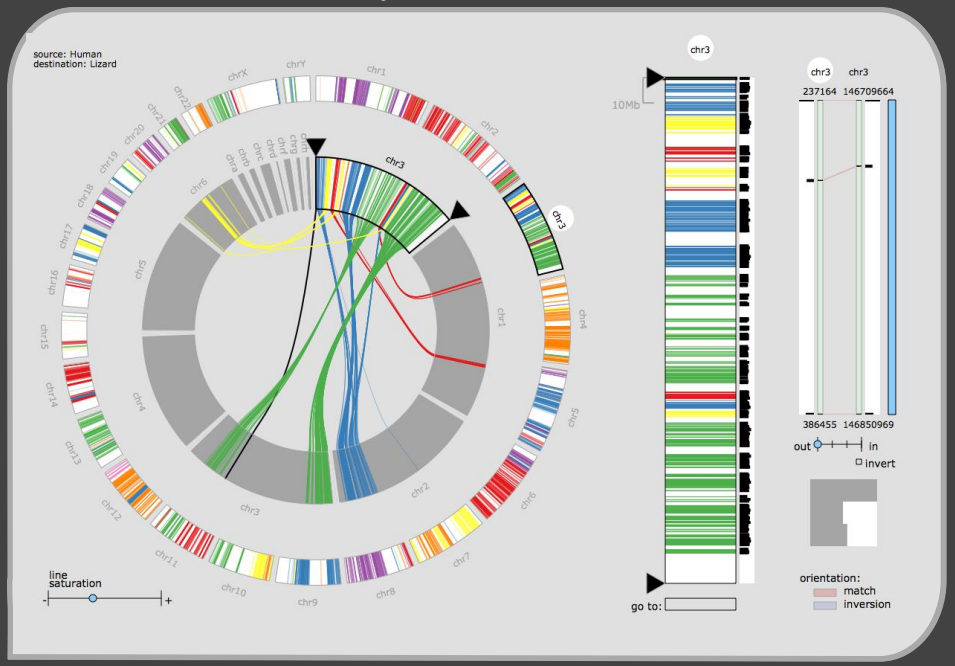


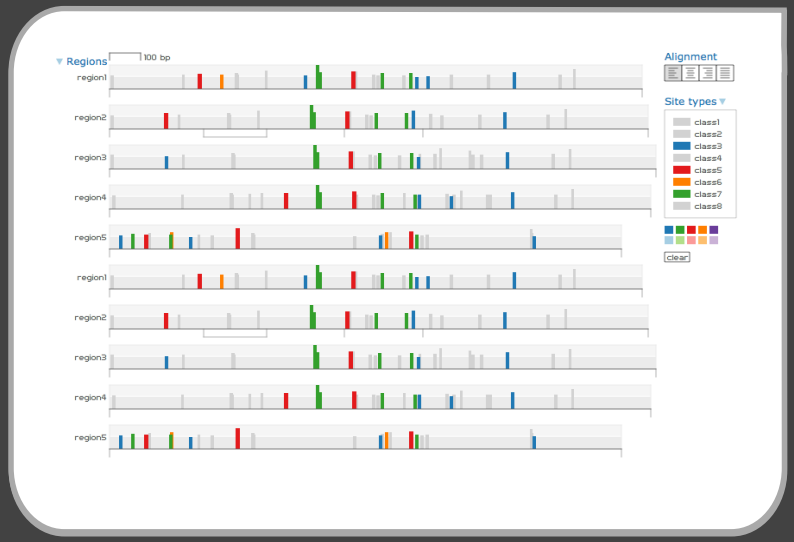
- target specific biological problem
- close collaboration with biologists
- rapid, iterative prototyping
- focus on genomic and molecular data

M. Meyer et al., EuroVis 2010. **Pathline**

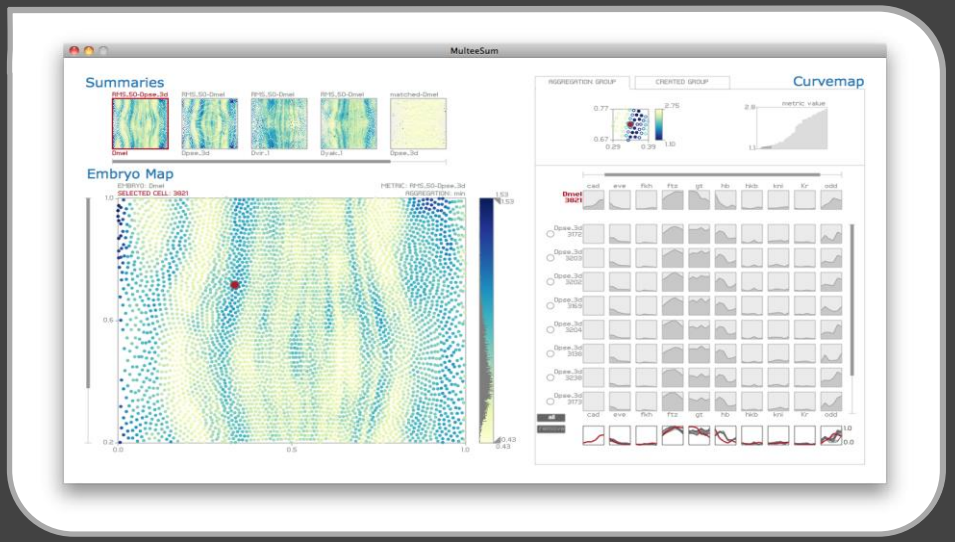


MizBee *M. Meyer et al., InfoVis 2009.*

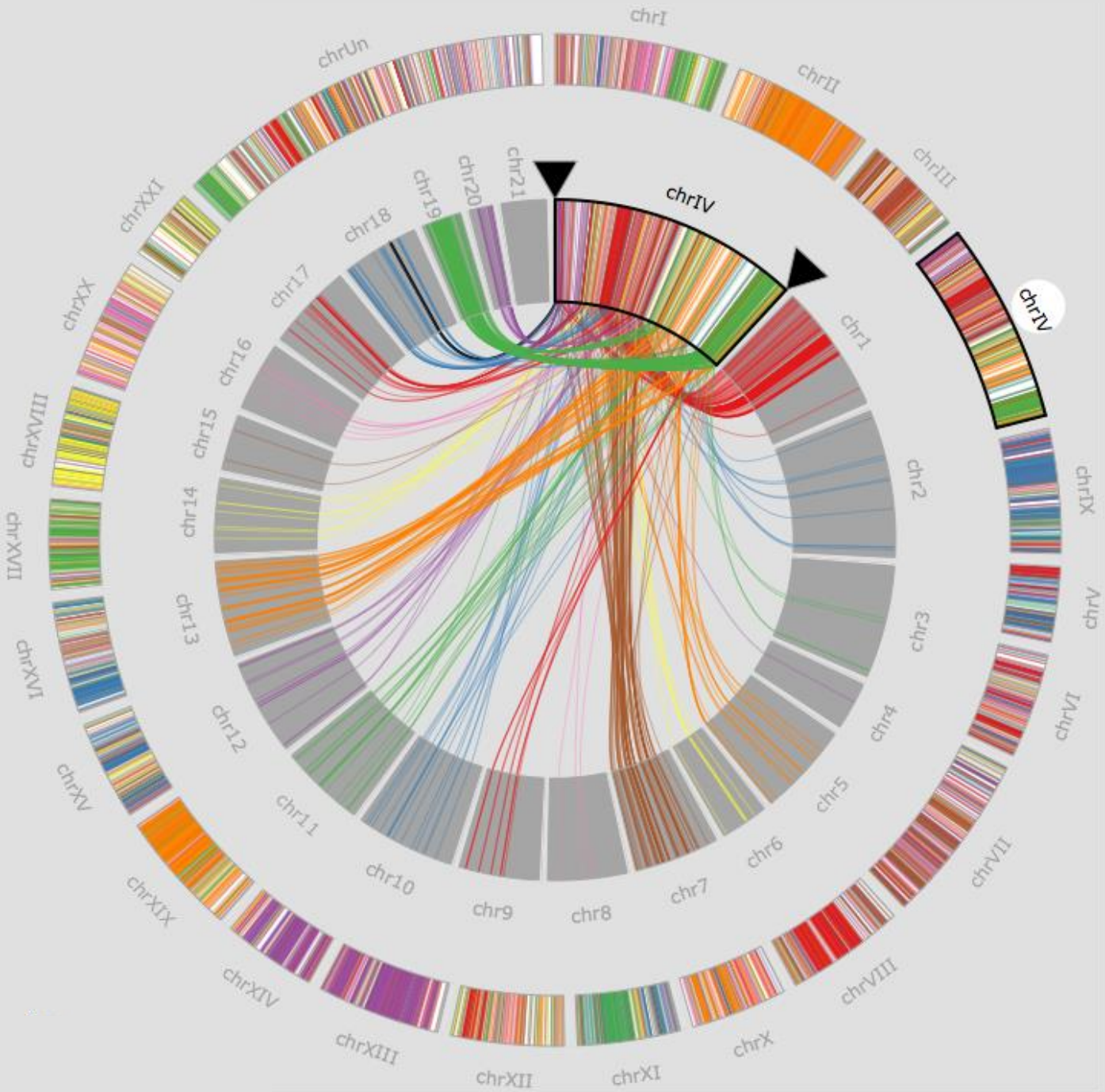


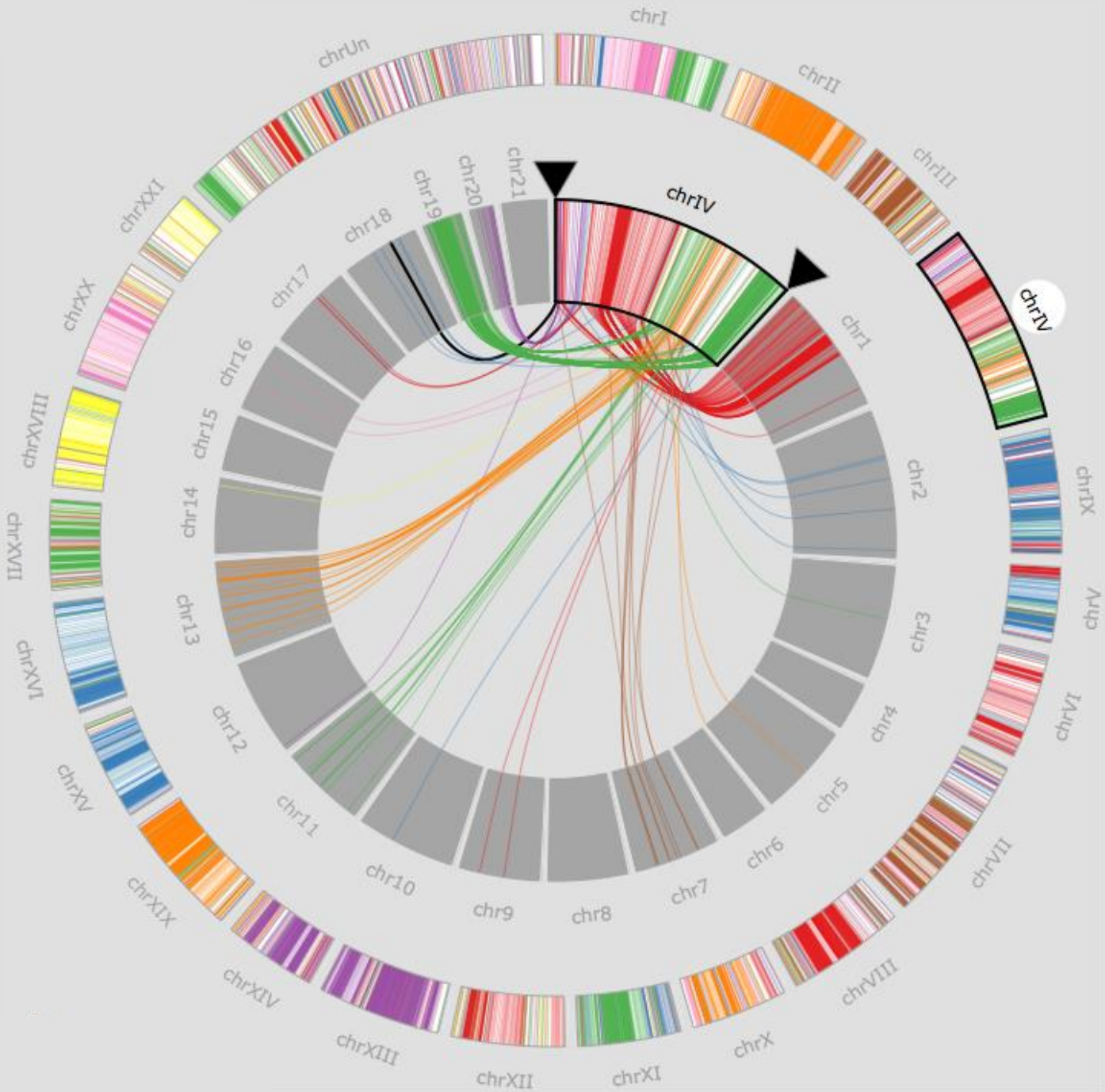


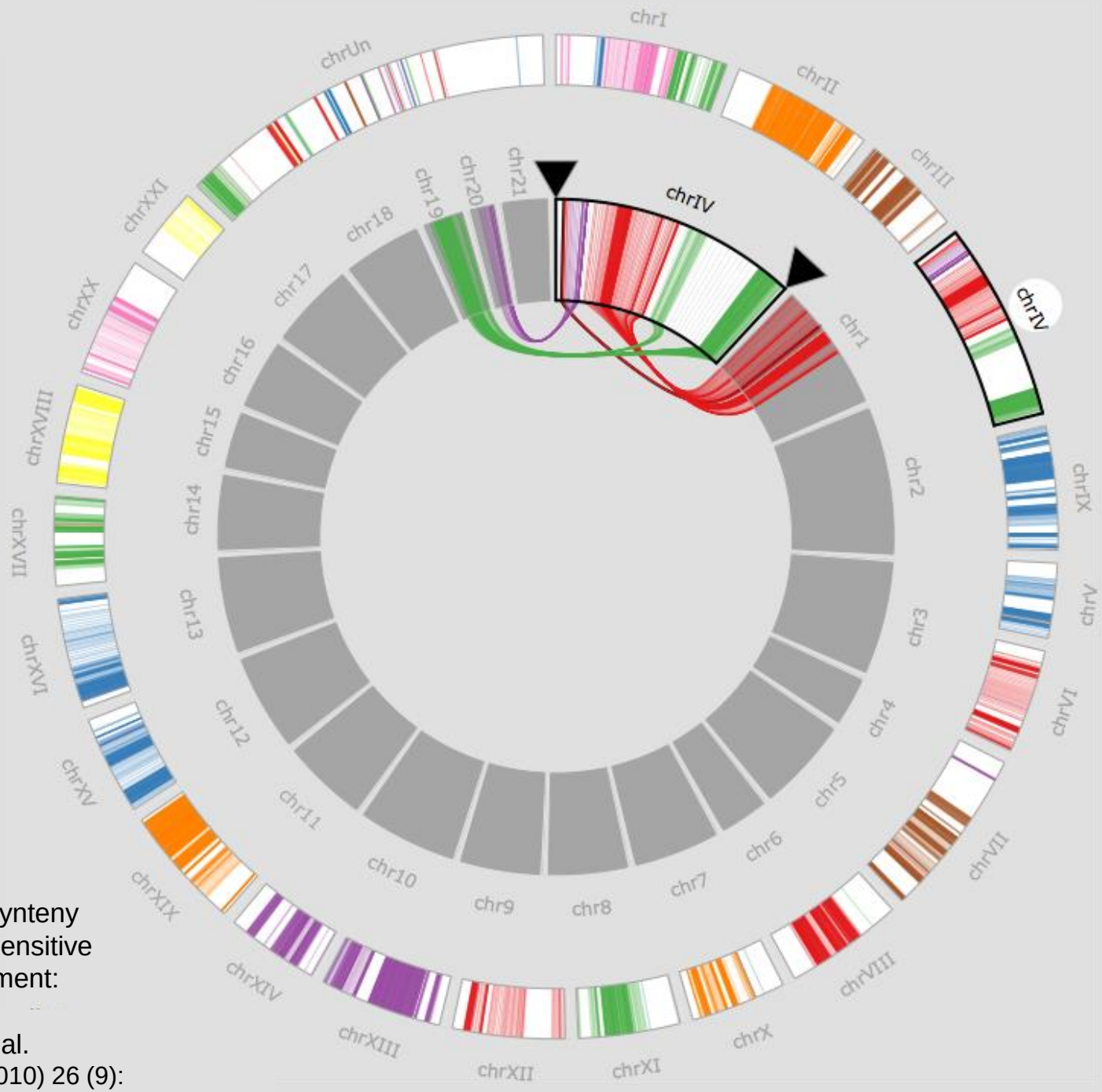
InSite



MulteeSum *M. Meyer et al., InfoVis 2010.*


chrI
chrII
chrIII
chrIV
chrV
chrVI
chrVII
chrVIII
chrX
chrXI
chrXII
chrXIII
chrXIV
chrXIX
chrXV
chrXVI
chrXVII
chrXVIII
chrXX
chrXXI
chrUn
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chrIV


chrI
chrII
chrIII
chrIV
chrIV
chrIX
chrV
chrVI
chrVII
chrVIII
chrX
chrXI
chrXII
chrXIII
chrXIV
chrXIX
chrXV
chrXVI
chrXVII
chrXVIII
chrXX
chrXXI
chrUn
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21

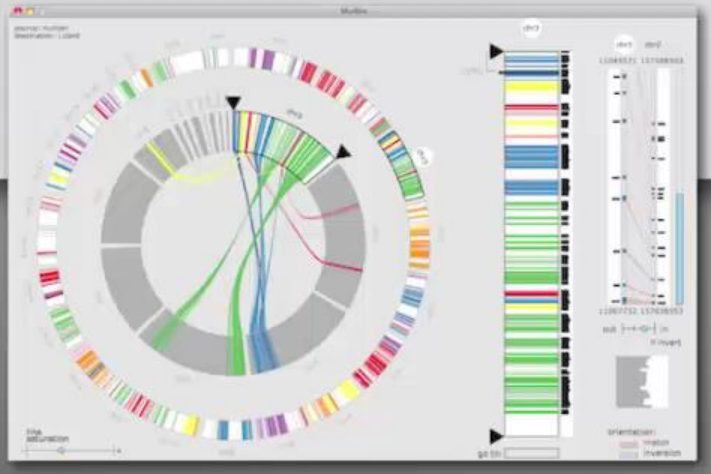


Genome-wide synteny through highly sensitive sequence alignment: Satsuma

M. Grabherr, et al.

Bioinformatics (2010) 26 (9): 1145-1151.

Visualization of Biological Data

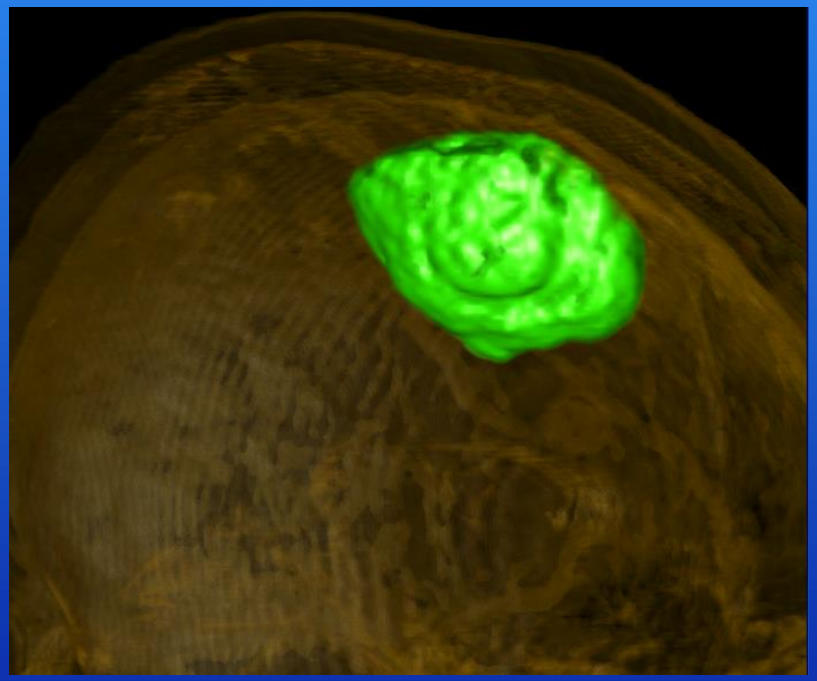


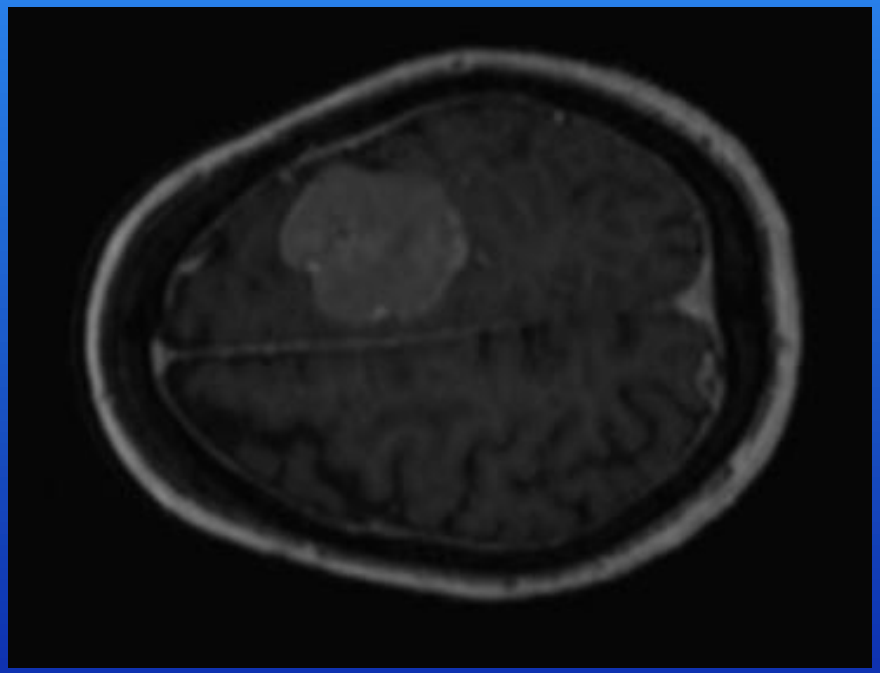
MizBee

Browser that enables analysis of comparative genomics data through visualization across multiple scales.

Uncertainty Visualization

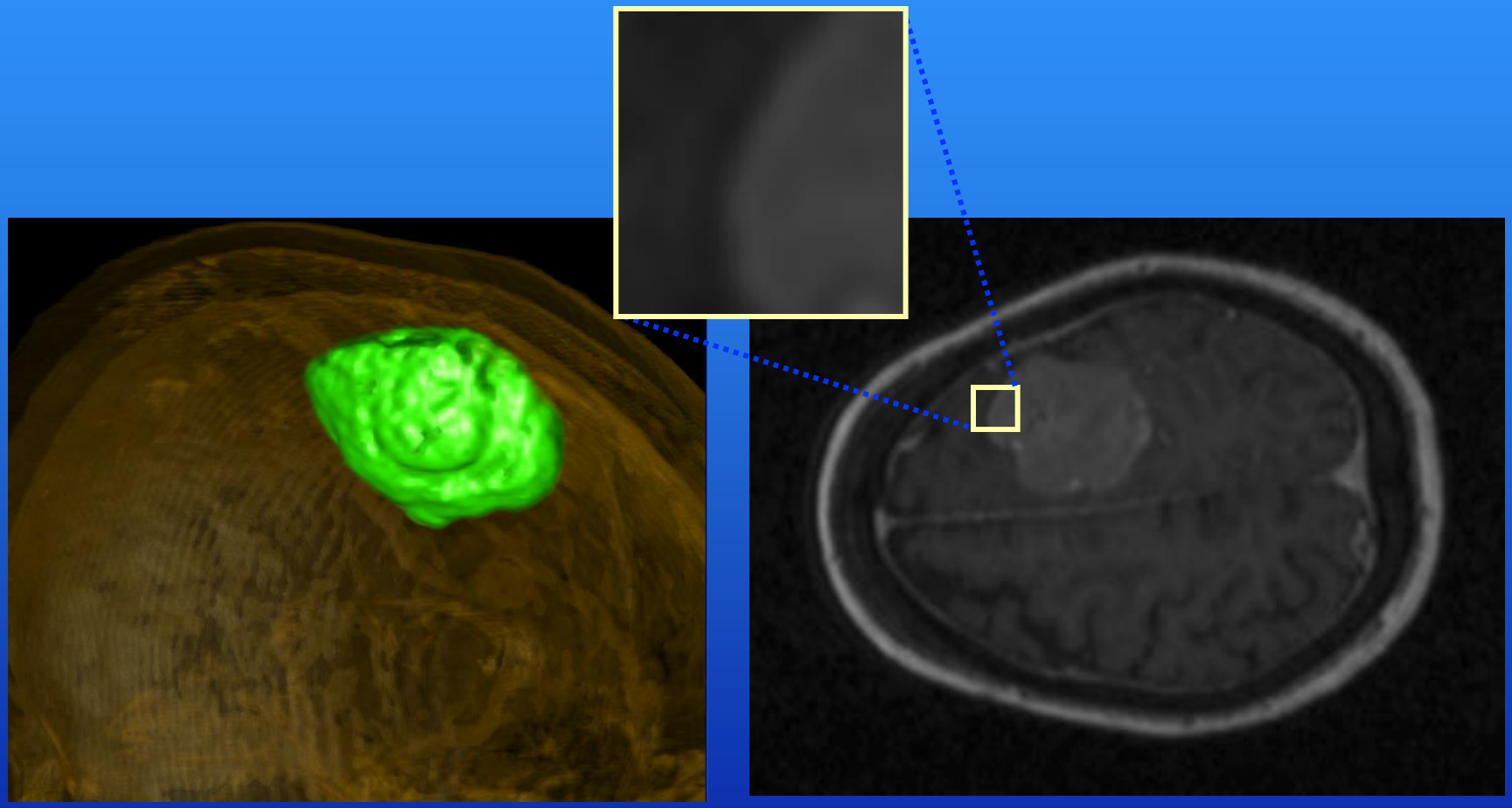
When is the last time you've seen an error bar in a 3D visualization?





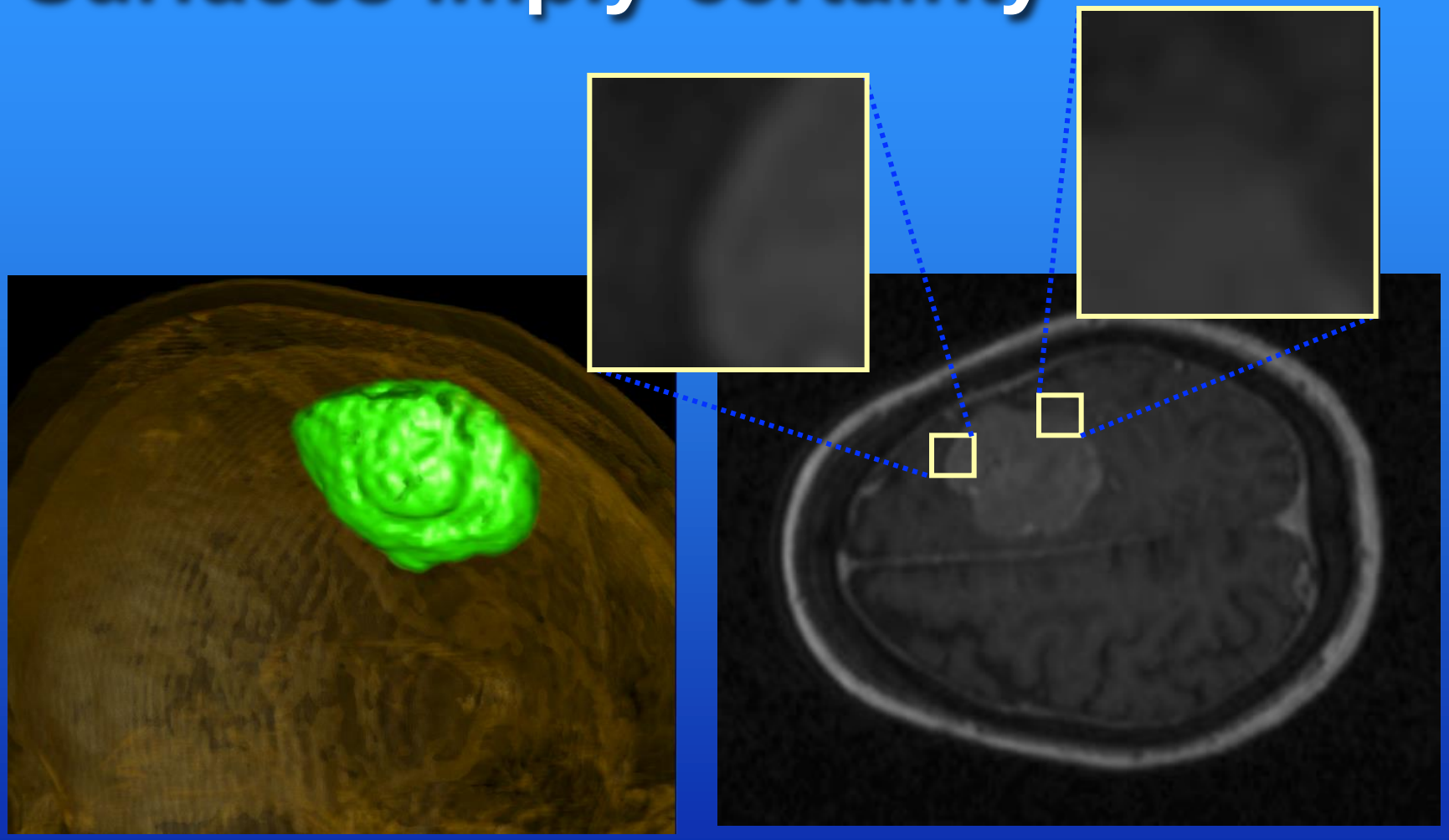
Uncertainty Visualization

Surfaces imply certainty



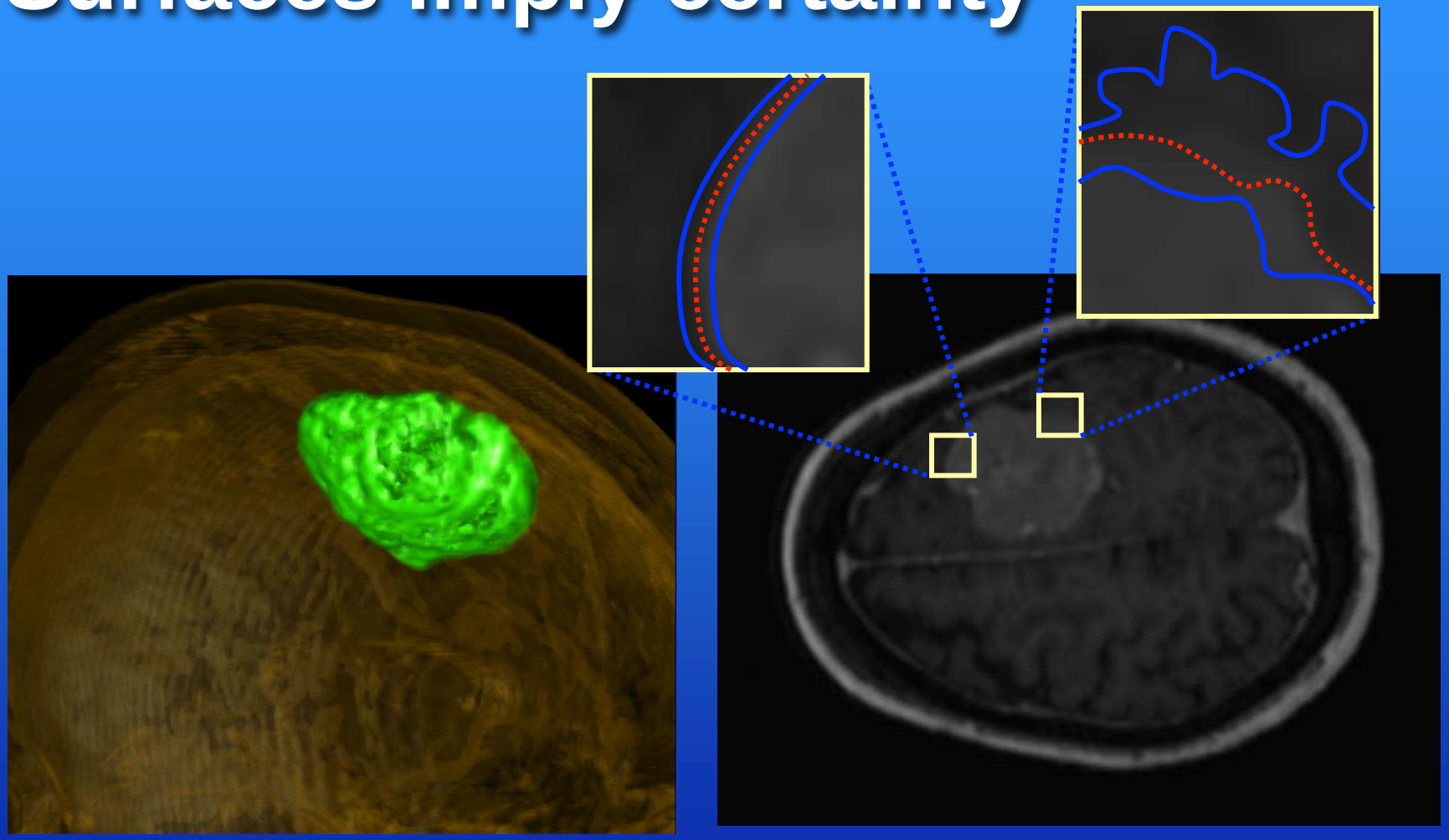
Uncertainty Visualization

Surfaces imply certainty

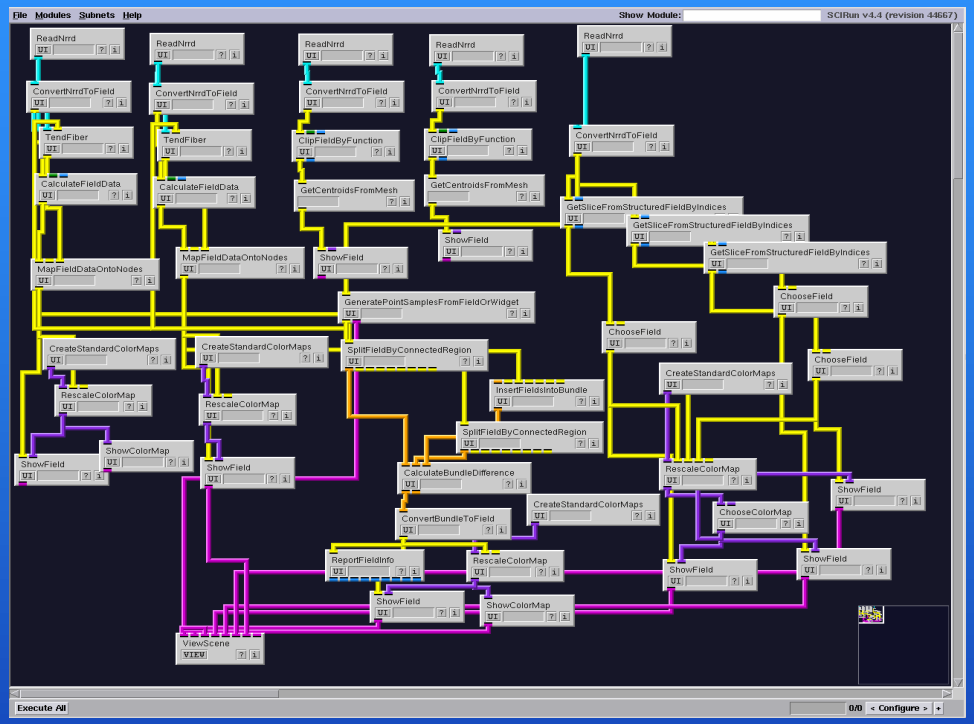


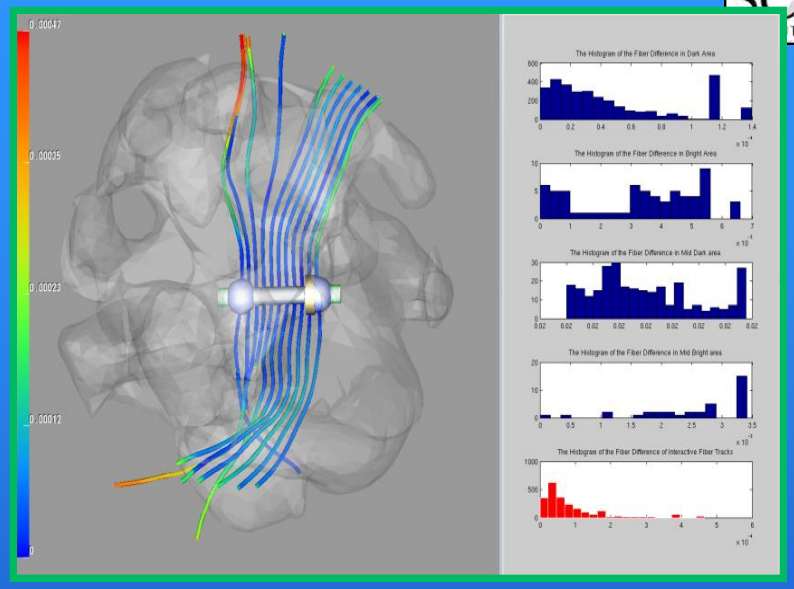
Uncertainty Visualization

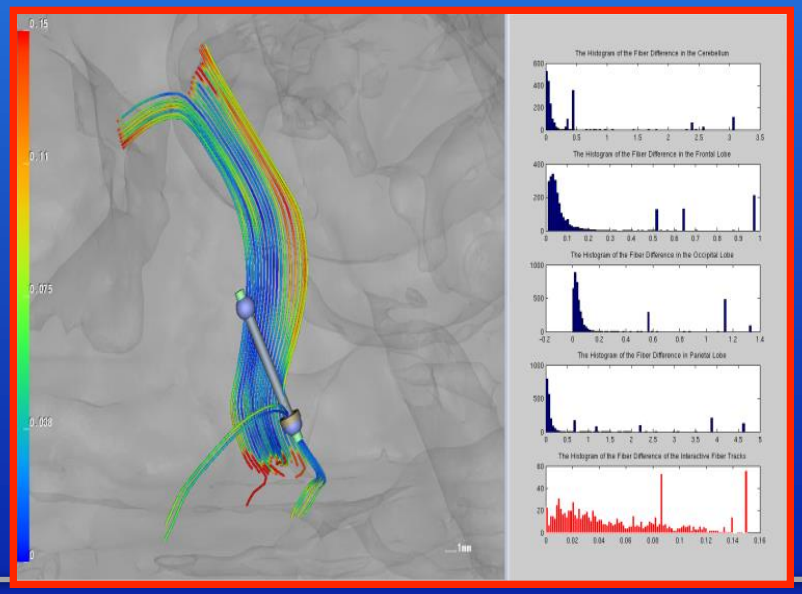
Surfaces imply certainty



iQuantVis


File Modules Subnets Help
Show Module:
SCIRun v4.4 (revision 44667)
ReadNrrd
ConvertNrrdToField
TendFiber
CalculateFieldData
MapFieldDataOntoNodes
CreateStandardColorMaps
RescaleColorMap
ShowColorMap
ShowField
ClipFieldByFunction
GetCentroidsFromMesh
GeneratePointSamplesFromFieldOrWidget
SplitFieldByConnectedRegion
InsertFieldsIntoBundle
CalculateBundleDifference
ConvertBundleToField
ReportFieldInfo
ViewScene
GetSliceFromStructuredFieldByIndices
ChooseField
ChooseColorMap
Execute All
< Configure >


The Histogram of the Fiber Difference in Dark Area
The Histogram of the Fiber Difference in Bright Area
The Histogram of the Fiber Difference in Mid Dark area
The Histogram of the Fiber Difference in Mid Bright area
The Histogram of the Fiber Difference of Interactive Fiber Tracks


The Histogram of the Fiber Difference in the Cerebellum
The Histogram of the Fiber Difference in the Frontal Lobe
The Histogram of the Fiber Difference in the Occipital Lobe
The Histogram of the Fiber Difference in Parietal Lobe
The Histogram of the Fiber Difference of the Interactive Fiber Tracks

Topological Uncertainty

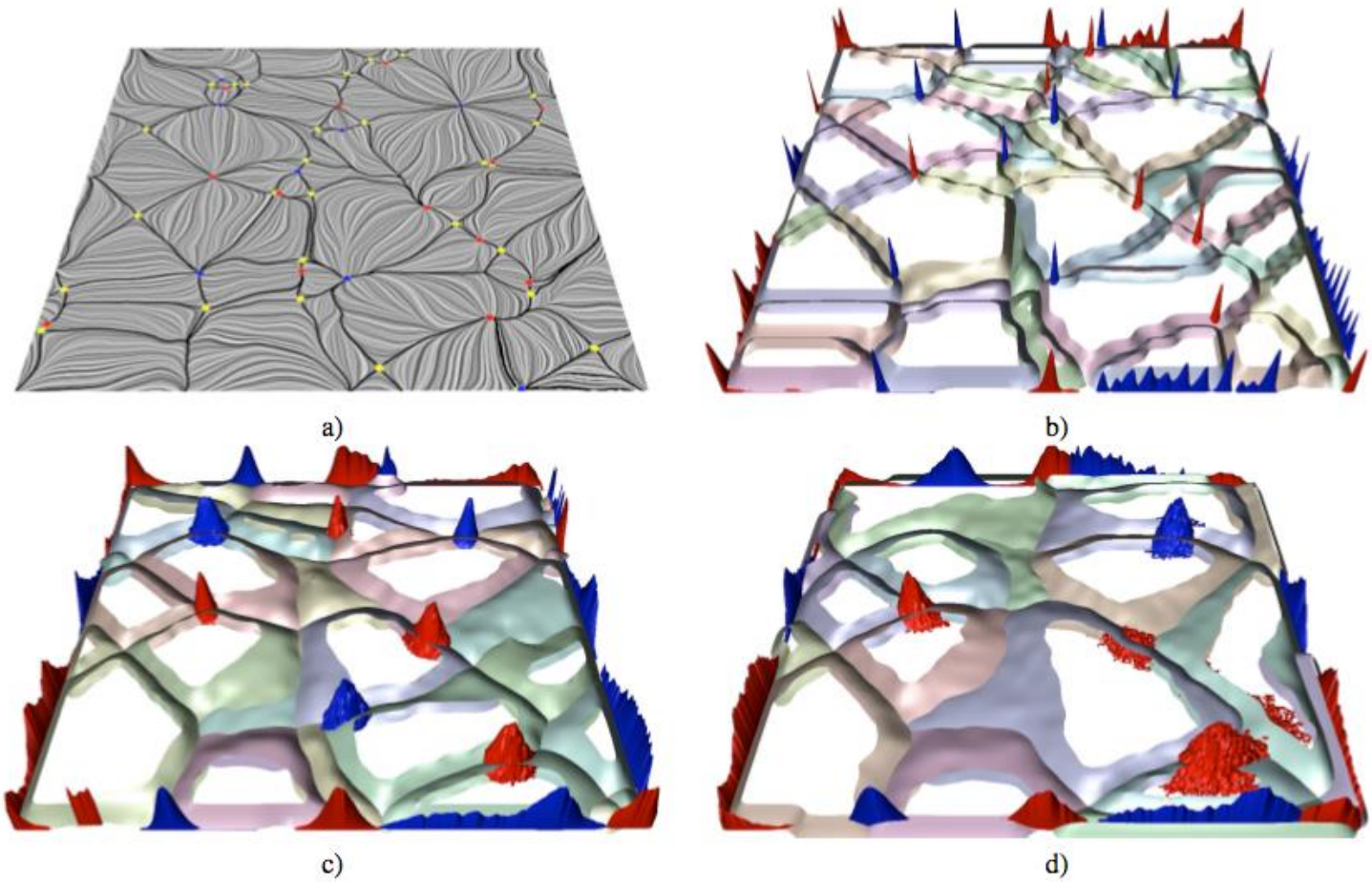
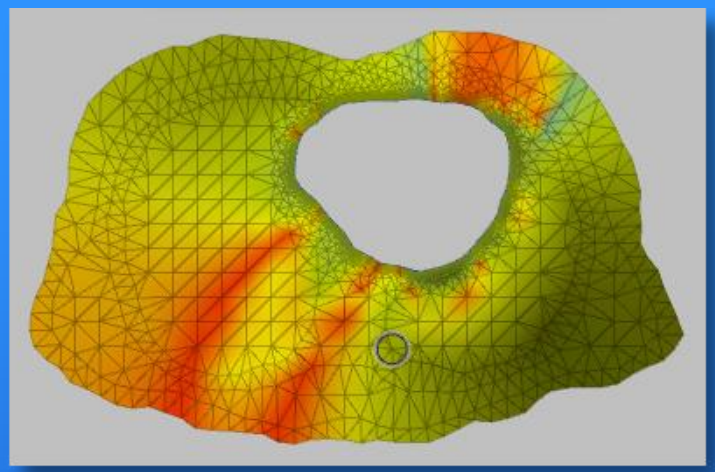
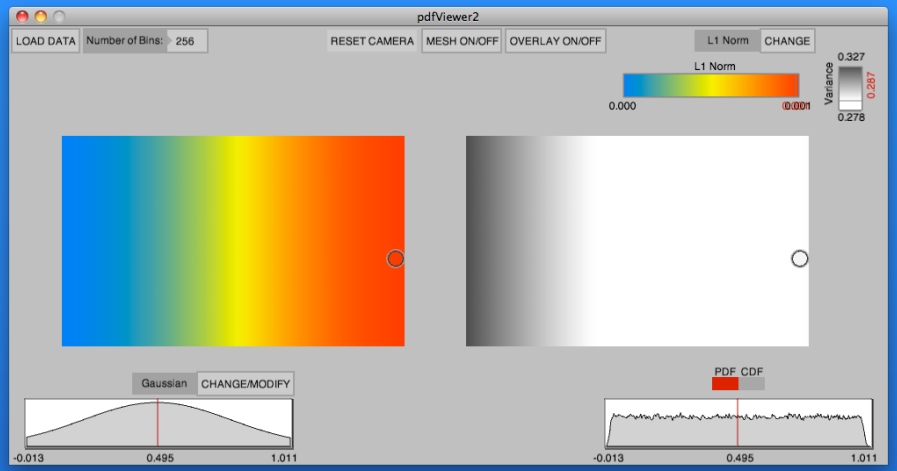


Figure 6: *Increasing the uncertainty of a random vector field: a) certain topology of mean vector field; b)* $\|\mathbf{T}\|_F = 0.2$*; c)* $\|\mathbf{T}\|_F = 2.0$*; d)* $\|\mathbf{T}\|_F = 5.0$.

M. Otto, T. Germer, H.C. Hege, H. Theisel. Uncertain 2D Vector Field Topology. In *CGF*, 29(2), 2010.

ProbVis



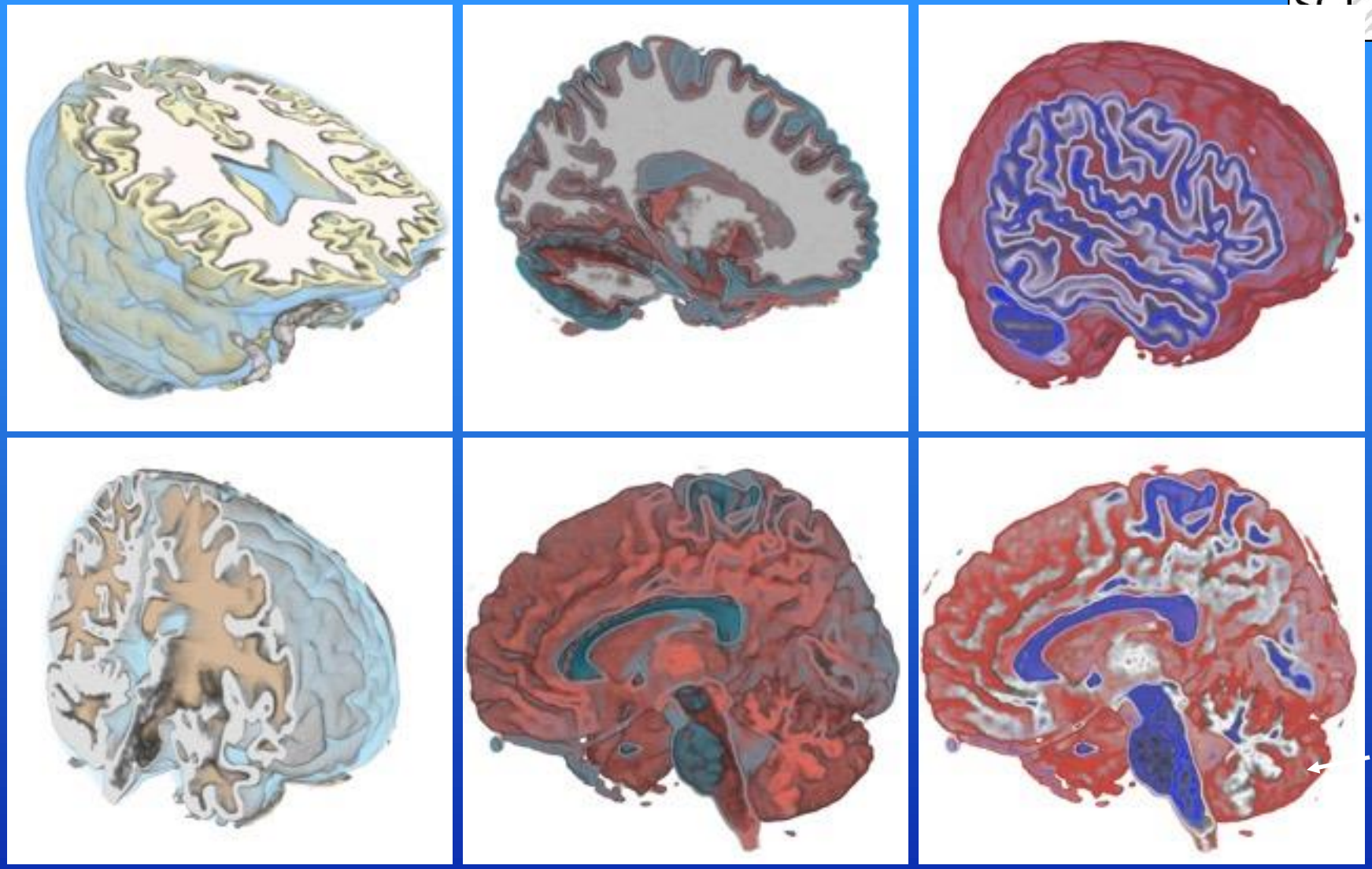


Visualization & Exploration for Distributions

- Show differences between PDFS
- Summarize all data in a single view

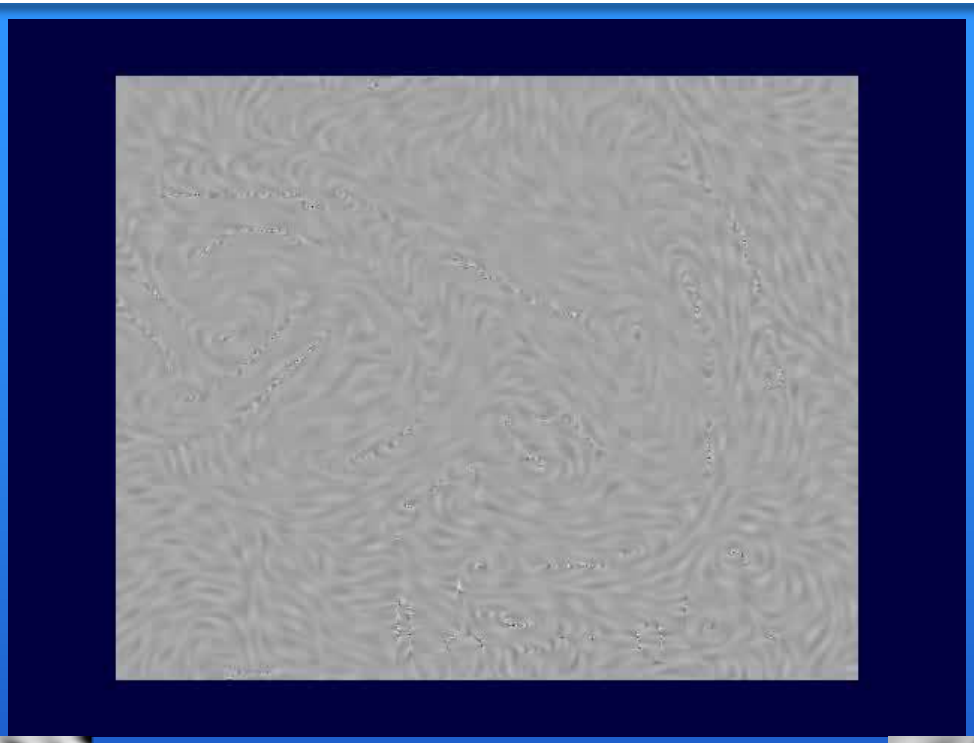
K. Potter, R.M. Kirby, D. Xiu, C.R. Johnson. "Interactive visualization of probability and cumulative density functions," In *International Journal of Uncertainty Quantification*, Vol. 2, No. 4, pp. 397--412. 2012.

Visualizing Uncertainty



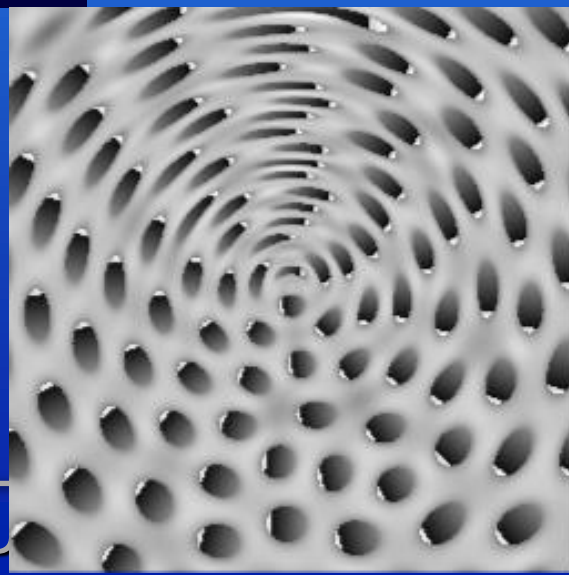
Fuzzy | Sensitivity | Confidence

Reaction Diffusion Vector Field Visualization



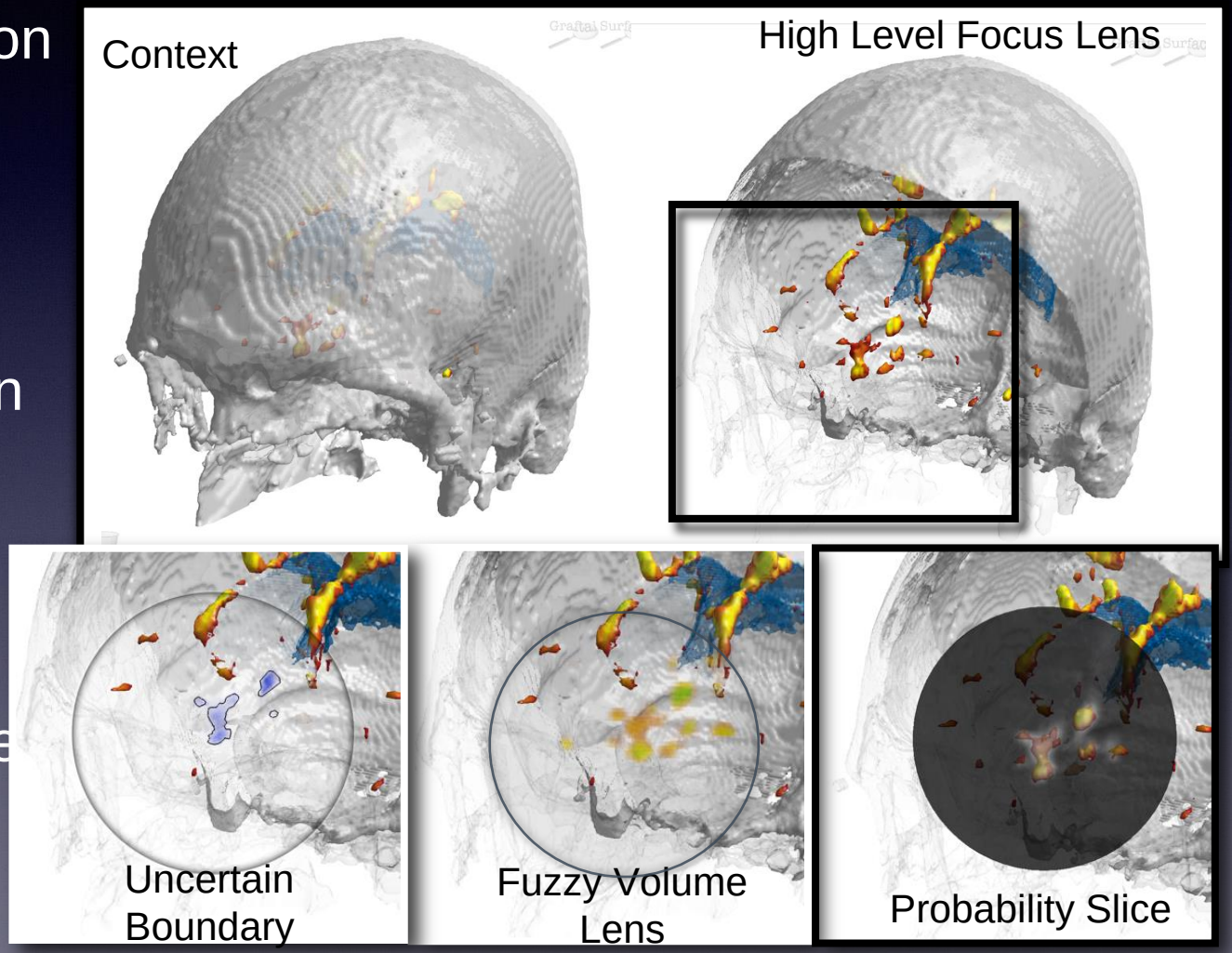
A.R. Sanderson, C.R. Johnson, R.M. Kirby. “Display of Vector Fields Using a Reaction Diffusion Model,” In *Proceeding of IEEE Visualization 2004*, pp. 115--122. 2004

A.R. Sanderson, R.M. Kirby, C.R. Johnson, L. Yang. “Advanced Reaction-Diffusion Models for Texture Synthesis,” In *Journal of Graphics Tools*, Vol. 11, No. 3, pp. 47--71. 2006.



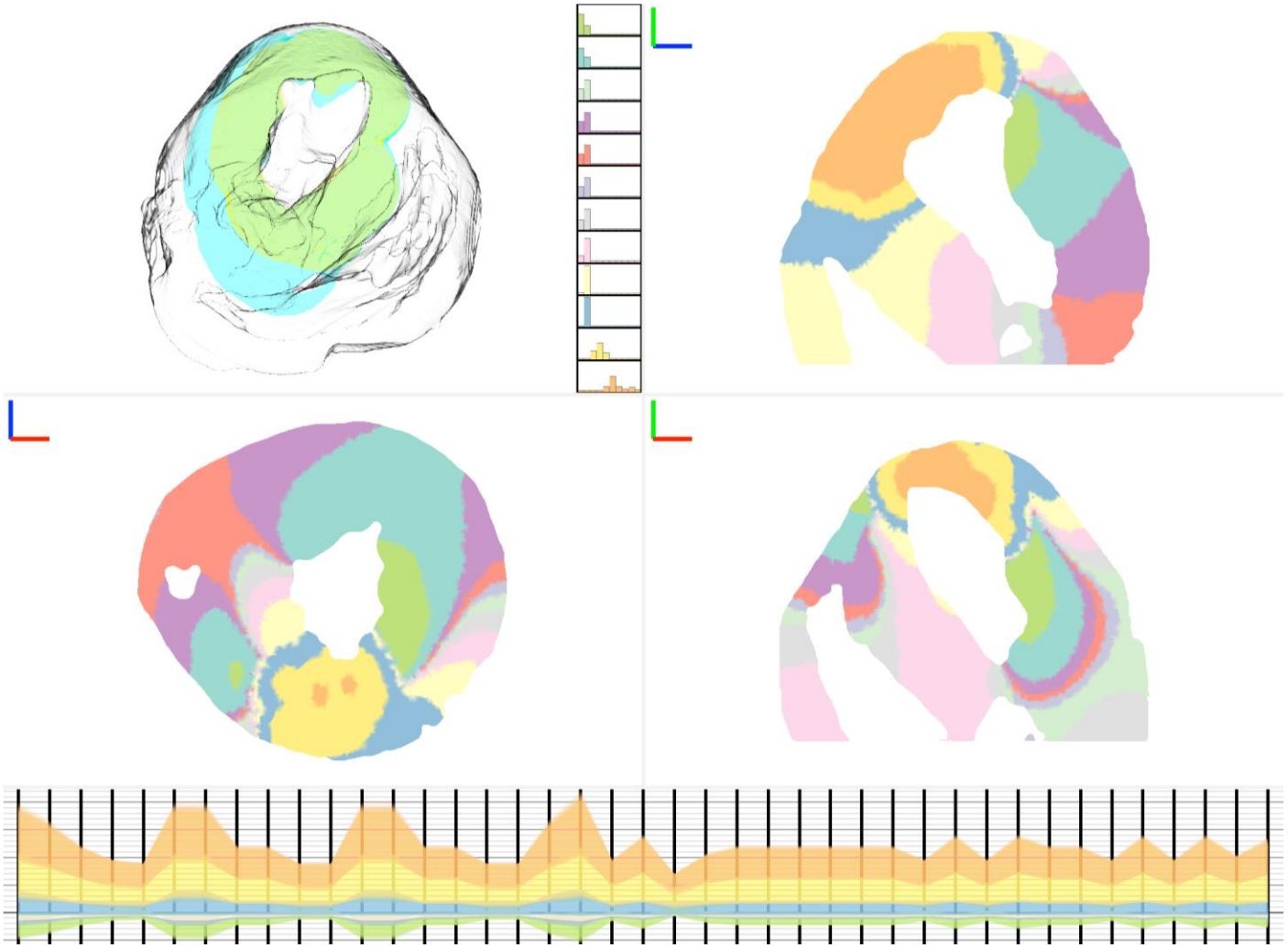
QuizLens: A Multi-lens approach for uncertainty exploration

- Global information important for qualitative evaluation & context
- Local information necessary for quantitative understanding
- Interchangeable lenses to explore various data characteristics



muView Visualization System

Visualizing uncertainty in cardiac ischemia simulations



P. Rosen, B. Burton, K. Potter, C.R. Johnson. "Visualization for understanding uncertainty in the simulation of myocardial ischemia," In *Proceedings of the 2013 Workshop on Visualization in Medicine and Life Sciences*, 2013.

Ensemble-Vis: A Framework for the Statistical Visualization of Ensemble Data

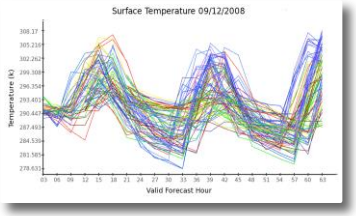
What is ensemble data?

Collection of data sets generated by computational simulations.

Used to simulate complex systems, mitigate uncertainty, unknowns in initial conditions, and parameter sensitivity.

These data sets are:

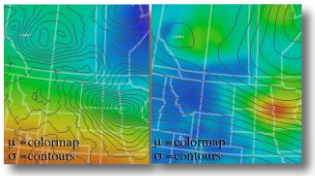
- Multidimensional
- Multivariate
- Multivalued



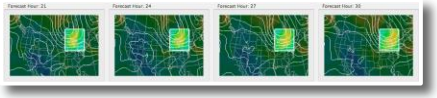
Ensemble-Vis Workflow

- User-driven
- Component-based

Ensemble Overviews

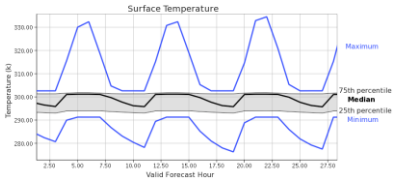


Spatial Overviews:
Mean and standard deviation encoded through colormaps and contours.

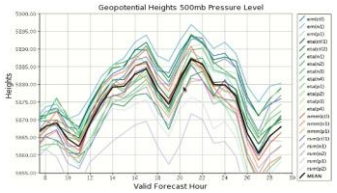


Temporal Overviews:
Filmstrip and animation.
Show evolution through time. Small multiples show every time step. User can select desired temporal location.

Trend Charts

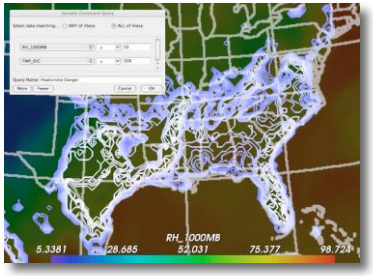


Quartile Charts:
Show minimum and maximum, innerquartile range.



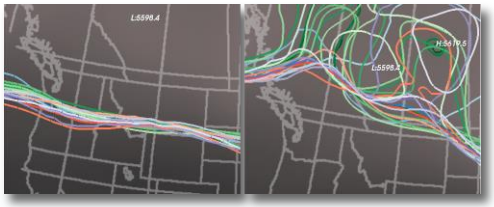
Plume Charts:
Show every member and mean.
Color coded based on model.
Deselect members to hide.
Drill-down to direct data display.

Query Contours



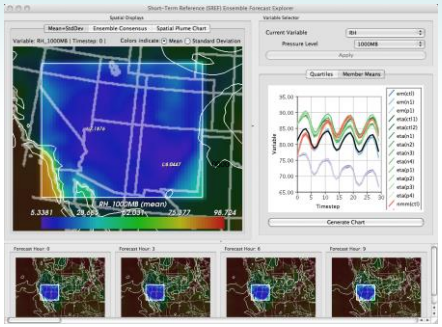
* User-driven query
* Select subset based on conditions
* Returns % of satisfying members
* Displayed as nested contours

Spaghetti Charts



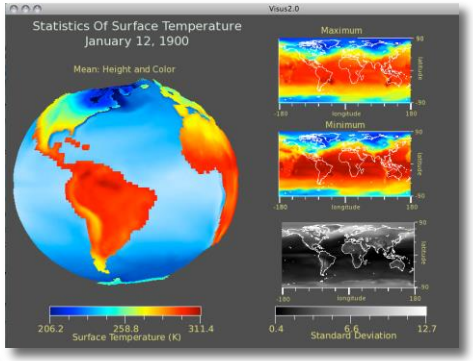
* Show variation across space
* User chosen contour value
* Isocontour for each desired member
* Highlights outliers and divergence

Implementation



SREF Weather Explorer

- VTK filters, Qt Widgets
- Relational database:
  - MySQL/ Netezza



ViSUS

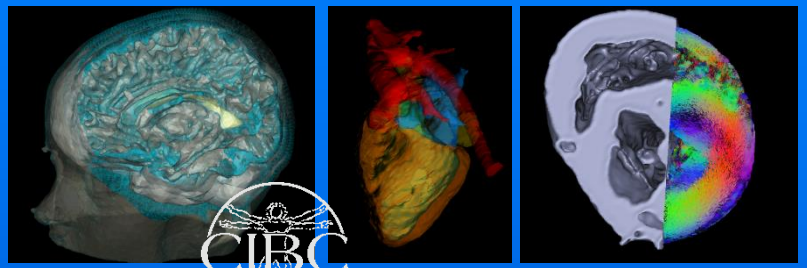
- Climate Data Analysis Tools (CDAT) integration
- C++, python, FLTK
- Out-of-core, streaming

The SCI Institute

Productivity Machines

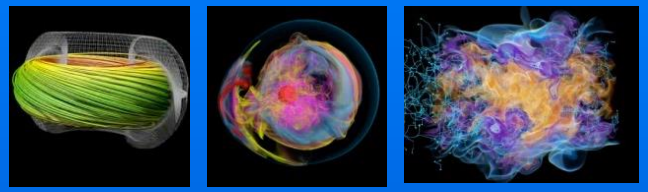
Acknowledgments

NIH/NIGMS Center for Integrative Biomedical Computing



SDAV

Scalable Data Management, Analysis and Visualization



Utah Center for Neuroimage Analysis

UTAH Center for Computational Earth Sciences

NIH NAMIC

Center for Extreme Data Management, Analysis, and Visualization

CEDMAV

IAMCS
Institute for Applied Mathematics and Computational Science

More Information

www.sci.utah.edu

crj@sci.utah.edu