Outcome Data, Links to Electronic Medical Records

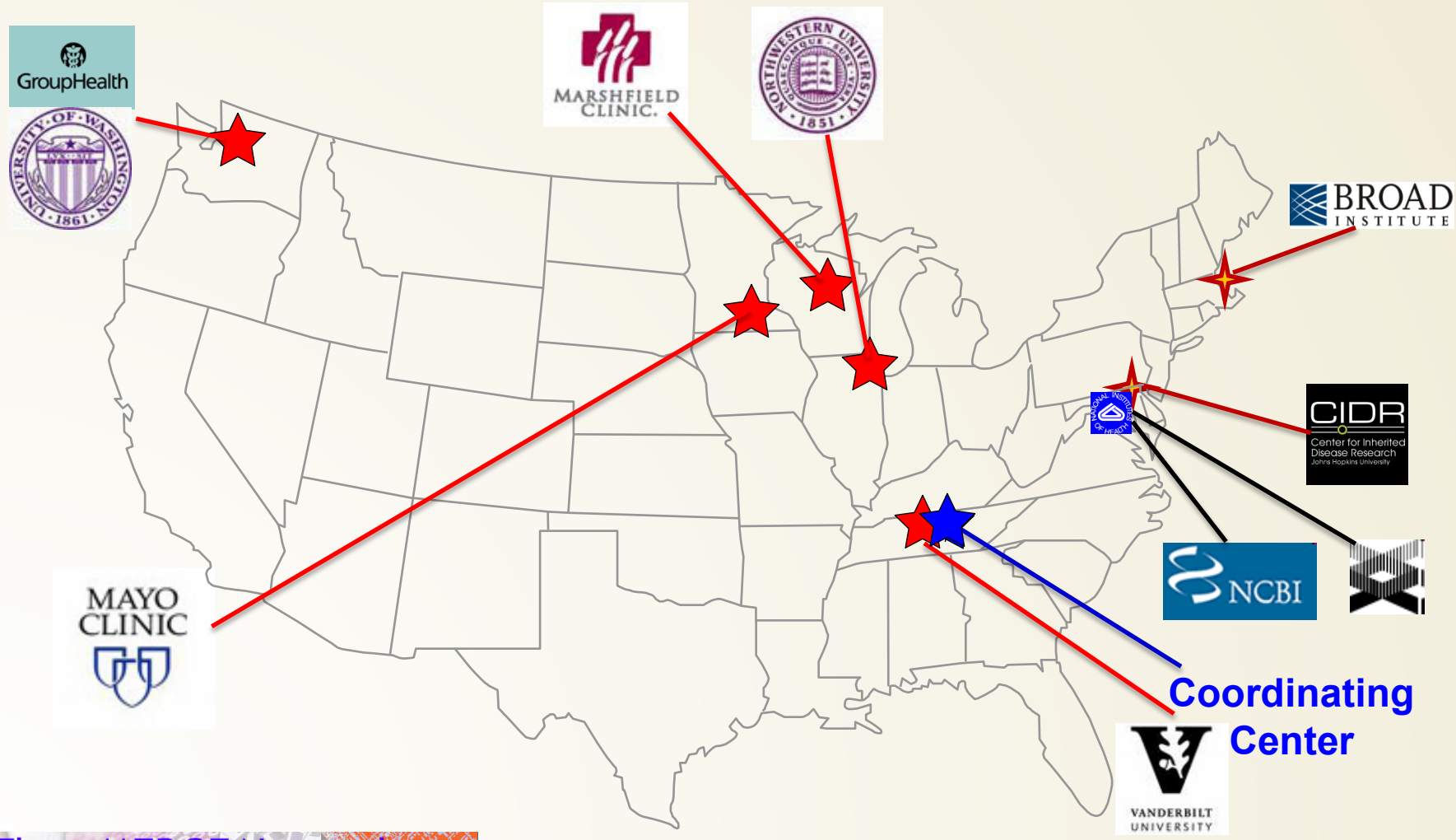
Dan Roden

Vanderbilt University

The *eMERGE* Network

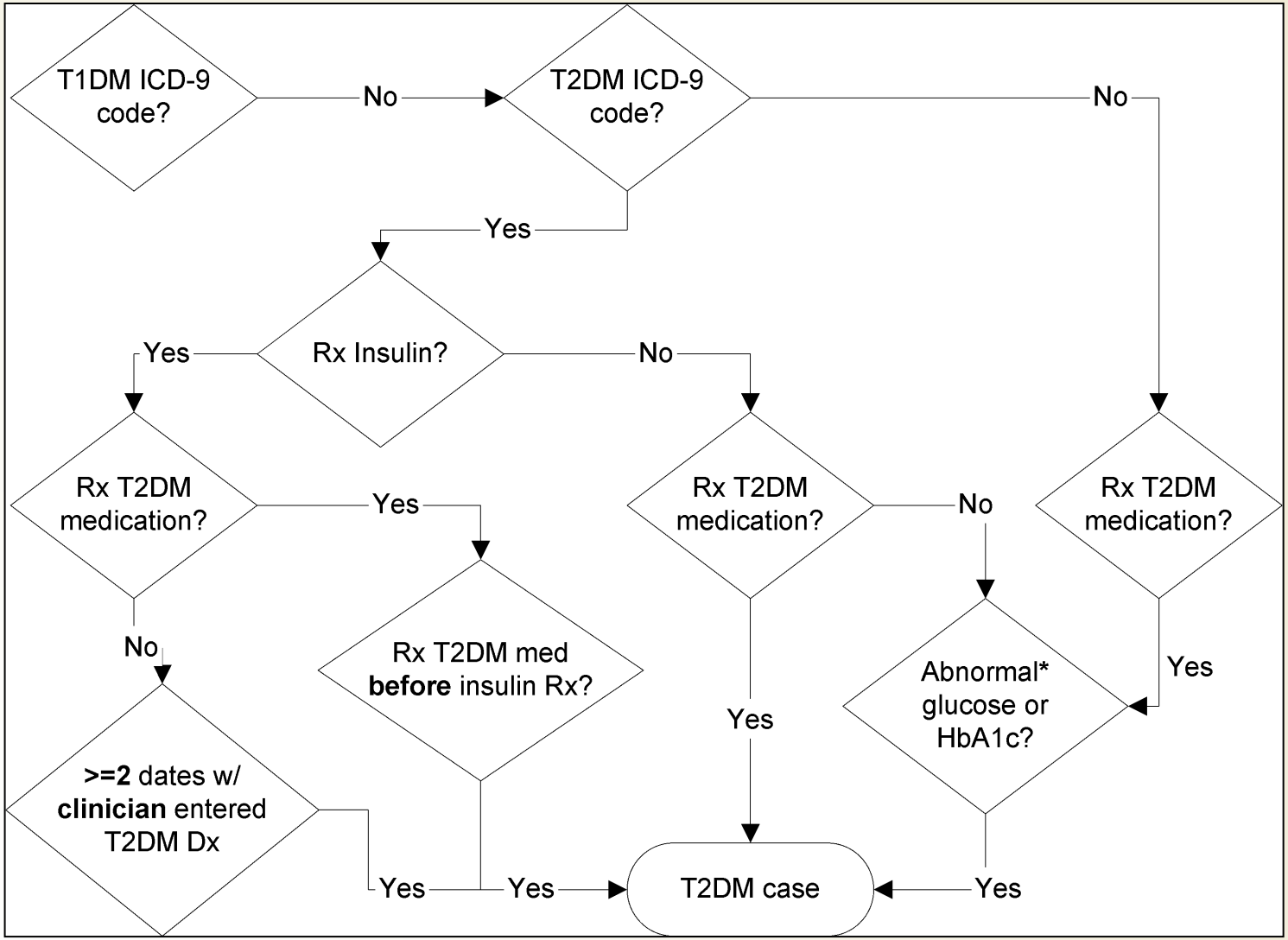
electronic Medical Records & Genomics

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



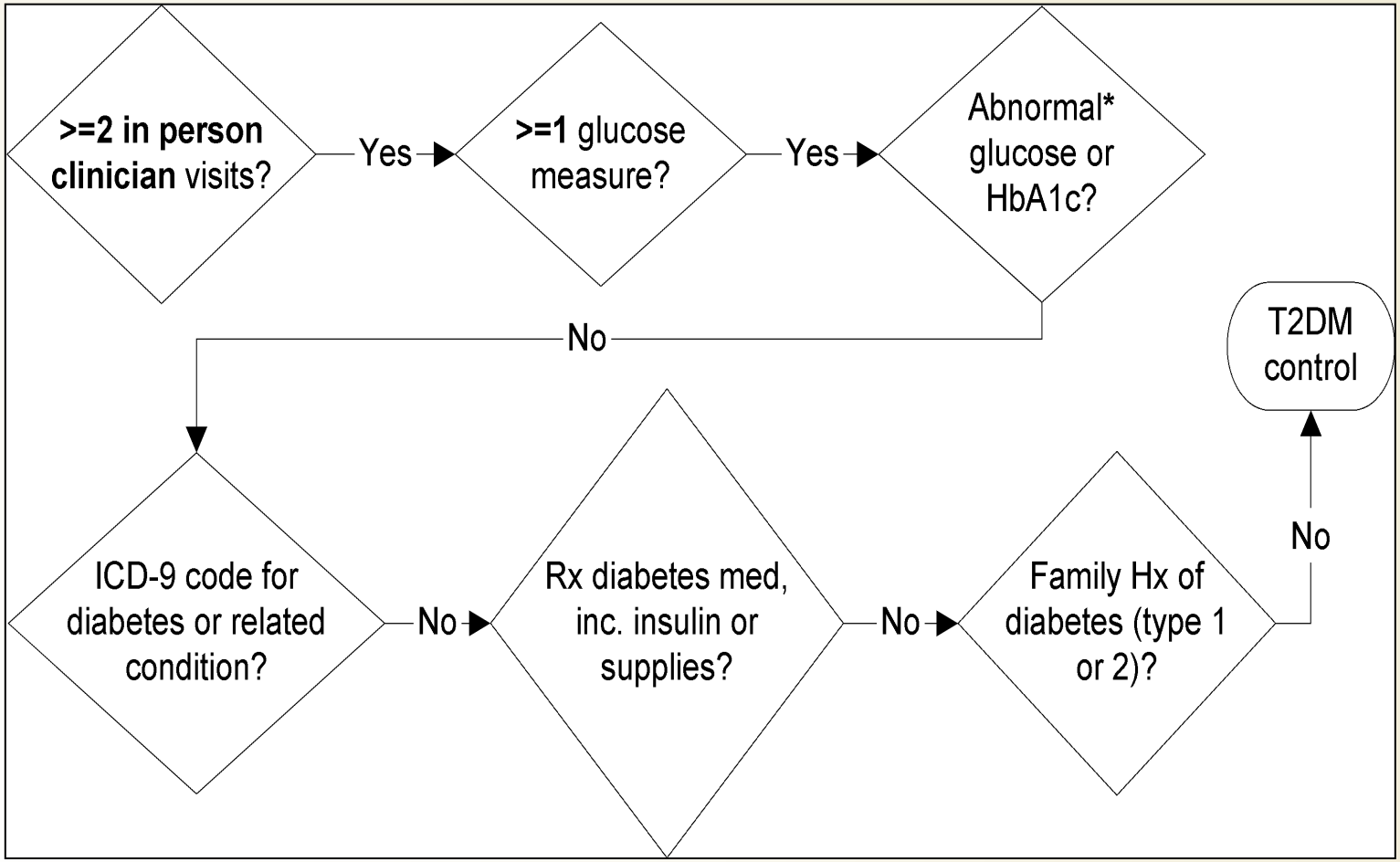
The *eMERGE* Network
electronic Medical Records & Genomics

Type II Diabetes Case Algorithm



*** Abnormal lab=** Random glucose > 200mg/dl, Fasting glucose > 125 mg/dl, or hemoglobin A1c ≥6.5%.

Type II Diabetes Control Algorithm



Type II Diabetes Chart Review

Cases

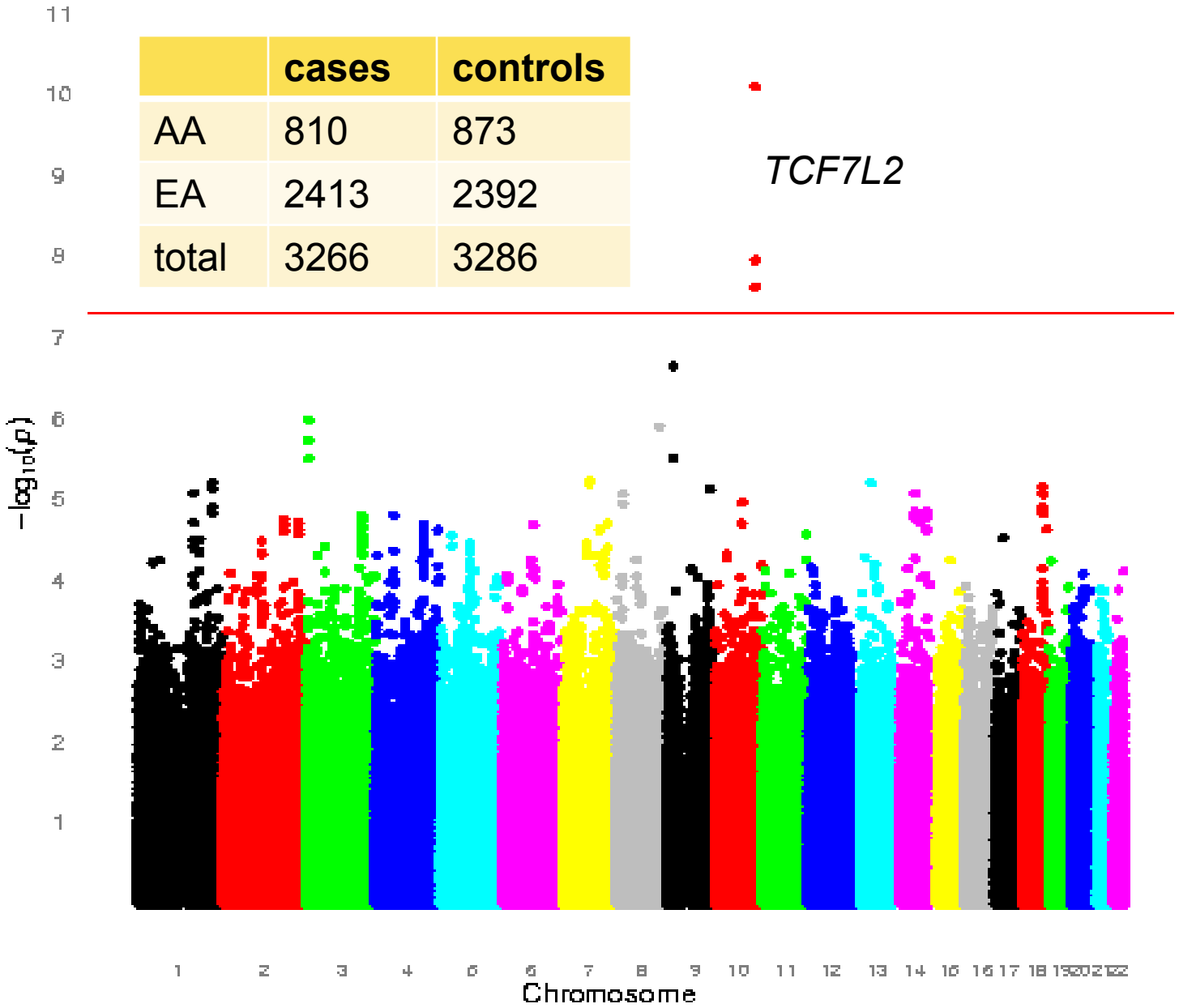
- Blinded clinician review of 100 random charts (50 cases & 50 controls)
- Case PPV = 98%
- Control PPV = 98%

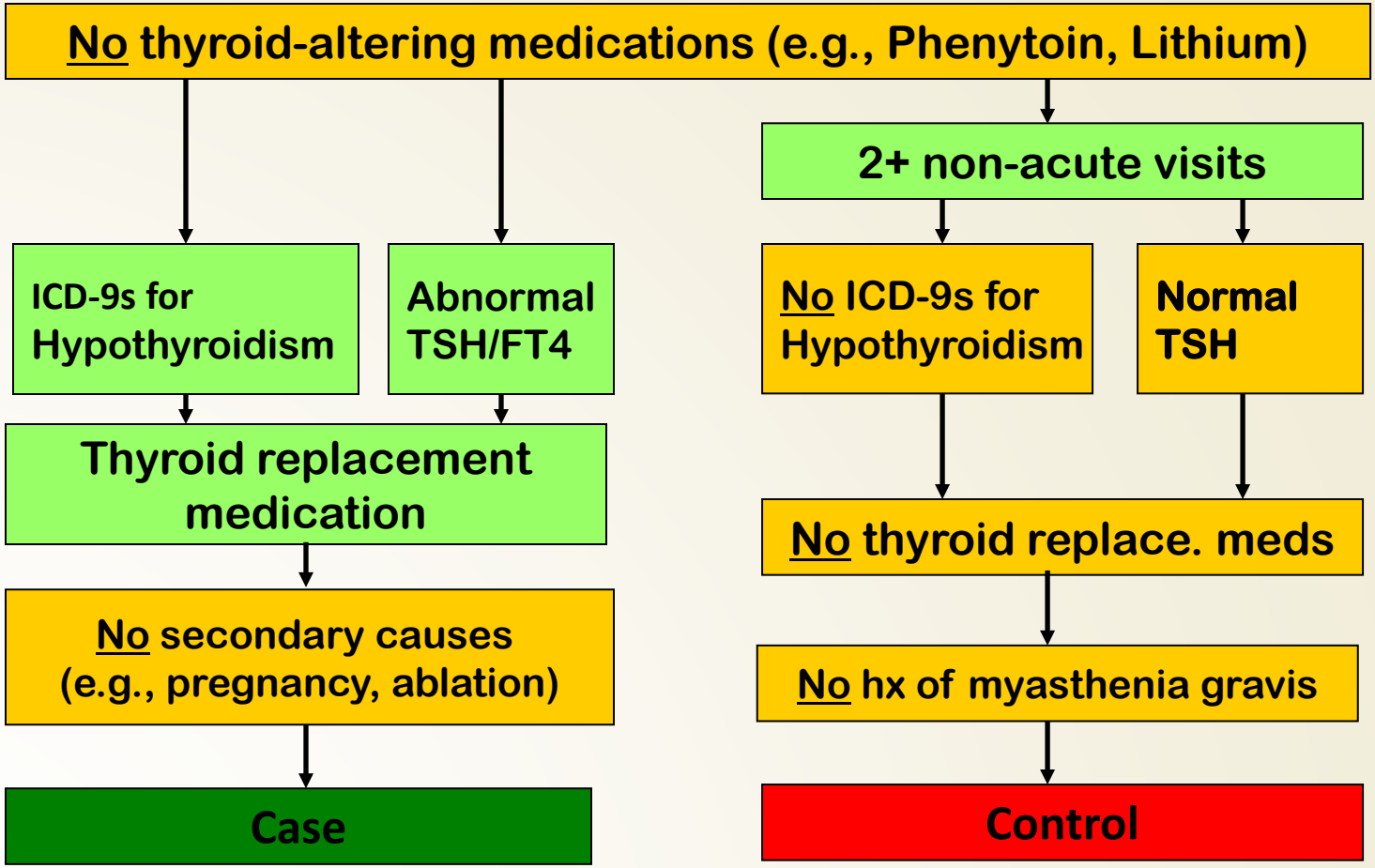
Controls

- Absence of data ≠ absence of condition
  - May simply be inadequate data capture in EMR for individual
  - Difference in comprehensiveness of data capture between sites
  - Valid criteria for controls requires same data elements as case

Imputed T2D Merged (Case/Ctrl) – 98GE SNPs, Adjusted Sex,Age,BMI,PC1,PC

	cases	controls
AA	810	873
EA	2413	2392
total	3266	3286




No thyroid-altering medications (e.g., Phenytoin, Lithium)
2+ non-acute visits
ICD-9s for Hypothyroidism
Abnormal TSH/FT4
No ICD-9s for Hypothyroidism
Normal TSH
Thyroid replacement medication
No thyroid replace. meds
No secondary causes (e.g., pregnancy, ablation)
No hx of myasthenia gravis
Case
Control

Phenotype Algorithm
Validation

Site	EMR-based Cases/Controls	Chart Review Cases/Controls	Case PPV (%)	Control PPV (%)
Group Health	397/1,160	50/50	98	100
Marshfield	514/1,187	50/50	91	100
Mayo Clinic	233/1,884	100/100	82	96
Northwestern	92/470	50/50	98	100
Vanderbilt	81/352	50/50	98	100
All sites (weighted)	1,317/5,053	—	92.4	98.5

Denny et al 2011

PheKB

a knowledgebase for discovering phenotypes from electronic medical records

What is the Phenotype KnowledgeBase?

The reuse of data from electronic medical records (EMRs) and other clinical data systems holds tremendous promise for improving the efficiency and effectiveness of health research. Clinical data in the EMR is a potential source of rich longitudinal data for research, and the recent government efforts to promote the use of EMRs in the clinical setting may further promote the use of such systems in the US healthcare system. As the use of EMRs expands, the demand for usable data from these systems for research has also expanded.

One such effort by the Electronic Medical Records and Genomics Network (eMERGE) has investigated whether data captured through routine clinical care using EMRs can identify disease phenotypes with sufficient positive and negative predictive values for use in genome-wide association studies (GWAS). Most EMRs captured key information (diagnoses, medications, laboratory tests) used to define phenotypes in a structured format; in addition, natural language processing tool has also been shown to improve case identification rates.*

PheKB is an outgrowth of that validation effort and provides a collaborative environment of building and validating electronic phenotype algorithms. On this site you can:

- View existing algorithms
- Enter or create new algorithms
- Collaborate with others to create or review algorithms
- View implementation details for existing algorithms

Phenotype algorithms can be viewed by data modalities or methods used:

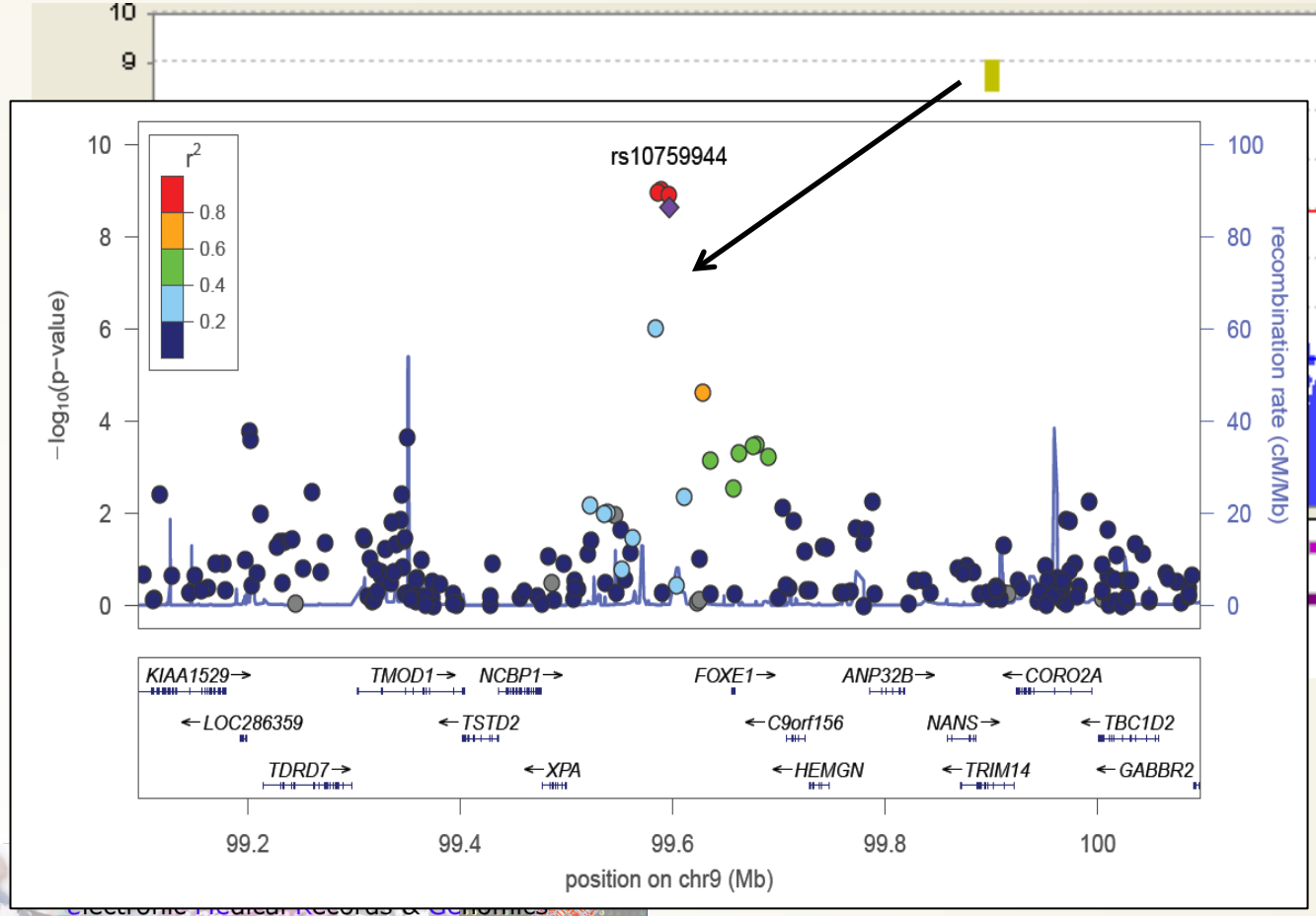
- CPT codes
- ICD 10 codes
- ICD 9 codes
- Laboratories

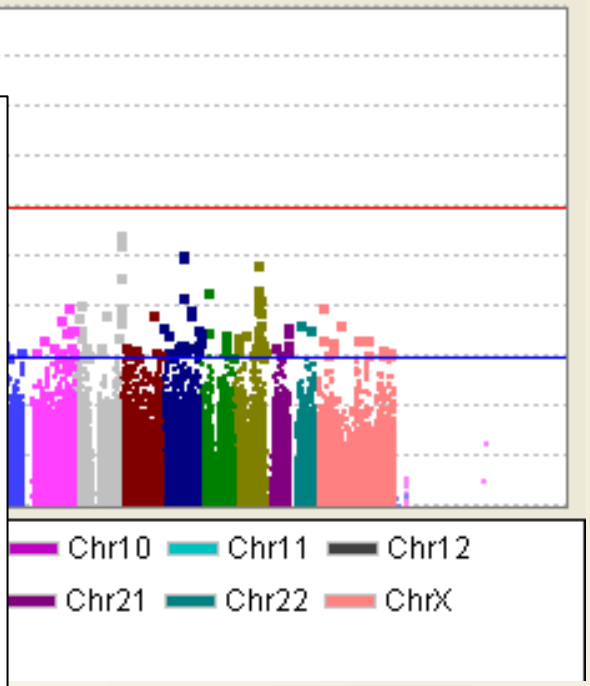
Most Recent Phenotypes

- White Blood Cell Indices
- Type II Diabetes Mellitus
- Red Blood Cells Indices
- Peripheral Arterial Disease
- Lipids

An eMERGE-wide phenotype analyzed with no extra genotyping: hypothyroidism

European Americans (1,306 cases and 5,013 controls)

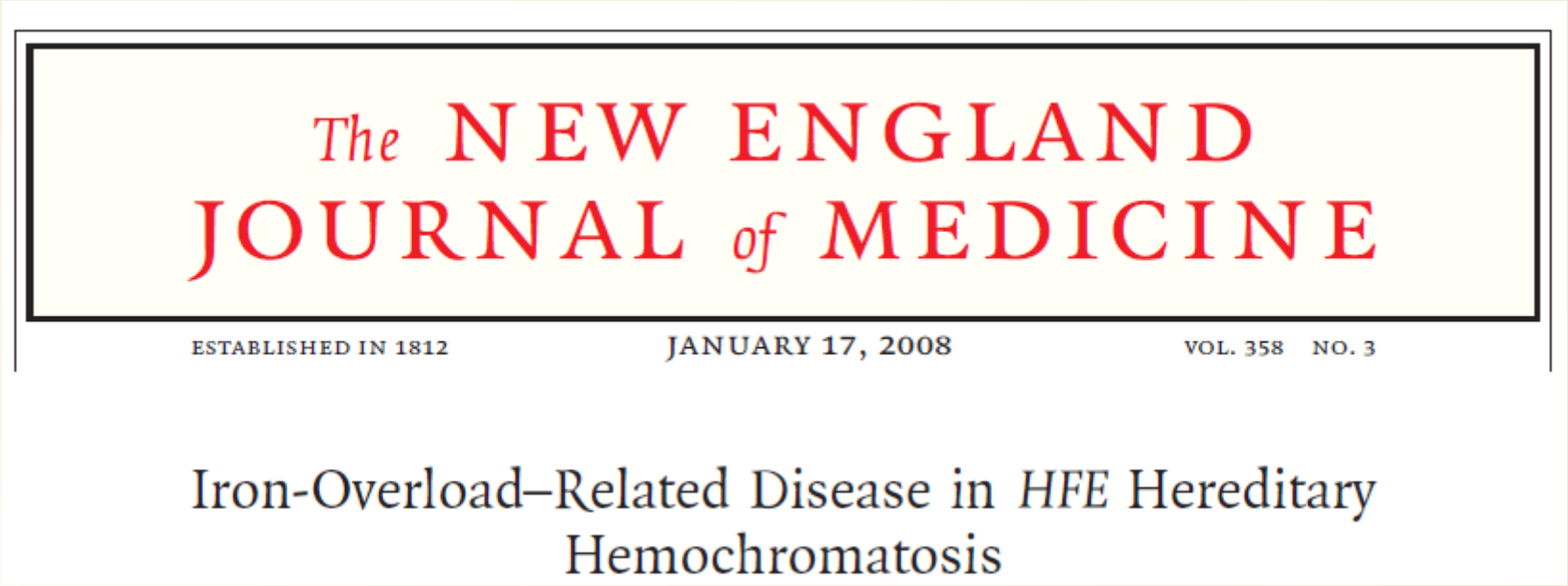




Denny et al., 2011

Phase I Phenotypes

	GHC/UW	Marshfield	Mayo	Northwestern	Vanderbilt
Primary					
Dementia	X	X			X
Cataract		X			X
PAD		X	X	X	X
Type 2 Diabetes		X	X	X	X
QRS Duration		X	X	X	X
Secondary					
WBC	X	X	X	X	X
Diabetic Retinopathy	X	X			X
RBC	X	X	X	X	
Lipids		X	X	X	
Height		X	X	X	
PheWAS					X
HDL	X	X	X		
Network					
Hypothyroidism	X	X	X	X	X
Resistant HTN	X	X	X	X	X



The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 JANUARY 17, 2008 VOL. 358 NO. 3

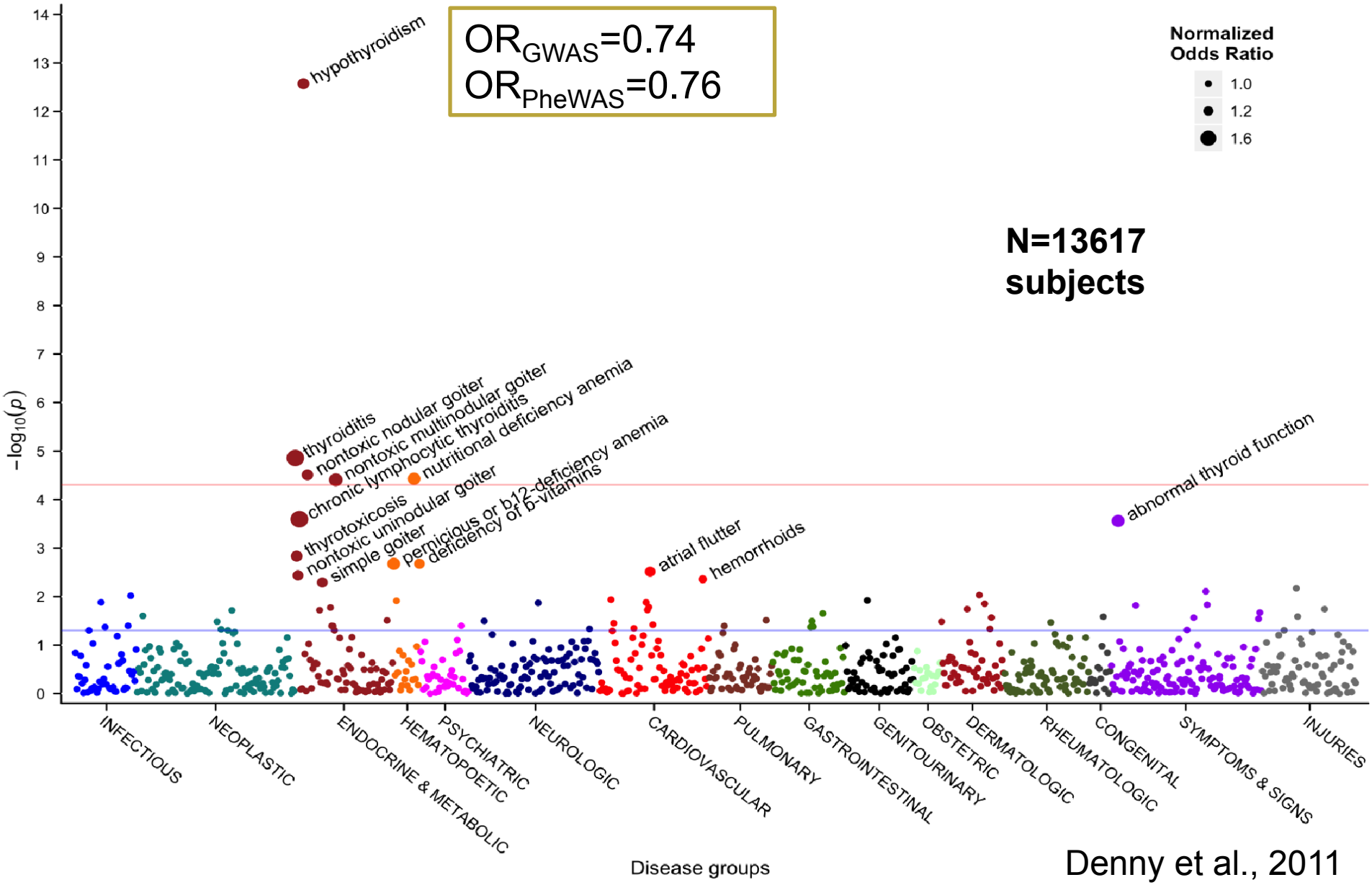
Iron-Overload–Related Disease in *HFE* Hereditary Hemochromatosis

- 203 C282Y homozygotes; n=31,192 Northern Europeans ages 40-69
- 28.4% of men and 1.2% of women had possible "iron-overload related disease"

VUMC BioVU data:

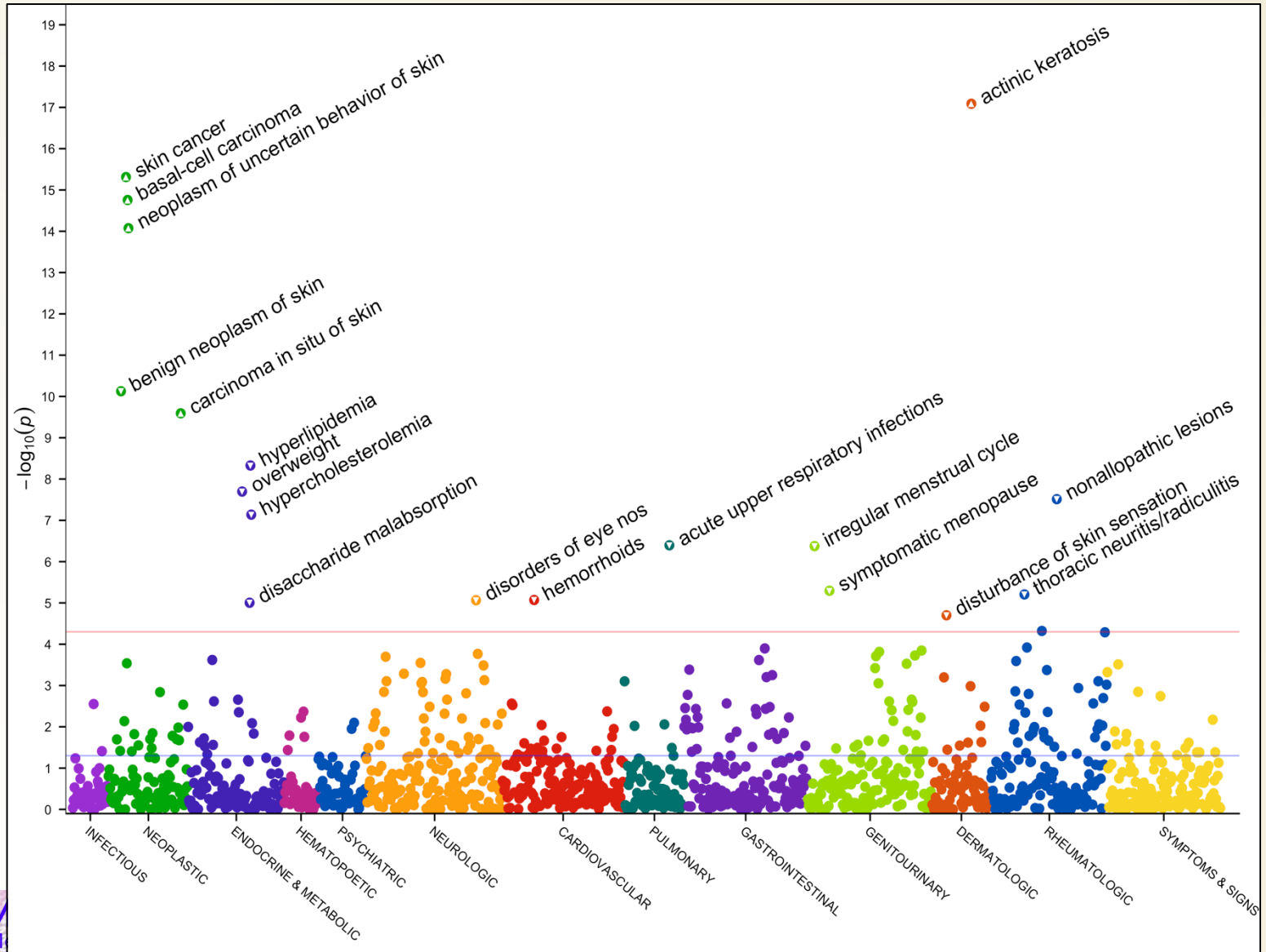
- 41/~5000 C282Y homozygotes
  - 9 diagnosed as having hemochromatosis
  - 32 undiagnosed:
    - 17 possible symptoms
    - 7 receiving iron supplements

PheWAS for rs10759944



Denny et al., 2011

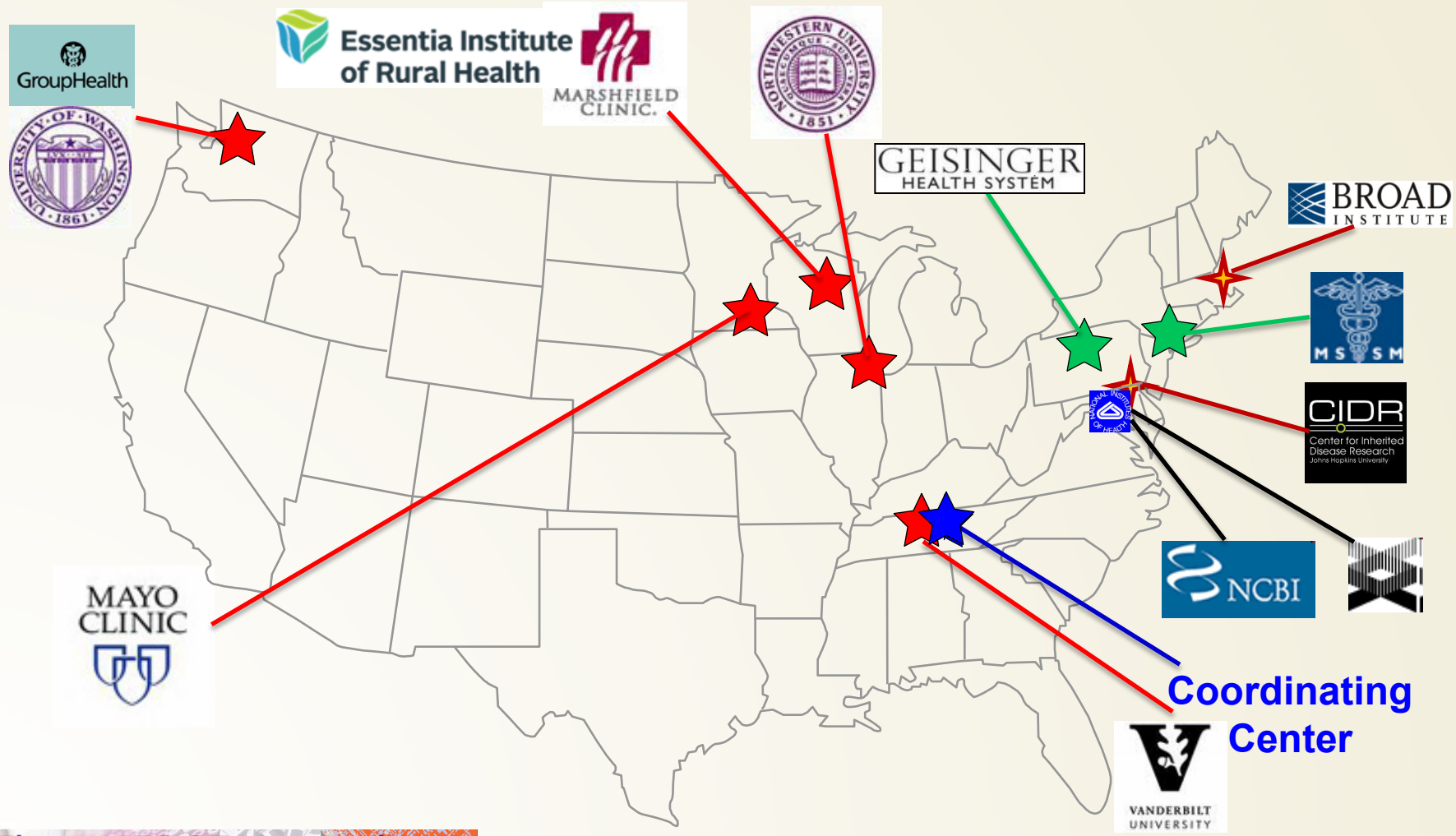
Pleiotropy: PheWAS associations with a skin color SNP



The *eMERGE* Network

electronic Medical Records & Genomics

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

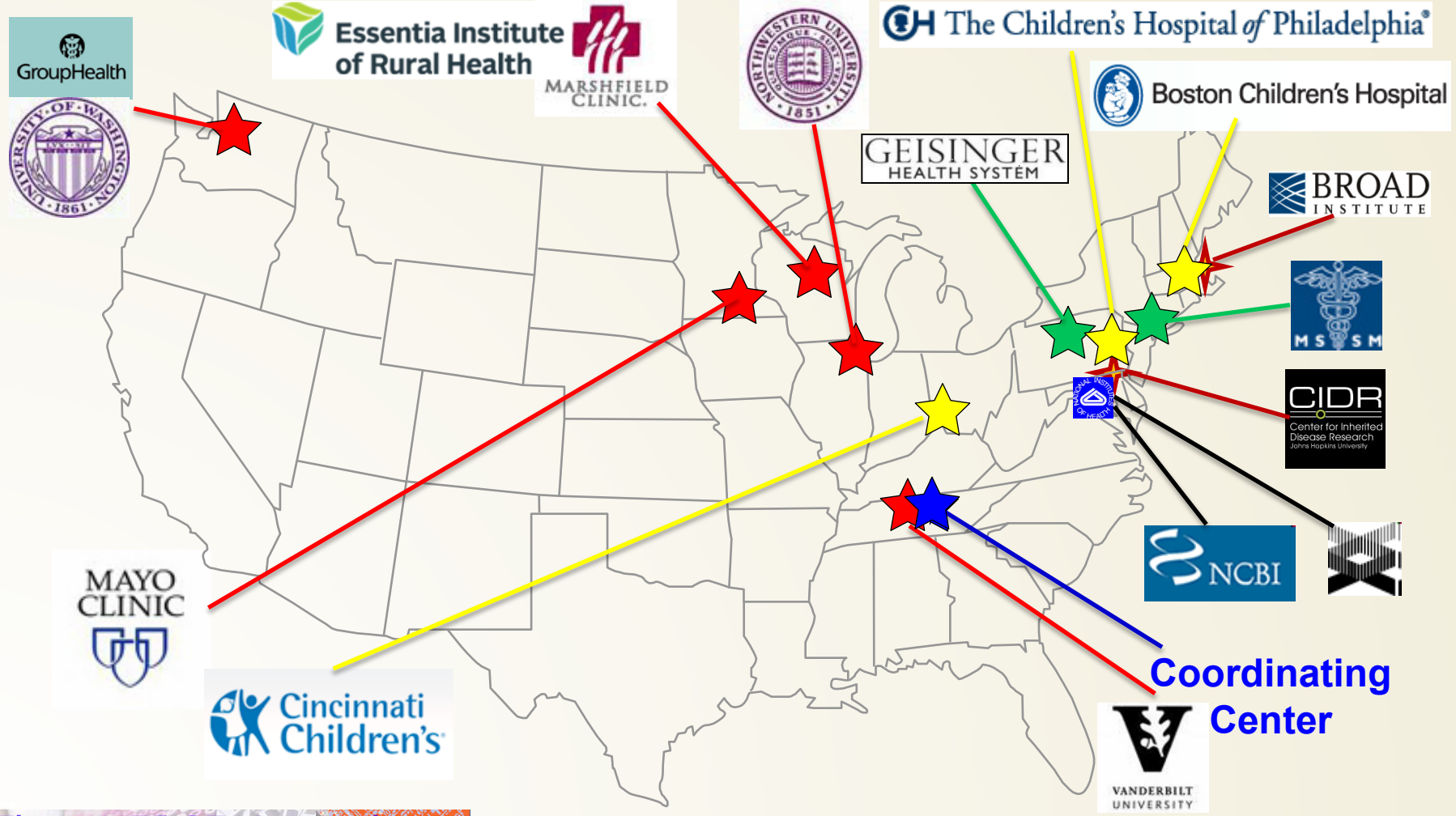


The *eMERGE* Network
electronic Medical Records & Genomics

The *eMERGE* Network

electronic Medical Records & Genomics

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



The *eMERGE* Network
electronic Medical Records & Genomics

eMERGE II Goals

- EHR-based Phenotyping
- Establish new Genotype-Phenotype associations
- Clinical Use:
  - Defining actionability/clinical utility/validity
  - Integration into EHR/Visualization/Clinical Decision support
- Physician and Patient attitudes/Education
- Consent/Regulatory
- Privacy/Security/CLIA/CAP

eMERGE-II Sample Size

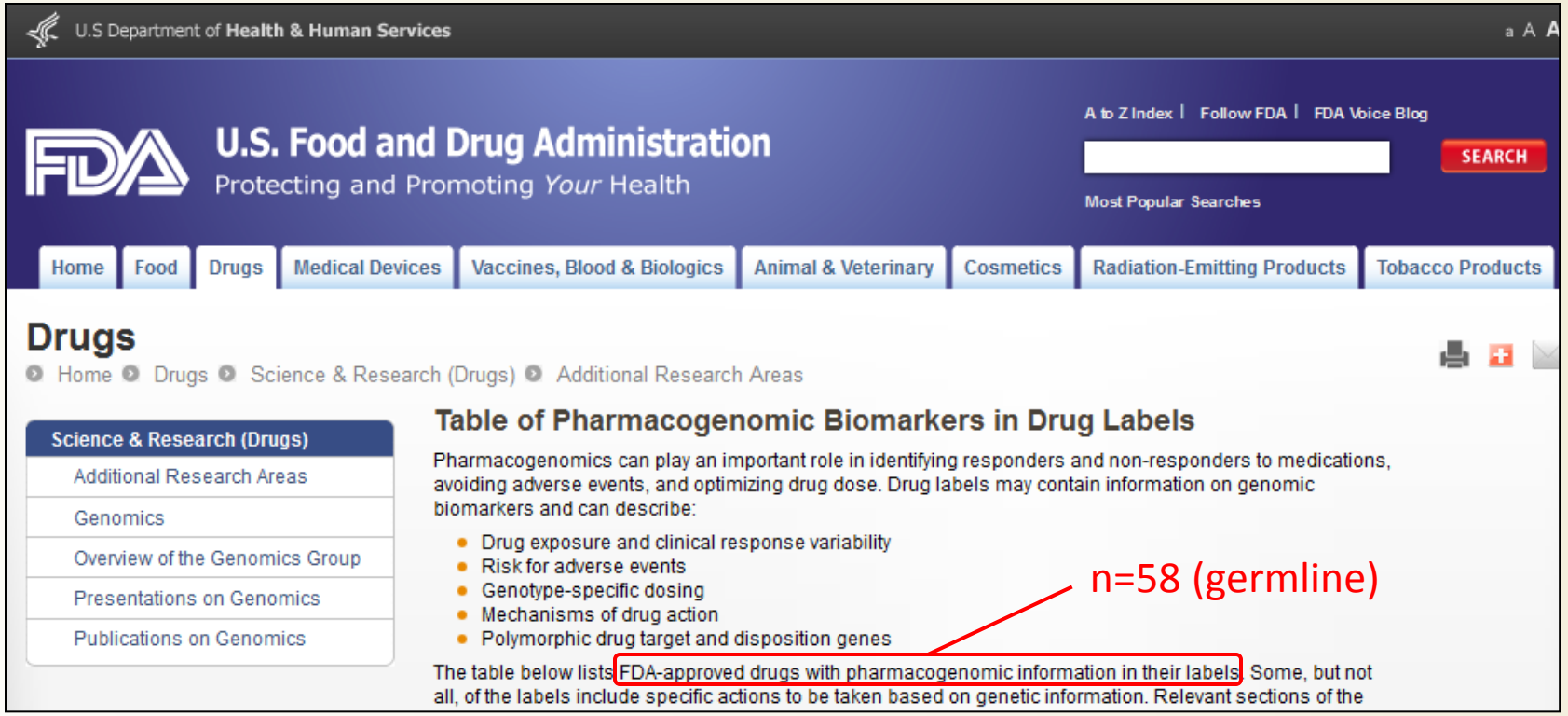
	Participants	Genotyped
GHC/UW	6,400	3,575
Marshfield	20,000	4,987
Mayo	19,000	6,940
NU	10,500	4,962
VU	140,000	33,228
Geisinger	19,700	4,191
Mt. Sinai	21,000	16,000
CCMC/CHB	40,100	5,586
CHOP	40,000	8,000
	316,600	**87,469**

Collins: Pharmacogenomics will undoubtedly become a very compelling part of medical practice. The limiting factor right now is that oftentimes, if you are ready to write a prescription, you do not want to wait a week to find out the genotype before you decide whether you've got the right dose and the right drug. But if everybody's DNA sequence is already in their medical record and it is simply a click of the mouse to found out all the information you need, then there is going to be a much lower barrier to beginning to incorporate that information into drug prescribing. If you have the evidence, it will be hard, I think, to say that this is not a good thing. And once you've got the sequence, it's not going to be terribly expensive. And it should improve outcomes and reduce adverse events.

Francis Collins, 9/16/2009

"Here's my sequence..."

New Yorker, 2000


U.S Department of Health & Human Services
a A A
U.S. Food and Drug Administration
Protecting and Promoting Your Health
A to Z Index | Follow FDA | FDA Voice Blog
SEARCH
Most Popular Searches
Home
Food
Drugs
Medical Devices
Vaccines, Blood & Biologics
Animal & Veterinary
Cosmetics
Radiation-Emitting Products
Tobacco Products
Drugs
Home Drugs Science & Research (Drugs) Additional Research Areas
Science & Research (Drugs)
Additional Research Areas
Genomics
Overview of the Genomics Group
Presentations on Genomics
Publications on Genomics
Table of Pharmacogenomic Biomarkers in Drug Labels
Pharmacogenomics can play an important role in identifying responders and non-responders to medications, avoiding adverse events, and optimizing drug dose. Drug labels may contain information on genomic biomarkers and can describe:
Drug exposure and clinical response variability
Risk for adverse events
Genotype-specific dosing
Mechanisms of drug action
Polymorphic drug target and disposition genes
n=58 (germline)
The table below lists FDA-approved drugs with pharmacogenomic information in their labels. Some, but not all, of the labels include specific actions to be taken based on genetic information. Relevant sections of the

Clopidogrel label revision March 2010

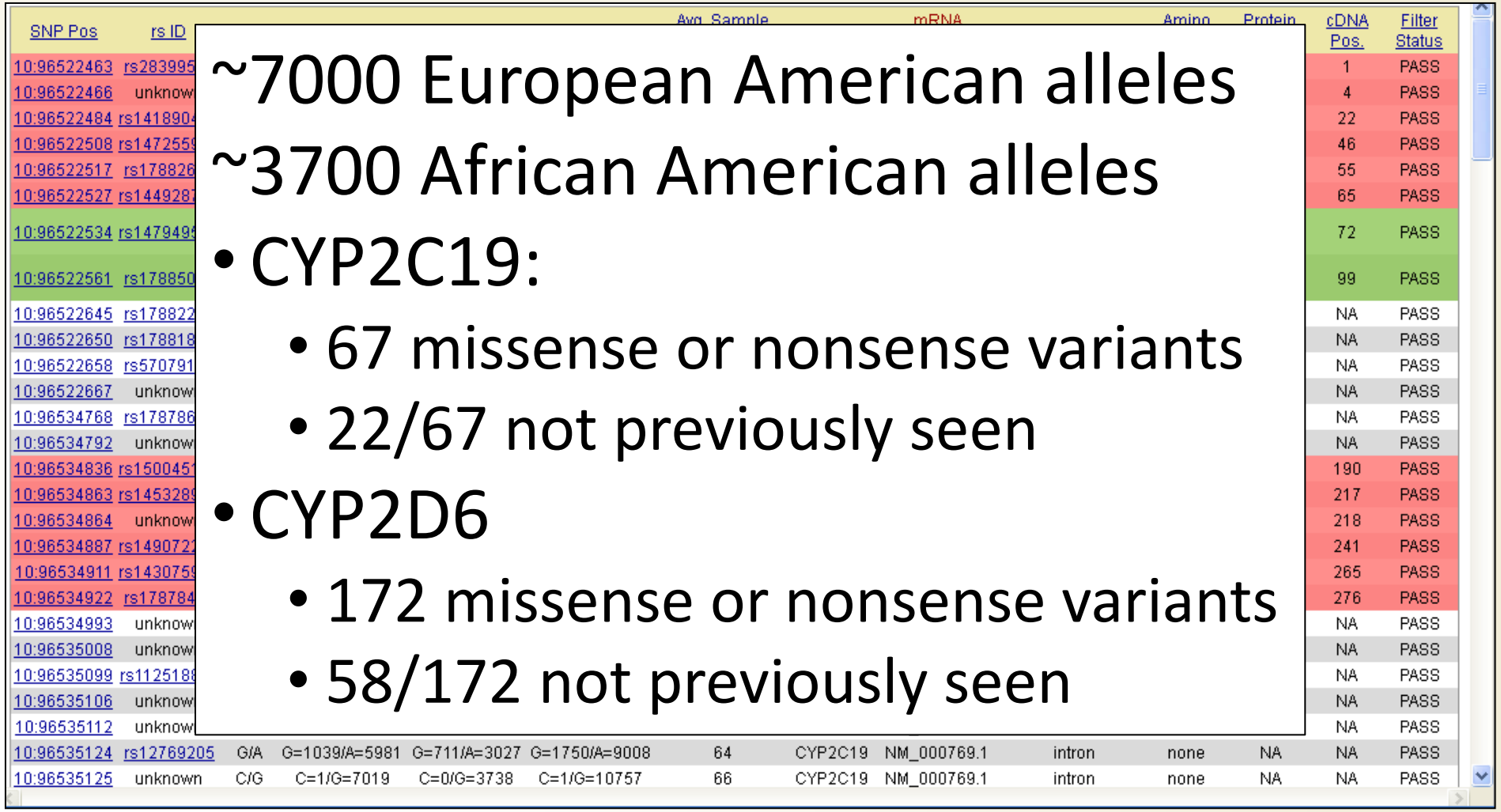
WARNING: DIMINISHED EFFECTIVENESS IN POOR METABOLIZERS

See full prescribing information for complete boxed warning.

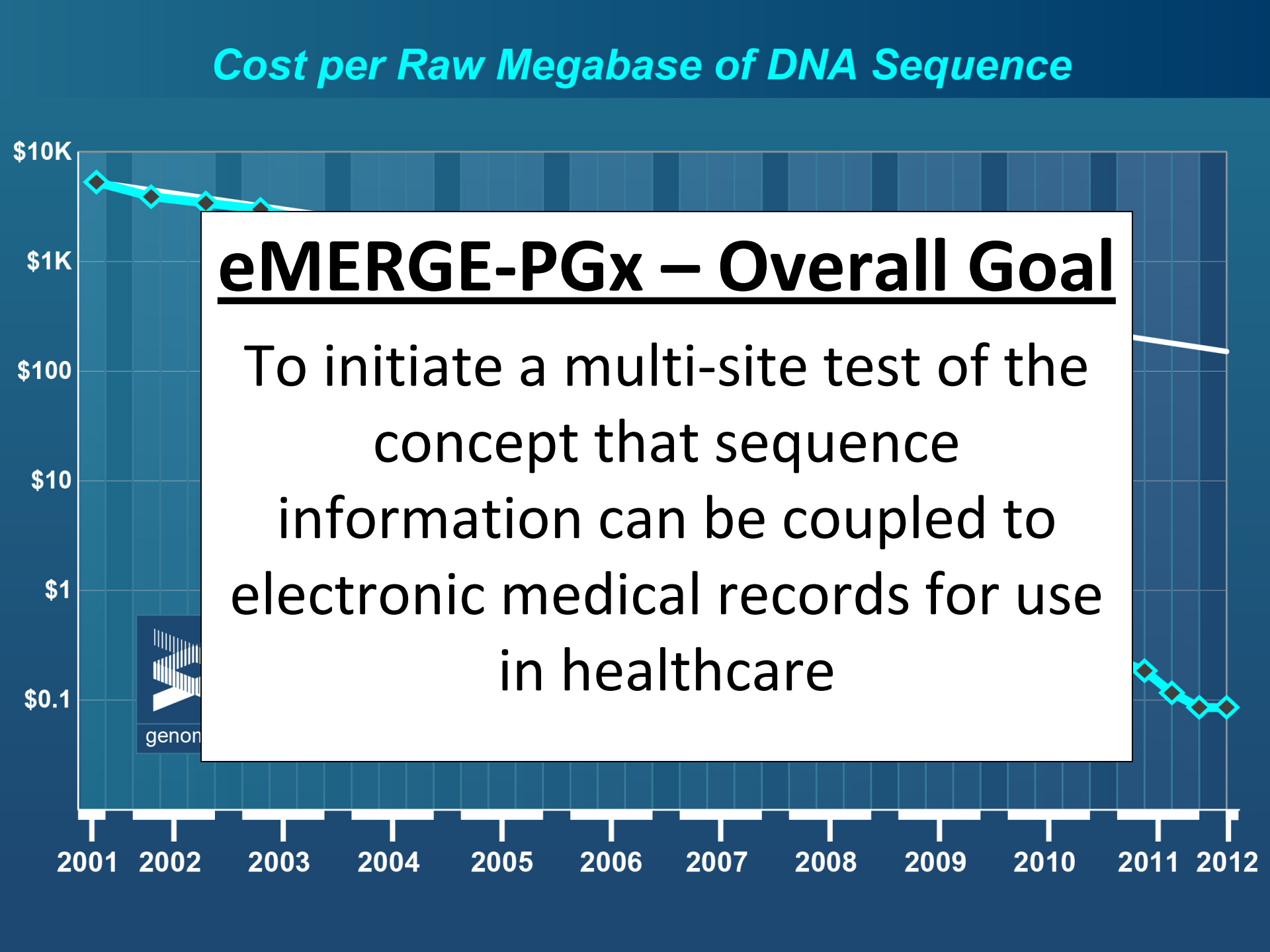
- Effectiveness of Plavix depends on activation to an active metabolite by the cytochrome P450 (CYP) system, principally CYP2C19. (5.1)
- Poor metabolizers treated with Plavix at recommended doses exhibit higher cardiovascular event rates following acute coronary syndrome (ACS) or percutaneous coronary intervention (PCI) than patients with normal CYP2C19 function. (12.5)
- Tests are available to identify a patient's CYP2C19 genotype and can be used as an aid in determining therapeutic strategy. (12.5)
- Consider alternative treatment or treatment strategies in patients identified as CYP2C19 poor metabolizers. (2.3, 5.1)

*1/*2 (18.8%) and *2/*2 (2.6%)

Beyond CYP2C19*2…



Exome Variant Server: http://evs.gs.washington.edu/EVS/


Cost per Raw Megabase of DNA Sequence
$10K
$1K
$100
$10
$1
$0.1
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
eMERGE-PGx – Overall Goal
To initiate a multi-site test of the concept that sequence information can be coupled to electronic medical records for use in healthcare

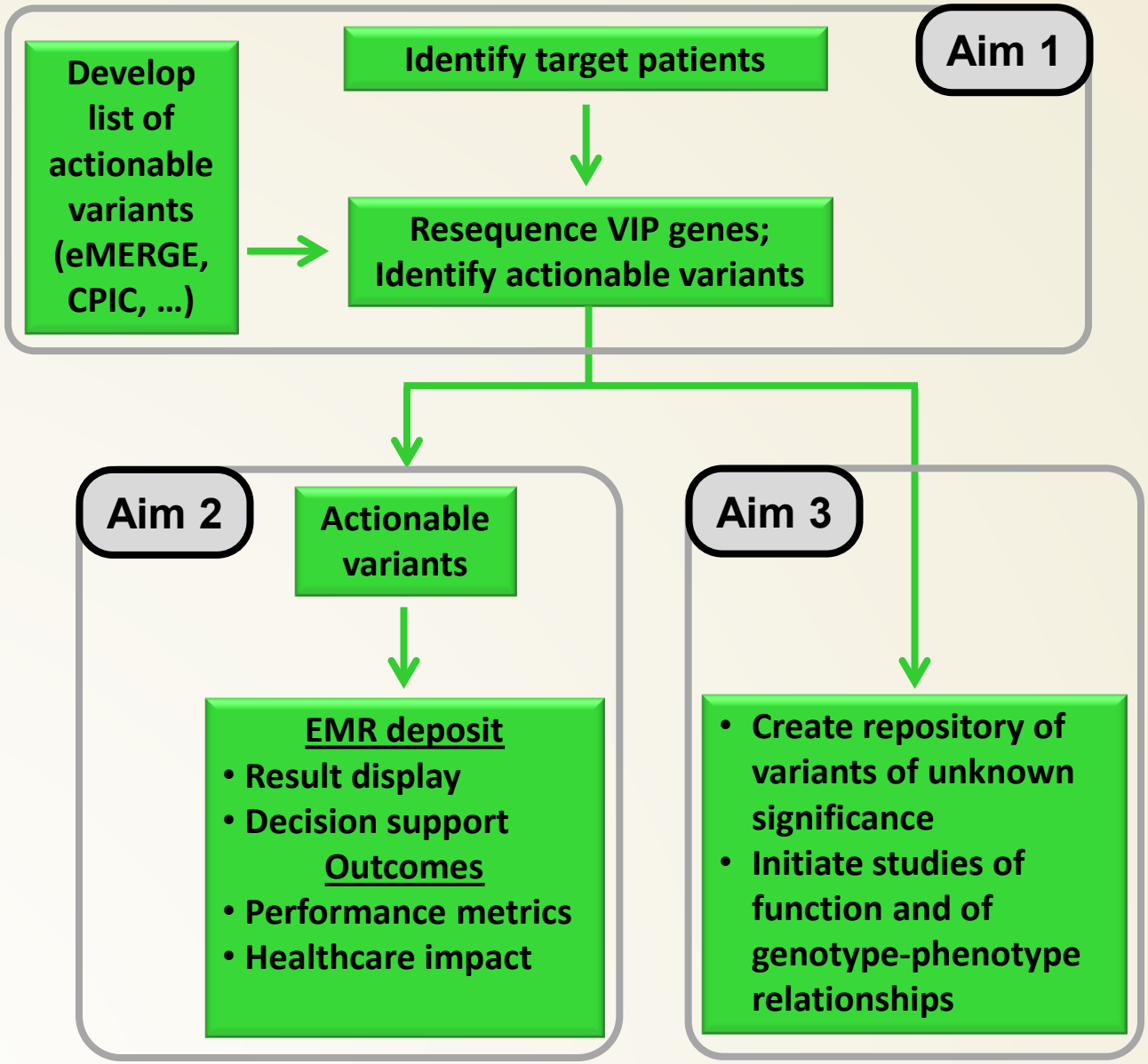
eMERGE-PGx: a PGRN-eMERGE alliance

PGRN

- Clinical Pharmacogenomics Implementation Consortium (CPIC)
- Translational Pharmacogenomics Project (TPP)
- PGRN-Seq and other platforms

eMERGE

- developing and validating electronic phenotyping algorithms (including for drug responses)
- developing and deploying clinical decision support


Aim 1
Identify target patients
Develop list of actionable variants (eMERGE, CPIC, …)
Resequence VIP genes; Identify actionable variants
Aim 2
Actionable variants
EMR deposit
• Result display
• Decision support
Outcomes
• Performance metrics
• Healthcare impact
Aim 3
• Create repository of variants of unknown significance
• Initiate studies of function and of genotype-phenotype relationships

What have we learned since 2008?

- Algorithms to accurately identify cases and controls for simple phenotypes can be developed and deployed across multiple EMRs
- In progress: more complex phenotypes
  - Multigene prediction tools
  - Complications of disease
  - Variable drug responses
- The PheWAS experiment is feasible
- Implementation is far from straightforward

Closing thoughts/"answers"

- Populations to be sequenced should include healthcare information
- Advantages of mining in EMRs
  - Real-world
  - Feasibility demonstrated
  - Enabling for implementation (eventually)
  - Rare/extreme phenotypes accessible
  - Potential for coupling to other datasets
- Disadvantages
  - The phenotype is what is in the EMR. Dense exquisite phenotypes need (a lot of) extra work.

Closing thoughts/answers

1. General considerations about power and sample size
2. Relationship of expected disease architecture to study design
3. Strengths and weaknesses of prospective cohort or retrospective case-control designs, family or extremes designs
4. Strengths and weaknesses, especially costs and benefit, of whole genome vs whole exome sequencing
5. Potential differences in analytic approaches for interpretation of WES/WGS sequence data from thousands of samples vs from small numbers of samples
6. Types of validation needed and how this should be factored into initial design
7. Additional data types would one want to gather (RNA, proteomics, etc.) and any impact their availability might have on design of the DNA sequencing component