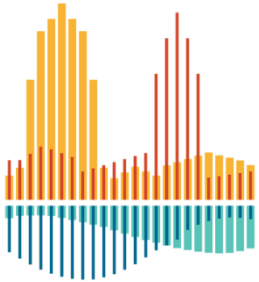
Interdisciplinary Collaborations and EHRs: The Electronic Medical Records and Genomics (eMERGE) network



INSTITUTE FOR COMPUTATIONAL BIOLOGY

SCHOOL OF MEDICINE
CASE WESTERN RESERVE UNIVERSITY

October 30, 2015

Dana C. Crawford, PhD
Associate Professor
Epidemiology and Biostatistics
Institute for Computational Biology

Dana C. Crawford, PhD

Graduated from Collierville High School in Tennessee

Vanderbilt University (BS in Molecular Biology)

Emory University (PhD, Genetics and Molecular Biology)

CDC (EIS Fellow 2000-2002)

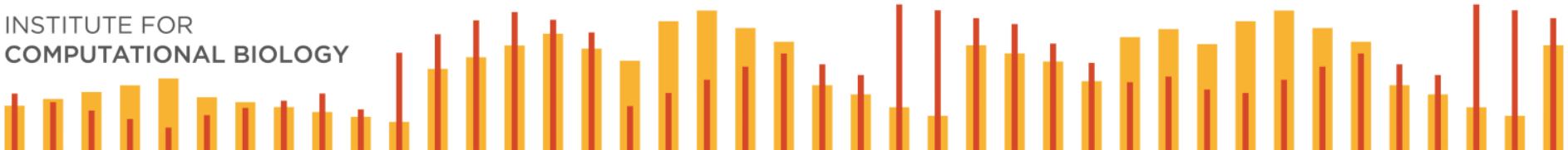
University of Washington (Senior Fellow in Genome Sciences)

Vanderbilt University [Assistant (2006)and Associate Professor (2012)]

SCHOOL OF MEDICINE
CASE WESTERN RESERVE UNIVERSITY

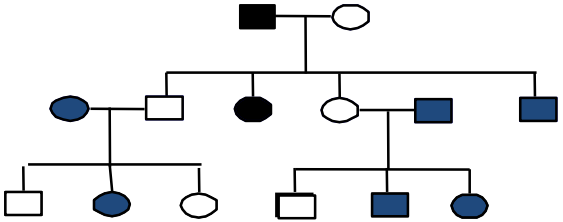
2014 - present

INSTITUTE FOR
COMPUTATIONAL BIOLOGY



Genomic Discovery Study Designs

Linkage Studies in Families

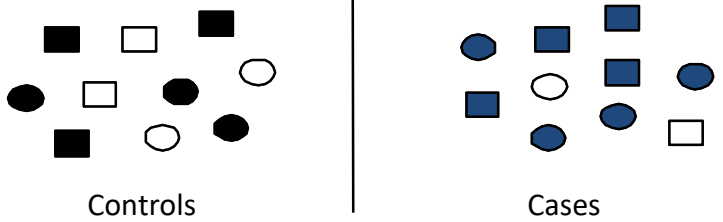


"Simple" Inheritance

Single Gene

Rare variants

Genetic Association Studies



Complex Inheritance

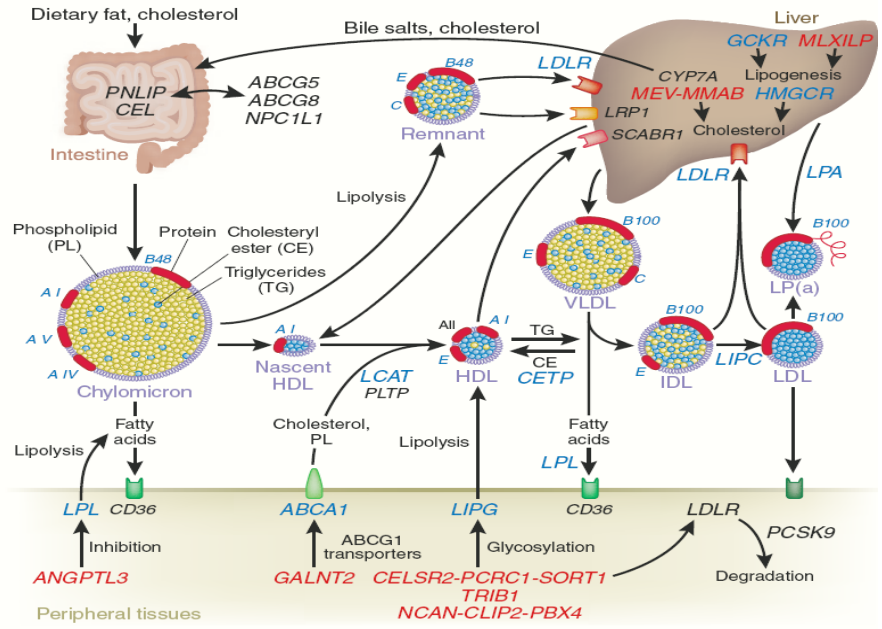
Multiple Genes

Common Variants

INSTITUTE FOR
COMPUTATIONAL BIOLOGY

Study Designs

Candidate gene/pathway



Lusis et al. (2008) *Nat Genet* 40:129-130

Genome-wide association study (GWAS)

Illumina Infinium assay

INSTITUTE FOR
COMPUTATIONAL BIOLOGY

GWAS as of 2015:

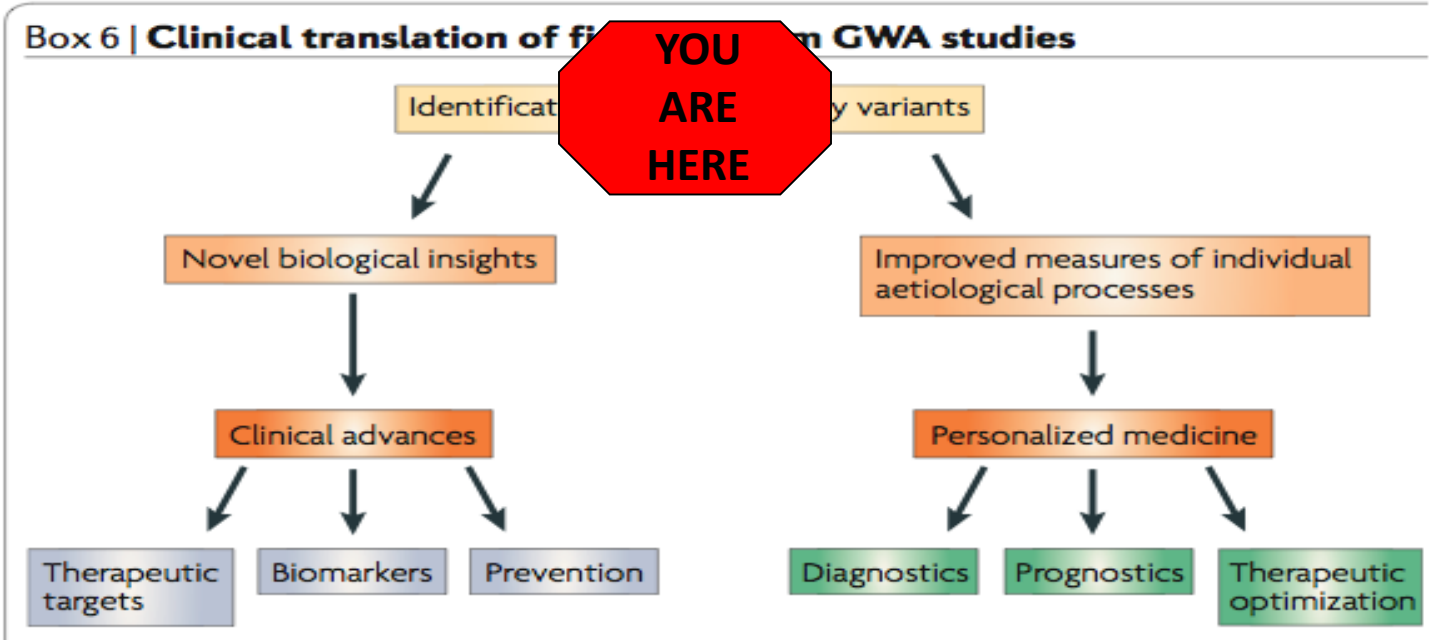
2,111 publications

15,396 associated SNPs (10^{-5})

https://www.ebi.ac.uk/gwas/

INSTITUTE FOR
COMPUTATIONAL BIOLOGY

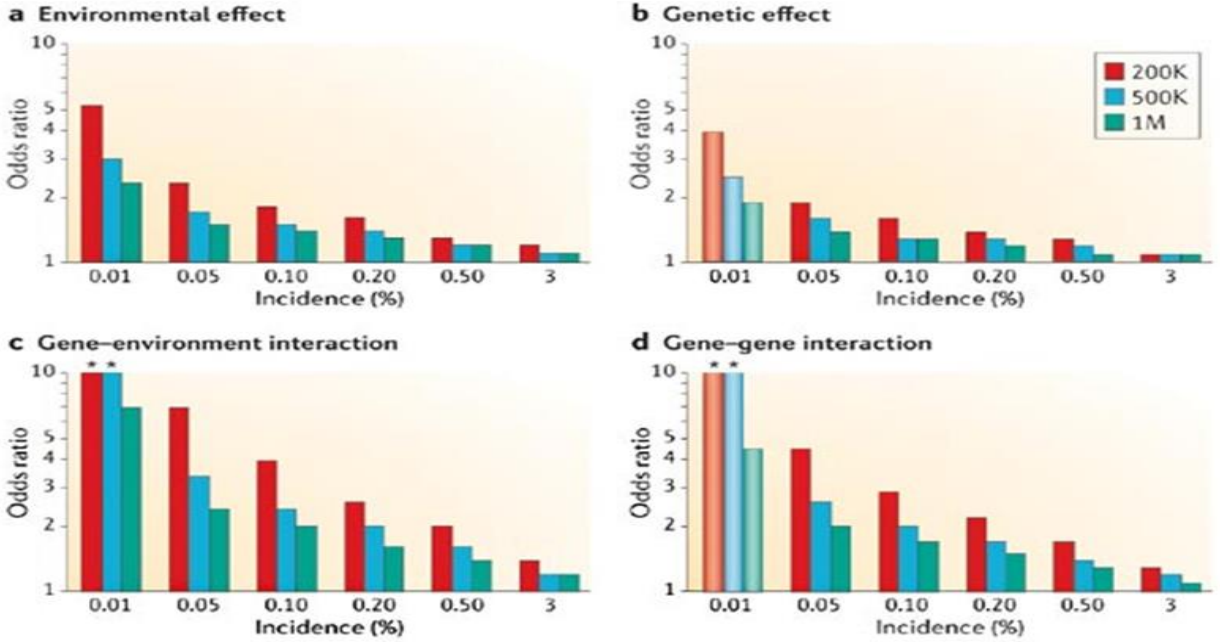
Beyond GWAS



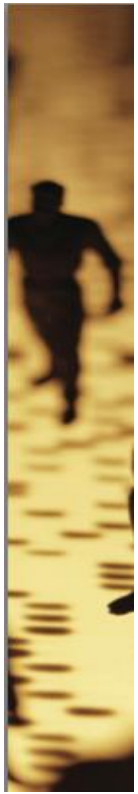
McCarthy et al (2008) *Nat Rev Genet* 9:356-369

INSTITUTE FOR
COMPUTATIONAL BIOLOGY

Genetic Association Studies Require Large Sample Sizes



Manolio et al. (2006) *Nat Rev Genet* 7:812–820



insight commentary

The case for a US prospective cohort study of genes and environment

Francis S. Collins

But…..

- Already several existing cohorts
- Very expensive

Collins (2004) *Nature* 429:475-477

American Journal of Epidemiology
Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health 2012.

Vol. 175, No. 9
DOI: 10.1093/aje/kwr453
Advance Access publication:
March 12, 2012

Commentary

New Models for Large Prospective Studies: Is There a Better Way?

Teri A. Manolio*, Brenda K. Weis, Catherine C. Cowie, Robert N. Hoover, Kathy Hudson, Barnett S. Kramer, Chris Berg, Rory Collins, Wendy Ewart, J. Michael Gaziano, Steven Hirschfeld, Pamela M. Marcus, Daniel Masys, Catherine A. McCarty, John McLaughlin, Alpa V. Patel, Tim Peakman, Nancy L. Pedersen, Catherine Schaefer, Joan A. Scott, Timothy Sprosen, Mark Walport, and Francis S. Collins

biobank uk
Improving the health of future generations

Manolio et al (2012) *Am J Epidemiol* 175:859-66

INSTITUTE FOR
COMPUTATIONAL BIOLOGY

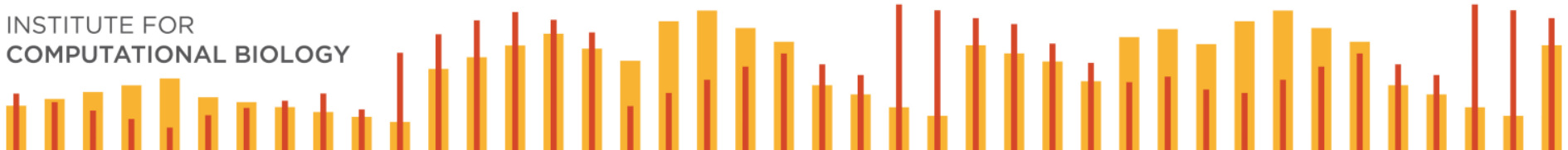
Biobanks Linked to EHRs

- Large and relatively inexpensive compared with cohorts
- Some are population-based
- Some are prospective
- Store more than DNA
- Ethical issues

eMERGE Network

Supported by NHGRI since 2007

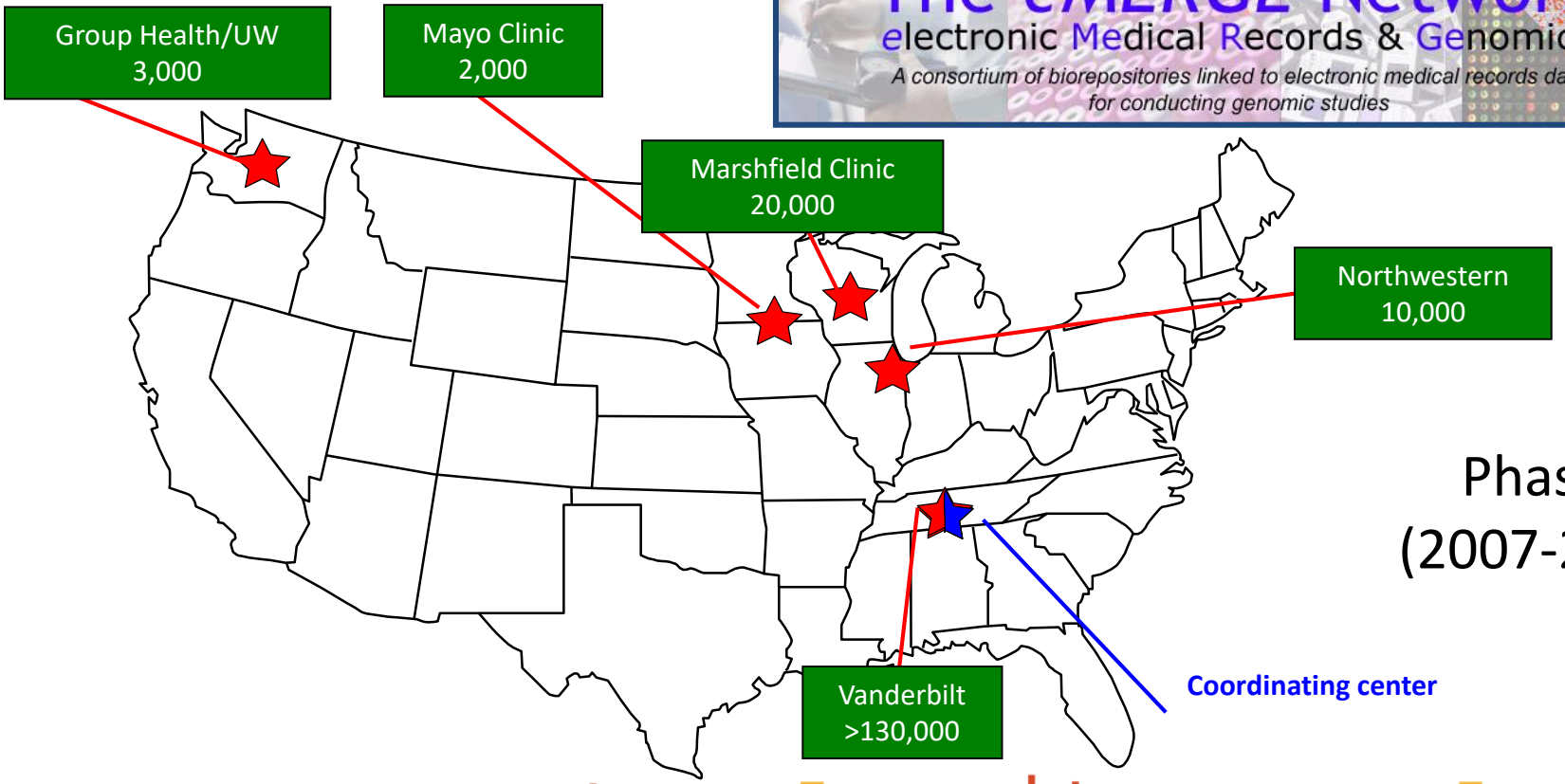
- Define phenotypes and traits using EHRs
- Perform genomic discovery
- Ethics (RROR)
- Clinical translation: assessment and model



The eMERGE Network

electronic Medical Records & Genomics

A consortium of biorepositories linked to electronic medical records data for conducting genomic studies



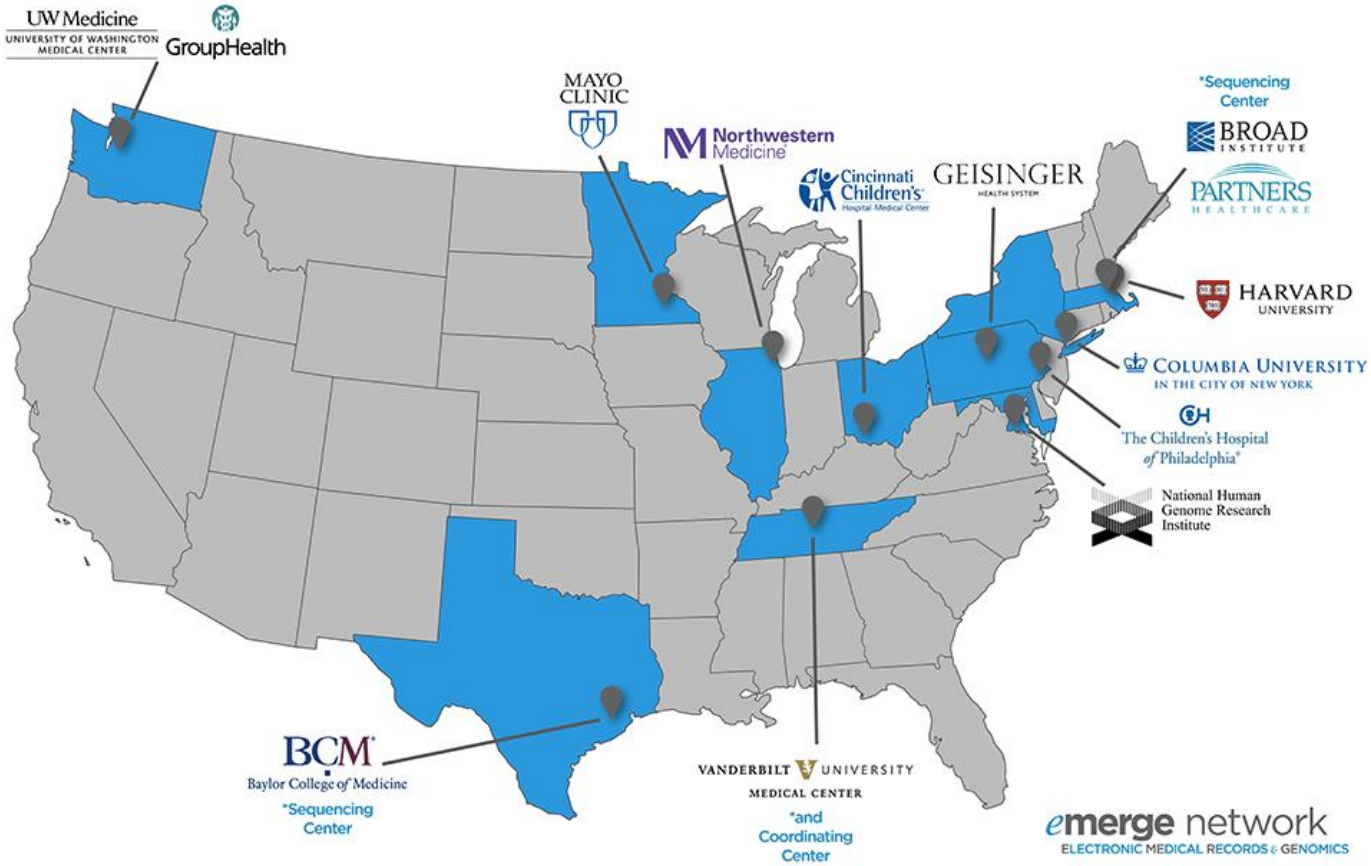
Phase I
(2007-2011)

INSTITUTE FOR
COMPUTATIONAL BIOLOGY

eMERGE Network

electronic medical records & genomics

Phase 2
(2011-2015)



Phase 3
(2015-)

INSTITUTE FOR
COMPUTATIONAL BIOLOGY

eMERGE Network

A Network of Work Groups

Genomics

Informatics (Phenotyping)

eMERGE PGx

Clinical Variant Annotation

Consent and Community Consultation (Consent, Education, Regulation, and Consultation)

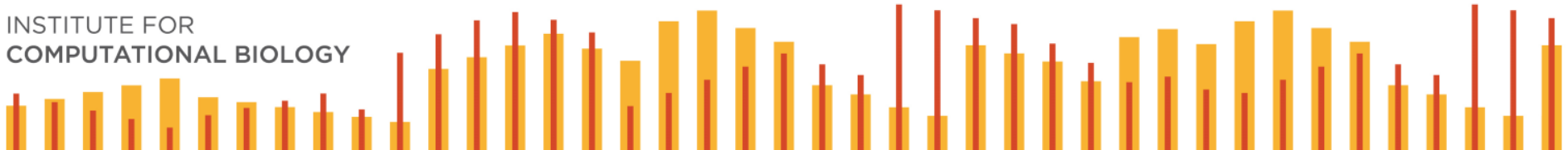
Pediatrics

Return of Results

EHR integration

Outcomes

INSTITUTE FOR
COMPUTATIONAL BIOLOGY

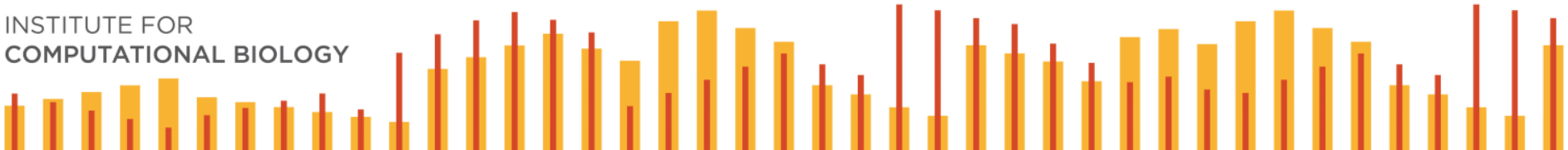


EHRs and Challenges in Research

Variable Data Density

Mean (SD) # of clinical visits	81.8 (107.8)
Range of # of clinical visits	1 to 1,456
Mean (SD) # of ICD9 codes	147.3 (230.4)
Range of # of ICD9 codes	1 to 3,617

Crawford et al (2015) *Hum Hered* 79(3-4):137-46

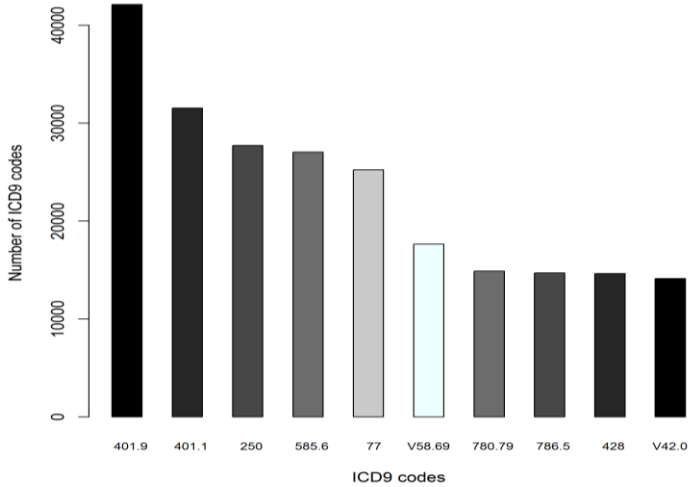


EHRs and Challenges in Research

Clinical versus Population-based

Common ICD-9-CM codes among African Americans (>18 years)

Hypertension (401.9, 401.1)
Diabetes Mellitus (250)
End-stage renal disease (585.6)



Crawford et al (2015) *Hum Hered* 79(3-4):137-46

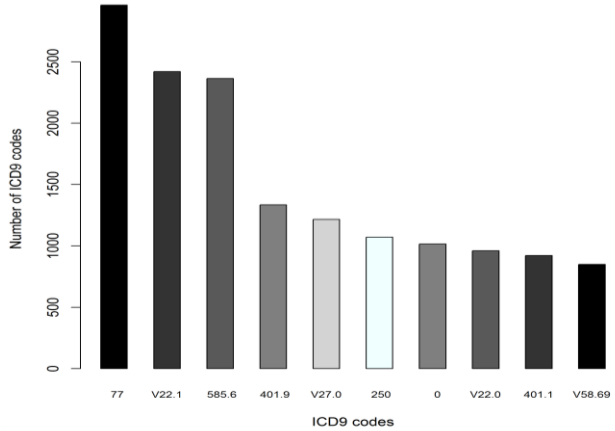
INSTITUTE FOR
COMPUTATIONAL BIOLOGY

EHRs and Challenges in Research

Clinical versus Population-based

Common ICD-9-CM codes among Hispanics (>18 years)

Sequestrectomy (77)
Supervision of other normal pregnancy (v22.1)
End-stage renal disease (585.6)
Hypertension (401.9, 401.1)
Outcome of delivery, single liveborn (V27.0)



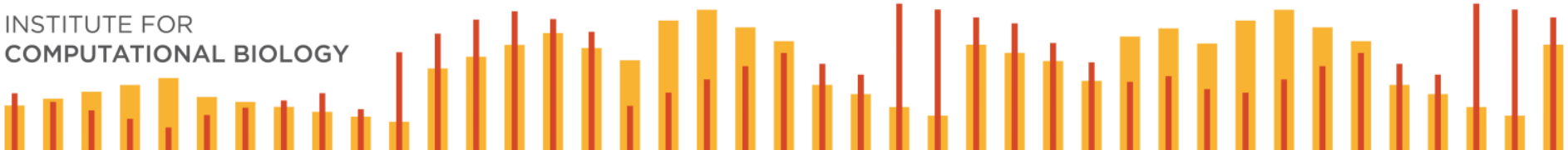
Crawford et al (2015) *Hum Hered* 79(3-4):137-46

INSTITUTE FOR
COMPUTATIONAL BIOLOGY

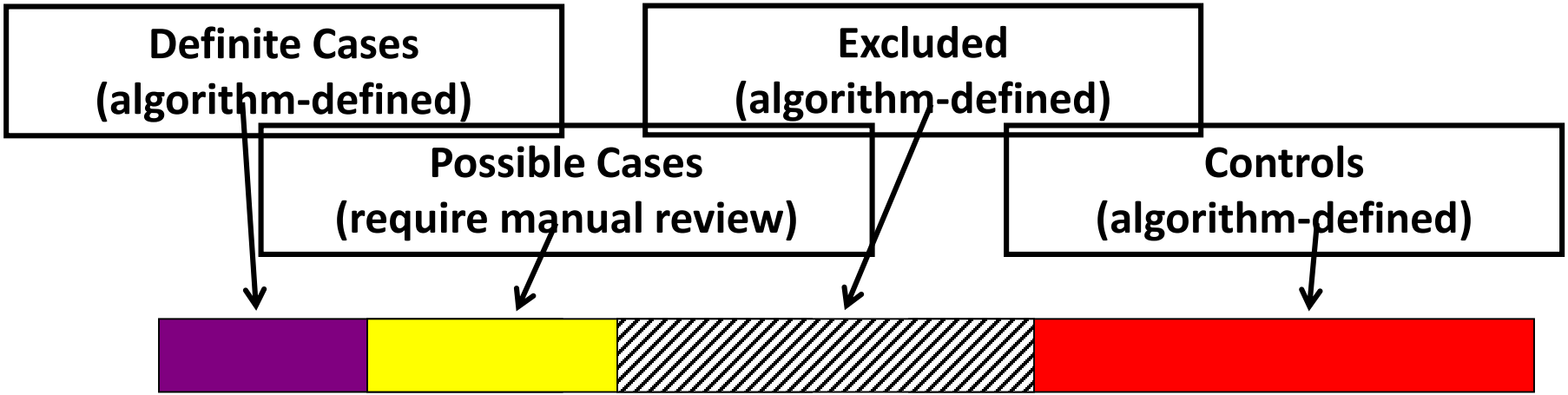
EHRs and Genetic Discovery

- Challenges
  - EHR is not built for research
  - Billing codes can be unreliable in defining cases and controls

- “Demonstration Project”
  - Can the EHR be used to define cases and controls for genetic association studies?
  - Can these EHR-defined cases and controls replicate known genotype-phenotype associations?



General algorithm for determining an EHR phenotype

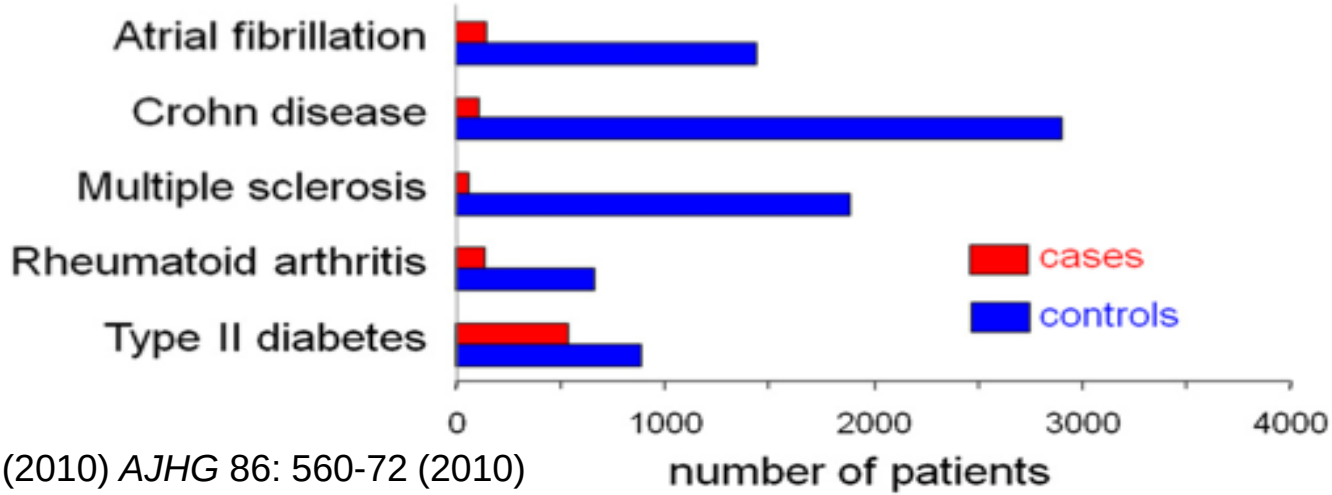


- Billing codes, procedure codes, labs, meds, free text
- Iterative process
- Strive for PPV>95%

ARTICLE

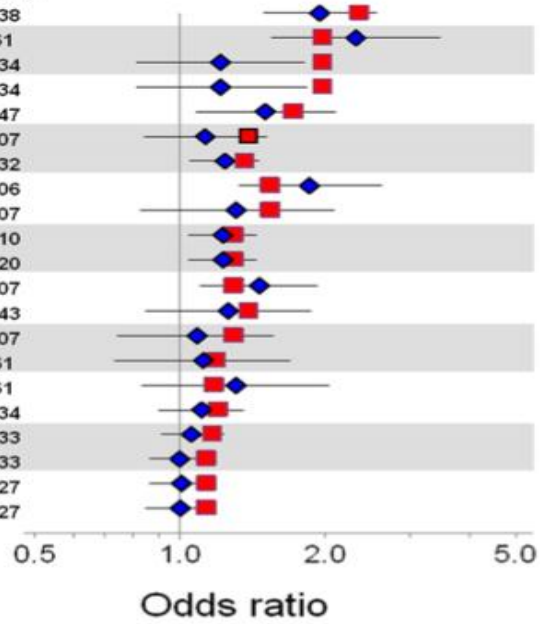
Robust Replication of Genotype-Phenotype Associations across Multiple Diseases in an Electronic Medical Record

Marylyn D. Ritchie,[2,7,9] Joshua C. Denny,[5,6,9] Dana C. Crawford,[2,7] Andrea H. Ramirez,[6] Justin B. Weiner,[6] Jill M. Pulley,[3] Melissa A. Basford,[1,3] Kristin Brown-Gentry,[2] Jeffrey R. Balser,[3,4,8] Daniel R. Masys,[5] Jonathan L. Haines,[2,7] and Dan M. Roden[1,6,8,*]



Ritchie, Denny, Crawford, et al (2010) *AJHG* 86: 560-72 (2010)

disease	marker	gene / region	number needed	number identified
RA	rs6457620	Intergenic Chr. 6	75	138
MS	rs3135388	DRB1*1501	108	61
RA	rs6679677	RSBN1	238	134
RA	rs2476601	PTPN22	238	134
AF	rs2200733	Chr. 4q25	292	147
CD	rs11805303	IL23R	493	107
T2D	rs4506565	TCF7L2	503	532
CD	rs17234657	Chr. 5	513	106
CD	rs1000113	Chr. 5	626	107
T2D	rs12255372	TCF7L2	745	510
T2D	rs12243326	TCF7L2	746	520
CD	rs17221417	NOD2	866	107
AF	rs10033464	Chr. 4q25	1046	143
CD	rs2542151	PTPN22	1104	107
MS	rs2104286	IL2RA	2133	61
MS	rs6897932	IL7RA	2263	61
T2D	rs10811661	CDKN2B	2406	534
T2D	rs8050136	FTO	2569	533
T2D	rs5219	KCNJ11	2792	533
T2D	rs5215	KCNJ11	2908	527
T2D	rs4402960	IGF2BP2	3111	527



Blue: Present study
Red: Previous report

Replicated 8/14 when previously reported odds ratio was >1.25

Ritchie, Denny, Crawford, et al (2010) *AJHG* 86: 560-72

INSTITUTE FOR
COMPUTATIONAL BIOLOGY

Genomic Discovery in eMERGE

- Five biobanks linked to EHRs with GWAS data on ~18K samples

- As a large network, can we
  - Define a phenotype algorithm at one site
  - Deploy algorithm at other sites
  - Collaborate to perform genomic discovery

ARTICLE

Variants Near *FOX1E* Are Associated with Hypothyroidism and Other Thyroid Conditions: Using Electronic Medical Records for Genome- and Phenome-wide Studies

Joshua C. Denny,[1,2,17,*] Dana C. Crawford,[3,4,17] Marylyn D. Ritchie,[1,3,4] Suzette J. Bielinski,[5] Melissa A. Basford,[6] Yuki Bradford,[4] High Seng Chai,[7] Lisa Bastarache,[1] Rebecca Zuvich,[3,4] Peggy Peissig,[8] David Carrell,[9] Andrea H. Ramirez,[2] Jyotishman Pathak,[7] Russell A. Wilke,[2] Luke Rasmussen,[8] Xiaoming Wang,[6] Jennifer A. Pacheco,[14] Abel N. Kho,[10] M. Geoffrey Hayes,[10] Noah Weston,[9] Martha Matsumoto,[7] Peter A. Kopp,[10,14] Katherine M. Newton,[8] Gail P. Jarvik,[11] Rongling Li,[12] Teri A. Manolio,[12] Iftikhar J. Kullo,[13] Christopher G. Chute,[7] Rex L. Chisholm,[14] Eric B. Larson,[9] Catherine A. McCarty,[15] Daniel R. Masys,[1] Dan M. Roden,[2,16] and Mariza de Andrade[7]

Denny, Crawford, et al (2011) *Am J Hum Genet* 89(4):529-42

INSTITUTE FOR
COMPUTATIONAL BIOLOGY

Hypothyroidism

- Symptoms
  - Fatigue and weight gain
- Risk Factors
  - Female sex, increased age, family hx
- Complications
  - Heart problems, goiter, depression, birth defects

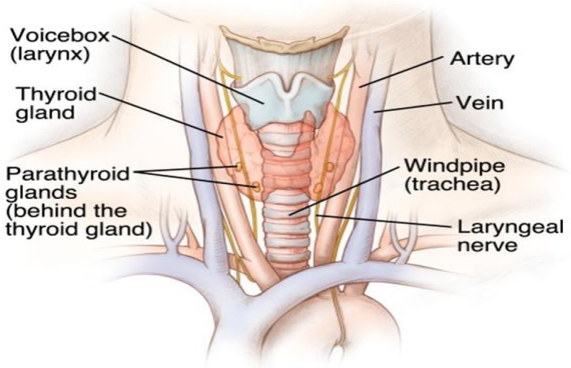
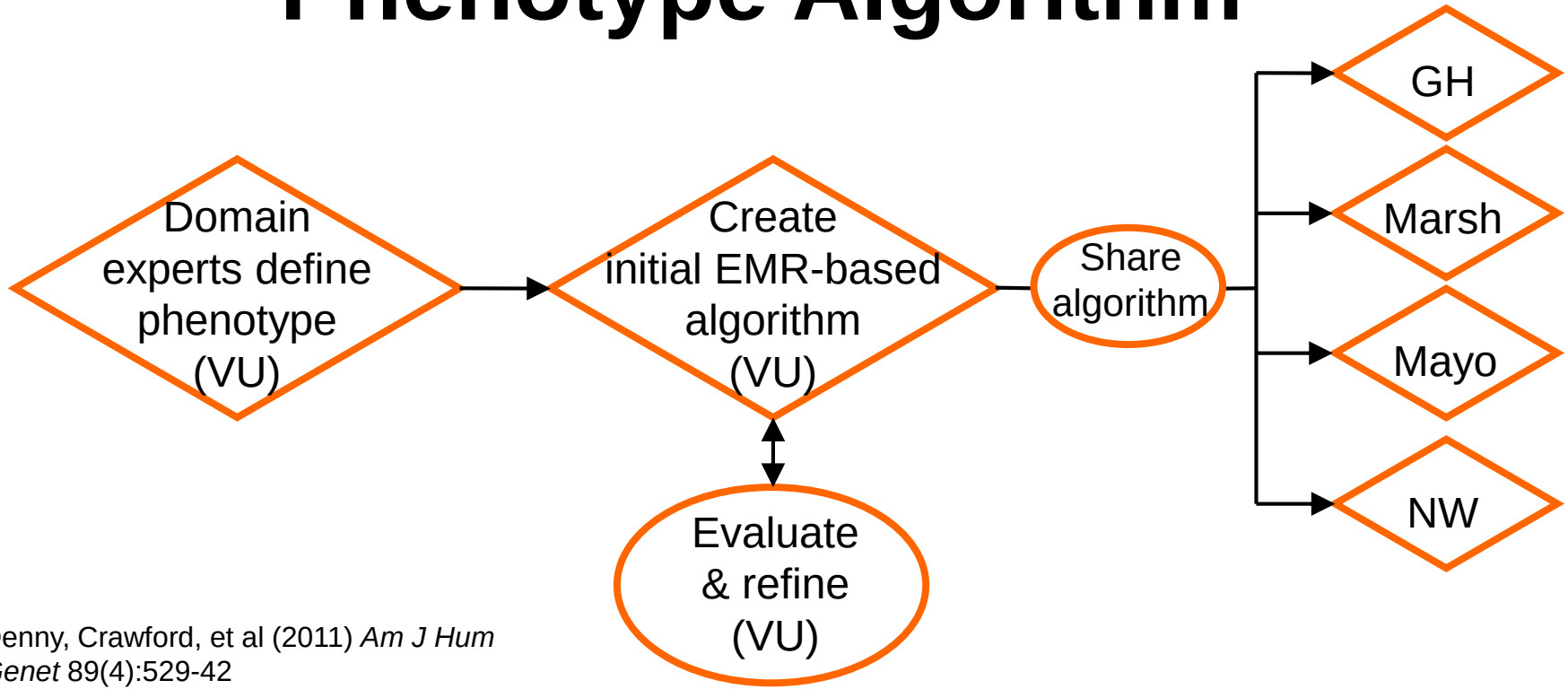


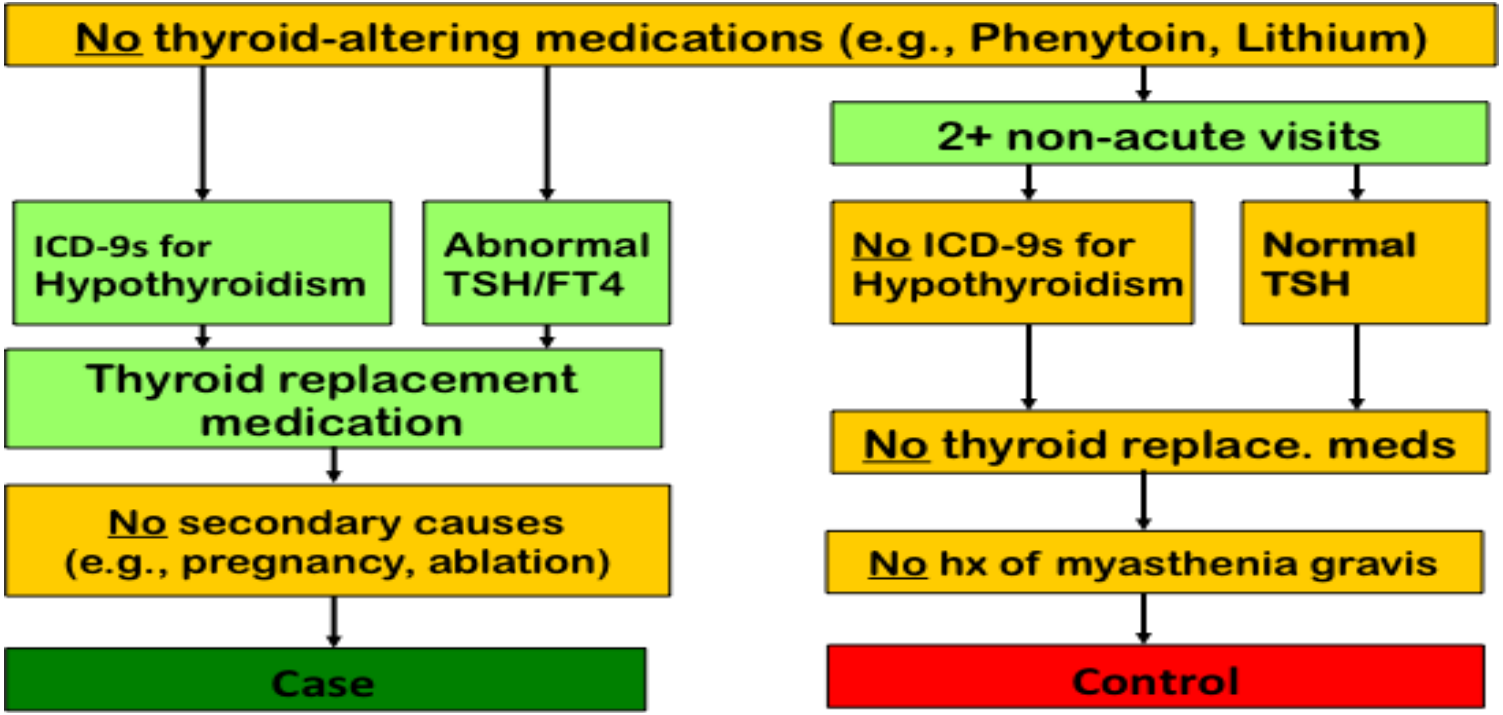
Image: http://www.yalemedicalgroup.org/stw/Page.asp?PageID=STW023081

Phenotype Algorithm



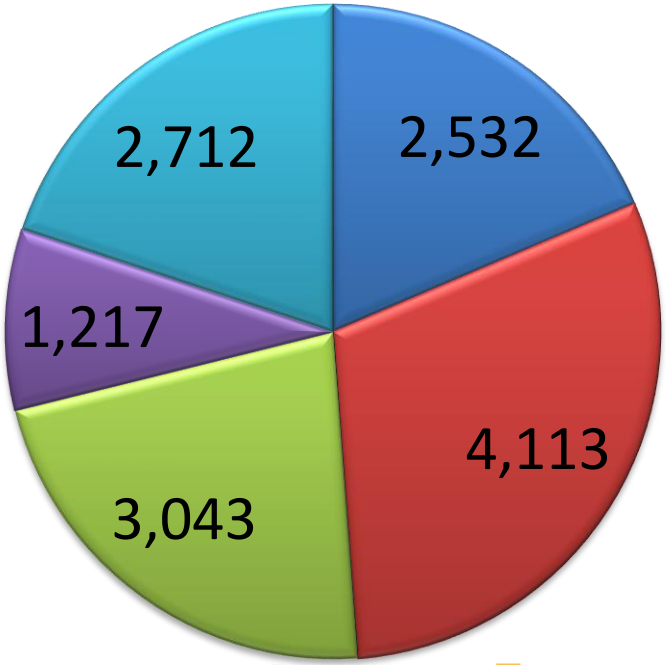
Denny, Crawford, et al (2011) *Am J Hum Genet* 89(4):529-42

Case/Control Algorithm



Sample Size, by Study

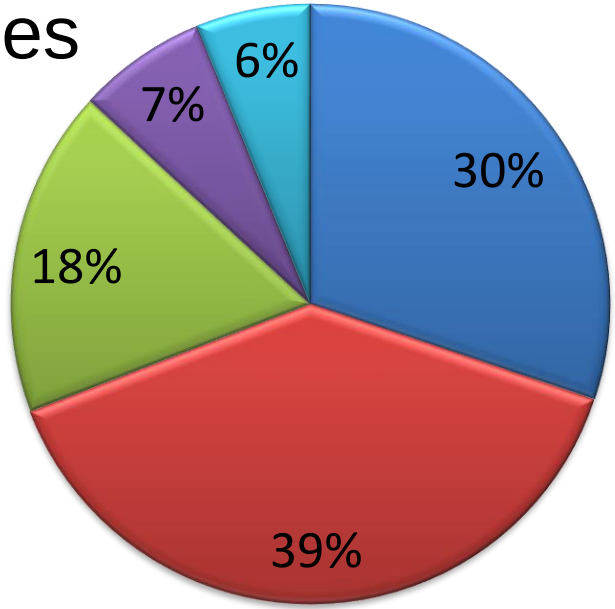
13,617 European Americans



INSTITUTE FOR COMPUTATIONAL BIOLOGY

Case Sample Size, by Study

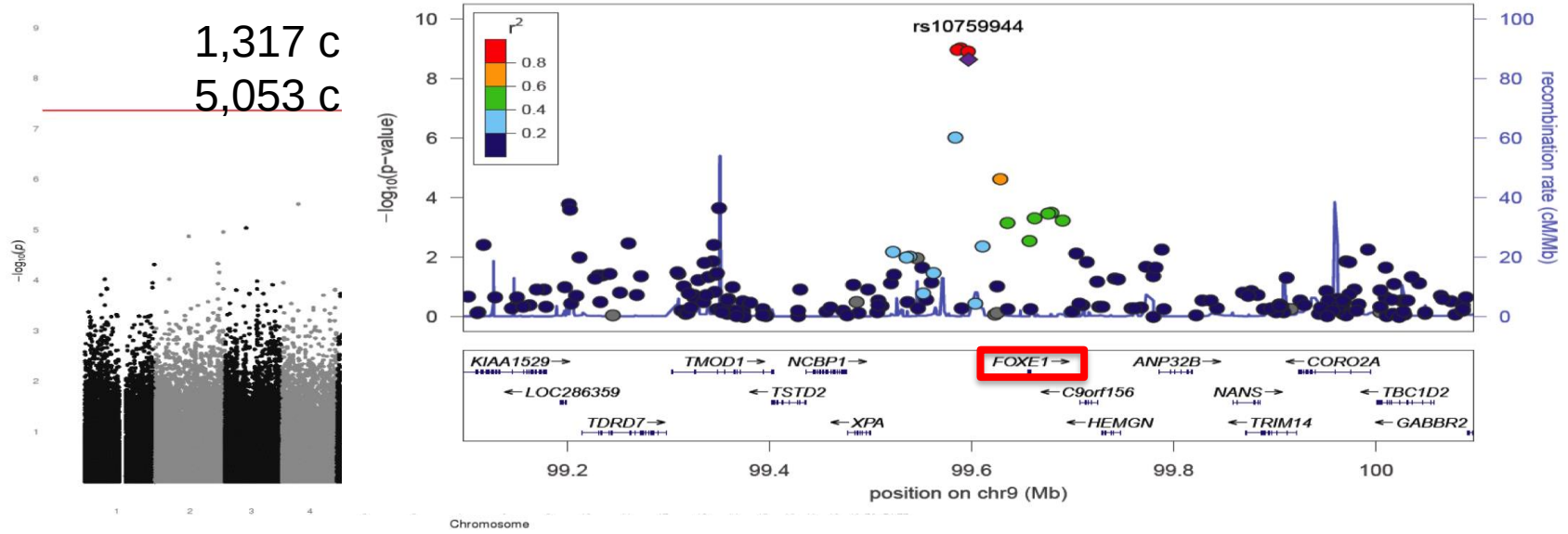
1,317 cases



- Group Health/Dementia
- Marshfield/Cataracts
- Mayo Clinic/PAD
- Northwestern/T2D
- Vanderbilt/Normal cardiac conduction

INSTITUTE FOR
COMPUTATIONAL BIOLOGY

eMERGE Discovery for Hypothyroidism



INSTITUTE FOR
COMPUTATIONAL BIOLOGY

Results

Unmatched Analysis

SNP	Odds ratio	95% CI	P-value
rs7850258	0.74	0.67-0.82	3.93×10^{-9}
rs965513	0.74	0.67-0.82	4.15×10^{-9}
rs925489	0.74	0.67-0.82	4.64×10^{-9}
rs10759944	0.75	0.68-0.83	8.13×10^{-9}

1,317 cases; 5,053 controls; adjusted for decade of birth, sex, site (±PC)

eMERGE and Electronic Phenotyping

PheKB a knowledgebase for discovering phenotypes from electronic medical records

Login | Request Account

Search

Home | Phenotypes | Implementations | Resources | Contact Us

Public Phenotypes

Public phenotypes are believed to be complete and final by their authors. When you are logged in you can view and edit phenotypes in your groups that are non public and in various stages of development.

Login To View Private Group Phenotypes

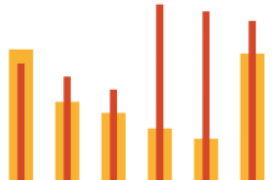
Institution

Beth Israel Deaconess Medical Center
CHOP
Cincinnati Children's Hospital Medical Center
Columbia University
Geisinger Health System
Group Health and University of Washington
Group Health Cooperative
Marshfield Clinic Research Foundation
Mayo Clinic

Apply

Title	Institution	Data Modalities and Methods Used	Owner Phenotyping Groups	View Groups	Has new content	Status
Abdominal Aortic Aneurysm (AAA)	Geisinger Health System	CPT Codes, ICD 9 Codes, Vital Signs	eMERGE Geisinger Group	eMERGE Geisinger Group, eMERGE Phenotype WG		Final
ACE Inhibitor (ACE-I) induced cough	Vanderbilt University	CPT Codes, ICD 9 Codes, Medications, Natural Language Processing	eMERGE Vanderbilt Group	eMERGE Phenotype WG		Final
ADHD phenotype algorithm	CHOP	ICD 9 Codes, Medications, Natural Language Processing	eMERGE CHOP Group	eMERGE Phenotype WG		Final
Appendicitis	Cincinnati Children's Hospital Medical Center	CPT Codes, ICD 9 Codes, Medications, Natural Language Processing	eMERGE CCHMC/BCH Group	eMERGE Phenotype WG		Final
Atrial Fibrillation - Demonstration Project	Vanderbilt University	CPT Codes, ICD 9 Codes, Natural Language Processing	Vanderbilt - SD/RD Group	Vanderbilt - SD/RD Group		Final
Autism	Cincinnati Children's Hospital Medical	ICD 9 Codes, Medications, Natural Language	eMERGE CCHMC/BCH Group	eMERGE Phenotype WG		Final

https://phekb.org/



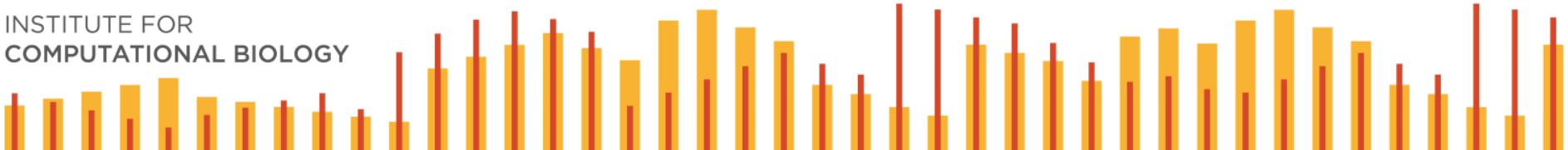
eMERGE Network

electronic medical records & genomics

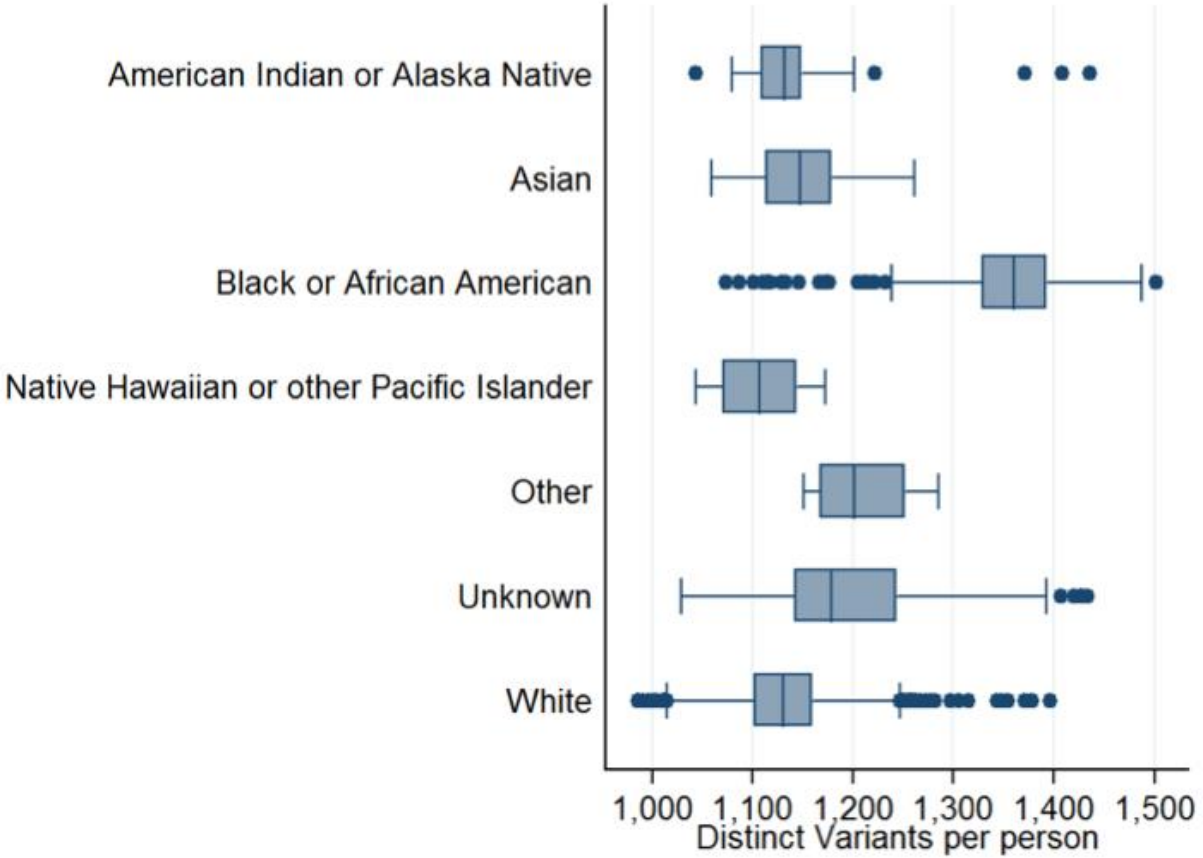
- Network-wide genomic data, eMERGE II
  ~87K samples representing diverse pediatric and adults

- Focused on pathway towards precision medicine

- How many patients have “actionable” genetic variants?
  Sequence patients for 84 pharmacogenes
  Make catalog available publically (SPHINX)

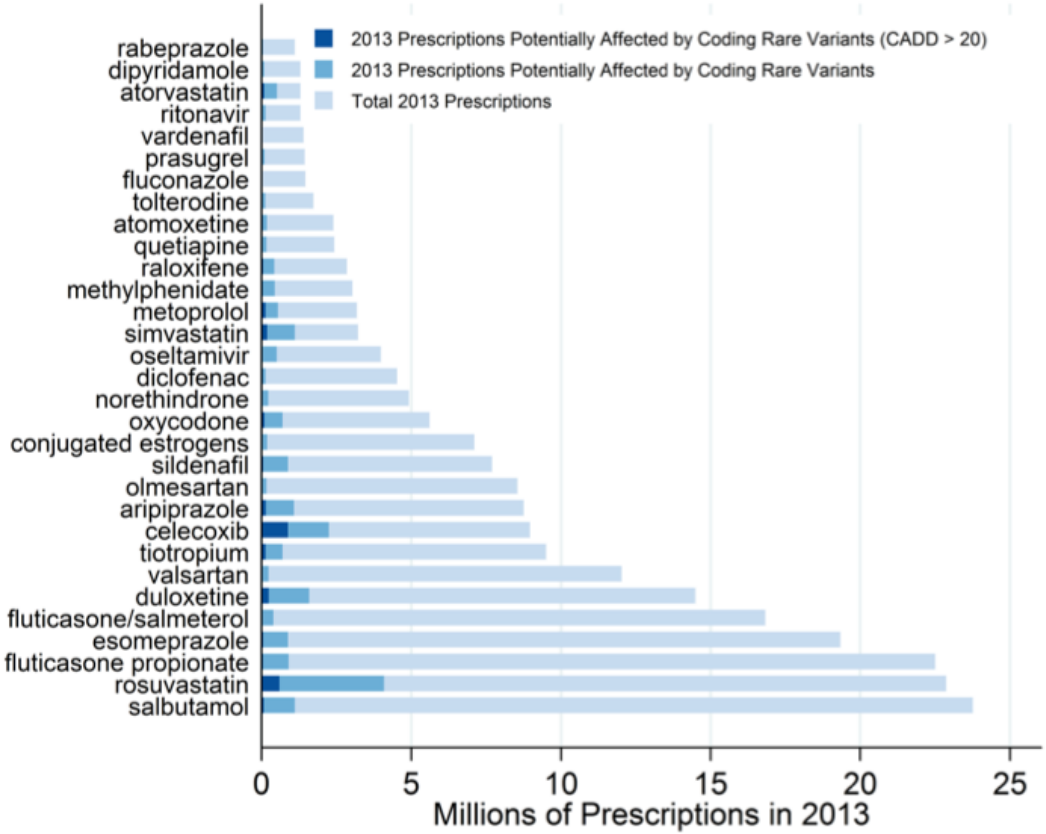


Almost 100% of patients sequenced had at least one actionable variant!



Bush, Crosslin, et al (submitted)

Nearly 75% of ***prescriptions*** were impacted by rare missense variants within pharmacogenes in 2013!



Bush, Crosslin, et al (submitted)

INSTITUTE FOR
COMPUTATIONAL BIOLOGY

www.emergesphinx.org/

INSTITUTE FOR
COMPUTATIONAL BIOLOGY

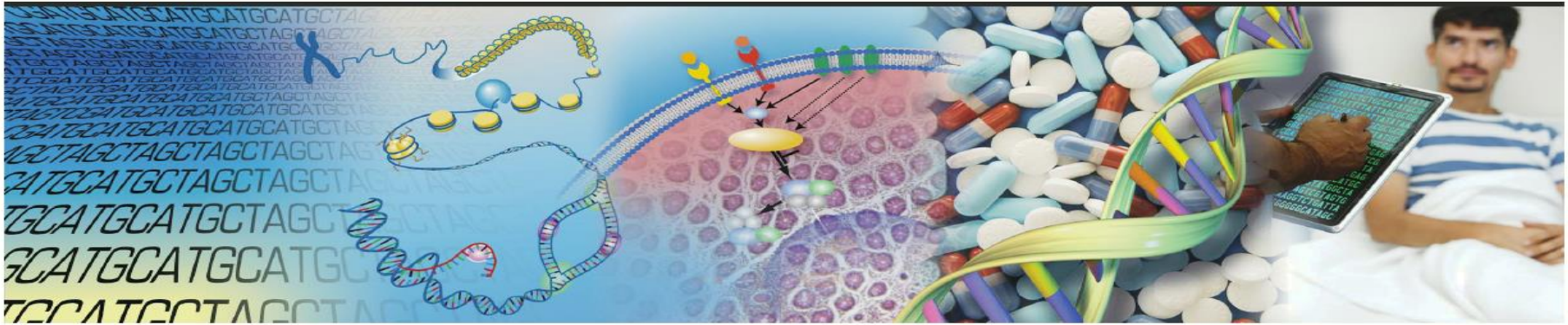
Understanding the structure of genomes

Understanding the biology of genomes

Understanding the biology of disease

Advancing the science of medicine

Improving the effectiveness of healthcare



eMERGE I, II, III

Green et al. (2011) *Nature* 470:204-13

INSTITUTE FOR
COMPUTATIONAL BIOLOGY

Large (1 M)

Linked to EHR and payer databases

Oversample subgroups

Recontact for return of results

Wide age range

Sophisticated lifestyle data collection

Rare and common disease

The Precision Medicine Initiative Cohort Program – Building a Research Foundation for 21st Century Medicine

Precision Medicine Initiative (PMI) Working Group Report to the Advisory Committee to the Director, NIH

September 17, 2015

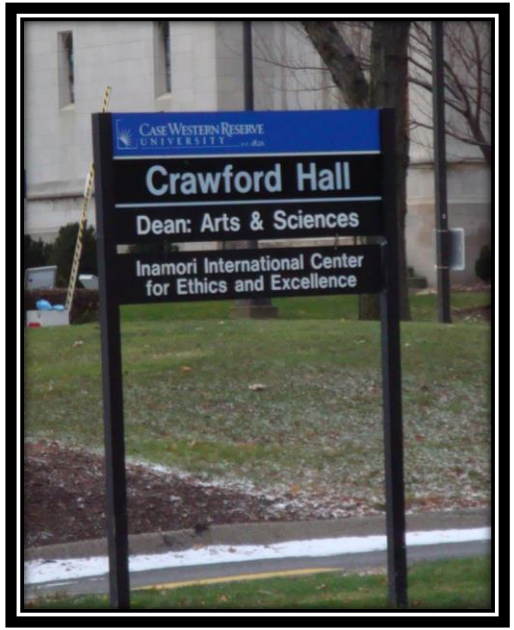
INSTITUTE FOR COMPUTATIONAL BIOLOGY

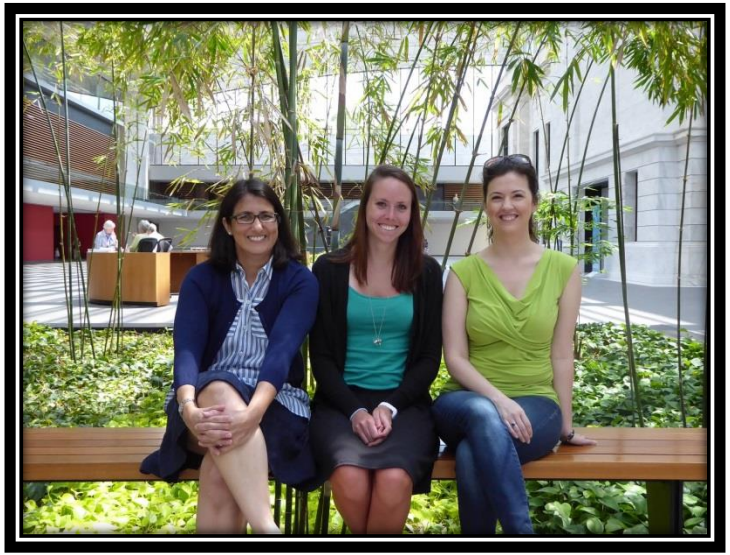
eMERGE Network

electronic medical records & genomics

INST
COM

Crawford Laboratory





Dana C. Crawford, PhD
Sarah Laper, medical school student
Nicole Restrepo, PhD

Brittany Hollister, PhD candidate

INSTITUTE FOR COMPUTATIONAL BIOLOGY