eMERGE Network
electronic medical records & genomics

Major US Genomic Medicine Programs: NHGRI's Electronic Medical Records and Genomics (eMERGE) Network

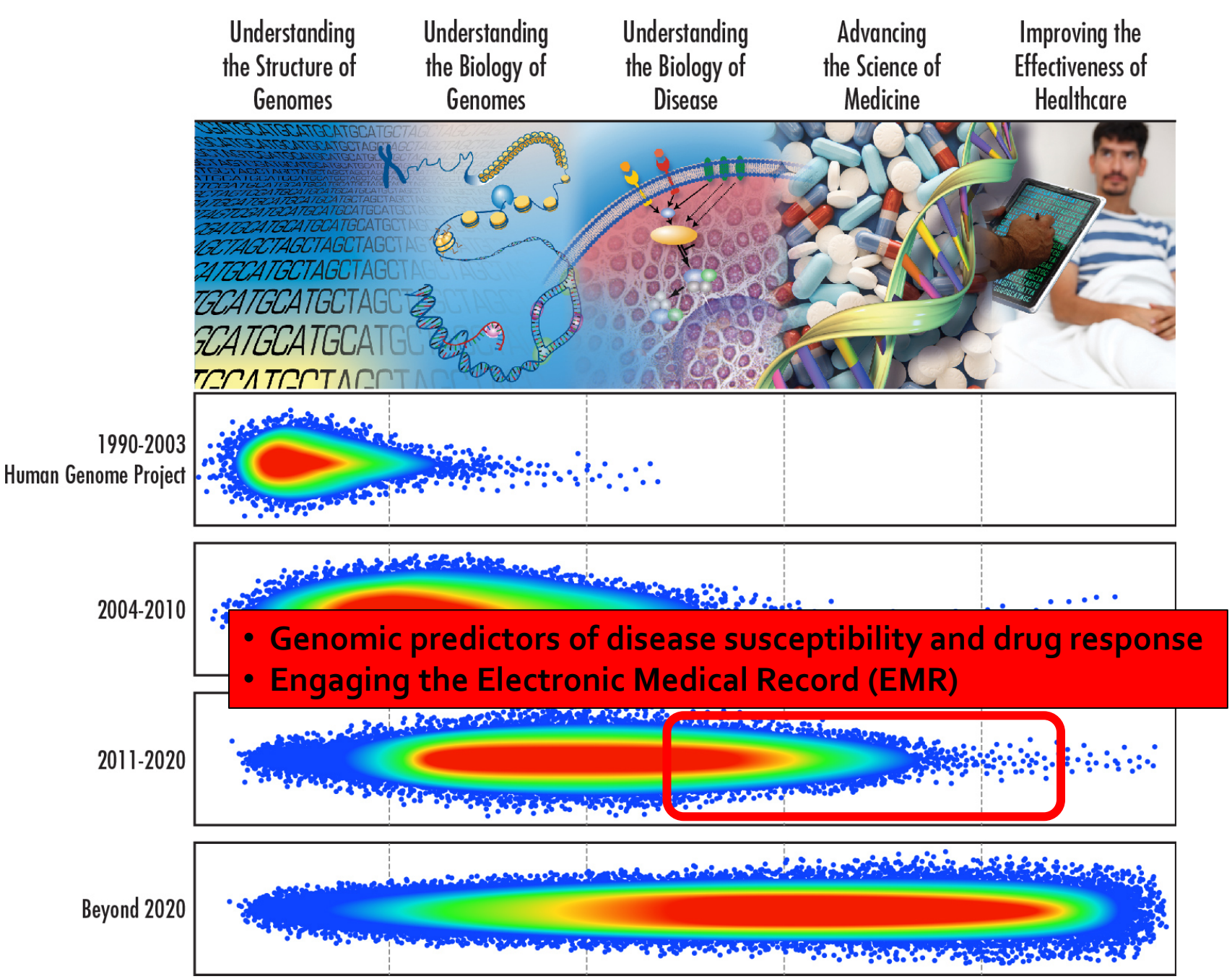
Dan Roden

Member, National Advisory Council For Human Genome Research Genomic Medicine Working Group

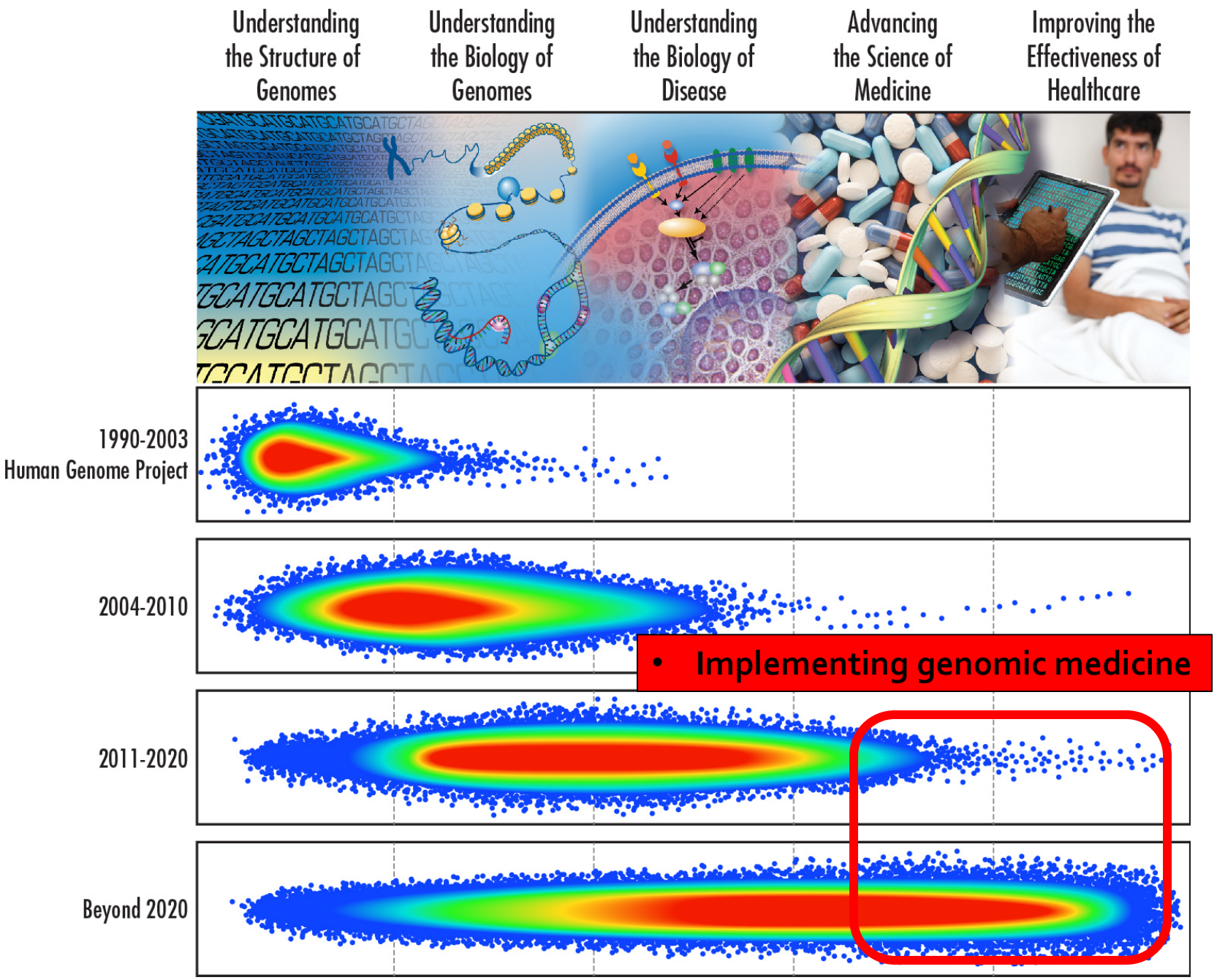
GroupHealth · Marshfield Clinic · University of Washington · Essentia Health · Mount Sinai School of Medicine · Geisinger · The Children's Hospital of Philadelphia · Mayo Clinic · Boston Children's Hospital · Northwestern University · Cincinnati Children's Hospital Medical Center · Vanderbilt University

Creating electronic medical records can enable

- Improved care of individual patients
- Identification of specific subsets of patients
- Discovery of new genotype-phenotype and phenotype-genotype associations
- Implementation of Genomic Medicine



Green et al. 2011

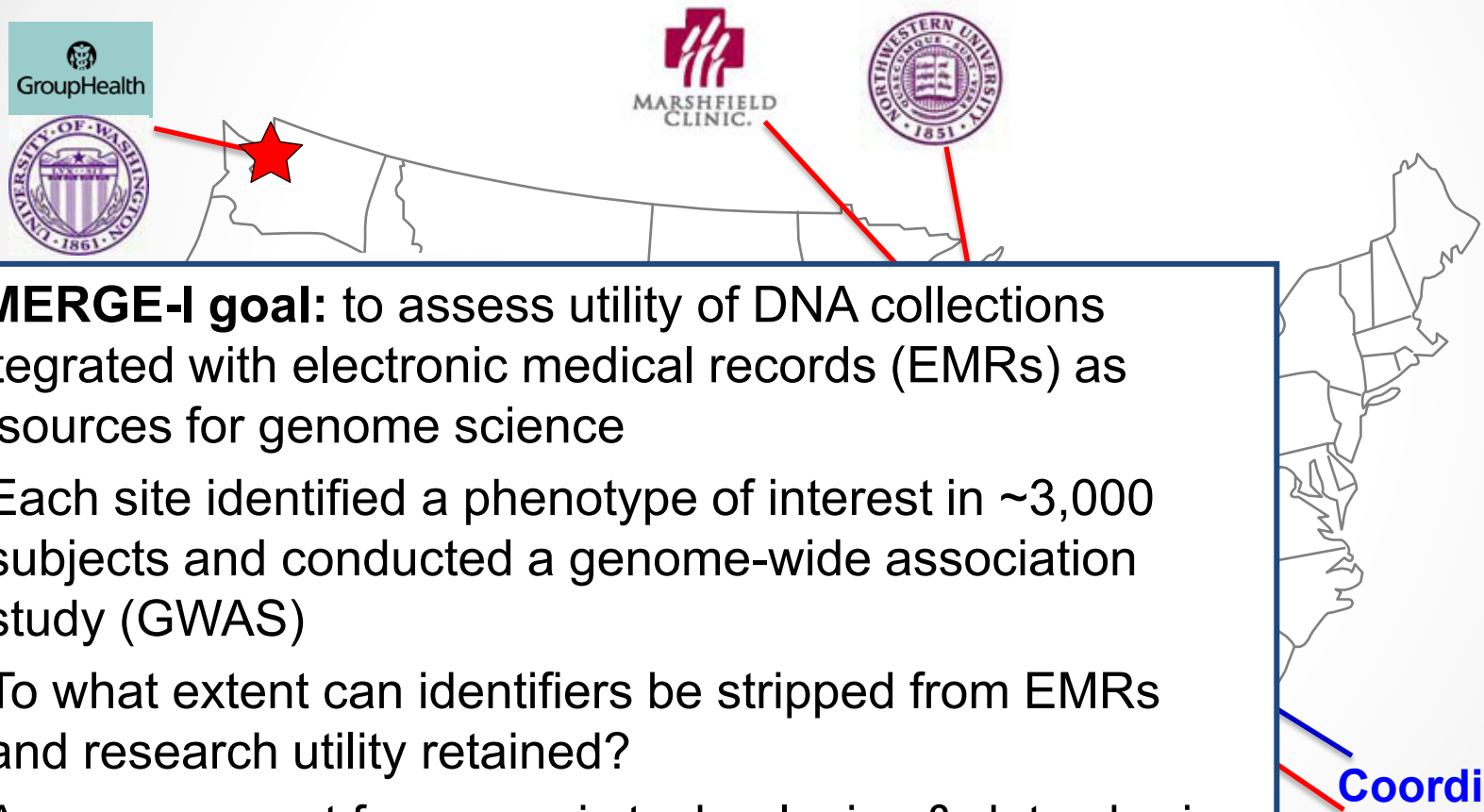


Green et al. 2011

eMERGE Network

electronic medical records & genomics

2007-2011: Phase I



eMERGE-I goal: to assess utility of DNA collections integrated with electronic medical records (EMRs) as resources for genome science

- Each site identified a phenotype of interest in ~3,000 subjects and conducted a genome-wide association study (GWAS)
- To what extent can identifiers be stripped from EMRs and research utility retained?
- Assess consent for genomic technologies & data sharing
- Develop and promulgate best practices for phenotyping and genomics in EMRs

eMERGE-I phenotypes

	# Genotype	
Phenotype	**Case**	**Control**
Cataract	2642	1322
Dementia	1241	2043
PAD	1641	1604
QRS	3034	-
T2 Diabetes	2706	1496

Data Sharing Memorandum of Understanding

- Each site has final authority regarding their data
- How data may be shared
- Privacy and Confidentiality agreements
- Limitations of Use

eMERGE-I phenotypes

`	# Genotype		Data Only	
Phenotype	**Case**	**Control**	**Case**	**Control**
Cataract	2642	1322	1386	1360
Dementia	1241	2043	14	-
PAD	1641	1604	1010	8743
QRS	3034	-	1019	-
T2 Diabetes	2706	1496	1101	912

Data Sharing Memorandum of Understanding

- Each site has final authority regarding their data
- How data may be shared
- Privacy and Confidentiality agreements
- Limitations of Use

eMERGE-I phenotypes

`	# Genotype		Data Only	
Phenotype	**Case**	**Control**	**Case**	**Control**
Cataract	2642	1322	1386	1360
Dementia	1241	2043	14	-
PAD	1641	1604	1010	8743
QRS	3034	-	1019	-
T2 Diabetes	2706	1496	1101	912
Platelet indices			13,582	
Red cell indices			16,915	
Hypothyroidism			1306	5013

Approach to electronic phenotyping

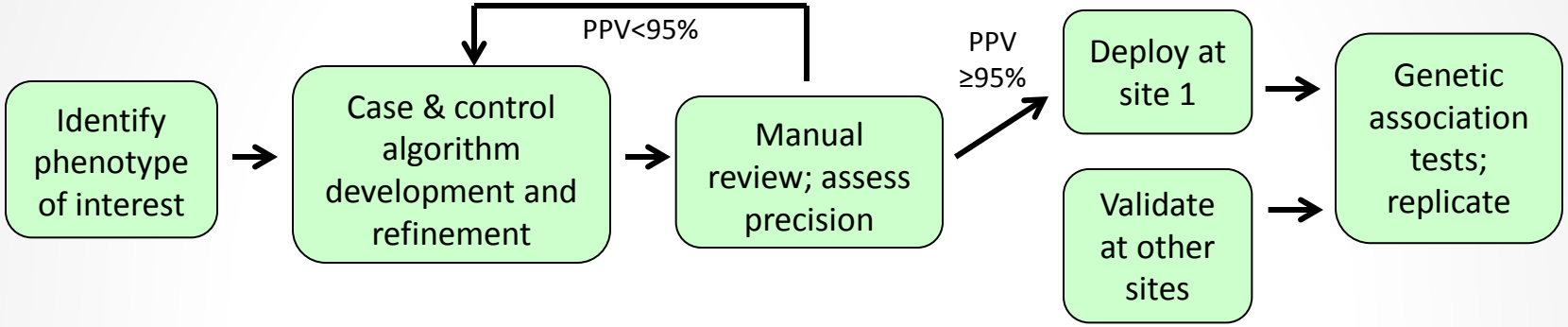


Table 1. Evaluation of Primary Hypothyroidism Algorithm at the Five eMERGE Sites

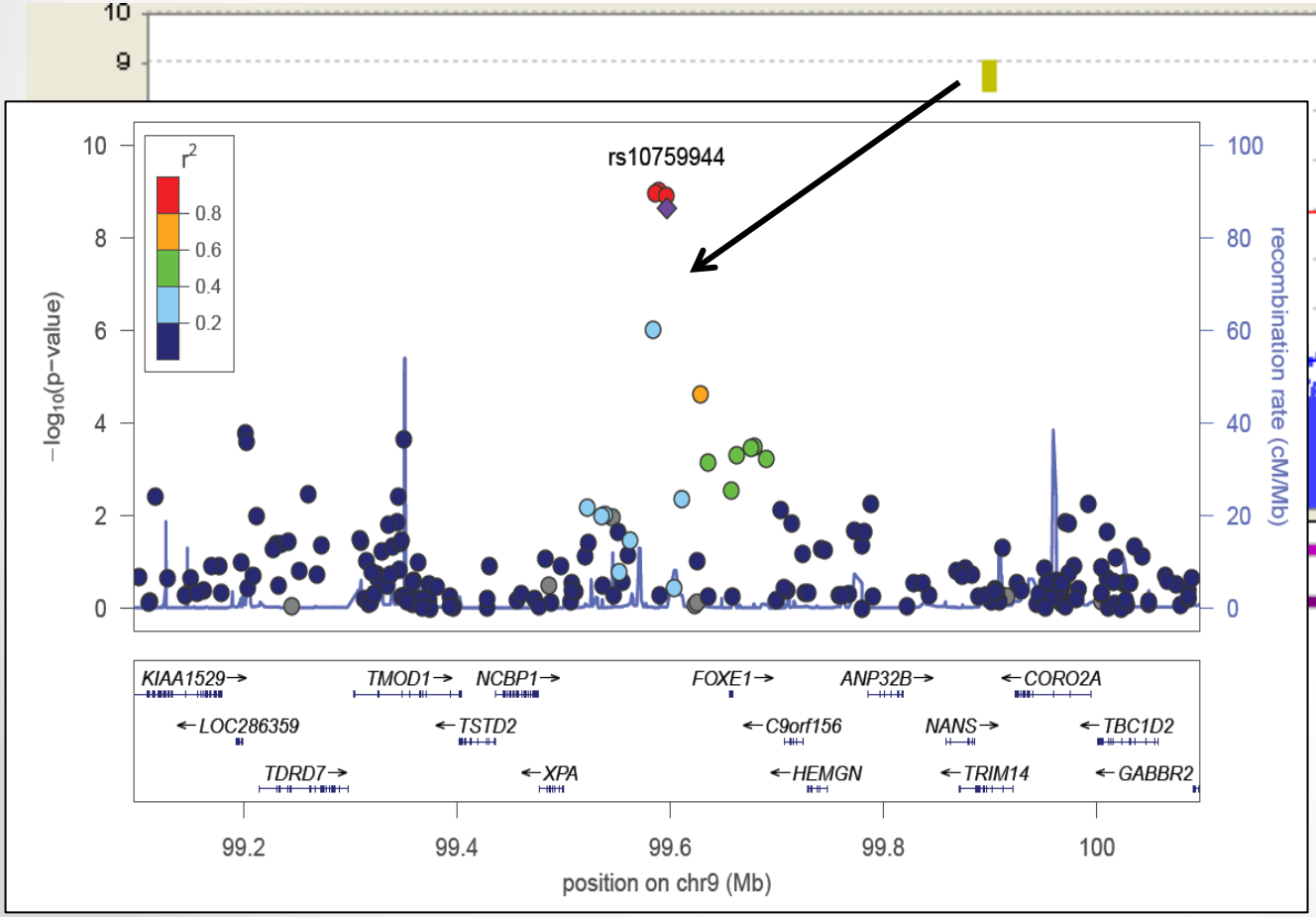
			Primary Hypothyroidism			
Site	**Primary Phenotype**	**Total Genotyped Subjects**	**Cases**	**Controls**	**Case PPV (%)**	**Control PPV (%)**
Group Health	dementia	2532	397	1,160	98	100
Marshfield	cataracts	4113	514	1,187	91	100
Mayo Clinic	peripheral arterial disease	3043	233	1,884	82	96
Northwestern	type 2 diabetes	1217	92	470	98	100
Vanderbilt	normal cardiac conduction	2712	81	352	98	100
All sites		13,617	1317	5053	92.4[a]	98.5[a]

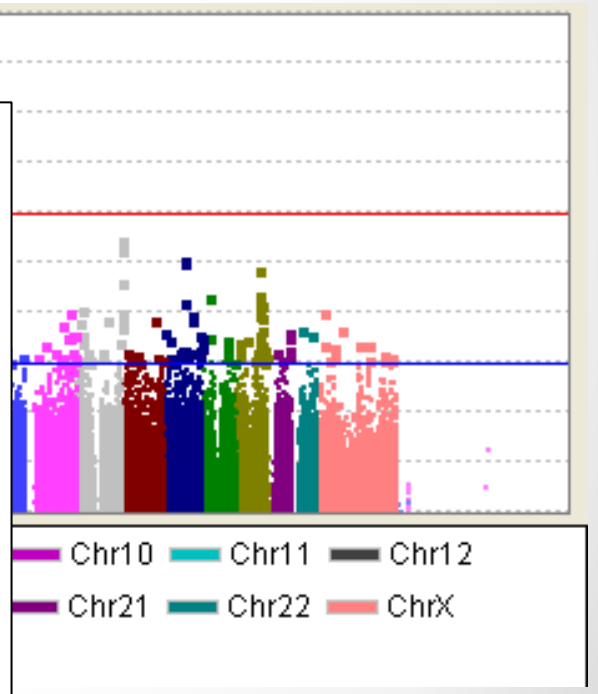
Genotype counts represent all subjects who were found by the hypothyroidism algorithms at each site and who were genotyped. Counts are limited to those classified as "white" in the electronic medical record of each site. PPV = positive predictive value.
[a] Average weighted for number of samples contributed to the total.

Denny et al., 2011

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

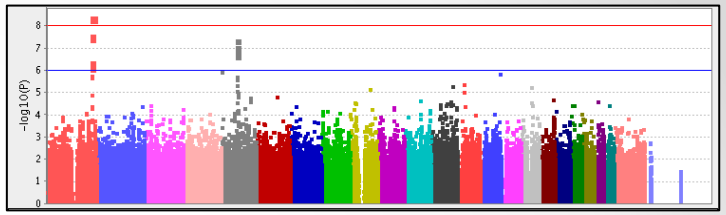
European Americans (1,306 cases and 5,013 controls)





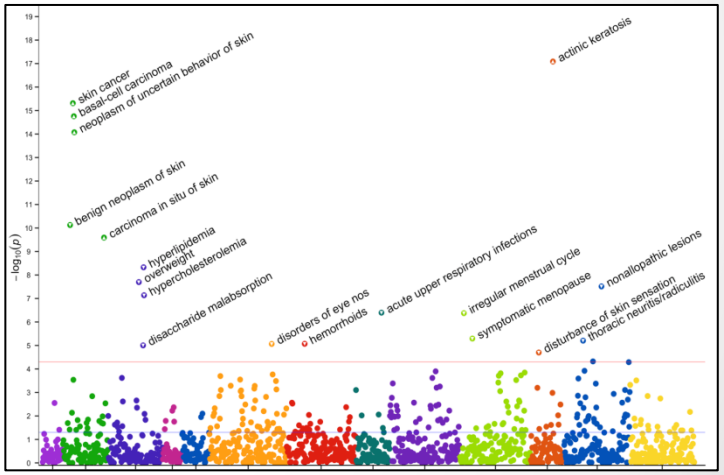
Denny et al., 2011

GWAS: Target phenotype → association P value



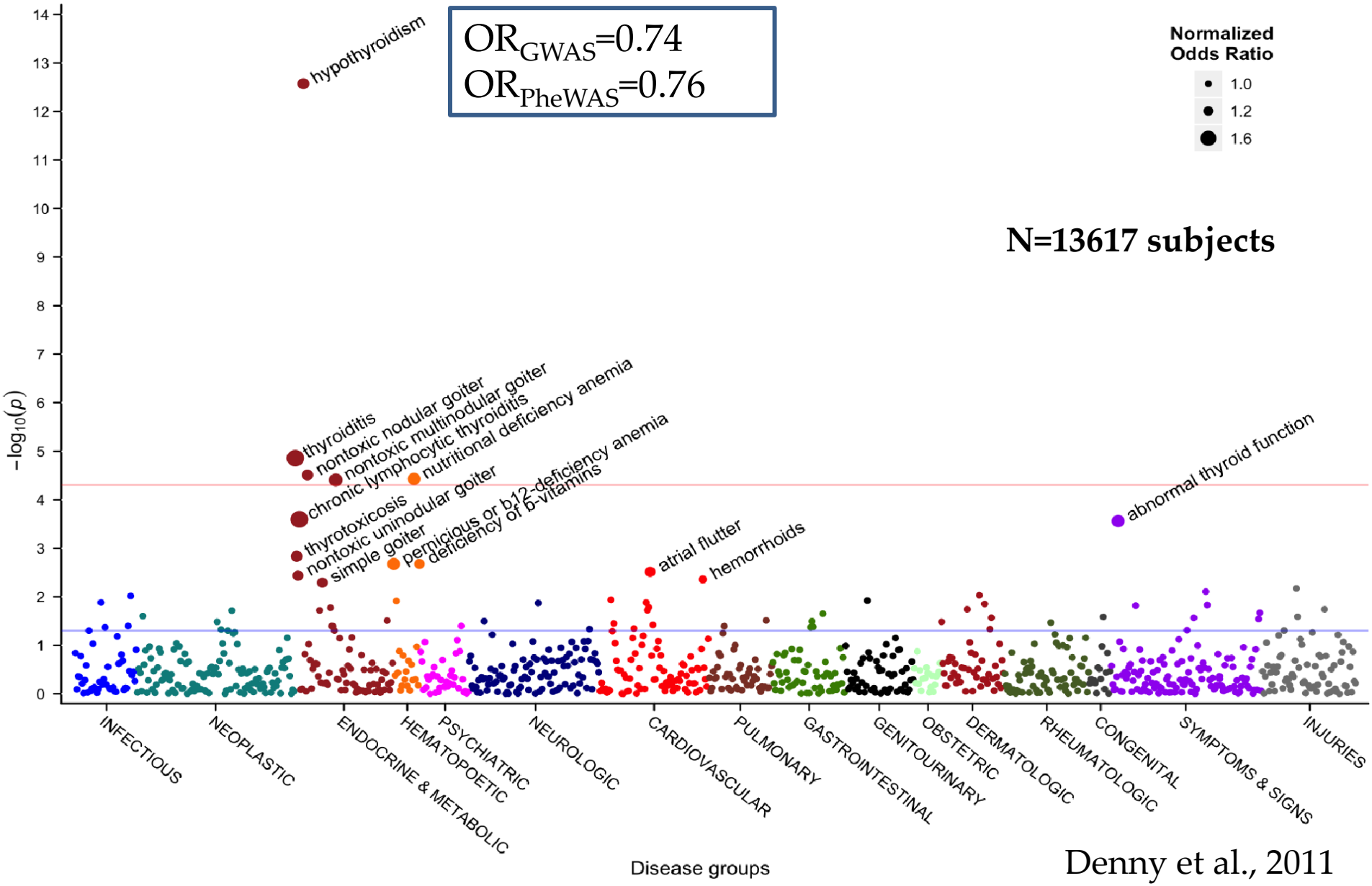
The phenome-wide association study

PheWAS (ΦWAS): Target genotype → association P value



PheWAS requirement: A large cohort of patients with genotype data and many diagnoses

PheWAS for rs10759944 near *FOXE1*



Denny et al., 2011

Pleiotropy: PheWAS associations with an IRF4 SNP previously associated with hair and eye color

- All SNPs in the GWAS catalog have now been analyzed by PheWAS
- PheWAS provides a replication tool for conventional GWAS and identifies potential new genetic associations
- All data are publically available at emrphewas.org

Denny et al., 2013

Firefox | What is the Phenotype KnowledgeBase? ... | phekb.org | Bing

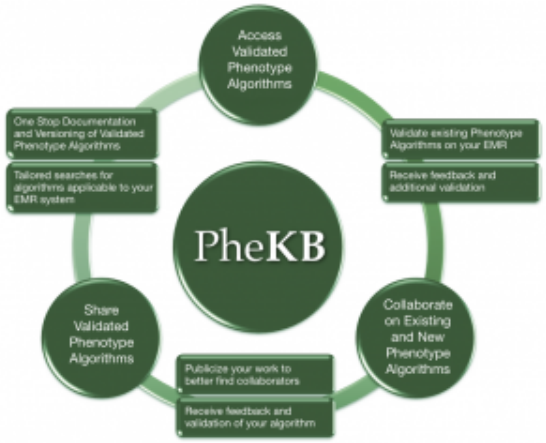
PheKB

a knowledgebase for discovering phenotypes from electronic medical records

Login | Register

Home | Phenotypes | Implementations | Groups | Institutions

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 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

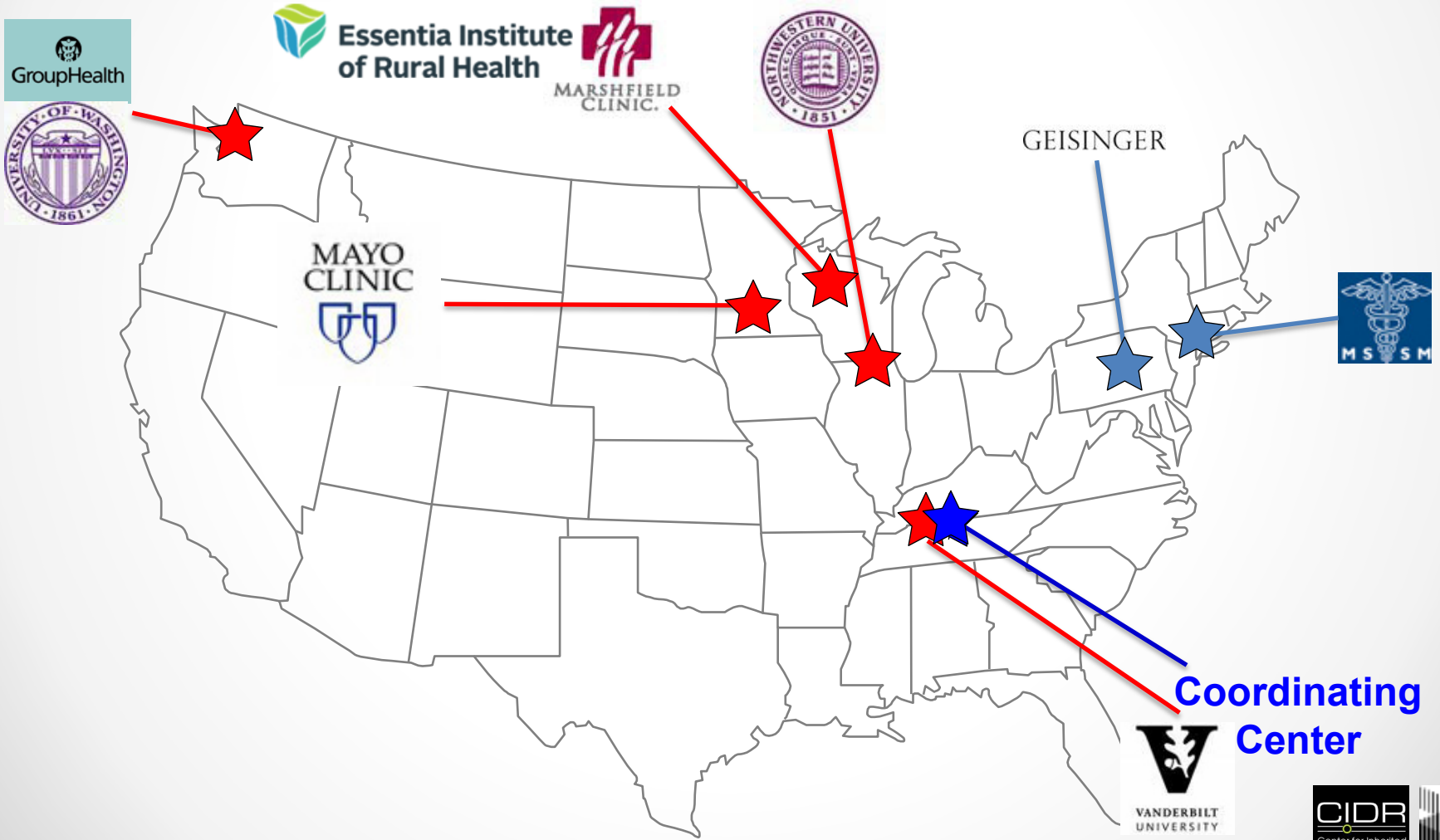
Most Recent Phenotypes

- Clopidogrel Poor Metabolizers
- Atrial Fibrillation - Demonstration Project
- Rhematoid Arthritis - Demonstration Project
- Multiple Sclerosis - Demonstration Project
- Crohn's Disease - Demonstration Project

eMERGE Network

electronic medical records & genomics

2011-2015: Phase II



eMERGE Network

electronic medical records & genomics

2011-2015: Phase II
+ pediatric sites

Essentia Institute

eMERGE-II goals

- Expand the electronic phenotyping library and apply to genotyped samples
- Initiate implementation of actionable variants into the EMR
  - Site-specific projects
  - Cross network initiatives
- Define actionability, clinical utility, validity
- Advance methods for integration of genomic information into EMRs, including methods for visualization and Clinical Decision Support
- Evaluate physician and patient attitudes and educational needs
- Continued focus on consent, regulatory, privacy, and security issues; extend to clinical laboratory implementation.

ordinating Center

CIDR Center for Inherited Disease Research Johns Hopkins University

Network-wide patient survey: biobanking consent

Questions:

- Do participants view specific consent to be a requirement for sharing biosamples and data for future research?
- Which biospecimen and biobanking-related research practices are likely to have the greatest impact on willingness to participate under broad consent?

Plan

- Survey 100,000 participants and patients across the eMERGE institutions to elicit a wide cross-section of patient perspectives.

Outcome

- Recommendations to inform future policy for the ethical conduct of human subject research

Network-wide return of results project: hemochromatosis

Site	C282Y/C282Y	C828Y/H63D	H63D/H63D	Sum
Geisinger	12	67	110	189
GHC/Seattle	17	60	72	149
Marshfield	15	52	87	154
Mayo	44	179	206	4
Mt. Sinai	1	12	29	42
Northwestern	19	64	81	164
Vanderbilt	39	152	141	332
Total	**147**	**586**	**726**	**1459**

- Do these patients carry the clinical diagnosis?
- Do they have clinical phenotypes?

Site-specific Genomic Medicine Implementation Pilot Projects

- **Developing genetic risk scores and evaluating their potential clinical impact**
  - Marshfield: Age-related Macular Degeneration (7 SNPs)
  - Mayo: Coronary Artery Disease (28 SNPs)
- **Genotyping specific variants and evaluating impact on physicians and patients of returning results:**
  - Mount Sinai: ApoL1 variants and development of renal dysfunction in hypertensives
  - Northwestern: Impact of genotyping for HFE and FVL variants in an Internal Medicine clinic

Site-specific Genomic Medicine Implementation Pilot Projects

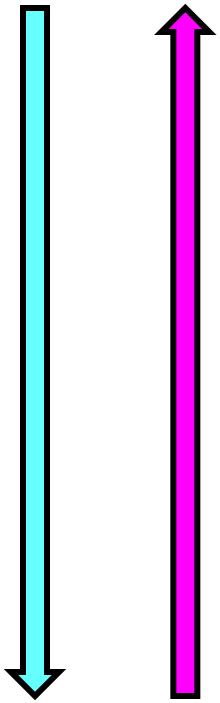
- **Whole genome sequencing**
  - Geisinger: WGS for undiagnosed disease in trios
- **Pharmacogenomics focus**
  - Geisinger: Preemptive genotyping for IL28B in patients with hepatitis C
  - Cincinnati Children's/Boston Children's: Assay CYP2D6 and provide results to parents and providers
  - Children's Hospital of Philadelphia: Response to beta-adrenergic agonists in children with asthma
  - Vanderbilt: Multiplexed preemptive pharmacogenomic testing

eMERGE-PGRN Partnership

PGRN Pharmacogenomics Research Network

<u>PGx capabilities:</u>

- CPIC guidelines
- Resequencing platform for 84 Very Important Pharmacogenes
- CLIA & QC standards

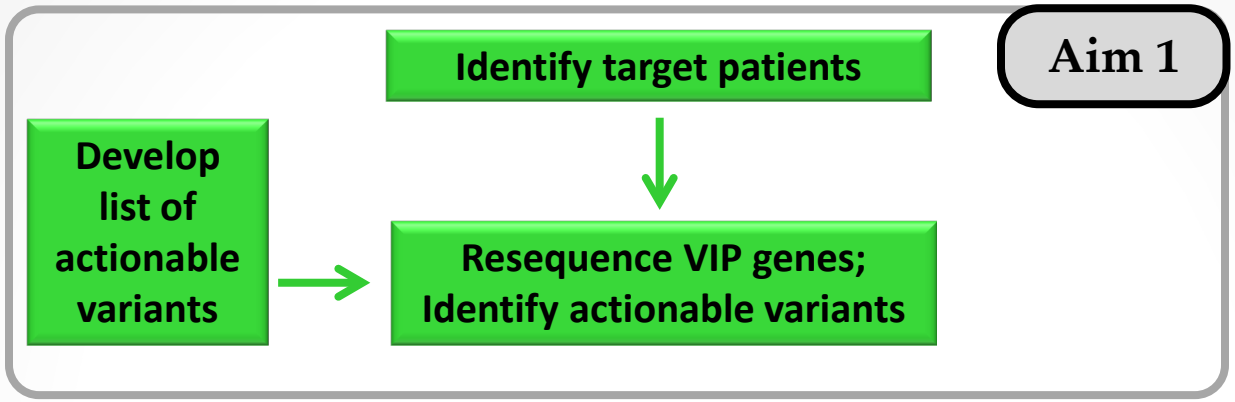


<u>EMR-informatics capabilities</u>

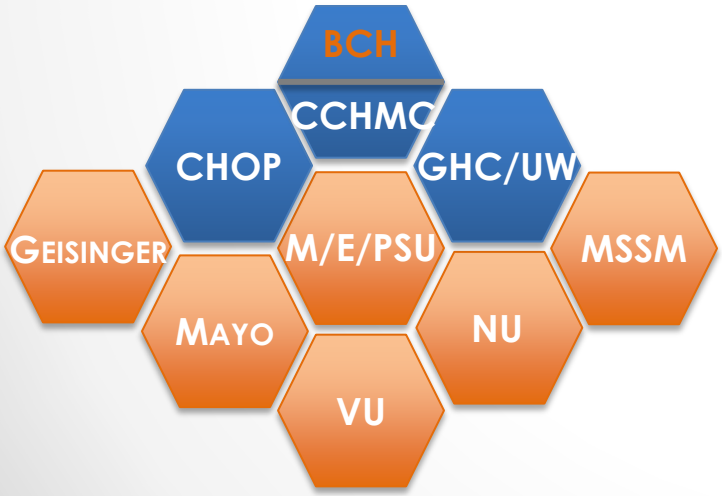
- Privacy
- Electronic phenotyping
- Large populations
- Decision support

eMERGE Network
electronic medical records & genomics

eMERGE-PGx project



Target: 9000 subjects

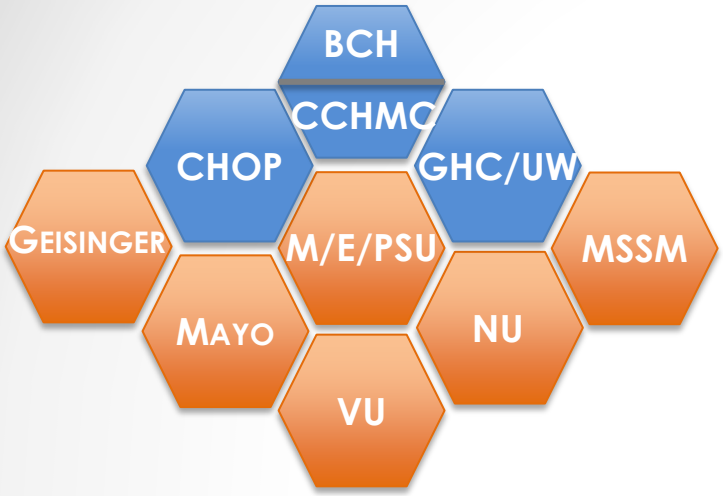


Drug-Genome pairs study

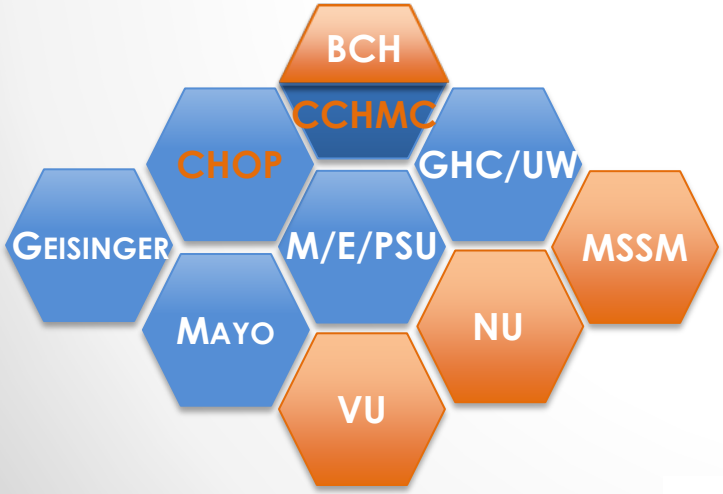
CYP2C19-Clopidogrel
VKORC1/CYP2C9-Warfarin*
SLCO1B1-Simvastatin

* BCH DGI only VKORC1/CYP2C9-Warfarin
* Geisinger and M/E/PSU also have CYP4F2-Warfarin

Identifying target patients



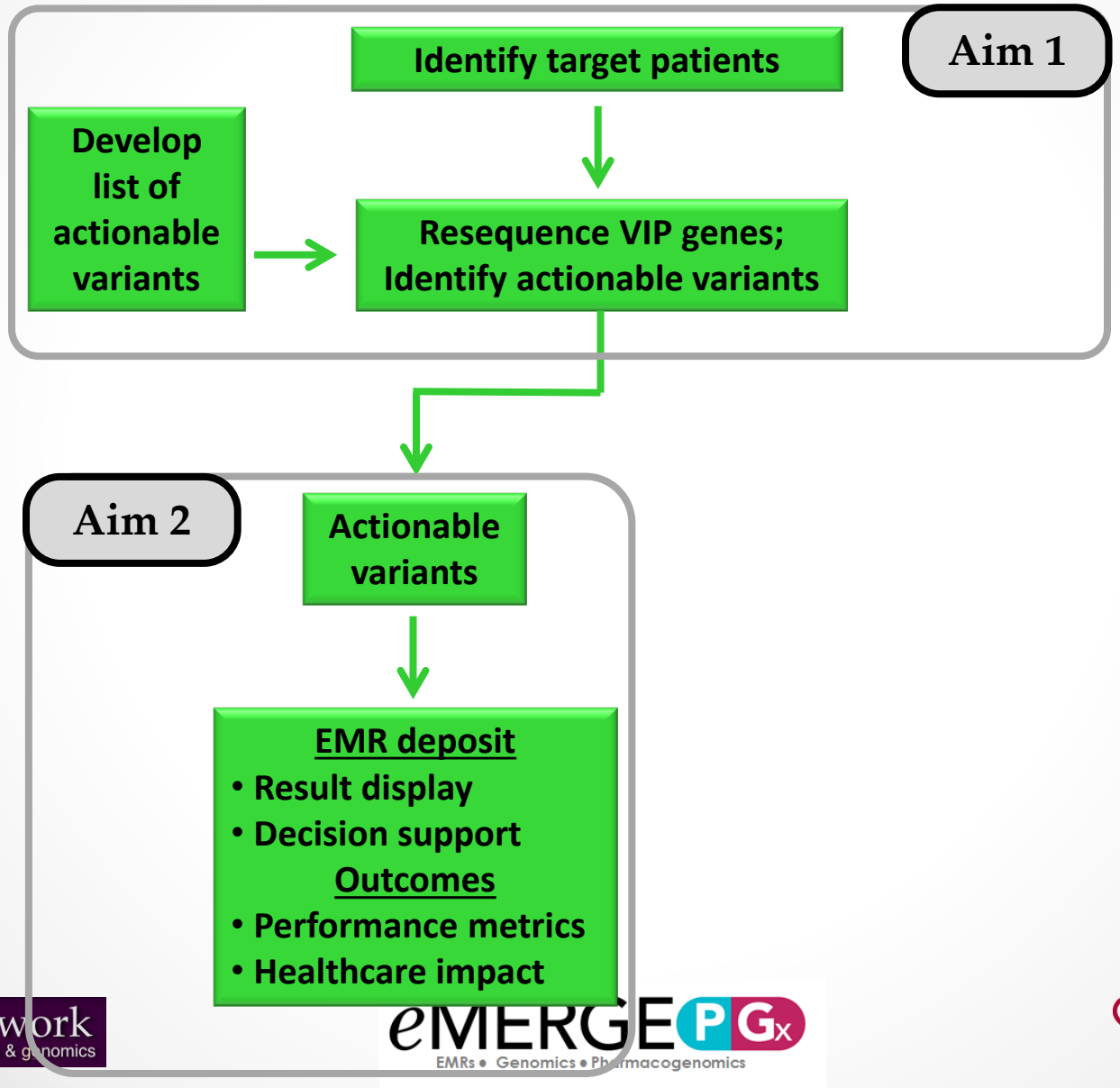
Using a Predictive Algorithm in Recruitment



Participants = Newly Recruited Subjects*

* CCHMC and CHOP have a hybrid approach with a new subject cohort and an existing subject cohort being reconsented.

eMERGE-PGx project

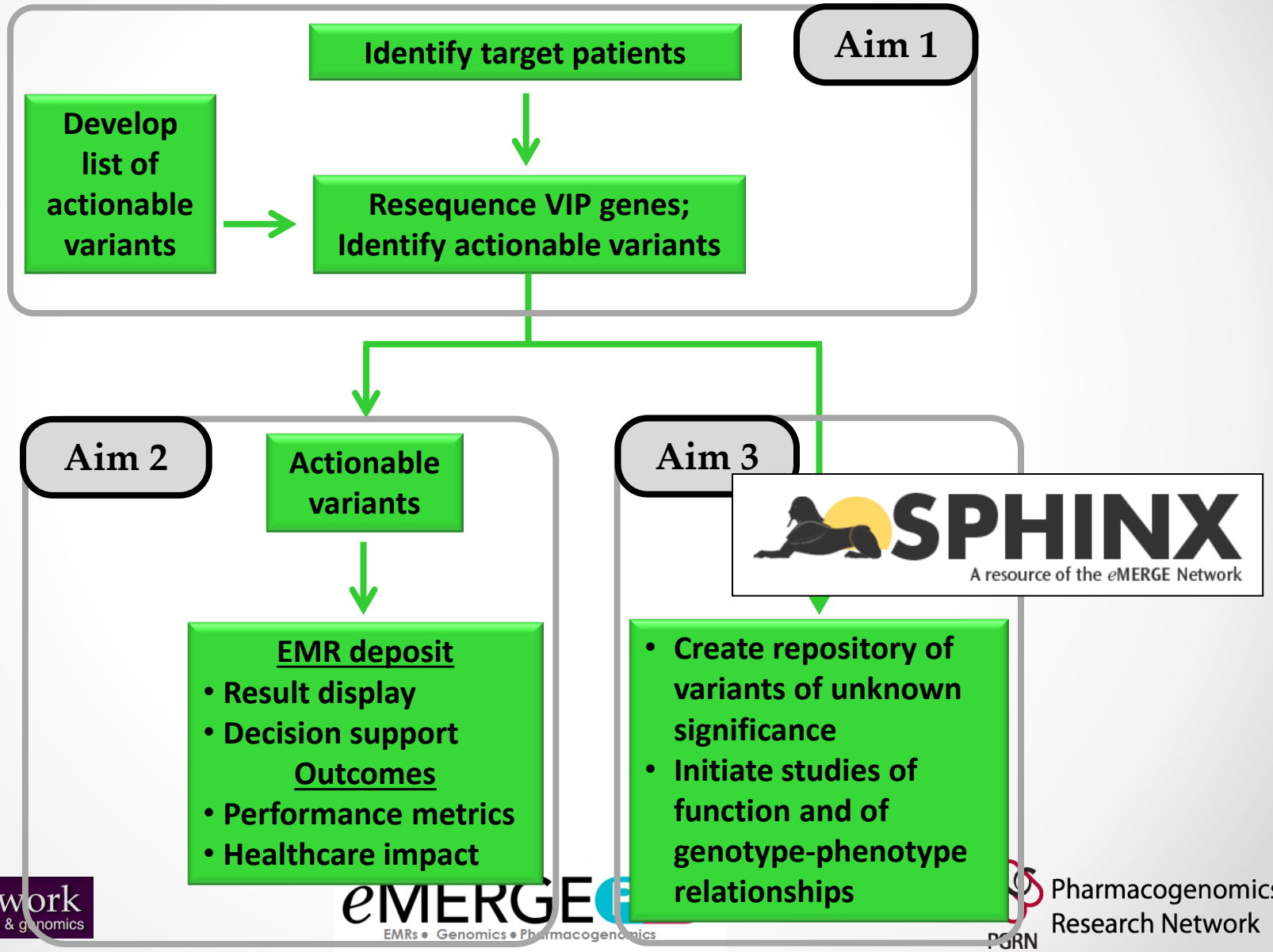


eMERGE Network
electronic medical records & genomics

eMERGE PGx
EMRs • Genomics • Pharmacogenomics

PGRN Pharmacogenomics Research Network

eMERGE-PGx project



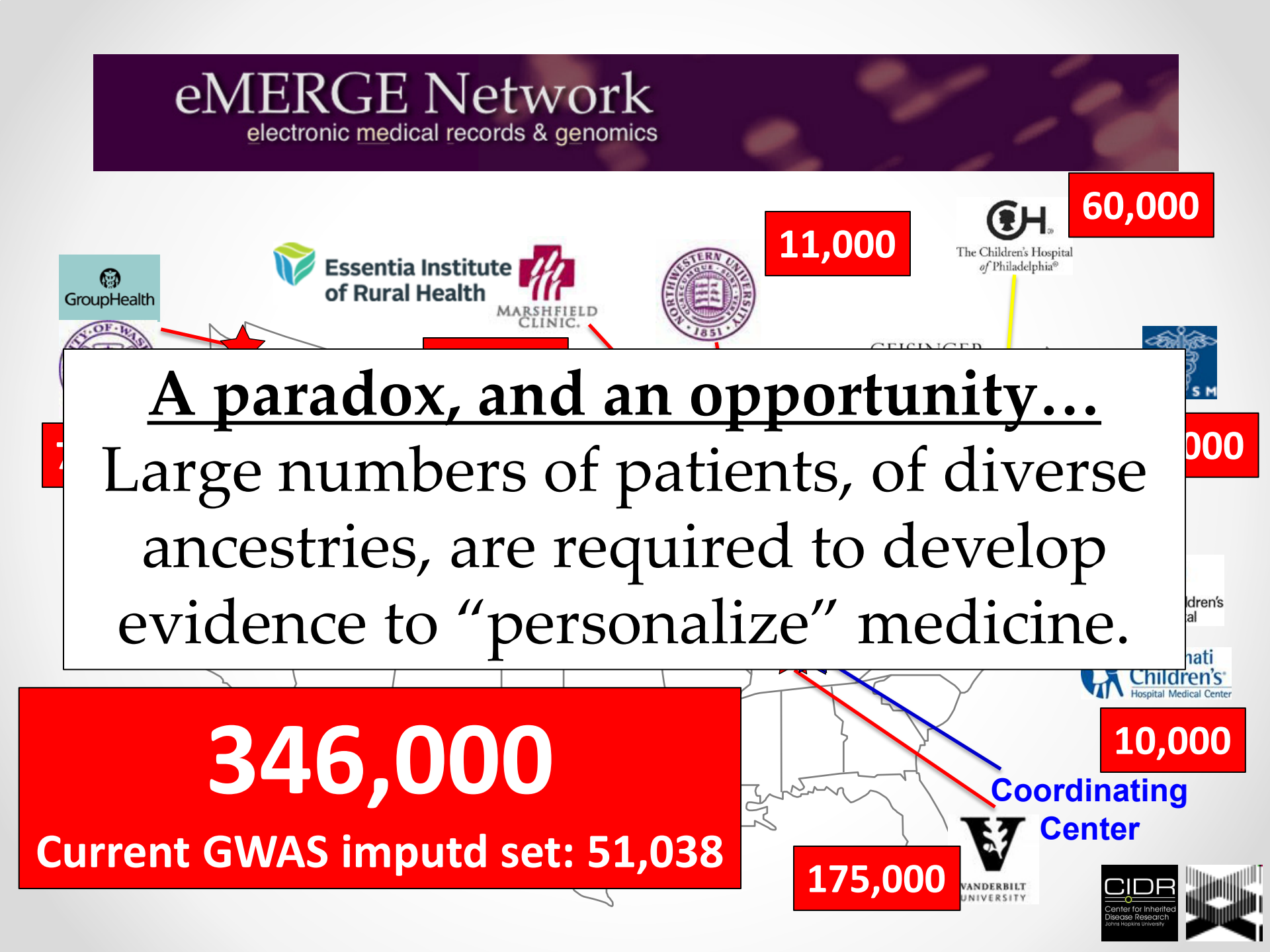
Outcomes

Process outcomes

- Recruitment
- PGRN-Seq Sequencing metrics
- Comparison to Validation Genotyping
- EMR Integration and Clinical decision support
- Returned Results
- Education: clinicians, patients

Healthcare outcomes

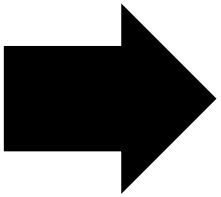
- Statins: Myopathy, Drug Switch
- Clopidogrel: Stent or ACS event? Within 30 days?
- Warfarin: Time to steady state? Time out of range? Bleeding? Thrombosis?
- Thiopurines: Blood counts, (disease outcome), …
- Return of results project: 6 ACMG "actionable" genes


eMERGE Network
electronic medical records & genomics
60,000
11,000
GroupHealth
Essentia Institute of Rural Health
MARSHFIELD CLINIC
The Children's Hospital of Philadelphia
A paradox, and an opportunity…
Large numbers of patients, of diverse ancestries, are required to develop evidence to "personalize" medicine.
346,000
Current GWAS imputd set: 51,038
10,000
Coordinating Center
175,000
VANDERBILT UNIVERSITY
CIDR
Center for Inherited Disease Research

Anonymizing records while enabling research

a) Original Records

Record	*ICD-9 Codes*
49532	{427.31, 401.00, 401.01}, {695.40}
579852	{810.03, 053.00}
778954	{681.11}, {427.31}, {810.03}
794456	{427.31}, {401.00}, {810.03}

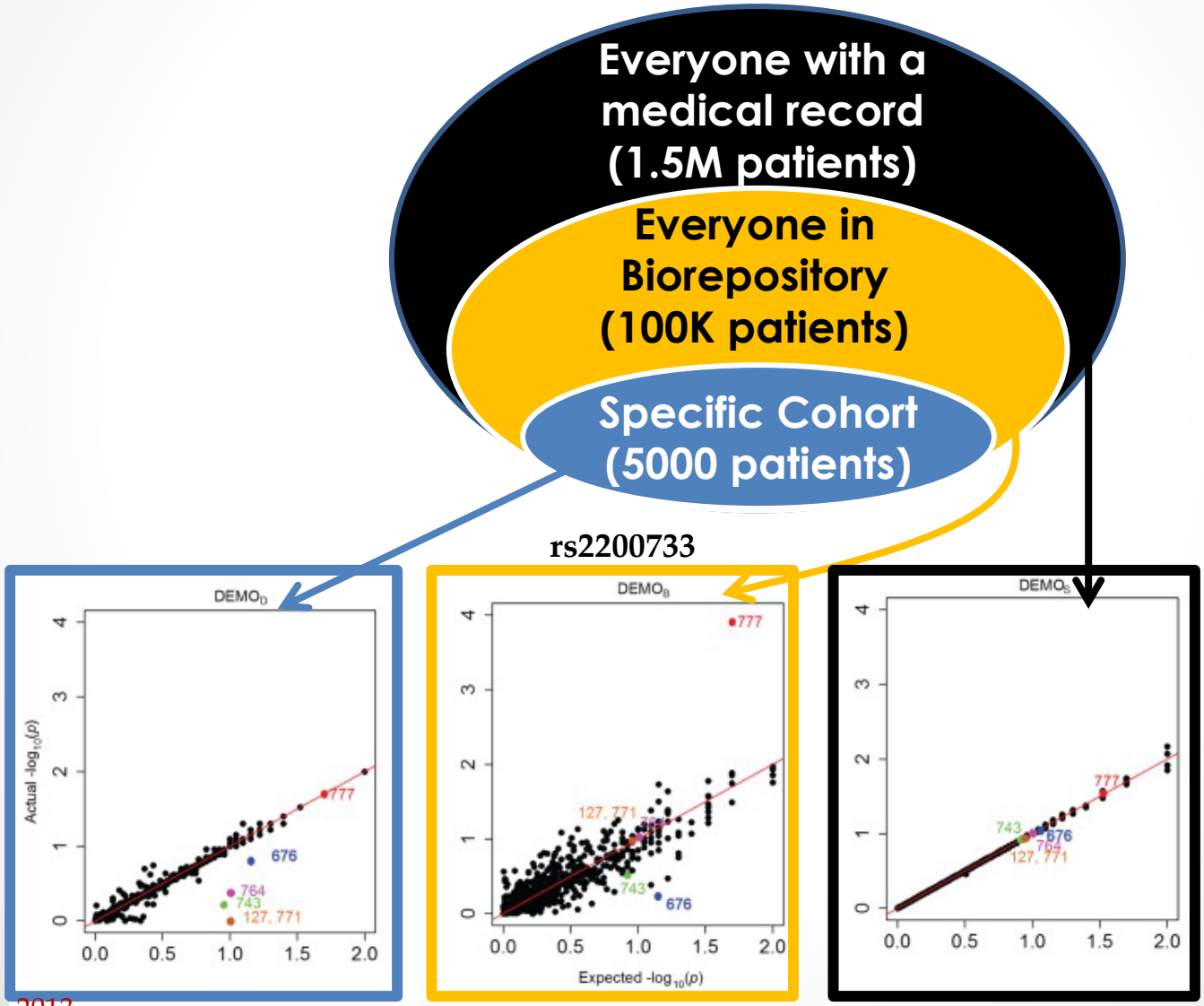


b) Anonymized Records

ID	*ICD-9 Codes*
1	427.31, 695.40, ⟨401.00, 401.01⟩
2	810.03
3	427.31, 695.40, 810.03
4	427.31, ⟨401.00, 401.01⟩, 810.03

- At least k subjects with specified code groups (k=2 in this example)
- Test this scheme by setting k=5 and examining 192 phenotype-genotype associations in
  - 5,944 k-anonymized records
  - 5,944 records drawn from 104,904 k-anonymized records (biobank)
  - 5,944 records drawn from 1,366,786 k-anonymized records (entire EMR)

Anonymizing records while enabling research



Heatherly, et al. 2013

Creation of the eMERGE Genomics dataset

- Creation of QC Pipeline – High throughput and high quality
- Generating a merged set across multiple genotyping platforms → Imputation
- eMERGE-I (5 sites)
  - 2 platforms: Illumina 660 & 1M
- eMERGE-II (10 sites)
  - Illumina 1M, Illumina 660W, Affymetrix 6.0, Illumina HumanOmni Express, Illumina Metabochip, ADME Illumina, Illumina Immunochip, Illumina Metabochip, Illumina OMNI 1, Illumina OMNI 5.

BEAGLE Imputed Data (Adult Sites only)

	# Genotyped Samples	# BEAGLE Imputed SNPs
Merged eMERGE-I 1M	2,634	
Merged eMERGE-I 660	16,029	
Adult sites (unmerged)	*19,625*	
Adult Site Total	**38,288**	**15,212,466**

Impute2 Imputed Data (Adult and Pediatric)

	# Genotyped Samples
Merged eMERGE-I 1M	2,634
Merged eMERGE-I 660	16,029
Geisinger	3,111
Group Health	731
Marshfield	500
Mayo	3121
Mt. Sinai	6,290
NU	2,951
Vanderbilt	3,461
BCH	1,038
CCHMC	4,322
CHOP	6,850
Total - All Impute2 Imputed Samples	**51,038**

EMR-linked biobanks in eMERGE-II

Site	#of subjects with DNA samples linked to EMRs
Geisinger	22,000
Mt. Sinai	22,000
CHOP	60,000
Cincinnati/Boston	10,000
TOTAL	346,000

A paradox, and an opportunity…

Large numbers of patients, of diverse ancestries, are required to develop evidence to "personalize" medicine.

EMR-linked biobanks in eMERGE-II

Site	#of subjects with DNA samples linked to EMRs
Group Health Seattle	7,000
Marshfield	20,000
Mayo	19,000
Northwestern	11,000
Vanderbilt	175,000
Geisinger	22,000
Mt. Sinai	22,000
CHOP	60,000
Cincinnati/Boston	10,000
TOTAL	346,000