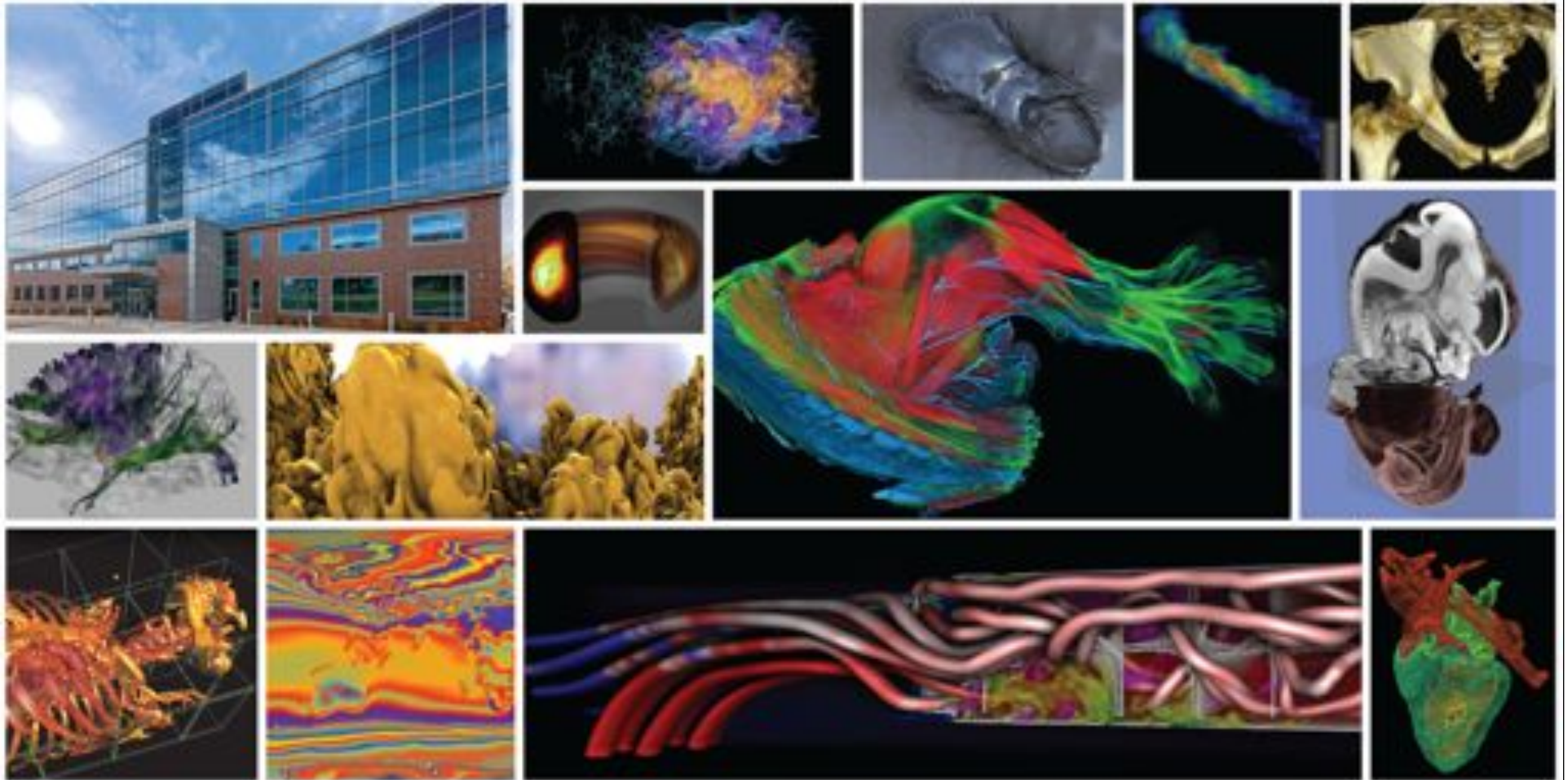
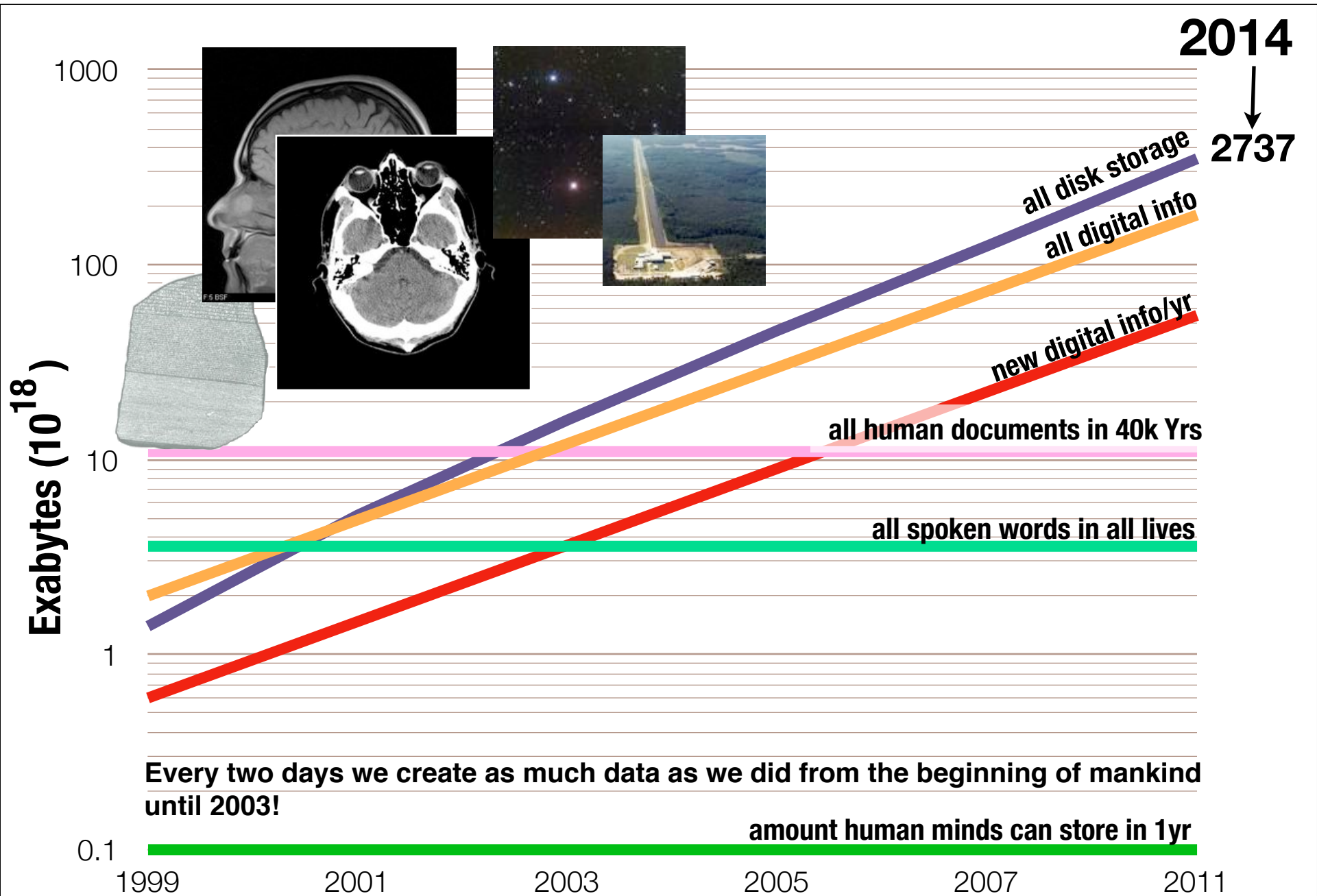


# Data to Insight to Change



**Chris Johnson**  
**Scientific Computing and Imaging Institute**  
**University of Utah**



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

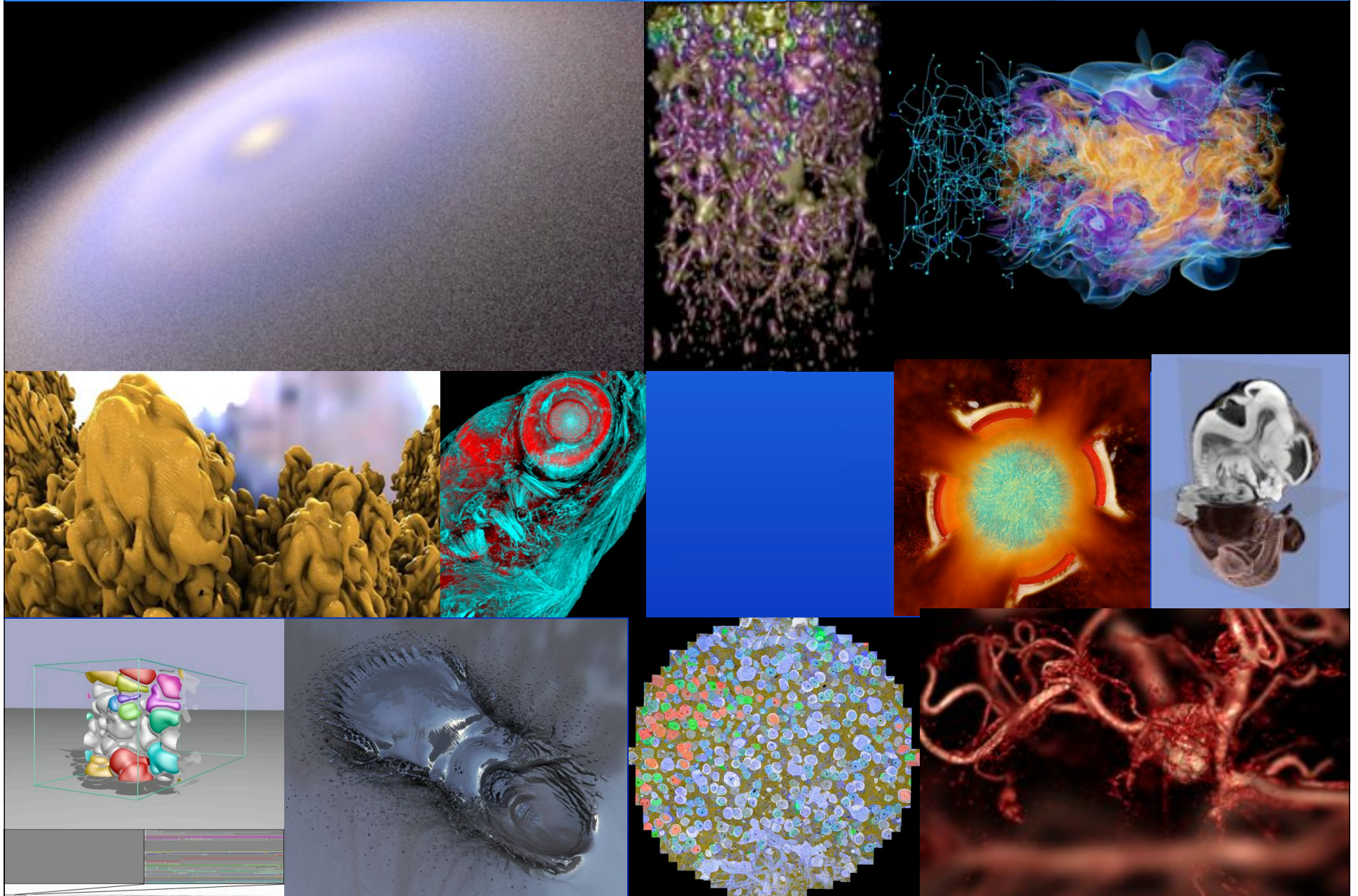
# Big Data

**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...**

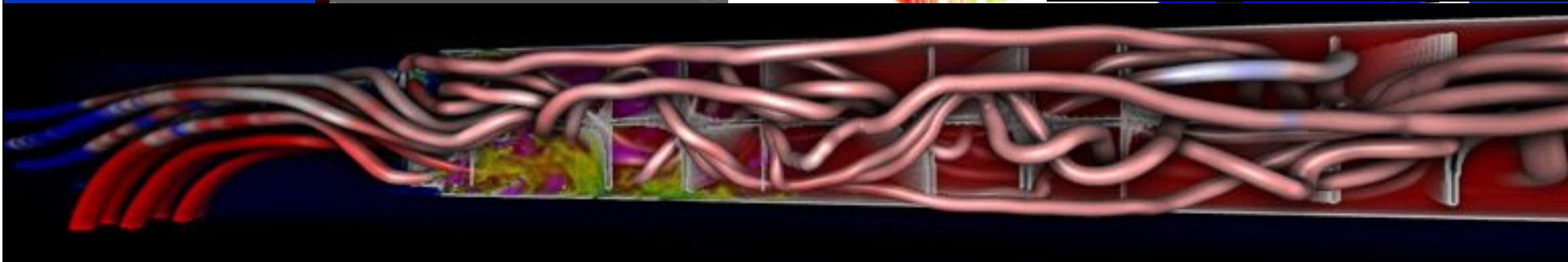
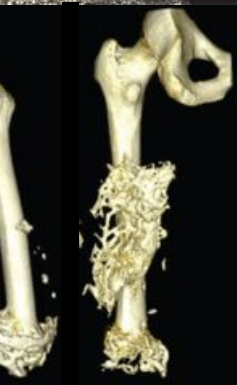
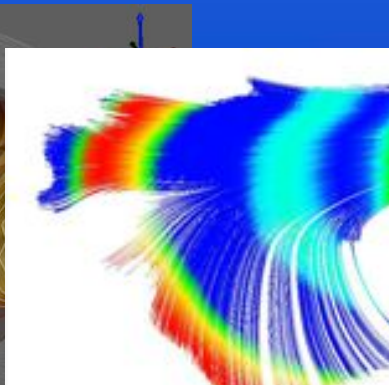
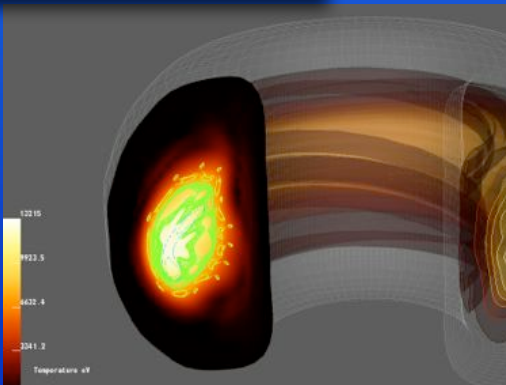
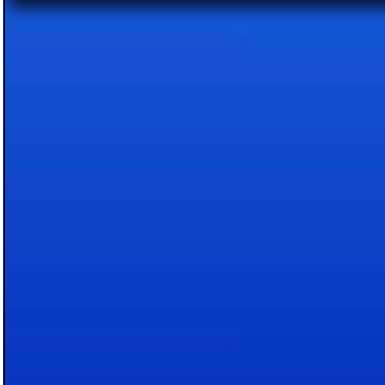
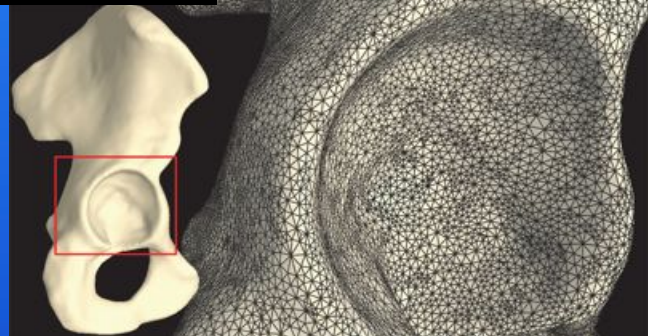
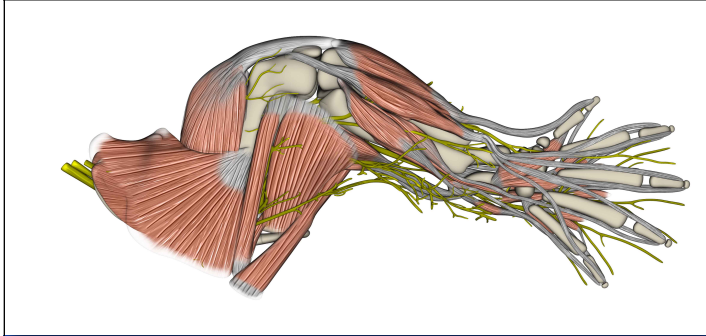
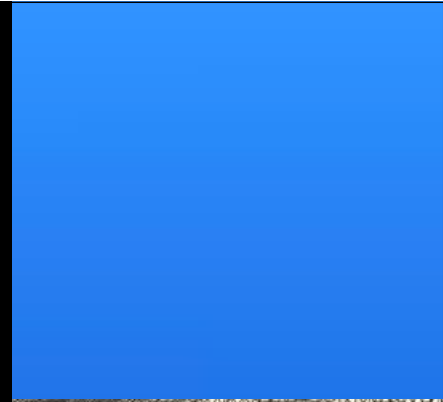
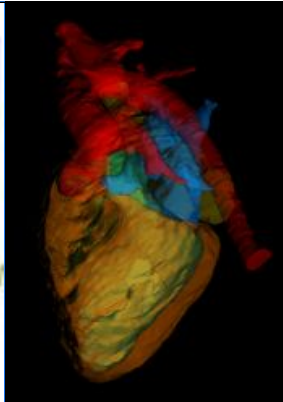
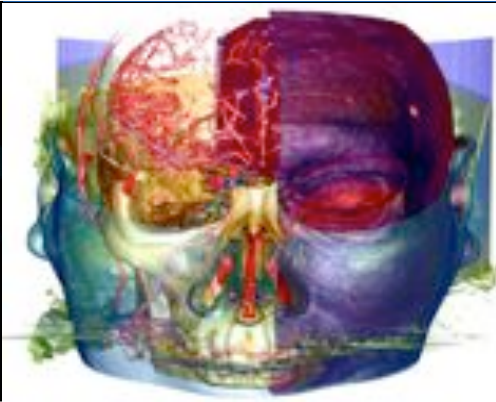
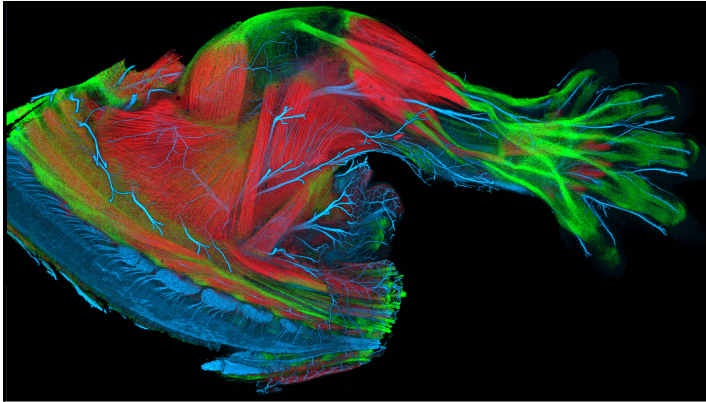
**Dan Ariely**



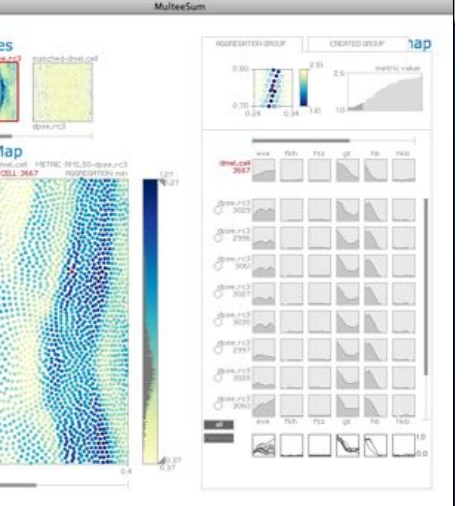
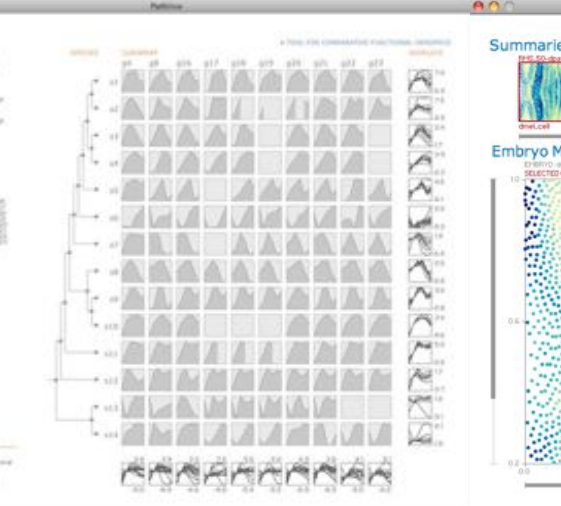
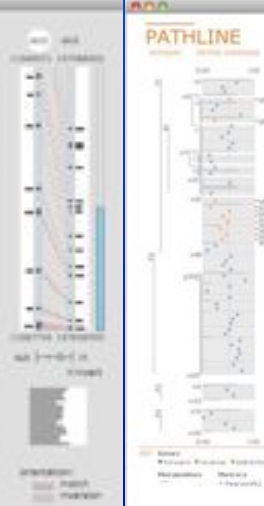
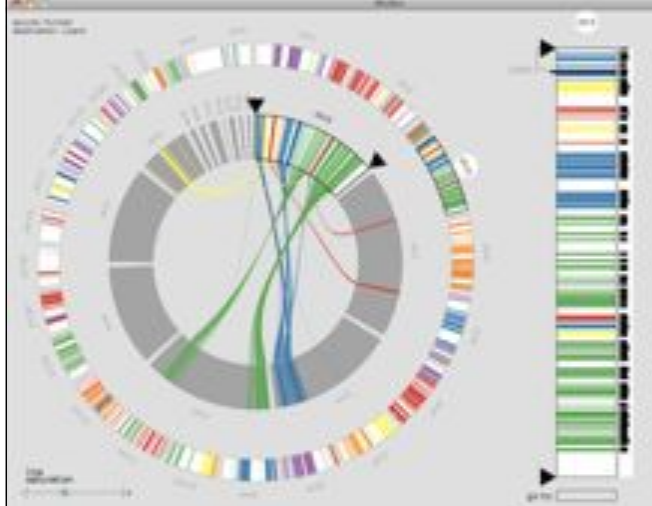
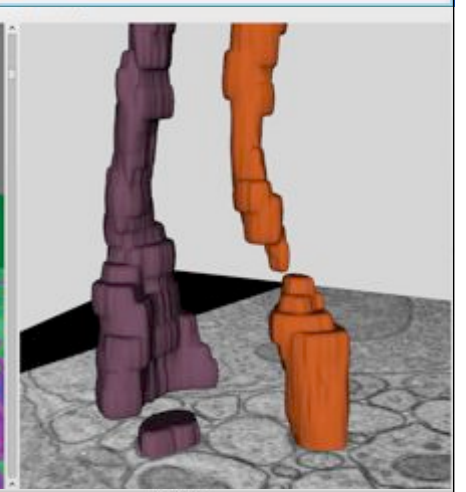
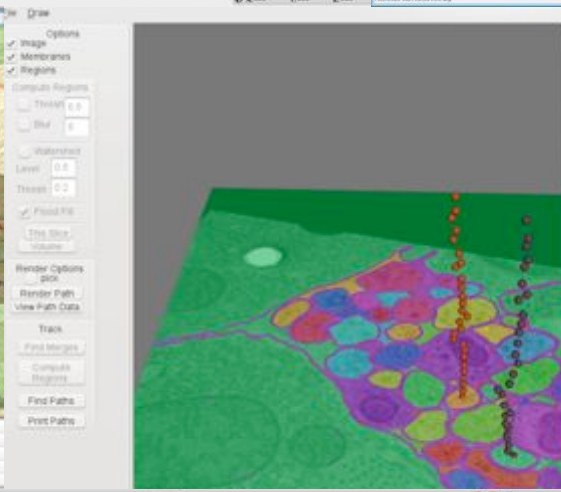
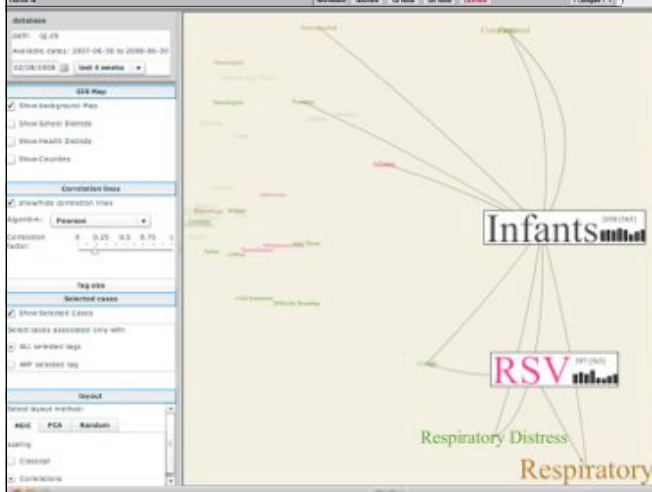
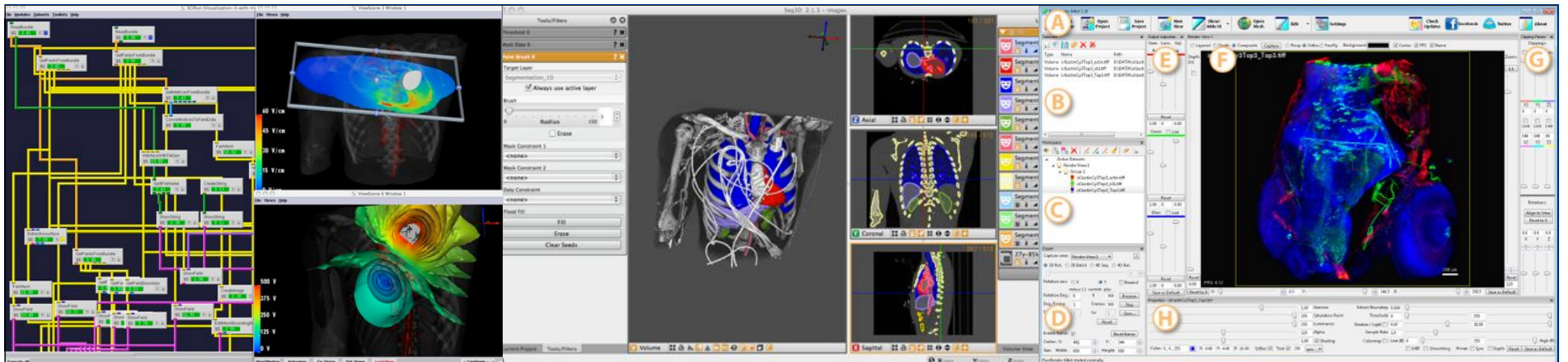
# New Visual Analysis Techniques

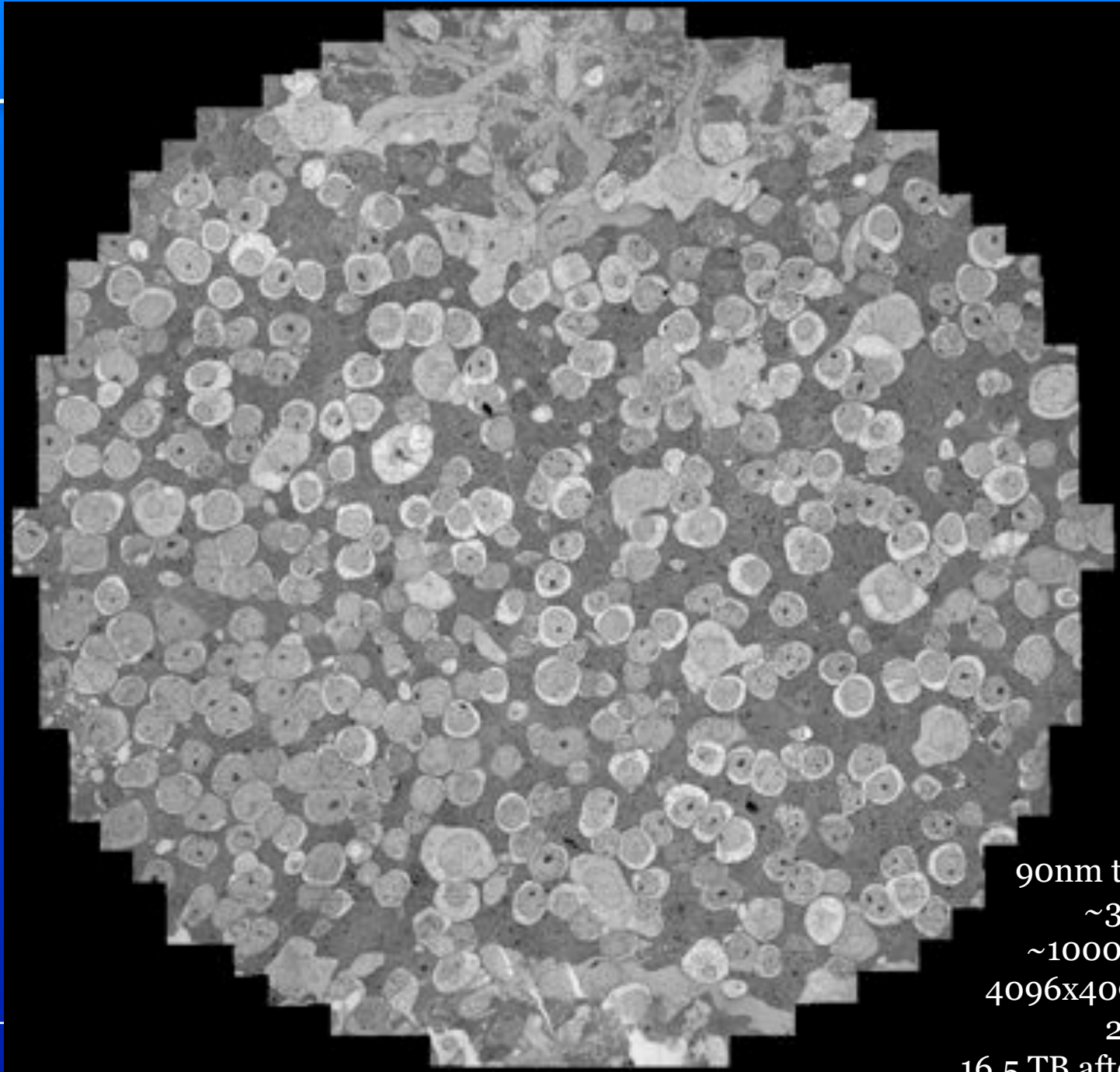












341 Sections  
90nm thick sections  
~32GB/Section  
~1000 tiles/section  
4096x4096 pixels/tile  
2.18 mm/Pixel  
16.5 TB after processing



# Antony van Leeuwenhoek (1632-1723)



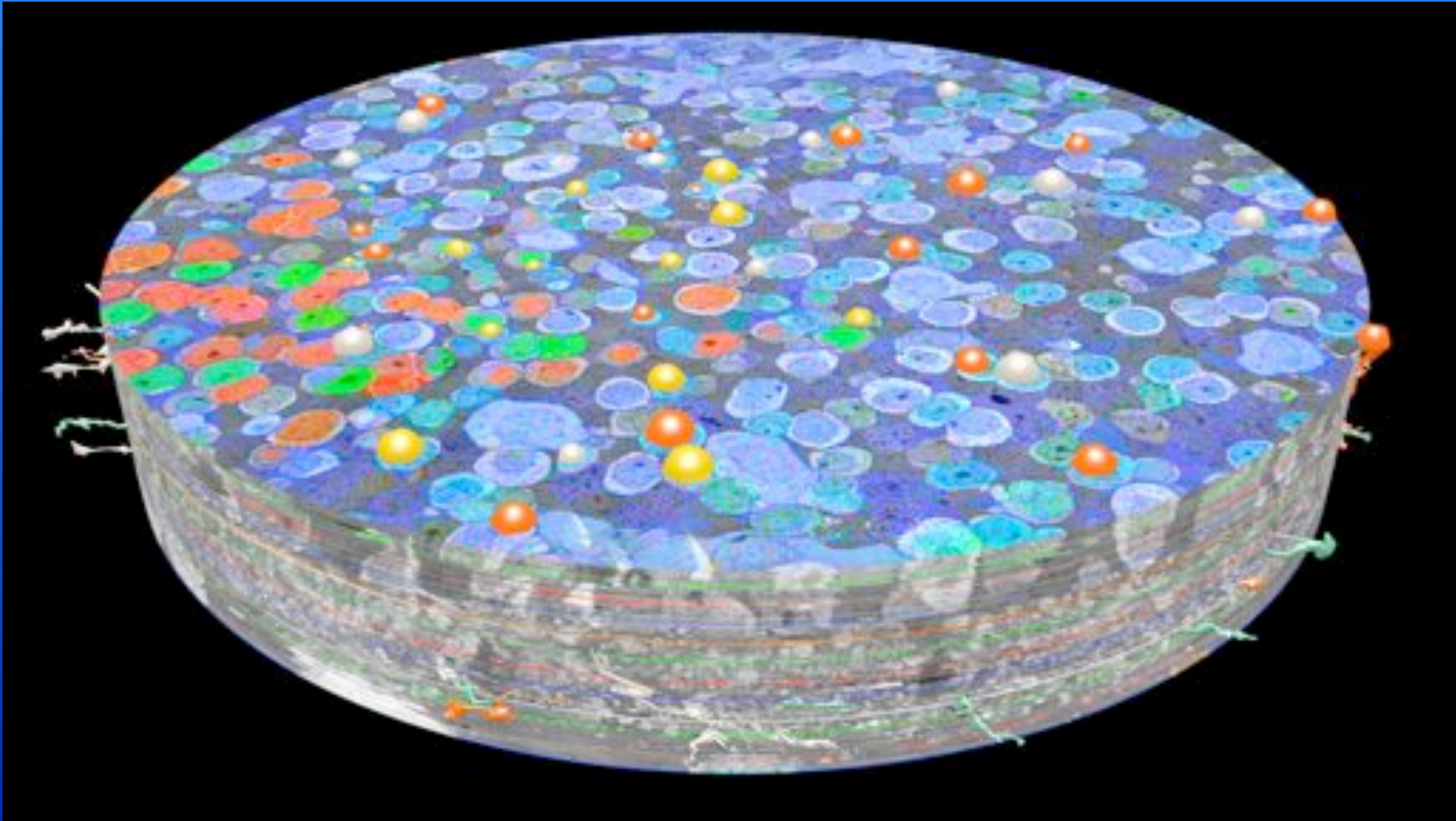
*... my work, which I've done for a long time, was not pursued in order to gain the praise I now enjoy, but chiefly from a craving after knowledge, which I notice resides in me more than in most other men. And therewithal, whenever I found out anything remarkable, I have thought it my duty to put down 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 biological data*

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

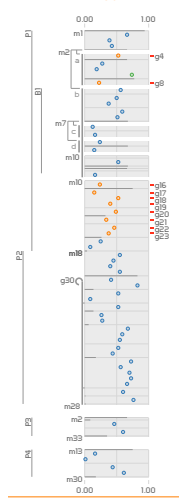


M. Meyer et al., EuroVis 2010.

# Pathline

## PATHLINE

PATHWAY METRIC OVERVIEW

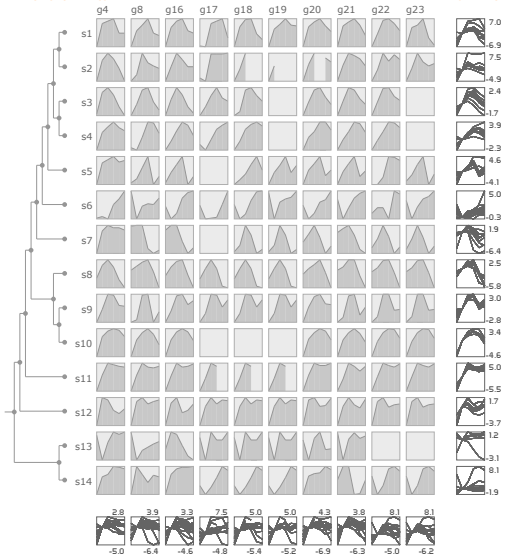


SPECIES

CURVEMAP

A TOOL FOR COMPARATIVE FUNCTIONAL GENOMICS

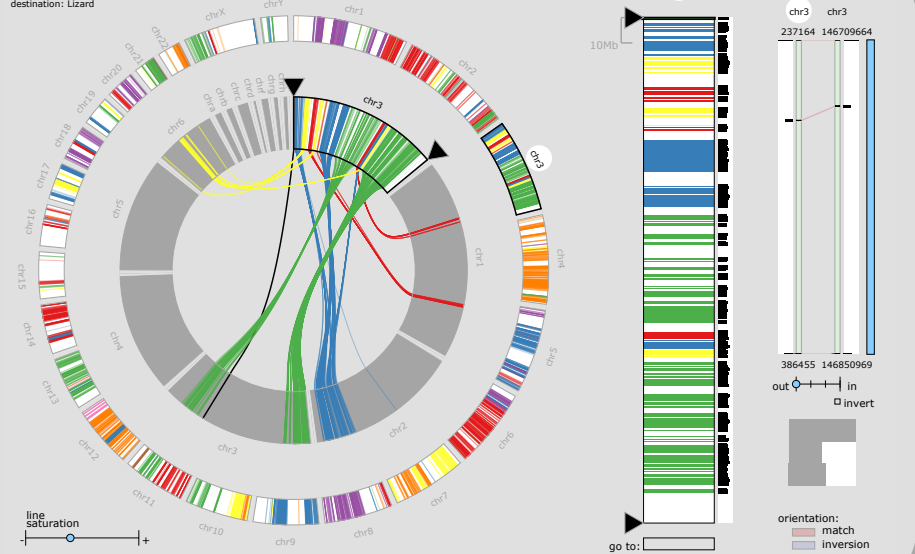
OVERLAYS



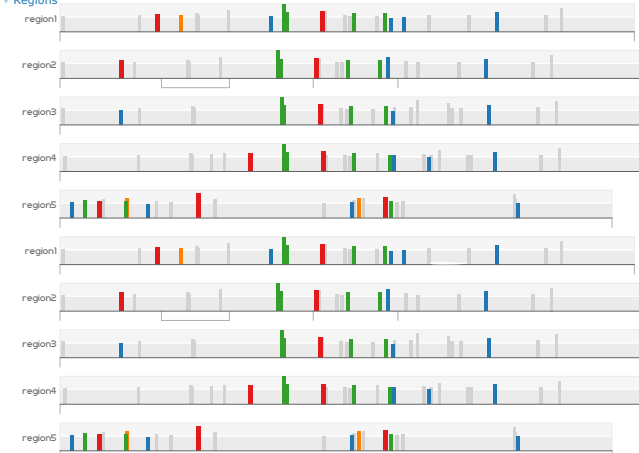
# MizBee

M. Meyer et al., InfoVis 2009.

source: Human  
destination: Lizard



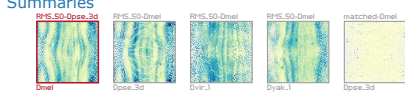
Regions



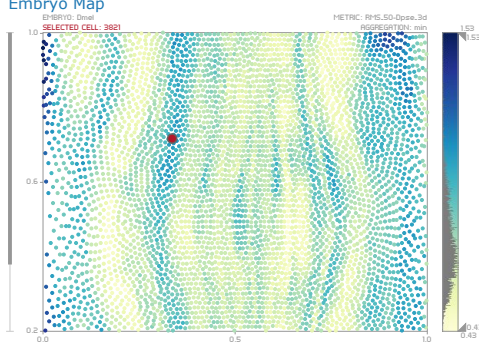
Alignment



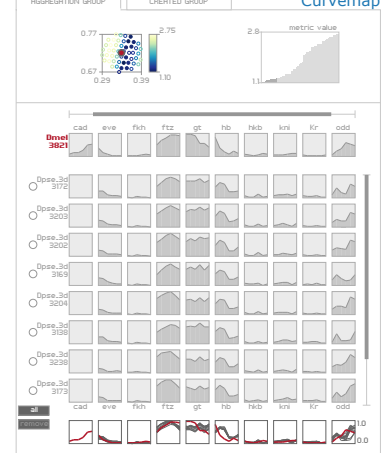
Summaries



Embryo Map



REGREGATION GROUP CREATED GROUP Curvemap



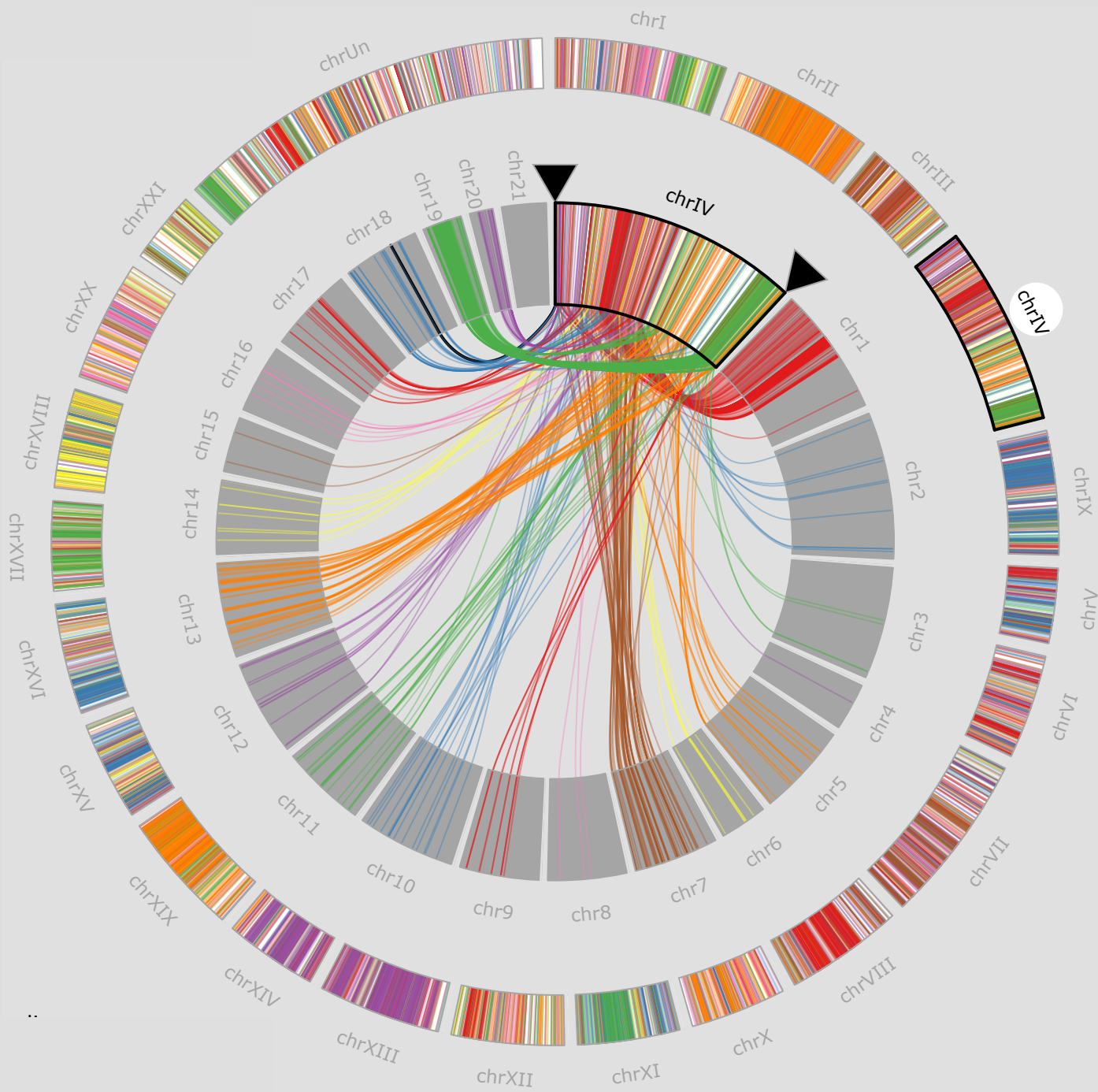
# InSite

# MulteeSum

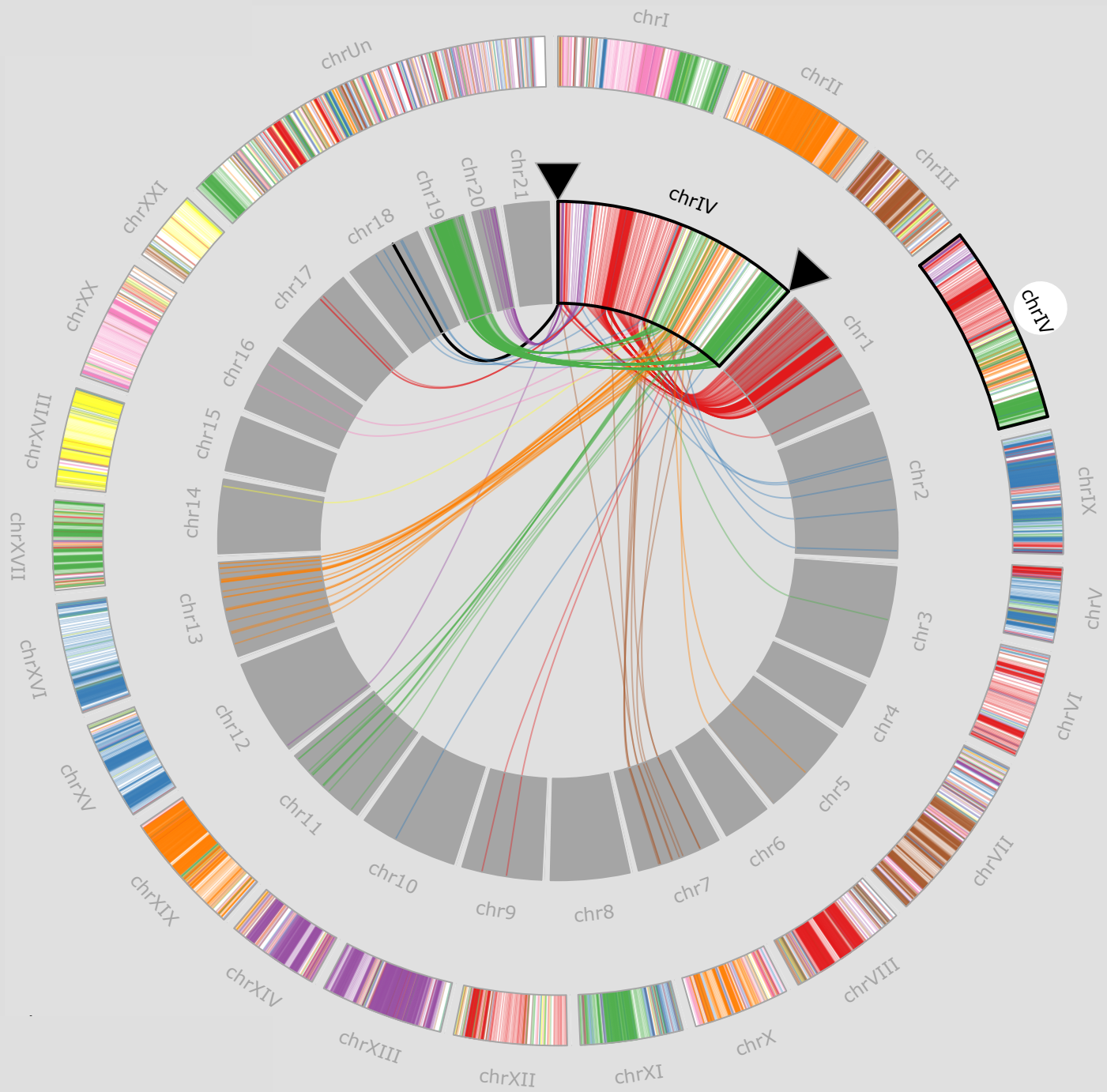
M. Meyer et al., InfoVis 2010.

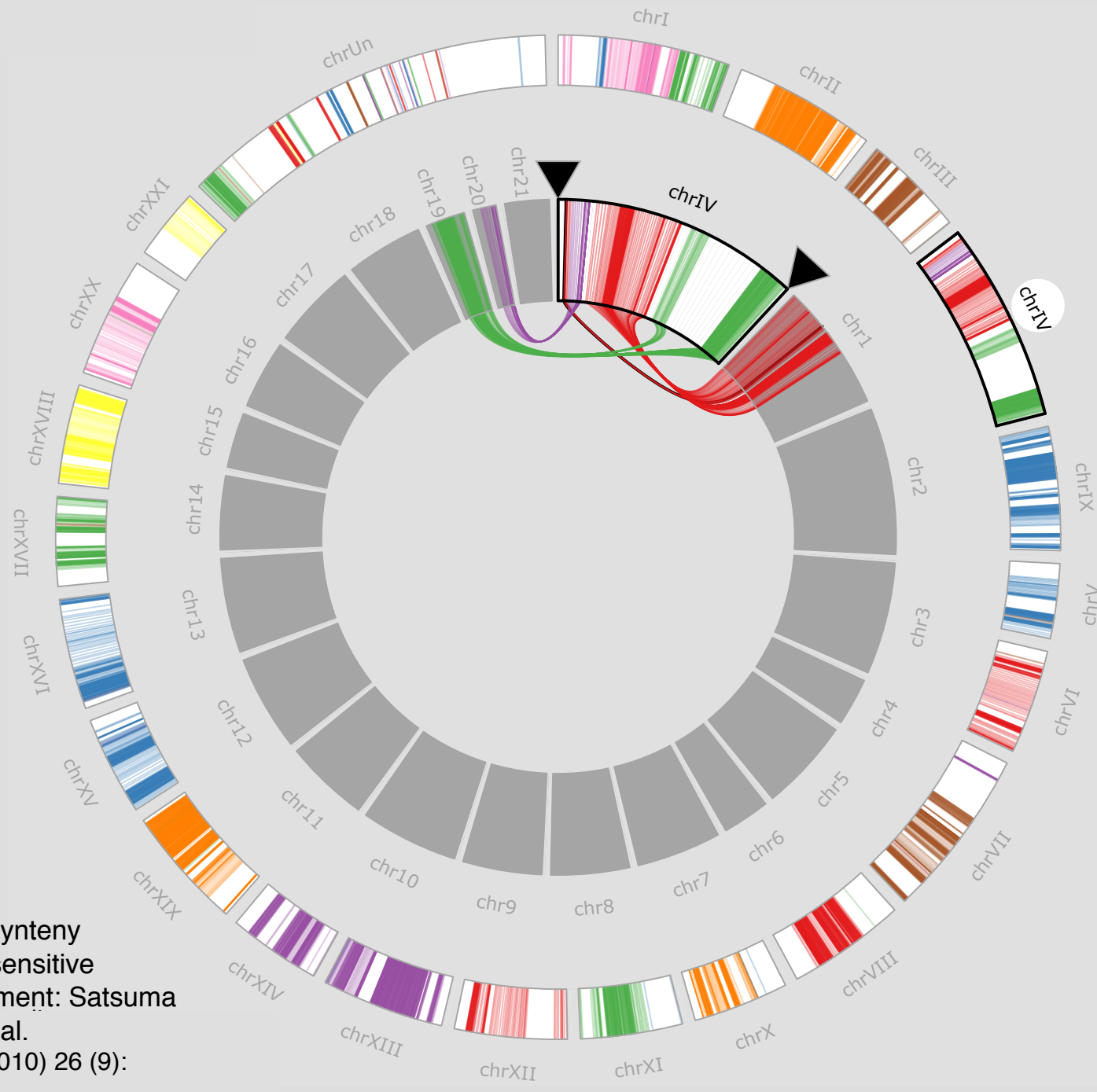








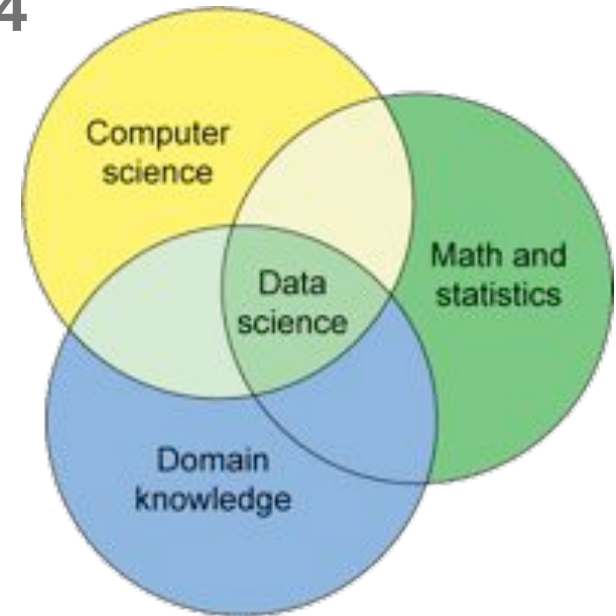




Genome-wide synteny through highly sensitive sequence alignment: Satsuma  
 M. Grabherr, et al.  
 Bioinformatics (2010) 26 (9): 1145-1151.

# Data Science Programs

- [http://analytics.ncsu.edu/?page\\_id=4184](http://analytics.ncsu.edu/?page_id=4184)
- 19 MS programs in Data Analytics
- 8 MS programs in Data Science
- 28 MS programs in Business Analytics
- Several additional “tracks” or “concentration” programs





# Big Data Curriculum

## Analytics Electives:

- **Data Mining** (required)
- **Machine Learning** (required)
- **Visualization** (required)
- **Artificial Intelligence.**  
Decision making under uncertainty.
- **Natural Language Processing.**  
Understanding textual data and language.
- **Probabilistic Modeling.**  
Advanced statistical techniques and tools (using R).
- **Image Processing.**  
Analysis and learning on image data.

# Big Data Curriculum

## Algorithmics Electives:

- **Advanced Algorithms** (required)
- **Models of Computation for Big Data.**  
How algorithmic bottlenecks change as data becomes very large;  
Relation to modern big data systems (e.g. MapReduce).
- **Computational Geometry.**  
Geometric interpretation of big data analysis and computation.
- **Computational Topology.**  
Topological data analysis and algorithms.

# Big Data Curriculum

## Management Electives:

- **Database Systems** (required)
- **Parallel Programming for Many-Core Architectures.**  
**Parallel Computing and High Performance Computing.**  
Scalable programming on GPUs, many-cores, and HPC clusters.
- **Advanced Computer Networks.**  
Large-scale network protocols, architectures, and applications.
- **Network Security.**  
Message integrity, access control, authentication, confidentiality.



# Piloting in Adobe (Lehi)

- Starting Fall 2014.
- Live 2-way streaming. Interaction across video.  
Fall 2014: Visualization: T-Th 9:10 - 10:30am  
Fall 2014: Adv. Algorithms: T-Th 10:45 - 12:05am  
*(plan for early evening, e.g. Data Mining M-W 5:15-6:35pm)*
- Potential for Instructor on site in future.  
Adobe lecture room open to others.

# The SCI Institute

---





# Productivity Machines

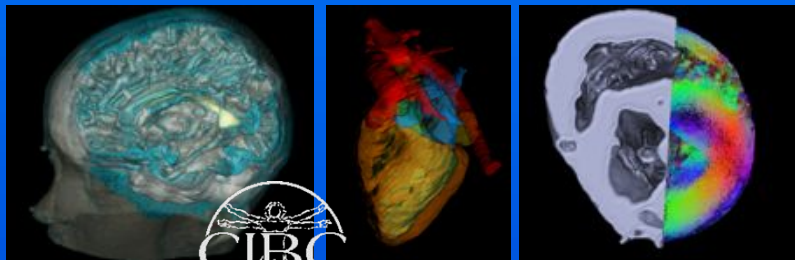




# Acknowledgments

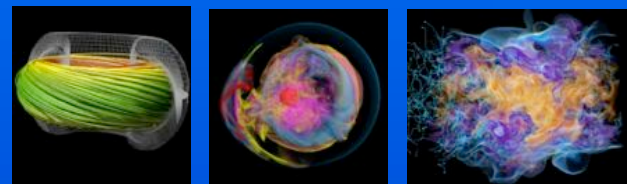


**NIH/NIGMS Center for Integrative Biomedical Computing**



# SDAV

**Scalable Data Management, Analysis and Visualization**



**Utah Center for Neuroimage Analysis**



**UTAH** Center for Computational Earth Sciences

**Center for Extreme Data Management, Analysis, and Visualization**



**NIH NAMIC**



**IAMCS**  
Institute for Applied Mathematics and Computational Science



# More Information

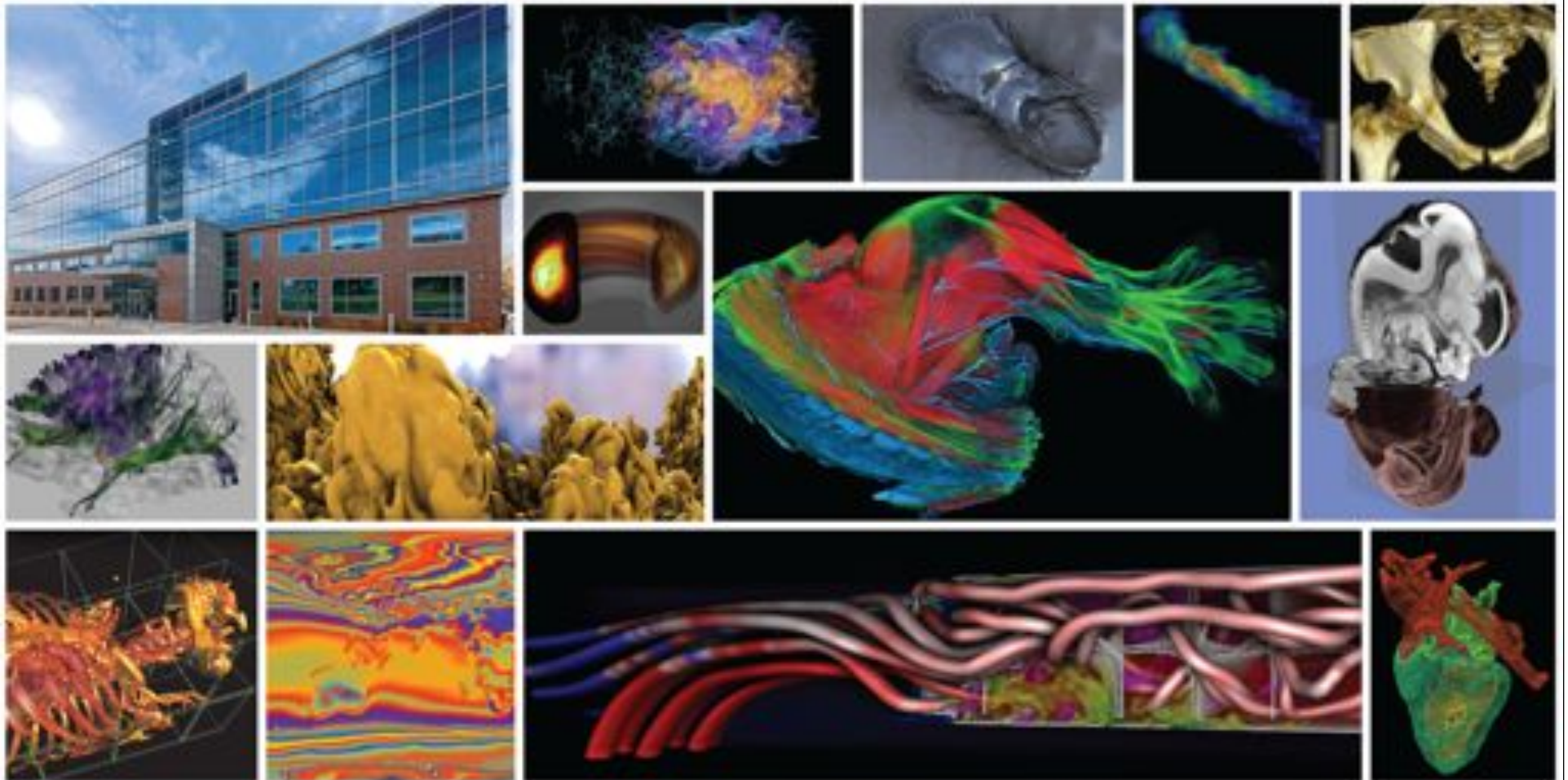
---



[www.sci.utah.edu](http://www.sci.utah.edu)

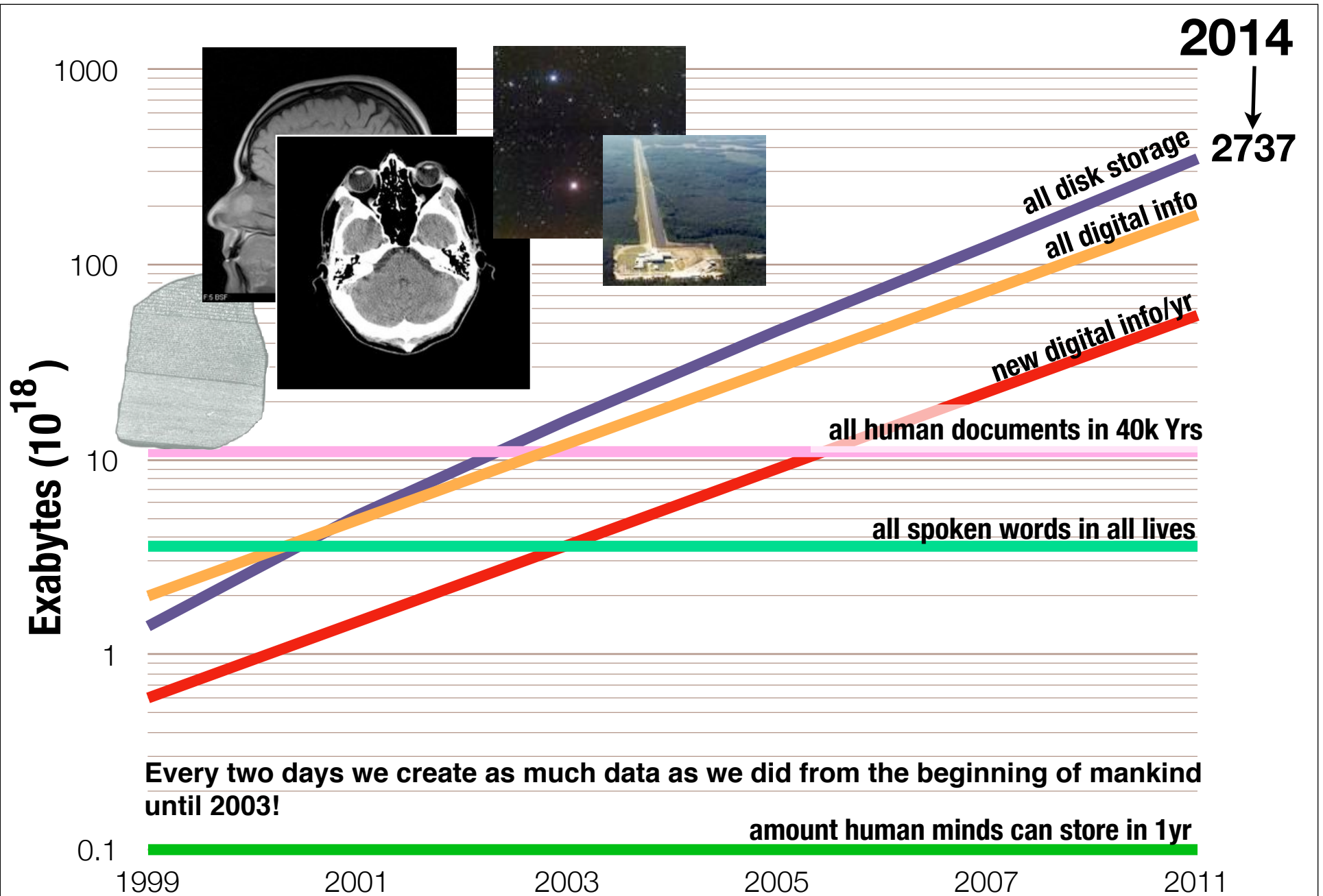
[crj@sci.utah.edu](mailto:crj@sci.utah.edu)

# Ecosystem Challenges Around Data Use



**Chris Johnson**  
**Scientific Computing and Imaging Institute**  
**University of Utah**





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

# 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**

# Panelists

**Leonid Zhukov - Director of Data Science, Ancestry.com**

**Vance Checketts - VP and GM of EMC**

**Edison Ting, Solutions Architect, Pivotal**