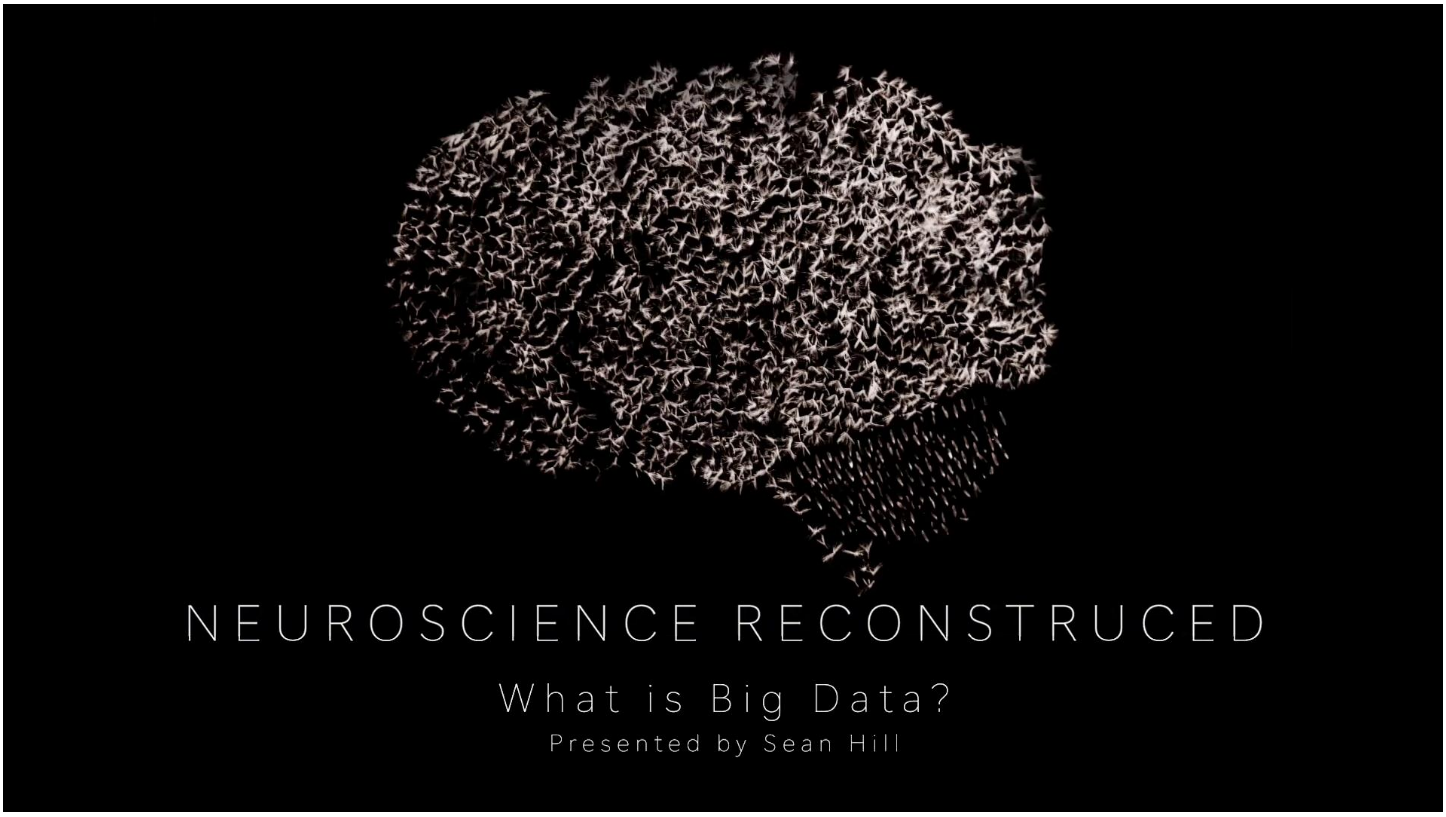
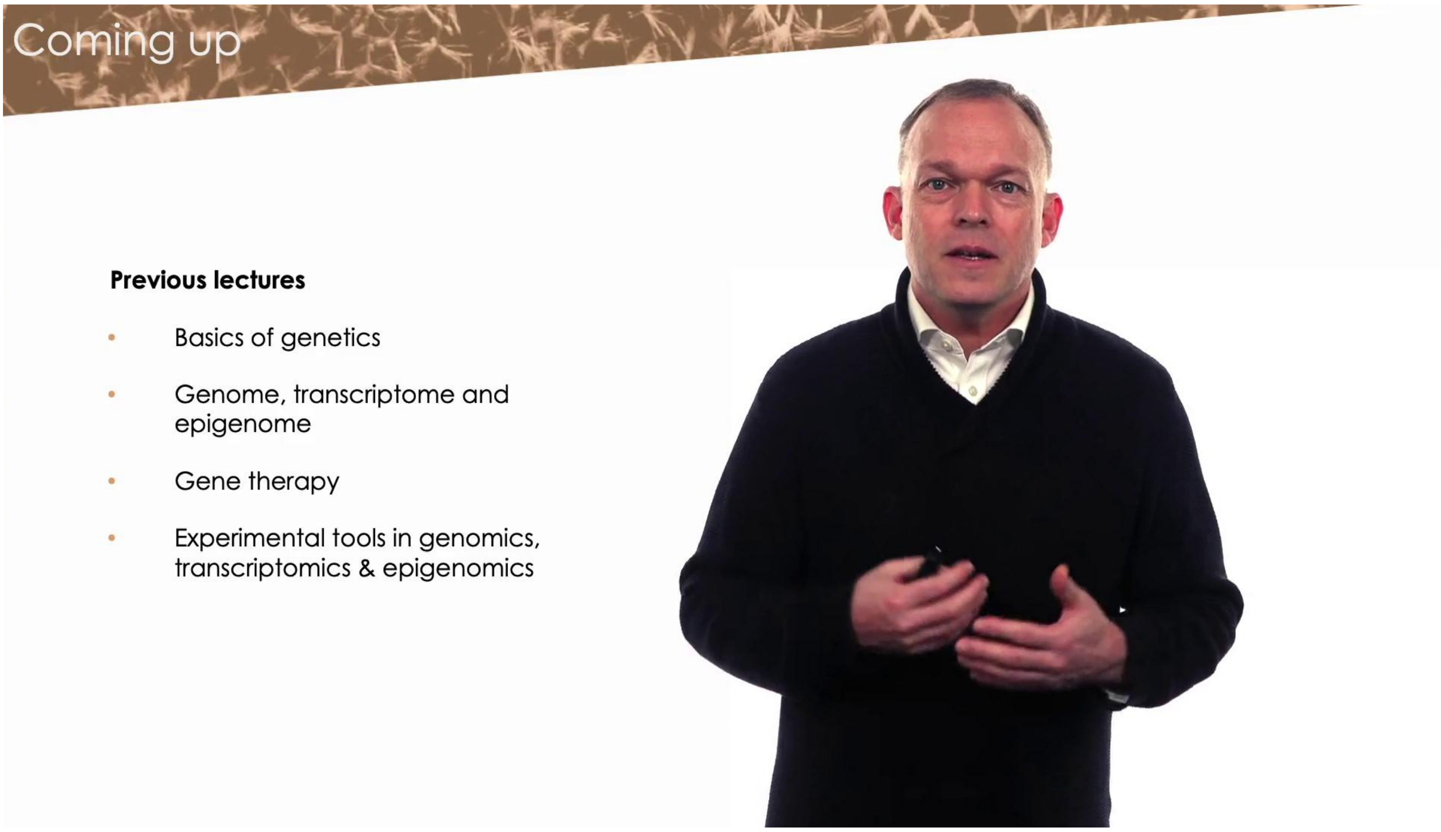


Search MOOC

Video

EPFL



Hello. My name is Sean Hill, and today I'll be talking to you about big data in neuroscience. I'll be giving you an introduction. In your previous lectures, you've learned about the basics of genetics and the genome and gene therapy.

Notes

Summary

0m 09s

More data – what for?

Developmental disorders

Adolescent disorders

Adult disorders

The Worldwide Cost

Disability-adjusted life years (DALY) is a commonly used public-health metric meant to express the number of years lost as a result of bad health, disability, or premature death.

Disorder	DALYs
Unipolar depressive disorders	65,000,000
Bipolar affective disorder	14,000,000
Schizophrenia	17,000,000
Epilepsy	8,000,000
Alcohol use disorders	24,000,000
Alzheimer's and other dementias	11,000,000
Drug use disorders	8,000,000
Post-traumatic stress disorder	3,000,000
Obsessive-compulsive disorder	5,000,000

Sources: Harvard School of Public Health and the World Economic Forum (2011)

MIT Technology Review

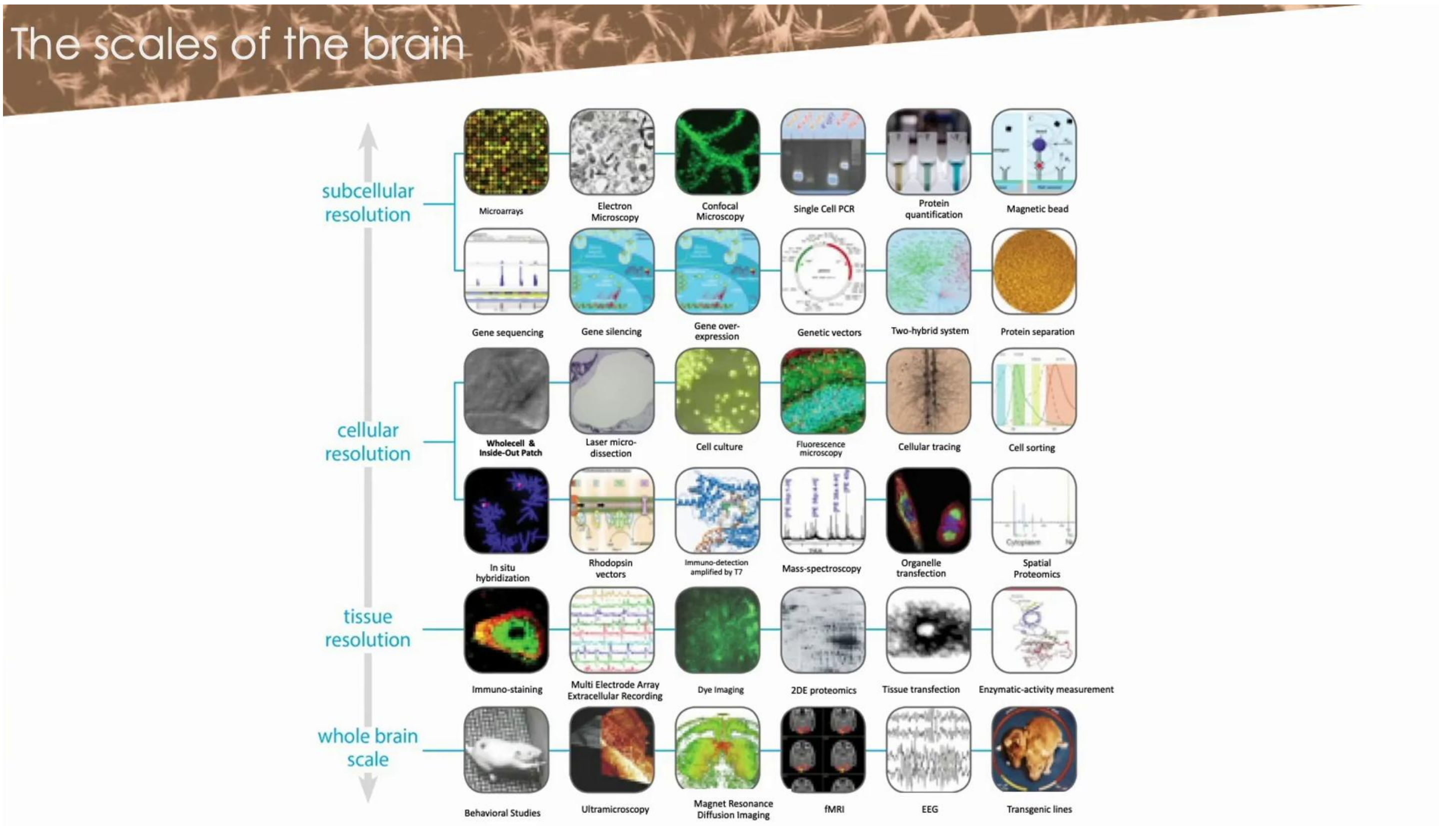
Glutamate
Nutrition
Dopamine
Genes
Sugar
GABA
Myelin
Serotonin
Metals
Dopamine
Toxins
Acetylcholine
Protein misfolding

But today, I want to talk about big data in neuroscience. What is that? What does that mean? How do we use text mining to get data? What is metadata and knowledge graphs? How are they used to organise data in neuroscience? What's the role of ontologies and atlases in organising data? In general, the big challenge is how do we actually make use of data effectively in neuroscience? What is data in neuroscience for? Why do we need more data? Well, the challenge is that worldwide, there are a tremendous number of brain disorders that are causing a tremendous cost, not only at the individual level and in terms of the challenges for the individual families and friends, but for our society, and the costs are tremendous worldwide. We can see that there's a broad range of different disorders that emerge at various points in the lifespan, and they can be caused by many different underlying causes. We really don't understand those well.

Notes

Summary

0m 26s

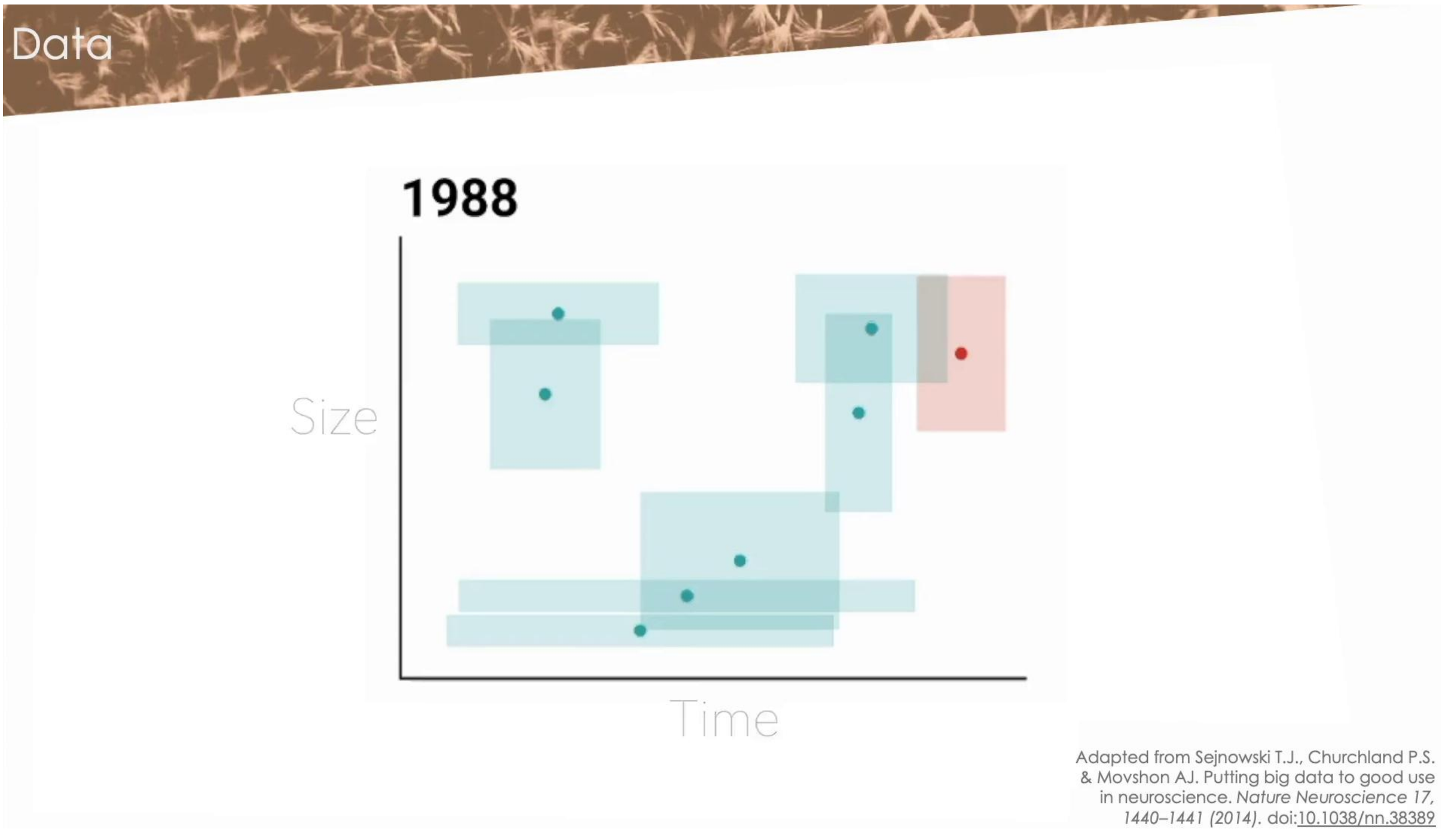


The data is actually coming from many different levels of the brain, and that's one of the core challenges. The data can be coming from the subcellular level, the cellular level, the tissue, the whole brain, individuals, and it's measured using many different techniques.

Notes

Summary

1m 39s

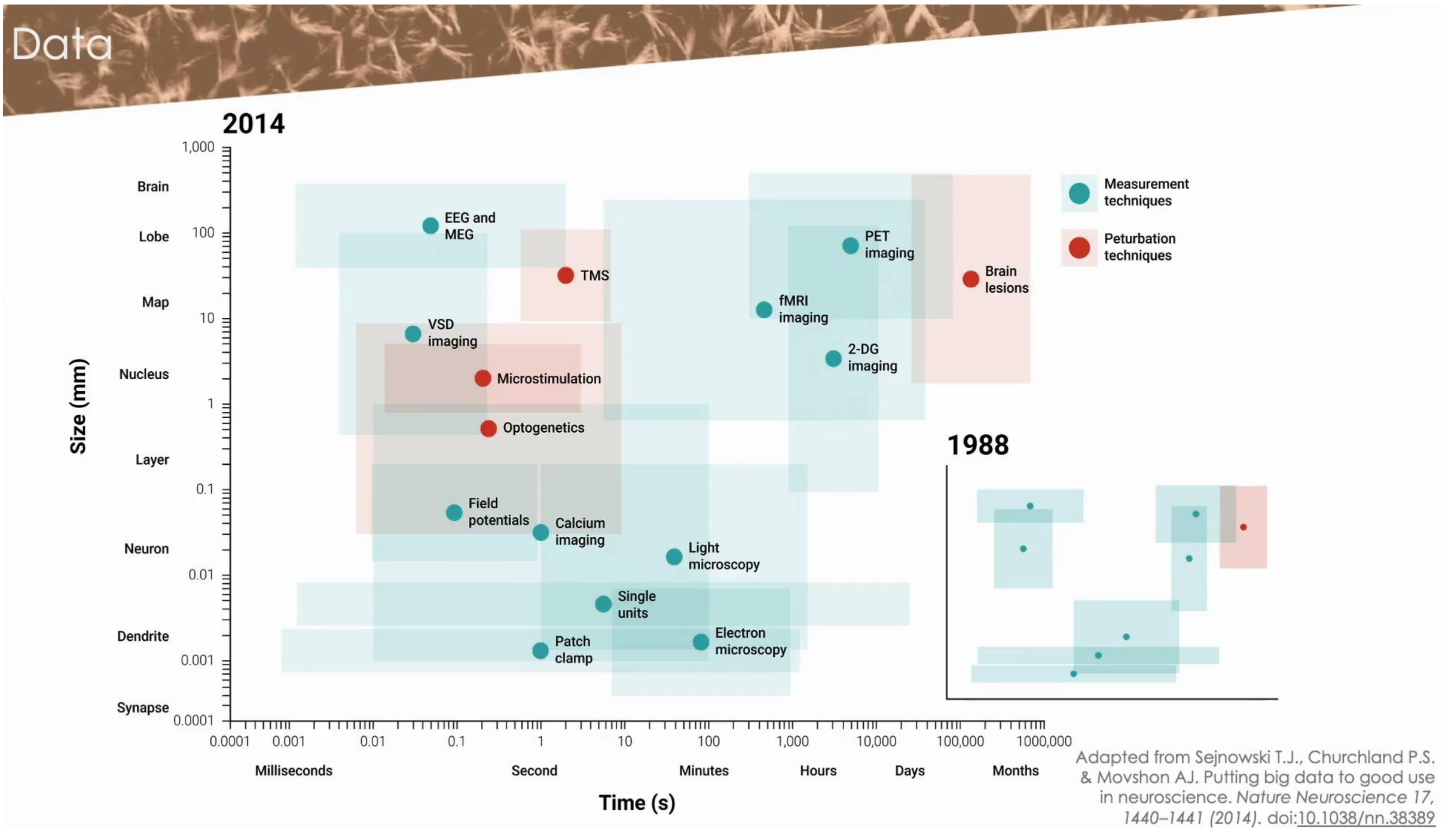


In fact, we've seen that the number of possible measurement techniques, as well as the ways in which we can probe the system through perturbations, have continued to evolve tremendously over the years.

Notes

Summary

2m 00s

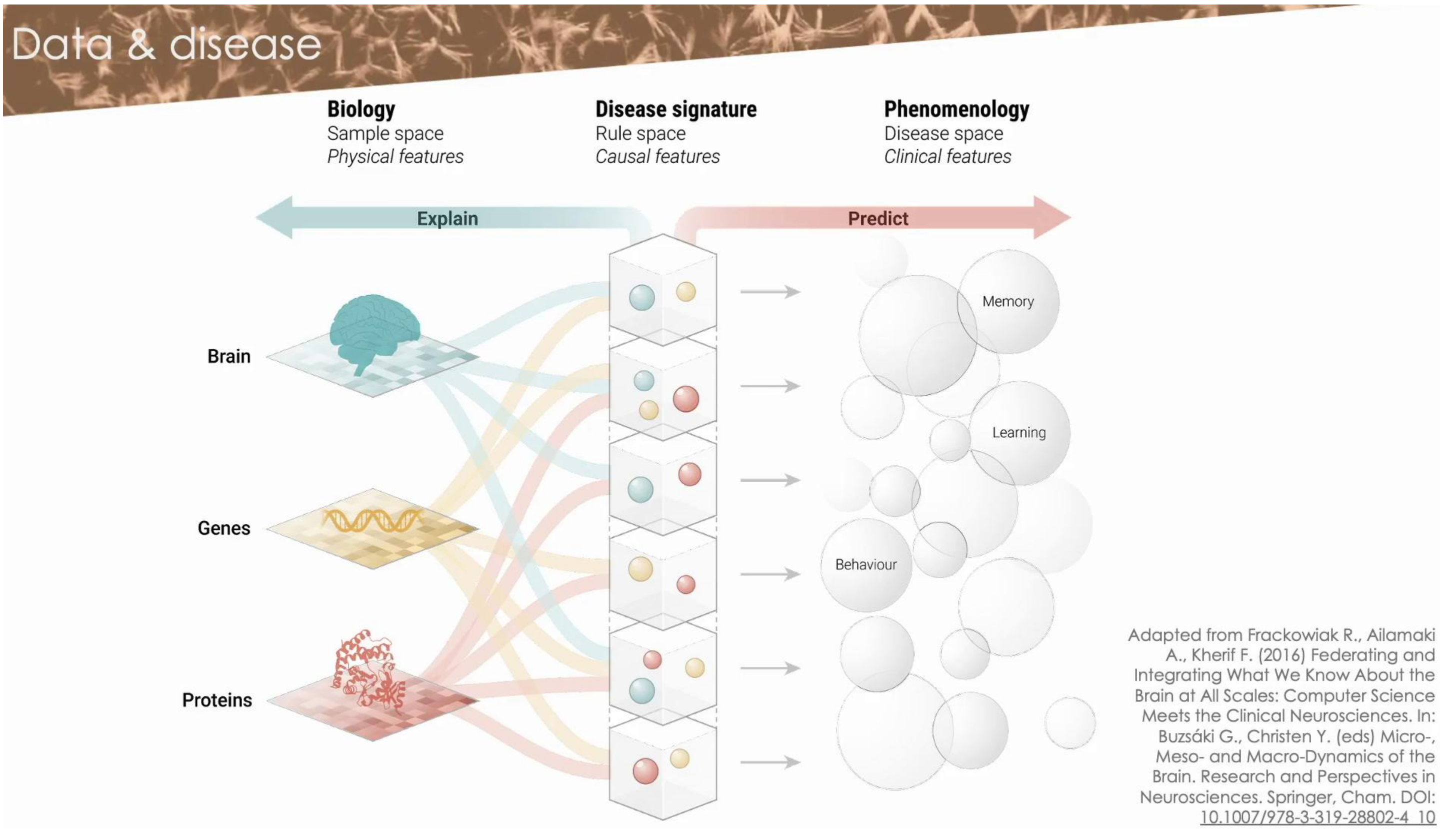


Comparing 1988 to 2014, we've got a tremendous new array of techniques. That's phenomenal from the point of being able to investigate the brain at many additional levels. But it's a big challenge in terms of how do we understand the data in the context from where it's coming, from all of these different levels? How do we start to put that together to understand how these levels of the brain relate to each other and help us build a more integral understanding of brain disorders?

Notes

Summary

2m 14s



That's really where we're working. It's a challenge to go from the level of the brain, the genes, the proteins, to how those different levels interact, the causal, we could call it a rule space? What are the different types of interactions and the perturbations in those interactions that occur in the context of a disease to create the symptoms that are currently used to actually define brain disorders and mental illness. Mental illness is primarily defined according to the symptoms that an individual exhibits. We don't really have the disease signatures yet. We don't really know the underlying causal systems and underlying biophysical basis of these mental disorders. That's a really grand challenge.

Notes

Summary

2m 47s

What is big data? Big data really consists of, of course, very large data. It can also consist of very variable data. It can be mainly generated through high throughput methods. There are tremendous new techniques, actually, single-cell transcriptomics, large brain imaging, that are producing tremendous amounts of data, but these need to be analysed computationally in order to extract the biological meaning and the key features that are useful to understanding the biology.

Notes

Summary

3m 40s

There's molecular data, there's cellular data, there's many different types of organisms, the environmental context, the experimental context, the properties of the brain. There's the Human Connectome Project, which has produced a huge amount of data from brain imaging data. The BRAIN Initiative Cell Census Network is producing tremendous amounts of individual cellular level data. The Allen Institute is producing data from all different levels of the brain, including in behaviour. Across many different... All of these efforts are working across many different levels and species. We've got to be able to organise that, make sense of it, and be able to extract from that meaning in order to understand the biological system. At the core of that are some key challenges. For example, how do we find relevant data? How do we search for it? You know how to search for something on Google. We don't have a way of searching for relevant data yet in neuroscience, and that means we've got to organise and describe the data, we've got to standardise how we represent and interpret that data, and we've got to be able to identify which data can we actually combine so that we can, for example, get greater statistical power or understand how large amounts of data can help us better understand the variability in the real biological systems.

Notes

Summary

4m 17s

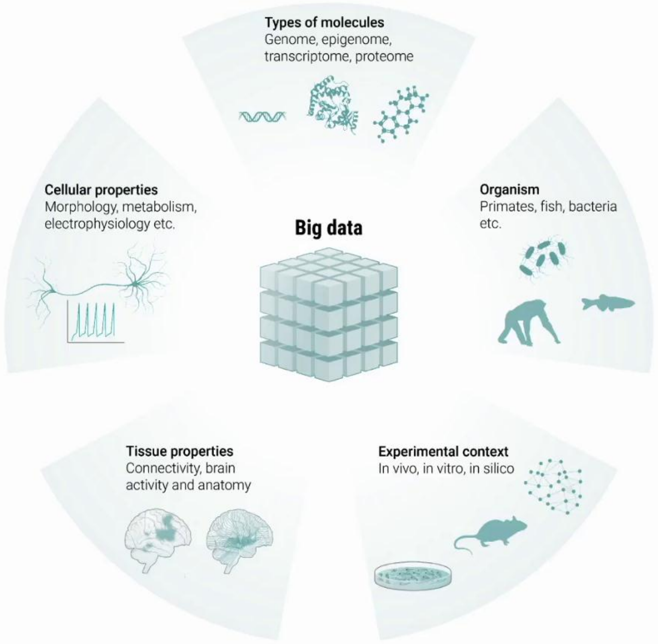
One of the ways that we do to navigate the complexity of all of this data at different levels and understanding the causal mechanisms between the levels is to build models or visualise that data in order to gain greater insight.

Notes

Summary

5m 50s

Challenges of Big Data



Challenges

- Finding relevant data
- Organizing and describing data
- Standardizing and interpreting data
- Integrating data
- Building models
- Visualizing data

Tools

- Text mining
- Metadata and knowledge graphs
- Ontologies
- Atlases

Now, some other ways to extract data, because a lot of data is actually in the publications, is text mining. We'll talk a bit about that. One way to organise this data is to have rich metadata, that is data about the data, that describes the data, describes properties of it, describes how it was captured, describes who produced it, for example, and organise that in knowledge graphs, and we'll talk a bit about that. One of the key ways to ensure that you have high quality metadata is to use ontologies or these controlled vocabularies with clear definitions of the terms and their relationships between other terms within your metadata. Atlases are yet another key way. Brain atlases give us another way to organise the data, and we'll talk a bit about that.

Notes

Summary

6m 07s