RE-ENGINEERING THE FUTURE OF HEALTH WITH PREDICTIVE MODELS

IMAG

MULTISCALE MODELING CONSORTIUM

NIH National Institute of Biomedical Imaging and Bioengineering

New York Scientific Data Summit (NYSDS)

Grace C.Y. Peng, Ph.D.

October 22, 2020

Virtual

The Future of Health

- Real-time measurements
- Digital Twins
- Salutogenesis

Predictive Modeling

- Types
- Process
- Credible practice

Biomedical and Behavioral Research Culture

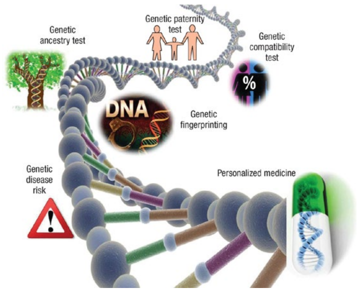
- Modeling approaches
- Data landscape
- Heterogeneity and uncertainty

How do we get there?

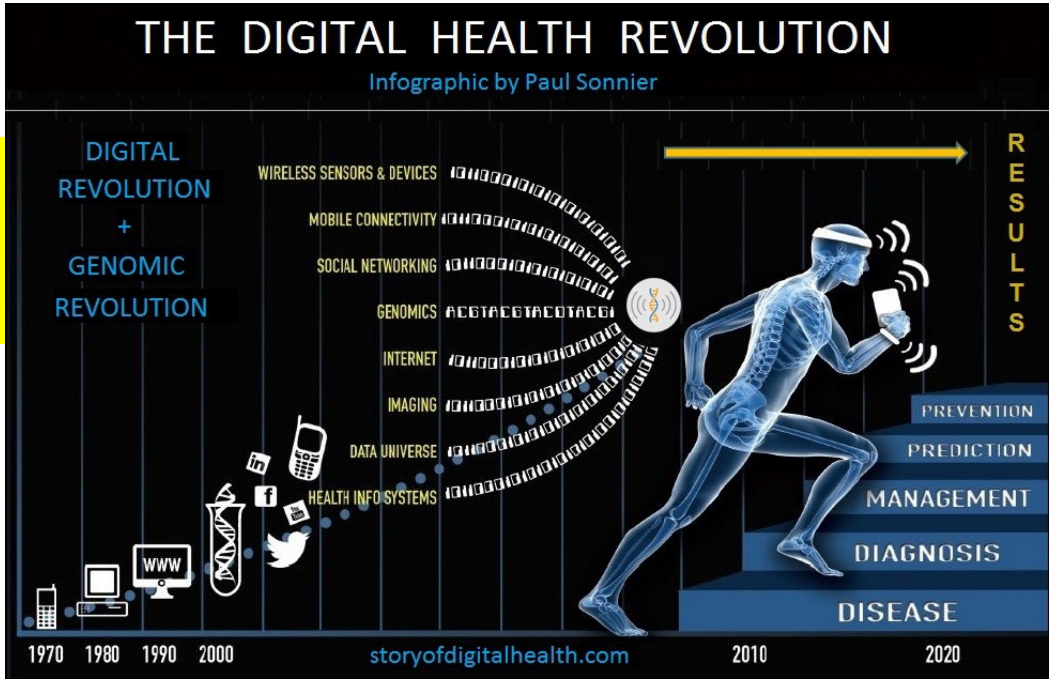
- Integrate heterogeneous data
- Standardize processes for credible, reusable models
- **Next Gen Models** to predict and provide insight

THE FUTURE OF HEALTH

The Future of Health



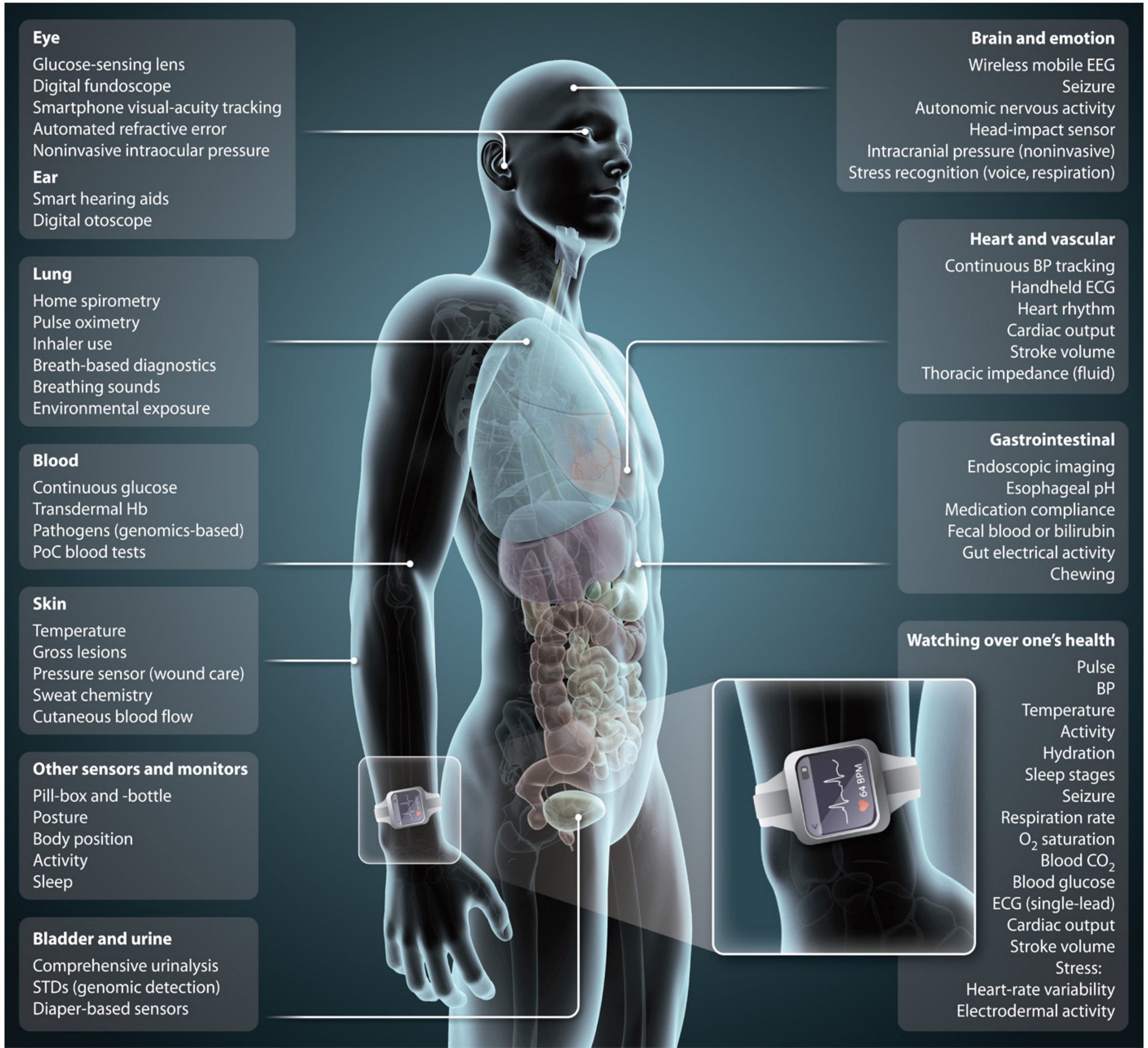
Future Medicine:
Dynamic data,
Digital Health

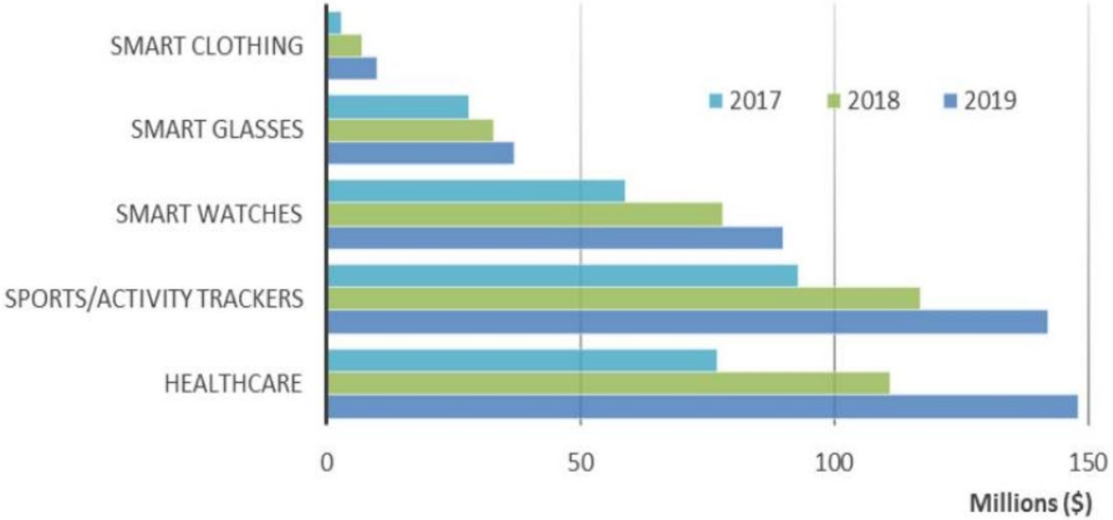


Time

Mobile Health & Sensor Technology

Wearable Health Devices – Vital Sign Monitoring, Systems, and Technologies

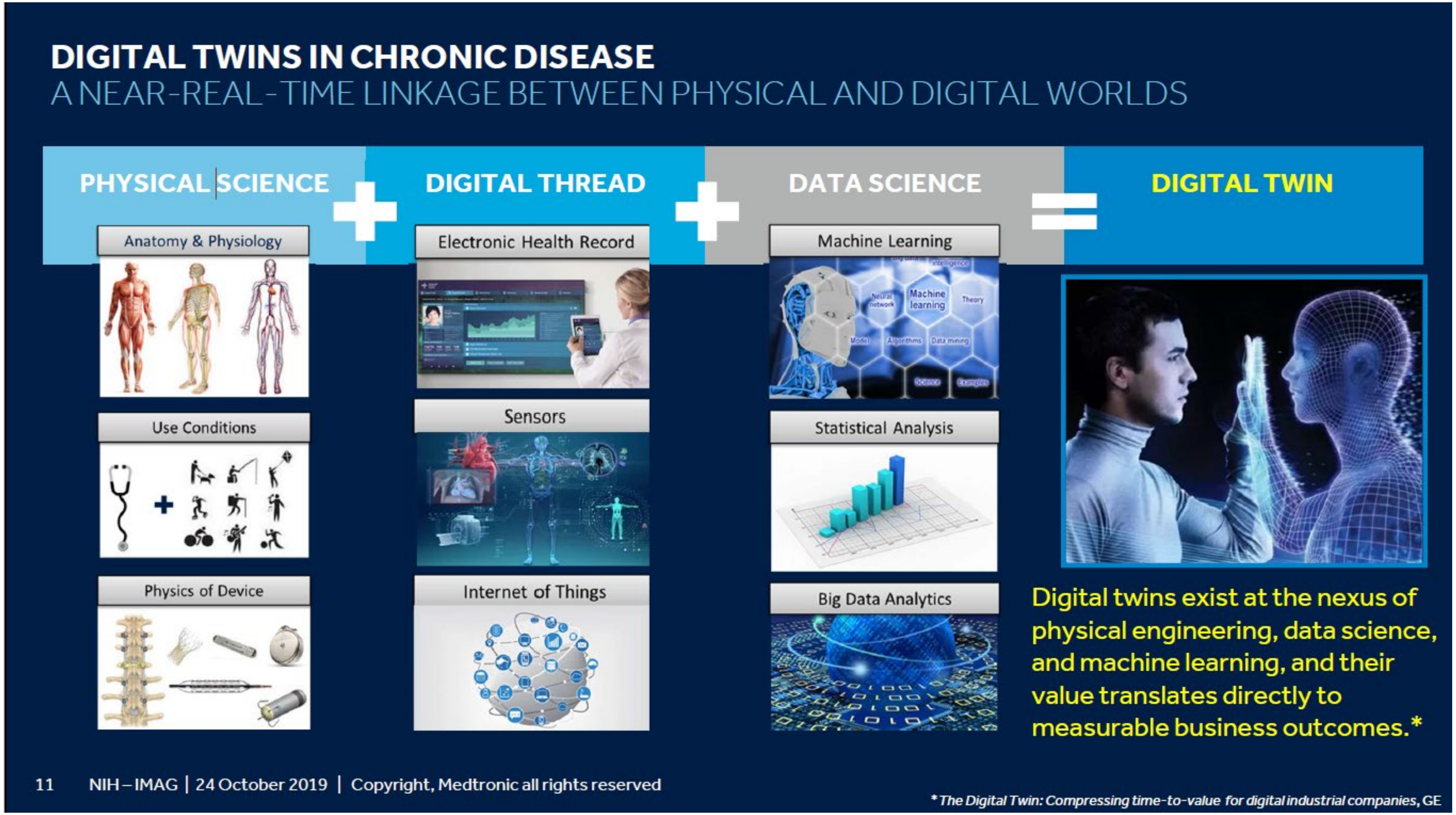




Trend of global market value of wearable computing devices, in millions, between 2017 and 2019.

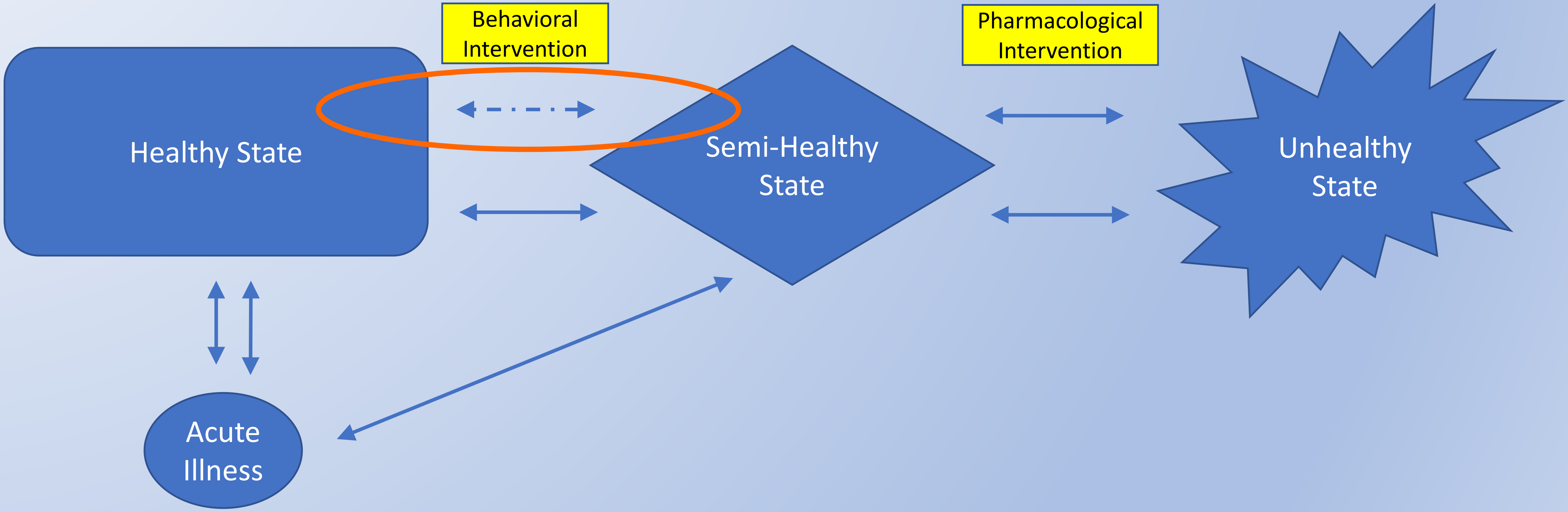
Duarte Dias et al. Sensors (Basel). 2018 Aug; 18(8): 2414.

Digital Twins



Curtesy of MARK PALMER, MD, PHD
Distinguished Scientist, Strategic Scientific Operations, Medtronic

Salutogenesis

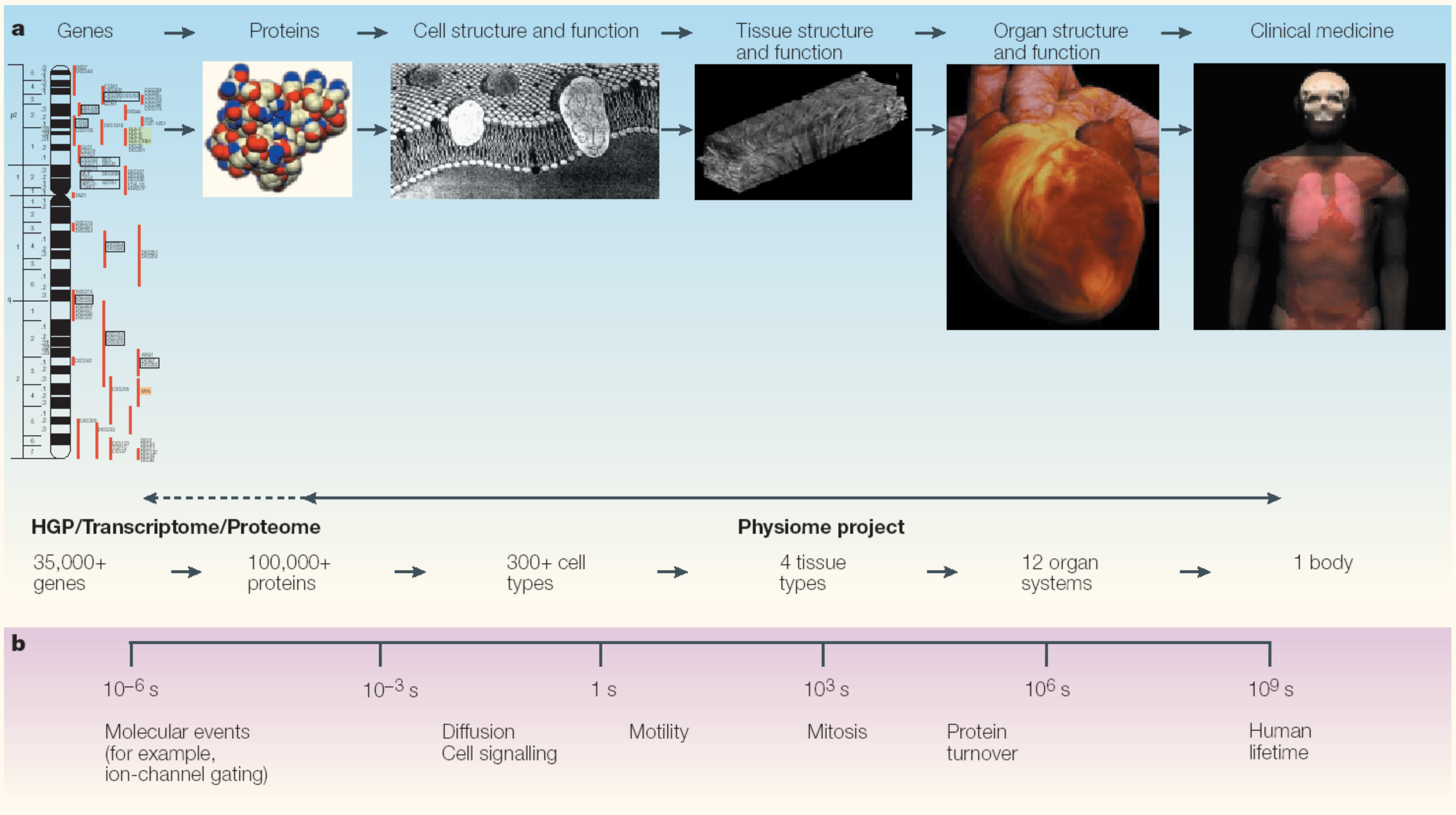


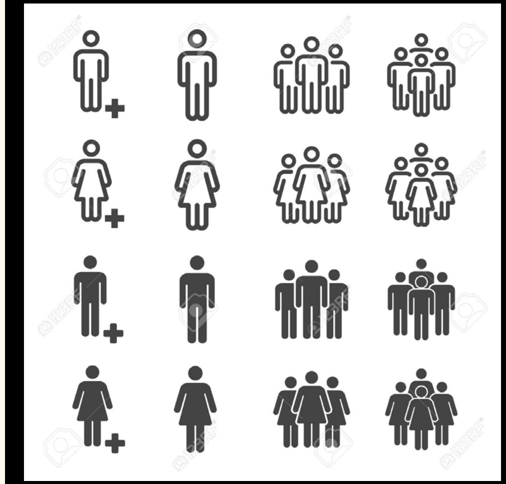
PREDICTIVE MODELING

IMAG

Interagency Modeling and Analysis Group

Biological Spatiotemporal Scales



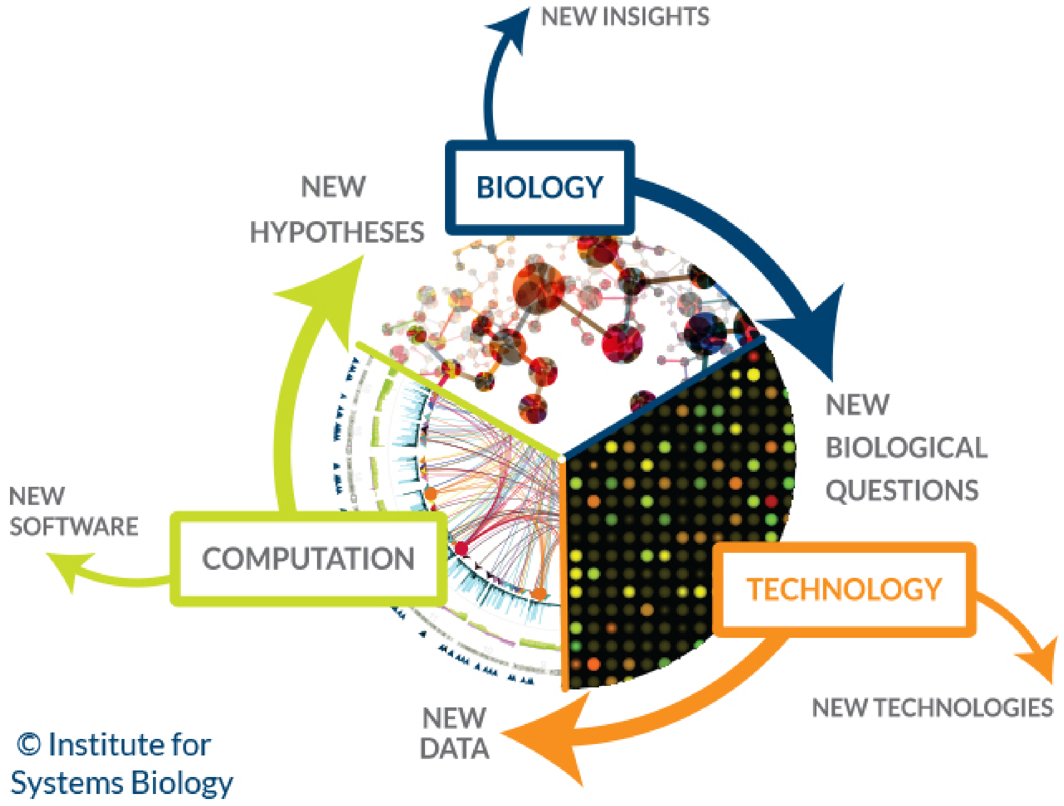


→ Individuals to Populations

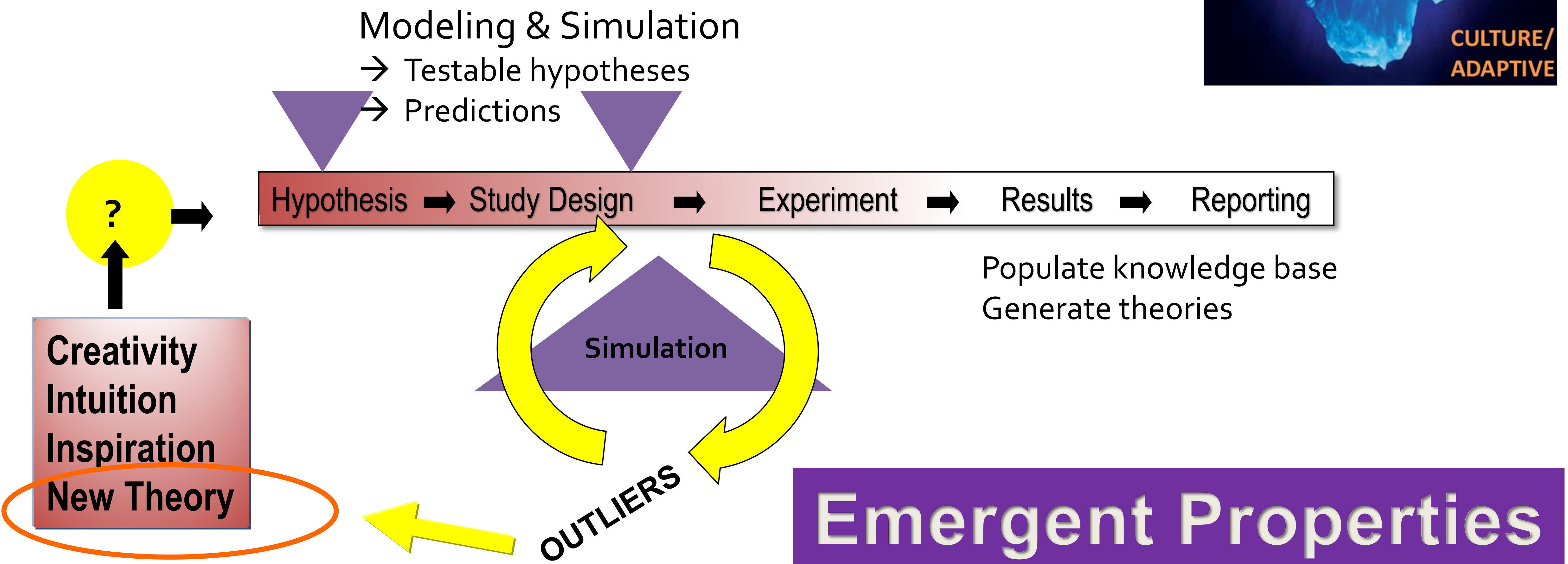
Hunter and Borg, Nature 2003

Why Model?

- Infrastructure for systematically archiving and transferring knowledge
  - Prevent reinventing processes year after year
- Only way to predict outcomes not otherwise testable
- Drive scientific discovery
  → emergent properties
- Extend insight and understanding beyond the cognitive capabilities of the human mind

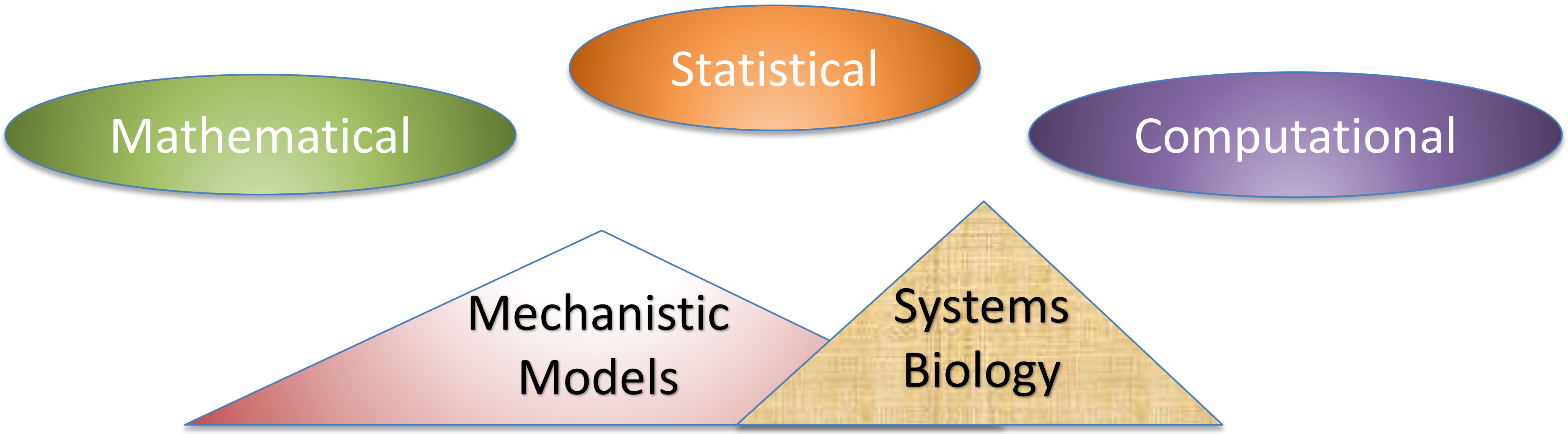


Model-Driven Science

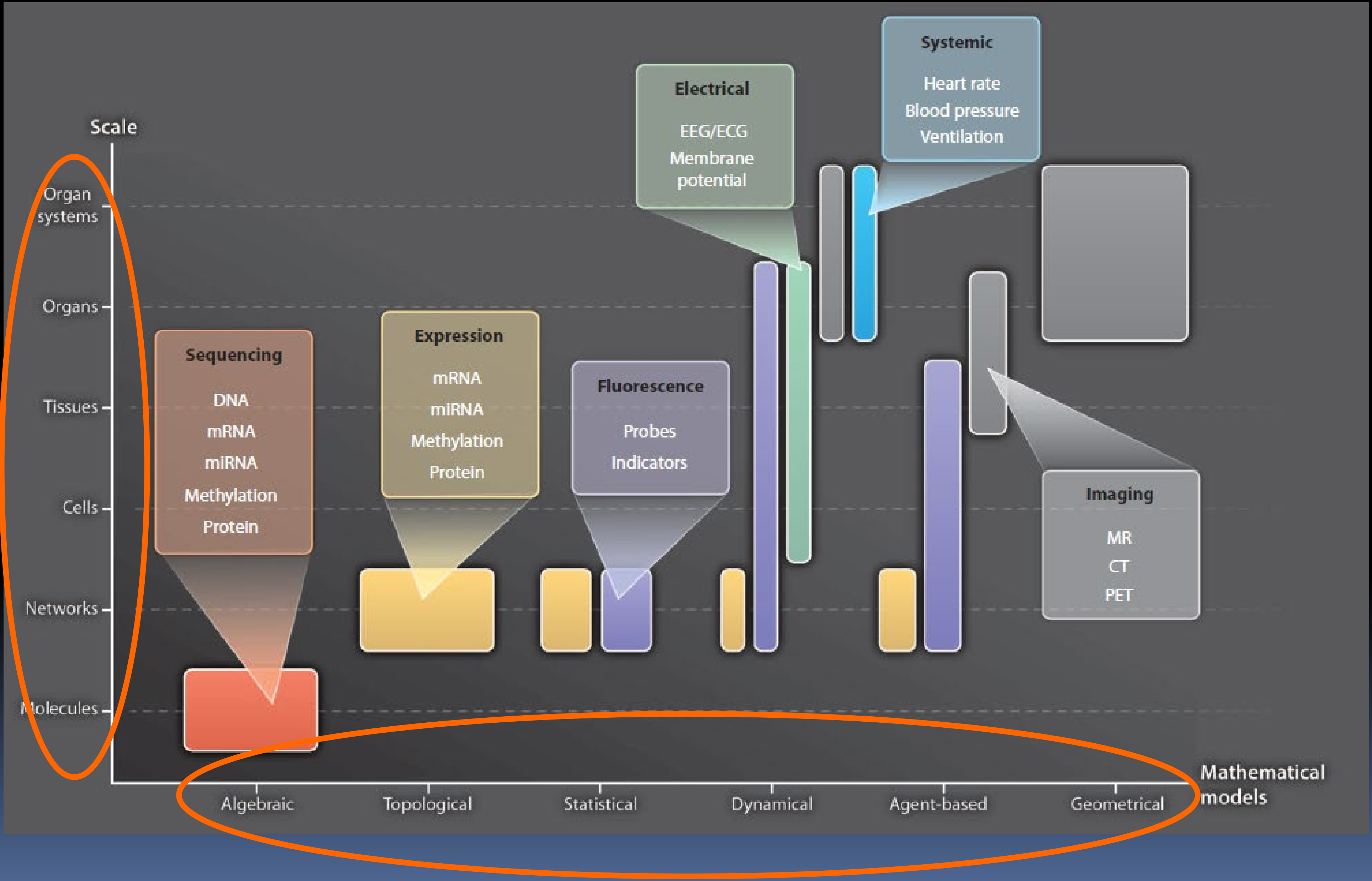


Predictive Modeling

Predictive models generate new hypotheses, and do not merely recapitulate the data that were used to build them.
[2003-present: IMAG Multiscale Modeling Initiative]



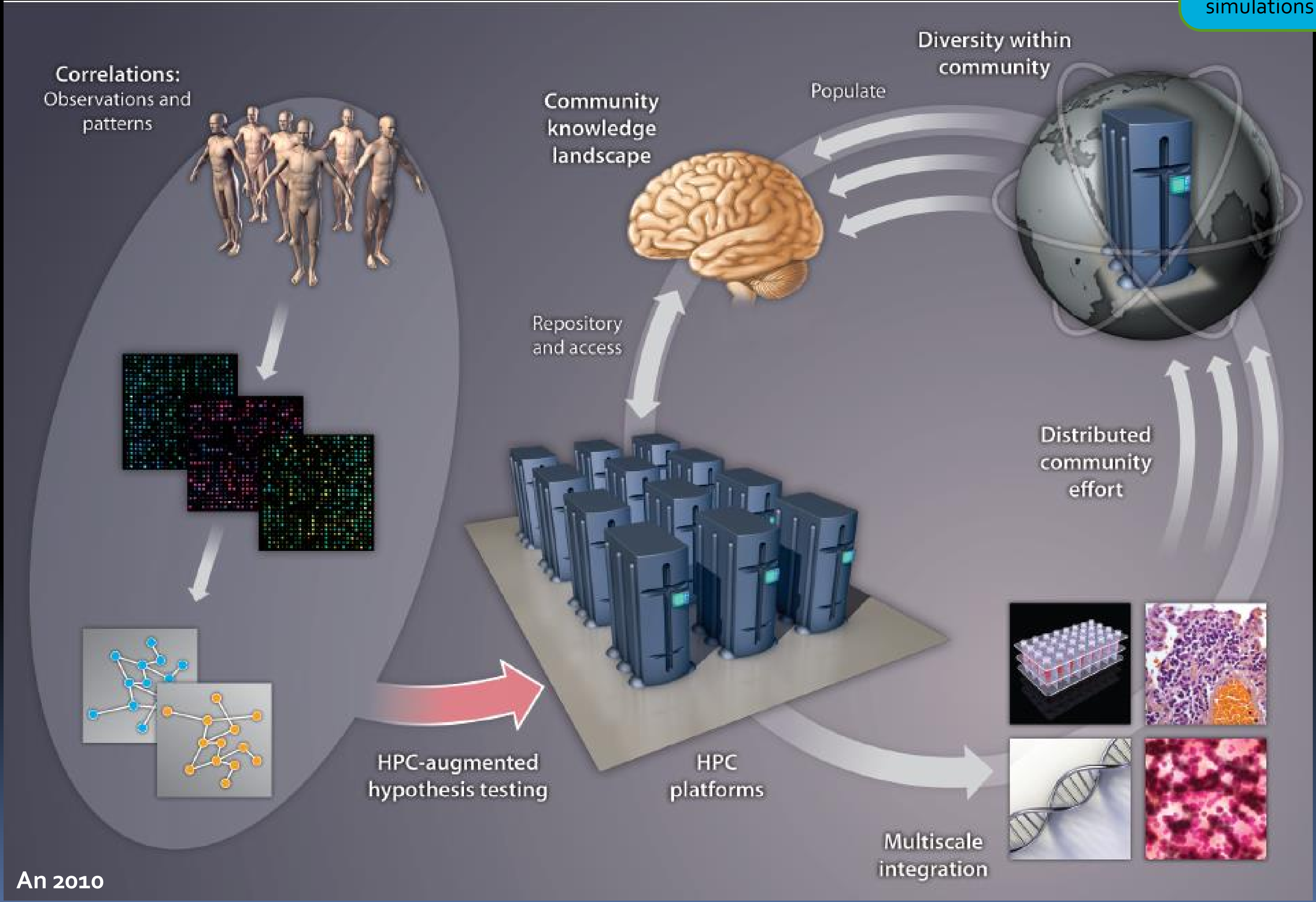
Quantitative Measures



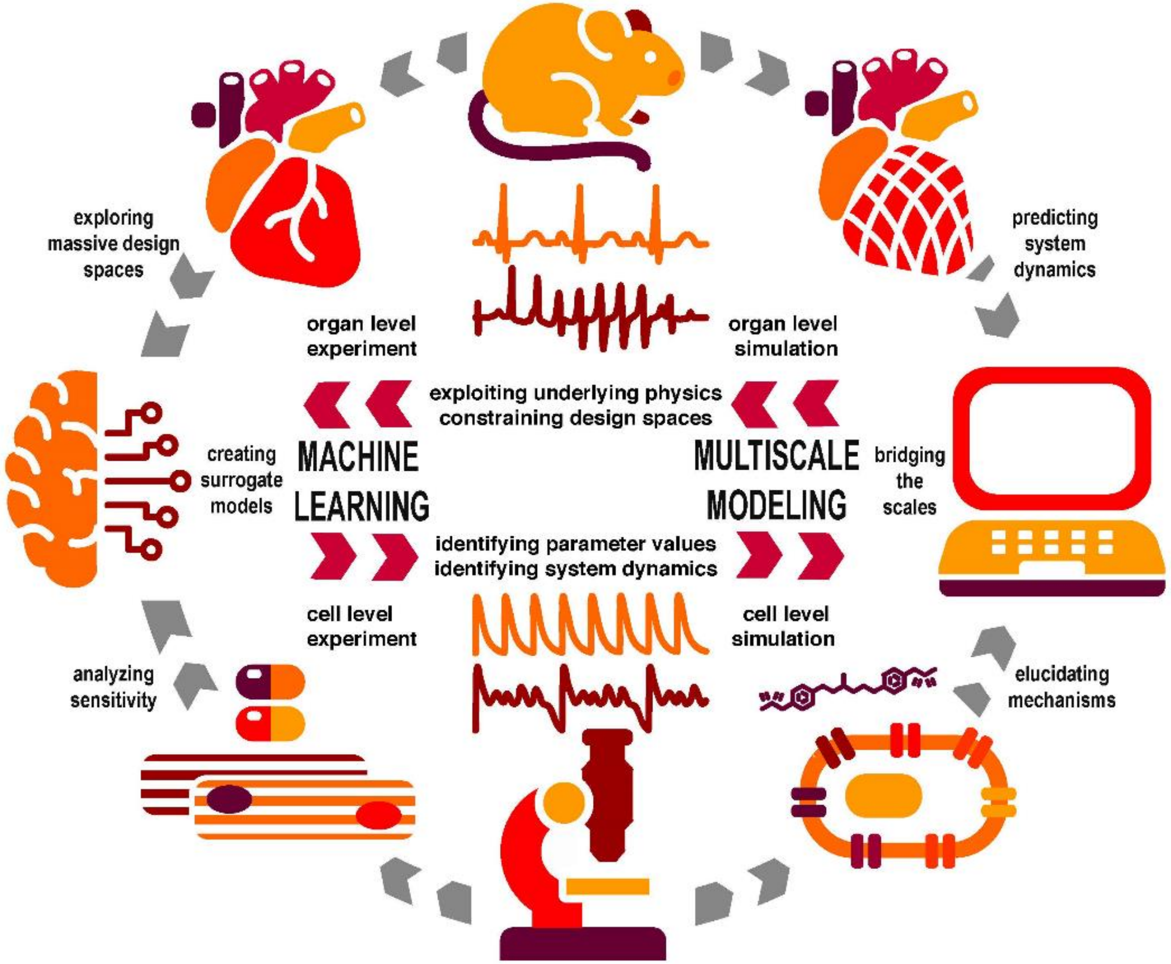
Winslow et al., 2012

High Performance Computing – the AI wave

Mechanistic Models to drive **better** simulations

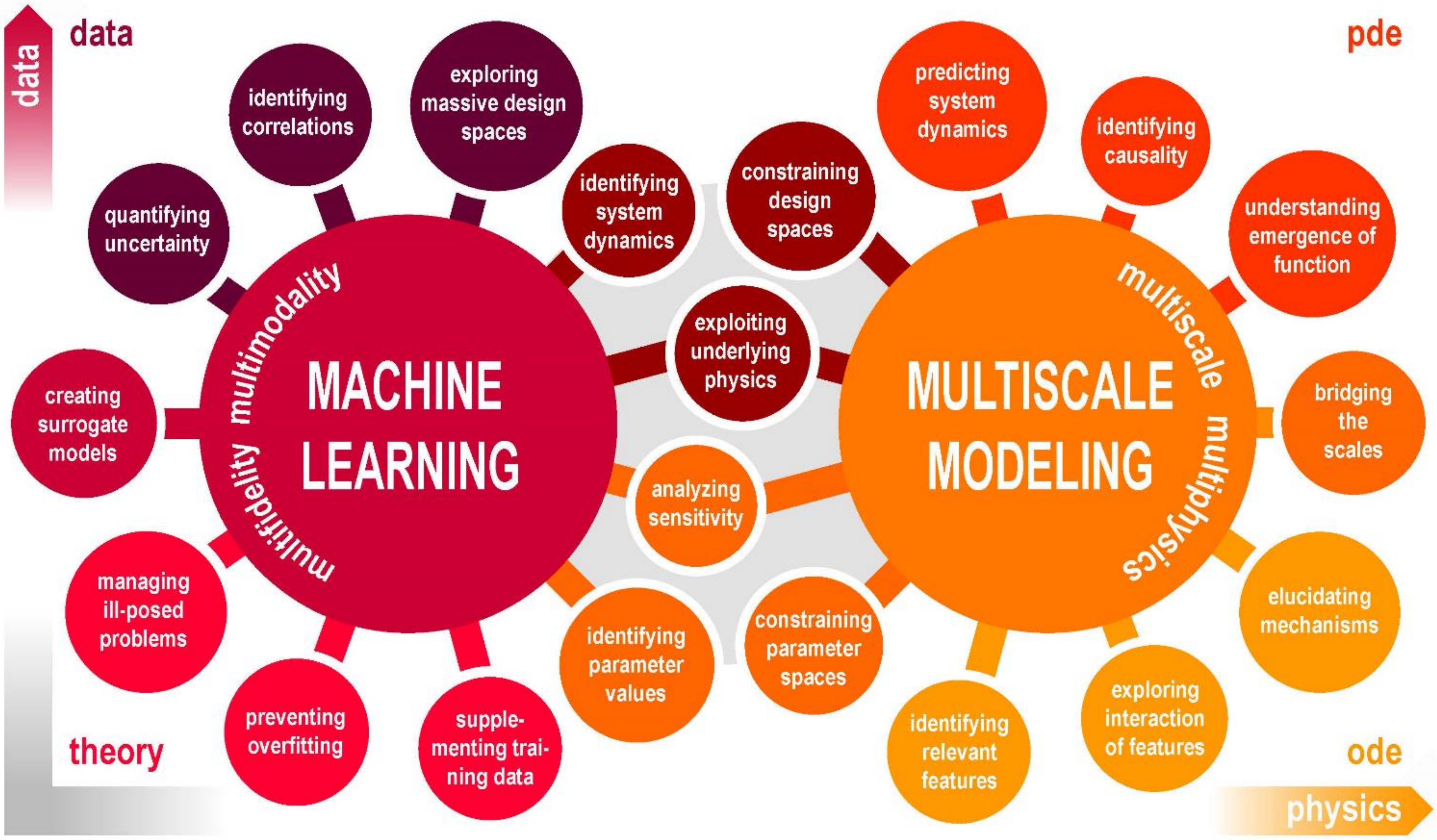


Alber, M., Buganza Tepole, A., Cannon, W.R. *et al.** **Integrating machine learning and multiscale modeling—perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences**. *npj Digit. Med.* **2,** 115 (2019). https://doi.org/10.1038/s41746-019-0193-y



*Ellen Kuhl

Peng, G.C.Y., Alber, M., Buganza Tepole, A. *et al.** **Multiscale Modeling Meets Machine Learning: What Can We Learn?**. *Arch Computat Methods Eng* (2020). https://doi.org/10.1007/s11831-020-09405-5



npj Digital Medicine (2019) 2:115
https://doi.org/10.1038/s41746-019-0193-y

SimTK

Search for projects

Projects About Sign Up Log In

Credible Practice of Modeling & Simulation in Healthcare

Follow (5)

Join (18)

About Downloads Documents Forums Wiki Source Code News

Wiki

Titles Text

RecentChanges F

Immutable Page Inf

Just Published!

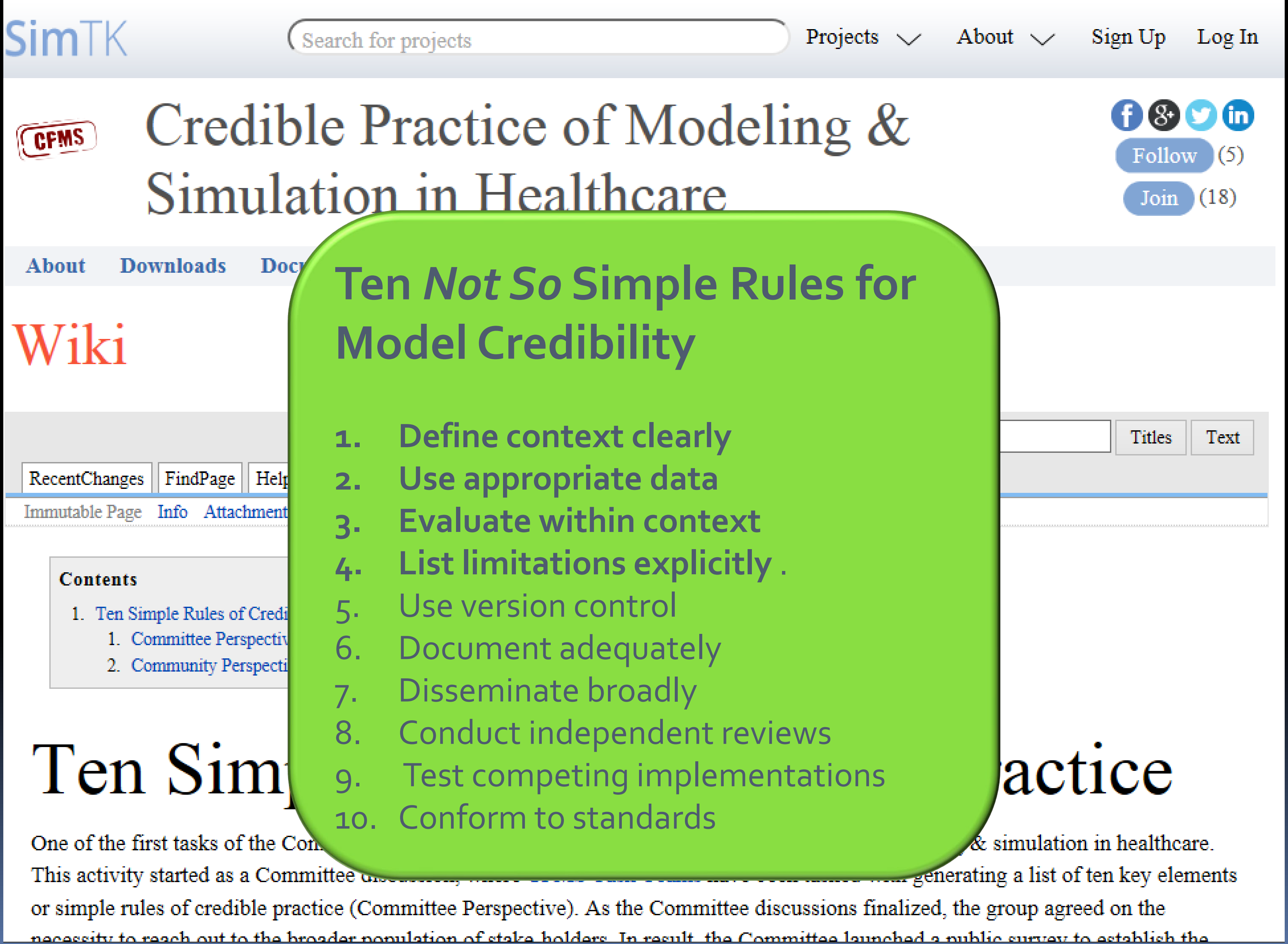
Erdemir, A., Mulugeta, L., Ku, J.P. *et al.* **Credible practice of modeling and simulation in healthcare: ten rules from a multidisciplinary perspective**. *J Transl Med* **18,** 369 (2020). https://doi.org/10.1186/s12967-020-02540-4

Contents

1. Ten Simple Rules of Credible Practice
   1. Committee Perspective
   2. Community Perspective

Ten Simple Rules of Credible Practice

One of the first tasks of the Committee was to identify best practices to enhance credibility of modeling & simulation in healthcare. This activity started as a Committee discussion, where CPMS Task Teams have been tasked with generating a list of ten key elements or simple rules of credible practice (Committee Perspective). As the Committee discussions finalized, the group agreed on the necessity to reach out to the broader population of stake holders. In result, the Committee launched a public survey to establish the



Ten *Not So* Simple Rules for Model Credibility

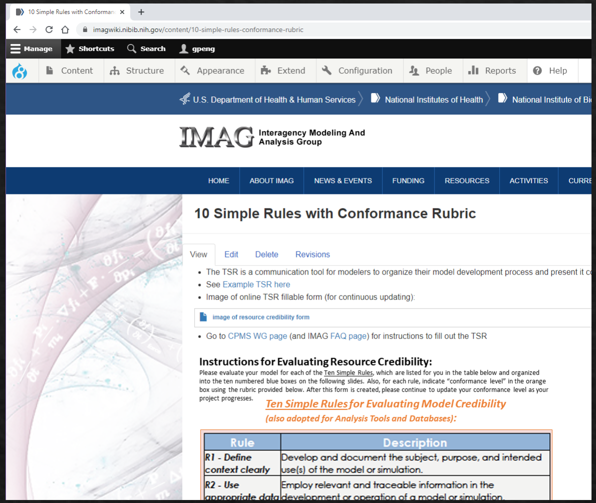
1. **Define context clearly**
2. **Use appropriate data**
3. **Evaluate within context**
4. **List limitations explicitly .**
5. Use version control
6. Document adequately
7. Disseminate broadly
8. Conduct independent reviews
9. Test competing implementations
10. Conform to standards

Now Applied to COVID-19

MULTISCALE MODELING CONSORTIUM

Ten Simple Rules for Model Credibility

1. Define context clearly
2. Use appropriate data
3. Evaluate within context
4. List limitations explicitly
5. Use version control
6. Document adequately
7. Disseminate broadly
8. Conduct independent reviews
9. Test competing implementations
10. Conform to standards



DARPA

NASA

CDC

NIH

IMAG

MSM Task Forces

MULTISCALE MODELING CONSORTIUM
IMAG

Task Force on Methodologies

- Cell-to-Macroscale Working Group
- High Performance Computing Working Group
- Multiscale Systems Biology Working Group
- Theoretical and Computational Methods
- Population Modeling Working Group

Task Force on Basic Science Applications

- Biomechanics Working Group
- Computational Neuroscience Working Group
- Integrated multiscale biomaterials experiment and modeling group (ImuBEAM)

Task Force on Dissemination

- Committee on Credible Practice of Modeling & Simulation in Healthcare Description
- Model and Data Sharing Working Group
- Public Dissemination and Education

Task Force on Clinical Translation

- Clinical and Translational Issues → Multiscale Modeling and Viral Pandemics
- MSM for Medical Devices

Greater than the sum of its parts

IMAG

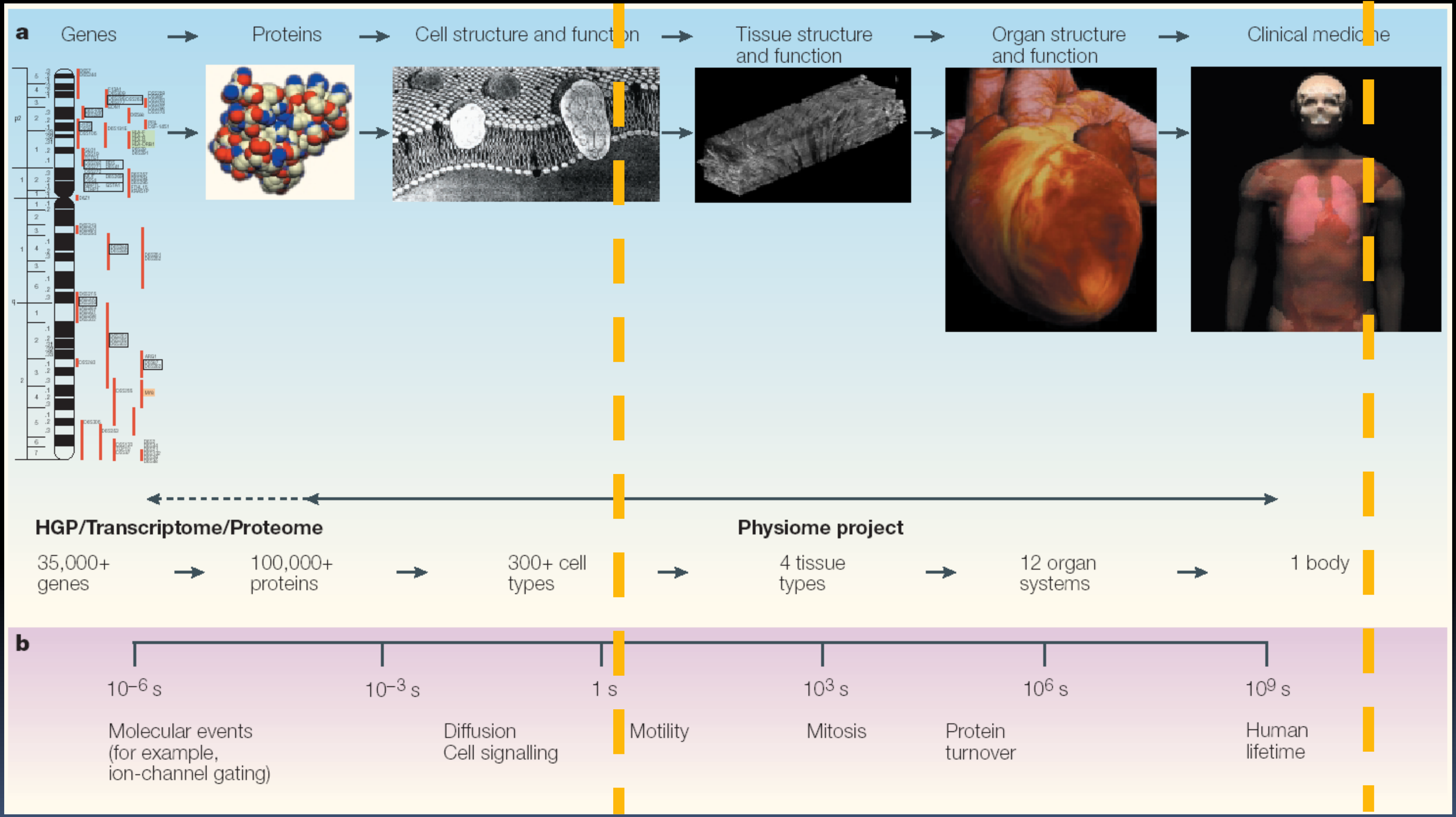
MULTISCALE MODELING CONSORTIUM

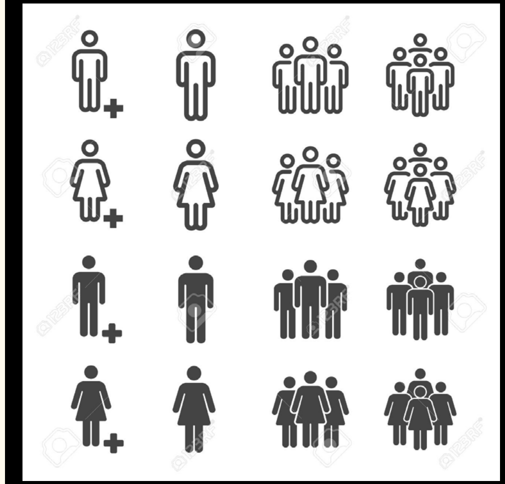
Interagency Modeling and Analysis Group (IMAG) Wiki

(Search: IMAG Wiki)

BIOMEDICAL AND BEHAVIORAL RESEARCH CULTURE

Biological Spatiotemporal Scales





→ Individuals to Populations

Hunter and Borg, Nature 2003

Microscale **Mesoscale** **Macroscale**

VIDEO: Understanding Back Pain

Robert H. Carter, MD, talks about the NIH HEAL Initiative's BACPAC program.

Tweet This

https://youtu.be/lmhuuQtliBI

heal.nih.gov/news/stories/bacpac-low-back-pain

U.S. Department of Health & Human Services | National Institutes of Health

NIH National Institutes of Health
HEAL Initiative

About | Research | Funding | News & Events | Resources | Search | Search

Piecing Together the Puzzle of Chronic Low Back Pain

A computer model may be able to inspire new insights and treatments.

Home / News & Events / Research Spotlights / Piecing Together the Puzzle of Chronic Low Back Pain

The human back is a complex structure with bones, nerves, tendons, discs, and more — all places where something can go wrong and cause pain, which, for many people, becomes a long-term or chronic problem. Life stresses and other medical and mental health conditions aggravate the problem.

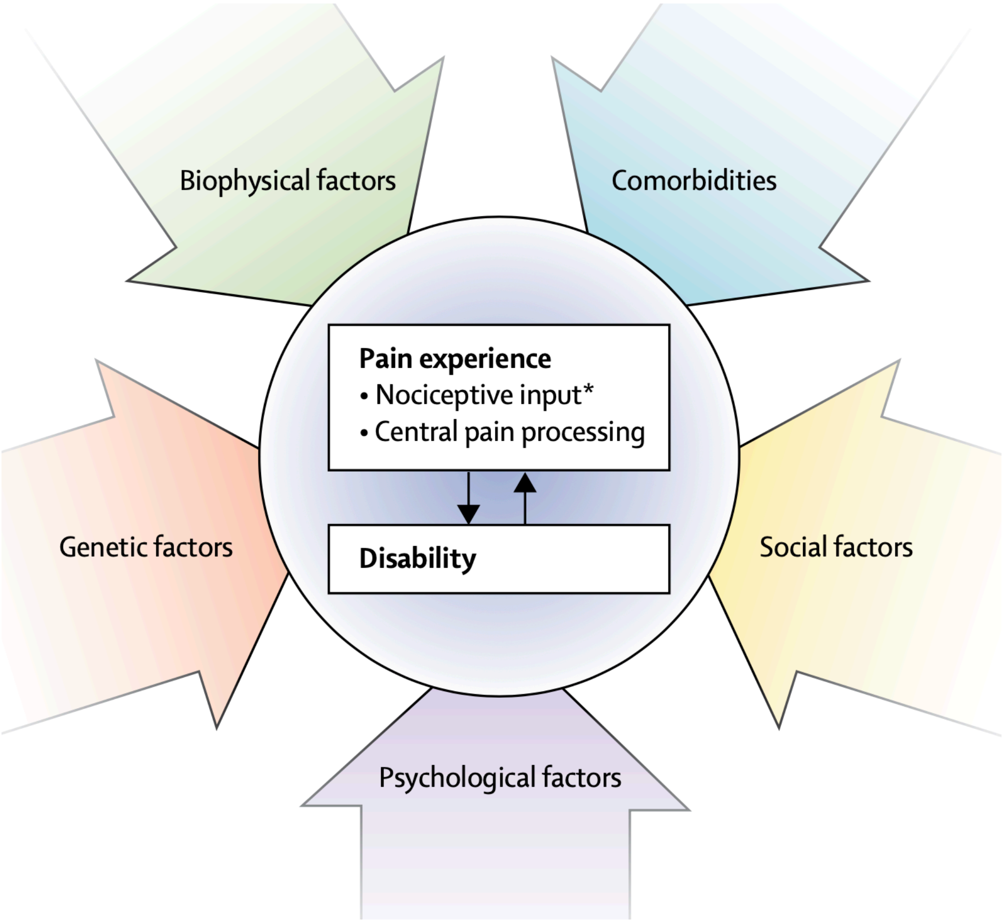
With so many pieces, it's hard to get a holistic view of the puzzle or pinpoint the cause of the pain.

"People tend to focus on one aspect or another," said Jeffrey Lotz, Ph.D., a medical engineer who studies back pain at the University of California, San Francisco. "Some people think it's largely in the mind; some people think

Models, such as this plastic replica of the spine, are representations

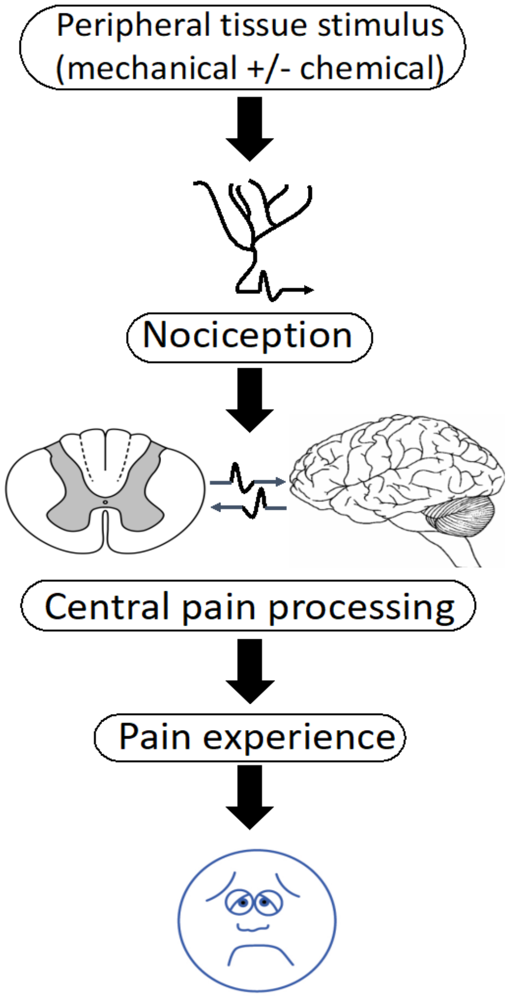
Lower Back Pain

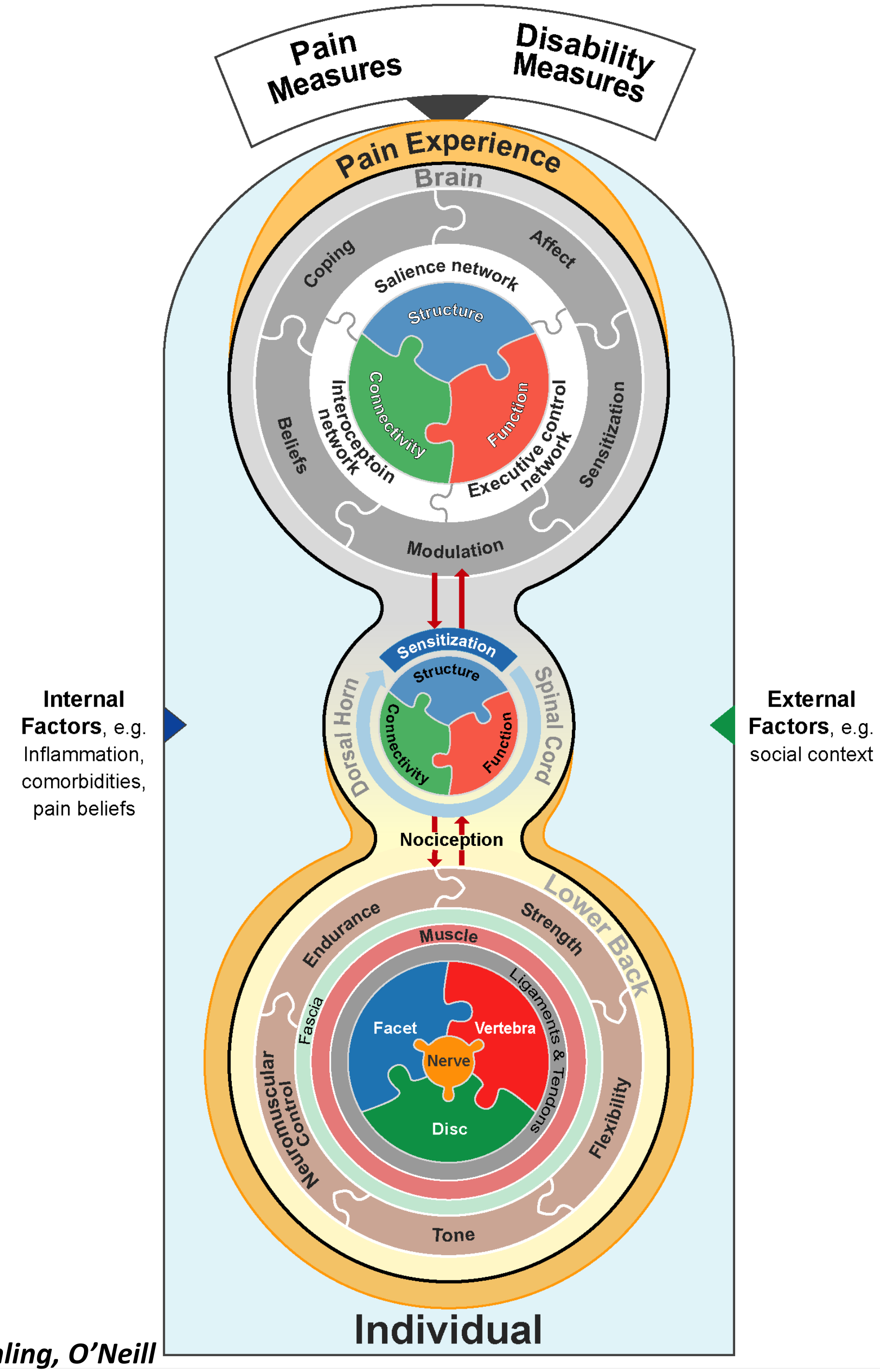
Built on the Foundation of the Biopsychosocial Concept



Hartvigsen et al., 2018

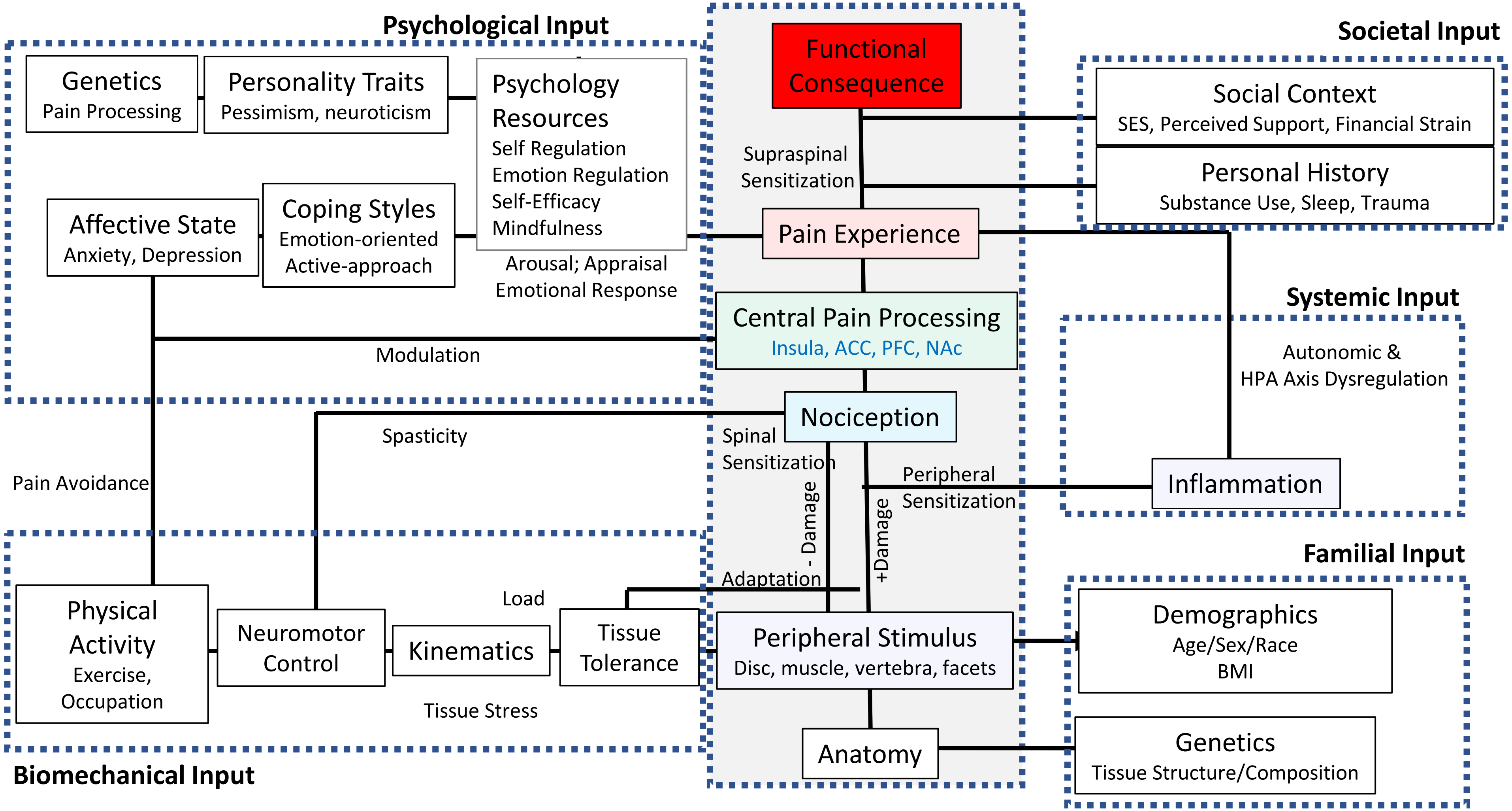
BACPAC
Back Pain Consortium





Hunt, Mehling, O'Neill

BACPAC Theoretical Model WG



BACPAC
Back Pain Consortium

BACPAC Research Agenda – Goal 1
State-of-the-Art Model for Chronic Low Back Pain

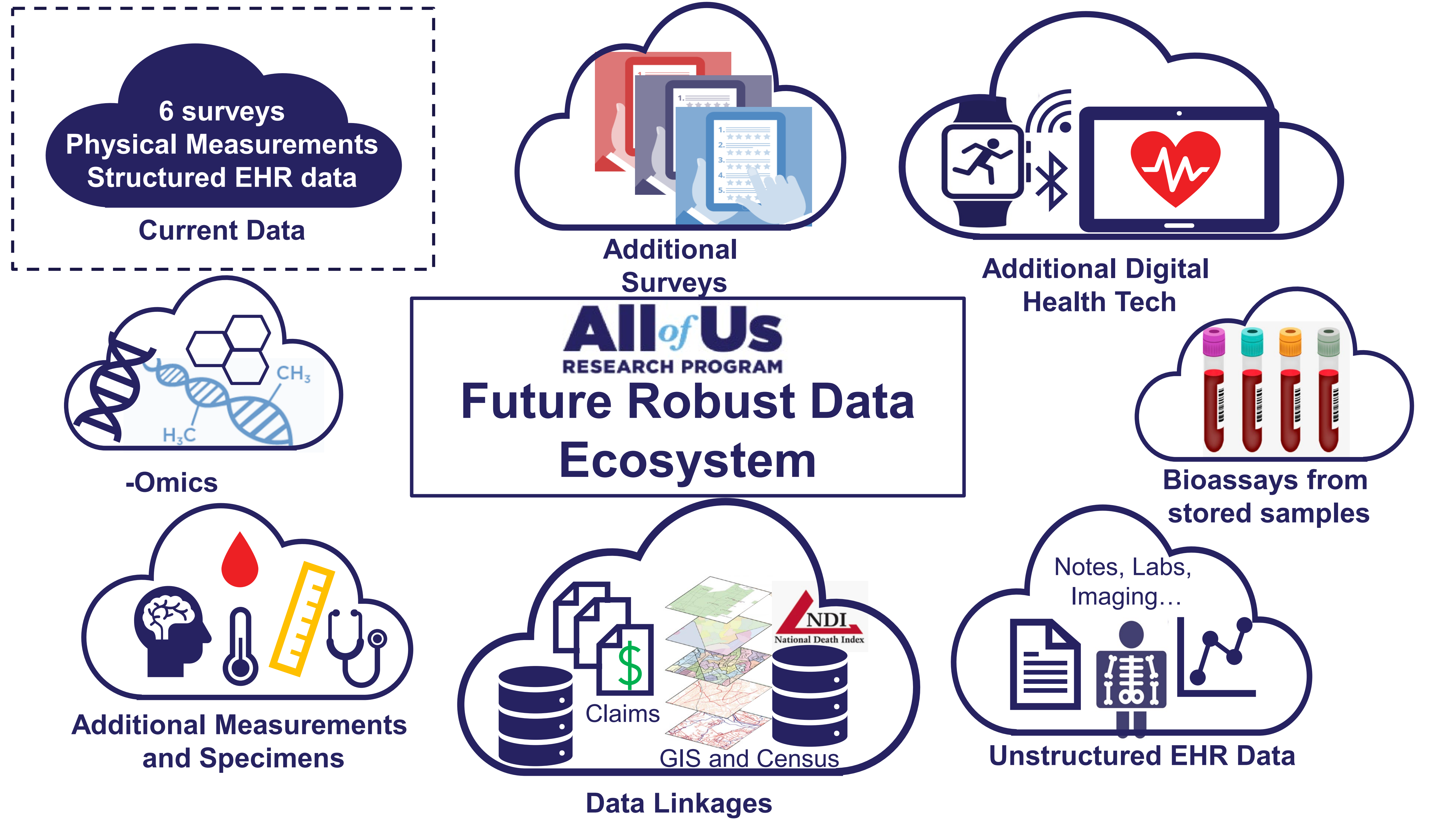
- Develop a theoretical model for chronic low back pain

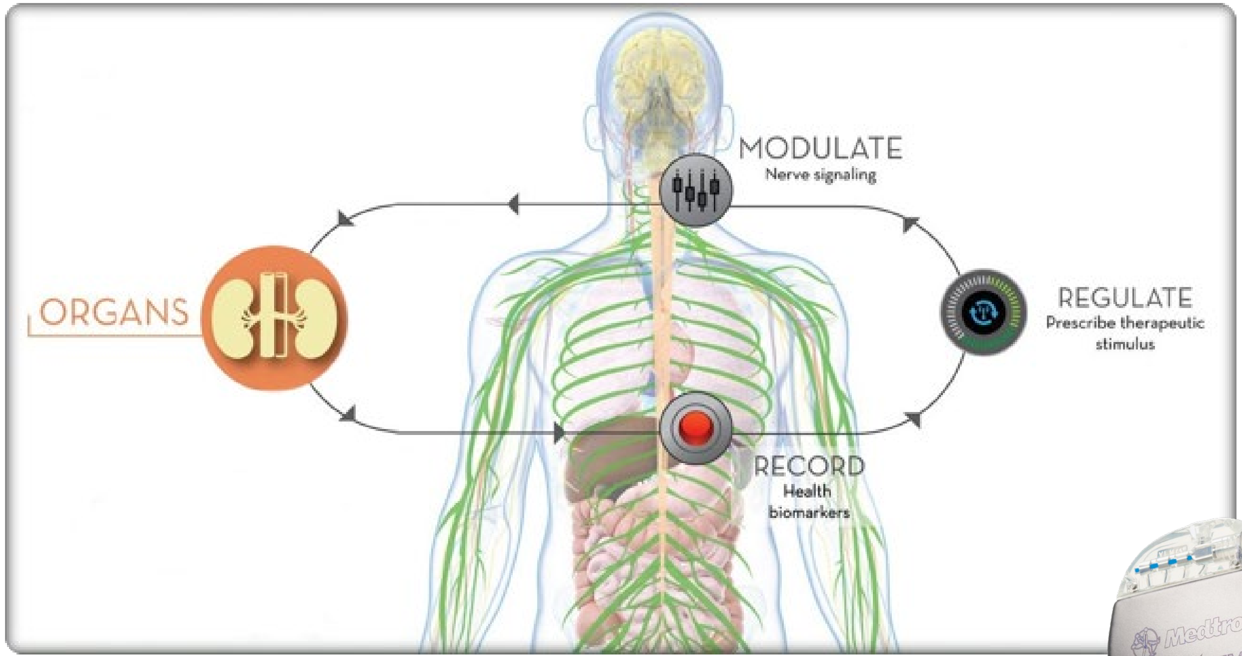
BACPAC Research Agenda – Goal 2
Identify factors that are predictive of treatment effectiveness for well-defined patient subpopulations

- Develop Testable Hypotheses

BACPAC Research Agenda – Goal 3
Develop an algorithm for multi-modal interventions for individuals with different phenotypes of chronic low back pain

- Design and conduct a large-scale adaptive cLBP trial that tests multiple bundled or sequential interventions


6 surveys
Physical Measurements
Structured EHR data
Current Data
Additional Surveys
Additional Digital Health Tech
All of Us
RESEARCH PROGRAM
Future Robust Data Ecosystem
CH3
H3C
-Omics
Bioassays from stored samples
Notes, Labs, Imaging…
Unstructured EHR Data
Additional Measurements and Specimens
Claims
NDI
National Death Index
GIS and Census
Data Linkages



Inputs

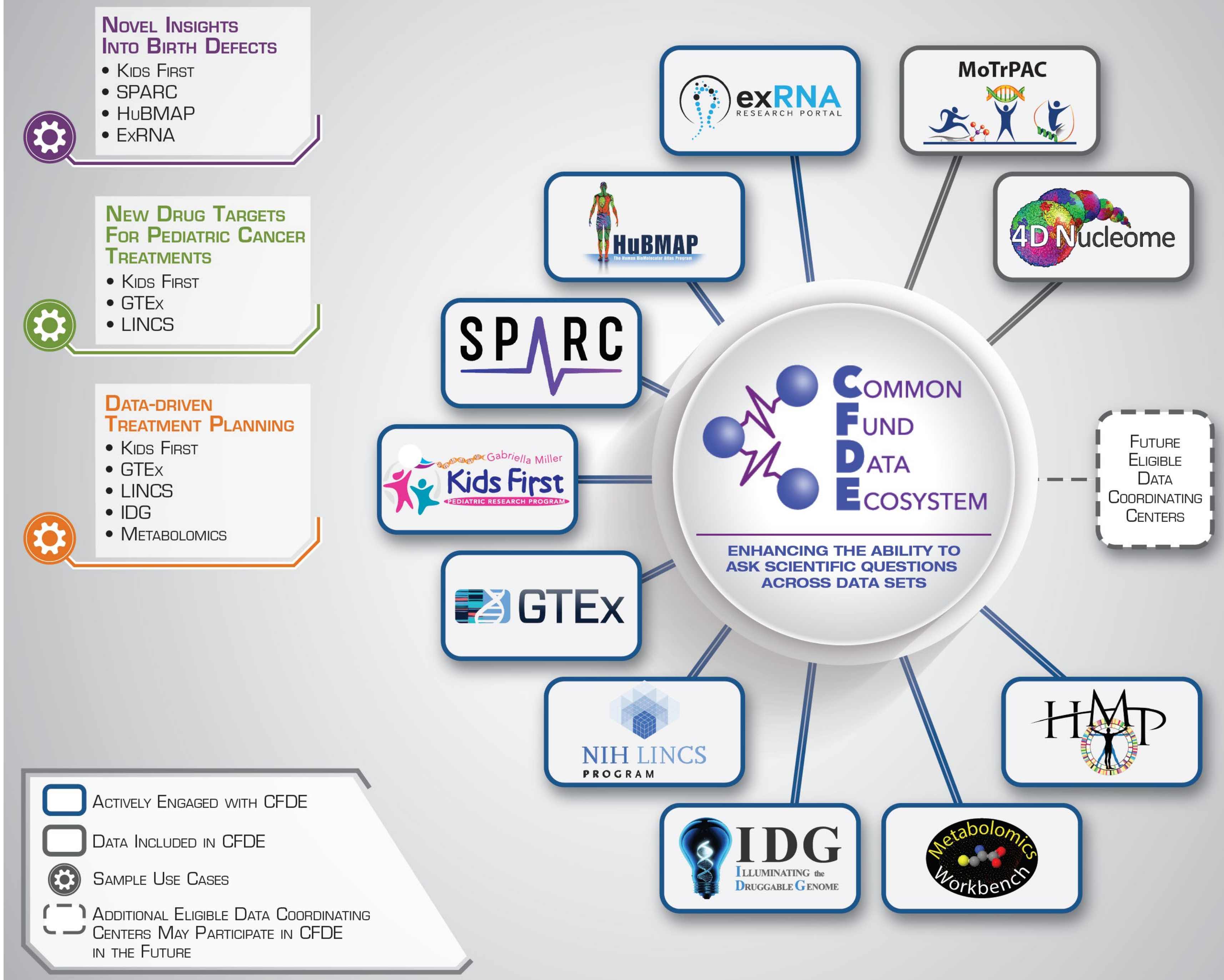
- Stimulus design
  - Frequency
  - Amplitude
  - Duration
  - Duty Cycle
- Electrode design and placement
- Animal model
- Disease model/phenotype
- Physiological recordings (for informing closed-loop stimulation)

Outputs

- Change in organ function (on- and off-target effects)
- Biomarkers
- Change in electrode interface (e.g. impedance)
- Plasticity/adaptation

CFDE

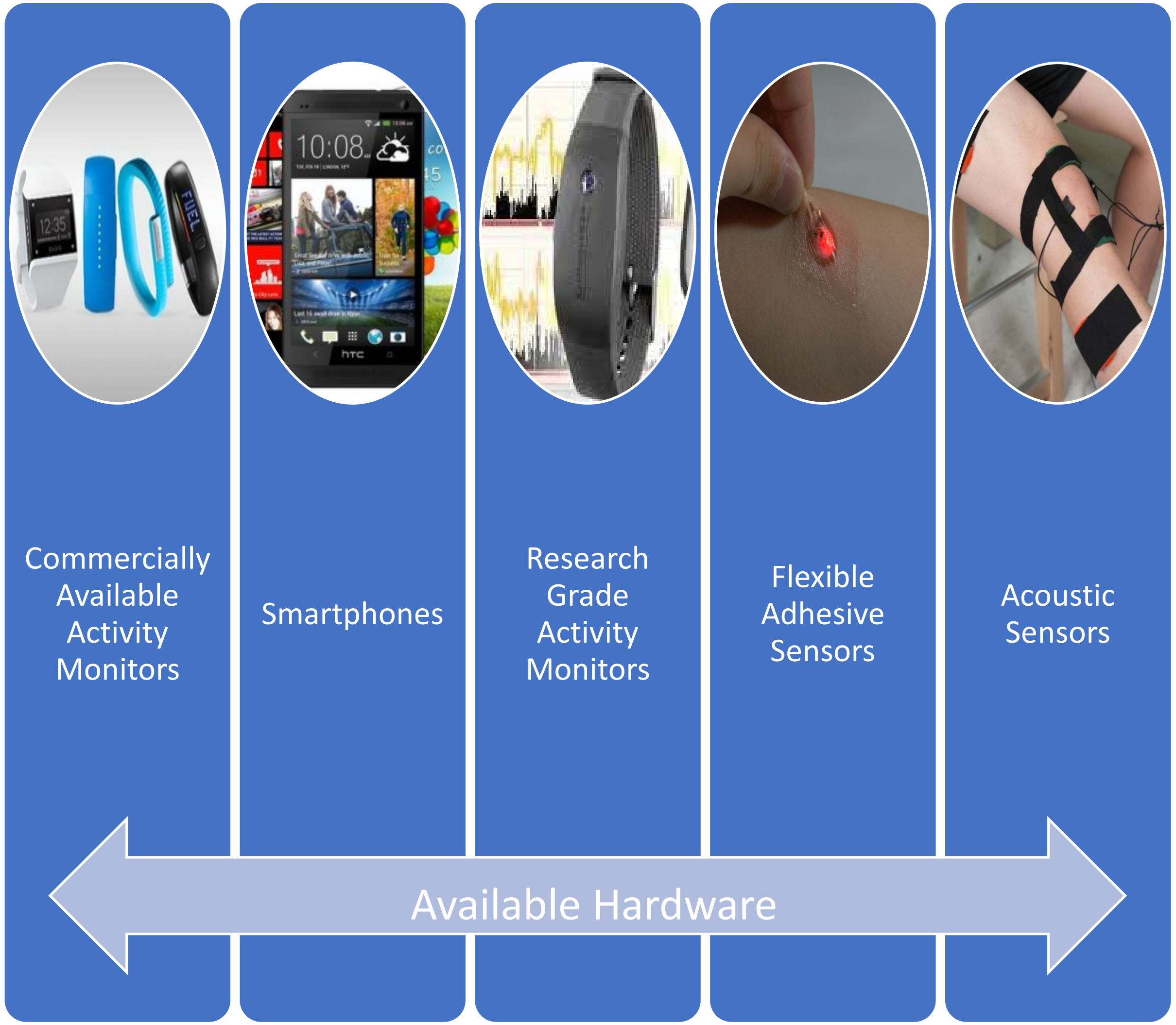
- Making data from Common Fund programs FAIR



Experiment types		GTEx	HMP/iHMP	LINCS	Metabolomics	SPARC	4D Nucleome	MoTrPAC	Kids First
Genomics	Whole genome/exome sequencing	●	●						●
	Omni SNP Arrays	●							
	16S metagenomic sequence		●						
	Epigenomics		●	●			●	●	
Transcriptomics	RNA/miRNA/mRNA-Seq	●	●	●			●	●	●
	Array or L1000 platform based	●		●					
Metabolomics	Mass Spectrometry (GCMS, LCMS)		●					●	
	NMR				●				
	Lipidomics		●		●			●	
Proteomics	Mass Spec based		●	●	●			●	
	Microarray based			●					
Imaging	Fluorescence based			●		●	●		
	Histology/Cytology	●				●			●
	Radiology (MRI)					●			●
In Vitro	Binding Assays			●					
	Immuno Assays			●		●	●		
In/ex vivo	Neural stimulation/neural activity recording					●			
	High resolution manometry					●			
Code	Analysis Workflow					●			

Hardware Exists, Research Methodologies Lag

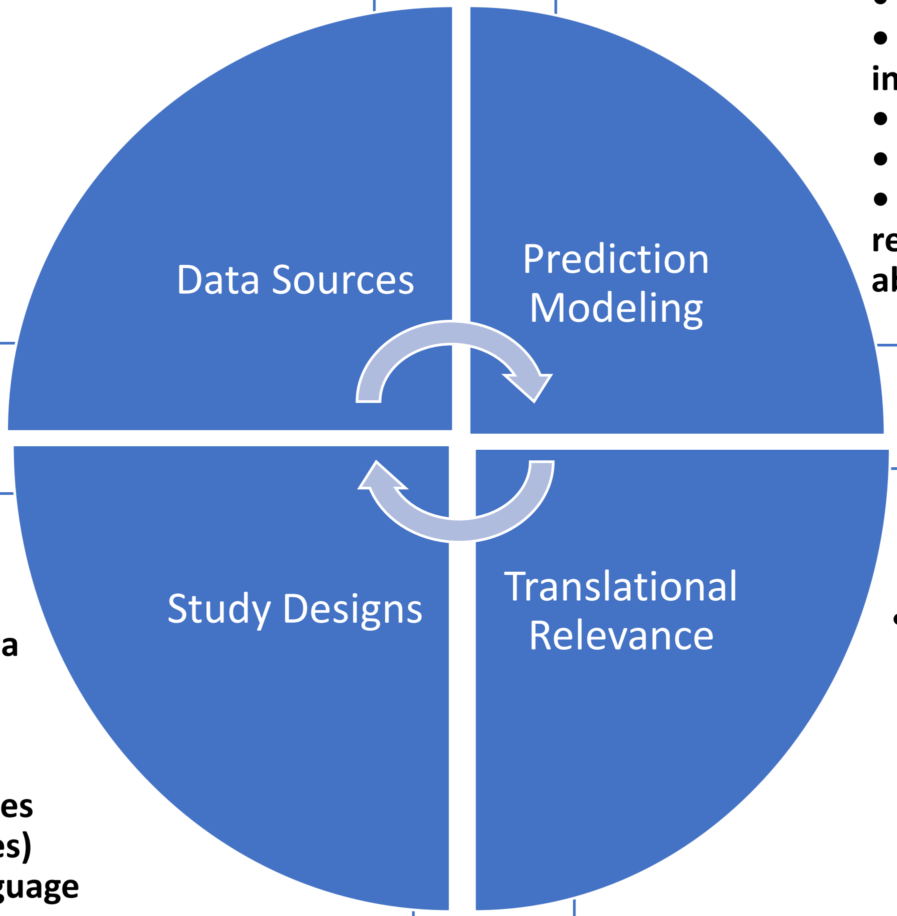
- Quantified self and health monitoring technologies are evolving rapidly; data is often proprietary
- Challenges
  - Methods for dealing with data size and complexity
  - Validation of models for application in clinical research and trials
  - Multidimensional interpretation of data
  - Ethics of use, provenance of data, and regulatory strategies



Biomedical and Behavioral Research Challenges

- Heterogeneous data sources
- Unstructured data sources
- Lack of data on social determinants of health
- Measurement issues (e.g. incompleteness, inaccuracy, imprecision in self-reported data)
- Privacy and security
- Cost
- Limited adoption of common data models

- Biases of all sorts (e.g. protopathic)
- Confounding
- Causal inference vs. mechanistic
- Black-boxes vs. white-boxes (i.e. interpretability vs. performance)
- Complexity-based model selection
- Benchmark development
- Pragmatic interoperability (reproducibility, replicability, generalizability)- different labs able to get the same result



- Semantic data integration (i.e. linking data elements by their meaning)
- Large longitudinal cohorts
- Ontology integration
- Ontology appropriateness (e.g. ontologies made for billing vs. for diagnostic purposes)
- Semantic interoperability- common language
- Automated study design

- Limited individual empowerment
- Disconnect from relevant clinical research
- Personal health record/health avatar (besides provider's electronic records)
- Acceptance of artificial intelligence as integral part of doctors' tools
- Learning systems
- Ethical usage and dissemination of modelling algorithms
- Redefining disease phenotype

from Prosperi et al. BMC Medical Informatics and Decision Making (2018)

HOW DO WE GET THERE?

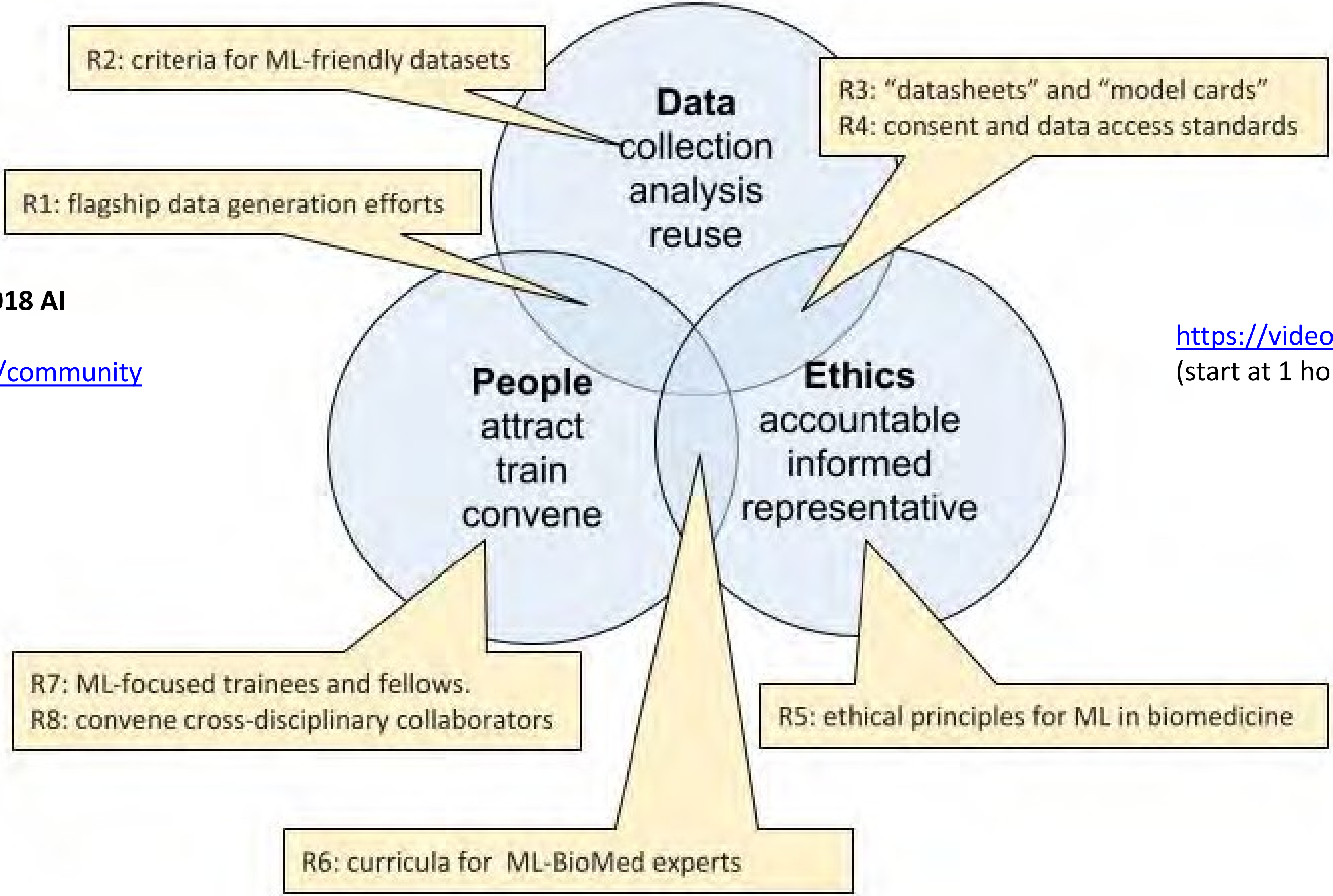
The NIH Artificial Intelligence Initiative through 2028

(acronym TBD)

→→ To Propel Progress in Biomedical Research through NEXT-GENERATION AI

NIH AI Working Group Recommendations (12/13/2019 ACD Report)

A follow-up from the **July 2018 AI Workshop**, https://datascience.nih.gov/community/2018biomedAI



https://videocast.nih.gov/watch=35426 (start at 1 hour, 2 minutes)

Parallel Revolutions:
Fusing Biomedicine and Machine Learning (12/13/2019 ACD Report)

Data Generation

more data about the biology and health of more individuals than ever before

Data Analysis

machine learning, other forms of artificial intelligence, cloud computing

ML-BioMed

- biomedical experiments* that are designed for ML
- ML that's designed for biomedical experiments*

*Note: Biomedical experiments include **biological** and **behavioral** studies

(1) Support flagship data generation efforts to propel progress by the scientific community.

Support flagship efforts that generate large-scale experimental data, with billions of data points designed to:

i. be well-suited for ML analysis and inference
ii. address key biomedical challenges
iii. stimulate new approaches in machine learning

And that implement processes designed to:

i. develop improved criteria and technical mechanisms for data access
ii. strengthen ethical criteria for dataset use (consent, privacy, accountability, ...)

Projects should:

- address key biomedical challenges using ML methods
- advance ML methods for future use in biomedicine
- produce transformative data sets, designed with ML in mind
- propel new ways to gather massive data in biomedicine
- **involve strong engagement from leading ML researchers**

Project review should:

- **incorporate expertise in ML as well as traditional biomedical domains**

Overall Initiative Goals

- Establish a launchpad for widespread adoption of **Next-Generation AI**
- Create next generation AI-driven scientific **design and assessment frameworks**

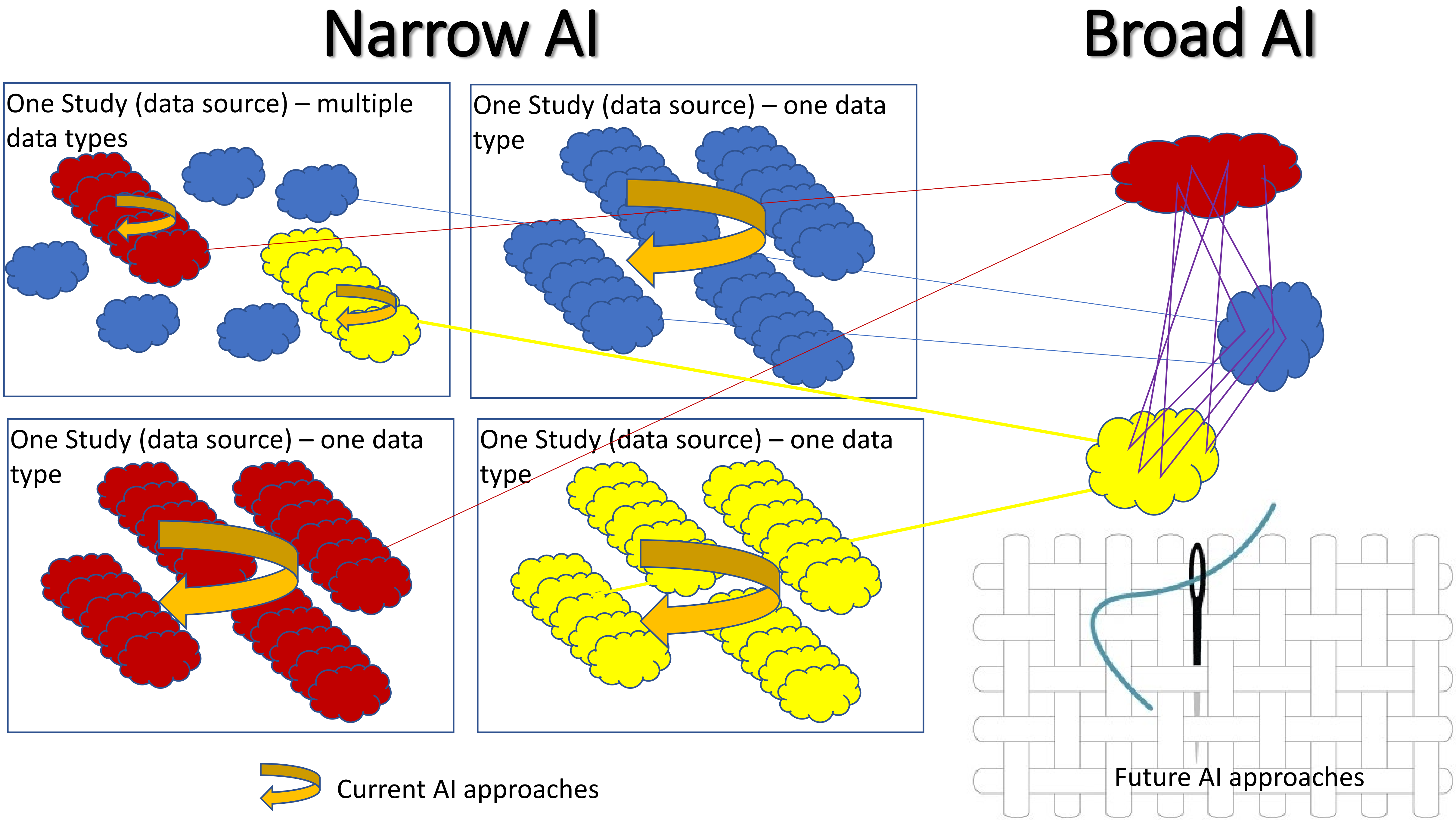
- Enable transformative data collection around **grand challenges** in biomedical research
  - Challenges that are currently beyond our human intuition and **require next-generation AI approaches** to solve
  - All future challenges to use adoptive framework

Artificial Intelligence – what's next-gen?

•Self-driving cars
•Facial recognition tools to predict depression and mental health
•Detection of cancerous pulmonary nodules on chest X-rays
•Improved accuracy and speed of interpretation of pathology slides
•Classification of skin cancer by image analysis
•Diagnose heart attack from ECG data
•Finding diminutive (<5mm) polyps in a colonoscopy
•Diagnosing eye conditions from retinal fundus photographs

1) Remove the black box - explainable AI
2) Remove fragile AI – impenetrable security
3) Remove bias – ethics in AI
4) Remove big data – learn from small data
5) Address physics of AI - infrastructure

Narrow AI → Broad AI → General AI

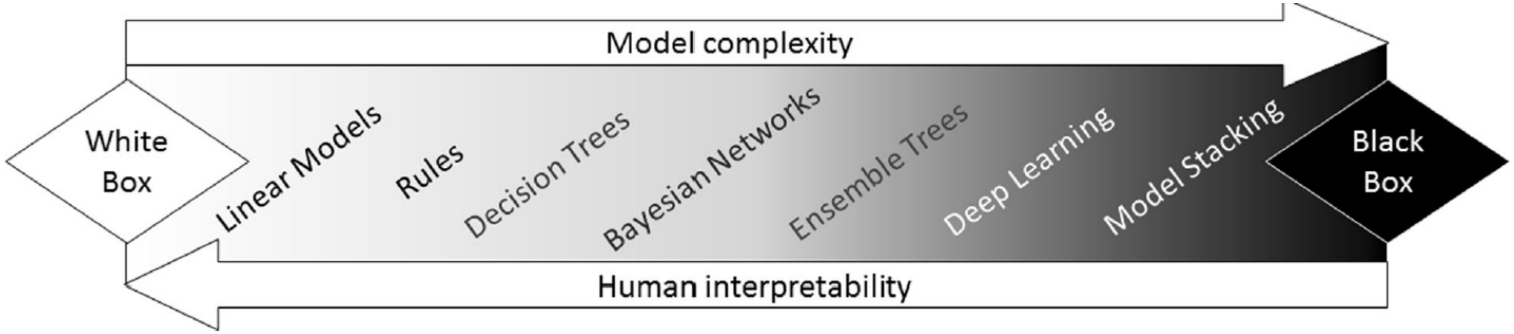

Narrow AI
Broad AI
One Study (data source) – multiple data types
One Study (data source) – one data type
One Study (data source) – one data type
One Study (data source) – one data type
Current AI approaches
Future AI approaches

Who/What needs to come together?

- Biologists/Biomedical Scientists
- Engineers
- Mathematicians
- Statisticians
- Computer Scientists
- Physicists
- Chemists
- Clinicians
- Economists
- Philosophers
- Anthropologists

Next-generation AI methods?

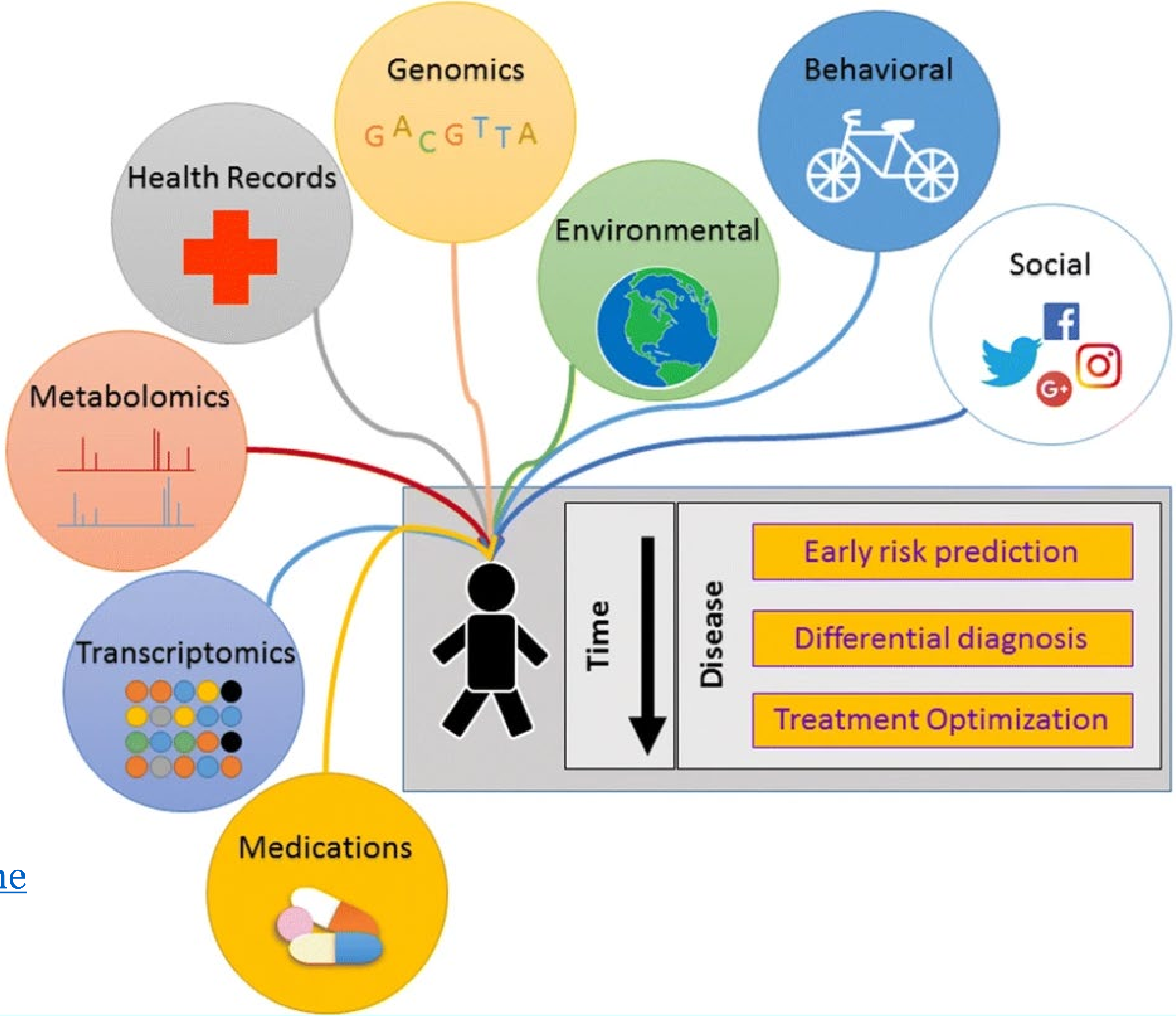
- Transfer learning
- Artificial general intelligence
- End-to-end learning (in DL)
- Tabula rasa learning theory
- Bayesian networks and inference
- Pearl-esque probabilistic causal learning
- Monte Carlo simulation and tree search
- Hypothesis-free, unsupervised DL
- High-scale modeling for prediction and forward simulation
- Quantum-inspired optimizations, including sampling, minimization, and training neural networks



Prosperi et al. BMC Medical Informatics and Decision Making (2018)

Grand Challenges

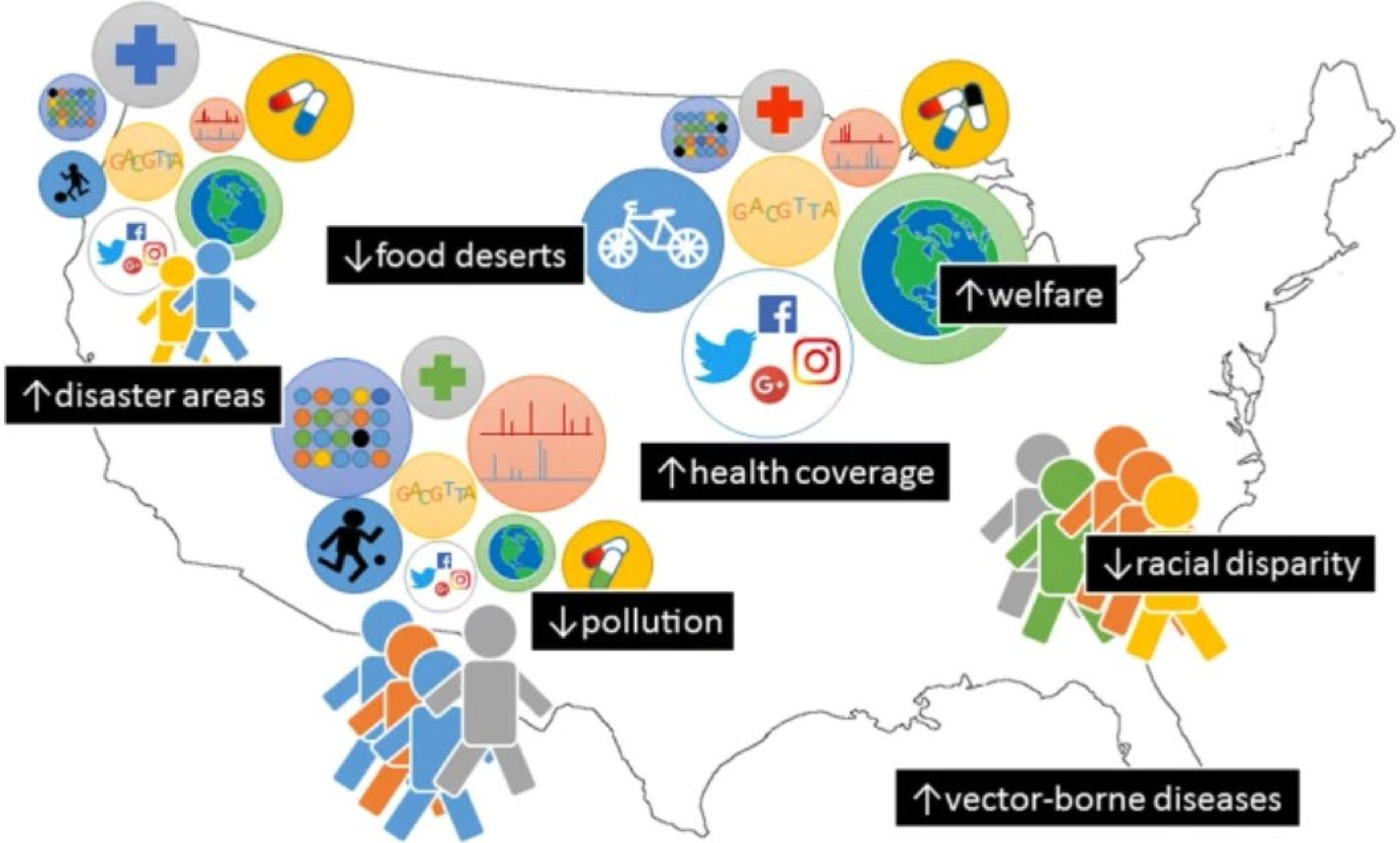
that integrate all types of biomedical and behavioral data to predict health outcomes



From: [Big data hurdles in precision medicine and precision public health]()

Prosperi et al. BMC Medical Informatics and Decision Making (2018)

Some Environmental Factors Influencing Outcomes



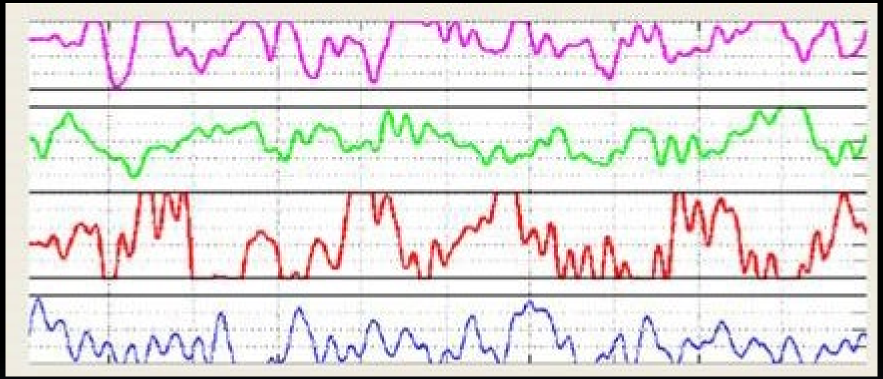
Precision public health. Community, societal and ecological factors must be accounted on top of the individual-based, fine-grained approach for precision medicine. The map is an edited version of a Wikimedia Commons image (https://commons.wikimedia.org/wiki/File:United_States_Administrative_Divisions_Blank.png, licensed under the Creative Commons Attribution-Share Alike 3.0 Unported)

Prosperi et al. BMC Medical Informatics and Decision Making (2018)

Grand Challenge Use Cases

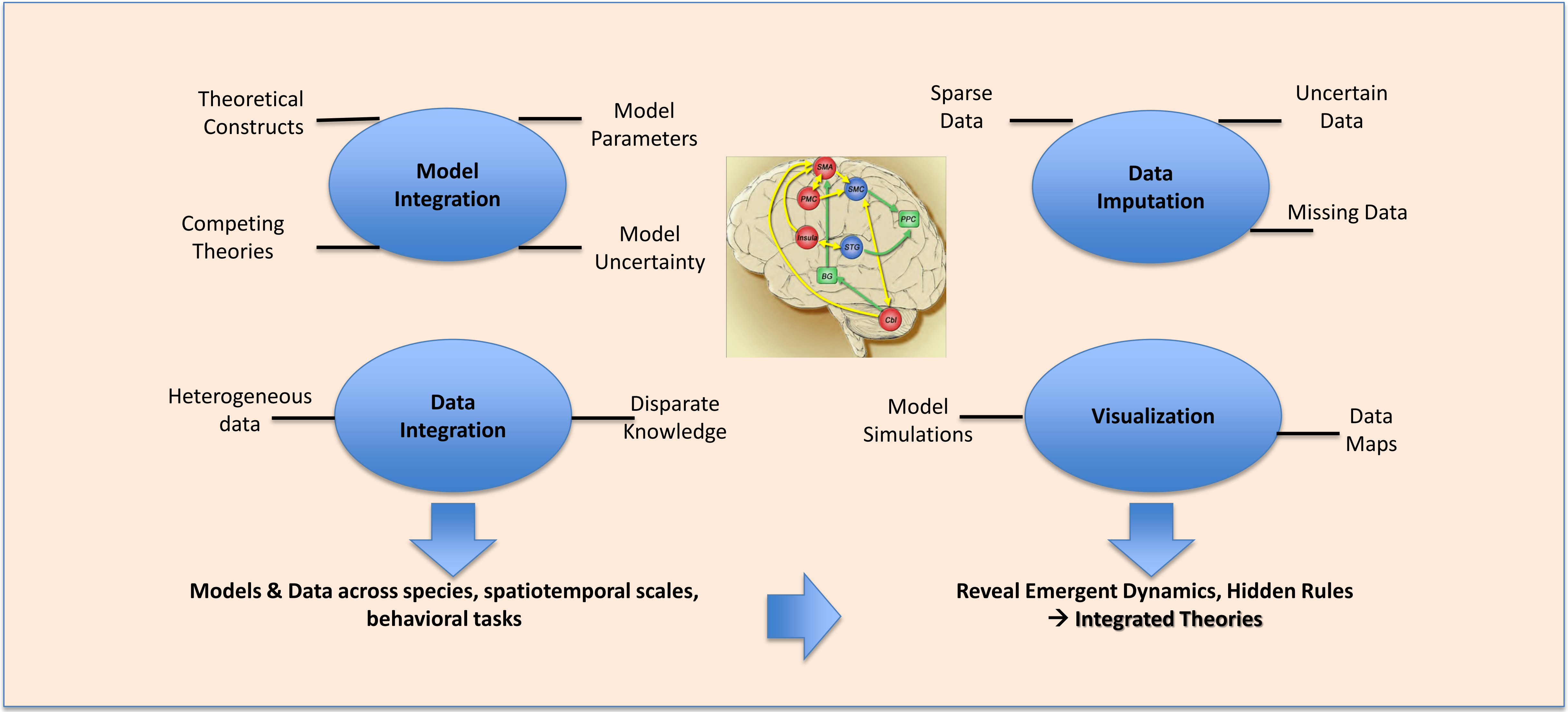
- Connecting multiple pieces of information for research
  - Integrating heterogenous, variable, uncertain data
  - Finding the hidden signals to derive knowledge and insight

- Wearables give a "movie for health"
  - How to read between the lines of data
  - Finding the signal we never saw before



- Generalizing from known diseases to rare diseases
- Finding the mechanisms of health restoration

Discovering Theories & Knowledge → INSIGHT



Development and implementation of a machine-driven knowledge integration process for discovery

1960: A Vision for Cognitive Assistance

"In not too many years, human brains and computing machines will be coupled together very tightly, and the resulting partnership will think as no human brain has ever thought!"

--JCR Licklider

Visionary psychologist and computer scientist

Funded research that led to most of modern computing

Google Car, 2018

Tesla Truck, November 2019

Oakley Radar Pace

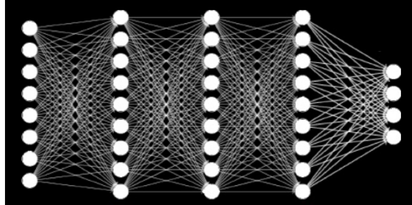
- https://www.youtube.com/watch?v=-S4V1TS4yFk

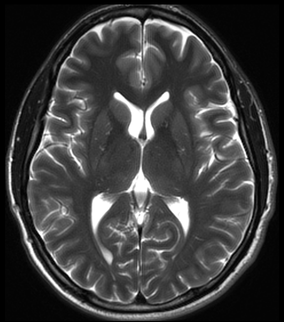
Radar Pace, available online

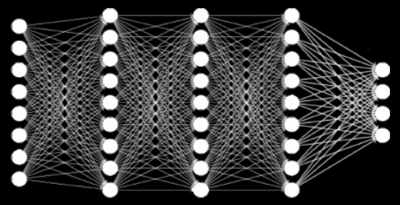
AI can change how we gather data, not just how we interpret it

Learned reconstruction with neural networks

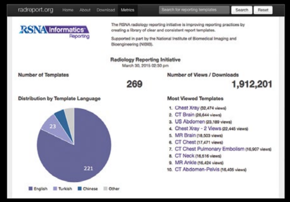
data







information



CAI²R

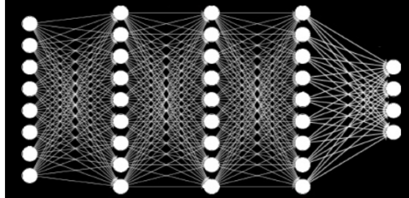
FAIR

DK Sodickson, P41 EB017183-05S1

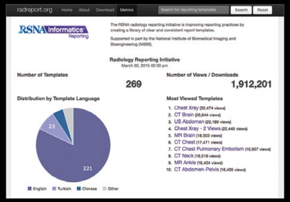
NYU Langone Health | CAI²R

Direct to information

data



information



DK Sodickson, P41 EB017183-05S1

AI to Extract Human Intelligence (not human stupidity)

- Integrating the right types of data
  - at the right time
  - within the right context at (this point in time)
- Avoiding human biases
  - coding for ethics and (human?) learning
  - AI research for ethics
- Incorporating the totality of the data
  - without collecting ALL the data
  - Stitching together SMALL data
- Re-engineering the Future of Health
  - Understanding the mechanisms of prevention, diagnosis and treatment
  - Lowering healthcare costs

Stay Tuned!

Vision: To Propel Progress in Biomedical Research through **NEXT-GENERATION AI** (beyond Narrow AI to Broad AI)
Culture Change: Biomedical experiments designed for next-gen AI → AI designed for biomedical experiments*
3 Pillars: People (multiple disciplines), Data (transformational), Ethics (data bias and transparency)
<u>Goals/Outcomes after 7 years (FY21-27, ~$200M):</u>

- AI Design Centers for the Biomedical Community –- data design and assessment, ethics and training
- New "Gold Data" that can be mined with future AI methods
- Ability to "stitch" Gold Data with existing data (across sites, protocols, processing methods)
- Next generation discoveries for biomedical research, powered by next-gen AI

Partnerships: DARPA, NSF, DOE, FDA, …

* Includes biological and behavioral studies

Immediate Timeline:

October 27-28, 2020: Community Workshop in partnership with DARPA Synergistic Discovery and Design (SD2) program

Fall 2020: Release of Funding Opportunities for DEFINE AI Design Centers inspired by Biomedical Grand Challenges

Fall-Winter 2021: Formation of Multidisciplinary Teams, Vet Grand Challenge Ideas → Online breakout groups for each Grand Challenge idea

Thank You!

Grace C.Y. Peng

grace.peng@nih.gov