



Please note our measures
concerning Coronavirus / Covid 19

SCHLOSS DAGSTUHL
Leibniz-Zentrum für Informatik



About Dagstuhl

Program

Publications

Library

dblp

Seminar Calendar

All Events

Dagstuhl Seminars

Dagstuhl Perspectives

GI-Dagstuhl Seminars

Summer Schools

Events

Research Guests

Expenses

Planning your visit

You are here: Program » Seminar Calendar » Seminar Homepage

<https://www.dagstuhl.de/21152>

April 11 – 16 , 2021, Dagstuhl Seminar 21152

Multi-Level Graph Representation for Big Data Arising in Science Mapping

Organizers

Katy Börner (Indiana University – Bloomington, US)

Stephen G. Kobourov (University of Arizona – Tucson, US)

For support, please contact

Susanne Bach-Bernhard for administrative matters

Shida Kunz for scientific matters

Documents

List of Participants

Shared Documents

Dagstuhl Seminar Wiki

Dagstuhl Seminar Schedule (Upload here)

(Use personal credentials as created in DOOR to log in)

Documentation

In the series Dagstuhl Reports each Dagstuhl Seminar and Dagstuhl Perspectives Workshop is documented. The seminar organizers, in cooperation with the collector, prepare a report that includes contributions from the participants' talks together with a summary of the seminar.

Download [📄](#) overview leaflet (PDF).

Publications

Furthermore, a comprehensive peer-reviewed collection of research papers can be published in the series Dagstuhl Follow-Ups.

Dagstuhl's Impact

Please inform us when a publication was published as a result from your seminar. These publications are listed in the category Dagstuhl's Impact and are presented on a

On-site participants

- **Katy Börner** (Indiana University – Bloomington, US) [dblp]
- **Ingo Günther** (Worldprocessor Studio – New York, US)
- **Francis Harvey** (Leibniz Institut für Länderkunde – Leipzig, DE) [dblp]
- **Michael Kaufmann** (Universität Tübingen, DE) [dblp]
- **Alexander Wolff** (Universität Würzburg, DE) [dblp]

Remote participants

- **Patrizio Angelini** (John Cabot University – Rome, IT) [dblp]
- **Michael A. Bekos** (Universität Tübingen, DE) [dblp]
- **Kevin Boyack** (SciTech Strategies Inc. – Albuquerque, US)
- **David Chavalarias** (CNRS – Paris, FR)
- **Giuseppe Di Battista** (University of Rome III, IT) [dblp]
- **Sara Irina Fabrikant** (Universität Zürich, CH) [dblp]
- **Jean-Daniel Fekete** (INRIA Saclay – Orsay, FR) [dblp]
- **Lynda Hardman** (CWI – Amsterdam, NL) [dblp]
- **Ben Jacobsen** (University of Arizona – Tucson, US)
- **Philipp Kindermann** (Universität Trier, DE) [dblp]
- **Karsten Klein** (Universität Konstanz, DE) [dblp]
- **Stephen G. Kobourov** (University of Arizona – Tucson, US) [dblp]
- **Thomas Köhler** (TU Dresden, DE)
- **Vincent Larivière** (University of Montreal, CA)
- **Tamara Mchedlidze** (Utrecht University, NL) [dblp]
- **Guy Melançon** (University of Bordeaux, FR) [dblp]
- **Staša Milojevic** (Indiana University – Bloomington, US)
- **Filipi Nascimento Silva** (Indiana University – Bloomington, US)

- **Martin Nöllenburg** (TU Wien, AT) [dblp]
- **Adam Ploszaj** (University of Warsaw, PL)
- **Sergey Pupyrev** (Facebook – Menlo Park, US) [dblp]
- **Chrysanthi Raftopoulou** (National Technical University of Athens, GR) [dblp]
- **Andrea Scharnhorst** (Royal Netherlands Academy of Arts and Sciences, NL) [dblp]
- **André Skupin** (San Diego State University, US) [dblp]
- **Cassidy Rose Sugimoto** (Indiana University – Bloomington, US)
- **Antonios Symvonis** (National TU – Athens, GR) [dblp]
- **Markus Wallinger** (TU Wien, AT)
- **Angela Zoss** (Duke University – Durham, US)

Motivation

For centuries, cartographic maps have guided human exploration. While being rather imperfect initially, they helped explorers find promised lands and return home safely. Recent advances in data, algorithms, and computing infrastructures make it possible to map humankind's collective scholarly knowledge and technology expertise by using topic maps on which “continents” represent major areas of science (e.g., mathematics, physics, or medicine) and zooming reveals successively more detailed subareas. Basemaps of science and technology are generated by analyzing citations links between millions of publications and/or patents. “Data overlays” (e.g., showing all publications by one scholar, institution, or country or the career trajectory of a scholar as a pathway) are generated by science-locating relevant publication records based on topical similarity. Despite the demonstrated utility of such maps, current approaches do not scale to the hundreds of millions of data records now available. The main challenge is designing efficient and effective methods to visualize and interact with more than 100 million scholarly publications at multiple levels of resolution.

This Dagstuhl Seminar will bring together researchers in cartography, information visualization, science of science, and graph drawing to discuss novel graph mining and layout algorithms and their application to the development of science mapping standards and services. We will also organize an exhibition of art contributed by scientists and science maps contributed by artists, and use this to stimulate discussion.

This Dagstuhl Seminar will bring together researchers in cartography, information visualization, science of science, and graph drawing to discuss novel graph mining and layout algorithms and their application to the development of science mapping standards and services. We will also organize an exhibition of art contributed by scientists and science maps contributed by artists, and use this to stimulate discussion.

One of our main goals is to create a special journal issue on the topics related to the development of the next generation science map, SciMap2020. Among others, we will aim to answer:

1. What data is most robust to design SciMap2020?
2. What information should the nodes and link encode in SciMap2020?
3. What user experience metaphors, functionality, and interactive user interfaces are best?
4. What user studies should be run to compare options and study knowledge gains?

On the algorithmic front, specific research problems include:

1. Construct a hierarchy of graphs, so that each represents the underlying large graph well
2. Study the complexity of underlying problem: computing multi-level graph sketches
3. Design efficient algorithms for computing multi-level graph sketches and clustering
4. Design graph layout algorithms, driven by multi-level sketches and clustering

We look forward to a stimulating week that brings together world-leading experts to tackle these research challenges.

Monday

07:30-08:45 Breakfast

09:00-09:15 Welcome by Katy Börner

09:15-10:00 Brief Introductions / Research Overviews

- Francis Harvey, University of Leipzig
- [David Chavalarias](#), ISC-PIF
- Mohammad Khaledur Rahman, IU, "[BatchLayout: A batch-parallel force-directed graph layout algorithm in shared memory](#)"

10-10:30 Overview Talk

- Katy Börner, IU: "Big Data Arising in Science Mapping" (30mins)

10:30-11:00 Coffee Break

11:00-12:00 Brief Introductions / Research Overviews

- Thomas Koehler, TU Dresden
- Kilian Buehling, TU Dresden
- Andrea Scharnhorst, DANS, NL; <https://pure.knaw.nl/portal/en/persons/andrea-scharnhorst>
- Ingo Günther, <https://ingogunther.com/worldprocessor>

12:15-13:00 Lunch

Brief Introductions / Research Overviews

- Title slide with name, affiliation(s), picture
- Main research areas
- **Research relevant for this Dagstuhl seminar**
- Why did you decide to attend this seminar?
- What paper might you contribute to the July/Aug 2022 special issue in *IEEE Computer Graphics and Applications* on “Multi-Level Graph Representations for Big Data in Science”?

12:15-13:00 Lunch

14:00 Introduction of Medina who will serve as the VCA Monday-Thursday 14:00-20:00.

14:10-15:00 Overview Talks (30mins each)

- Stephen Kobourov, UA: Welcome by Stephen & Multi-Level Graph Representations
- Sara Fabrikant, Department of Geography, University of Zürich: Cognitive research on landmarks, color, uncertainty & [The Future of Geographic Information Displays](#)
Maybe better for this late in the evening: [Susceptibility of domain experts to color manipulation indicate a need for design principles in data visualization](#)

15:00-15:30 Coffee Break / Katy welcomes all to [Amatria's Virtual Birthday Party](#)

15:30-16:30 Brief Introductions / Research Overviews (10mins each)

- Andreas Bueckle, IU: [Science Map Metaphors: A Comparison of Network Versus Hexmap-Based Visualizations](#)
- Angela Zoss, Duke U: XXX
- Pino di Battista: Schematic Representation of Large Biconnected Graphs
- Alexander Wolff, U Würzburg: Using the Metro Map Metaphor for Drawing Hypergraphs
- Guy Melançon, Vice President of Digital Transformation, Université de Bordeaux*
- Adam Ploszaj, University of Warsaw

16:30-18:00 Challenges & Opportunities and Discussion of Topics for Special Issue in IEEE Computer Graphics and Applications (Katy and Stephen)

18:00-19:00 Dinner

19:00-20:00 **WEBINAR** Special Event: [Dagstuhl Exhibit Debut](#) with Francis Harvey & Katy Börner

Call for Papers: Special Issue on Multi-level Graph Representations for Big Data in Science

CG&A seeks submissions for this upcoming special issue.

The July/Aug 2022 special issue in *IEEE Computer Graphics and Applications* on “Multi-Level Graph Representations for Big Data in Science”

Articles due for review:
December 29, 2021

Guest editors:

- Katy Börner, Indiana University, Bloomington, US
- Stephen G. Kobourov, University of Arizona, Tucson, US

<https://www.computer.org/digital-library/magazines/cg/call-for-papers-special-issue-on-multi-level-graph-representations-for-big-data-in-science>

For centuries, cartographic maps have guided human exploration. While being rather imperfect initially, they helped explorers find promised lands and return home safely. Recent advances in data, algorithms, and computing infrastructures make it possible to map humankind's collective scholarly knowledge and technology expertise by using topic maps on which “continents” represent major areas of science (e.g., mathematics, physics, or medicine) and zooming reveals successively more detailed subareas. Basemaps of science and technology are generated by analyzing citations links between millions of publications and/or patents. “Data overlays” (e.g., showing all publications by one scholar, institution, or country or the career trajectory of a scholar as a pathway) are generated by science-locating relevant publication records based on topical similarity. Despite the demonstrated utility of such maps, current approaches do not scale to the hundreds of millions of data records now available. The main challenge is designing efficient and effective methods to visualize and interact with more than 100 million scholarly publications at multiple levels of resolution.

This special issue invites researchers in cartography, data visualization, science of science, graph drawing, and other domains to submit novel and promising new research on graph mining and layout algorithms and their application to the development of science mapping standards and services. Topics of interest include:

- Science of science user needs and applications
- Efficient multi-level graph algorithms
- Network visualizations
- Effective user interfaces to large-scale data visualizations

Deadlines

Submissions due: 29 December 2021

Preliminary notification: 2 March 2022

Revisions due: 6 April 2022

Final notification: 11 May 2022

Final version due: 25 May 2022

Publication: July/August 2022

15:00-15:30 Coffee Break / Katy welcomes all to [Amatria's Virtual Birthday Party](#)



Mapping Science Exhibit

<http://scimaps.org>





101st Annual Meeting of the Association of American Geographers, Denver, CO.
April 5th - 9th, 2005 (First showing of Places & Spaces)



University of Miami, Miami, FL.
September 4 - December 11, 2014.



Duke University, Durham, NC.
January 12 - April 10, 2015

<http://scimaps.org>

The David J. Sencer CDC Museum, Atlanta, GA.
January 25 - June 17, 2016.

Places & Spaces: Mapping Science Exhibit

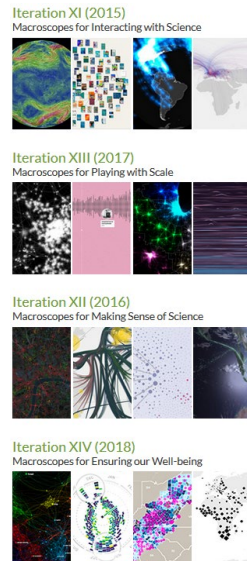
1st Decade (2005-2014)

Maps



2nd Decade (2015-2024)

Macroscopes



100

MAPS
in large format, full color, and
high resolution.

248

MAPMAKERS
from fields as disparate as art,
urban planning, engineering,
and the history of science.



43



MACROSCOPE MAKERS
including one whose job title is
"Truth and Beauty Operator."

20

MACROSCOPES
for touching all kinds of data.

382

DISPLAY VENUES
from the Cannes Film Festival
to the World Economic Forum.

354

PRESS ITEMS
including articles in *Nature*,
Science, *USA Today*, and *Wired*.



<http://scimaps.org>

Map of Scientific Collaborations from 2005-2009

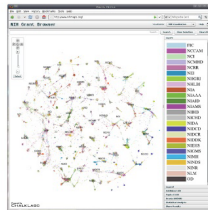


Computed Using Data from Elsevier's Scopus

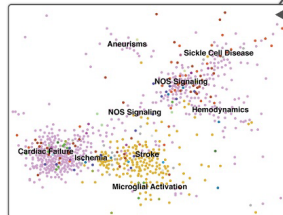
A Topic Map of NIH Grants 2007

Bruce W. Herr II (Chalklabs & IU), Gully Burns (ISI), David Newman (UCI), Edmund Talley (NIH)

The National Institutes of Health (NIH) is organized as a multitude of Institutes and Centers whose missions are primarily focused on distinct diseases. However, disease etiologies and therapies flout scientific boundaries, and thus there is tremendous overlap in the kinds of research funded by each Institute. This creates a daunting landscape for decisions on research directions, funding allocations, and policy formulations. Shown here is devised an interactive topic map for navigating this landscape, online at www.nihmaps.org. Institute abbreviations can be found at www.nih.gov/icd.



Topic modeling, a statistical technique that automatically learns semantic categories, was applied to assess projects in terms used by researchers to describe their work, without the biases of keywords or subject headings. Grant similarities were derived from their topic mixtures, and grants were then clustered on a two-dimensional map using a force-directed simulated annealing algorithm. This analysis creates an interactive environment for assessing grant relevance to research categories and to NIH Institutes in which grants are localized.

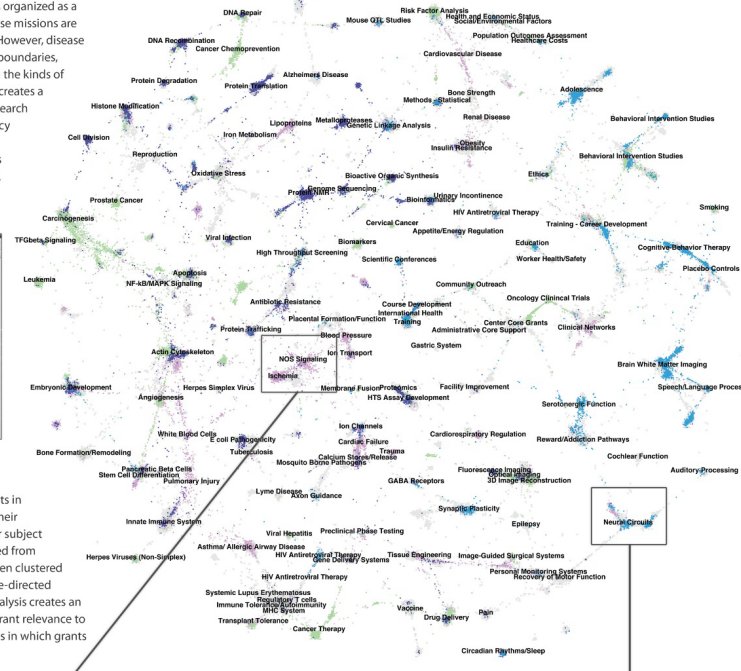
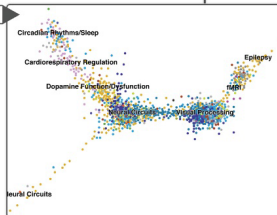


Cardiac Diseases Research

An area of the map focused on cardiovascular function and dysfunction. Cardiac Failure (primarily funded by NHLBI) is typically clustered next to Stroke (NINDS), since these are the two major medical emergencies associated with ischemia, which results from a restricted blood supply. Also localized in this area are grants focused on Nitric Oxide (NOS) Signaling, a major biochemical pathway for vasodilation, and grants on Hemodynamics, Sickle Cell Disease, and Aneurysms.

Neural Circuits Research

An area of the map focused on neural circuits, which shows the diversity of topics and NIH Institutes that fund research in this area, such as: Cardiorespiratory Regulation, primarily funded by NHLBI; Visual Processing, primarily funded by NINDS; Epilepsy, primarily funded by NINDS. For color coding, see legend in the upper-left inset.



National Cancer Institute (NCI)

TOP 10 TOPICS

- 1 Oncology Clinical Trials
- 2 Cancer Treatment
- 3 Cancer Therapy
- 4 Carcinogenesis
- 5 Risk Factor Analysis
- 6 Cancer Chemotherapy
- 7 Metastasis
- 8 Leukemia
- 9 Prediction/Prognosis
- 10 Cancer Chemoprevention

National Institute of General Medical Sciences (NIGMS)

TOP 10 TOPICS

- 1 Bioactive Organic Synthesis
- 2 X-ray Crystallography
- 3 Protein NMR
- 4 Computational Models
- 5 Yeast Biology
- 6 Metalloproteases
- 7 Enzymatic Mechanisms
- 8 Protein Complexes
- 9 Invertebrate/Zebrafish Genetics
- 10 Cell Division

National Heart, Lung, and Blood Institute (NHLBI)

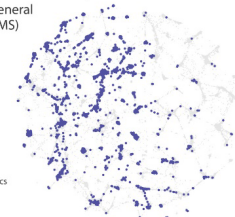
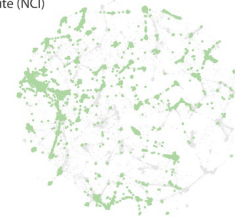
TOP 10 TOPICS

- 1 Cardiac Failure
- 2 Pulmonary Injury
- 3 Genetic Linkage Analysis
- 4 Cardiovascular Disease
- 5 Atherosclerosis
- 6 Hemostasis
- 7 Blood Pressure
- 8 Asthma/ Allergic Airway Disease
- 9 Gene Association
- 10 Lipoproteins

National Institute of Mental Health (NIMH)

TOP 10 TOPICS

- 1 Mood Disorders
- 2 Schizophrenia
- 3 Behavioral Intervention Studies
- 4 Mental Health
- 5 Depression
- 6 Cognitive-Behavior Therapy
- 7 AIDS Prevention
- 8 Genetic Linkage Analysis
- 9 Adolescence
- 10 Childhood



The Structure of Science

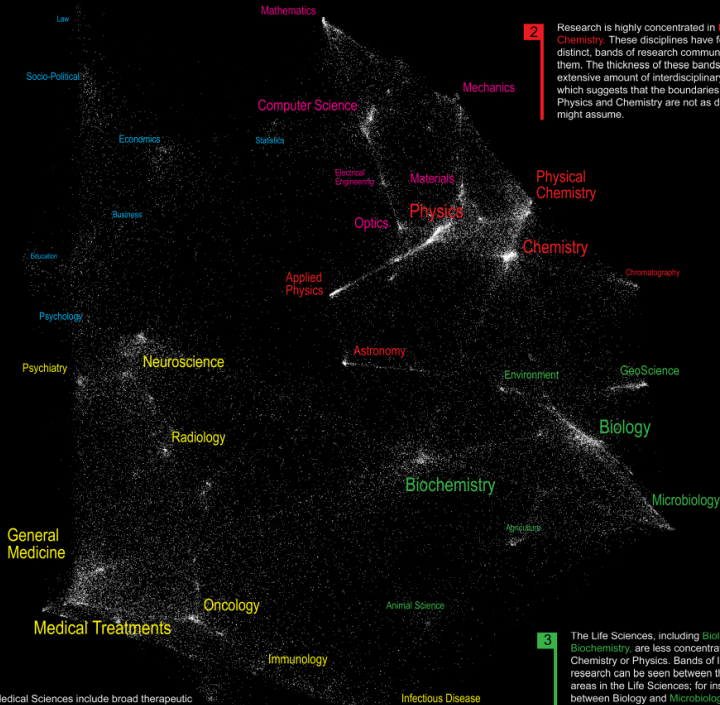
5 The Social Sciences are the smallest and most diffuse of all the sciences. **Psychology** serves as the link between **Medical Sciences (Psychiatry)** and the **Social Sciences, Statistics** serves as the link with **Computer Science** and **Mathematics**.

1 **Mathematics** is our starting point, the purest of all sciences. It lies at the outer edge of the map. **Computer Science, Electrical Engineering,** and **Optics** are applied sciences that draw upon knowledge in **Mathematics** and **Physics**. These three disciplines provide a good example of a linear progression from one pure science (**Mathematics**) to another (**Physics**) through multiple disciplines. Although applied, these disciplines are highly concentrated with distinct bands of research communities that link them. Bands indicate interdisciplinary research.

2 Research is highly concentrated in **Physics** and **Chemistry**. These disciplines have few, but very distinct, bands of research communities that link them. The thickness of these bands indicates an extensive amount of interdisciplinary research, which suggests that the boundaries between **Physics** and **Chemistry** are not as distinct as one might assume.

3 The **Life Sciences**, including **Biology** and **Biochemistry** are less concentrated than **Chemistry** or **Physics**. Bands of linking research can be seen between the larger areas in the **Life Sciences**; for instance between **Biology** and **Microbiology**, and between **Biology** and **Environmental Science**. **Biochemistry** is very interesting in that it is a large discipline that has visible links to disciplines in many areas of the map, including **Biology, Chemistry, Neuroscience,** and **General Medicine**. It is perhaps the most interdisciplinary of the sciences.

4 The **Medical Sciences** include broad therapeutic studies and targeted areas of **Treatment** (e.g. central nervous system, cardiology, gastroenterology, etc.) Unlike **Physics** and **Chemistry**, the medical disciplines are more spread out, suggesting a more multi-disciplinary approach to research. The transition into **Life Sciences** (via **Animal Science** and **Biochemistry**) is gradual.



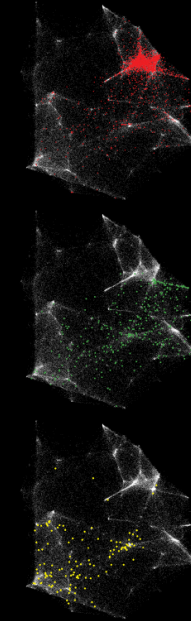
We are all familiar with traditional maps that show the relationships between countries, provinces, states, and cities. Similar relationships exist between the various disciplines and research topics in science. This allows us to map the structure of science.

One of the first maps of science was developed at the Institute for Scientific Information over 30 years ago. It identified 41 areas of science from the citation patterns in 17,000 scientific papers. That early map was intriguing, but it didn't cover enough of science to accurately define its structure.

Things are different today. We have enormous computing power and advanced visualization software that make mapping of the structure of science possible. This galaxy-like map of science (left) was generated at Sandia National Laboratories using an advanced graph layout routine (VxOrd) from the citation patterns in 800,000 scientific papers published in 2002. Each dot in the galaxy represents one of the 96,000 research communities active in science in 2002. A research community is a group of papers (9 on average) that are written on the same research topic in a given year. Over time, communities can be born, continue, split, merge, or die.

The map of science can be used as a tool for science strategy. This is the terrain in which organizations and institutions locate their scientific capabilities. Additional information about the scientific and economic impact of each research community allows policy makers to decide which areas to explore, exploit, abandon, or ignore.

We also envision the map as an educational tool. For children, the theoretical relationship between areas of science can be replaced with a concrete map showing how math, physics, chemistry, biology and social studies interact. For advanced students, areas of interest can be located and neighboring areas can be explored.



Nanotechnology

Most research communities in nanotechnology are concentrated in **Physics, Chemistry, and Materials Science**. However, many disciplines in the **Life and Medical Sciences** also have nanotechnology applications.

Proteomics

Research communities in proteomics are centered in **Biochemistry**. In addition, there is a heavy focus in the tools section of chemistry, such as **Chromatography**. The balance of the proteomics communities are widely dispersed among the **Life and Medical Sciences**.

Pharmacogenomics

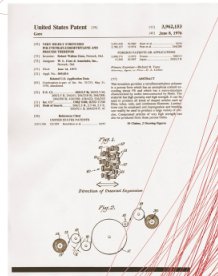
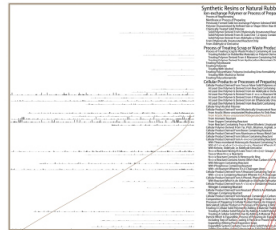
Pharmacogenomics is a relatively new field with most of its activity in **Medicine**. It also has many communities in **Biochemistry** and two communities in the **Social Sciences**.

Impact

The United States Patent and Trademark Office does scientists and industry a great service by granting patents to protect inventions. Inventions are categorized in a taxonomy that groups patents by industry or use, proximate function, effect or product, and structure. At the time of this writing there are 160,252 categories in a hierarchy that goes 15 levels deep. We display the first three levels (13,529 categories) at right in what might be considered a textual map of inventions.

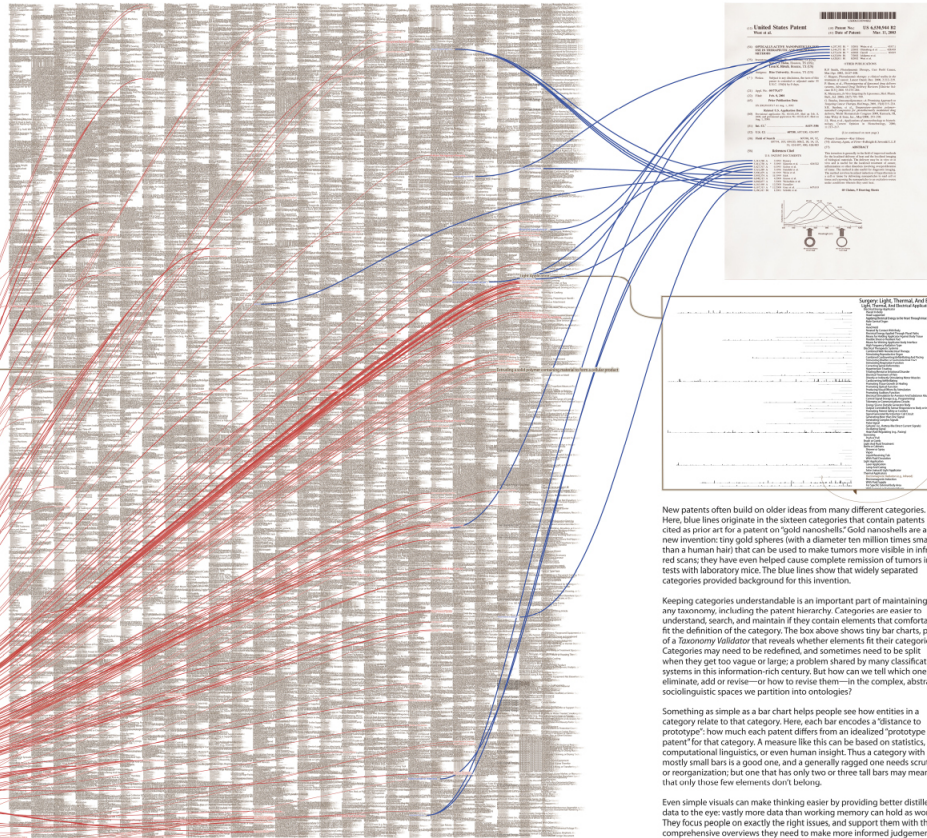
Patent applications are required to be unique and non-obvious, partially by revealing any previous patents that might be similar in nature or provide a foundation for the current invention. In this way we can trace the impact of a single patent, seeing how many patents and categories it affects.

The patent on Goretex—a lightweight, durable synthetic fiber—is an example of one that has had significant impact. The box below enlarges the section of the hierarchy where it is filed, and the red lines (arranged to start along a time line from 1961 to 2006) point to the 130 categories that contain 182 patents, from waterproof clothing to surgical cosmetic implants, that mention Goretex as "prior art".



The US Patent Hierarchy

Prior Art



New patents often build on older ideas from many different categories. Here, blue lines originate in the sixteen categories that contain patents cited as prior art for a patent on "gold nanoshells." Gold nanoshells are a new invention: tiny gold spheres (with a diameter ten million times smaller than a human hair) that can be used to make tumors more visible in infrared scans; they have even helped cause complete remission of tumors in tests with laboratory mice. The blue lines show that widely separated categories provided background for this invention.

Keeping categories understandable is an important part of maintaining any taxonomy, including the patent hierarchy. Categories are easier to understand, search, and maintain if they contain elements that comfortably fit the definition of the category. The box above shows tiny bar charts, part of a *Taxonomy Validator* that reveals whether elements fit their categories. Categories may need to be redefined, and sometimes need to be split when they get too vague or large; a problem shared by many classification systems in this information-rich century. But how can we tell which ones to eliminate, add or revise—or how to revise them—in the complex, abstract sociolinguistic spaces we partition into ontologies?

Something as simple as a bar chart helps people see how entities in a category relate to that category. Here, each bar encodes a "distance to prototype": how much each patent differs from an idealized "prototype patent" for that category. A measure like this can be based on statistics, computational linguistics, or even human insight. Thus a category with mostly small bars is a good one, and a generally ragged one needs scrutiny or reorganization; but one that has only two or three tall bars may mean that only those few elements don't belong.

Even simple visuals can make thinking easier by providing better distilled data to the eye: vastly more data than working memory can hold as words. They focus people on exactly the right issues, and support them with the comprehensive overviews they need to make more informed judgements.

Science related Wikipedian ACTIVITY

This visualization explores the activity of science, math, and technology (SMT) related articles in the English-language Wikipedia (<http://en.wikipedia.org>). The central image shows 659,388 articles (circles). Overlaid is a 37 x 37 grid of relevant half-inch sized images.

Blue, green, and yellow circles represent the 3,599 math, 6,874 science, and 3,164 technology related articles respectively. The larger the size of a circle the higher the likelihood it is that type of article. The four corners show activity patterns of SMT articles.

Article Edit Activity

Articles are size coded based on how frequently they have been edited from Feb. 6, 2001 to April 6, 2007. More consideration is given to current and major edits. Larger circles have been edited more frequently than smaller circles.

2007 Major Edits

Articles are size coded based on how many major edits they received from January 1st, 2007 to April 6th, 2007. Larger circles have received more edits than smaller circles. The highest number of major edits was 2,627.

For the central image, each article is size coded based on the likelihood that it is math, science, or technology related.



All five images are color coded based on type. Transparency is used for legibility, and creates different colors when nodes overlap.



Article Popularity

Articles are size coded based on the number of Wikipedia articles referencing it. Larger circles are receiving more links from other articles than smaller circles. The highest number of references to an article was 142,602.

Number of Bursts

Articles are size coded based on the number of bursts, i.e. sudden increases, of edit activity that occurred during the article's lifetime. Larger circles have had more bursts in activity than smaller circles. The most bursts an article had was 9.



Diseasome

The Human Disease Network

Explore online at <http://diseasome.eu>

Statistics

of Nodes: 516
of Edges: 1188
Density: 0.0009
Average Degree: 9.20
Diameter: 15
Average Shortest Path: 6,5

Top 5 Diseases

1. Deafness
2. Eukemia
3. Colon Cancer
4. Retinitis Pigmentosa
5. Diabetes Mellitus

Top 5 Genes

1. TP53
2. PAK5
3. FGFR2
4. RTN1
5. MSH2

Description

This map presents a network of 516 diseases linked by 1188 known disorder-gene associations, indicating the common genetic origin of many diseases.

HOW WE GOT LABELS

The map offers a visual reference of the genetic links between disorders and a valuable global perspective for physicians, genetic counselors, and biomedical researchers alike. The size represents the total number of genetic disorders reported by their affected genes, whereas the color indicates the cause of disease, and the number of partner genes.

NETWORK VISUALIZATION TECHNICAL SPECS

The map was drawn using the force-directed layout algorithm ForceAtlas2 in Gephi. Node sizes correspond to the disorder class to which the disorder belongs, and the size is proportional to the number of genes that are associated to both disorders and connected with the shared gene between those two target genes. Colored circles are not shown and only the gene component has been kept. The circles also indicate most remarkable disorder classes and their respective major clusters.

The Disorder Class Interaction graph below shows the interaction level between disorder classes, representing the number of shared genes, up to 80.

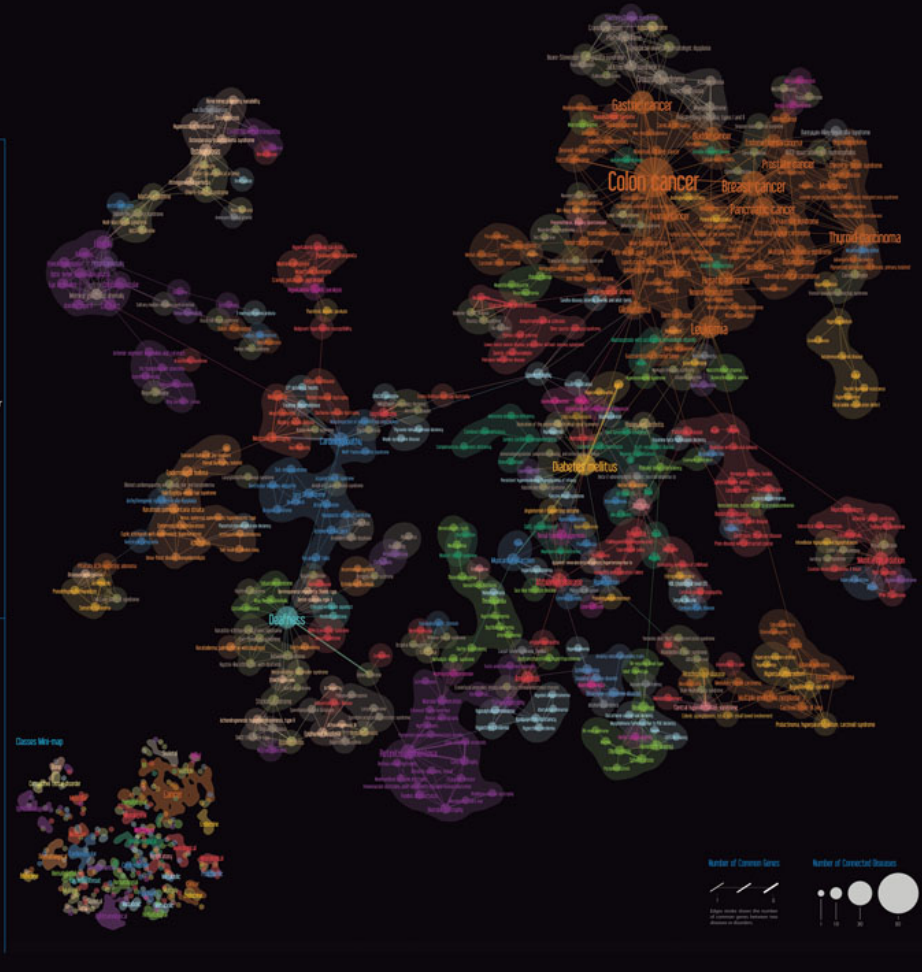
Disorder Class Network
See G. Bastien, M. Bastien, S. Heymann, B. Hainaut, M. Beaulieu, G. D. D'Amico
PLoS Biol 14(12): e1002402 (2016)

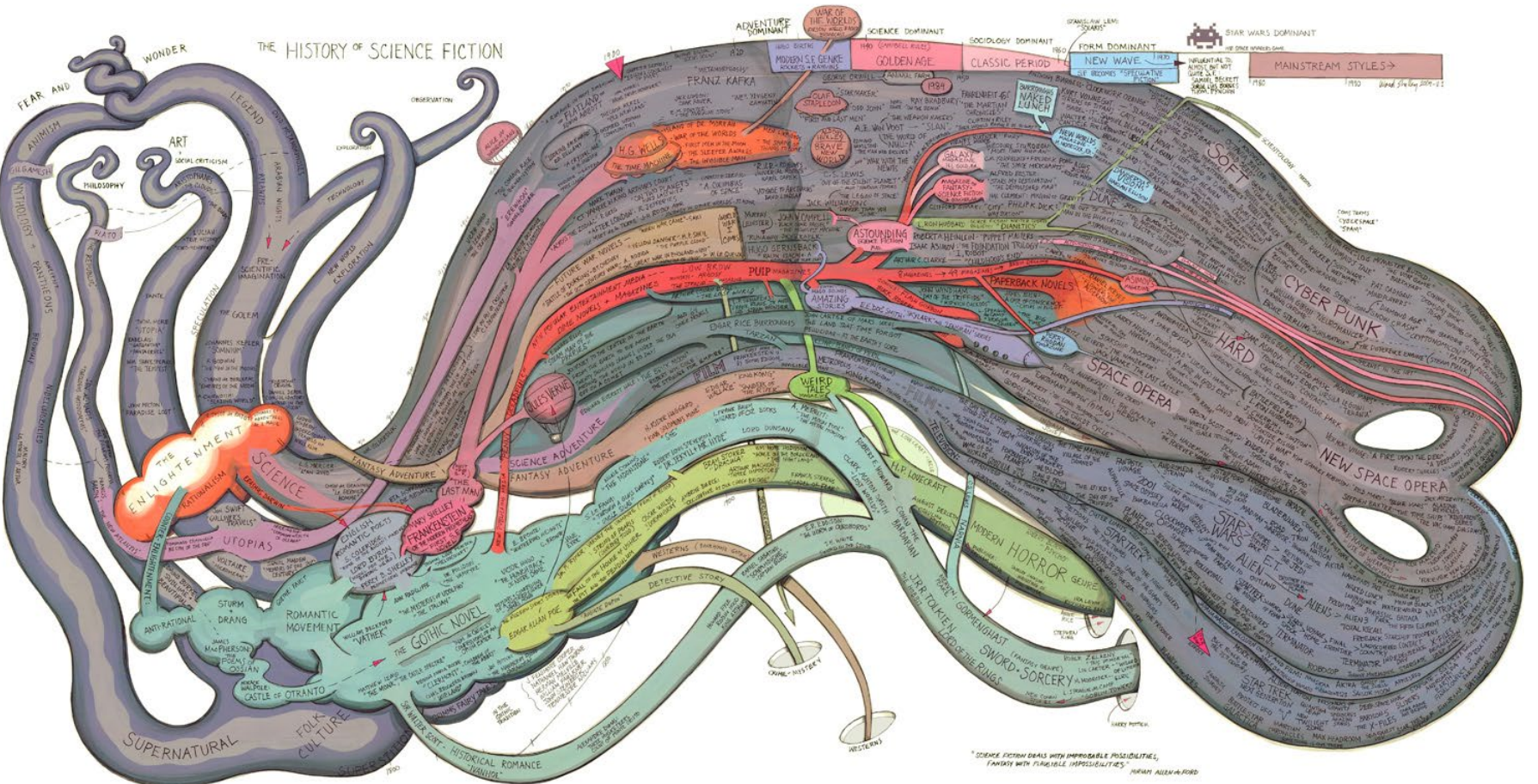
Disorder Class Interactions



Disorder Class

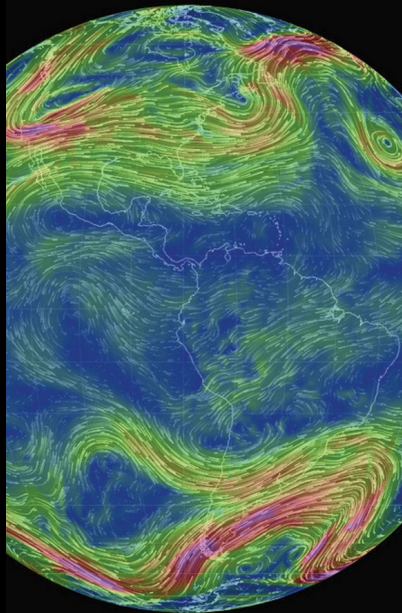
- Cancer
- Endocrine
- Ear, Nose, Throat
- Ophthalmological
- Neurological
- Hematological
- Cardiovascular
- Muscular
- Immunological
- Dermatological
- Nutritional
- Connective Tissue Disorder
- Renal
- Psychiatric
- Metabolic
- Bone
- Skeletal
- Developmental
- Gastrointestinal
- Respiratory
- Multiple
- Unclassified







MACROSCOPES FOR INTERACTING WITH SCIENCE



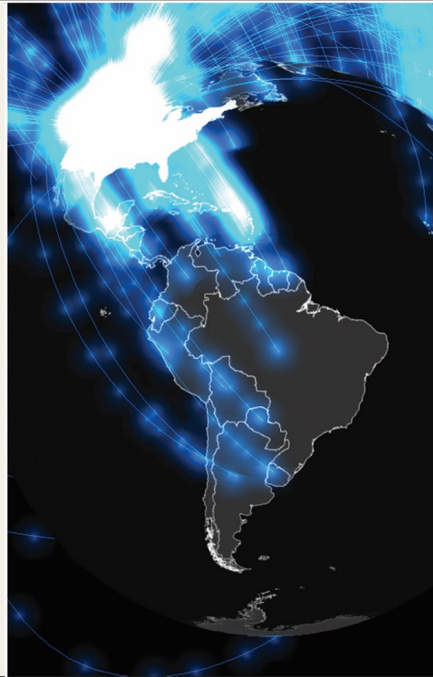
Earth

Weather on a worldwide scale



AcademyScope

Exploring the scientific landscape



Mapping Global Society

Local news from a global perspective

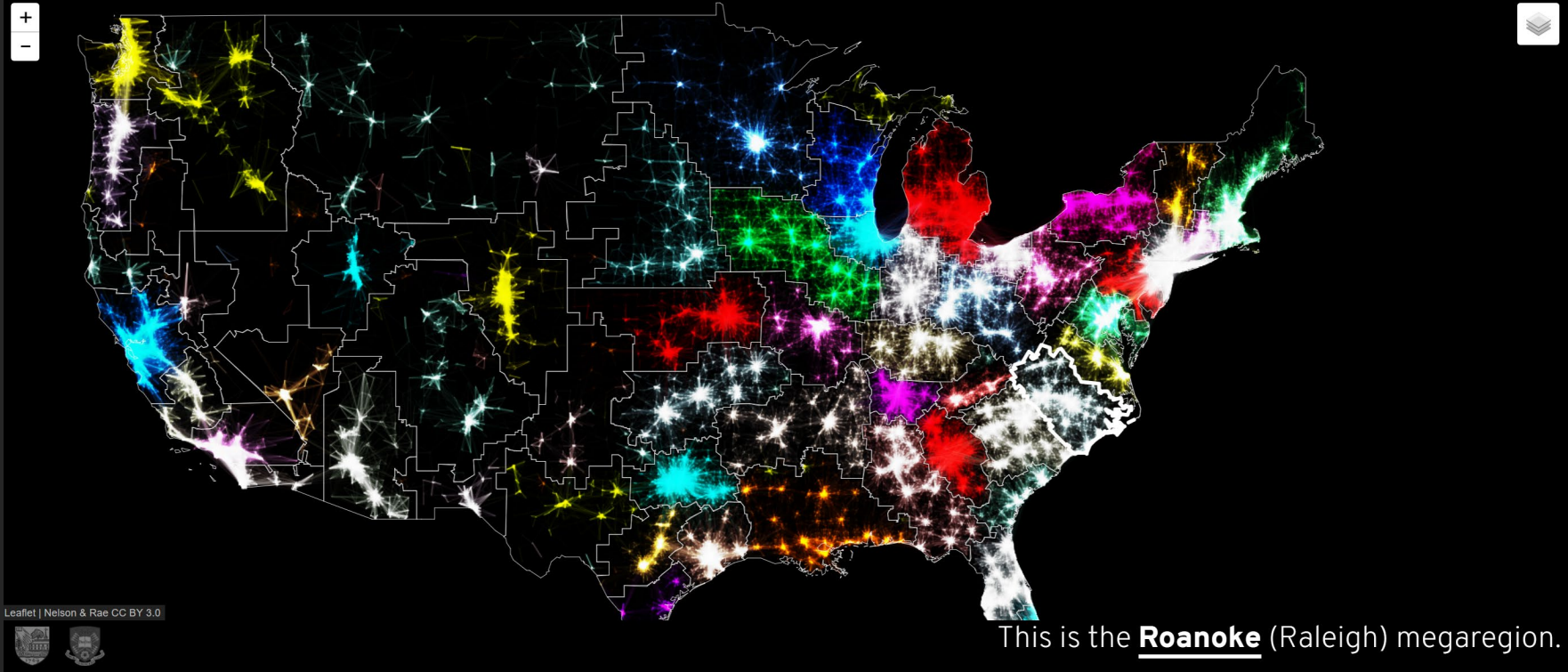


Charting Culture

2,600 years of human history in 5 minutes

THE MEGAREGIONS OF THE US

Explore the new geography of commuter connections in the US.
Tap to identify regions. Tap and hold to see a single location's commuteshed.



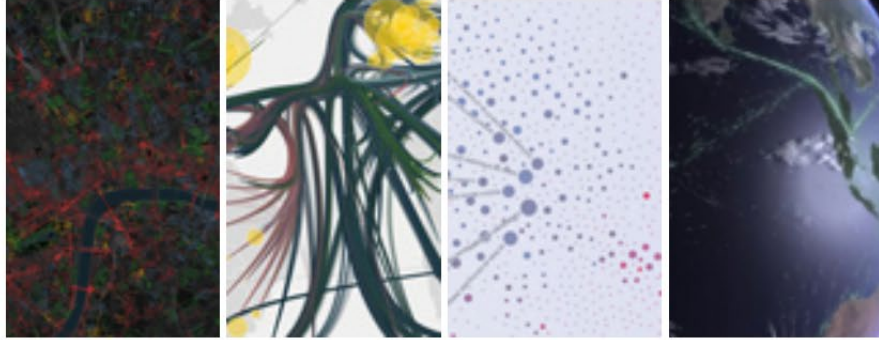
SMELLY MAPS



Smelly Maps – Daniele Quercia, Rossano Schifanella, and Luca Maria Aiello – 2015

Iteration XII (2016)

Macrosopes for Making Sense of Science



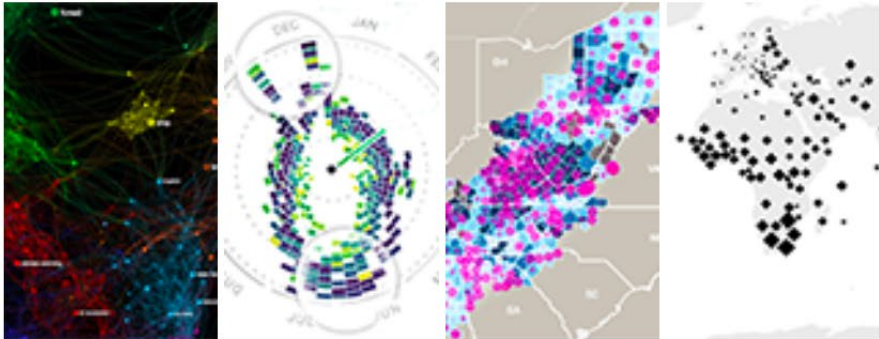
Iteration XIII (2017)

Macrosopes for Playing with Scale



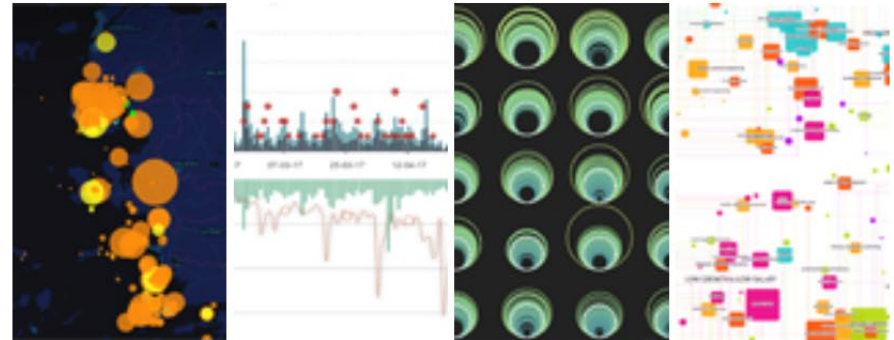
Iteration XIV (2018)

Macrosopes for Ensuring our Well-being

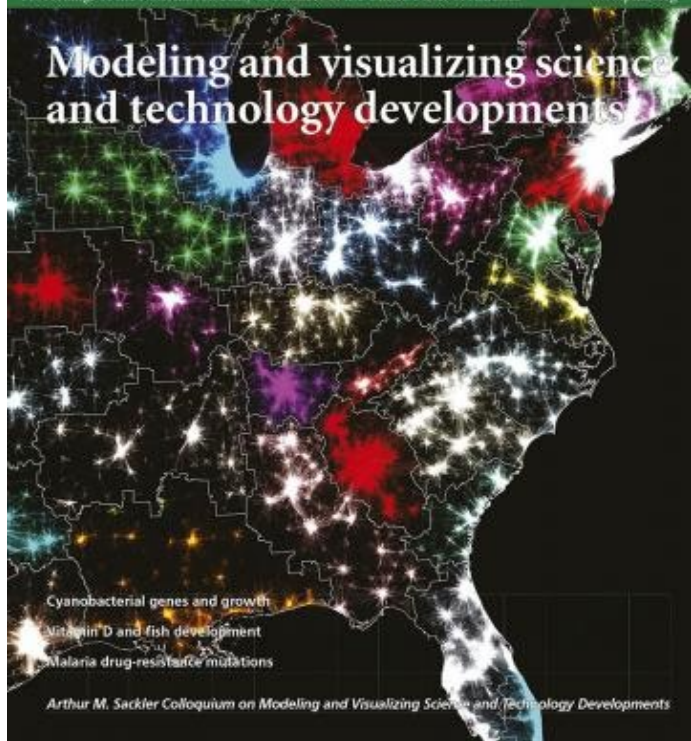


Iteration XV (2019)

Macrosopes for Tracking the Flow of Resources



Modeling and visualizing science and technology developments



<https://www.pnas.org/modeling>

Atlas of Forecasts

Modeling and Mapping Desirable Futures

Katy Börner



<https://mitpress.mit.edu/books/atlas-forecasts>

Acknowledgments

Exhibit Curators



The exhibit team: Lisel Record, Katy Börner, and Todd Theriault.

<http://scimaps.org>

Plus, we thank the more than 250 authors of the 100 maps and 16 interactive macroscopes.

Exhibit Advisory Board



Gary Berg-Cross
Cognitive psychologist (PhD, SUNY-Stony Brook). Potomac, MD, USA



André Skupin
Associate Professor of Geography at San Diego State University, California



Donna J. Cox, MFA, Ph.D.
Director of the **Advanced Visualization Laboratory** at the National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, IL, USA



Moritz Stefaner
Freelance designer on the crossroads of data visualization, information aesthetics, and user interface design in Germany



Bonnie DeVarco
Media X Distinguished Visiting Scholar at Stanford University, Palo Alto, CA, USA



Olga Subirós
Curator of Big Bang Data and Founder of **Olga Subirós Studio** in Barcelona, Spain



Peter A. Hook
Head of Digital and Scholarly Services and LawArXiv Administrator, Cornell Law Library. Ithaca, NY, USA



Stephen Uzzo
Vice President of Science and Technology for the **New York Hall of Science**



Francis Harvey
Professor of Visual Communication in Geography at the Leibnitz Institute for Regional Geography, Leipzig University, Germany



Benjamin Wiederkehr
Founding Partner and Managing Director of **Interactive Things** in Zürich, Switzerland



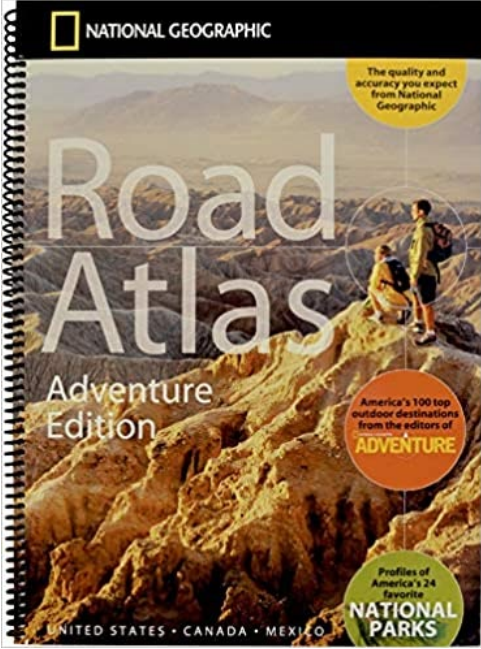
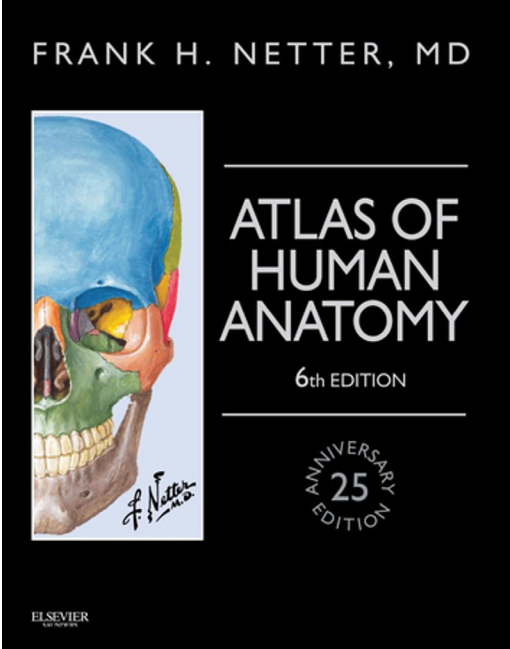
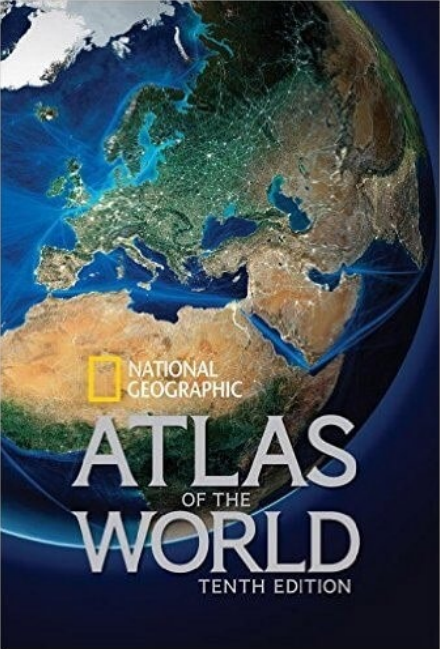
Lev Manovich
Professor, **The Graduate Center**, City University of New York; Director, **Software Studies Initiative** (big data, digital humanities, visualization)



HuBMAP Reference Atlas: Toward a human [reference] map



An **atlas** is an oversized, bound book of maps.
It has descriptive text, an index, possibly other data visualizations.



An human cell **atlas**
might show a landscape
of all cells, or

Maps of cells per tissue
type/anatomical structure.

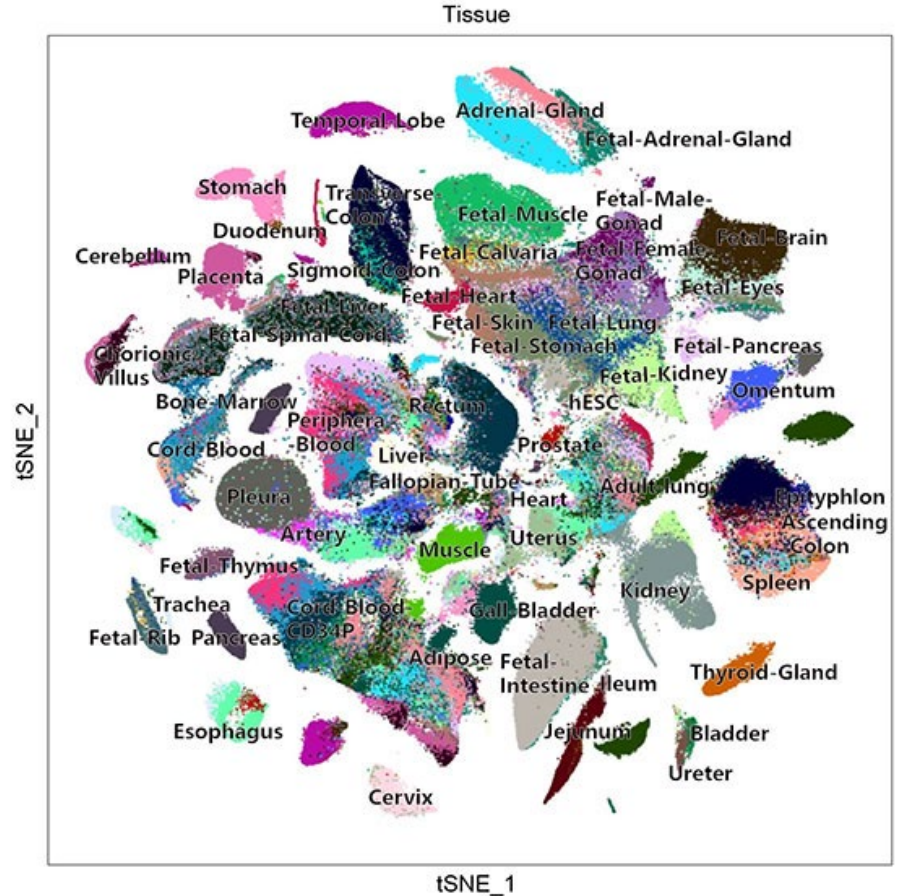
Article | Published: 25 March 2020

Construction of a human cell landscape at single-cell level

Xiaoping Han , Ziming Zhou, [...] Guoji Guo 

Nature **581**, 303–309(2020) | [Cite this article](#)

55k Accesses | **32** Citations | **409** Altmetric | [Metrics](#)



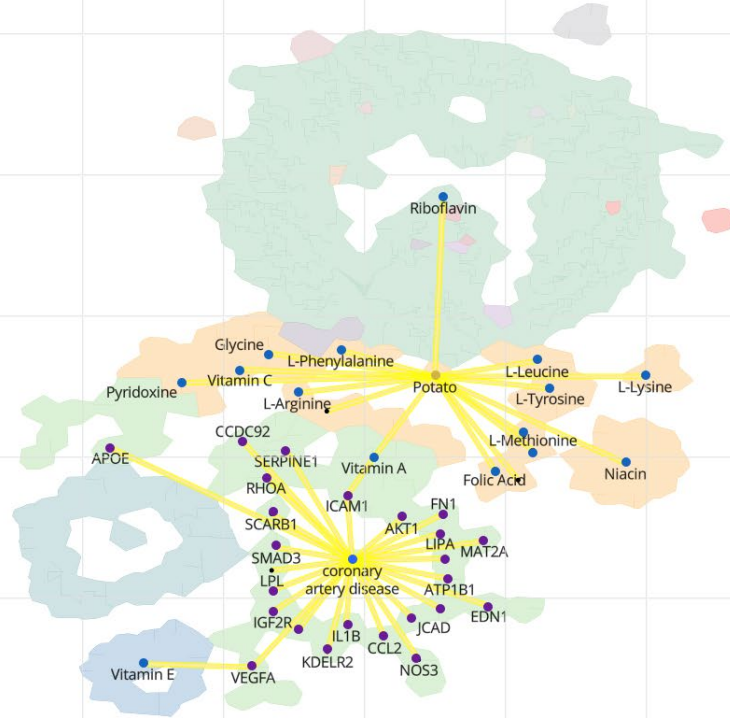
potato ×

coronary artery disease ×

SEARCH

Favorite Similar Send to phone Share

Show Overview



● Disease ● Compound ● Food ● Gene

Send feedback



H

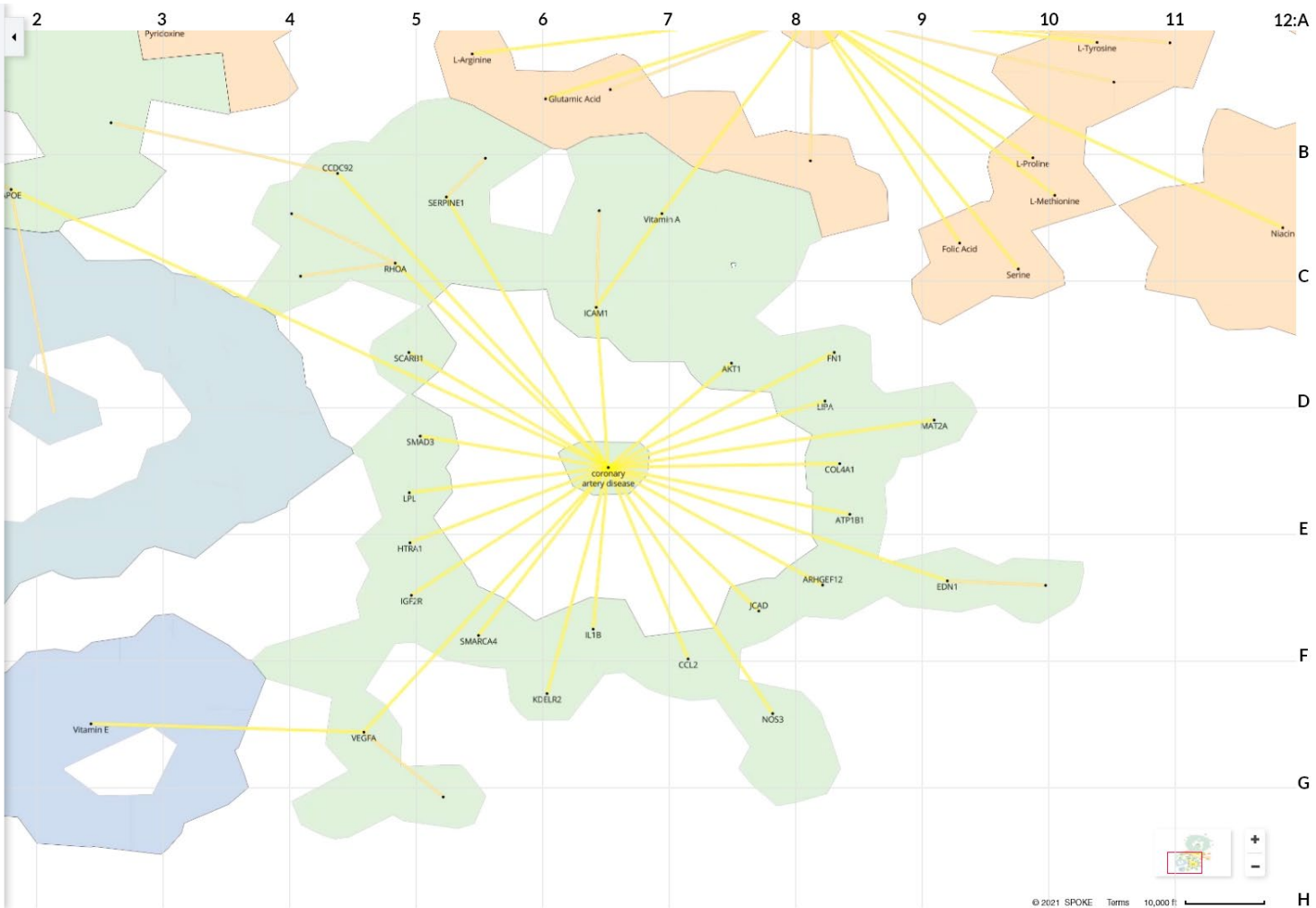
potato ×

coronary artery disease ×

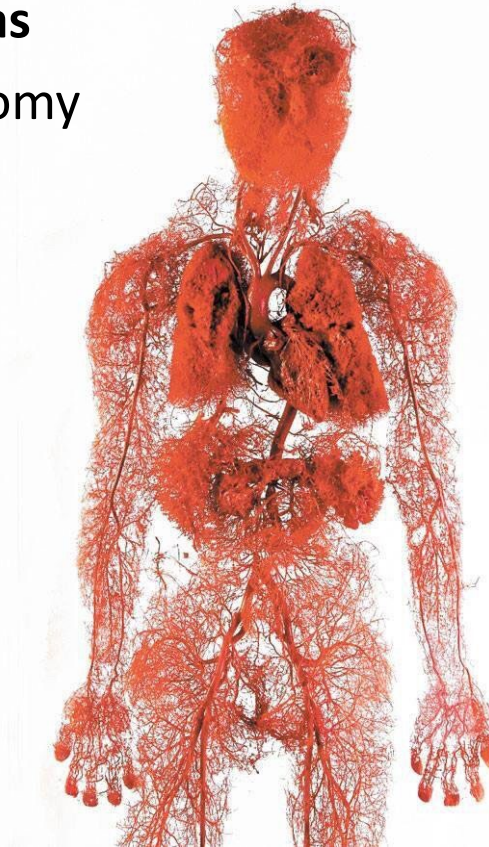
SEARCH

Favorite Similar Send to phone Share

Show Overview

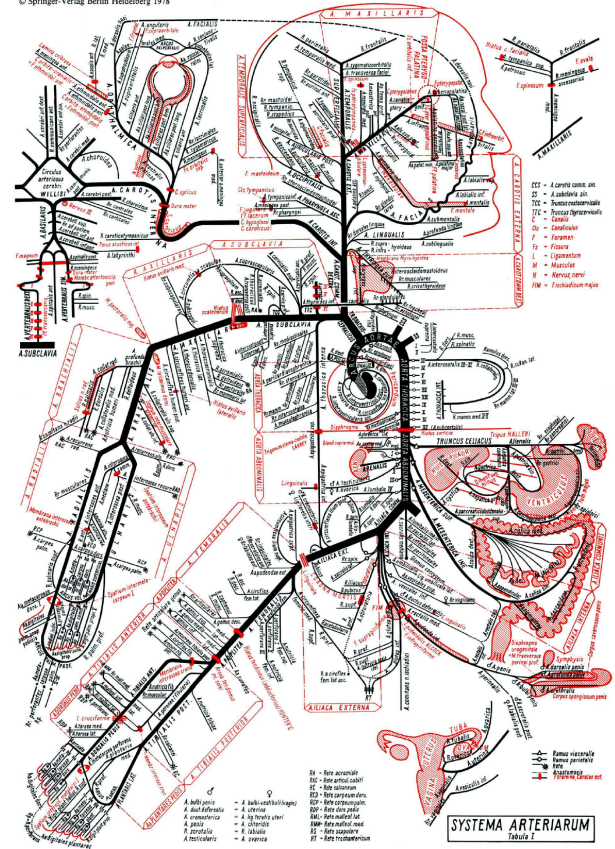


A human reference **atlas**
might use human anatomy
as a 'basemap,' or
an abstract space.



<https://bodyworlds.com>

E. M. W. Weber:
Schemata der Leitungsbahnen des Menschen
© Springer-Verlag Berlin Heidelberg 1978



Weber, 1978

The Human Body at Cellular Resolution: The NIH Human Biomolecular Atlas Program.
 Snyder et al. *Nature*. 574, p. 187-192.

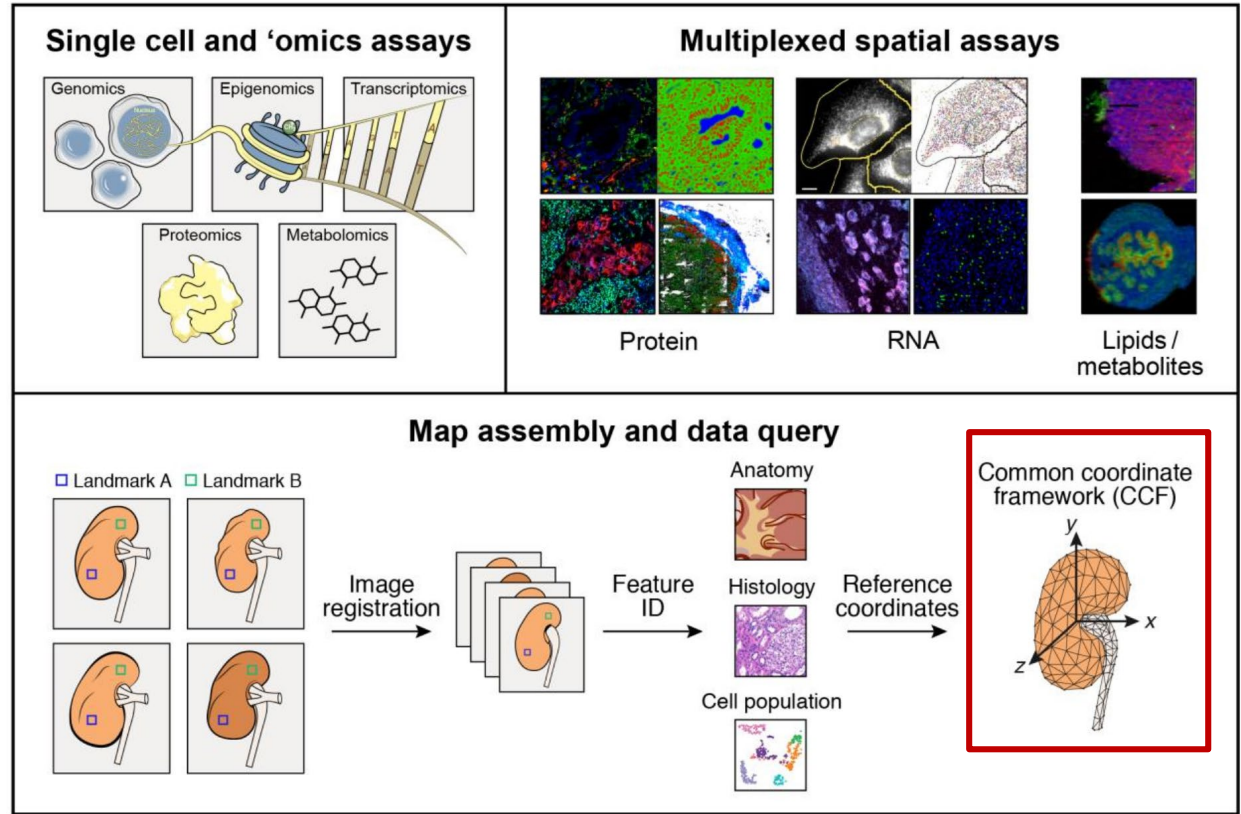


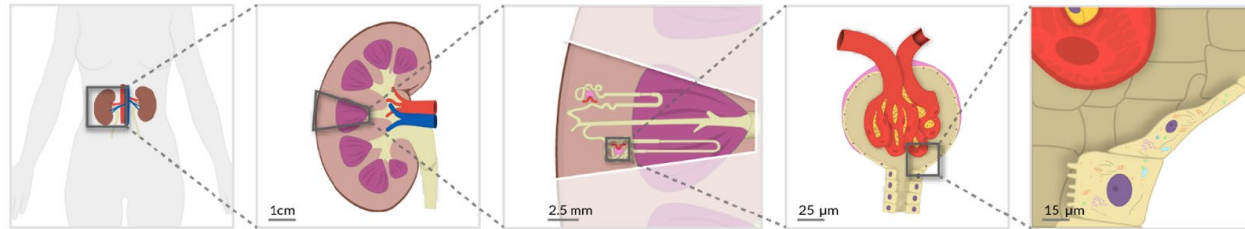
Fig. 3 | Map generation and assembly across cellular and spatial scales. HuBMAP aims to produce an atlas in which users can refer to a histological slide from a specific part of an organ and, in any given cell, understand its contents on multiple 'omic levels—genomic, epigenomic, transcriptomic, proteomic, and/or metabolomic. To achieve these ends, centres will apply a combination of imaging, 'omics and mass spectrometry

techniques to specimens collected in a reproducible manner from specific sites in the body. These data will be then be integrated to arrive at a high-resolution, high-content three-dimensional map for any given tissue. To ensure inter-individual differences will not be confounded with collection heterogeneity, a robust CCF will be developed.

CCF Requirements

The CCF must capture major **anatomical structures, cell types, and biomarkers** and their interrelations across **multiple levels of resolution**.

It should be **semantically explicit** (using existing ontologies, e.g., Uberon, CL) and **spatially explicit** (e.g., using 3D reference organs for registration and exploration).



Body

- Body
- Kidney (Left, Right)
- Aorta
- Renal artery
- Renal vein
- Ureter

Organ

- Renal capsule
- Renal pyramid
- Renal cortex
- Renal medulla
- Renal calyx
- Renal pelvis

Functional Tissue Unit

- Nephron
- Renal corpuscle
- Proximal convoluted tubule
- Loop of Henle
- Distal convoluted tubule
- Connecting tubule
- Collecting duct

FTU Sub-structure(s)

- Bowman's capsule
- Glomerulus
- Efferent arteriole
- Afferent arteriole

Cellular

- Parietal epithelial cell
- Capillary endothelial cell
- Mesangial cell
- Podocyte

ASCT+B Tables

Anatomical Structures, Cell Types, and Biomarkers (ASCT+B) tables aim to capture the partonomy of anatomical structures, cell types, and major biomarkers (e.g., gene, protein, lipid or metabolic markers).

Structure/Region	Substructure/Sub region	Cell Type	Subset of Marker Genes
Renal Corpuscle	Bowman's Capsule	Parietal epithelial cell	<i>CRB2*</i> , <i>CLDN1*</i>
	Glomerulus	Podocyte	<i>NPHS2*</i> , <i>PODXL*</i> , <i>NPHS1*</i>
		Capillary Endothelial Cell	<i>EHD3*</i> , <i>EMCN*</i> , <i>HECW2*</i> , <i>FLT1*</i> , <i>AQP1*</i>
		Mesangial Cell	<i>POSTN*</i> , <i>PIEZO2*</i> , <i>ROBO1*</i> , <i>ITGA8*</i>

Partial ASCT+B Table from

- El-Achkar et al. A Multimodal and Integrated Approach to Interrogate Human Kidney Biopsies with Rigor and Reproducibility: The Kidney Precision Medicine Project. bioRxiv. 2019, Updated Aug 2020. doi:10.1101/828665

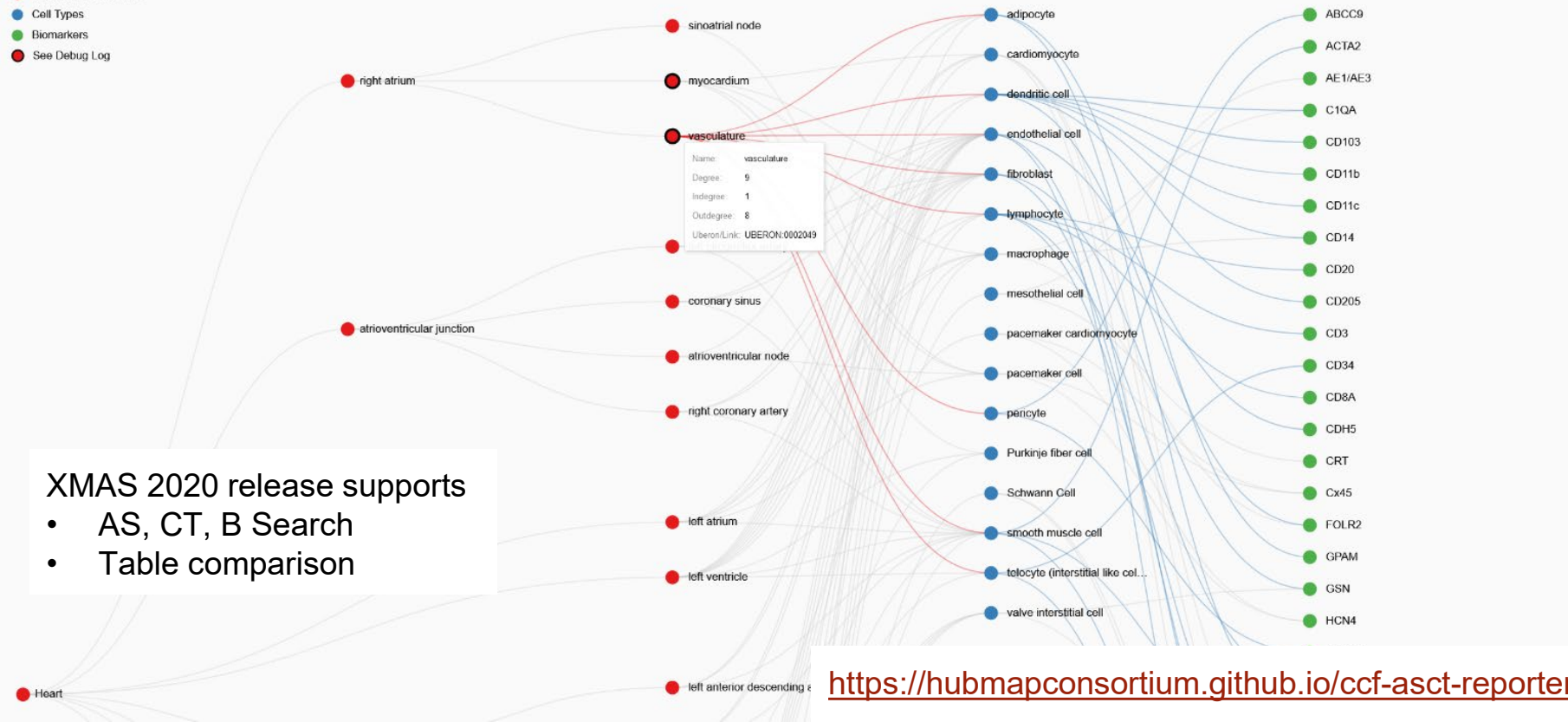
Anatomical Structures

Cell Types

Biomarkers

Legend

- Anatomical Structures
- Cell Types
- Biomarkers
- See Debug Log



XMAS 2020 release supports

- AS, CT, B Search
- Table comparison

<https://hubmapconsortium.github.io/ccf-asct-reporter>

Anatomical Structures

Cell Types

Biomarkers

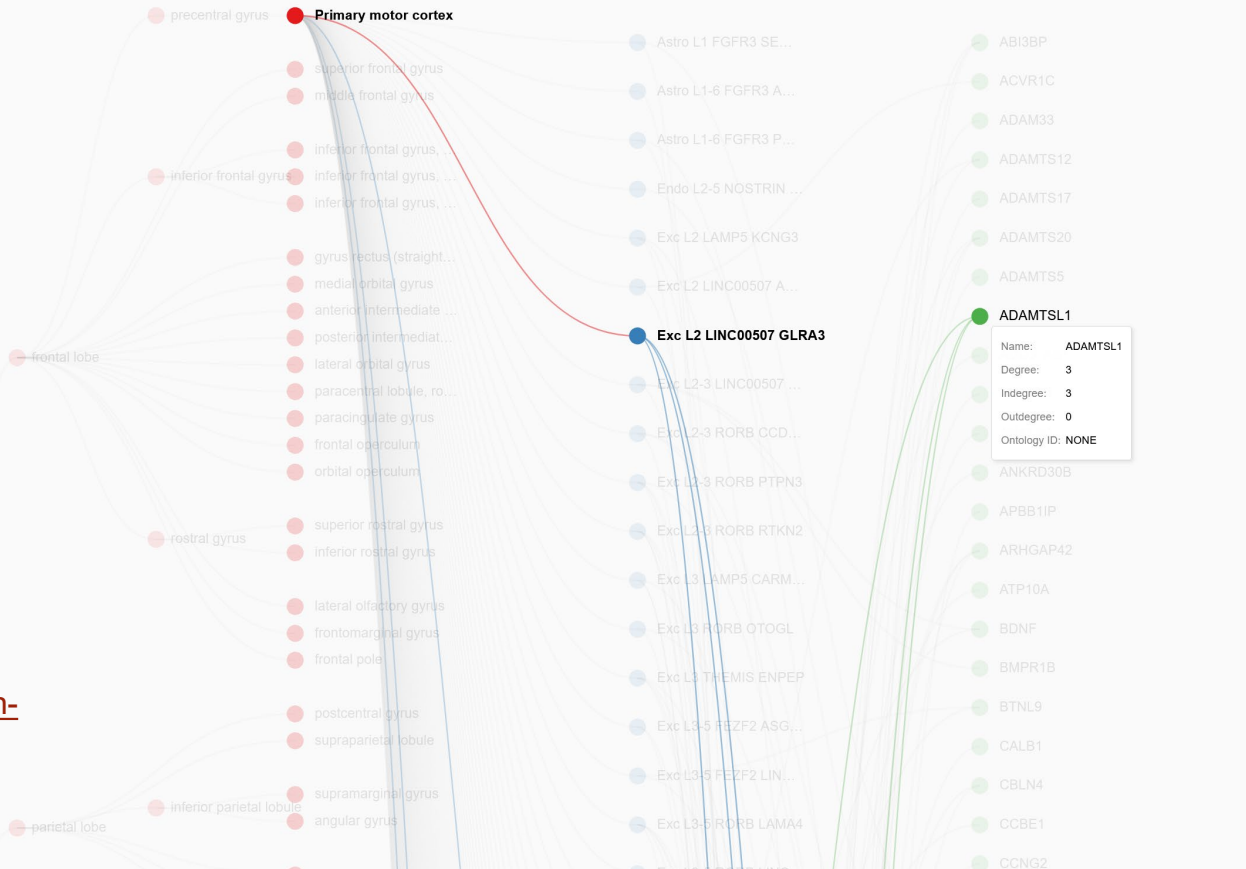
Legend

- Anatomical Structures
- Cell Types
- Biomarkers
- Multi-parent Nodes



<https://community.brain-map.org/t/allen-human-reference-atlas-3d-2020-new/>

Visualization Controls

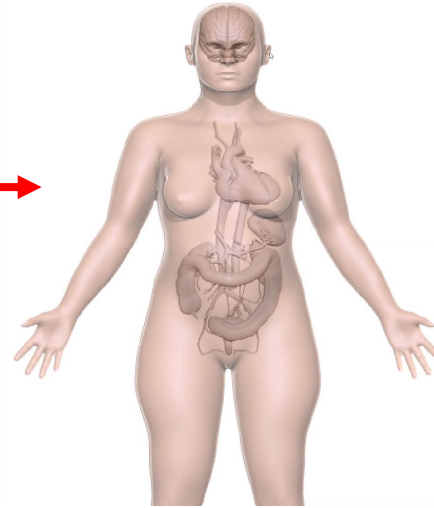


ASCT+B Tables vs. 3D Reference Organs

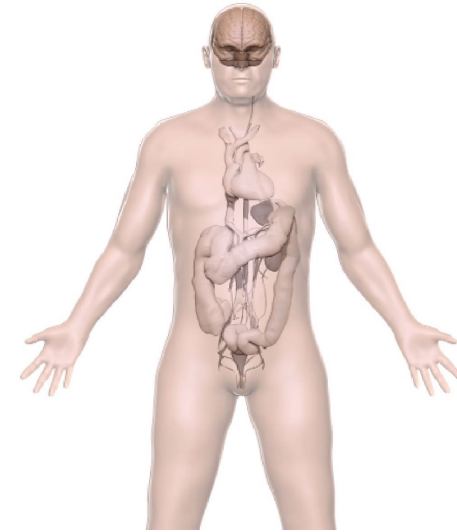
ASCT+B for 10 organs on 11/30/2020, 5:37pm ET:

Organ	#AS	#CT	#B	#AS-CT	#CT-B
Brain	184	127	254	127	346
Heart	23	16	35	73	42
Kidney	39	53	83	63	131
Large Intestine	22	33	45	306	72
Liver	16	27	34	29	35
Lung	18	62	103	110	128
Lymph nodes	34	30	50	63	110
Skin	14	32	57	37	99
Small Intestine	20	32	48	195	55
Spleen	33	26	46	48	72

Female



Male



Vasculature 751

<https://hubmapconsortium.github.io/ccf/pages/ccf-anatomical-structures.html>

<https://hubmapconsortium.github.io/ccf/pages/ccf-3d-reference-library.html> (NLM VH organs)

<https://community.brain-map.org/t/allen-human-reference-atlas-3d-2020-new/> (brain)

<https://www3.cs.stonybrook.edu/~ari/> (male colon)

CCF Exploration User Interface (EUI)

HuBMAP Sex: Both Age: 1-110 BMI: 13-83 Login

Search ontology terms ...

- body
 - heart
 - lung
 - kidney
 - right kidney
 - left kidney
 - kidney capsule
 - cortex of kidney
 - renal medulla
 - renal column
 - renal pyramid
 - hilum of kidney
 - kidney interstitium
 - kidney calyx
 - renal pelvis
 - ureter
 - renal papilla
 - renal fat pad
 - nephron

body

- 2 Centers
- 27 Donors
- 41 Samples

10x Female, Age 14, BMI 14.7
HBM894.MPVN.828
TMC-Florida
First case collected. Incomplete d...

CODEX Male, Age 18, BMI 27.1
HBM436.GHWX.449
TMC-Florida
section is 190um from block surface

Male, Age 56, BMI 32.5
HBM696.XTVL.498
TMC-Vanderbilt
Age 56, White Male

Male, Age 53, BMI 26.5
HBM652.VRLD.292
TMC-Vanderbilt
Age 53, Black Male

Male, Age 58, BMI 22.0
HBM477.CJKM.888
TMC-Vanderbilt
107-111

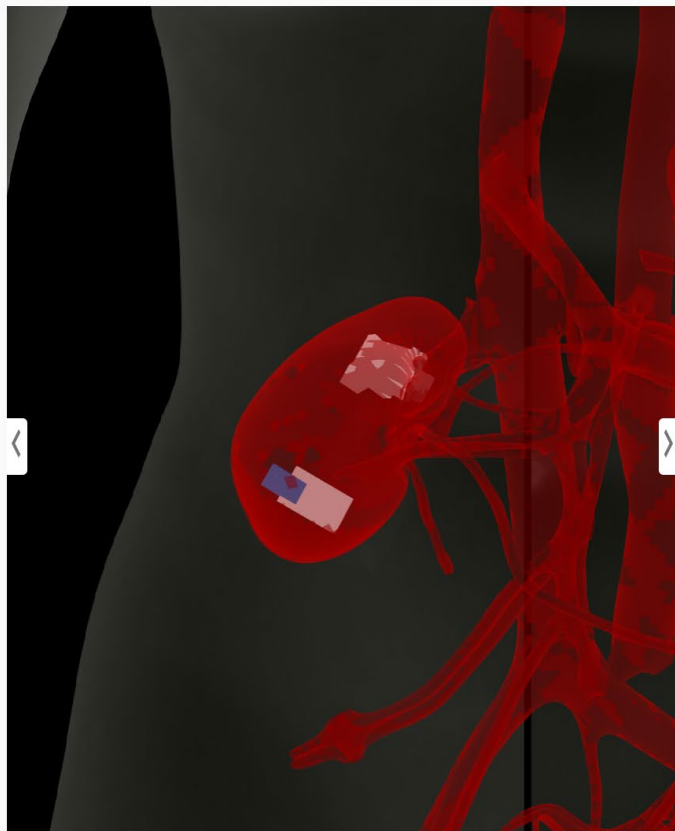
CODEX Male, Age 18, BMI 25.5
HBM473.VKCM.878
TMC-Florida
section is 255um from block surface

LC Male, Age 55, BMI 25.4
HBM824.BLXF.883
TMC-Vanderbilt
13-16

Search ontology terms ...



- body
 - heart
 - lung
 - kidney
 - right kidney
 - left kidney
 - kidney capsule
 - cortex of kidney
 - renal medulla
 - renal column
 - renal pyramid
 - hilum of kidney
 - kidney interstitium
 - kidney calyx
 - major calyx
 - minor calyx
 - renal pelvis
 - ureter
 - renal papilla
 - renal fat pad
 - nephron
 - spleen
 - colon



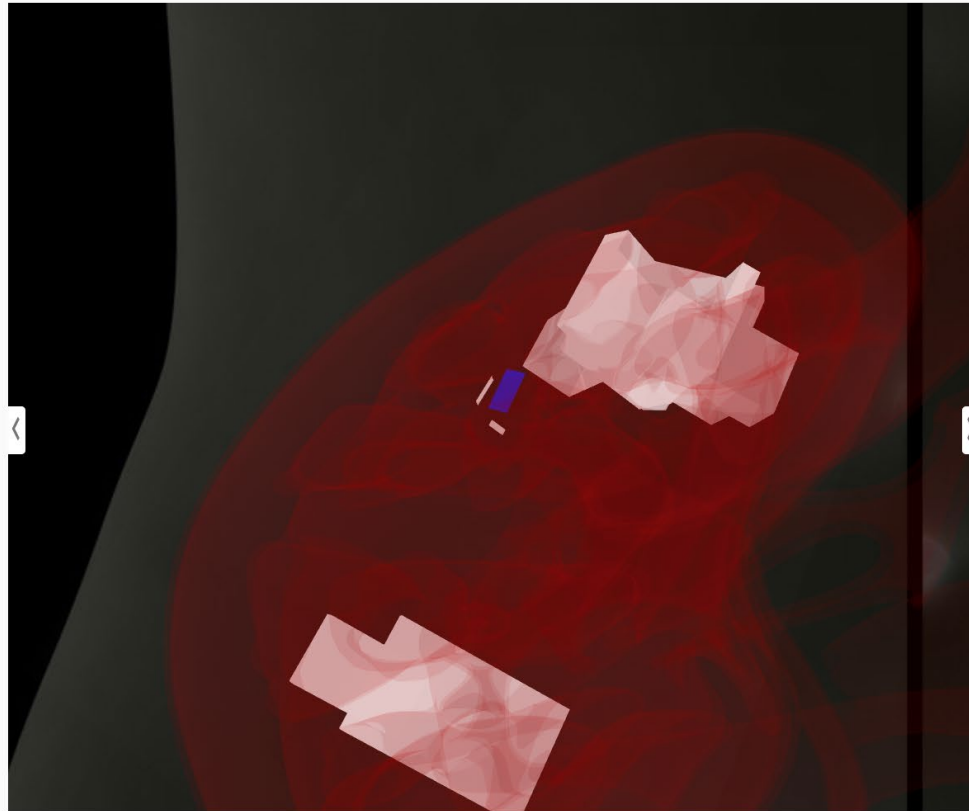
body

- 1 Centers
- 9 Donors
- 40 Samples

	Male, Age 55, BMI 25.4 HBM695 RTLJ.484 TMC-Vanderbilt 13-16	
	Male, Age 21, BMI 21.8 HBM634 MIMGK.572 TMC-Vanderbilt Age 21 , White Male, Trauma Patient	
	Female, Age 44, BMI 28.0 HBM457 NNQN.252 TMC-Vanderbilt Age 44, white female	
	Female, Age 44, BMI 28.0 HBM465 VKHL.532 TMC-Vanderbilt Age 44, white female	
	Male, Age 21, BMI 21.8 HBM693 HFFJ.752 TMC-Vanderbilt Age 21 , White Male, Trauma Patient	
	Female, Age 58, BMI 23.0 HBM536 LDTZ.757 TMC-Vanderbilt Age 58, White Female	
	Male, Age 48, BMI 35.3 HBM334 GCCX.874 TMC-Vanderbilt Age 48, White Male	
	Male, Age 31, BMI 32.6 HBM776 PKJF.786 TMC-Vanderbilt Age 21, White Male	
	Female, Age 66, BMI 31.3 HBM284 TRCV.726	























Search ontology terms ... 

- body
 - heart
 - lung
 - kidney
 - right kidney
 - left kidney
 - kidney capsule
 - cortex of kidney
 - outer cortex of kidney
 - renal medulla
 - outer medulla
 - inner medulla
 - renal column
 - renal pyramid
 - hilum of kidney
 - kidney interstitium
 - kidney calyx
 - major calyx
 - minor calyx
 - renal pelvis
 - ureter
 - renal papilla
 - renal fat pad
 - nephron
 - spleen
 - colon
 - small intestine

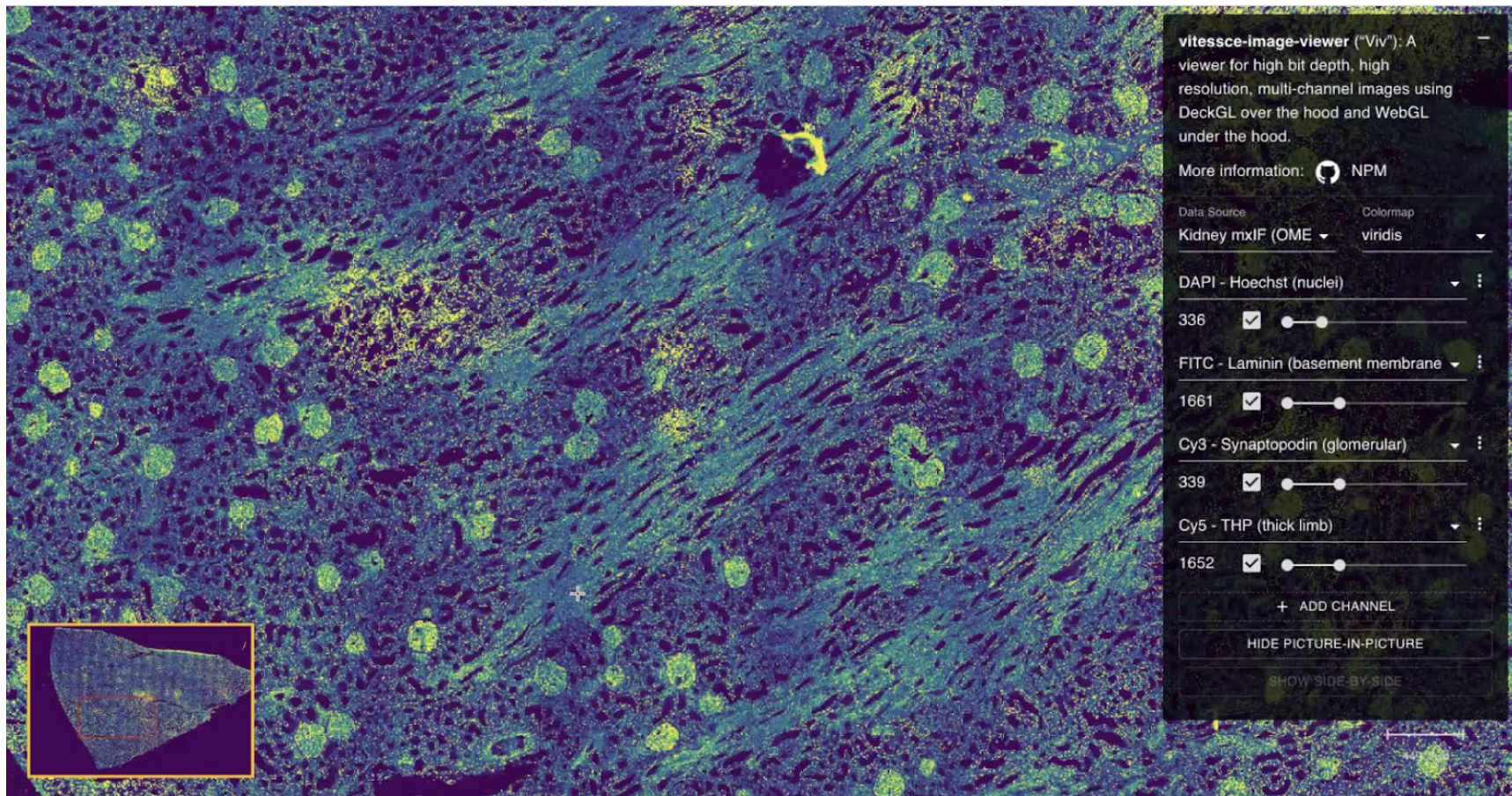


body

2 Centers
9 Donors
14 Samples

	CoverNephrectomy 10.1016/j.jrst.2017.07.006 KPMP-IJOSU Isolated as a part of a kidney st...	
	Patient B Cortical biopsy 10.1681/ASN.2016091027 KPMP-IJOSU Biopsy from Nephrology bioban...	
	Patient A Cortical biopsy 10.1681/ASN.2016091027 KPMP-IJOSU Biopsy from Nephrology bioban...	
	Male, Age 55, BMI 25.4 HBM824.BLXF.883 TMC-Vanderbilt 13-16	
	Female, Age 66, BMI 31.3 HBM554.ZRCG.496 TMC-Vanderbilt 21-24	
	Female, Age 58, BMI 23.0 HBM926.VBJV.597 TMC-Vanderbilt Age 58, White Female	
	Male, Age 62, BMI 34.9 HBM947.VLDP.894 TMC-Vanderbilt Kidneys 153-156	
	Female, Age 44, BMI 28.0 HBM457.NNQN.252 TMC-Vanderbilt Age 44, white female.	
	Male, Age 21, BMI 21.8 HBM693.HFJ.752 TMC-Vanderbilt Age 21, White Male, Trauma Pat...	
	Female, Age 58, BMI 23.0 HBM536.LDTZ.757 TMC-Vanderbilt Age 58, White Female	
	Male, Age 48, BMI 35.3	

Register your data via <https://hubmap-ccf-ui.netlify.app/rui/> so it can be spatially/semantically explored in EUI.



<http://gehlenborglab.org/research/projects/vitessce/>

Acknowledgements

HuBMAP Consortium (<https://hubmapconsortium.org>)



Thanks go to all the **patients** that agreed to volunteer healthy tissue and open use of their data.



TMCs



Jeffrey Spraggins
TMC-Vanderbilt
Vanderbilt University



Sanjay Jain
TMC-UCSD
Washington University,
St. Louis



Clive Wasserfall
TMC-UFL
University of Florida



Marda Jorgensen
TMC-UFL
University of Florida



Kristen Browne
Medical Imaging and
3D Modeling Specialist
NIAID

3D Models

MC-IU HIVE Team



Katy Börner
MCIU PI
CHS Director



Griffin Weber
Assoc. Professor of Medicine
Harvard Medical School



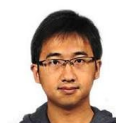
Lisel Record
MCIU PI
CHS Associate Director



Bruce Herr II
Sr. Systems Architect/PI



Ellen Quardokus
Sr. Research Analyst



Yingnan Ju
PhD Candidate



Andreas Bueckle
PhD Candidate



Leonard Cross
Sr. UI/UX Designer



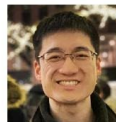
Matthew Martindale
Center Assistant



Daniel Bolin
Software Developer



Adam Phillips
Software Developer



Edward Lu
Software Developer



Paul Hrishikesh
Research Assistant



Leah Scherschel
Research Assistant



Avinash Boggana
Research Consultant



Yashvardhan Jain
Research Assistant



Kasturi Nikharge
Software Developer

References

Börner, Katy, Chen, Chaomei, and Boyack, Kevin. (2003). **Visualizing Knowledge Domains**. In Blaise Cronin (Ed.), *ARIST*, Medford, NJ: Information Today, Volume 37, Chapter 5, pp. 179-255. <http://ivl.slis.indiana.edu/km/pub/2003-borner-arist.pdf>

Shiffrin, Richard M. and Börner, Katy (Eds.) (2004). **Mapping Knowledge Domains**. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl_1). http://www.pnas.org/content/vol101/suppl_1

Börner, Katy (2010) **Atlas of Science: Visualizing What We Know**. The MIT Press. <http://scimaps.org/atlas>

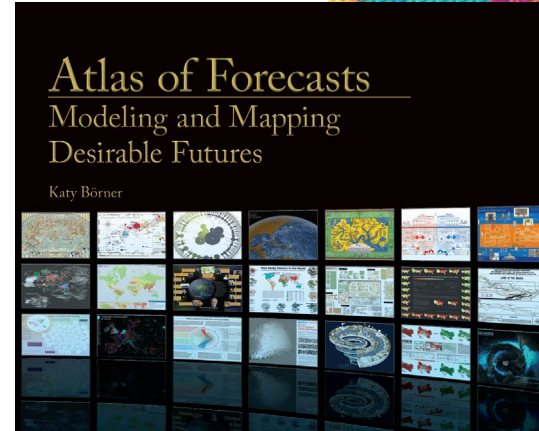
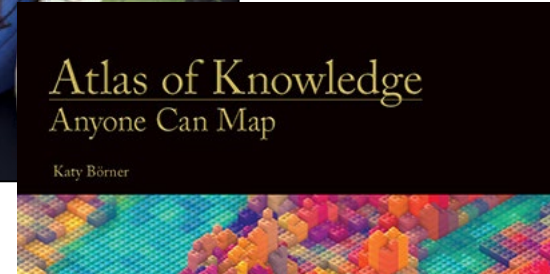
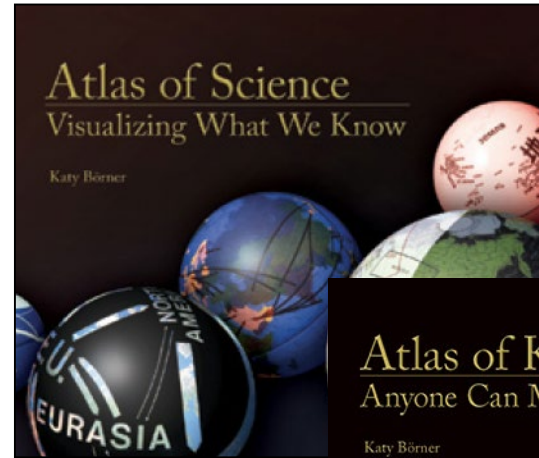
Scharnhorst, Andrea, Börner, Katy, van den Besselaar, Peter (2012) **Models of Science Dynamics**. Springer Verlag.

Katy Börner, Michael Conlon, Jon Corson-Rikert, Cornell, Ying Ding (2012) **VIVO: A Semantic Approach to Scholarly Networking and Discovery**. Morgan & Claypool.

Katy Börner and David E Polley (2014) **Visual Insights: A Practical Guide to Making Sense of Data**. The MIT Press.

Börner, Katy (2015) **Atlas of Knowledge: Anyone Can Map**. The MIT Press. <http://scimaps.org/atlas2>

Börner, Katy (2021) **Atlas of Forecasts: Modeling and Mapping Desirable Futures**. The MIT Press.





Q&A

