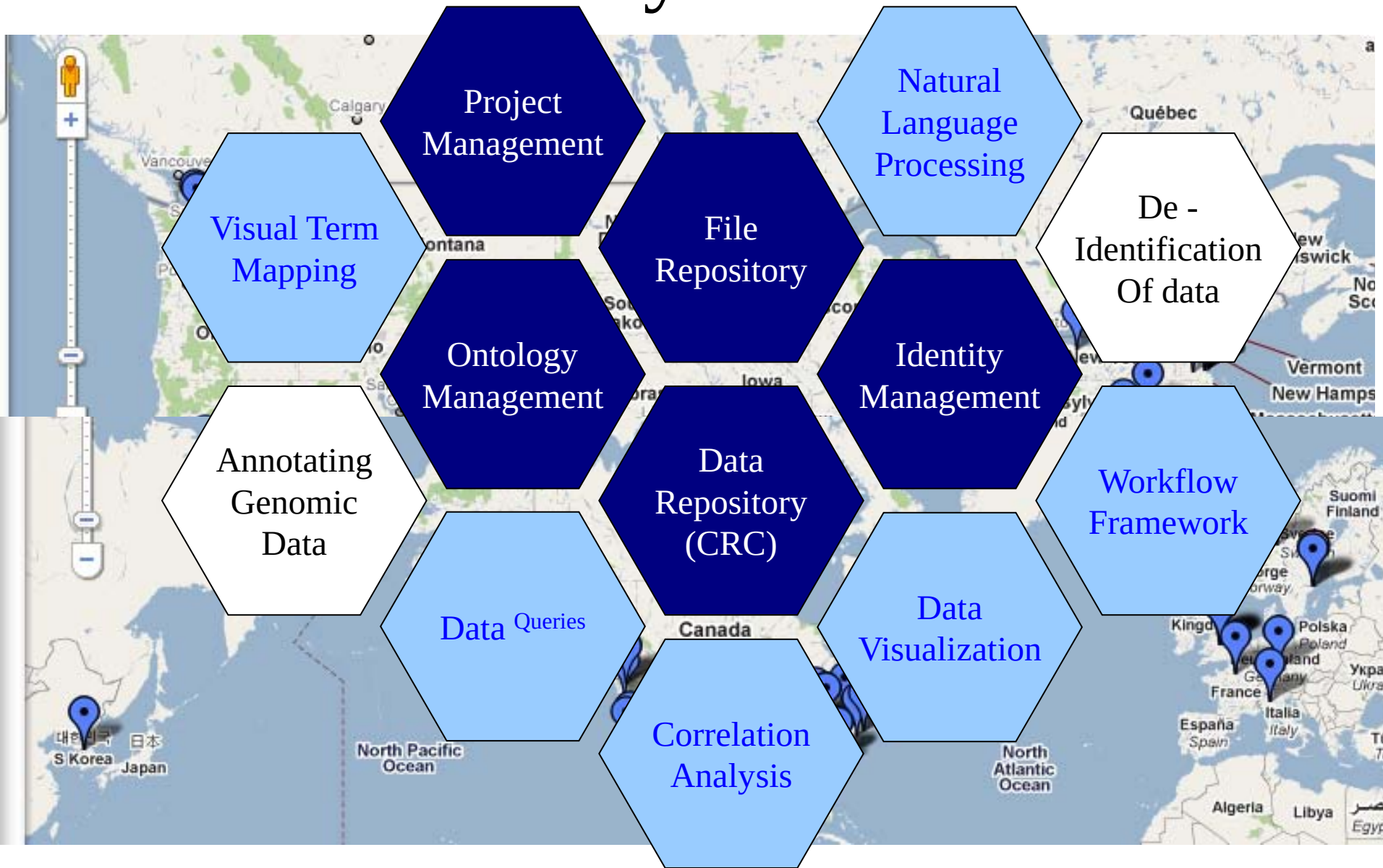
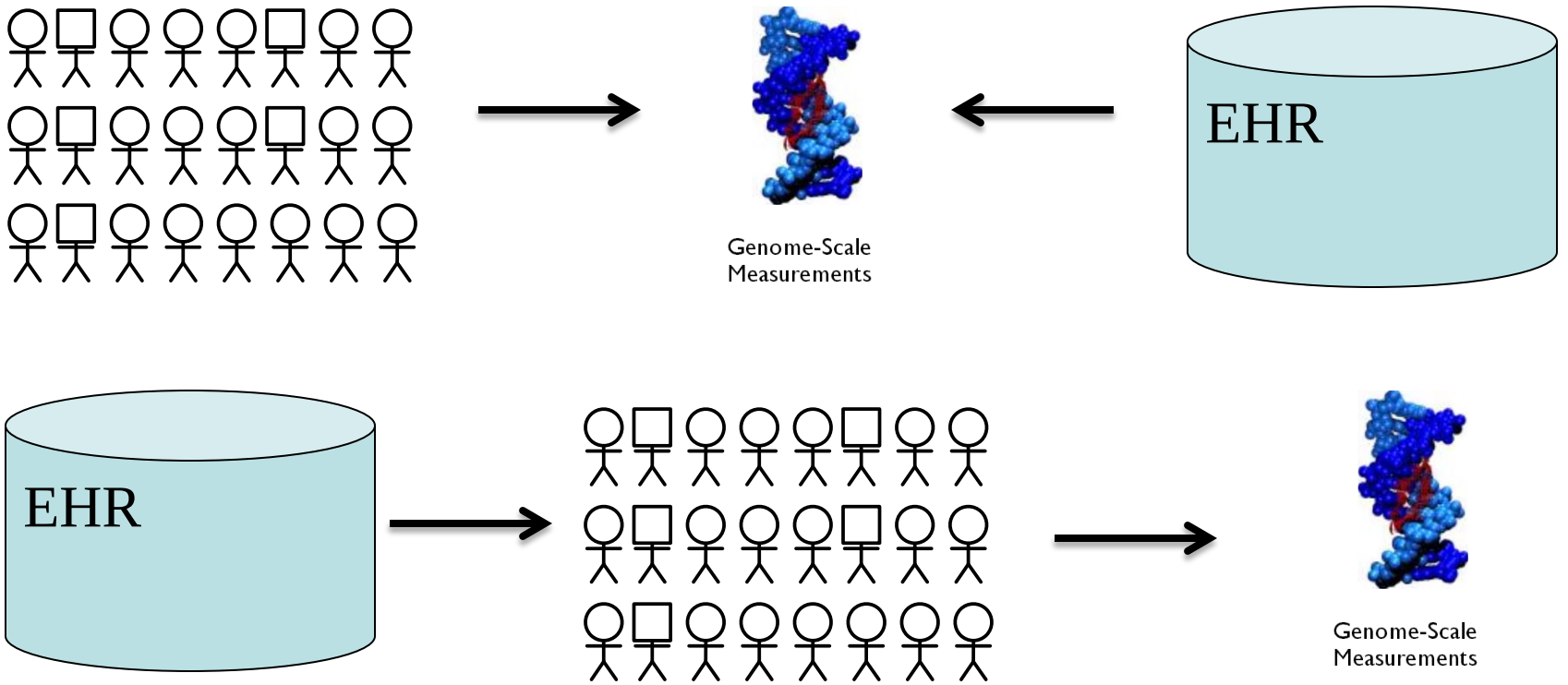
EHR Driven Genomic Research (EDGR)

Isaac “Zak” Kohane, MD, PhD

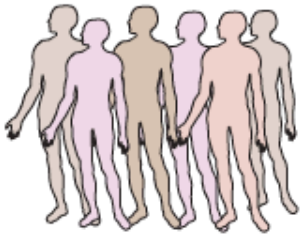
i2b2: Instrumenting the Enterprise for Discovery Research



Major Modes


EHR
Genome-Scale Measurements
EHR
Genome-Scale Measurements

Major Modes (II)



Entry route
Patients seen for healthcare

System Biobank

System-wide research database

Entry route
Subject selection

- History
- Physical exam
- Enviromental exposures
- Clinical measures

Biological/genomic measurement

Natural language processing

EHR-derived datamart

Study-specific datamart

Electronic health record

Extraction of codified data

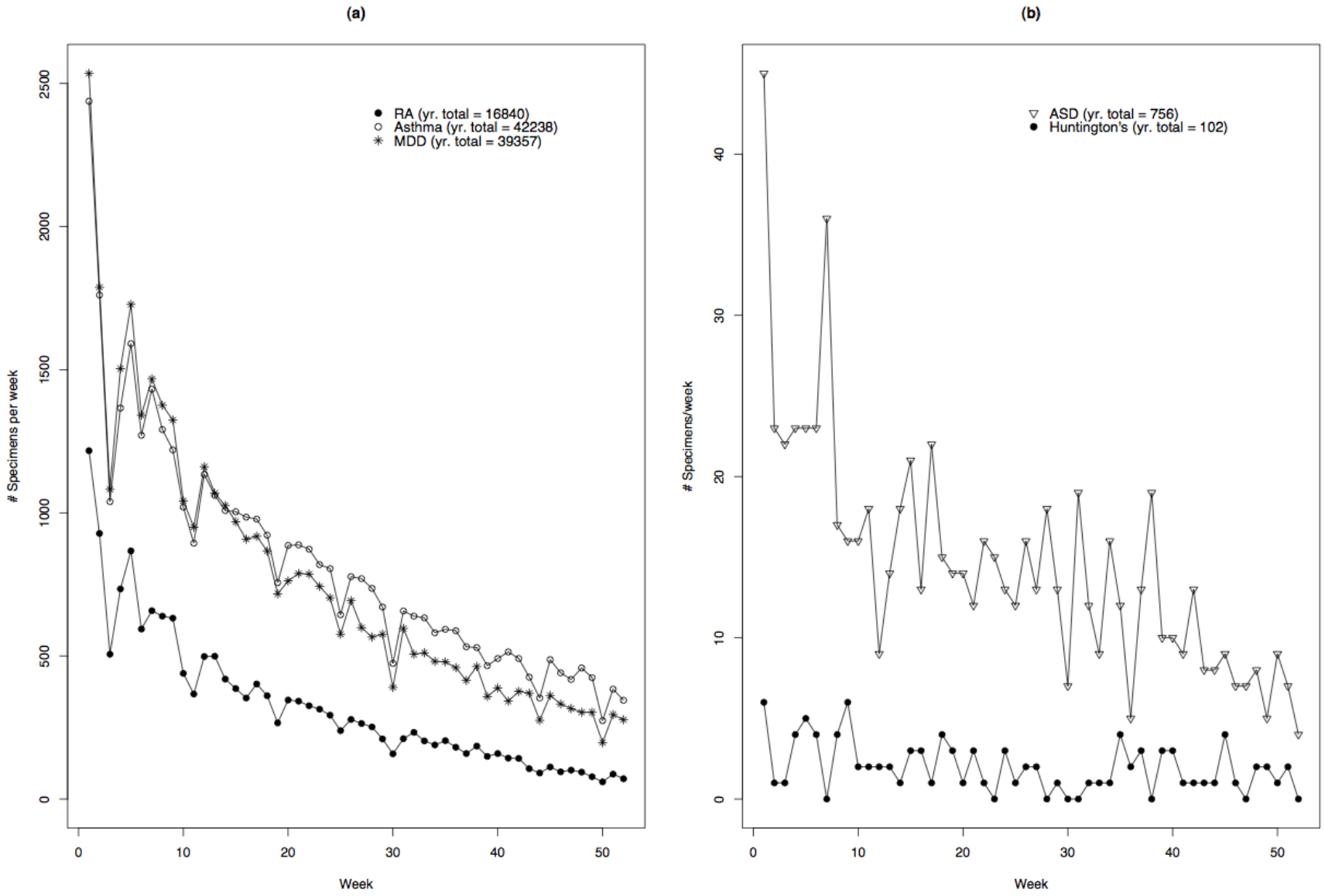
Clinical laboratory information system

Study-specific Biobank

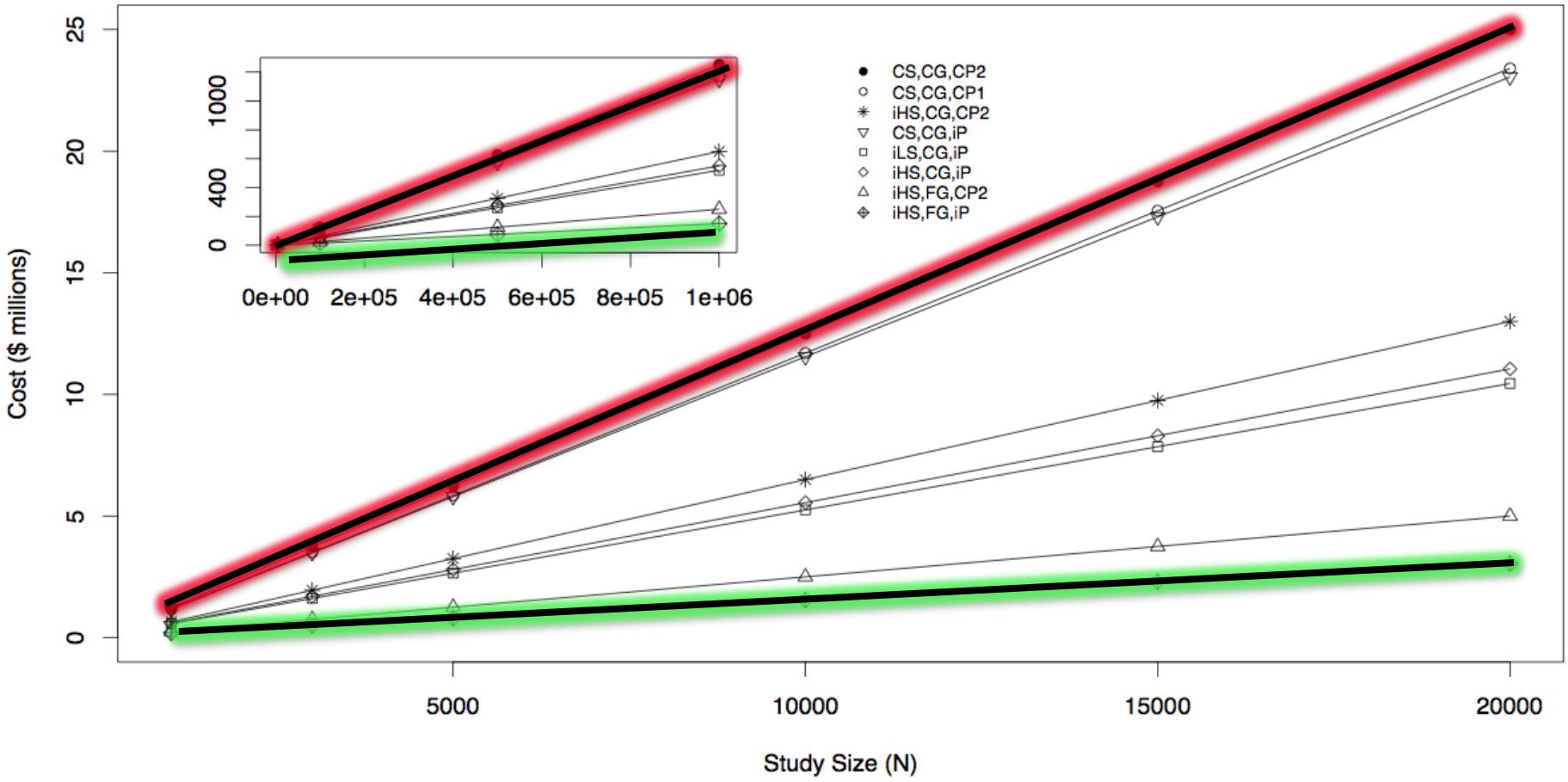
Nature Reviews | Genetics

EDGR Advantages

- Timeliness
- Clinical Relevance
- Underserved populations
- Controls
- Co-morbidity recognition (e.g. PheWAS)



Murphy *et al* Genome Research, 2009



Murphy *et al* Genome Research, 2009

EDGR Challenges

- Consent (None/Opt-in/Opt-Out)
- Cost of EHRs
- Quality of EHR data
- Lack of Family History codification
- Lack of EHR standardization
- Cultural gulf between clinical informatics and bioinformatics.
  - Translational Bioinformatics

But it works…

Kurreeman, AJHG 2011

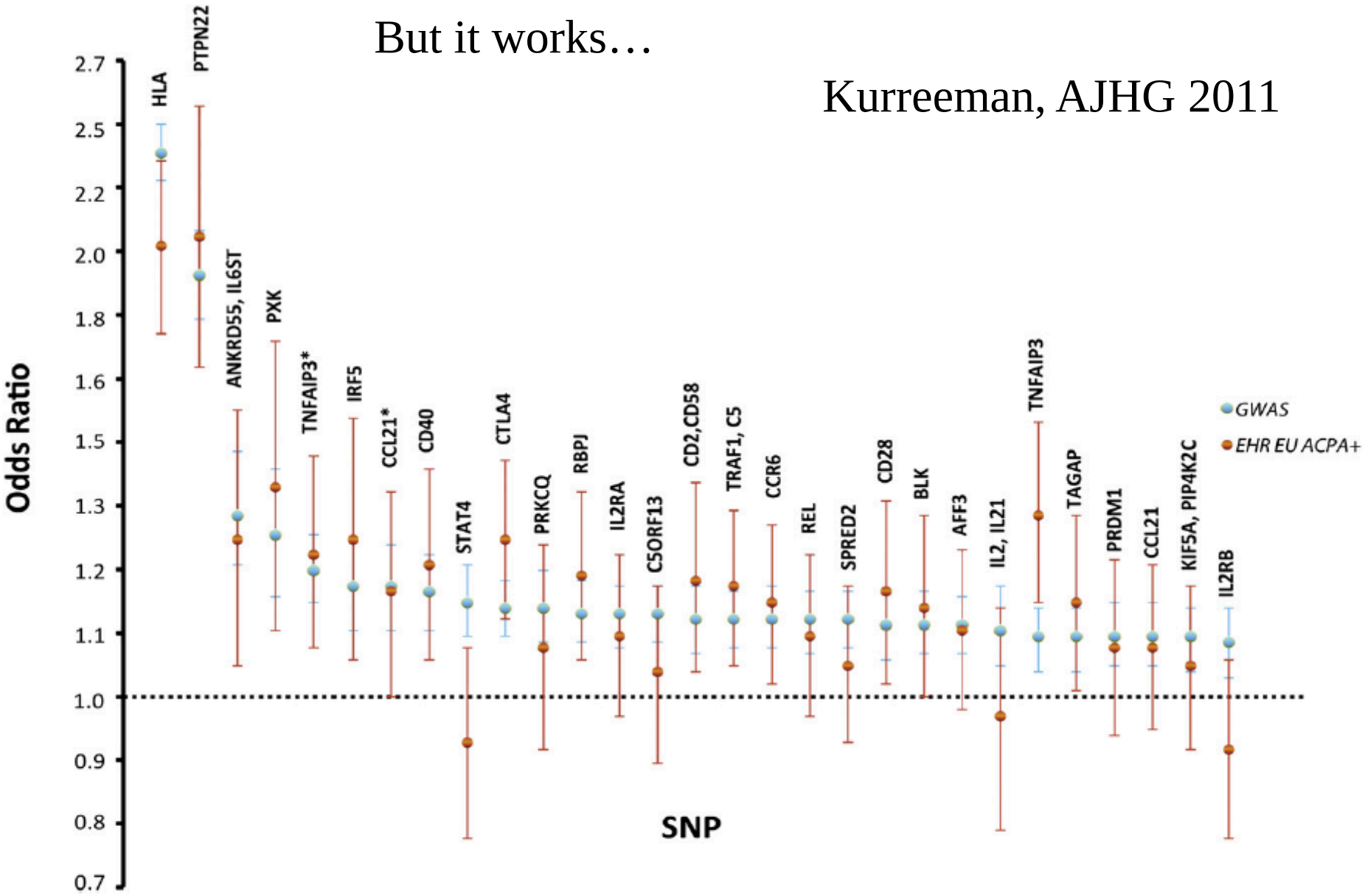


Figure 2. Overlap of Odds Ratio and 95% Confidence Intervals between Previous GWAS Meta-Analysis Dataset and ACPA+ European Subset from EHR Cohort

Kurreeman, AJHG 2011

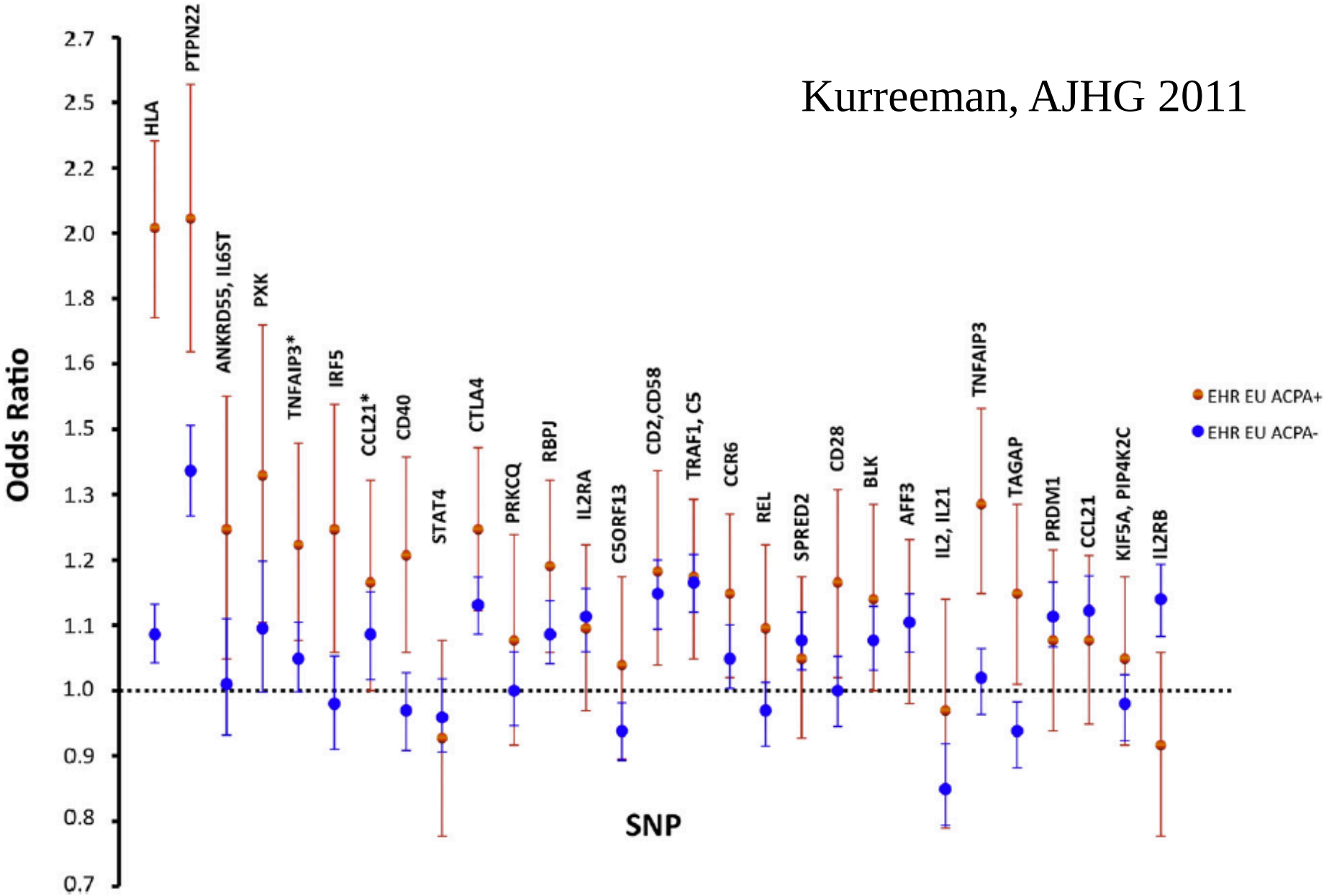


Figure 4. Overlap of Odds Ratio and 95% Confidence Intervals between European ACPA+ and ACPA− Subsets from the EHR Cohort

Early Results

- Validation of Prior GWAS
- Extension of GWAS
- PheWAS

Turning EDGR into Partnership

Kohane et al, Science, 2007

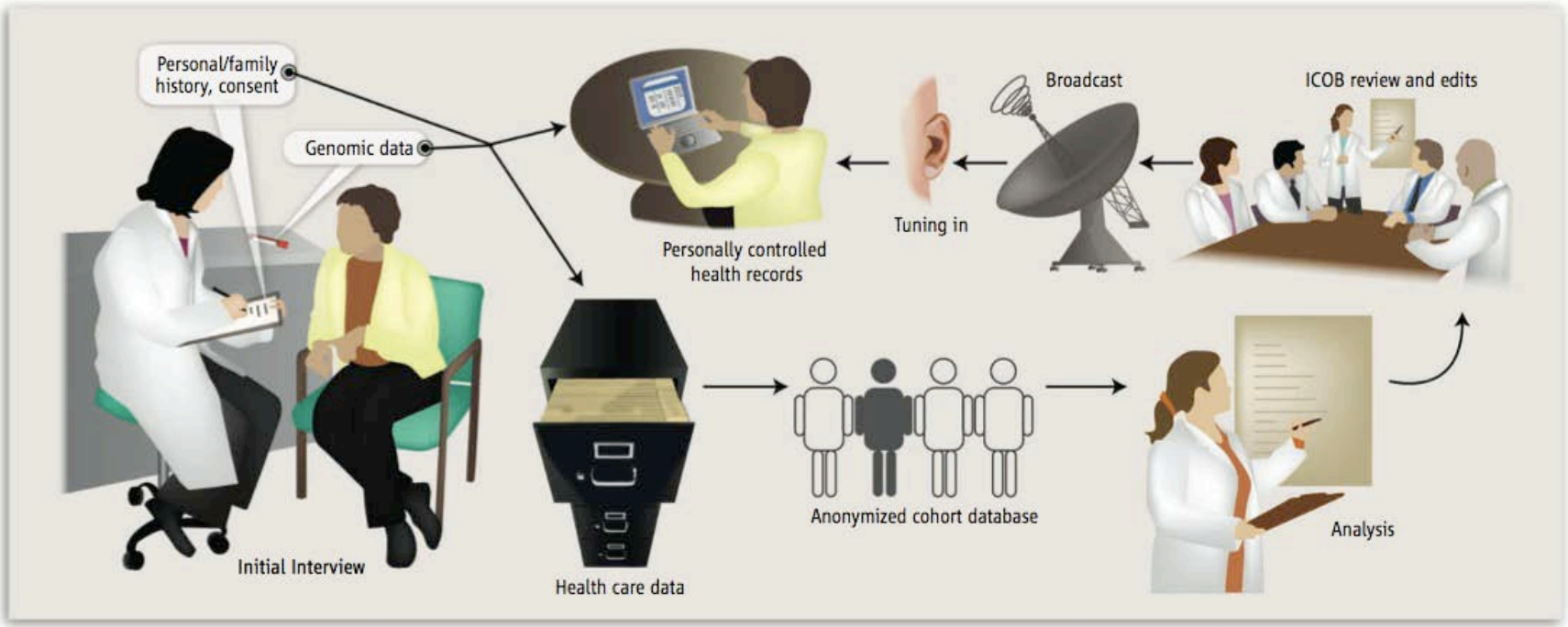
POLICYFORUM

MEDICINE

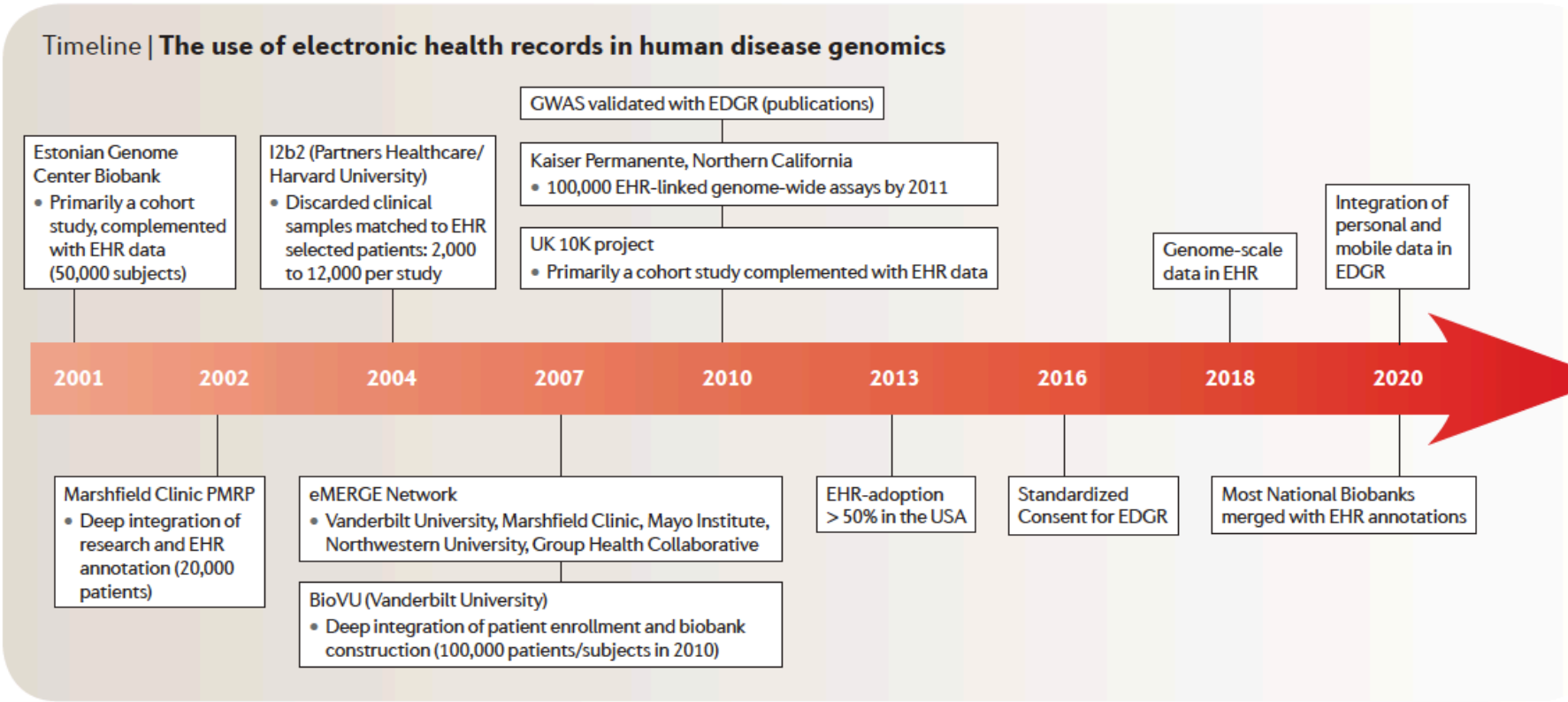
Reestablishing the Researcher-Patient Compact

Well-intentioned regulations protecting privacy are denying important information to patient subjects. Advances in information technology mean that a better approach to clinical research is possible.

Isaac S. Kohane,[1,2,3]* Kenneth D. Mandl,[1,2,3] Patrick L. Taylor,[2,4] Ingrid A. Holm,[2,5] Daniel J. Nigrin,[1,2,3] Louis M. Kunkel[2,5,6]



Timeline


Timeline | The use of electronic health records in human disease genomics
GWAS validated with EDGR (publications)
Estonian Genome Center Biobank
• Primarily a cohort study, complemented with EHR data (50,000 subjects)
I2b2 (Partners Healthcare/ Harvard University)
• Discarded clinical samples matched to EHR selected patients: 2,000 to 12,000 per study
Kaiser Permanente, Northern California
• 100,000 EHR-linked genome-wide assays by 2011
UK 10K project
• Primarily a cohort study complemented with EHR data
Genome-scale data in EHR
Integration of personal and mobile data in EDGR
2001
2002
2004
2007
2010
2013
2016
2018
2020
Marshfield Clinic PMRP
• Deep integration of research and EHR annotation (20,000 patients)
eMERGE Network
• Vanderbilt University, Marshfield Clinic, Mayo Institute, Northwestern University, Group Health Collaborative
BioVU (Vanderbilt University)
• Deep integration of patient enrollment and biobank construction (100,000 patients/subjects in 2010)
EHR-adoption > 50% in the USA
Standardized Consent for EDGR
Most National Biobanks merged with EHR annotations