Envisioning the future: The future of data platforms for human genetics

Josh Denny, MD MS

11/28/2018

All of Us RESEARCH PROGRAM | The Precision Medicine Initiative

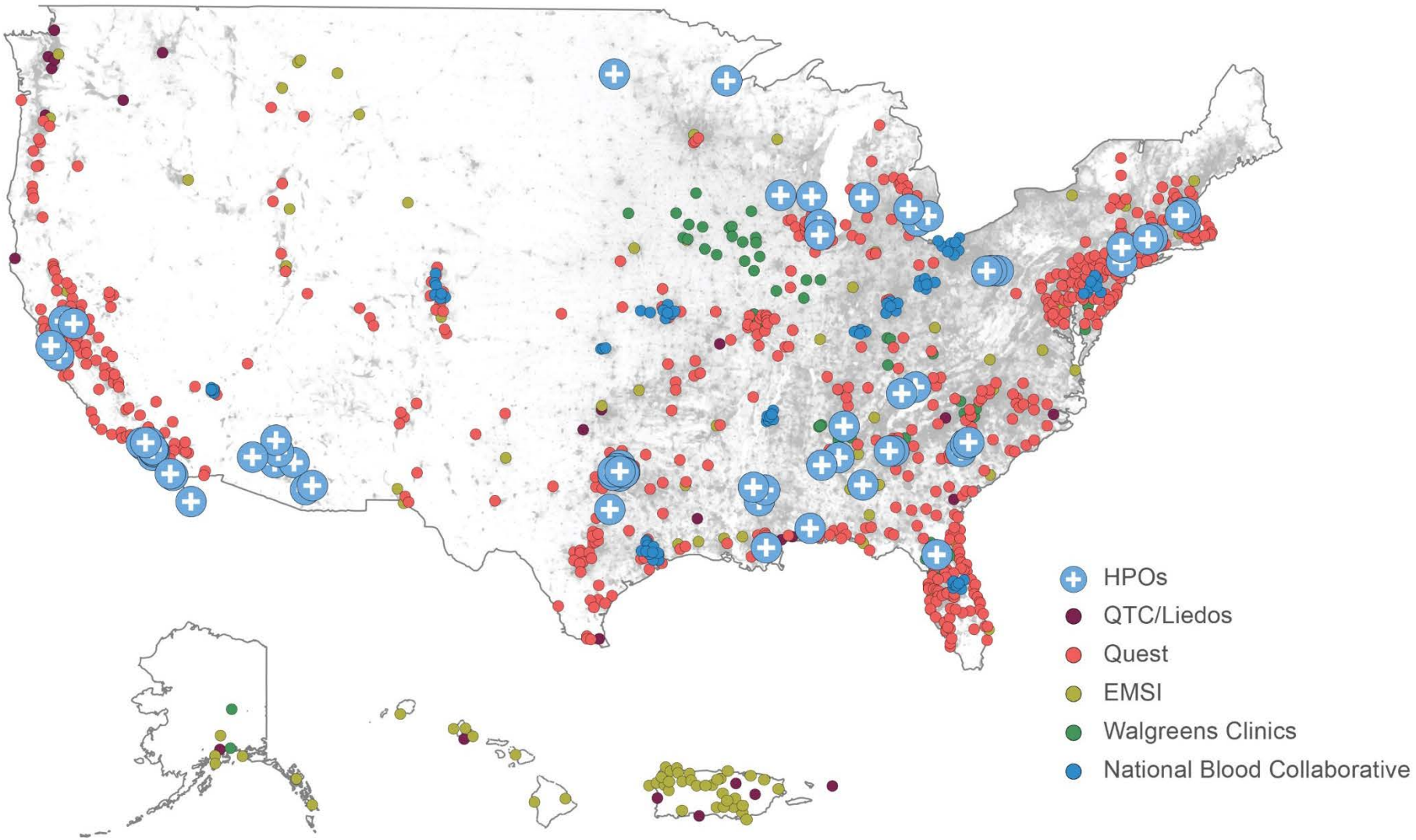
The *All of Us* Research Program – Breaking Down Data Silos

All of Us Research Program – Summary of Protocol and Status

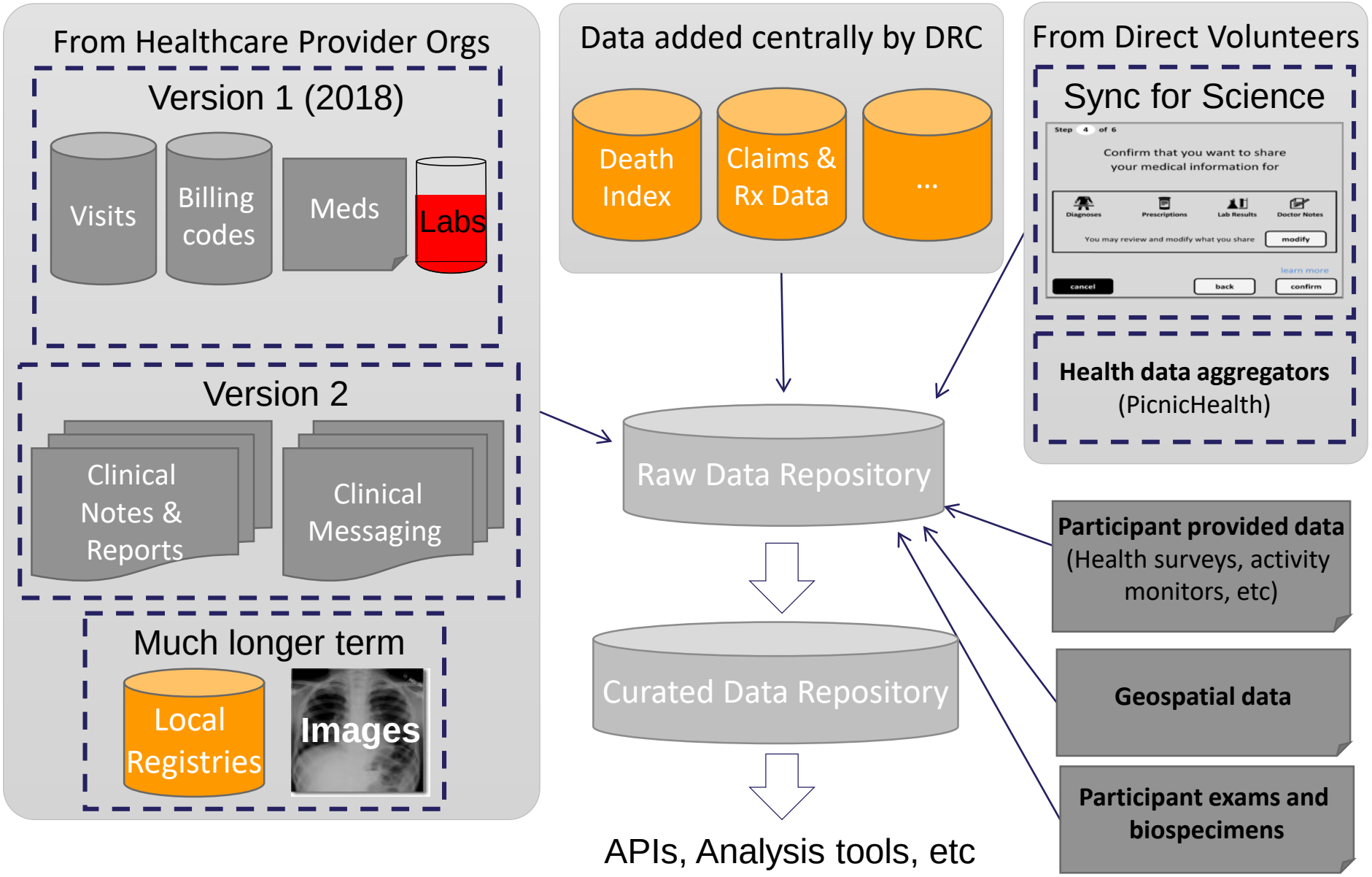
Key elements:

- **Goal**: 1 million or more diverse participants
- National launch May 6, 2018
- **Current**: ~83k "full" participants, ~140k have started, >45% nonwhite, >75% underrepresented in research
- **Participants will get data back** (genetics, EHR)
- Longitudinal, **recontactable**

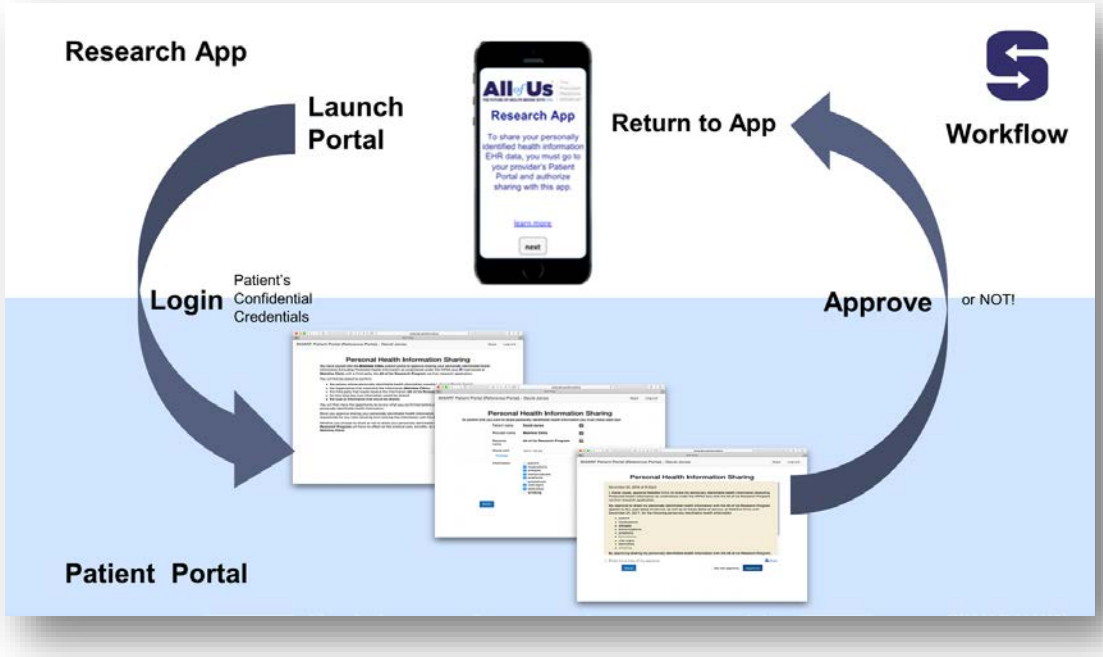
Current and planned in-person enrollment centers



All of Us will aggregate data from many sources



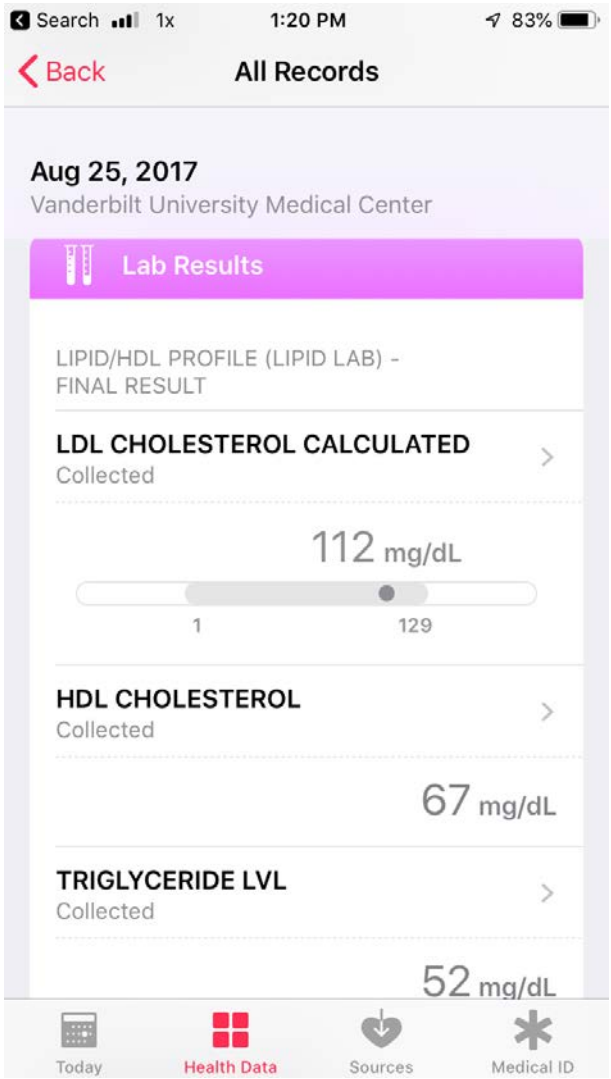
Putting the patient in charge of data mobility - Sync 4 Science (S4S)



Sync 4 Science:

- FHIR-based
- Starting with Meaningful Use Common Clinical Data set
- Piloting now

…or on your smartphone

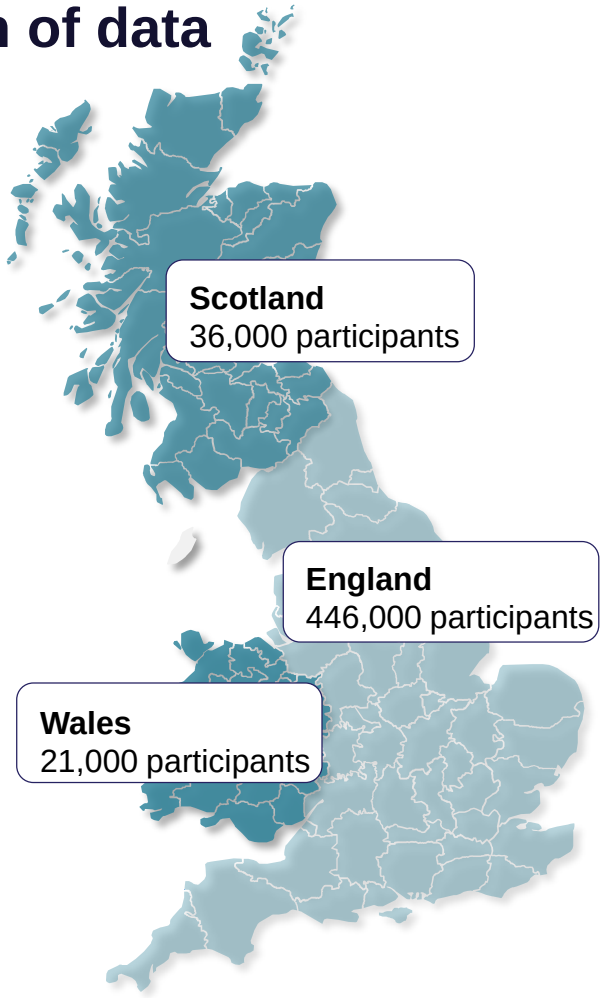


Scientific priorities for *All of Us*

Scientific Priority	Timeline				
	End of 2018	End of 2019	2020-2022	2023-2027	>2027
Expected cohort size	90,000	>200,000	<650,000	>1,000,000	>1,000,000
Returning data to participants	+	++	+++	+++	++++
Discover disease risk factors across genes and environment			++	+++	++++
Better predict therapeutic safety and efficacy			++	+++	+++
Discover disease biomarkers			++	+++	+++
Connect mHealth with clinical outcomes			++	+++	+++
Evaluate genetic variants			++	+++	+++
Develop new disease classifications			+	+++	++++
Support clinical trials			+	+++	+++
Enable machine learning applications			++	+++	++++
Better understand health disparities			++	+++	+++
Enable new therapeutics					++

UK Biobank in a nutshell: an open access prospective study with large size and extensive breadth and depth of data

- 500,000 UK men and women aged 40-69 years when recruited during 2006-2010
- Consent for all types of health research by both academic and commercial researchers, and for long term follow-up through all health-related records
- Extensive baseline questions and measurements, with biological samples stored for future assays
- Open access for approved research for the benefit of the public's health: see www.ukbiobank.ac.uk



biobank uk

Courtesy of Dr. Cathie Sudlow

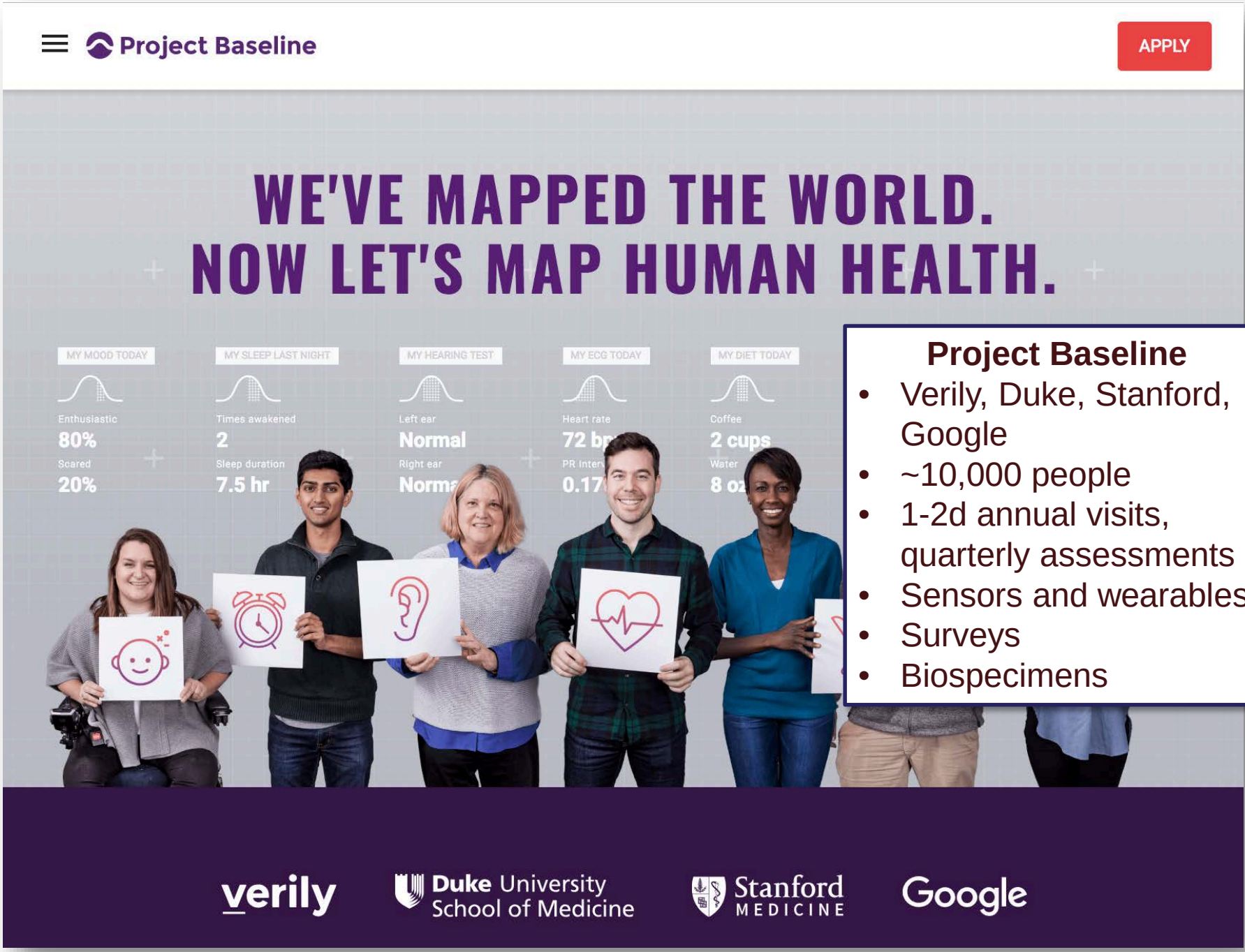
Assays

biobank uk

- Complete:
  - GWAS
  - Standard panel of biochemical assays (e.g. lipids; hormones; metabolic)
- Underway:
  - Whole exome sequencing
  - Whole genome sequencing
  - Infectious disease assays
  - Nightingale metabolomics (lipidomics)
- Other:
  - Whole-body Imaging (100k)
  - Activity monitoring (100k)

All of Us RESEARCH PROGRAM | The Precision Medicine Initiative

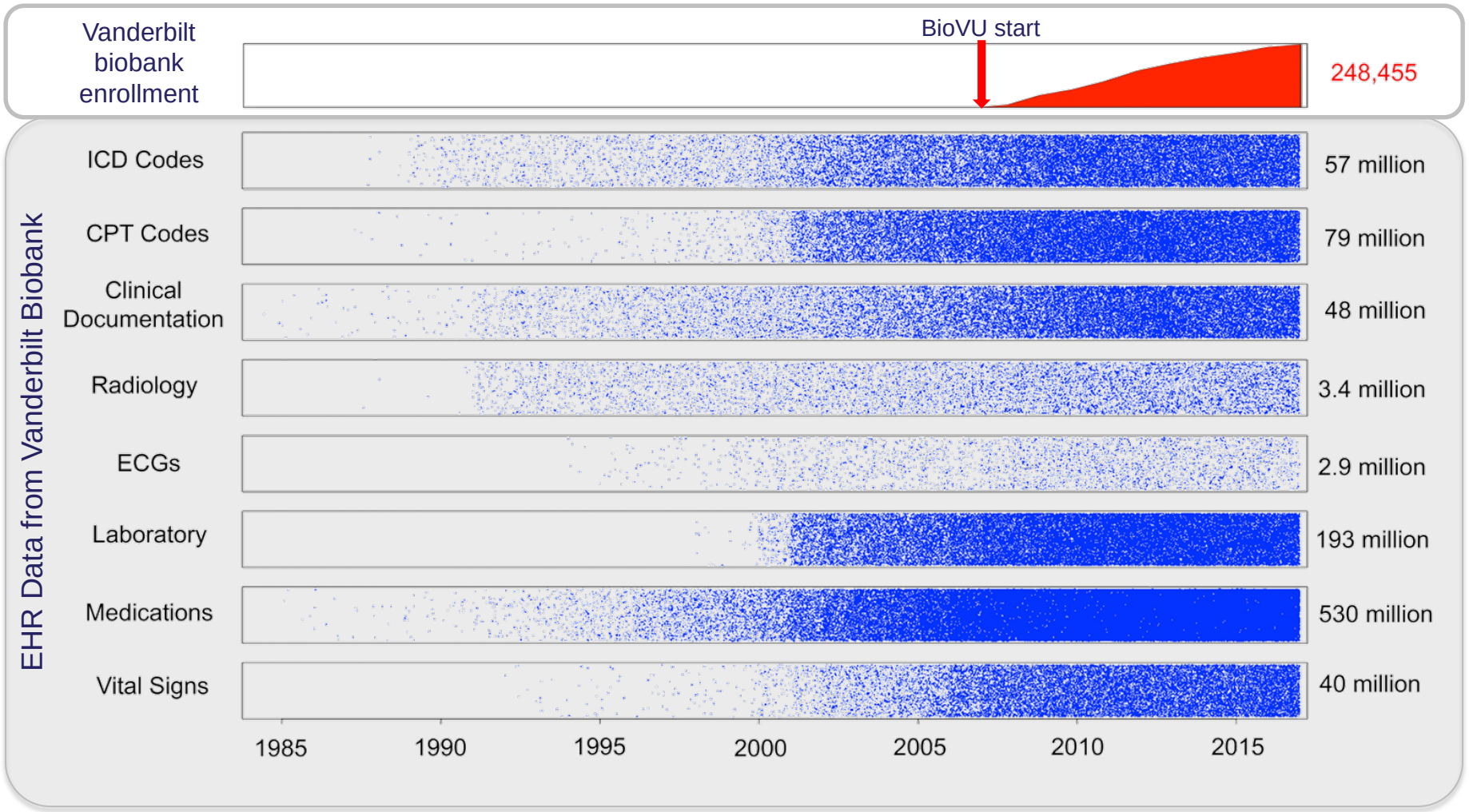
- Planned to start in 2019:
  - GWAS
  - Whole genome sequencing
  - Assay pilot (a few clinically relevant labs)
- Other planned
  - 10k FitBit pilot
  - “BYOD” data donation for FItBit, Apple Watch, etc
  - GIS data linkages
- Many more under consideration



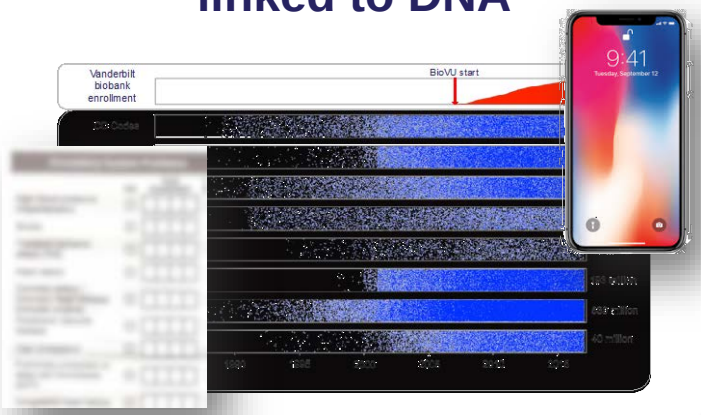
Project Baseline

- Verily, Duke, Stanford, Google
- ~10,000 people
- 1-2d annual visits, quarterly assessments
- Sensors and wearables
- Surveys
- Biospecimens

Passive data follow-up via EHRs provides dense resource for efficient discovery: Vanderbilt DNA Biobank example

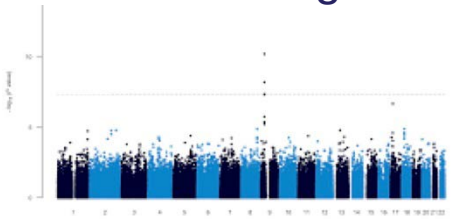


EHRs, surveys, wearables, etc. linked to DNA

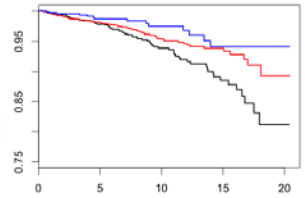


Phenotyping (codes, labs, NLP, etc)

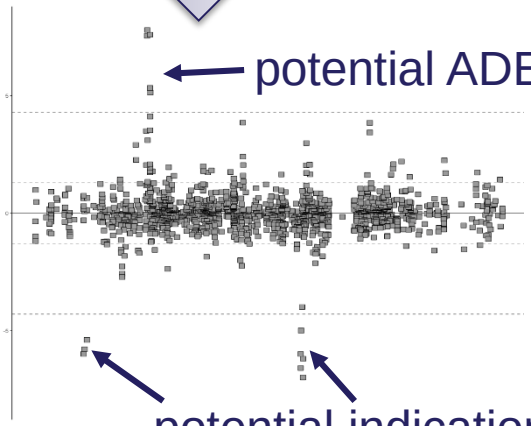
GWAS identifying new disease targets



Pharmacogenomic discovery



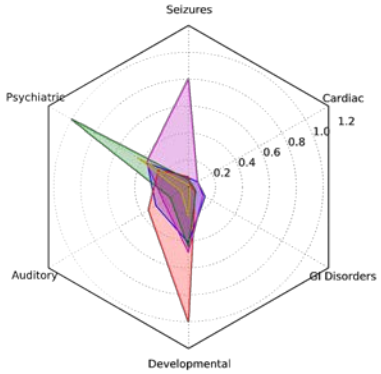
PheWAS

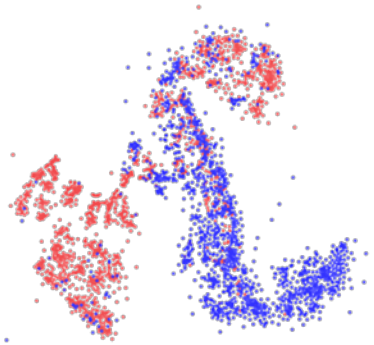


potential ADEs

potential indications

AI, clustering



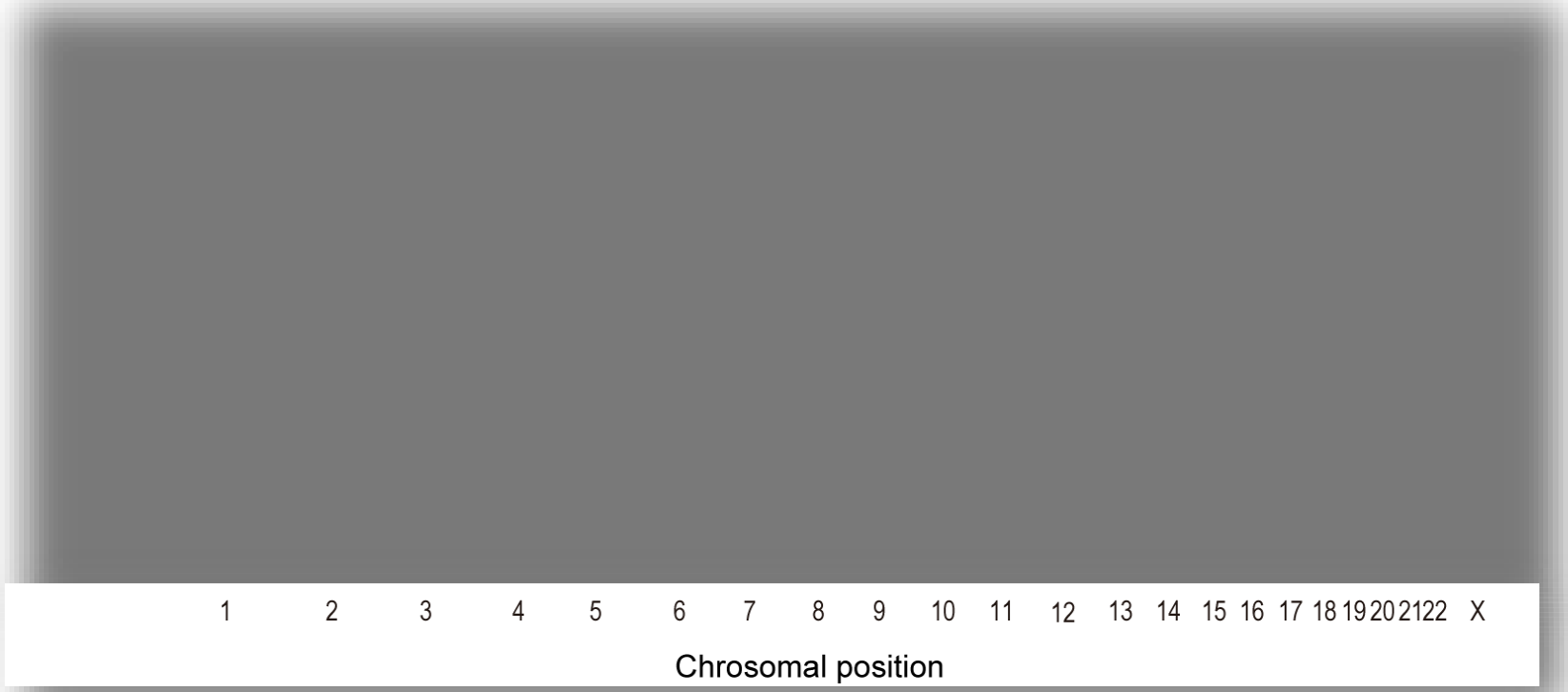


New disease clusters, subtypes, and learned phenotypes

Genetics of rheumatoid arthritis contributes to biology and drug discovery

Okada et al., *Nature* 2014

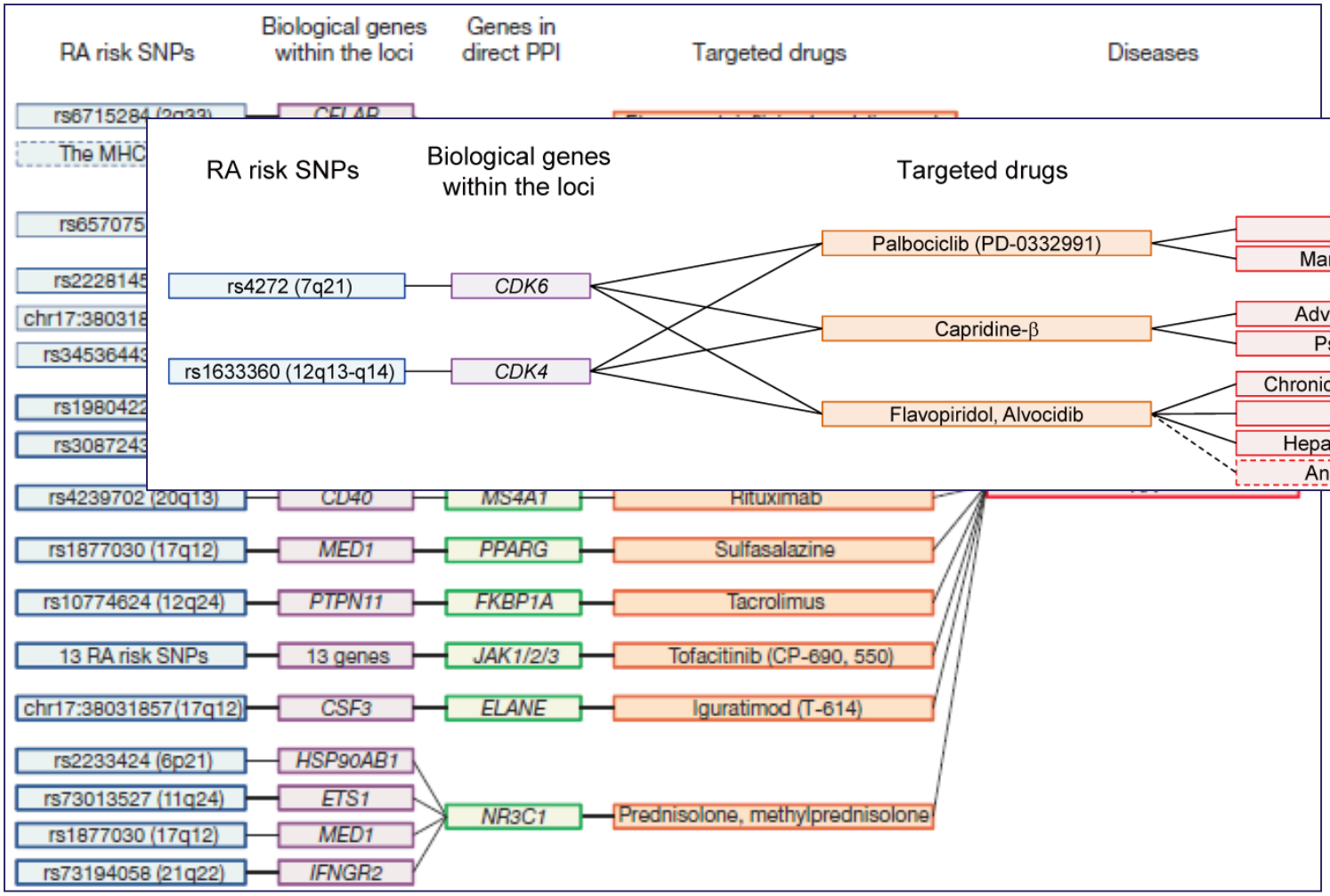
- >100,000 cases and controls from RA studies and biobanks, multiple ethnicities
- 101 RA loci → 98 candidate genes



Genetics of rheumatoid arthritis contributes to biology and drug discovery

Okada et al., *Nature* 2014

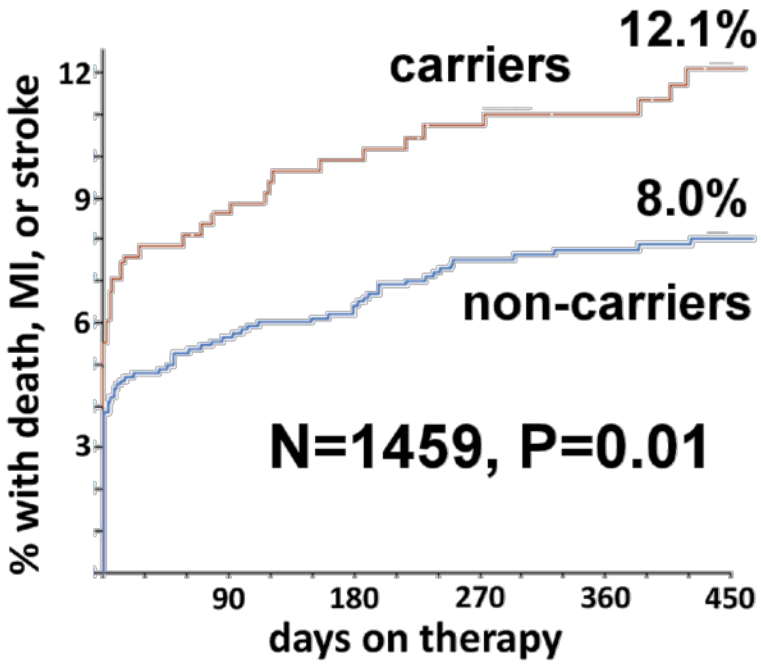
- Encode known targets for RA drugs, and drugs used in other disease
- ~2/3 display associations with diseases beyond RA



Finding drug response in cohorts: Clopidogrel adverse events associated with *CYP2C19* status

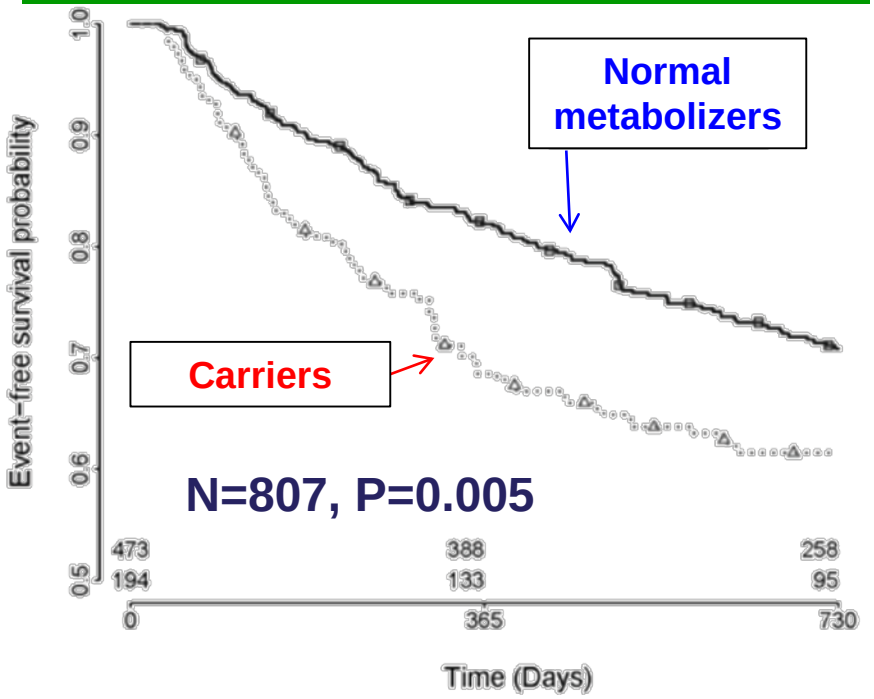
clopidogrel (*inactive*) —*CYP2C19*→ **2-oxoclopidogrel** (*active*)

From clinical trials



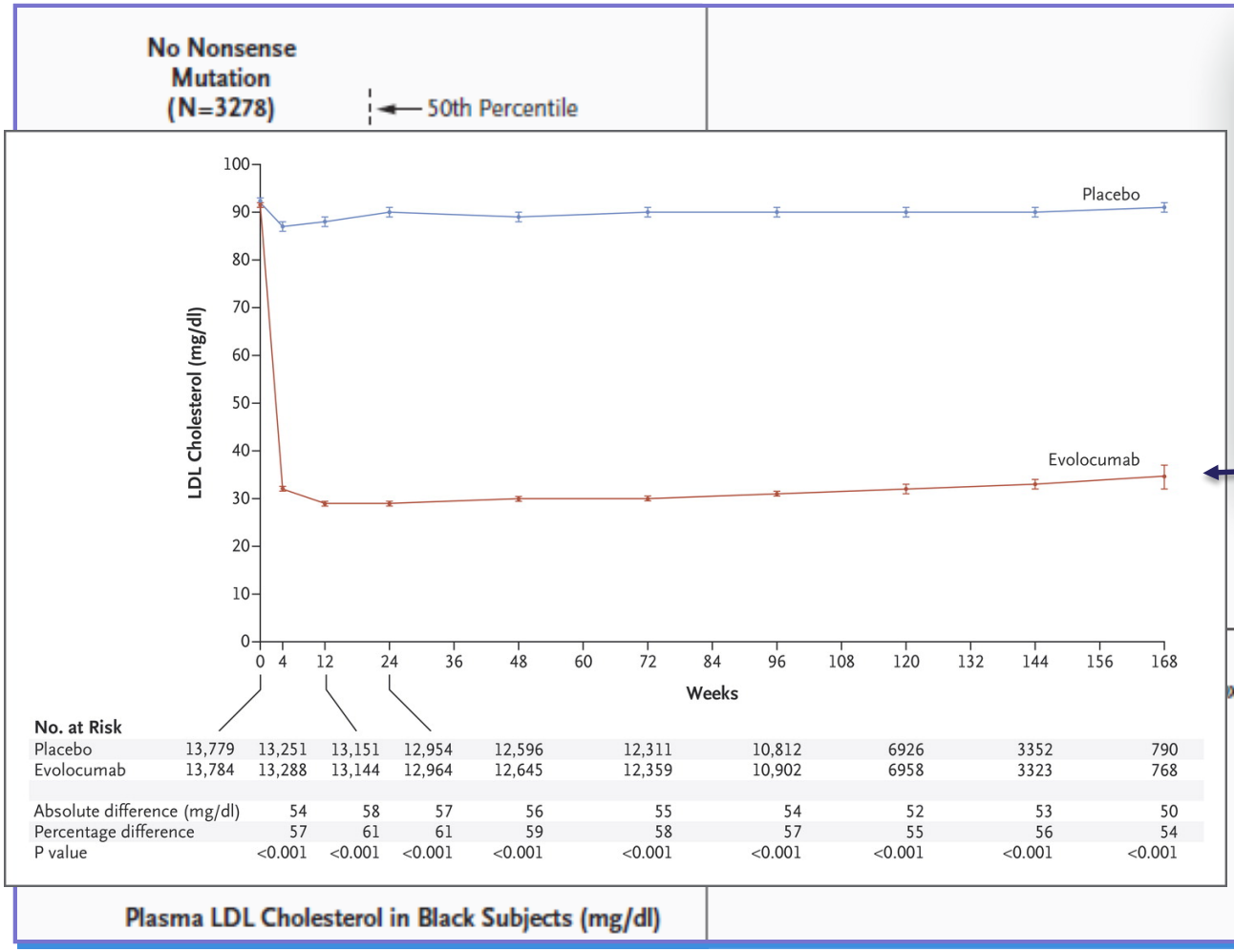
Mega et al., *NEJM* 2009

From the EHR



Delaney et al. *Clin Pharm Ther.* 2012

Identify new drug targets: The story of *PCSK9*



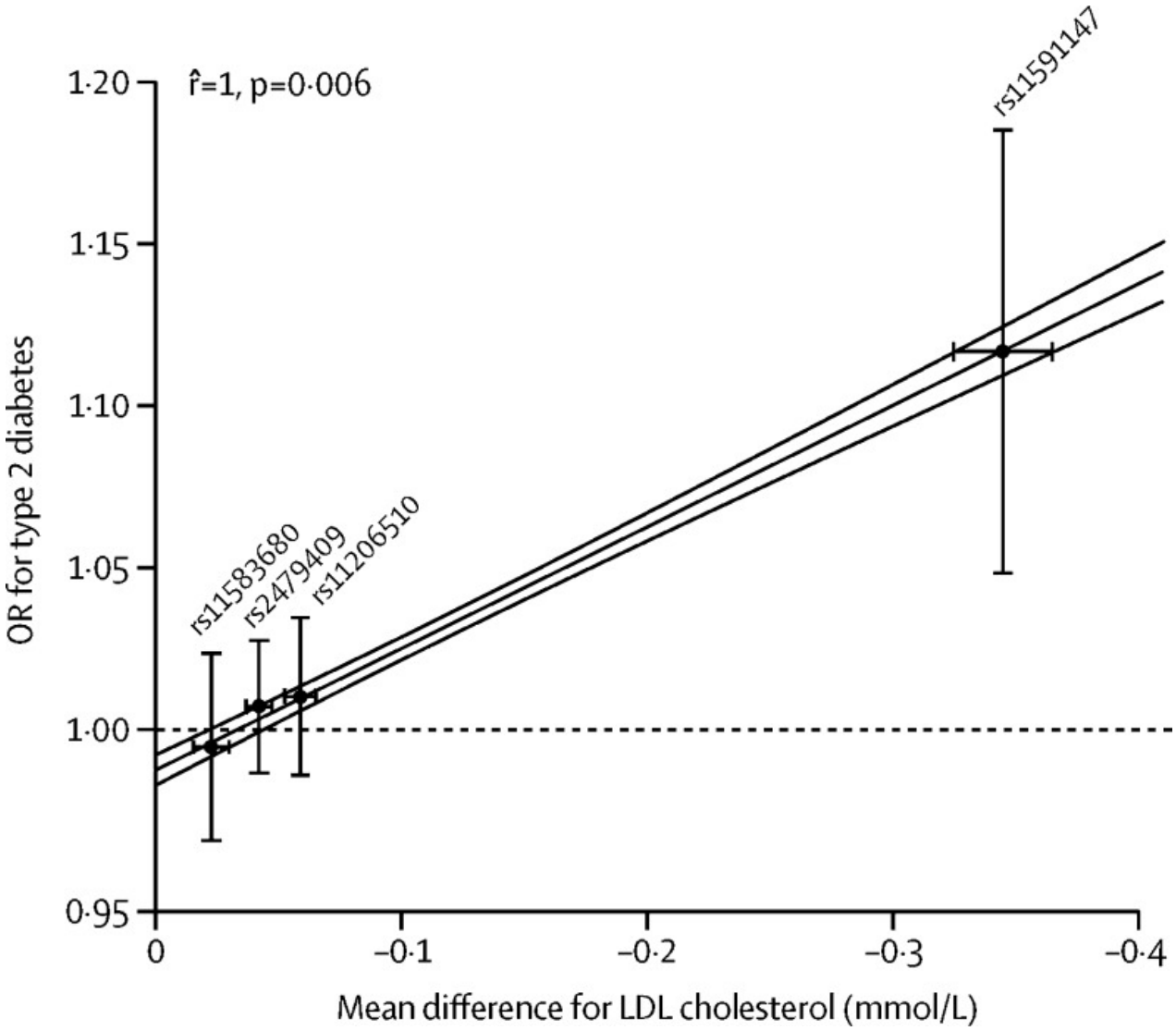
Nonsense mutations in *PCSK9* result in very low LDL and protect against coronary disease

PCSK9 inhibitors are now on the market (alirocumab, evolocumab)

Cohen et al., NEJM 2006
Sabatine NEJM 2017

What do cohorts tell us about potential risk of PCSK9 inhibitors?

(a Mendelian randomization study)



Lancet Diab Endocrinol. 2017; 5: 97–105.

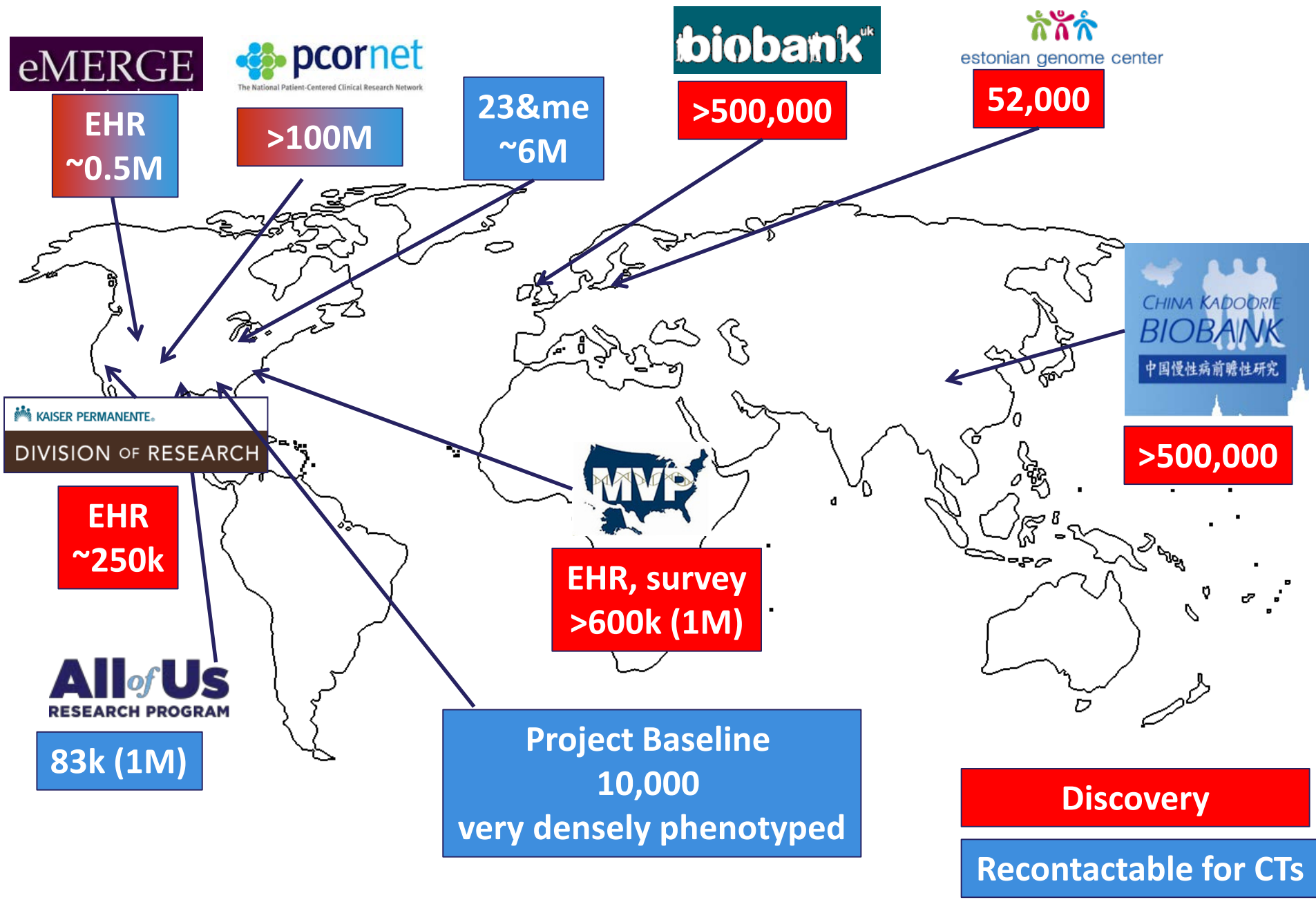
Summary: how large cohorts may accelerate clinical trials

- Basic discovery
  - E.g., new genetic loci predicting cholesterol levels (*PCSK9*) or drug effects (*CYP2C19* and clopidogrel efficacy)

- Direct recruitment and facilitate assessment via extant infrastructure
  - Some cohorts are **recontactable**
  - More cost effective recruitment
  - More cost effective follow-up (EHRs, basic assessment infrastructure and extant molecular data, digital devices)
  - Faster recruitment of the best targeted clinical trials (e.g., specific *CFTR* mutations)

- More intelligent trial design
  - Use large biobanks to generate hypotheses of possible indications and adverse effects for trials
  - Pursue the hypotheses that appear most promising, search for side effects most suggested

The cohort landscape is expanding…

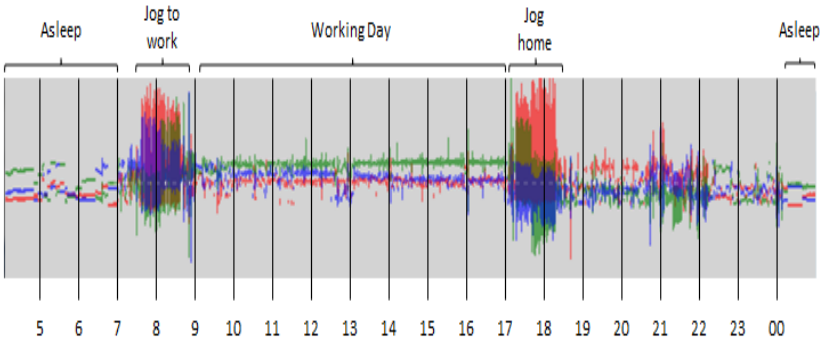


All of Us Data and Research Center Team

First *All of Us* Network meeting June 2016

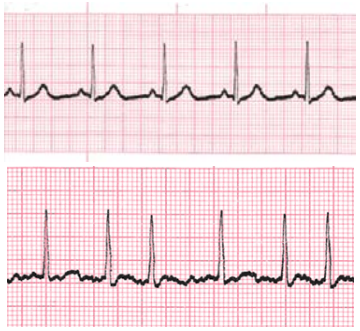
Data from portable wearable devices

Accelerometry data: 100,000 participants



Prospective design and **large size** enable well-powered studies of (causal) associations between accelerometry and cardiac rhythm measures and later onset disease

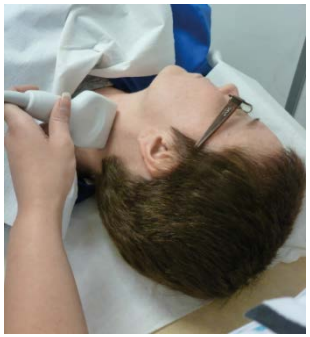
Continuous ECG monitoring: 20,000 + participants



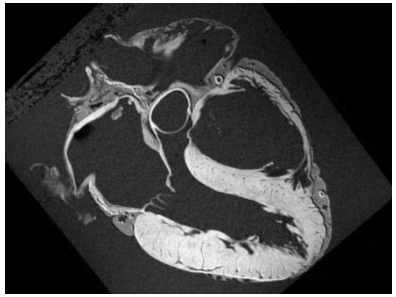
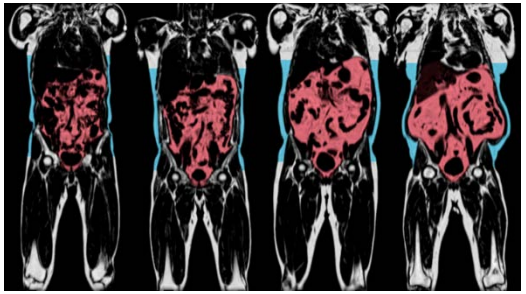
Need **scalable methods** of analysing **complex data** to derive measures for large scale analyses

Courtesy of Dr. Cathie Sudlow

Multimodal imaging of 100,000 participants


>30,000 imaged so far

Prospective design and **large size** enable well-powered studies of (causal) associations between structure and function of organs and later onset disease. Need **scalable methods** of analysing **complex data** to derive measures for large scale analyses

Courtesy of Dr. Cathie Sudlow