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

The Genetic Landscape of MS

1996
First generation genome-wide linkage studies
(400 markers)

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)

The Genetic Landscape of MS

IMSGC. Genes Immun 10:11, 2009
The Genetic Landscape of MS

2010
Whole genome sequencing of MS twins

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

The International Multiple Sclerosis Genetics Consortium & the Multiple Sclerosis Genetics Consortium 2
The Genetic Landscape of MS

2013
ImmunoChip
(>15,000 patients)
The Genetic Landscape of MS

2019
Meta-analysis 3.0/MSchip
(>40,000 patients)
Public databases

...Really?

- SwissProt
- GWAS Catalog
- CHEMBL
- Diseases
- Drugbank
- KEGG
- STRING
- SIDER
**SPOKE:**
A scalable precision medicine open knowledge network
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?

- astronomy
- finance
- E-commerce
- Bio-medicine
Hierarchical organization of biological complexity

Physiological/pathological Process (health/disease)
- neurodegenerative
- autoimmunity
- cancer
- metabolic
- infectious

Cell cycle
- apoptosis
- Cell differentiation

Cellular process
- pathways
- proteins

Genomic expression

Orders of magnitude
4x in space!
5x in time!!
SPOKE
(Scalable Precision Medicine Open Knowledge Engine)
A database of databases

- microbiome
- KO
- pathway
- protein
- gene
- disease
- anatomy
- symptoms
- compounds
- exposome
- side effects

interacts, encodes, upregulates, downregulates, associates, binds, participates, causes, treats, presents, influences

25+ databases
> 3 M nodes
> 5 M edges
SPOKE Apps model

- OncoKb
- Industry Apps
- Repurposing
- SEA
- NLP
- Precision Px/Dx
- Flu/ Sepsis
- TBI
- Web interface
- visualization
Network Signature for Patient$_i$ at time point j

Most Similar Diseases to the Average PTSD Patient at UCSF

Bipolar Disorder, Nicotine Dependence, Intracranial Aneurysm, Nephrolithiasis, Anemia, Endogenous Depression, Schizophrenia, Pancreatitis, Alcohol Dependence, Panic Disorder

Least Similar: Atopic Dermatitis
Highest Intensity Pixels for PTSD
NSF Convergence Accelerator (C-Accel)