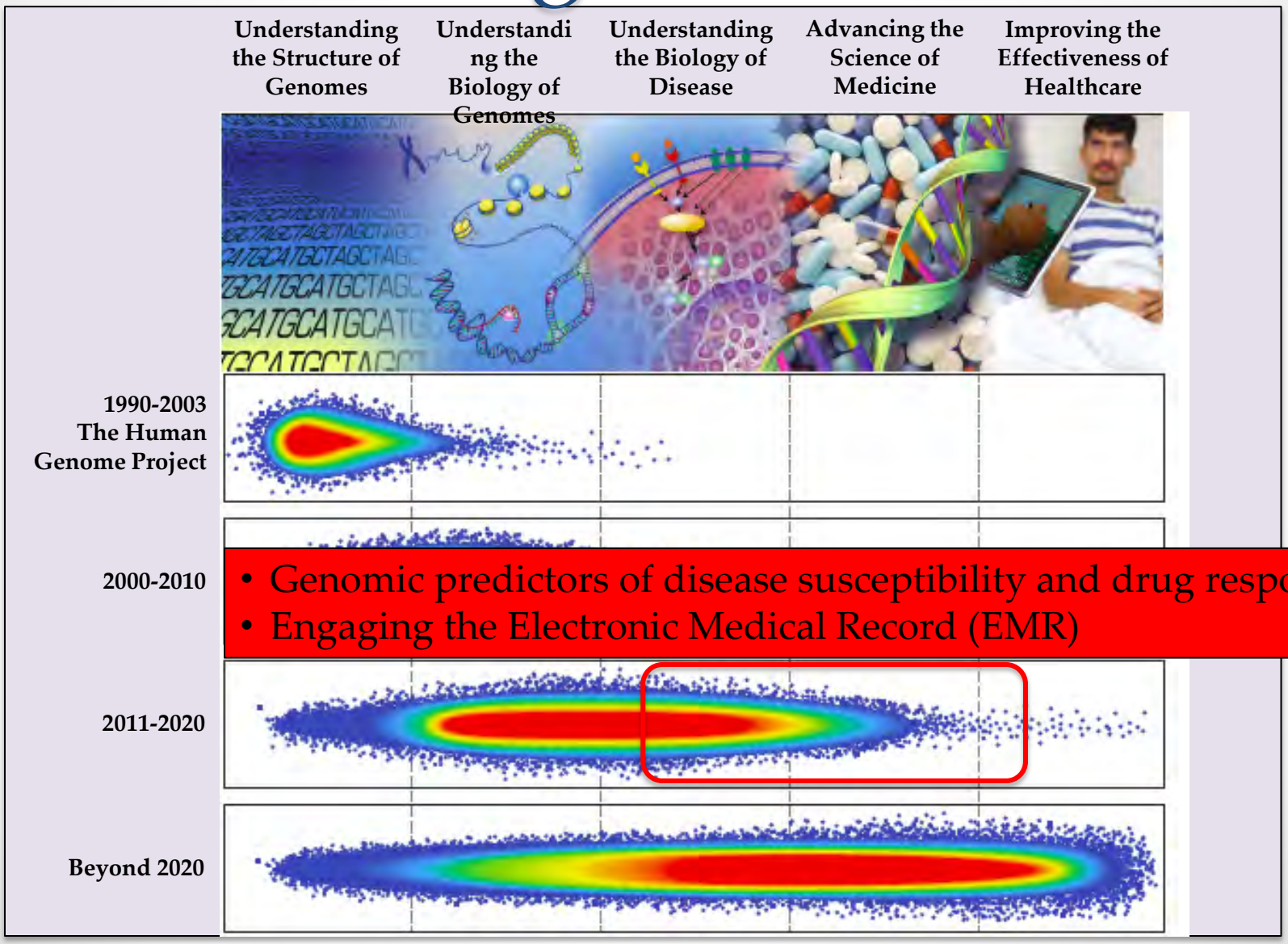
eMERGE Network
electronic medical records & genomics

The eMERGE Network

Rex Chisholm, Steering Committee Chair
On Behalf of the eMERGE network
January 22, 2014

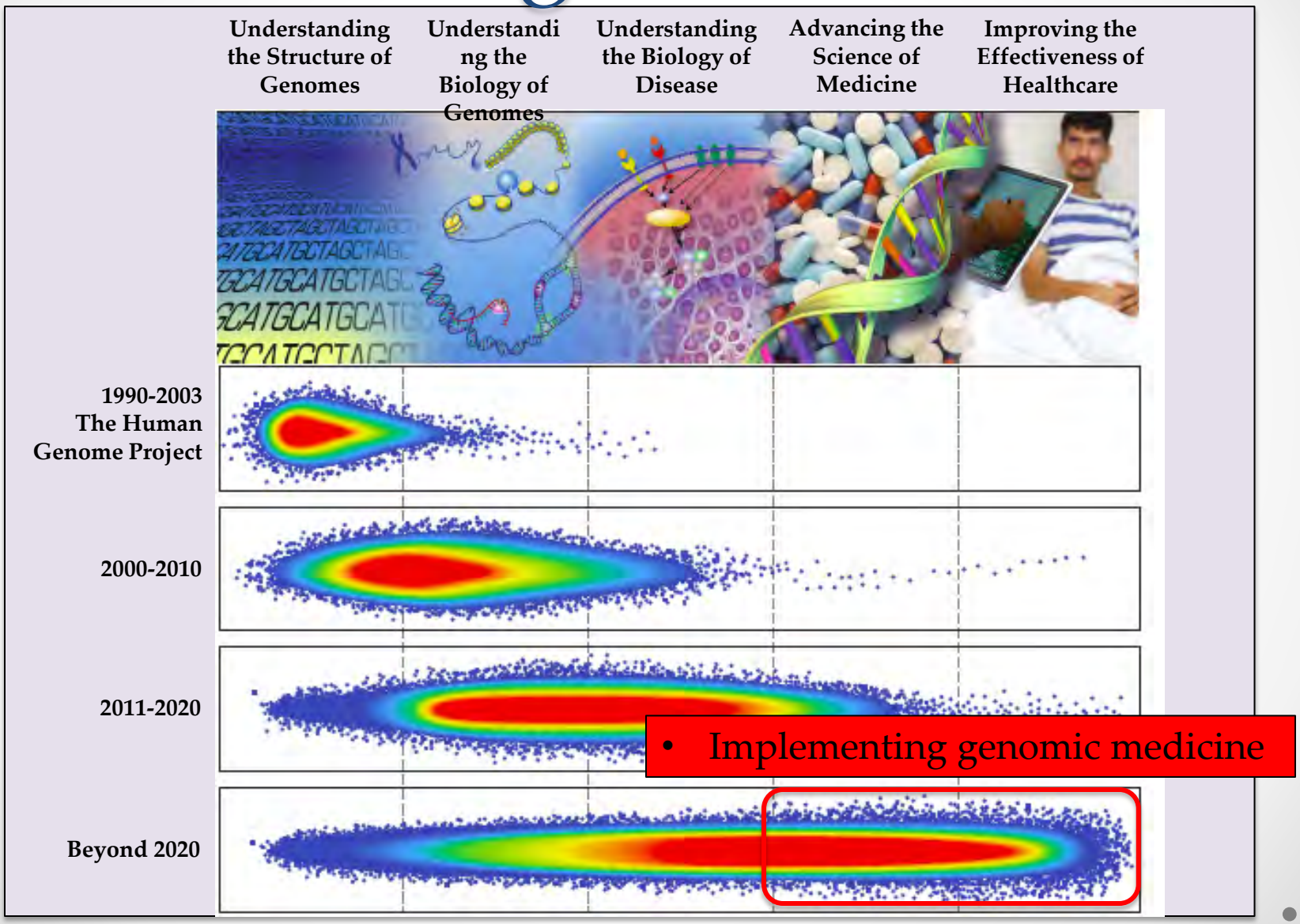
GroupHealth · Marshfield Clinic · University of Washington · Essentia Health · Mount Sinai School of Medicine · Geisinger · The Children's Hospital of Philadelphia · Mayo Clinic · Boston Children's Hospital · Northwestern University · Cincinnati Children's Hospital Medical Center · Vanderbilt University

Domains of genomic research



Green ED, Guyer MS. *Nature* 2011; 470:20

Domains of genomic research



Green ED, Guyer MS. *Nature* 2011; 470:20

eMERGE Network

electronic medical records & genomics

Phase I: Sept. 2007 – July 2011

Mayo, Vanderbilt, Group Health Cooperative/Univ. Washington, Marshfield Clinic, Northwestern

Phase II: Aug. 2011 – July 2015

Phase I sites plus Geisinger Health and Mt. Sinai

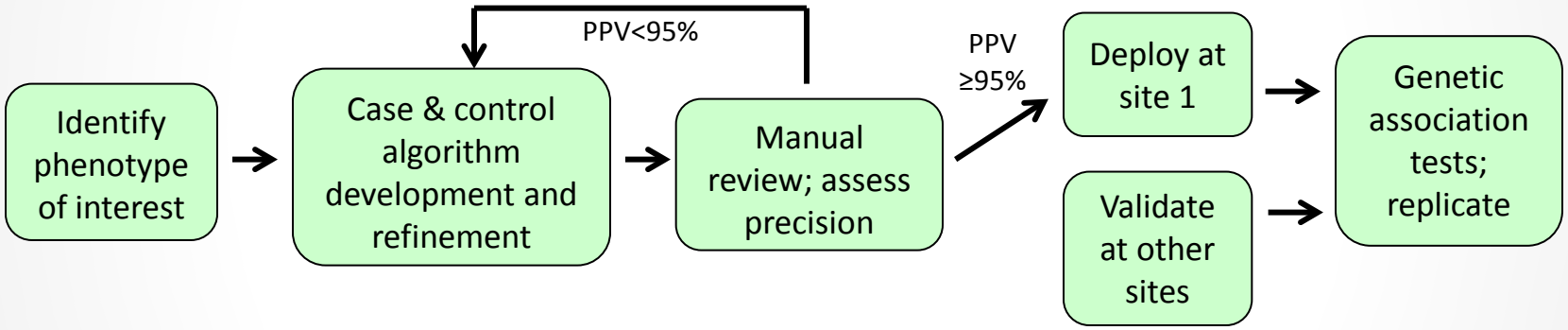
Pediatric sites (CHOP and BCH/CCH) added 2012

eMERGE I: Discovery

Goals: Test the ability to leverage EMRs and biobanks for genomic research

- Electronic Phenotyping: Develop & validate electronic algorithms for 1° and 2° phenotypes
- Genomics: Conduct association studies of genome-wide data with EMR-derived phenotypes and share data widely through dbGaP
- Assess adequacy of existing consent for genomic technologies & data sharing
- Develop best practices for GWAS in the areas of electronic phenotyping, genomics & analytics, and ELSI topics

Approach to electronic phenotyping



eMERGE-I phenotypes

`	Primary Site # Genotyped		Addl. Data from Network	
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

Phase I Phenotypes

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

Primary Phenotype-Gene Associations in eMERGE I

Associations between 19 phenotypes and 38 genes

Disease Phenotype	Gene
Cardiac Conduction	SCN5A, SCN10A
Hypothyroidism	FOXE1
LDL Cholesterol	APOE, TRIB1, LPL, ABCA1
Platelet Count & Volume	5 Chromosomes Associated with PLT & 8 with MPV
Glaucoma, Primary Open-A	
Glaucoma, Optic Nerve De	
Red Blood Cell Traits, Erythroid Differentiation and Cell Cycle Regulation	THRB, PTPLAD1, CDT1
RBC Traits, Erythrocyte Sedimentation Rate (ESR)	CR1
RBC Traits, Malaria Resistance	HBB, HBA1/HBA2, G6PD
RBC Traits, Peripheral Artery Disease (PAD)	SLC17A1, BLS1/MYB, TMPRSS6, HFE
White Blood Cell Count	DARC, GSDMA, MED24, PSMD3

….and MORE

Merged Genotype Dataset

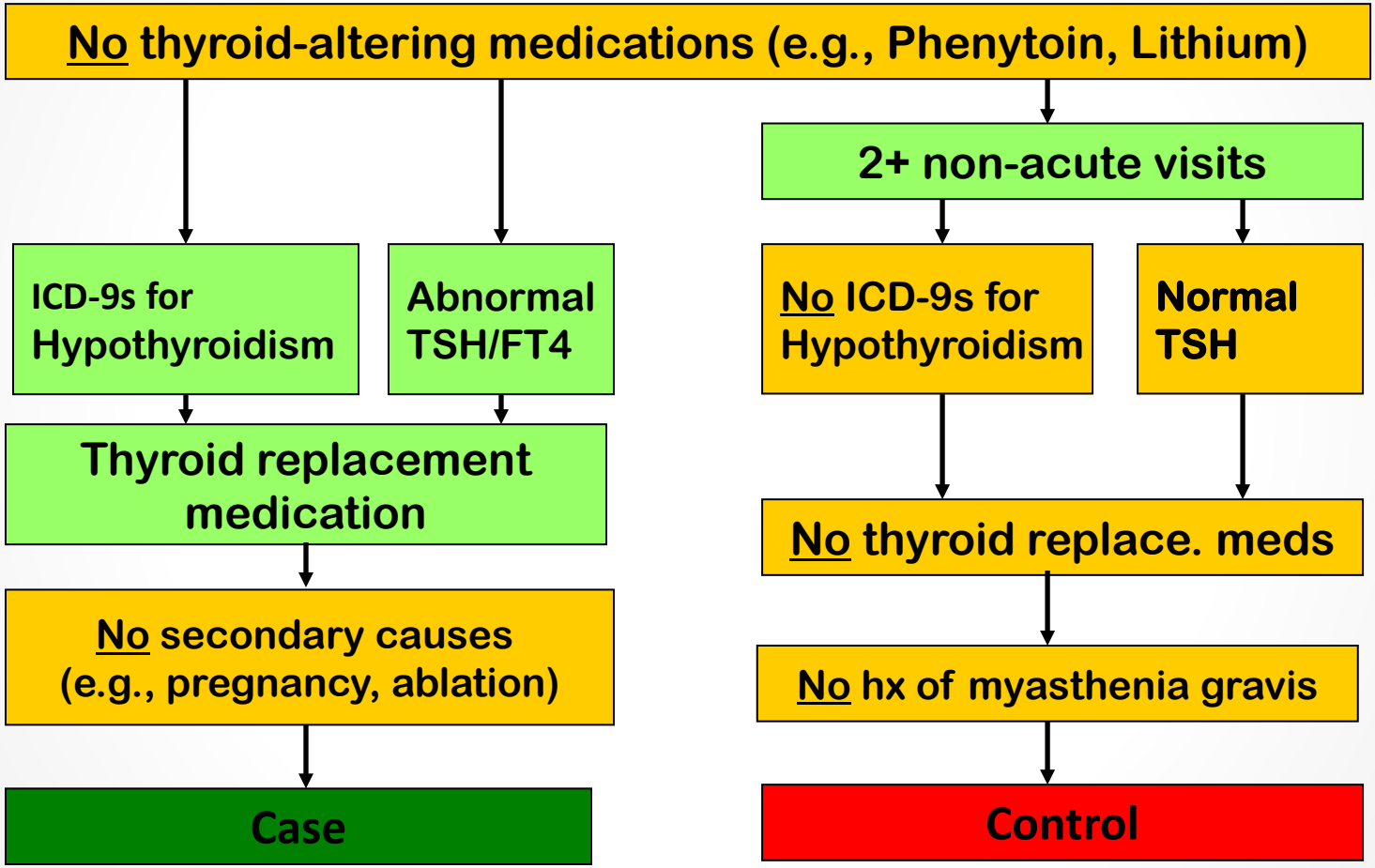
- 17,046 eMERGE samples with GWAS data
- Majority of samples genotyped using 660W
- Samples collected for various studies
  - GH – Dementia
  - Marshfield – Cataracts and HDL-C
  - Mayo – PAD
  - NW – T2D
  - VU – Normal ECGs

Can we use existing dataset for another experiment?

Primary Hypothyroidism

- Most common form is chronic lymphocytic hypothyroidism (Hashimoto's thyroiditis)
- More common in females (~10x)
- Other associated factors
  - age
  - race/ethnicity
  - family history of thyroid disease
- No published* GWAS (as of 07/23/2011)

Hypothyroidism Algorithm



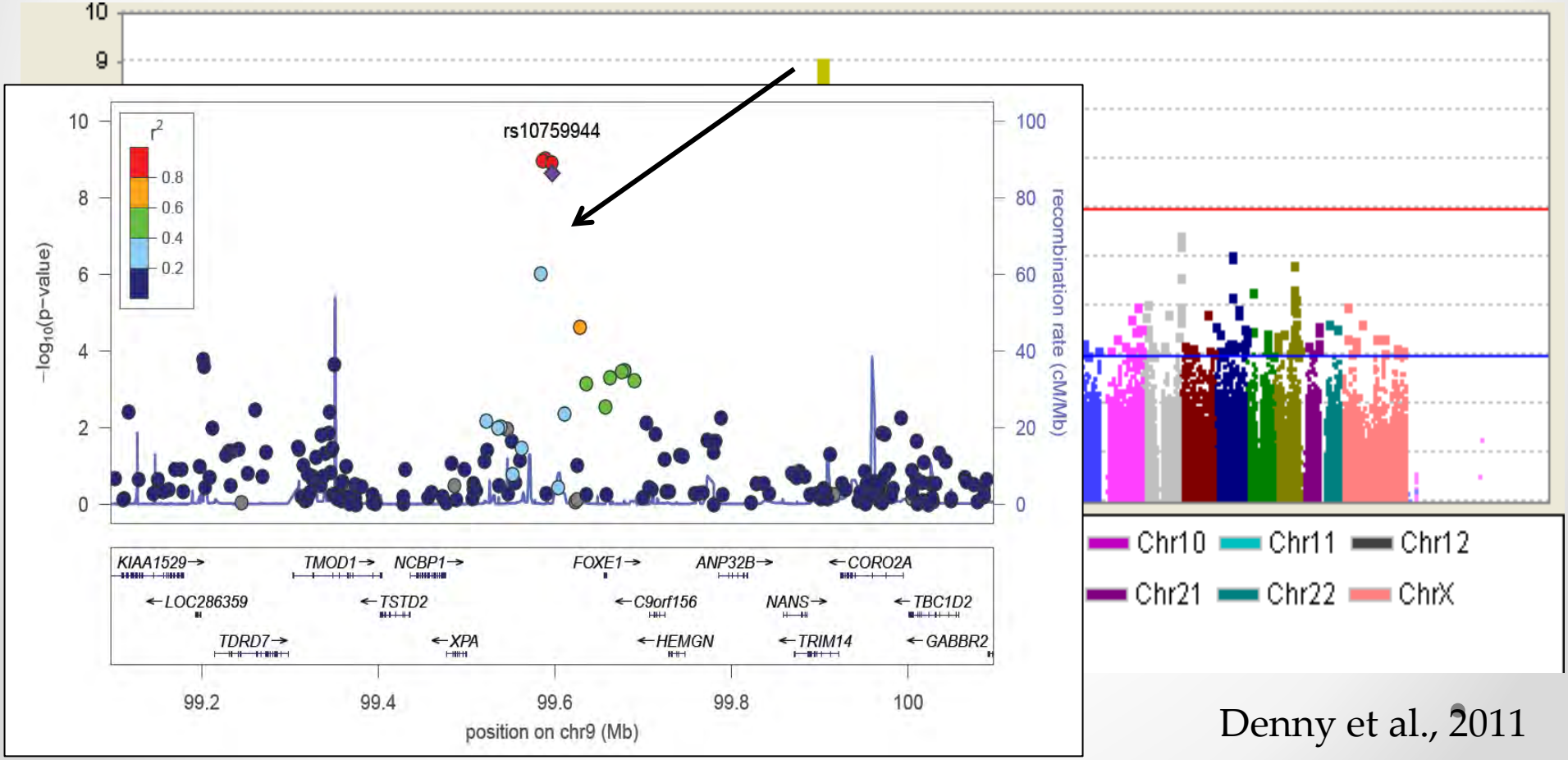
Phenotype Algorithm Validation

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

Am J Hum Genet. 2011 Oct 7;89(4):529-42

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

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



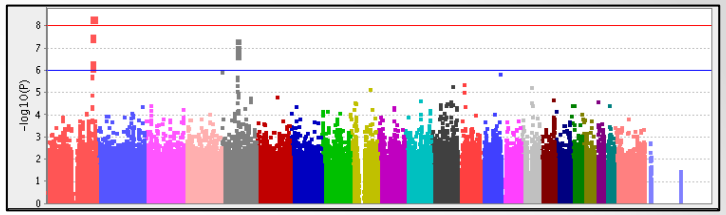
Denny et al., 2011

FOXE1

- Forkhead box E1 (thyroid transcription factor 2)
- ~3.46kb intronless gene
- Thyroid transcription factor which likely plays a crucial role in thyroid morphogenesis
- Mutations associated with congenital hypothyroidism and cleft palate with thyroid dysgenesis
- The map localization of this gene suggests it may also be a candidate gene for squamous cell epithelioma and hereditary sensory neuropathy type I

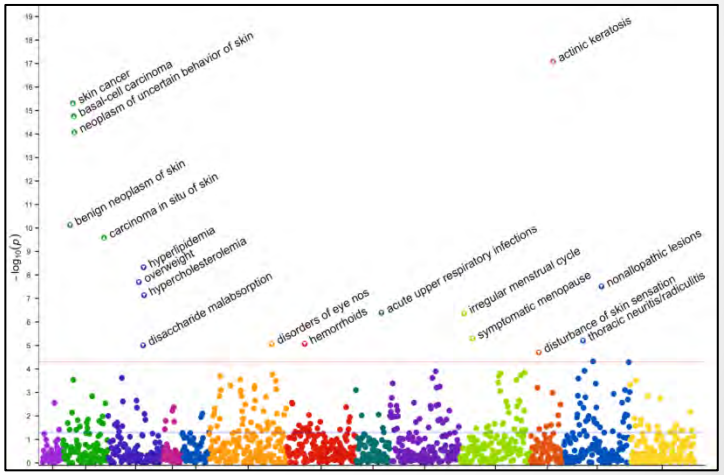
From EntrezGene

GWAS: Target phenotype → association P value



The phenome-wide association study (PheWAS)

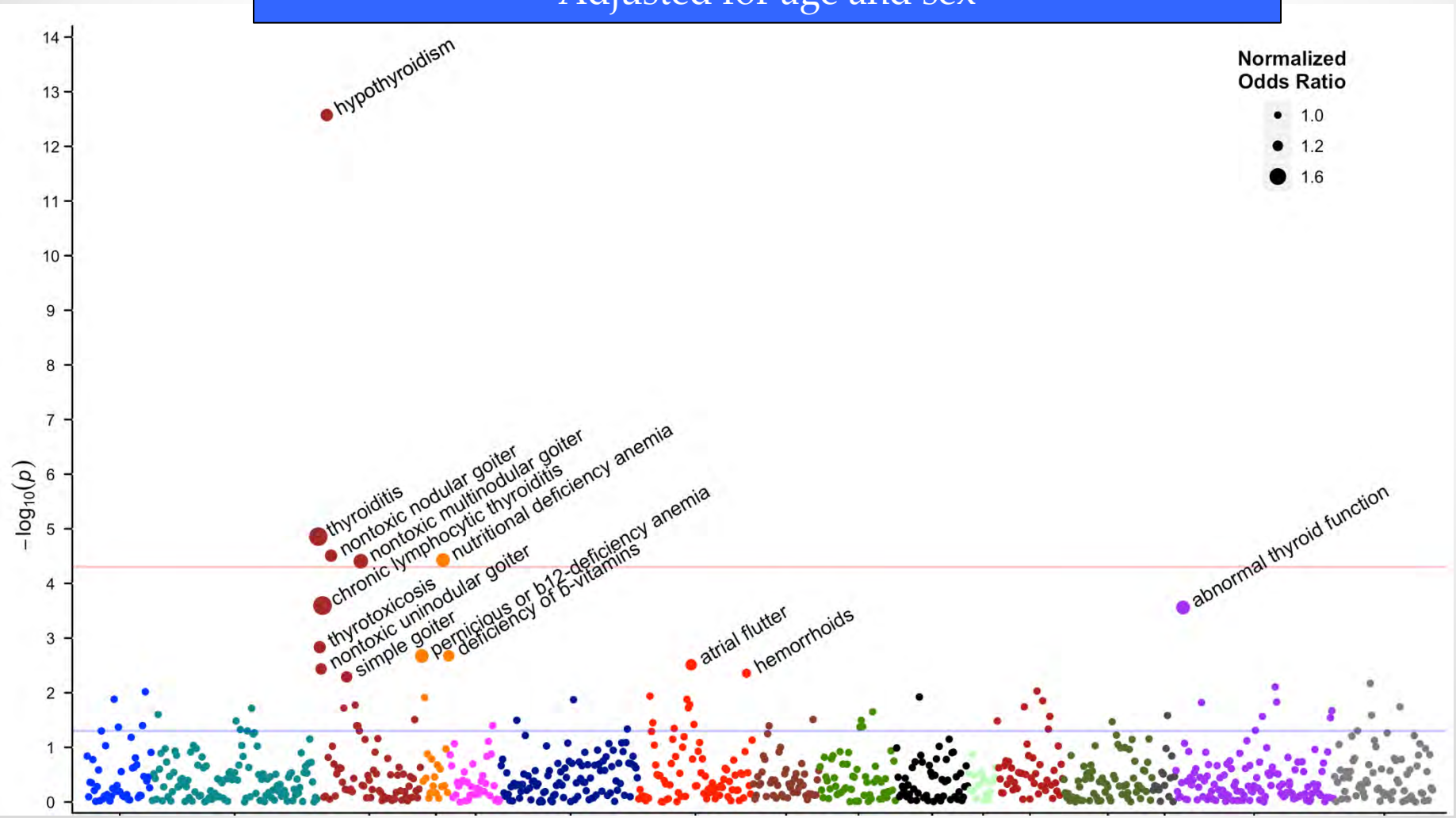
PheWAS (ΦWAS): Target genotype → association P value



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

PheWAS for rs965513 (*FOXE1*)

Analysis of 866 phenotypes in 13,617 European Americans
Adjusted for age and sex



Hypothyroidism Conclusions

- “No new genotyping”experiment yields new genetic associations
- Associations overlap with thyroid cancer GWAS
- Associations replicate in external dataset
- *FOXE1* likely candidate, may be associated with other thyroid diseases

Discovery to Practice: Integrating Genomics into the EMR

Phase 1

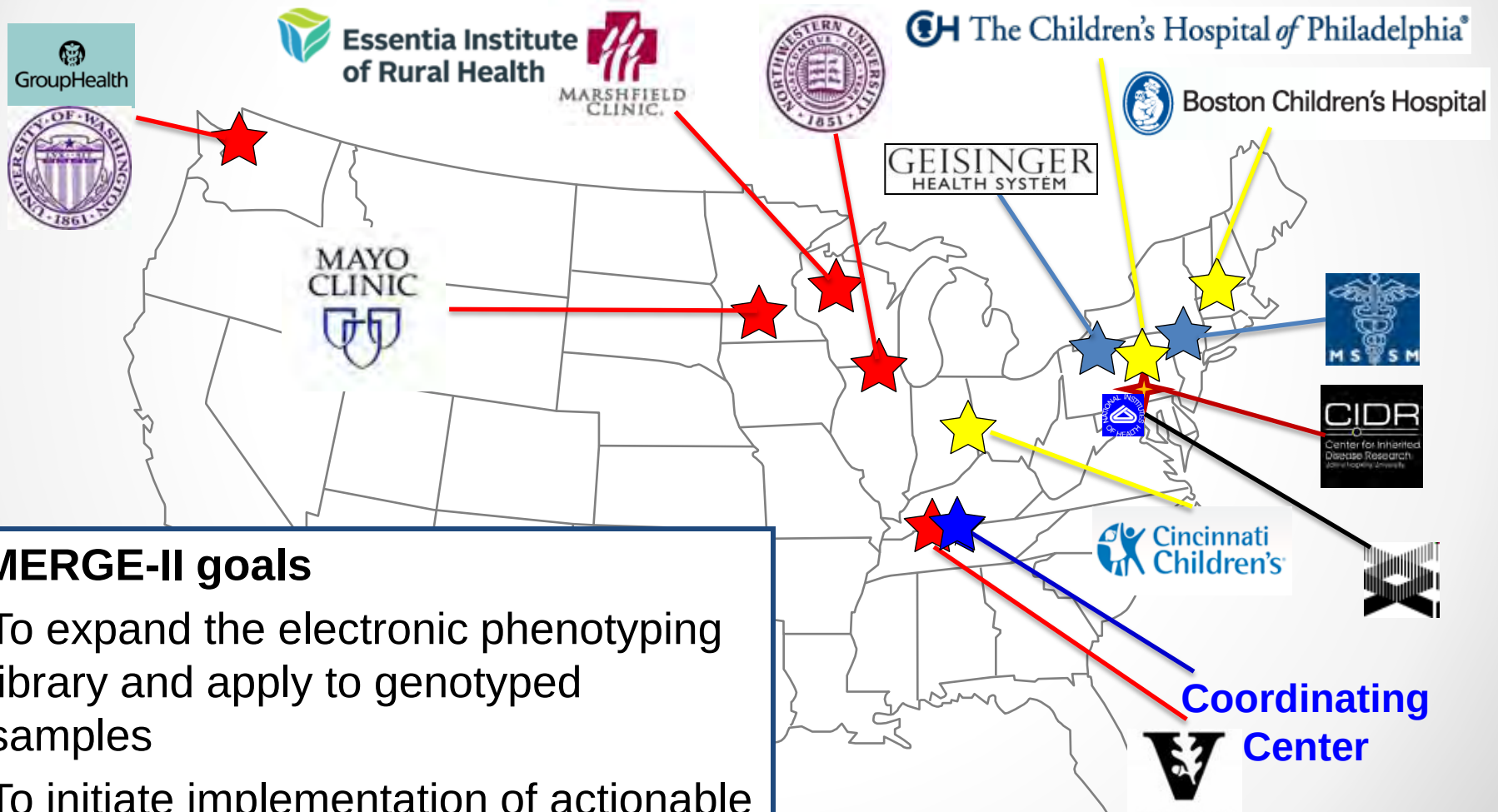
- Use EHR data for genome/phenome associations
- Cross institution phenotypes

Phase 2

- More sites
- Pediatrics
- Faster, better, cheaper EHR-based genomic science
- **Integration of genomic information back into EHR and clinical care**

eMERGE Network

electronic medical records & genomics



eMERGE-II goals

- To expand the electronic phenotyping library and apply to genotyped samples
- To initiate implementation of actionable variants into the EMR

eMERGE II: Discovery & Implementation

- 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: Discovery & Implementation

- ***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

EMR-linked biobanks in eMERGE-II

Site	Participants	(GWAS)-Genotyped Samples
Group Health Seattle	6,381	3,606
Marshfield	20,000	4,693
Mayo	19,000	6,934
Northwestern	11,000	4,987
Vanderbilt	158,514	27,173
Geisinger	22,000	4,191
Mt. Sinai	22,000	6,290
CHOP	60,000	45,000
Cincinnati/Boston	10,000	5,360
TOTAL	328,895	108,234

eMERGE Phenotyping: Sharable, High-Throughput

- Methods
  - **44 phenotypes** (complete or in development)
  - Sharable formats – KNIME, QDM
  - Machine Learning algorithms
  - Portable Methods
- Tools
  - PheKB with data standardization & validation tools
  - eMERGE RC
  - PheWAS
  - Downloadable NLP Tools – cTAKES, MedEx

Phase II Phenotyping

Phenotyping Algorithm Development

Phenotype	Lead Site	Secondary Site	Status
Clostridium difficile	GroupHealth	Vanderbilt	Complete
Abdominal aortic aneurysm	Geisinger	Mayo	Complete
Venous Thromboembolism	Mayo	Vanderbilt	Complete
Ocular Hypertension	MC/EIRH/PSU	Geisinger	Complete
Diverticulosis	Northwestern	Vanderbilt	Complete
Glaucoma	MC/EIRH/PSU	Geisinger	Complete
Herpes Zoster	GroupHealth	Vanderbilt	Complete
ACE-Inhibitor Induced Cough	Vanderbilt	Northwestern	Complete
Cardio Respiratory Fitness	Mayo	Geisinger	Complete
Extreme Obesity	Geisinger	MC/EIRH/PSU	Complete
Asthma	CHOP	MC/EIRH/PSU	Complete
Child Obesity	CCHMC/BCH	CHOP	Complete
Drug Induced Liver Injury	Mt. Sinai/Columbia	Mayo	Complete
Heart Failure	Mayo	GroupHealth	In process (exp. Jan. 2014)
Colon Polyps	Northwestern	MC/EIRH/PSU	In process (exp. Jan. 2014)
Autism	CCHMC	BCH	In process (exp. Jan. 2014)
Statins for MACE	Vanderbilt	MC/EIRH/PSU	In process (exp. Jan. 2014)
Age-related Macular Degeneration	MC/EIRH/PSU	Northwestern	In process (exp. Jan. 2014)
Atopic Dermatitis	CHOP	MC/EIRH/PSU	In process (exp. Jan. 2014)
Remission of Diabetes after ROUX-EN-Y	CHOP	MC/EIRH/PSU	In process (exp. Feb. 2014)
CAAD as Quantitative Measure	Geisinger	Northwestern	In process (exp. Feb. 2014)

Upcoming Phenotypes: Upper GI/PUD, GERD, Appendicitis, Epilepsy, Lipids, Pulmonary HTN, Diabetic Hypertensive CKD, Rapid Renal Decline in Diabetic HTN Nephropathy, caMRSA, ADHD

Standard Pediatric Measures

Milestones:

- Height, Weight, BP, Heart Rate
- Head Circumference (Brain Growth)
- Gross/Fine Motor Function (Coordination, Locomotion)
- Cognitive Development (Language, Reasoning)
- Communication (Receptive and Expressive)
- Sensory (Vision and Auditory)
- Social and Emotional Development
- Self Care (Sleep, Activity Level)

- ~Standard Time-Points (Well-Child Visits):
  - 1 month
  - 3 months
  - 6 months
  - 1 year
  - 2 years
  - 3 years
  - 4 years
  - 5 years

eMERGE Record Counter

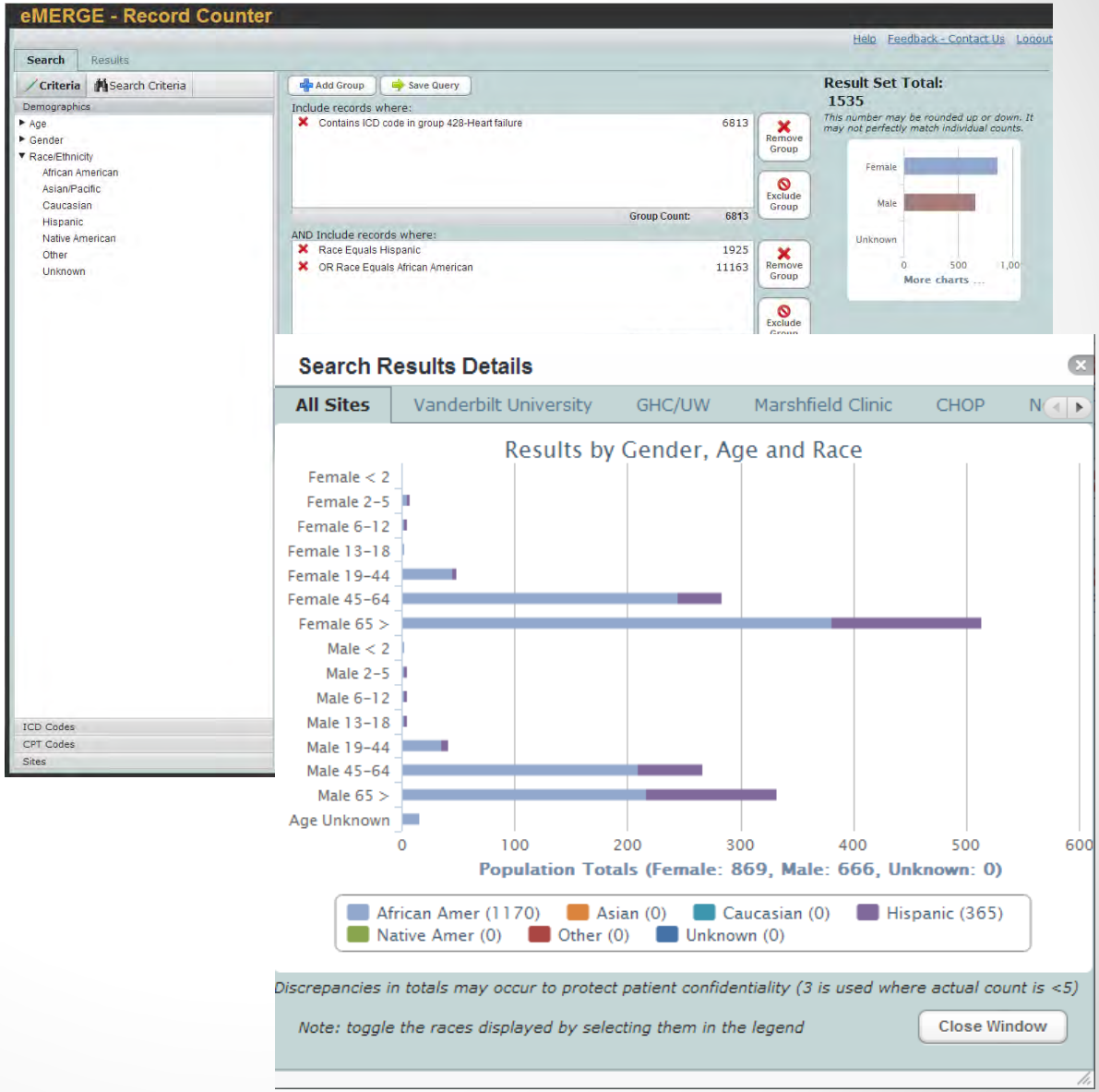
The Record Counter is a web-based tool that provides exploratory data figures to the members of the eMERGE community for research planning purposes and feasibility assessment.

What is returned?

Data counts stratified by sex, ancestry, age and site.

What data is available?

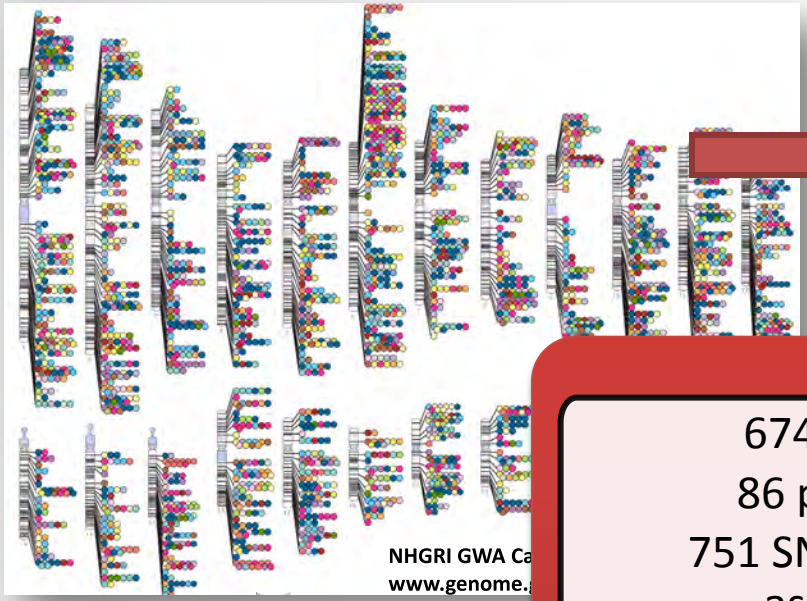
Over **52,000 records are available** in the Record Counter. The data incorporated in the eRC includes both the eMERGE 1 merged set data and eMERGE 2 data from Mt. Sinai, Essentia/Marshfield/PSU, Vanderbilt, Group Health/UW, Mayo Clinic, Northwestern, Geisinger, Boston Children's, Cincinnati Children's, and CHOP.



PheKB.org: a knowledgebase for discovering phenotypes

- Public Side of the Website
  - Final phenotypes algorithms and implementation results
- Private Side of the Website
  - Ability to create phenotype workspaces and collaborate on development, testing, and validation on algorithms
  - Version history
  - Access controls
    - Only viewable by authors
    - Shared with a collaborative groups(s)
    - Public ("Final" algorithms only)
  - Comment on Phenotypes shared with you
  - Receive alerts on Phenotypes you are following

PheWAS of "all" NHGRI GWAS Catalog SNPs



3,144 SNPs with prior GWAS-discovered associations

674 SNPs with
86 phenotypes
751 SNP-phenotype associations

Test for replication of 751 associations using PheWAS

Replication Arm

3,144 SNPs

PheWAS for each SNP to discovery pleiotropy

Replication of novel associations

Discovery Arm

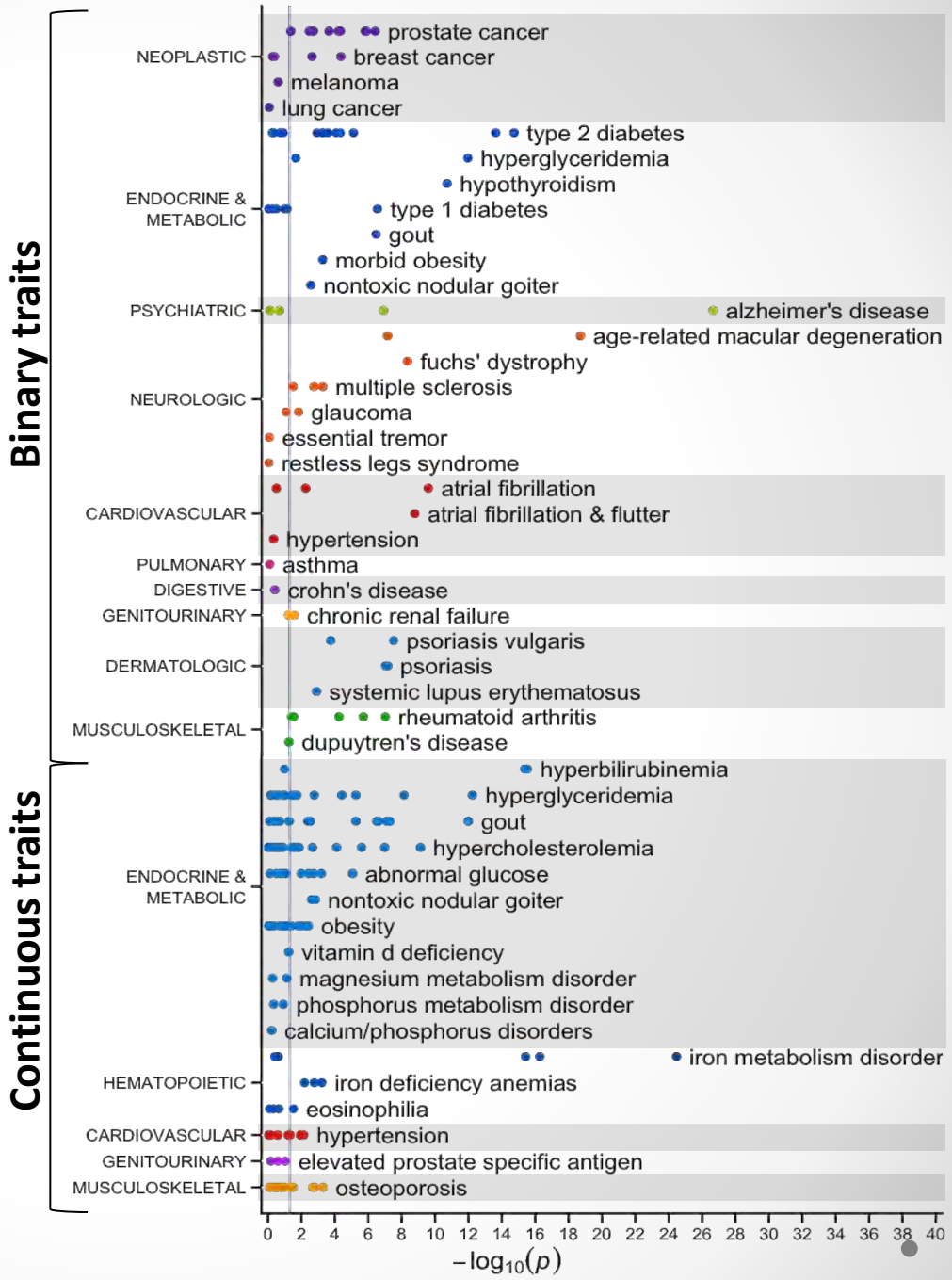
Denny et al, Nat Biotech 2013

Replications of NHGRI GWAS associations via PheWAS

Probability of replicating:

- All - 210/751: $2x10^{-98}$
- Powered - 51/77: $3x10^{-47}$

Denny et al, Nat Biotech 2013



eMERGE II: Discovery & Implementation

- 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

Genomics: Creation of the eMERGE Dataset

- Creation of **QC Pipeline** – High throughput and high quality
- Different genotyping platforms required a solution for a merged dataset: **Imputation**
- eMERGE-I
  - included 5 sites
  - 2 plattforms: Illumina 660 & 1M
- eMERGE-II
  - Included 7 adult sites, 3 pediatric
  - Illumina 1M, Illumina 660W, Affymetrix 6.0, Illumina HumanOmni Express, Illumina Metabochip, ADME Illumina, Illumina Immunochip, Illumina Metabochip, Illumina OMNI 1, Illumina OMNI 5, etc.

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)

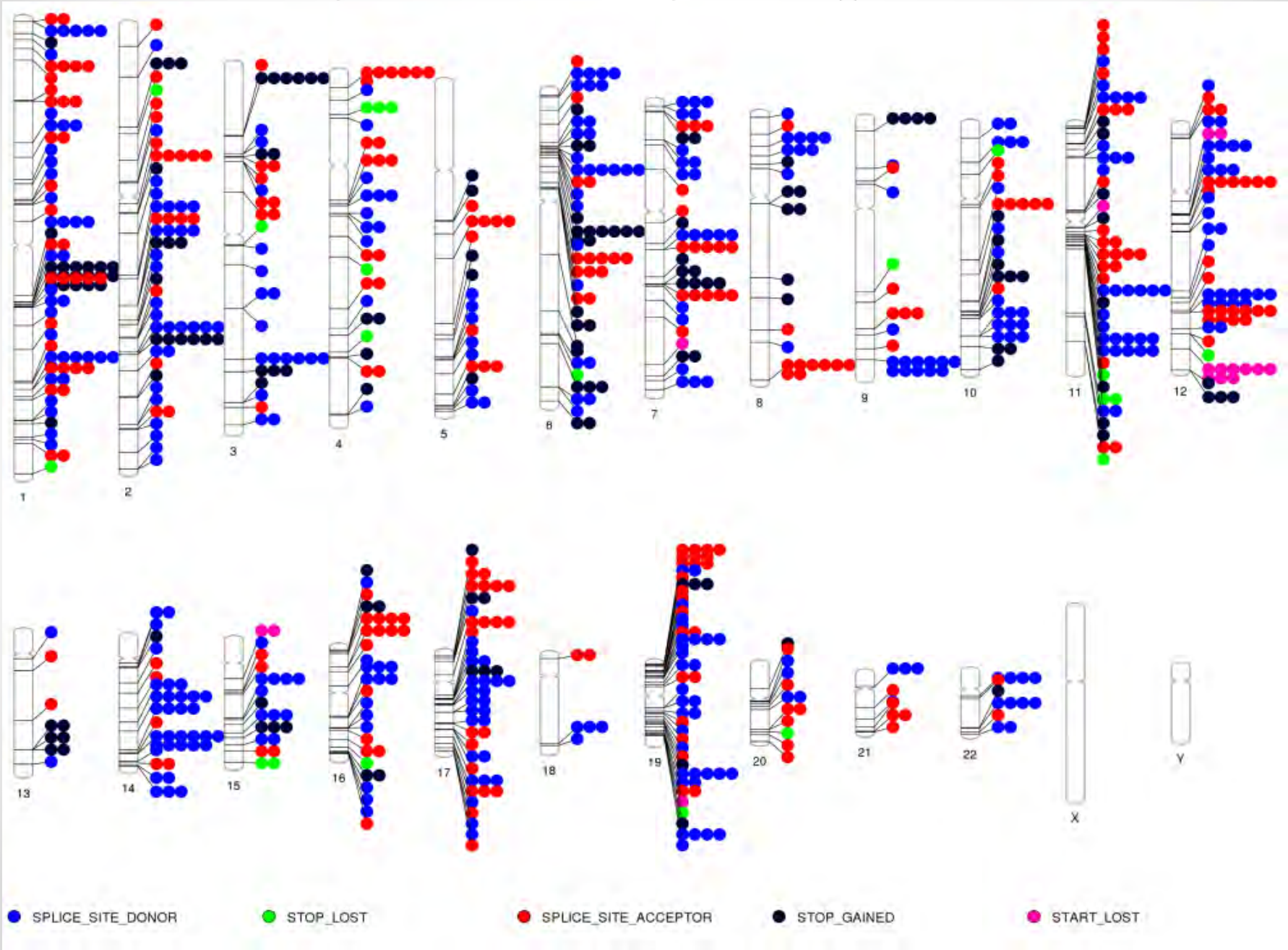
	# Genotyped Samples
Merged eMERGE-I 1M	2,634
Merged eMERGE-I 660	16,029
Geisinger	3,111
Group Health	731
Marshfield	500
Mayo	3121
Mt. Sinai	6,290
NU	2,951
Vanderbilt	3,461
BCH	1,038
CCHMC	4,322
CHOP	6,850
Total - All Impute2 Imputed Samples	**51,038**

Loss of Function (Null) Variants Project

Hypothesis: Use large eMERGE dataset to identify multiple rare, null variants and look for correlation with clinical traits

- Use bioinformatics approach to predict null variants
- Explore genotyped and imputed datasets for the occurrence of null variants
- Run PheWAS analysis on high impact variants to look for association with clinical characteristics

Phenogram of High Impact Variants



eMERGE II: Discovery & Implementation

- 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 Site-Specific Genomic Medicine Implementation Pilots

Site	Goal
CCHMC	CYP2D6 variants and post-operative opioids
CHOP	TPMT and LOF variants impacting adverse response to thiopurine drugs
Geisinger	IL28B variants and chronic hepatitis C treatment; WGS for undiagnosed diseases
Marshfield	CFH and risk of age-related macular degeneration
Mayo	RCT of 42 SNP-genomic risk score for CHD vs Framingham score alone for attitudes, behaviors
Mount Sinai	RCT of APOL1 genotype vs clinical risk factors for hypertensive nephropathy prevention, management
Northwestern	Effect of return of HFE and FVL risk variants on physician and patient attitudes, behaviors
Vanderbilt	Expanded PGx testing

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?

Returning Results: Hemochromatosis penetrance

- Most common genetic disorder in Caucasians
  - 1:220 individuals
  - 1/9 risk of allele carrier
  - Common alleles C282Y and H63D
- Iron over-absorption leads to end organ damage, easily screened for and treated
- Low penetrance
  - Estimated as 8% in males and <2% in females
  - Impact on screening and treatment
- Plan: Find penetrance for relevant phenotypes in eMERGE (charts)

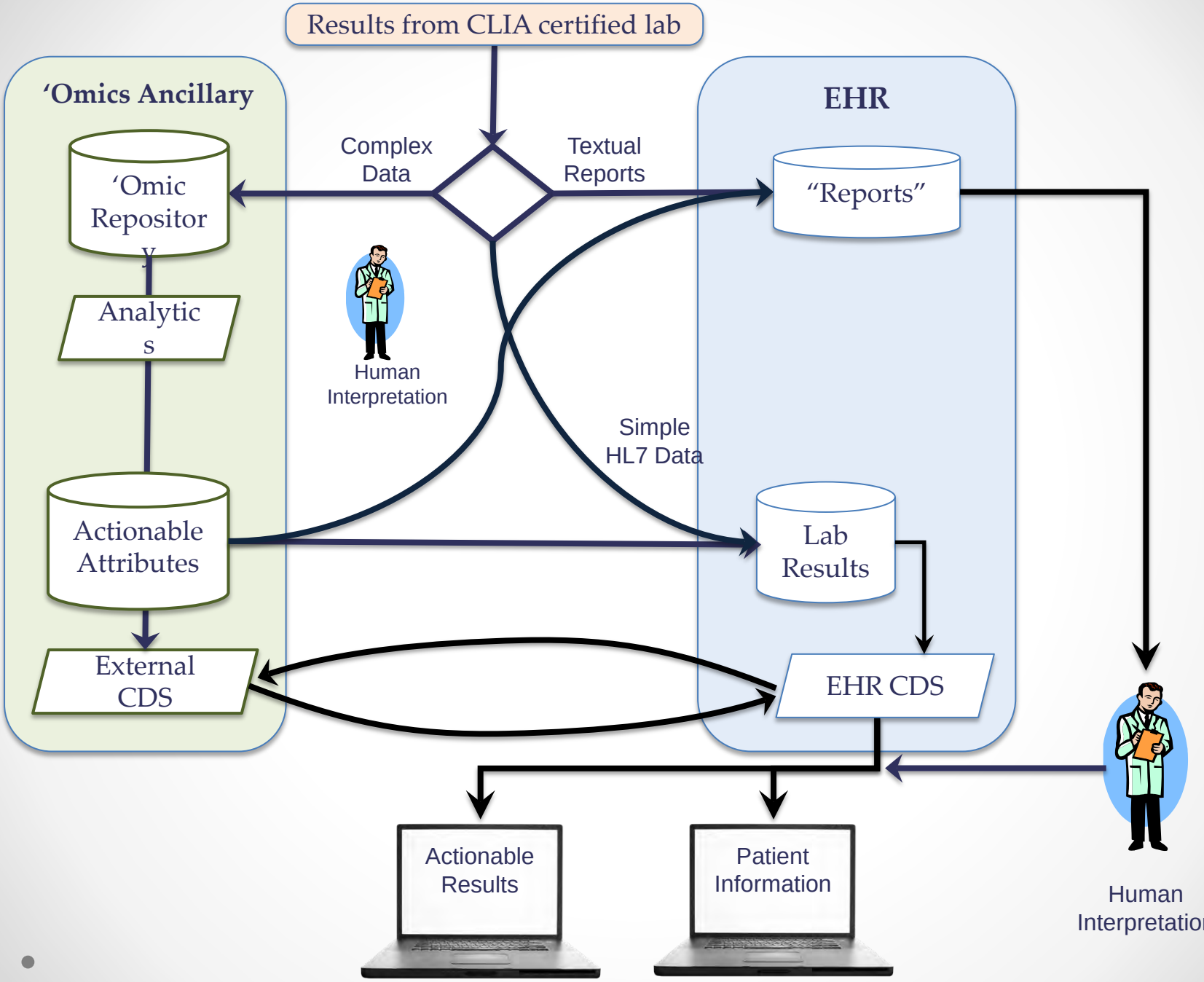
eMERGE II: Discovery & Implementation

- 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

EHR Infrastructure

Institution	EHR Platform for Genomic Integration
Group Health	Epic Ambulatory
Mayo	GE Centricity + Cerner
Marshfield	Internally Developed
Northwestern	Epic + Data Warehouse
Vanderbilt	Internally Developed
Geisinger	Epic + Data Warehouse
Mt. Sinai	Epic
CCHMC	TBD
BCH	TBD
CHOP	EPIC + Data Wareh

Variation in infrastructure and project goals makes common CDS development impractical.
Engaged, purposeful sharing of information and best practices.

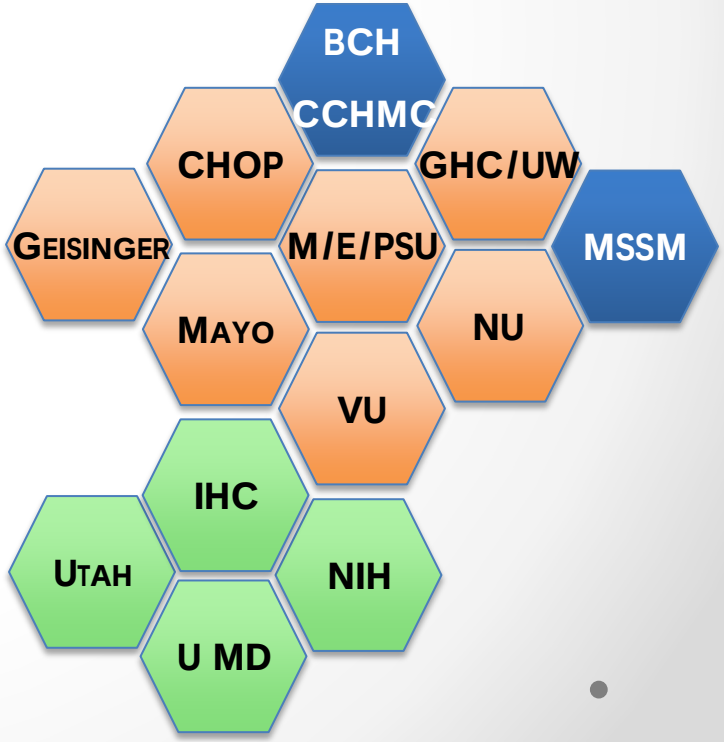

Results from CLIA certified lab
'Omics Ancillary
EHR
Complex Data
Textual Reports
'Omic Repository
Analytics
"Reports"
Human Interpretation
Simple HL7 Data
Actionable Attributes
Lab Results
External CDS
EHR CDS
Actionable Results
Patient Information
Human Interpretation

eMERGE Infobutton Project

Develop a new information resource based on the HL7 Infobutton standard

1. Required by Meaningful Use Stage 2
2. Multiple sites contribute content, all sites can access.

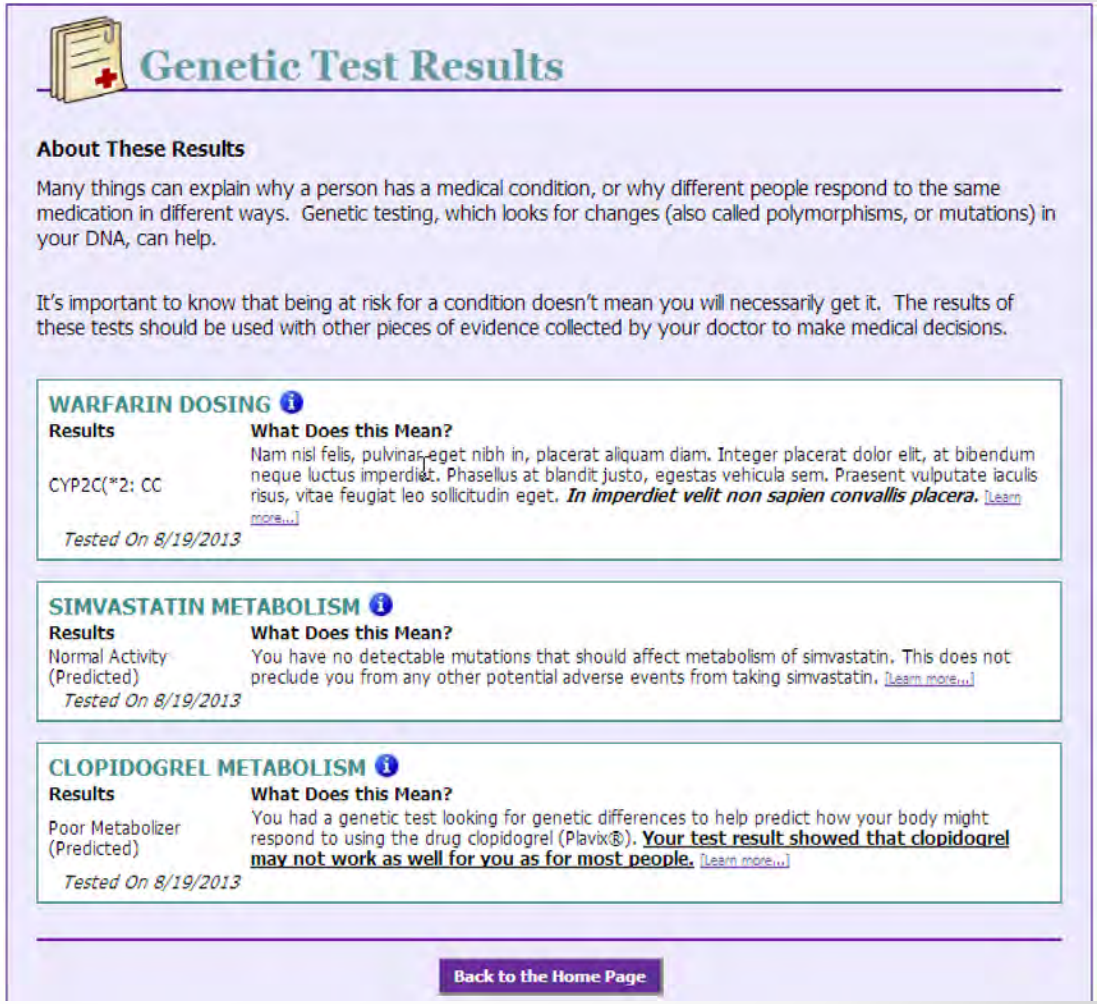
- Representation from many eMERGE sites
  - Crosses EHRI and CERC workgroups
- External collaborators
  - University of Utah – Dr. Guilherme del Fiol
  - Intermountain Healthcare – Dr. Nathan Hulse
  - NIH Clinical Center – Dr. Jim Cimino



eMERGE Infobutton Project

Implement infobuttons within EHRs at eMERGE institutions

- Identify supporting platforms
- Develop systems that don't exist, or contribute features to existing OSS projects
- Assist with implementation and configuration
- Evaluate usage



eMERGE II: Discovery & Implementation

- 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

Challenges in Research Ethics and Policy

- Consent
  - Biobanking → Genomics in the EMR
- Data Sharing
  - Biobank data → Clinical (genomic) data
- Community Engagement → Stakeholder Engagement, Patient & Physician Education

Collaboration: Interactions with CSER

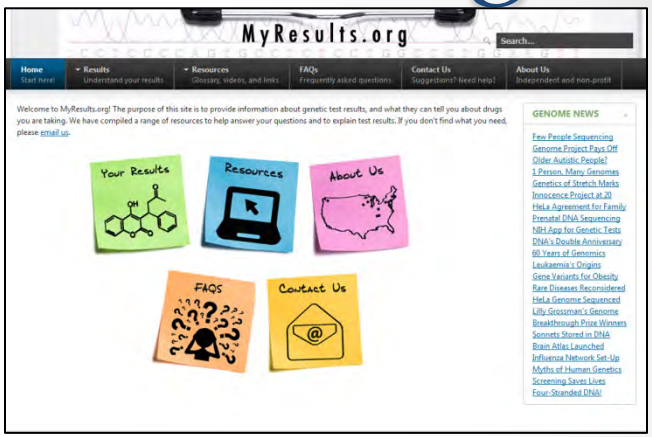
- Combined meeting
- Templated reporting of WGS and WES results
- PGx IFs in the ACMG list
- PGx reporting in CSER
- Contrast MD and patient responses to WES/WGS vs PGx ROR
- CDS

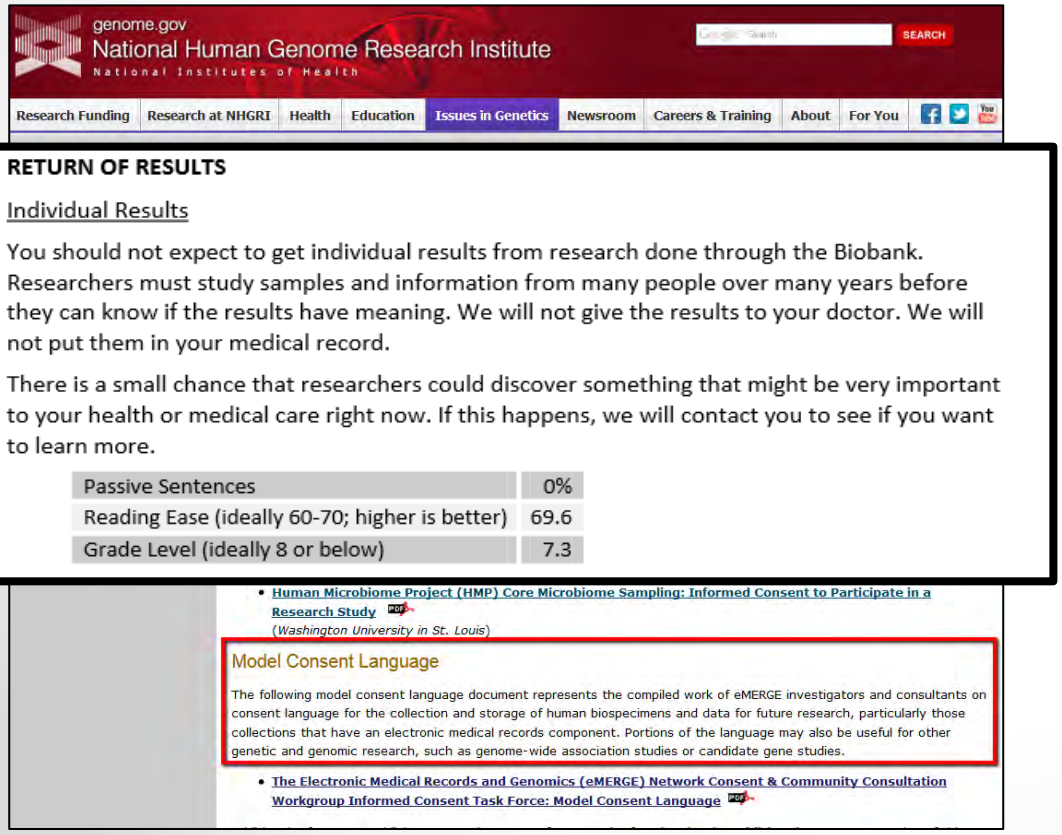
eMERGE II: Discovery & Implementation

- 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

Tools: Model Consent Language and MyResults.org

MyResults.org is a centralized repository of patient education materials as related to eMERGE studies.





The 2009 Model Consent Language document is hosted on NHGRI's Informed Consent website and represents the Network's compiled work on consent language for the collection and storage of human biospecimens and data for future.

Patient survey: biobanking consent issues

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.

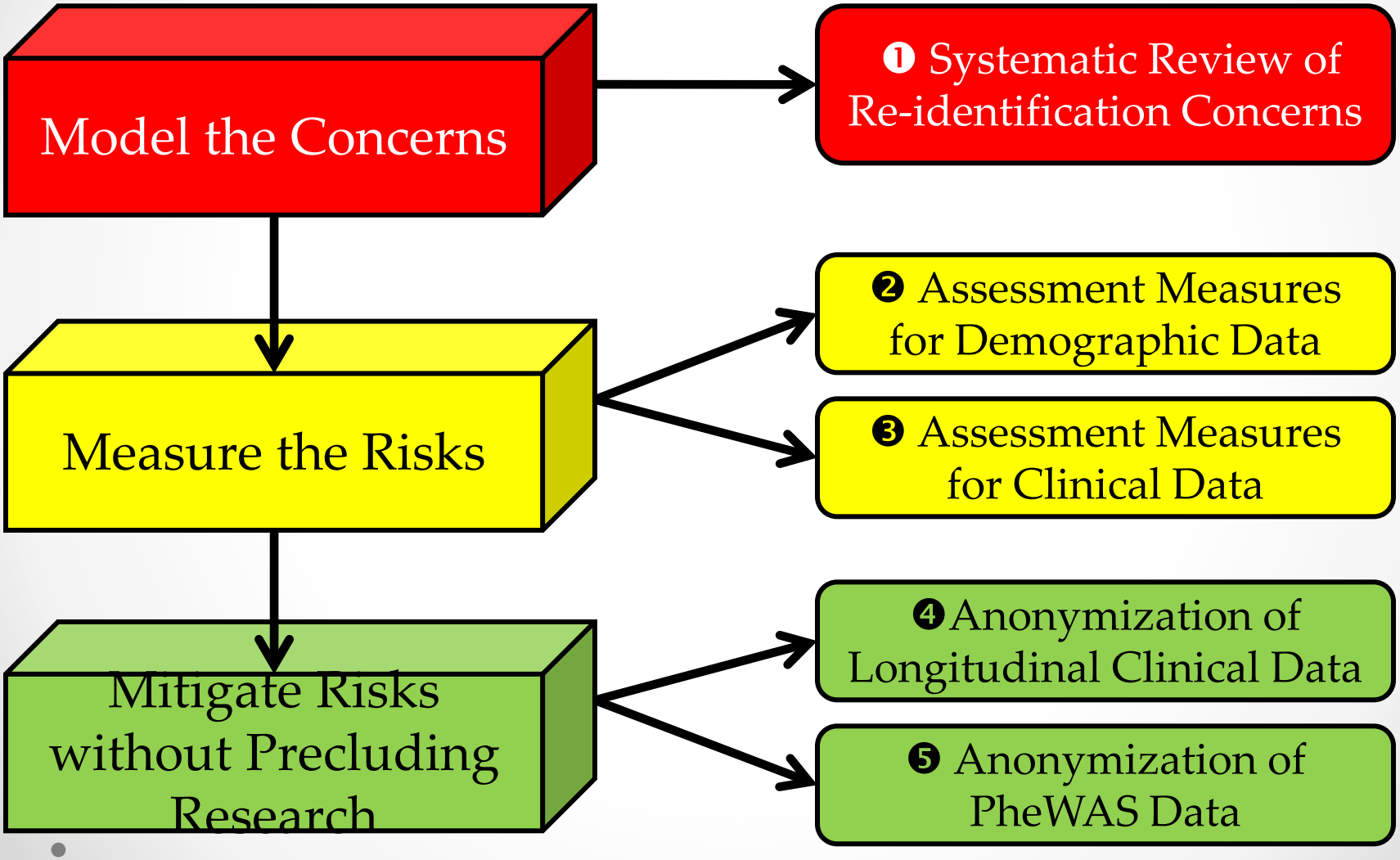
Outcome

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

eMERGE II: Discovery & Implementation

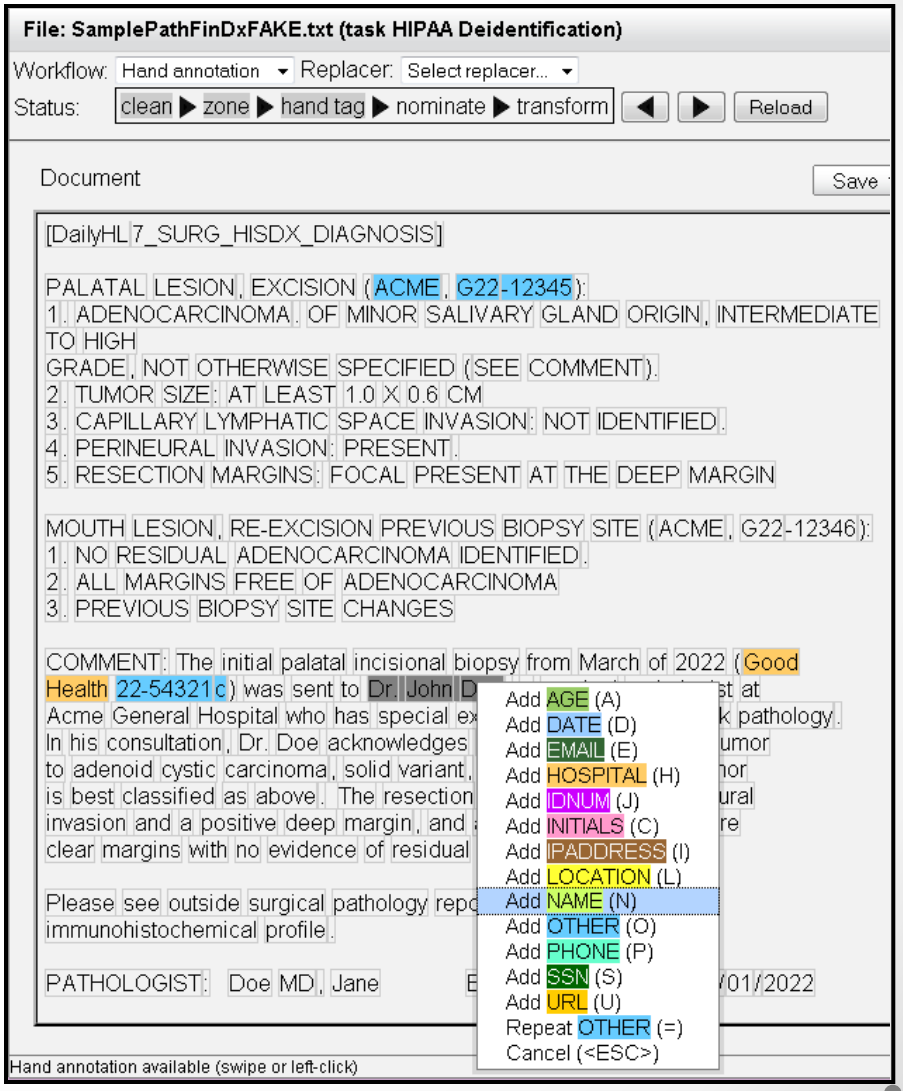
- EHR-based Phenotyping
- Establish new Genotype-Phenotype associations
- Clinical Use:
  - Defining actionability/clinical utility/validity
  - Integration into EHR/Visualization/Clinical Decision support
- Physician and Patient attitudes/Education
- Consent/Regulatory
- ***Privacy/Security/CLIA/CAP***

eMERGE Data Privacy Projects



Natural Language De-identification for Local Sites

- Many eMERGE sites using free text from EMRs for phenoytyping
- Freely-available open source software (MIST) developed as partnership between eMERGE members and MITRE
- Machine learning strategy: Annotate some training data and let the system learn a *model* of identifiers



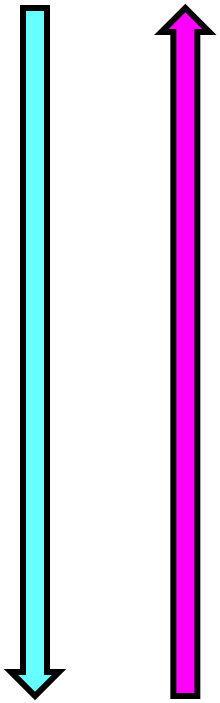
Aberdeen et. al., International Journal of Medical Informatics, 2010

eMERGE-PGRN Partnership

PGRN Pharmacogenomics Research Network

<u>PGx capabilities:</u>

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



<u>EMR-informatics capabilities</u>

- Privacy
- Electronic phenotyping
- Large populations
- Decision support

eMERGE Network
electronic medical records & genomics

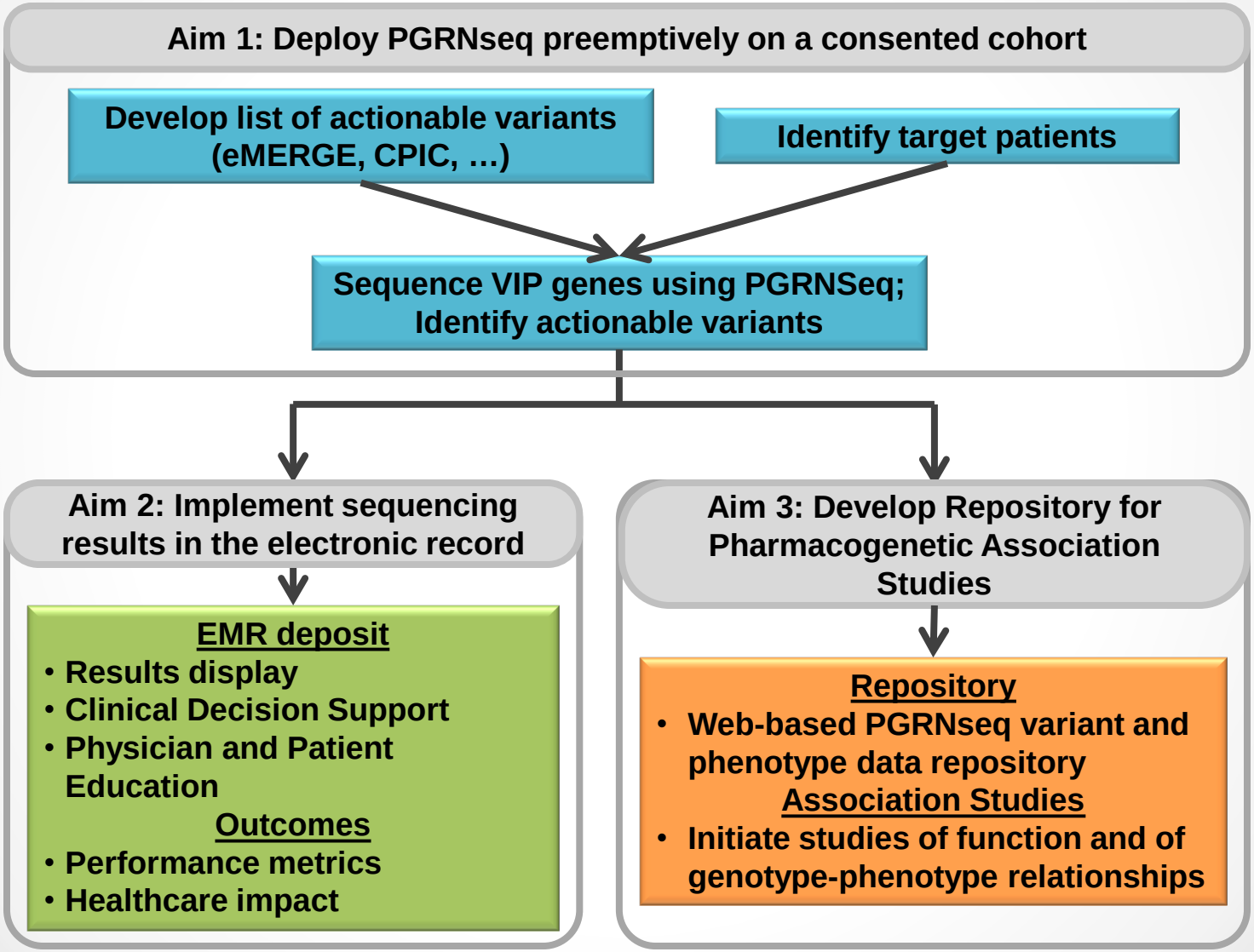
eMERGE PGx Project

- Deploy VIP platform developed by PGRN investigators
- Apply to participants enriched for encountering drugs for which there is a CPIC guideline
- Return appropriate genotype results through appropriate decision support tools
- Archive novel variants for further study

PGx candidate drug-gene pairs

Gene	Drug	Comment
CYP2C19	clopidogrel	Best evidence in patients with coronary stents
CYP2C9 *VKORC1* *CYP4F2*	Warfarin	Algorithms to predict starting dose available. Vary by ancestry
*HLA B*5701*	Abacavir	
*HLA B*1502*	carbamazepine	Higher frequency variant in Asian subjects
SLCO1B1	Simvastatin	Especially at higher dosages or with interacting drugs
TPMT	6-MP, azathioprine	
CYP2D6	Codeine	PM status predicts non-response
CYP2D6	some SSRIs	PM status predicts intolerance of effective dosages.

eMERGE PGx – Overview by Aim



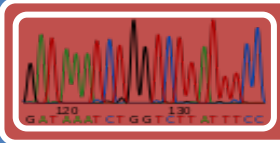
eMERGE PGx - Overview by Project

Specific Aim 1

Recruit / Collect Samples

PGRN-Seq Sequencing

Specific Aim 2

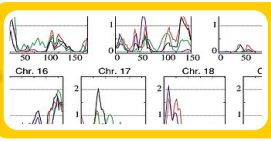


Clinical Variant Validation

Return Results: EHRIntegration and CDS

Patient & Clinician Education

Specific Aim 3



Populate Variant and Phenotype Data Repository (SPHINX)

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

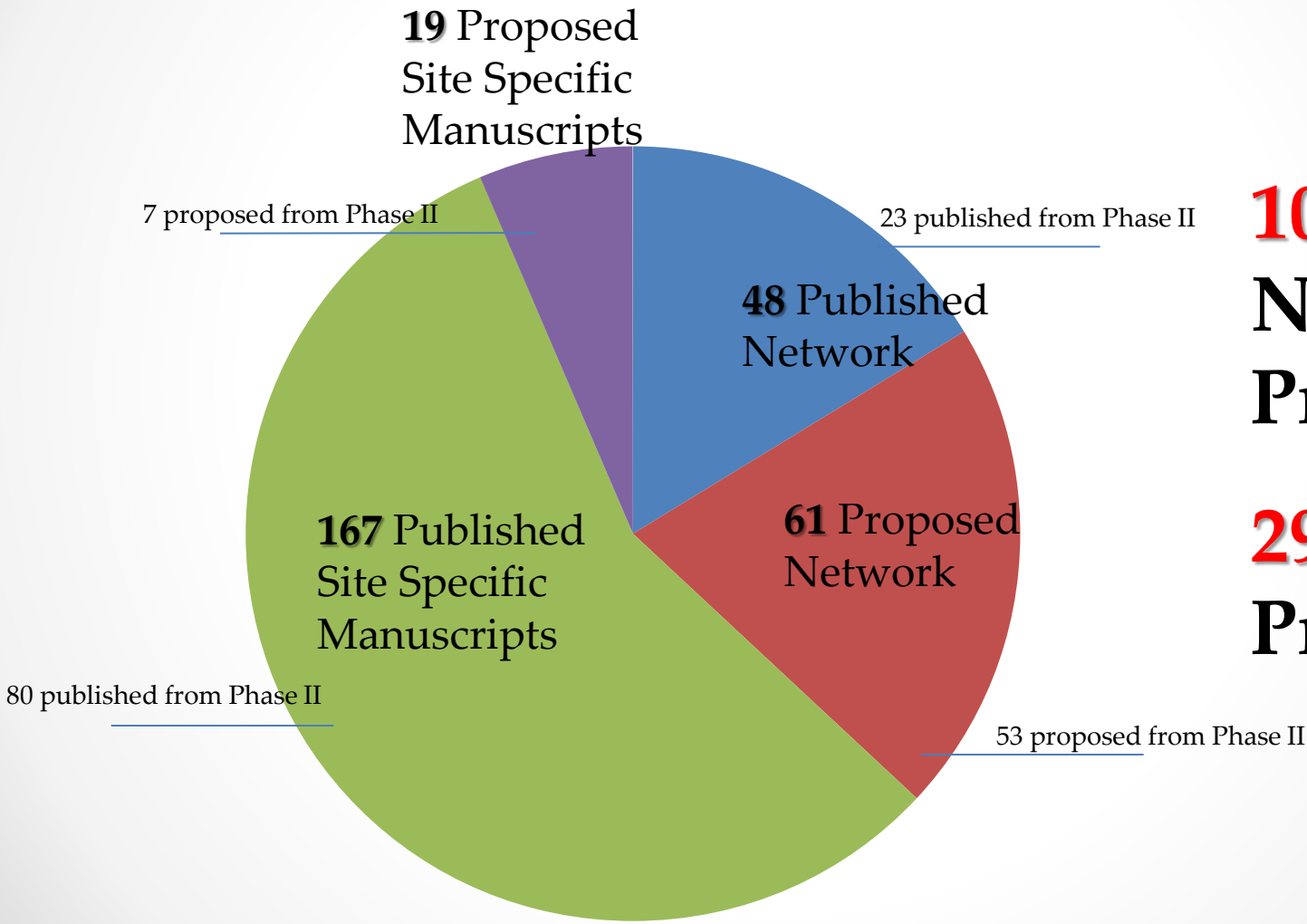
- Statins: Myopathy, Drug Switch
- Clopidogrel: Stent or ACS event? Within 30 days?
- Warfarin: Time to steady state? Time out of range? Bleeding? Thrombosis?
- Thiopurines: Blood counts, (disease outcome), …
- Return of results project: 6 ACMG "actionable" genes

PGRNSeq-Incidental Findings (IFs)

Gene Symbol	Gene Name	Phenotype
CACNA1S	Calcium channel, voltage-dependent, L type, alpha 1S subunit	malignant hyperthermia
KCNH2	Potassium voltage-gated channel, subfamily H (eag-related), member 2	long QT arrhythmia
LDLR	Low density lipoprotein receptor	hyperlipidemia
RYR1	Ryanodine receptor 1 (skeletal)	malignant hyperthermia
RYR2	Ryanodine receptor 2 (cardiac)	catecholaminergic polymorphic ventricular tachycardia
SCN5A	Sodium channel, voltage-gated, type V, alpha subunit	long QT arrhythmia

Measuring Productivity

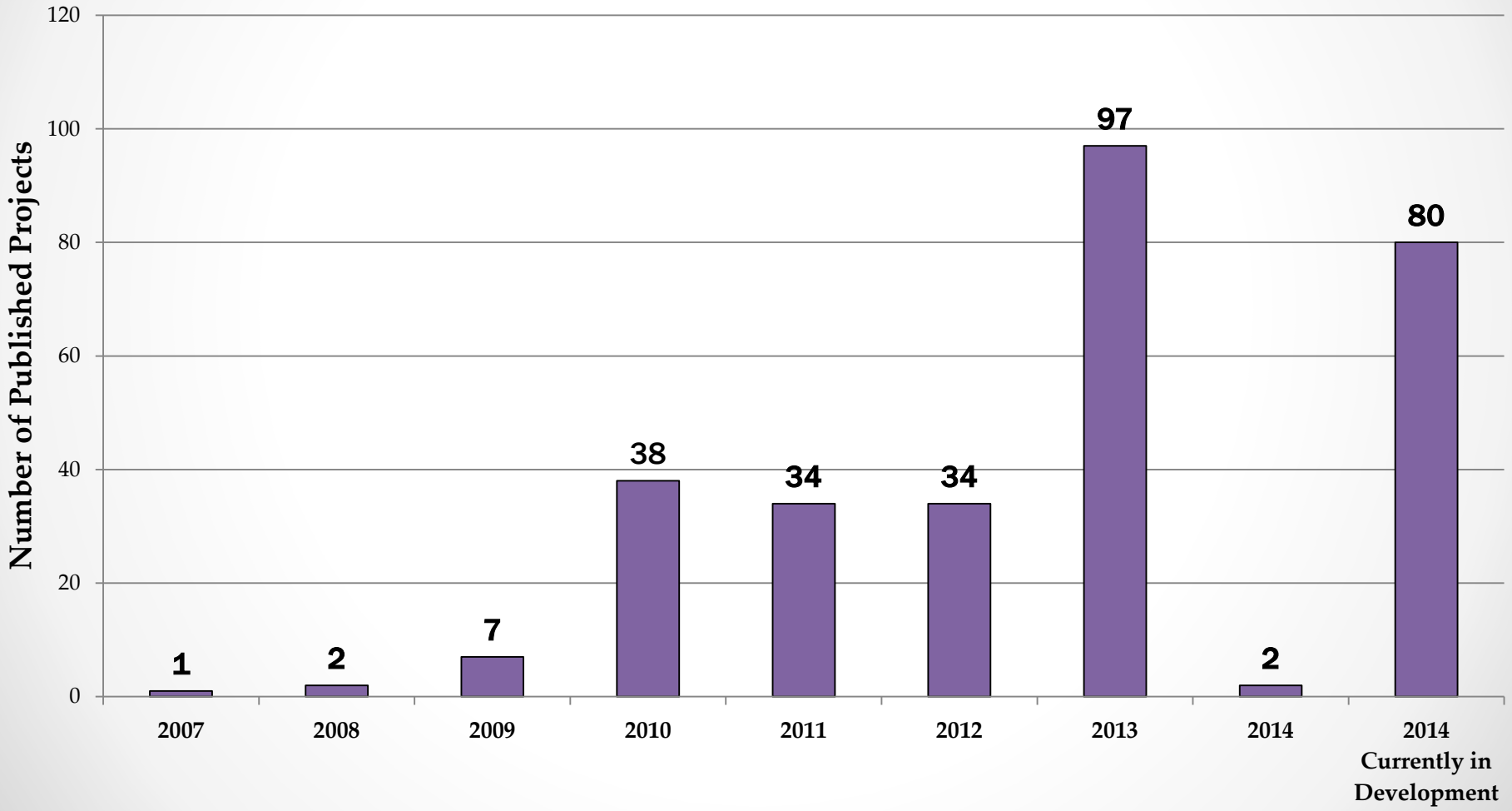
eMERGE Publications through January 2014



109 Total Network Projects

295 Total Projects

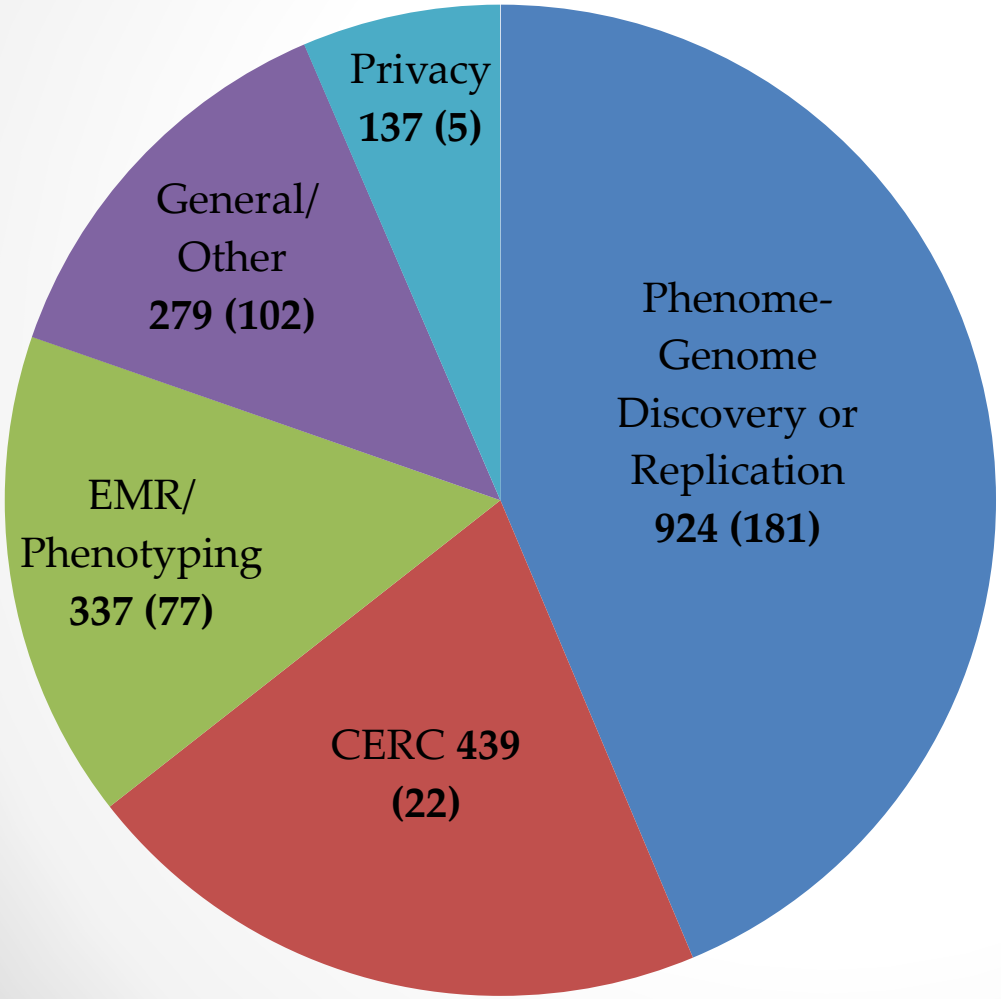
eMERGE Projects Published by Year through January 2014



eMERGE

Citation Analysis

eMERGE Publications: Number of Times Cited through January 2014



Total Number of Times Cited

- 2007-January 2014: **2,116**
  - Phase I Publications Only: **1,729**
  - Phase II Publications Only: **387**

*** Chart totals in parenthesis reflect number of times cited for Phase II publications only**

Genetics inMedicine
Official Journal of the American College of Medical Genetics and Genomics

Volume 15, Issue 10 (October 2013)

INTRODUCTION | Genetics inMedicine

Integration of genomics into the electronic health record: mapping terra incognita

Joseph M. Kannry, MD[1] and Marc S. Williams, MD[2]

eMERGE Tools

	Released Tools Available to Public
Phenotyping	**eleMAP (Mayo/CC)** *113 users representing 70 institutions* **PheKB (CC)** *233 users representing; 24 institutions; 2,936 unique visitors to the website since release*
Genotyping	**PennCNV (CHOP)** *Implements a hidden Markov model that integrates multiple sources of information to infer CNV calls for individual genotyped samples, Widely-used: 630 Citations to date* **NGS data Analysis Pipeline (CHOP)** *Over 1,700 subjects whole exome sequencing with over 100 different rare medical disorders, resolved over 30 rare disorders; sequencing an average of 50 exomes per week at 70x coverage, with an average turn-around time of four to six weeks from sample-processing to variant identification* **Biofilter, BioBin (Penn State/ Marshfield)** *Provides methods for prioritizing and analyzing variants singly or in groups, over 50 downloads since June, 2013* **PLATO, ATHENA (Penn State/ Marshfield)** *Provides platforms for QC and integrating multiple methods of analysis, over 50 downloads since June, 2013* **Synthesis-View PheWas-View, Phenogram (Penn State/ Marshfield)** *Provides visualization tools for genome and phenome-wide data, over 900 unique visitors*
Consent	**eMERGE Model Consent Language (eMERGE Network)** *Freely available on NIH website*
CDS	**ANNOVAR (CHOP)** *Annotates genetic variants detected from genomes that can shortlist SNVs and insertions/deletions, examine and report functional consequence, infer cytogenetic bands, find variants in conserved regions, and identify variants from the 1000 Genomes project and dbSNP, Widely used: 513 citations to date* **MyResearch, integration between Registrar and MyChart (NU)** *502 patients registered as of end of August*
NLP	**cTAKES (Mayo/Boston Children's)** *66 subscribers to User listserv, 75 subscribers to developers listserv, 207 unique hits to download cTAKES in last 30 days*

eMERGE tools

	In Development	Beta Testing	Early (Limited) Release
Phenotyping	**Phenotype Dictionary and File Validation Tool (eMERGE Network)**	**PheWAS R Package (VU)** **Web-based Ophthalmology Data Collection (Marshfield/ESS/PSU)** *Evaluation in progress at one institution*	**Phewas 1.0 (VU)** *3944 downloads from website*
Genotyping			**eRecordCounter (CC)** *Released to all eMERGE participants* **ParseCNV (CHOP)** *Takes CNV calls as input & creates SNP based statistics for CNV occurrence in cases & controls then calls CNVRs based on neighboring SNPs of similar significance; One citation, released Mar, 2013* **Genome-wide genotype quality control pipeline (CC)** *publish manuscript*
Privacy	**DARRT (VU)**	**MIST, with HIPS implementation (GHC/VU)** *Testing at 2 sites with MITRE*	
Consent			**Computer Based Training Consent (Marshfield/ESS/PSU)** *enrolled 70 participants currently active at one site*
CDS	**Clinical Knowledge Manager (NU)** **Ancillary Genomics System (AGS) (NU)**		
NLP	**cTakes Machine Learning Patient Vectors (cMPV) (CCHMC/BCH)**		**Medex (VU)** *Actively used by 6 institutions*
Clinical Integration		**Clinical Utility of Pharmacogenetic Research Opioid CYP2D6 Results: Physician Survey (CCHMC)**	**Research Opioid CYP2D6 Panel Templates (CCHMC)** *Early version shared with PGx workgroup* **Research Pharmacogenetics Results & Incidental Findings Parent Survey adapted from the National Pharmacogenetics Survey(CCHMC/BCH)** *Using for data collection; Shared with eMERGE ROR & CERC groups*

Conclusions

- Biobanks linked to EHR have found new genomic associations with disease and therapeutic outcomes
- eMERGE has demonstrated the efficiency of having genotyped, EHR linked samples for discovery
- EHR Phenotyping can be applied across varied EHRs
- eMERGE developed tools are widely used for both discovery and implementation
- PheWAS approach validated

Conclusions

- eMERGE has pioneered best practices for data sharing, privacy, consent standards and community engagement
- eMERGE PGx and Genomic Medicine pilot projects are at the vanguard of implementation
- eMERGE has enabled a 100K person assessment that will inform best practices for human subjects research