Sergio E. Baranzini, PhD

Weill Institute for Neurosciences

Department of Neurology

Graduate Program in Bioinformatics

Institute for Human Genetics

UCSF Microbiome Center

ImmunoX

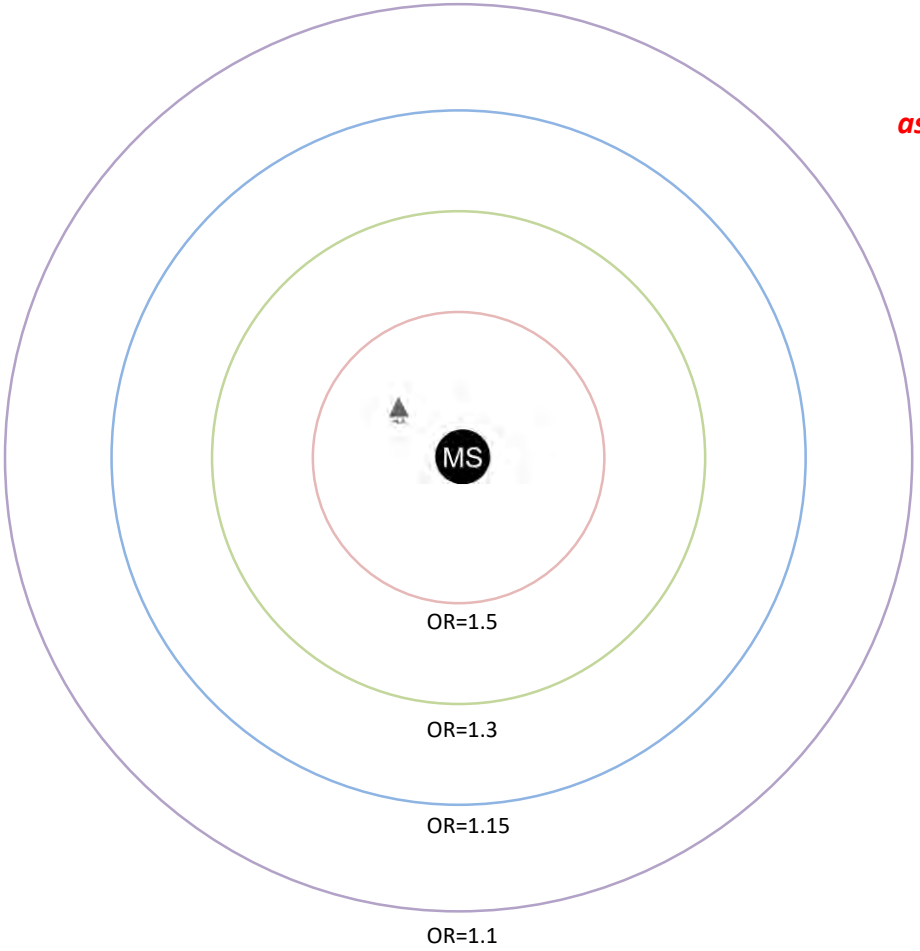
Bakar Computational Sciences Institute

University of California San Francisco

The value (and challenges) of shared work

- International multiple sclerosis genetics consortium
- MultipleMS Consortium
- BRAVEinMS Consortium
- International MS microbiome study

The Genetic Landscape of MS

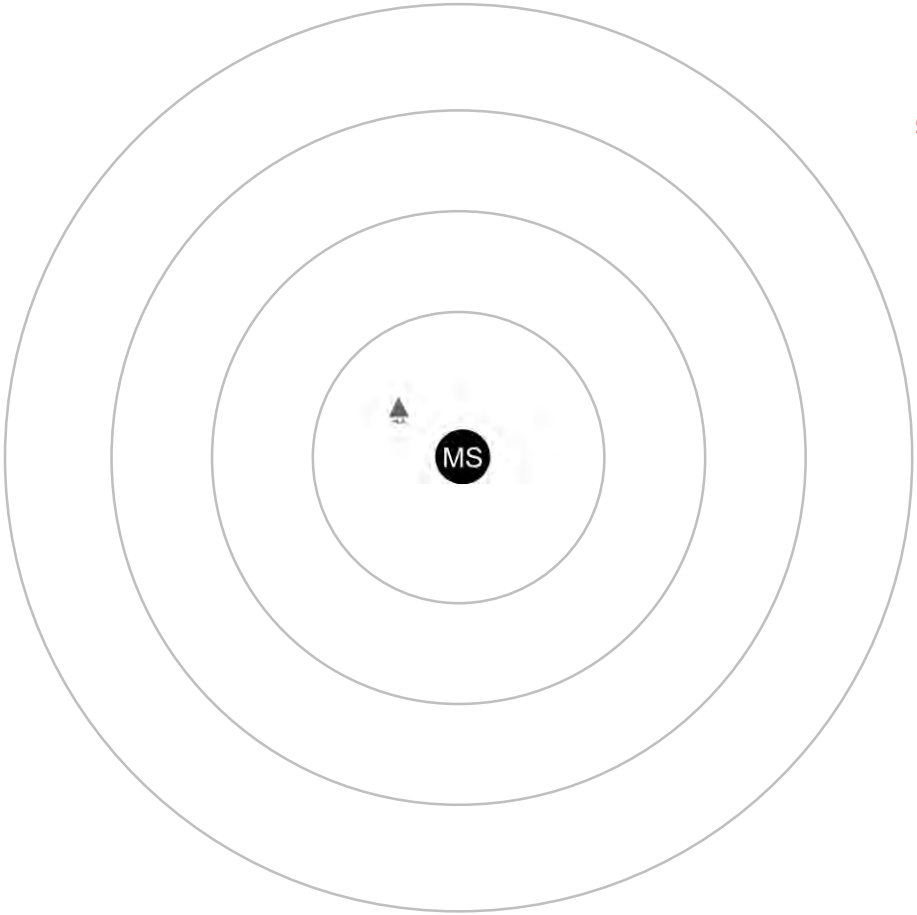


1972

First reported association between MS and HLA

Jersild et al. *Lancet* 1:1240, 1972

The Genetic Landscape of MS

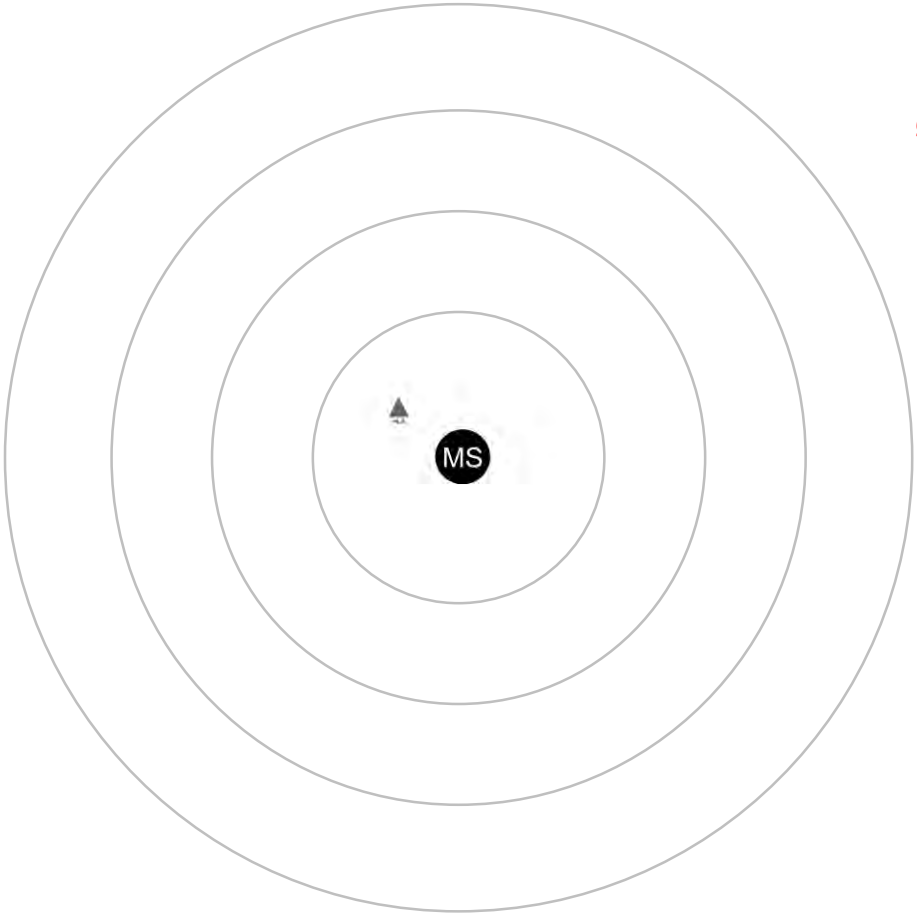


1996

First generation genome-wide linkage studies (400 markers)

Sawcer et al. *Nat Genet* 13:464, 1996
Haines et al. *Nat Genet* 13:469, 1996
Ebers et al. *Nat Genet* 13:472, 1996

The Genetic Landscape of MS



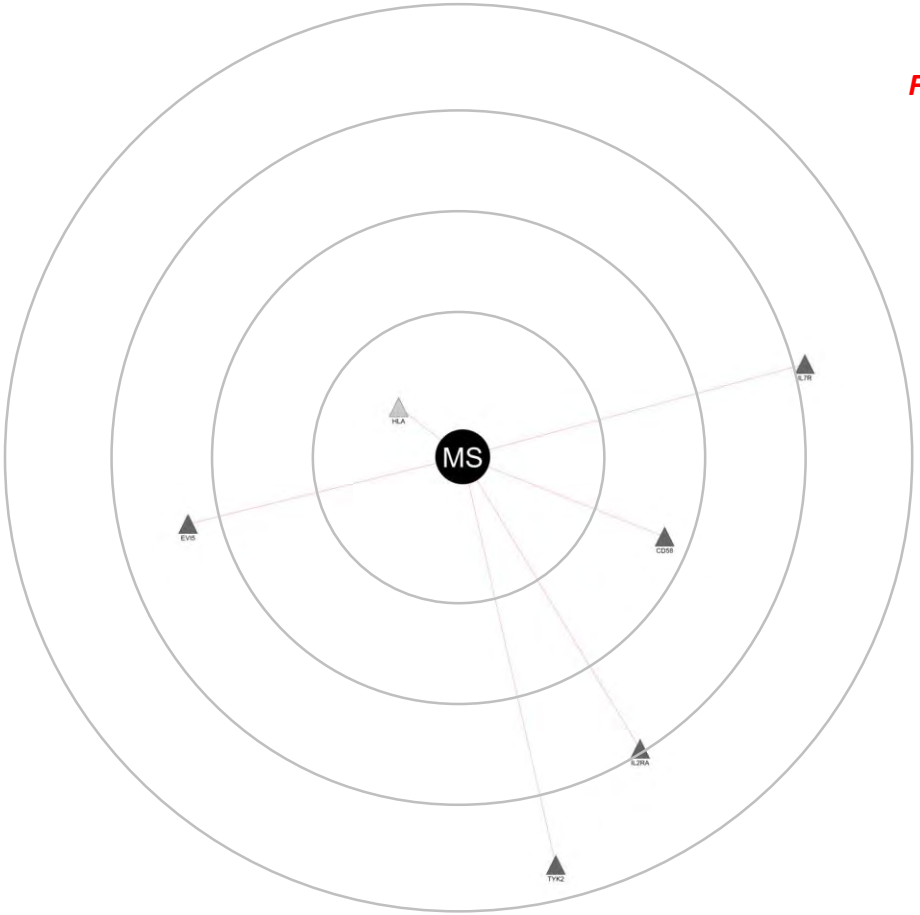
2005

Second generation genome-wide linkage study (5000 markers)

Sawcer et al. *Am J Hum Genet* 77:454, 2005

The International MS Genetics Consortium (IMSGC)

The Genetic Landscape of MS



2007

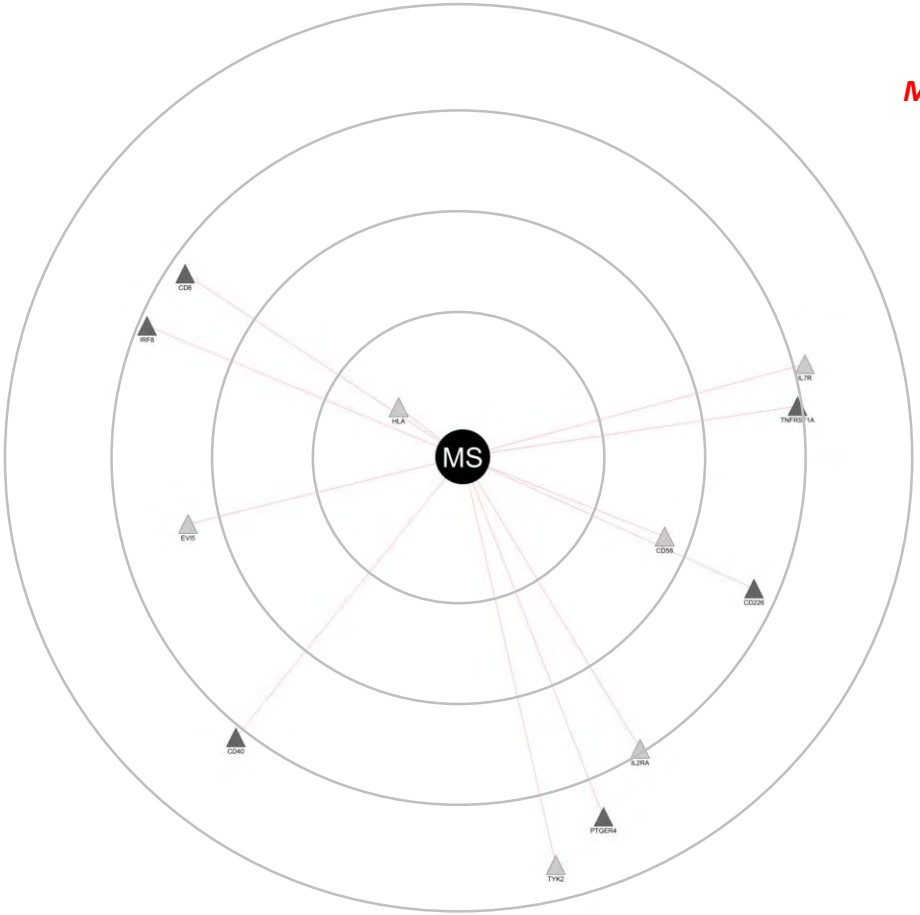
First generation GWAS
(1000 patients)

IMSGC. *N Engl J Med* 357:851, 2007

The Genetic Landscape of MS

2009

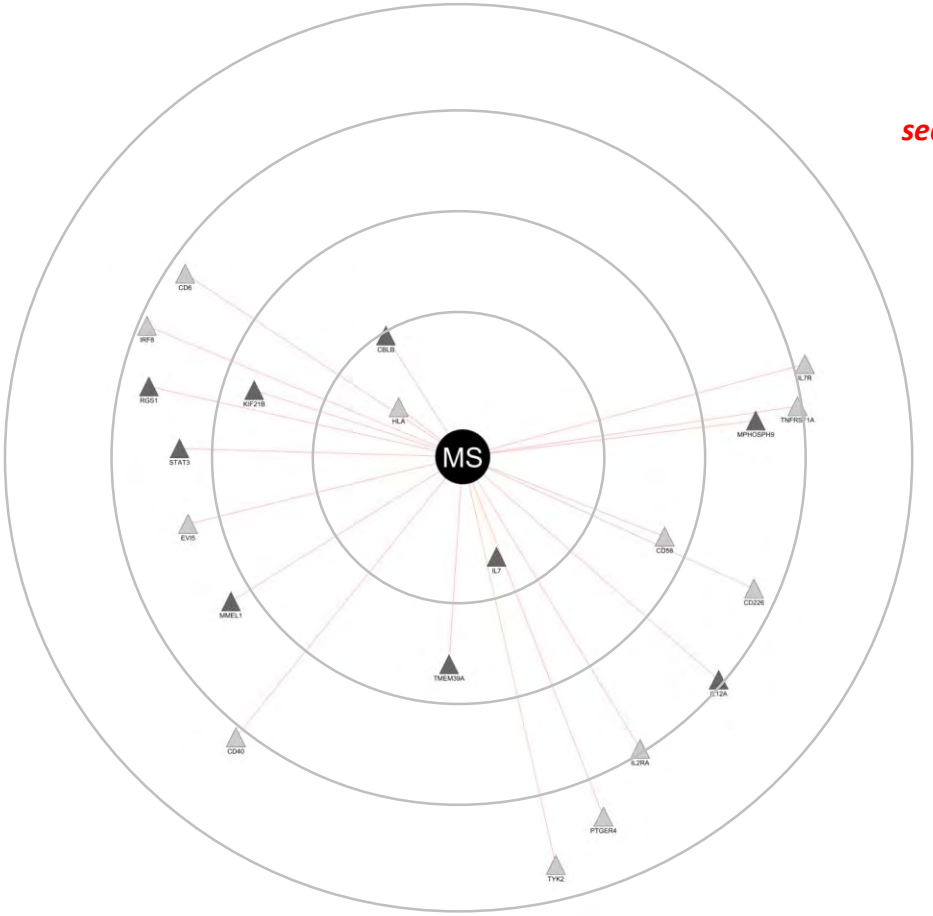
Meta-analysis of GWAS



IMSGC. *Genes Immun* 10:11, 2009

DeJager et al. *Nat Genet* 41: 776 , 2009

The Genetic Landscape of MS

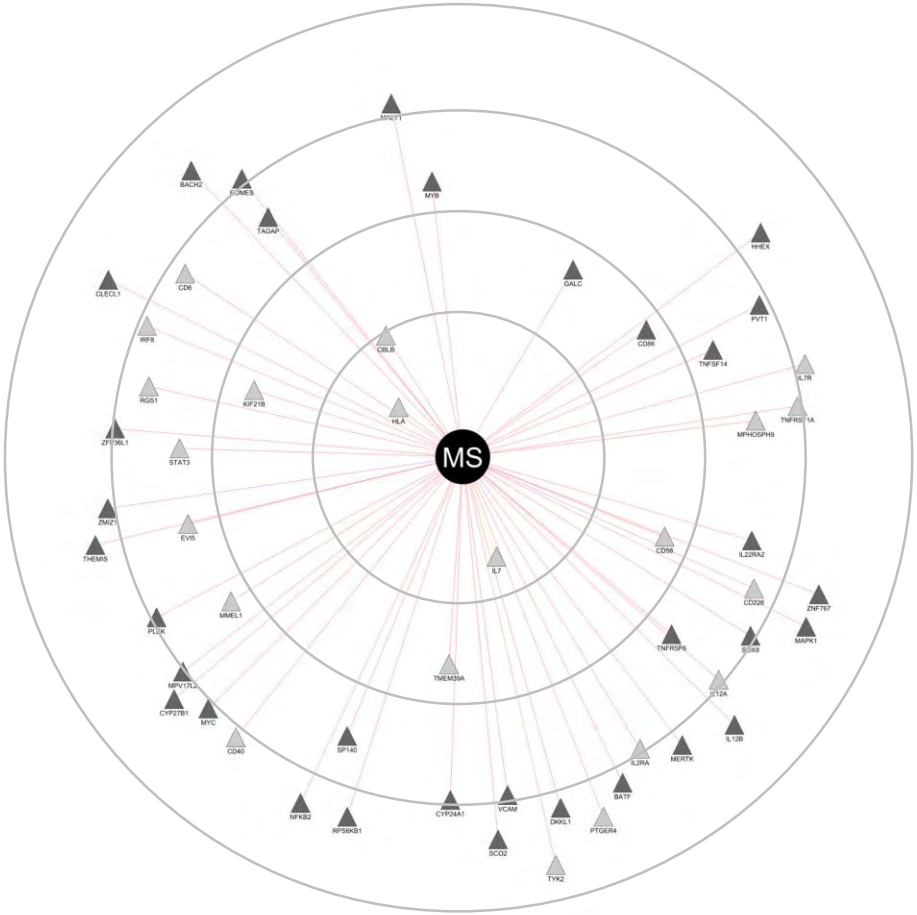


2010

Whole genome sequencing of MS twins

Ban et al. *Genes Immun* 11:660, 2010
Zuvich et al. *Hum Genet* 127:525, 2010
IMSGC. *Hum Mol Genet* 19:953, 2010
IMSGC. *Nat Genet* 42:469, 2010

The Genetic Landscape of MS



2011

Second generation GWAS (10,000 patients), GWAS Follow-ups

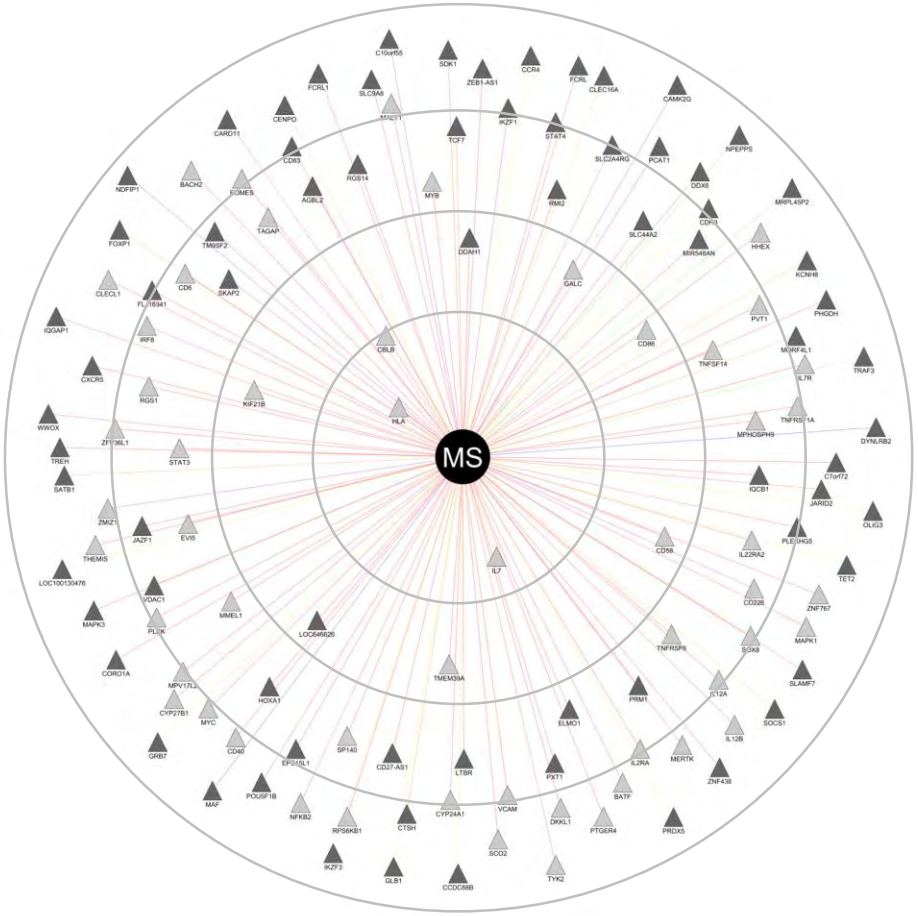
LETTER

doi:10.1038/nature10251

Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis

The International Multiple Sclerosis Genetics Consortium* & the Wellcome Trust Case Control Consortium 2*

The Genetic Landscape of MS



2013

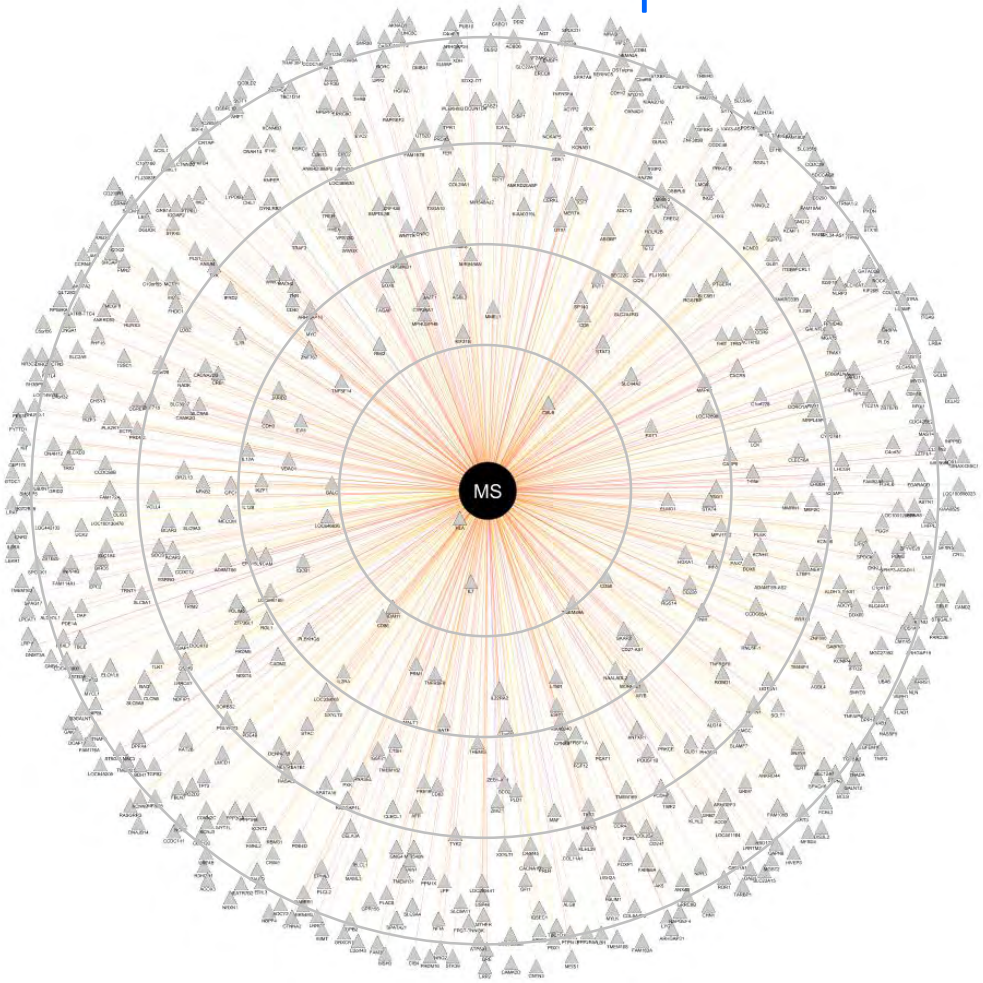
ImmunoChip
(>15,000 patients)

Published in final edited form as:
Nat Genet. 2013 November ; 45(11): 1353–1360. doi:10.1038/ng.2770.

Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis

A full list of authors and affiliations appears at the end of the article.

The Genetic Landscape of MS



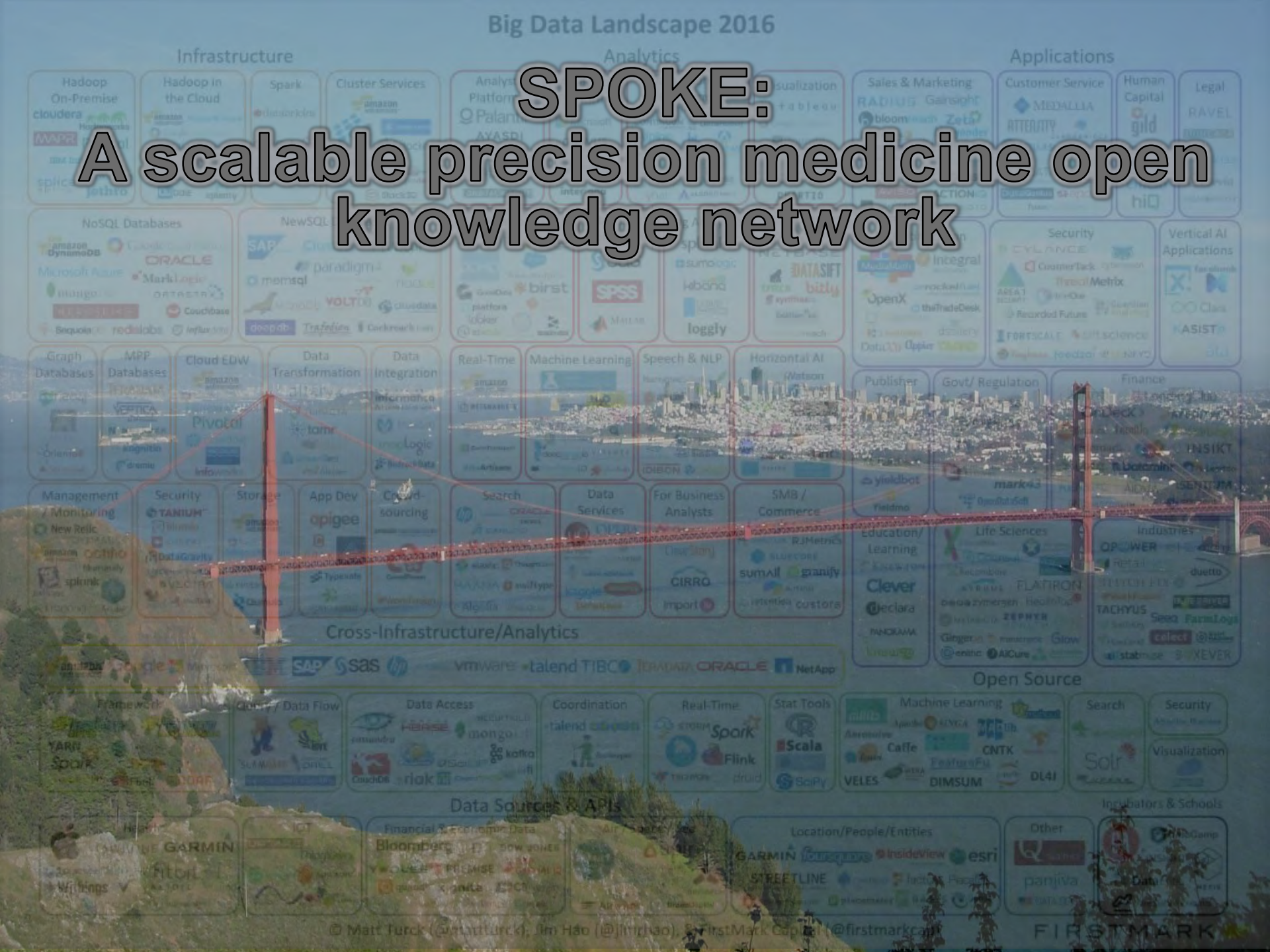
2019

Meta-analysis 3.0/MSchip
(>40,000 patients)

IMSGC. Science. In revision, 2018

Public databases
…Really?

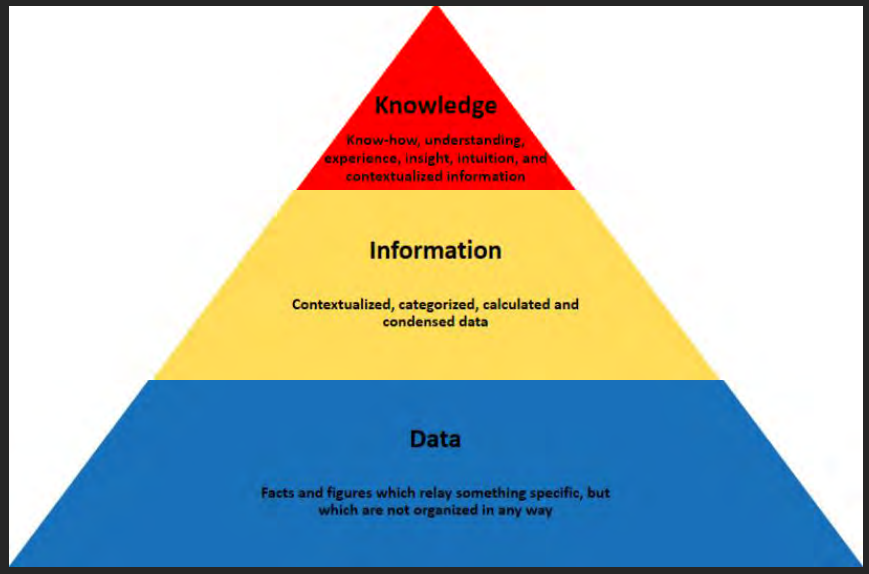
- SwissProt
- GWAS Catalog
- CHEMBL
- Diseases
- Drugbank
- KEGG
- STRING
- SIDER

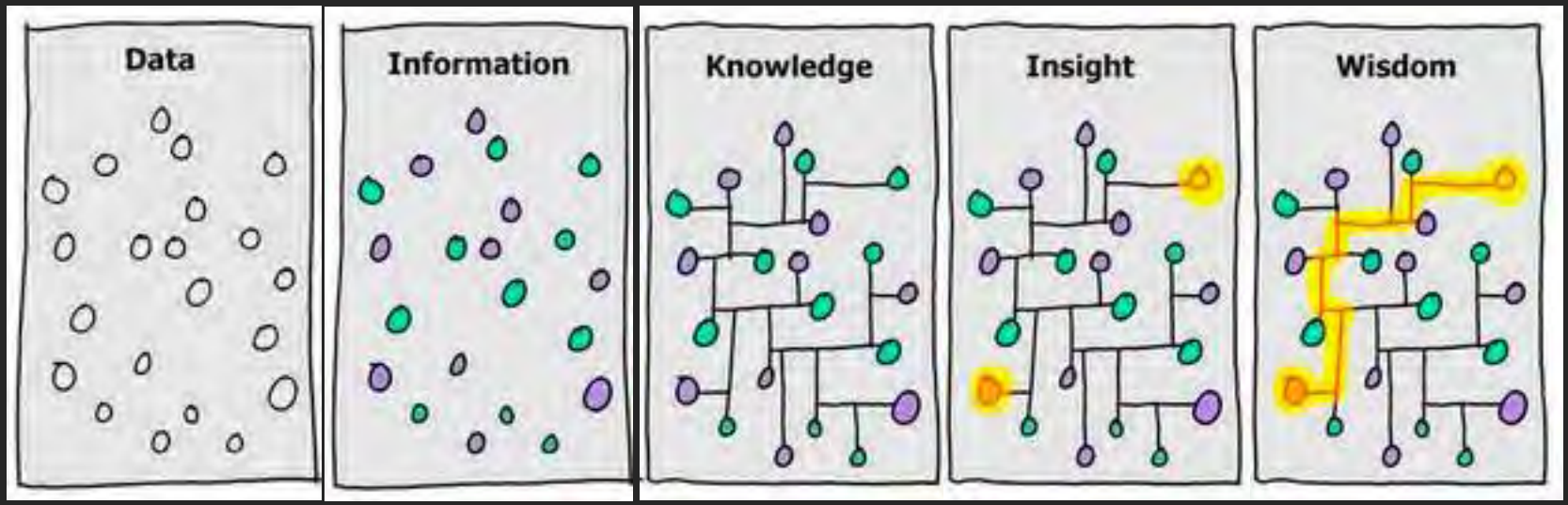

Big Data Landscape 2016
Infrastructure
Analytics
Applications
SPOKE:
A scalable precision medicine open knowledge network
Cross-Infrastructure/Analytics
Open Source
Data Sources & APIs

Building “Google maps” for health

- To develop a first-in-class AI system to track and predict disease outcomes
- System is based on a multi-scale, data-driven knowledge graph

Data -> information -> Knowledge

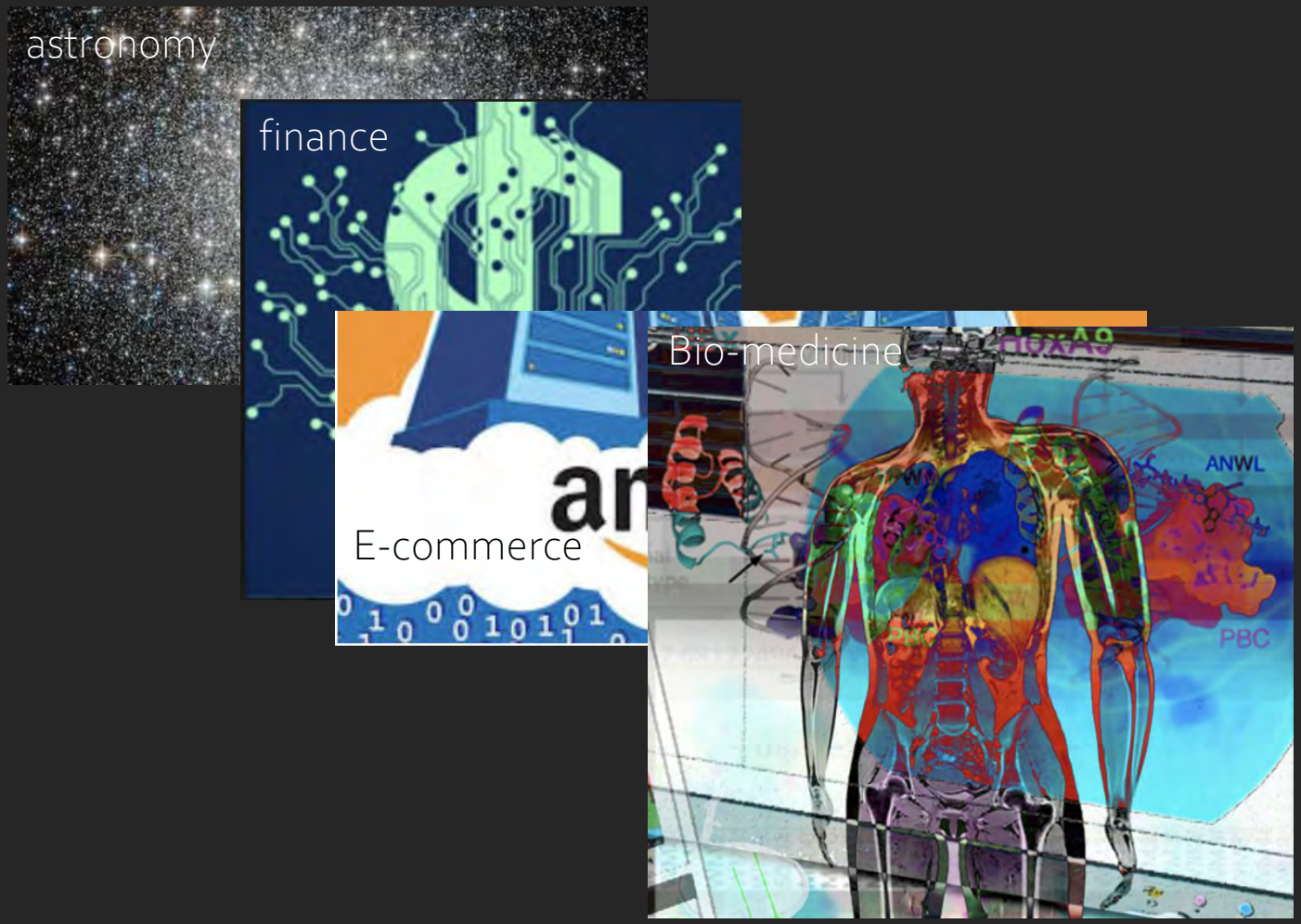




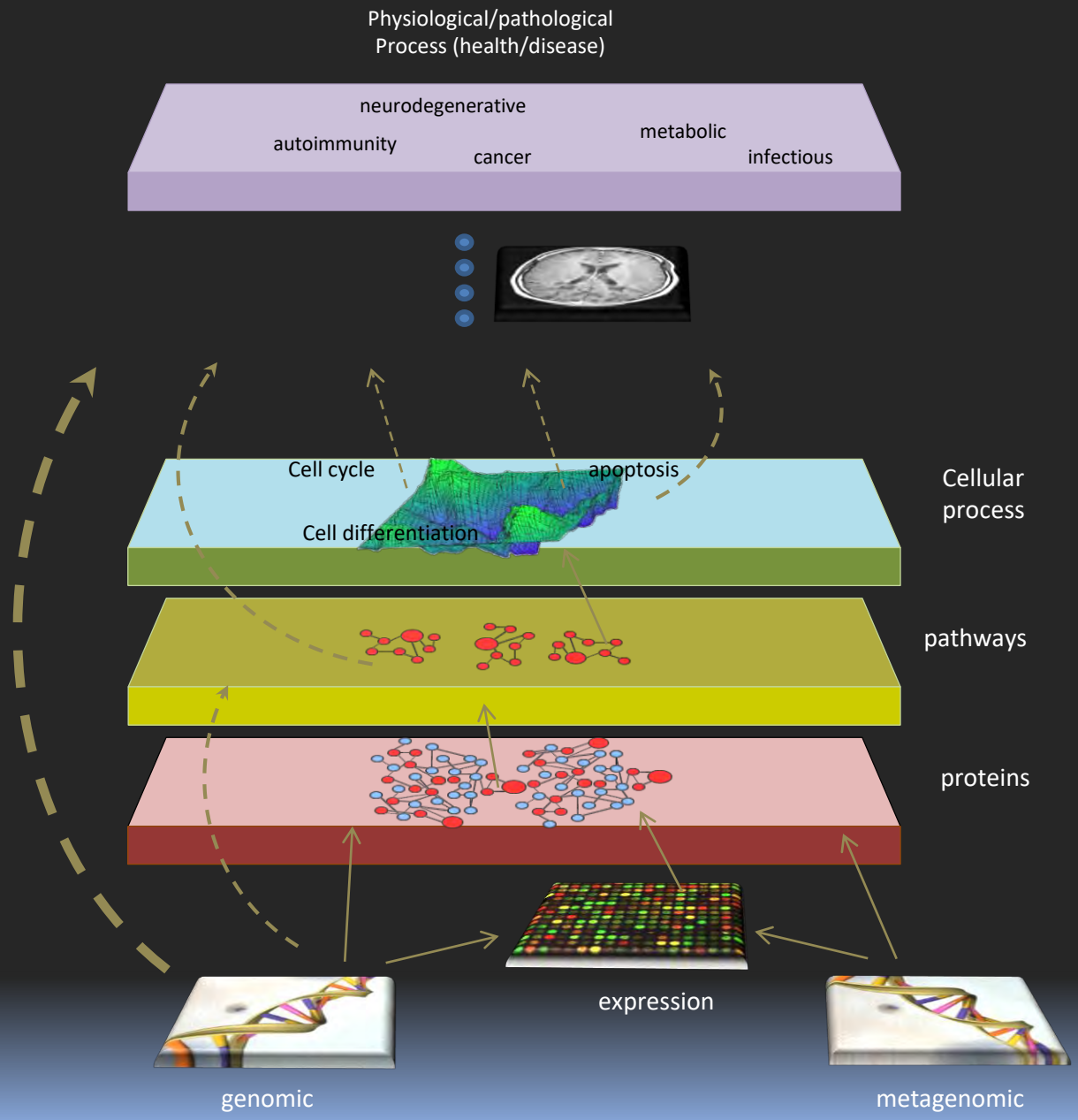
How much data is out there?

- ~33 Zettabytes in 2018 (~175 by 2025)
- more data was generated in 2017 than in the history of humanity
- 90% of the world's data was created in last 2 years

Who has the biggest problem to solve with Big data?



Hierarchical organization of biological complex



Orders of magnitude

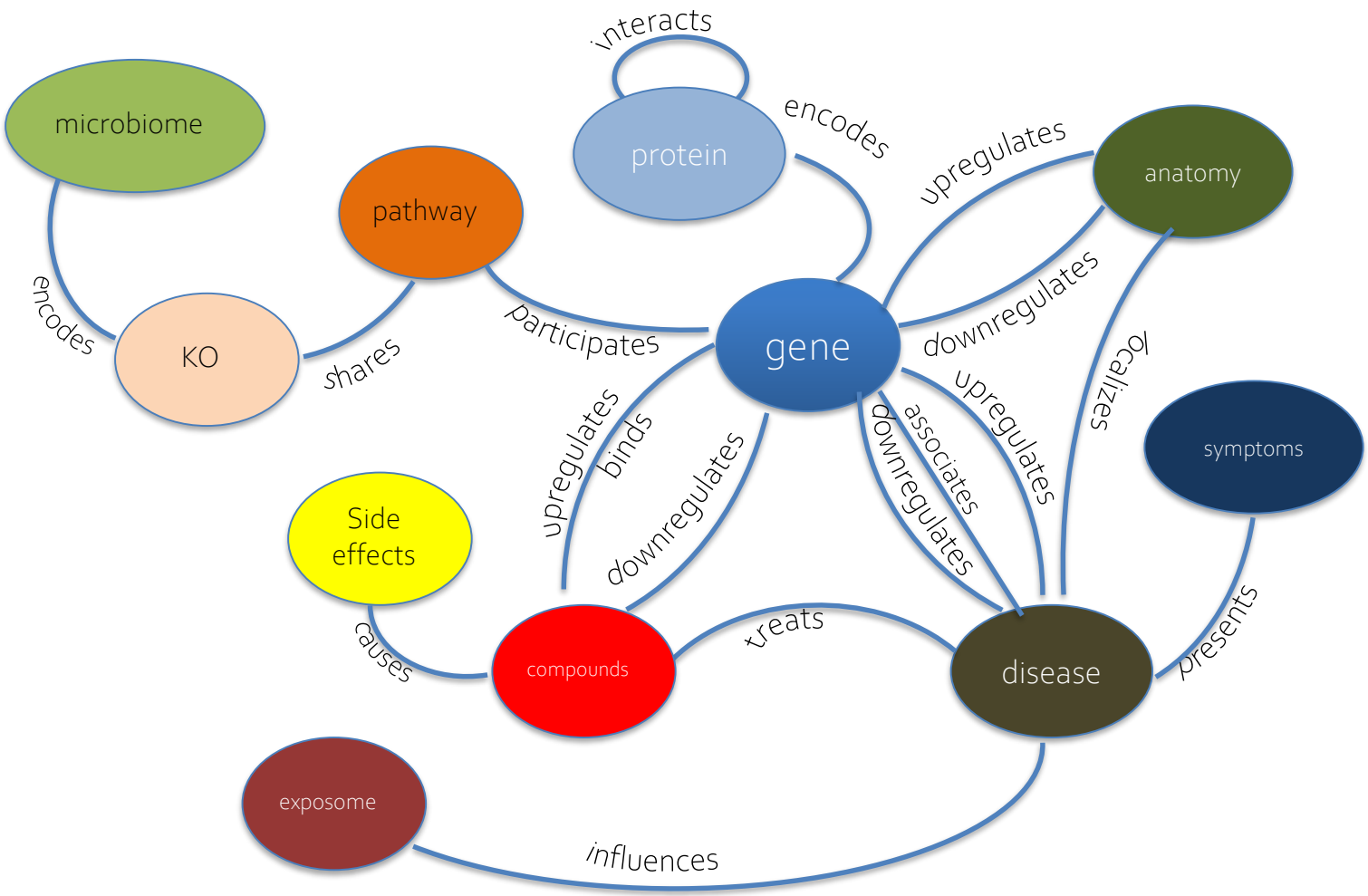
4x in space !

5x in time !!

SPOKE

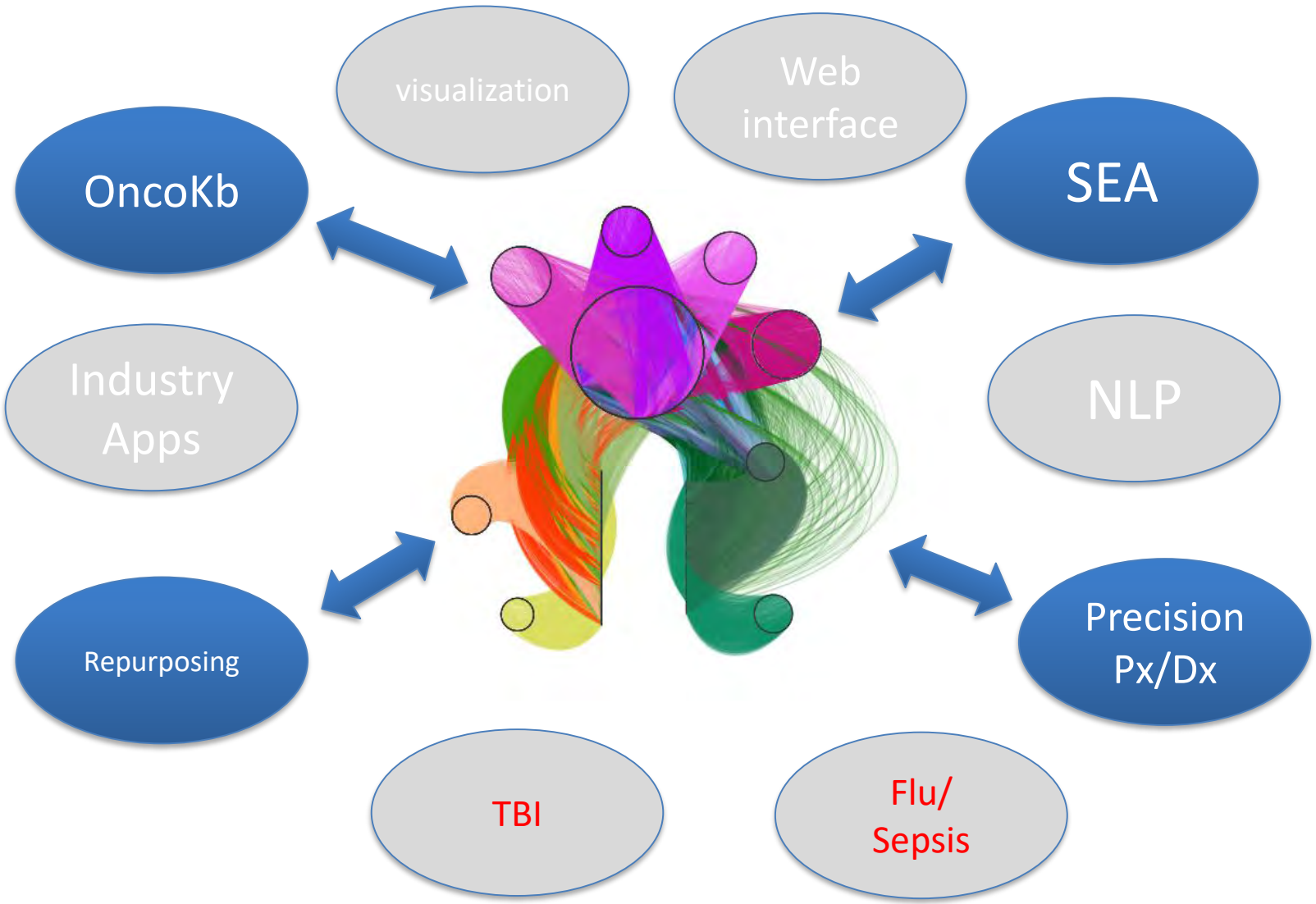
(Scalable Precision Medicine Open Knowledge Engine)

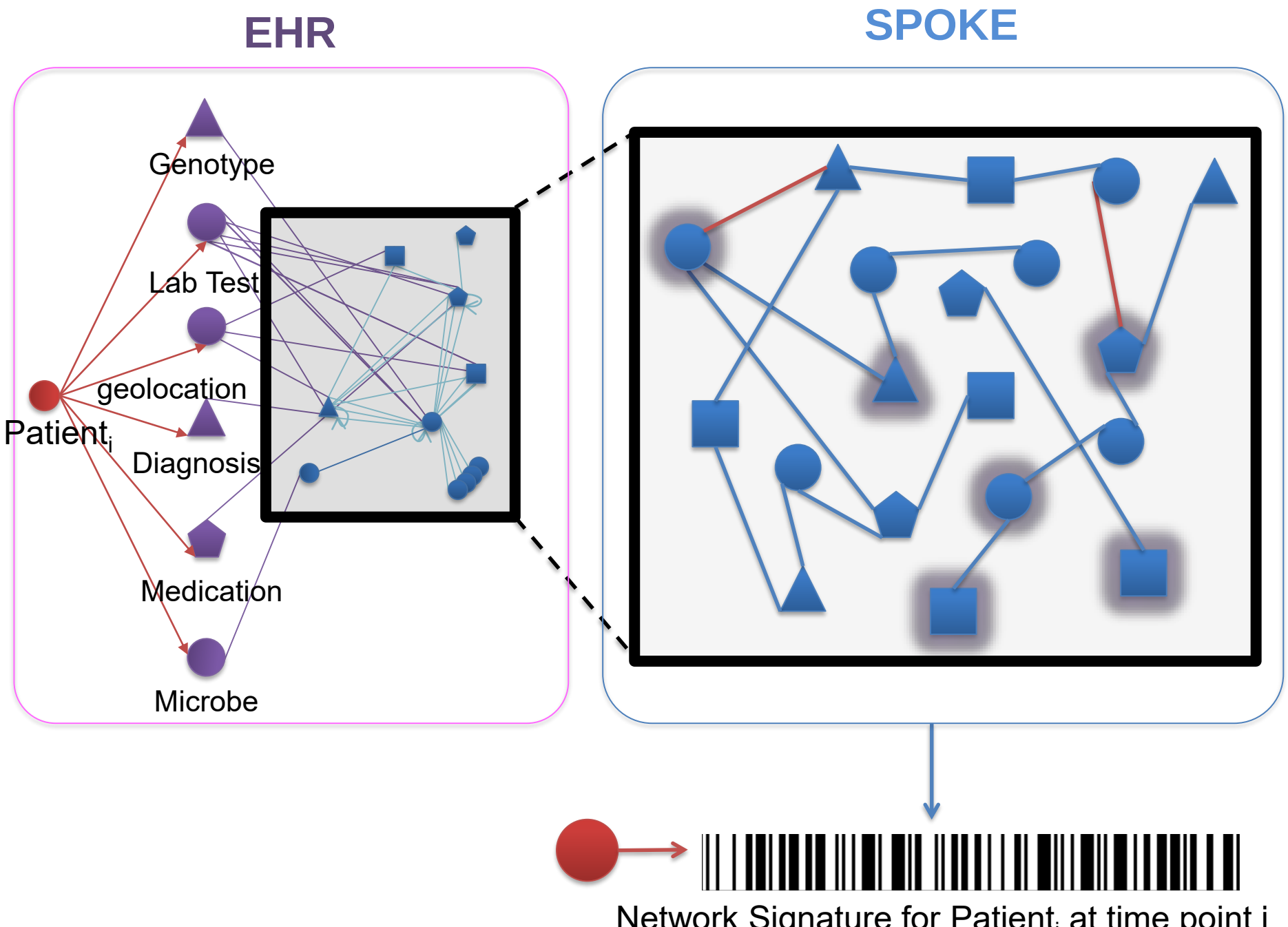
A database of databases



25+ databases
> 3 M nodes
> 5 M edges

SPOKE Apps model

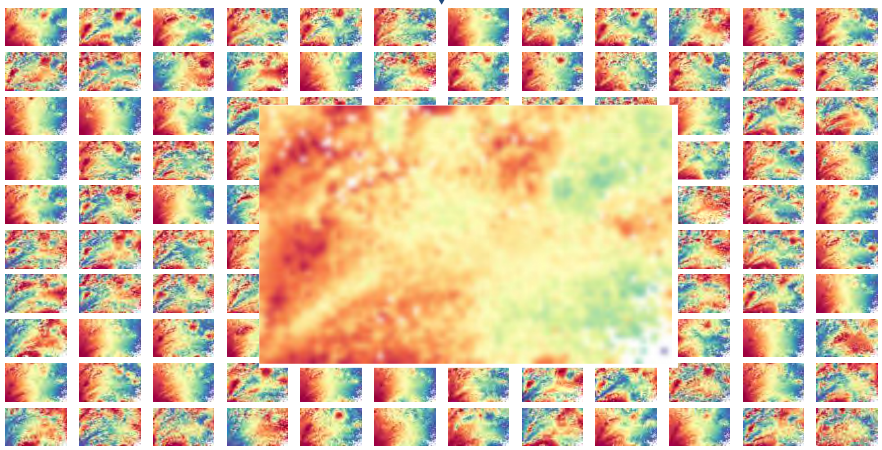




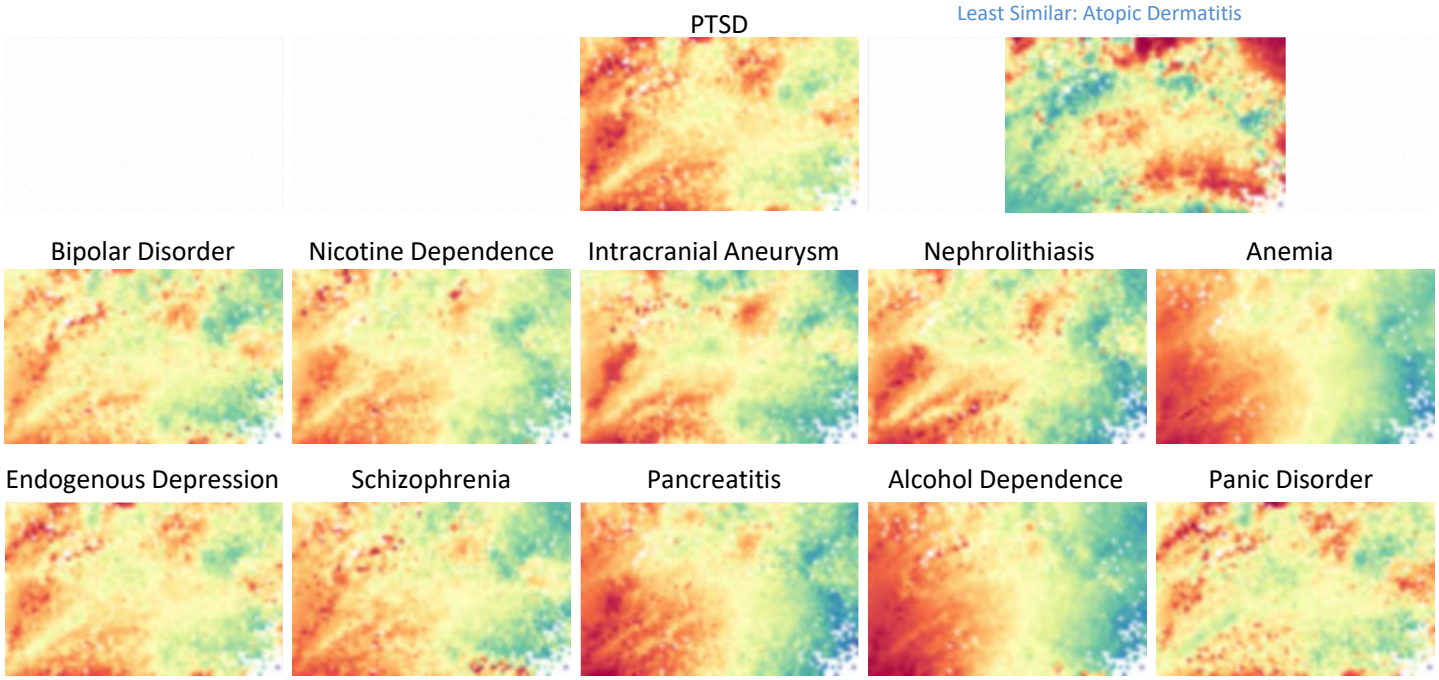
Nelson et al., Nat Comm, 2019.

UCSF Patient Cohort

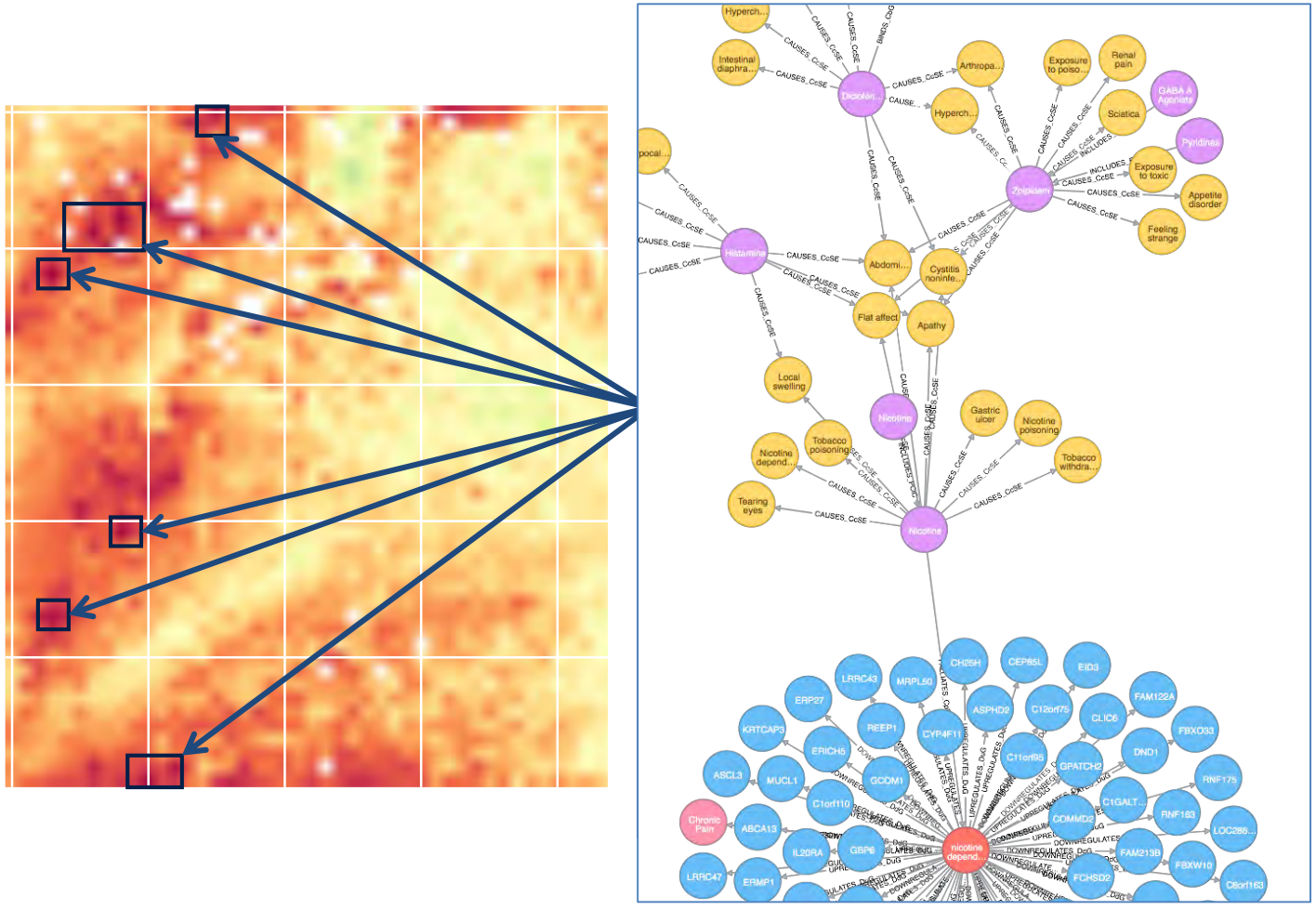
4,357 PTSD Patients



Most Similar Diseases to the Average PTSD Patient at UCSF



Highest Intensity Pixels for PTSD



National Science Foundation
WHERE DISCOVERIES BEGIN

Contact | Help

Search

NSB | Research Areas | Funding | Awards | Document Library | News | About NSF

Office of Integrative Activities (OIA)

- Office of Integrative Activities (OIA) Home
  - About
  - Programs
  - Staff
  - Funding
  - Awards
  - News
  - Events
  - Additional Resources

Home › Research Areas › Office of Integrative Activities

Email | Print | Share

NSF Convergence Accelerator (C-Accel)