Northwestern Medicine®

Progress Toward Precision Medicine and the Challenges of Intergrating Genomics into Electronic Health Records.

Rex L. Chisholm, PhD
Adam and Richard T. Lind Professor of Medical Genetics
Vice Dean for Scientific Affairs
Feinberg School of Medicine
Associate Vice President for Research
Northwestern University

Critical factors for Defining Genetic Contributions to Disease: A key to Precision Medicine

- Methods to measure genetic variation in an individual human genome

- *Large numbers of well phenotyped human genomes*

- Development of standards of care / best practices

- Methods to deploy genomics based decision support

Northwestern Medicine®

Northwestern's Biobank: NUgene

- Launched in fall 2002

- IRB approved

- Voluntary participation via informed consent

- Collect biological specimens under a broad consent

- One-time questionnaire at enrollment

- Longitudinal medical information captured from EMR

- Secure application and database

- Enables High-throughput phenotyping

- Recontact option for additional research

- Resource facilitating genetic research institution-wide

- Genotype data returned to NUgene

Northwestern Medicine®

NUgene today

- About 11,000 participants are enrolled

  - Male: 42% Female: 58%

  - Median age: 51

  - Age distribution: 18 - 85+

  - Ethnic breakdown similar to census data for 6-county area

  - Over half enrolled through &/or seen at primary care clinics

  - Average participant has 31 distinct diagnoses (ICD9 non V/E codes), & an average of 16 distinct diagnoses assigned at least twice**

  - Average patient followed over 8 years, some patients >20 years

- Overall participation rate is 28%

  - Uptake rate is ~52% if physician mentions the study

- 92% of participants agree to be contacted for future research or additional health information

Northwestern Medicine®

***assigned the same ICD-9 code on 2 or more dates**

Data Sources

- **Questionnaire (self-report):**
  - Completed once, at time of enrollment:
    - Demographic information
    - Environmental exposures
    - Medications
    - Self-reported family and medical history for select conditions
- **Electronic billing record data**
  - Retrospective and prospective diagnosis (ICD9) & procedure (CPT/ICD9CM) codes
- **Electronic medical record data**
  - Retrospective and prospective:
    - Medical history and diagnoses
    - Lab tests and results, including pathology reports
    - Medications and therapies
    - Family and social history
    - Free text physician notes

Northwestern Medicine®

Participant Counts Per Selected Top 3-Digit ICD9s*

ICD9	ICD9 Description	# Pts**	% Pts**
401	Essential hypertension	3310	34%
272	Disorders of lipoid metabolism	3269	34%
530	Diseases of esophagus	2259	23%
427	Cardiac dysrhythmias	1825	19%
451	Phlebitis and thrombophlebitis	1688	17%
278	Overweight, obesity and other hyperalimentation	1566	16%
715	Osteoarthrosis and allied disorders	1411	15%
477	Allergic rhinitis	1408	14%
216	Benign neoplasm of skin	1380	14%
250	Diabetes mellitus	1367	14%
311	Depressive disorder, not elsewhere classified	1261	13%
692	Contact dermatitis and other eczema	1238	13%
244	Acquired hypothyroidism	1102	11%
493	Asthma	1043	11%
366	Cataract	1010	10%

From billing, encounter, problem list, med Hx.

Top Laboratory Tests within Population

Laboratory Tests	# Participants*	% Participants*	# of Tests
Glucose	8,509	88	265,963
Creatinine	8,422	87	124,848
Calcium	8,392	86	258,308
Potassium	8,387	86	265,022
Sodium	8,384	86	262,436
Chloride	8,384	86	261,752
Mean Corpuscular Volume (MCV)	8,364	86	247,740
Red Cell Distribution Width (RDW)	8,364	86	247,733
Mean Corp. Hemoglobin Conc. (MCHC)	8,364	86	247,740
Mean Corpuscular Hemoglobin (MCH)	8,364	86	247,740
Blood Urea Nitrogen	8,366	86	245,748
Bicarbonate	8,332	86	243,499
Hemoglobin	8,320	86	227,208
Hematocrit	8,320	86	227,201
Red Cell Count	8,296	85	224,380

Northwestern Medicine®

How Can This EHR-linked Biobank be used to discover new gene-disease associations, improve diagnoses and measure therapeutic outcomes?

The eMERGE Network
electronic Medical Records & Genomics

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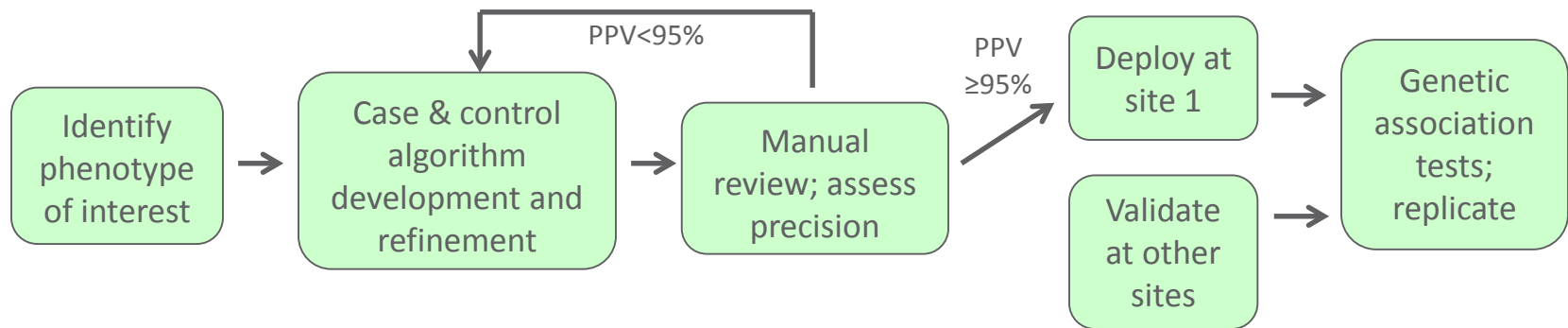
Coordinating Center

An NHGRI funded consortium

eMERGE I Goals

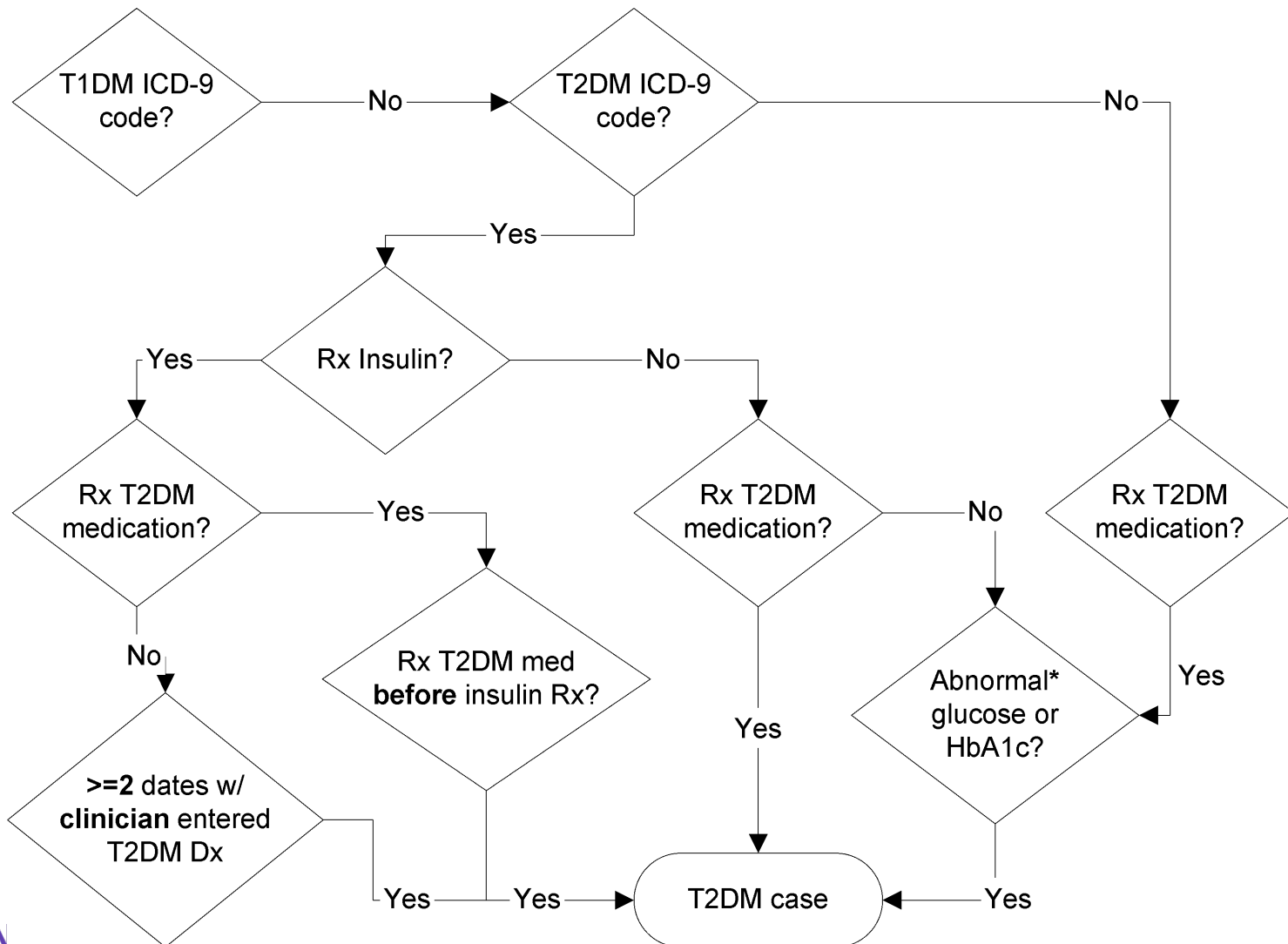Test the ability to leverage EMRs and biobanks for genomic research

- Evaluate validity & utility of EMR phenotypes for Genomics
- Develop & validate electronic phenotyping algorithms
- Conduct association studies of genome-wide data with EMR-derived phenotypes

Northwestern Medicine®

Site	eMERGE Phase I		eMERGE Phase II		eMERGE I & II
	Participants	Genotyped	Participants (Still enrolling)	Genotyped	Genotyped
GHC	2,820	2,789	5,291	739	3,528
Marshfield	20,000	4,210	20,000	777	4,987
Mayo	3,769	3,755	6,916	6,306	10,061
NU	10,500	1,907	12,000	3,030	4,937
VU	70,000	6,055	155,000	3,565	9,620
Geisinger	N/A	N/A	22,000	4,085	4,085
Mt. Sinai	N/A	N/A	25,000	6,290	6,290
CCHMC/BCH	N/A	N/A	11,799	5,799	5,799
CHOP	N/A	N/A	60,000	6,623	6,623
TOTAL	107,089	18,716	347,090	37,214	55,930

Approach to electronic phenotyping

Identify phenotype of interest → Case & control algorithm development and refinement → Manual review; assess precision

PPV<95% (loop back to Case & control algorithm development and refinement)

PPV ≥95% → Deploy at site 1 → Genetic association tests; replicate

Validate at other sites → Genetic association tests; replicate

Denny et al., 2011

Type II Diabetes Case Algorithm

T1DM ICD-9 code? — No → **T2DM ICD-9 code?** — No →

T2DM ICD-9 code? — Yes → **Rx Insulin?**

Rx Insulin? — Yes → **Rx T2DM medication?**

Rx Insulin? — No → **Rx T2DM medication?**

T2DM ICD-9 code? — No → **Rx T2DM medication?**

Rx T2DM medication? — Yes → **Rx T2DM med before insulin Rx?**

Rx T2DM medication? — No → **>=2 dates w/ clinician entered T2DM Dx**

Rx T2DM med before insulin Rx? — Yes → **T2DM case**

>=2 dates w/ clinician entered T2DM Dx — Yes → Yes → **T2DM case**

Rx T2DM medication? — Yes → **T2DM case**

Rx T2DM medication? — No → **Abnormal* glucose or HbA1c?**

Rx T2DM medication? — Yes → **Abnormal* glucose or HbA1c?**

Abnormal* glucose or HbA1c? — Yes → **T2DM case**

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

Genomic Analysis Identifies the same genes as Purpose Built Cohorts

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

TCF7L2
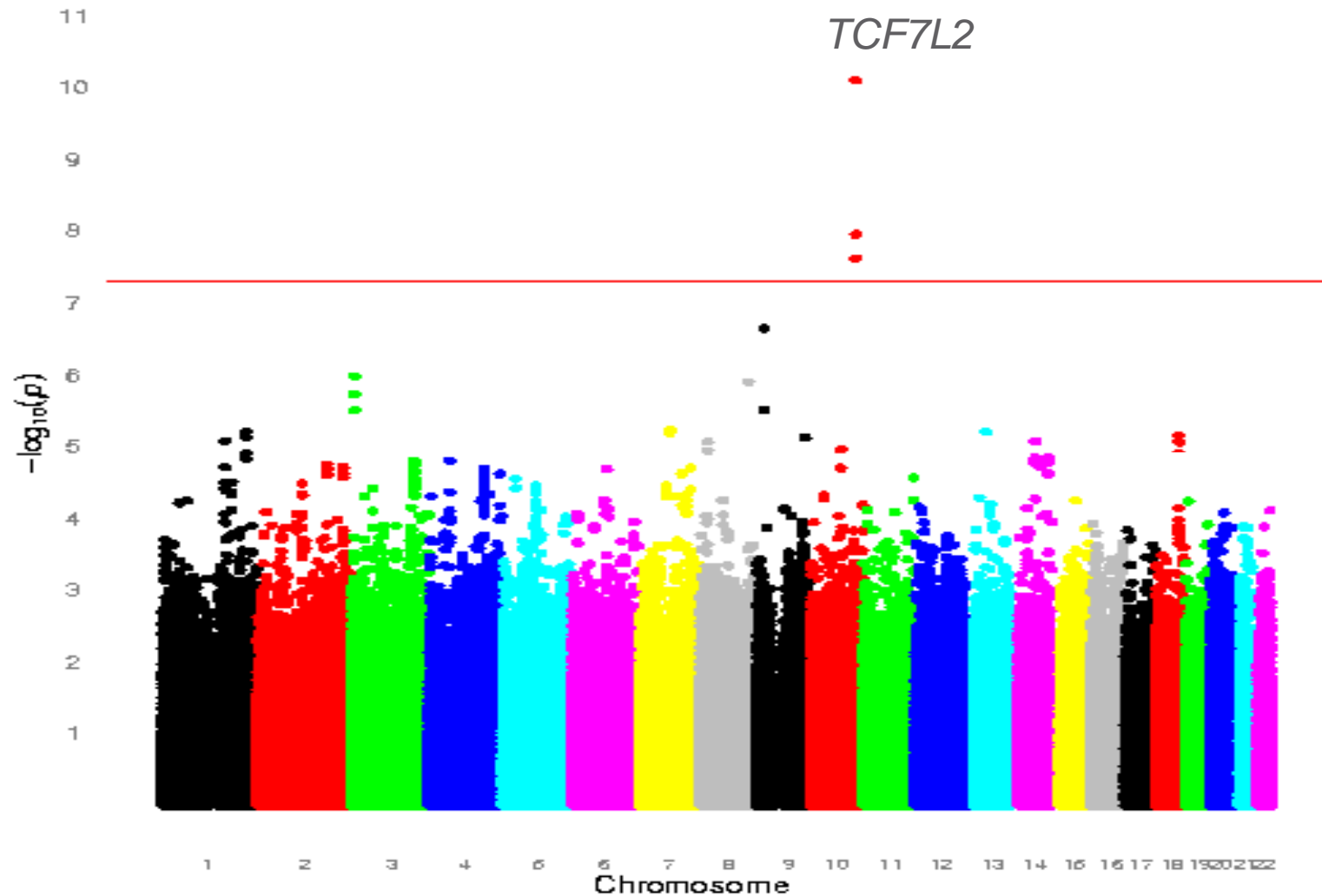
Phase I Phenotypes

	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

Approach to electronic phenotyping

Identify phenotype of interest → Case & control algorithm development and refinement → Manual review; assess precision

PPV<95% (loops back to Case & control algorithm development and refinement)

PPV ≥95% → Deploy at site 1 → Genetic association tests; replicate

Validate at other sites → Genetic association tests; replicate

Denny et al., 2011
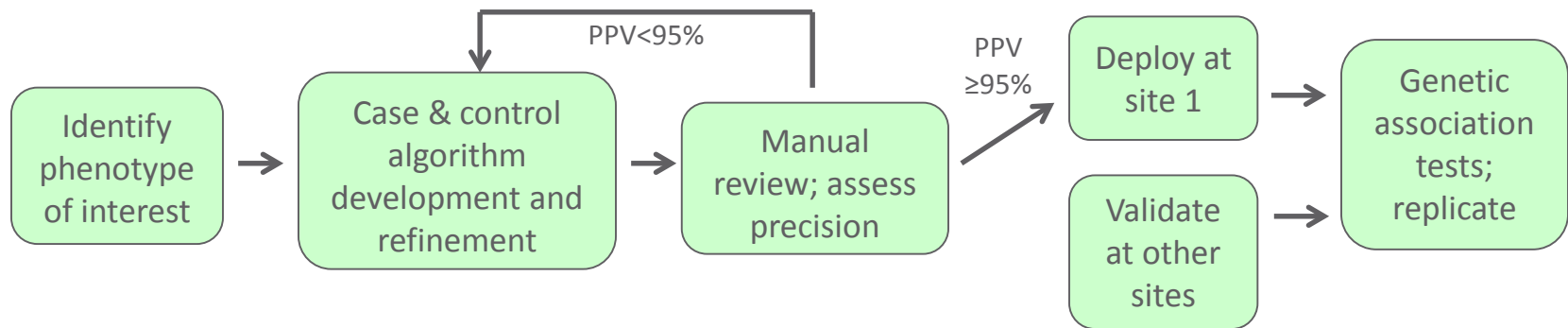
Merged Genotype Dataset

- 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?

Algorithm for Hypothyroidism

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

Northwestern
Medicine®
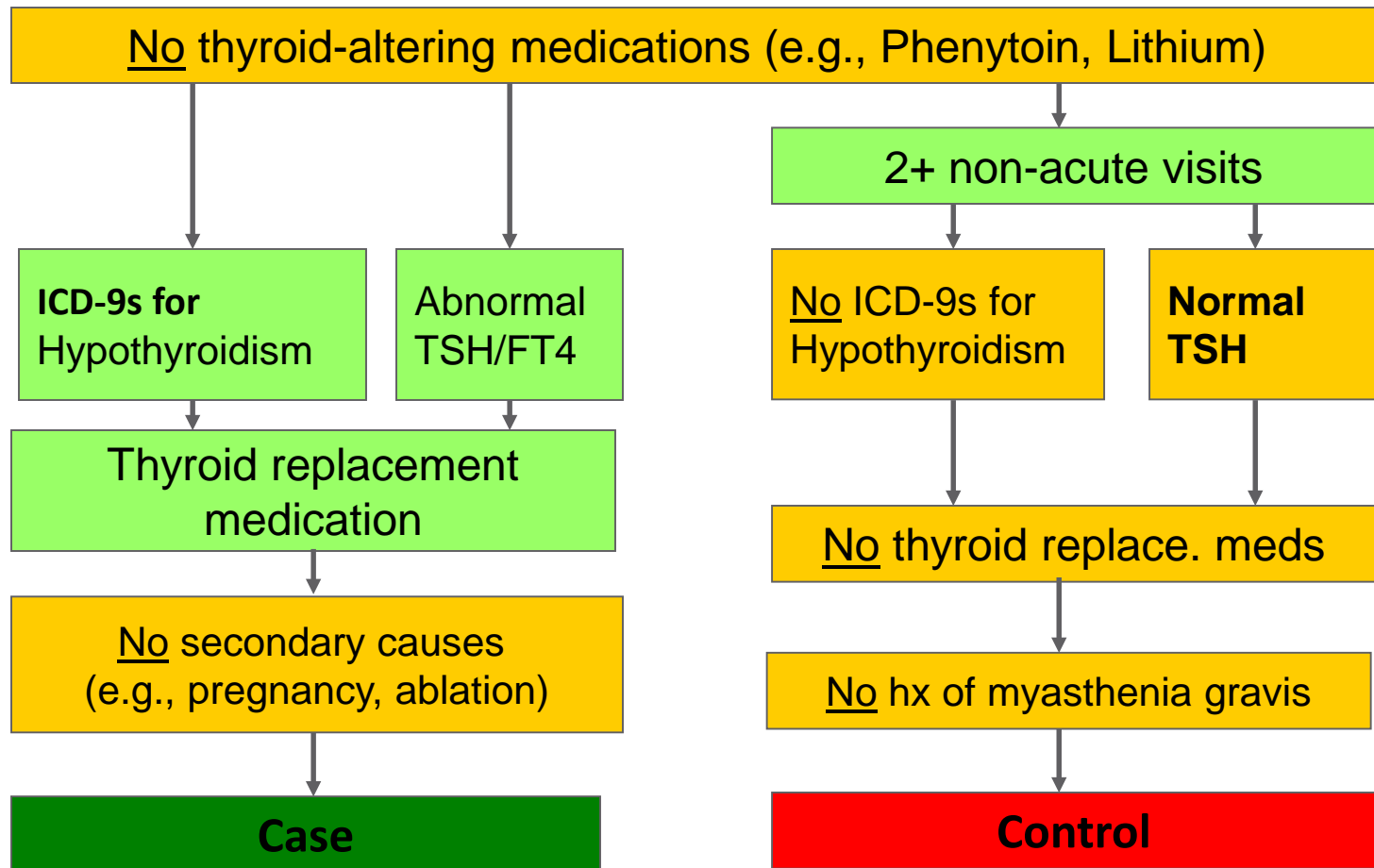
Genomic Analysis
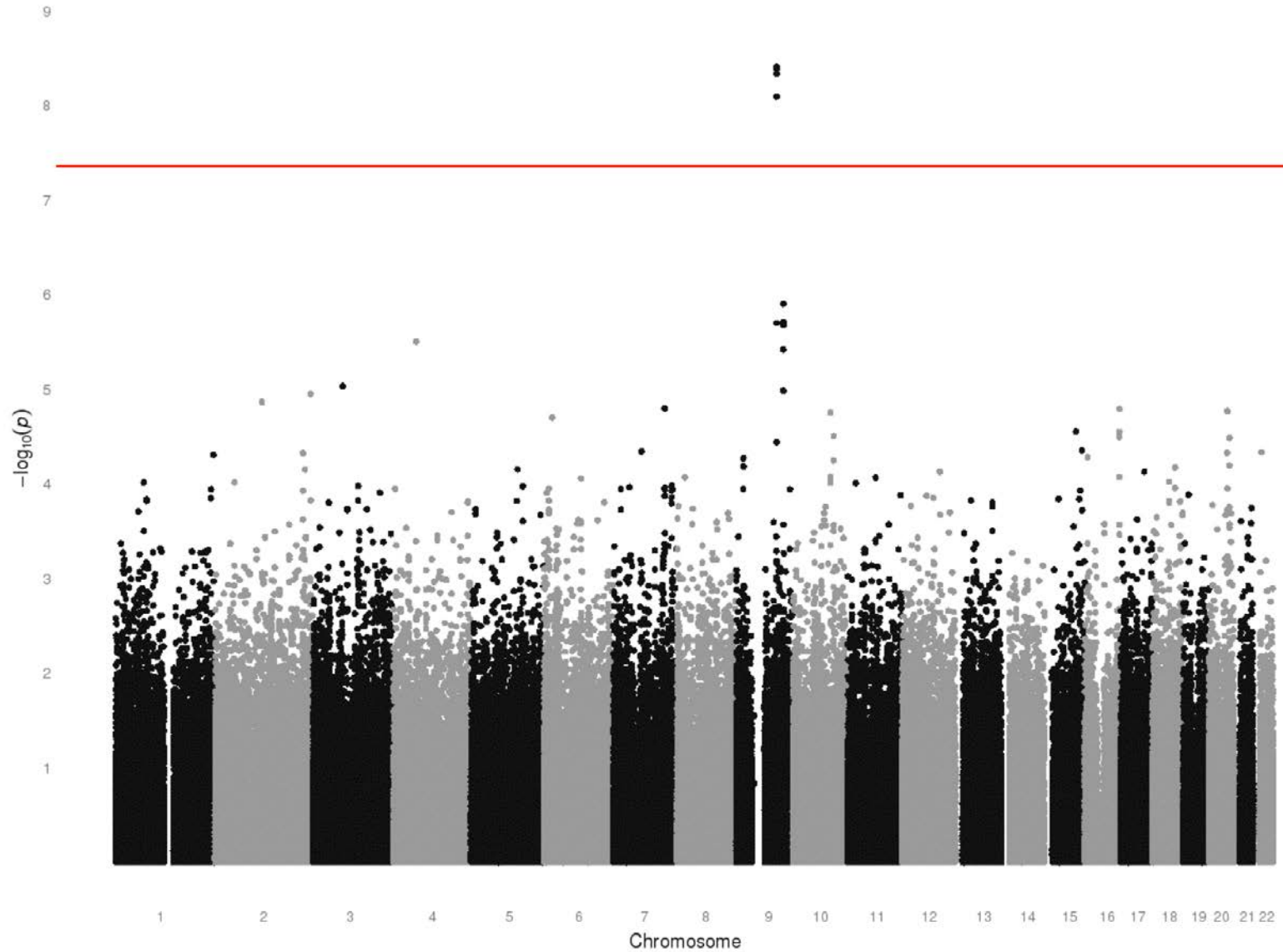
FoxE1 is associated with Hypothyroidism
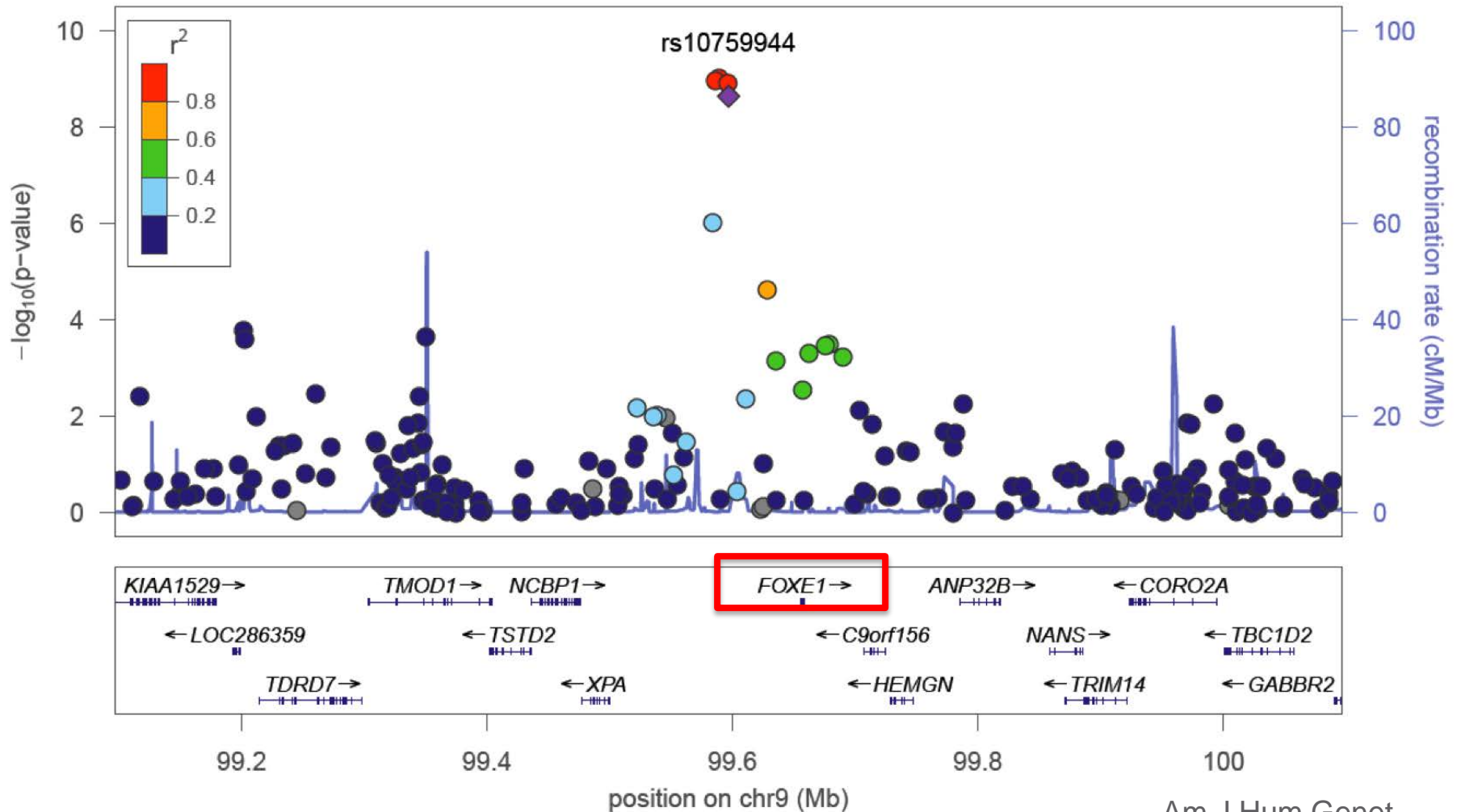
GWAS: Target phenotype
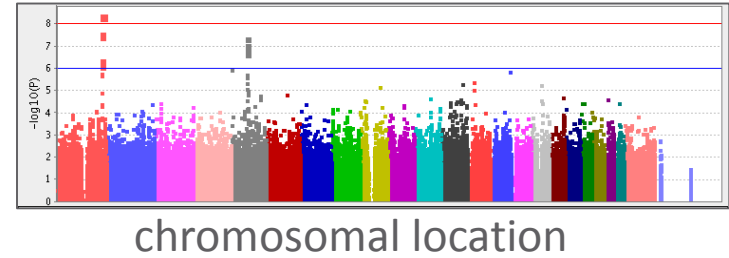
The phenome-wide association study (PheWAS)

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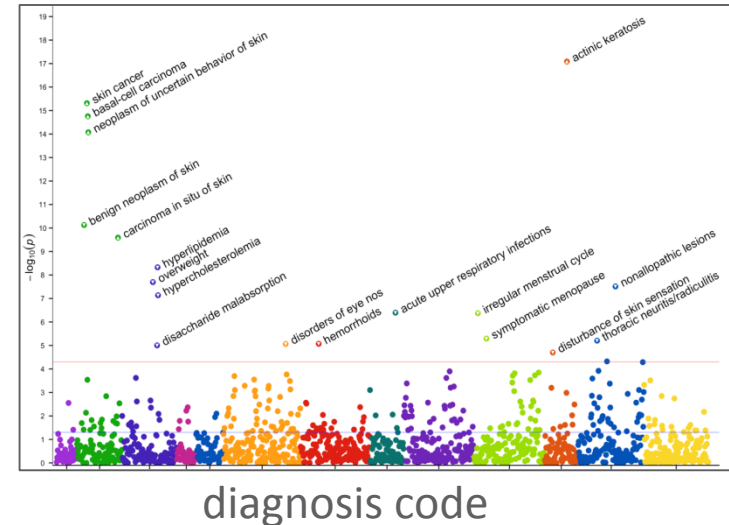
PheWAS requirement: A large cohort of patients with genotype data and many diagnoses

PheWAS for rs965513 (*FOXE1*)

Analysis of 866 phenotypes in 13,617 European Americans
Adjusted for age and sex
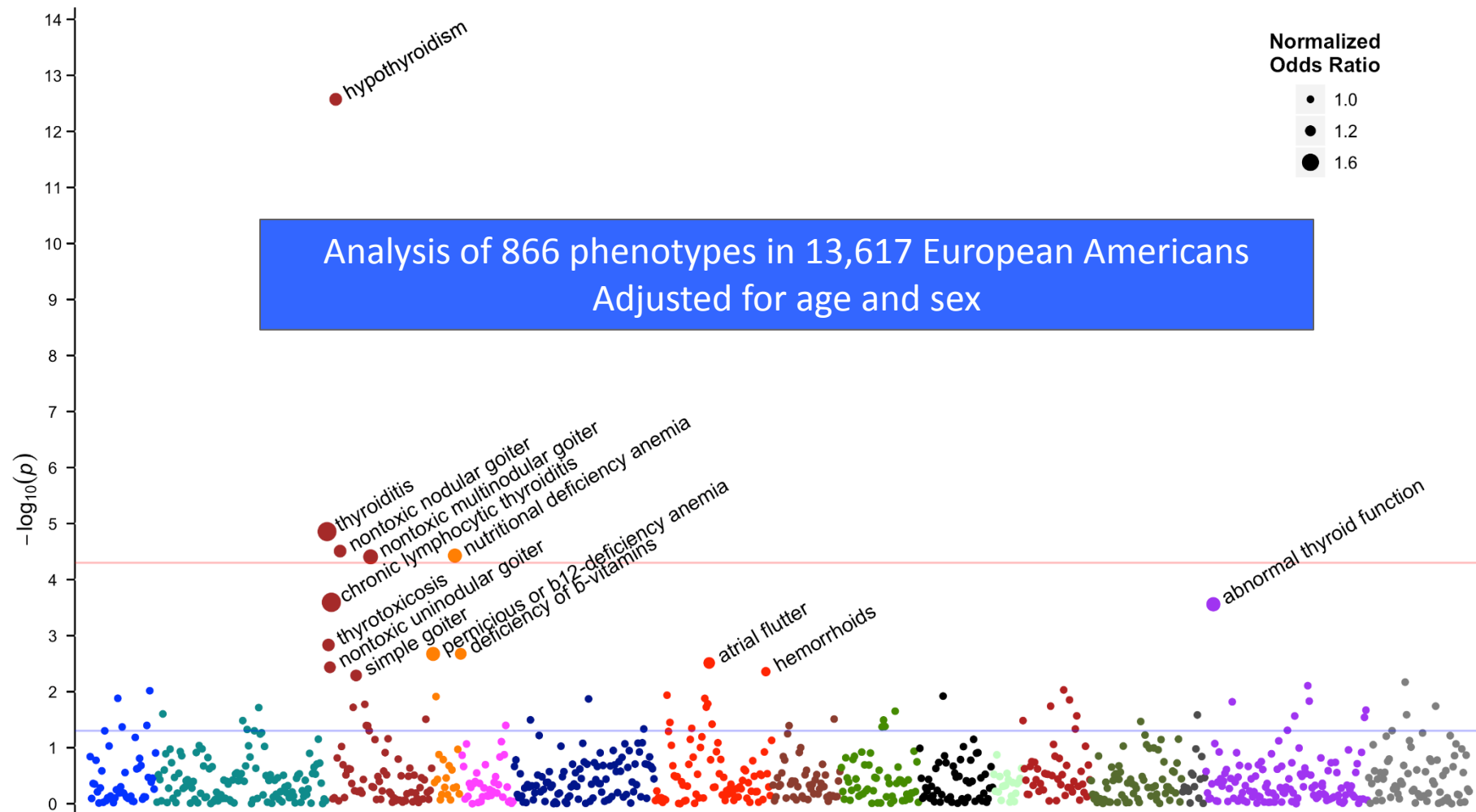
eMERGE Phenotyping: Sharable, High-Throughput

- 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

Northwestern Medicine®

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
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

Northwestern Medicine®

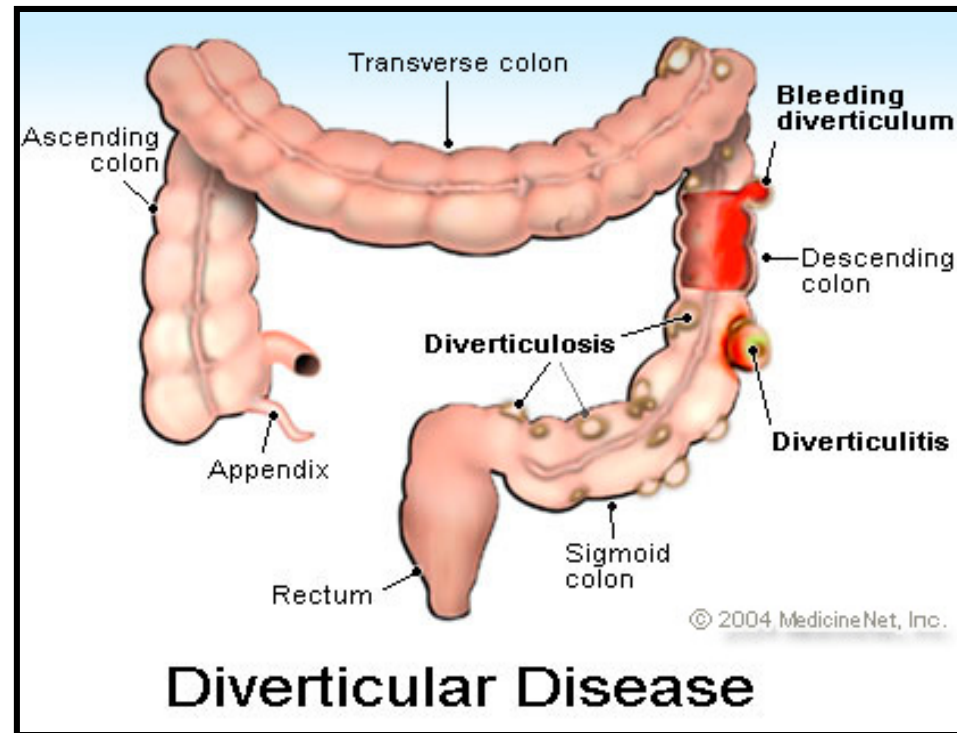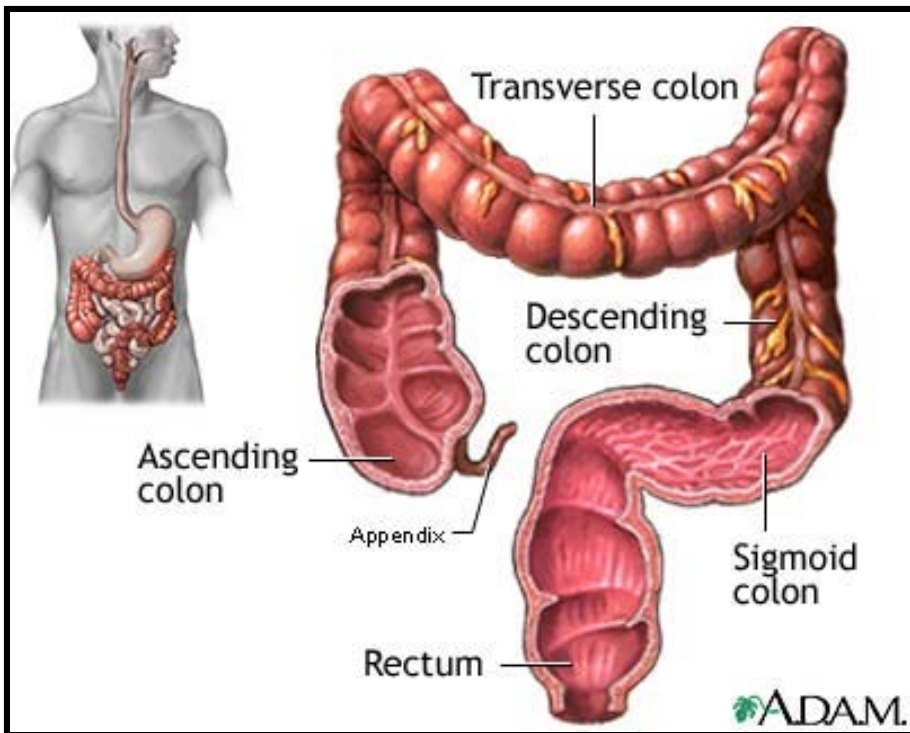PheKB — a knowledgebase for discovering phenotypes from electronic medical records
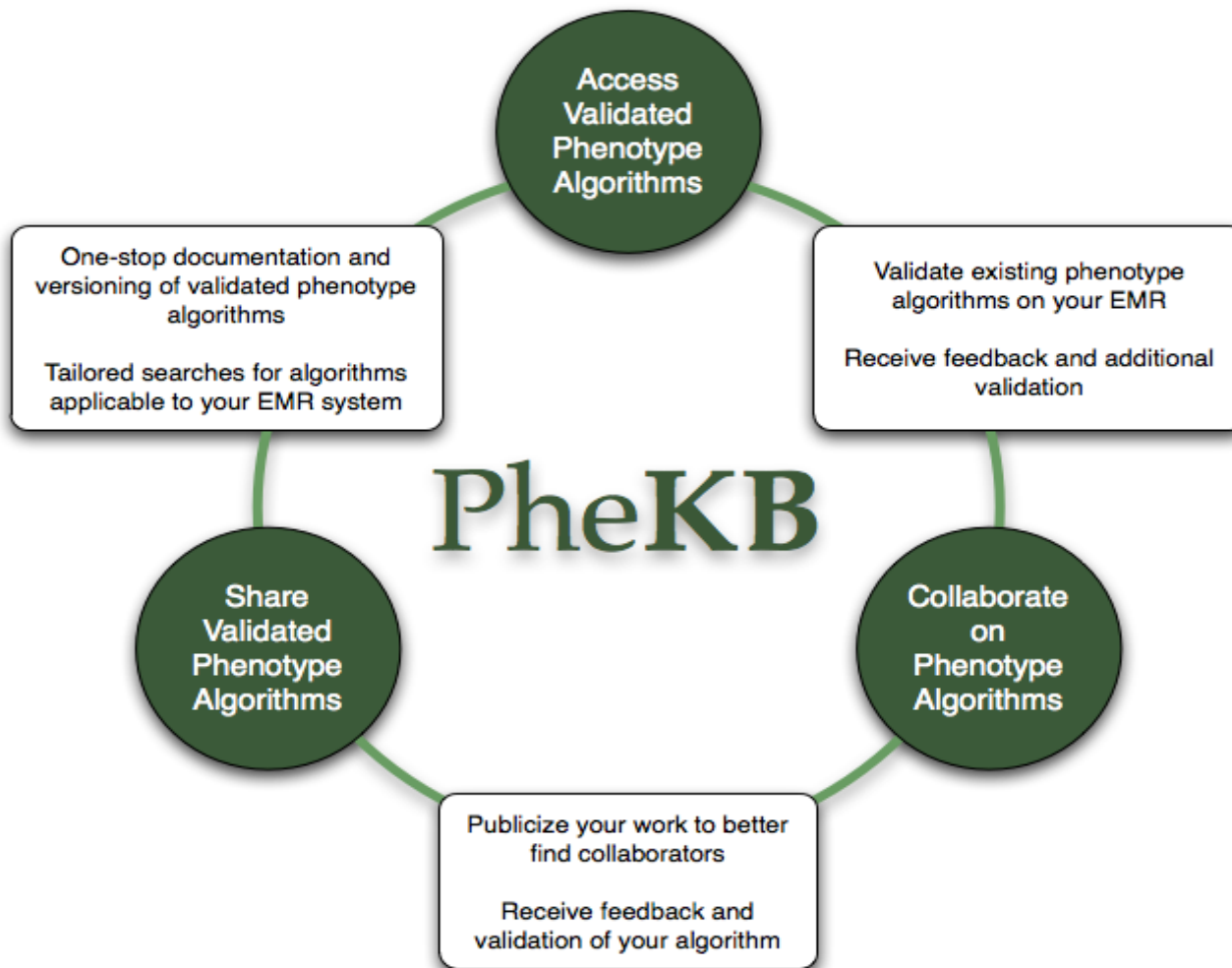
- View existing algorithms

- Enter or create new algorithms

- Collaborate to create or review algorithms

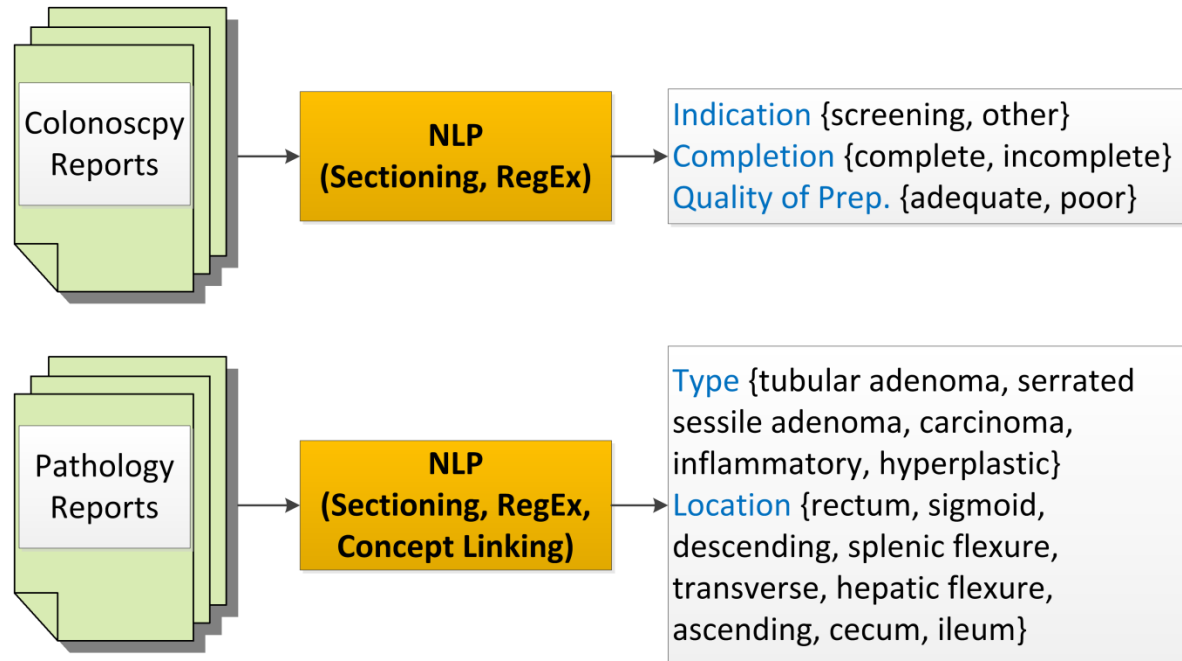- View implementation details for existing algorithms

Colon Polyps

Diverticulosis
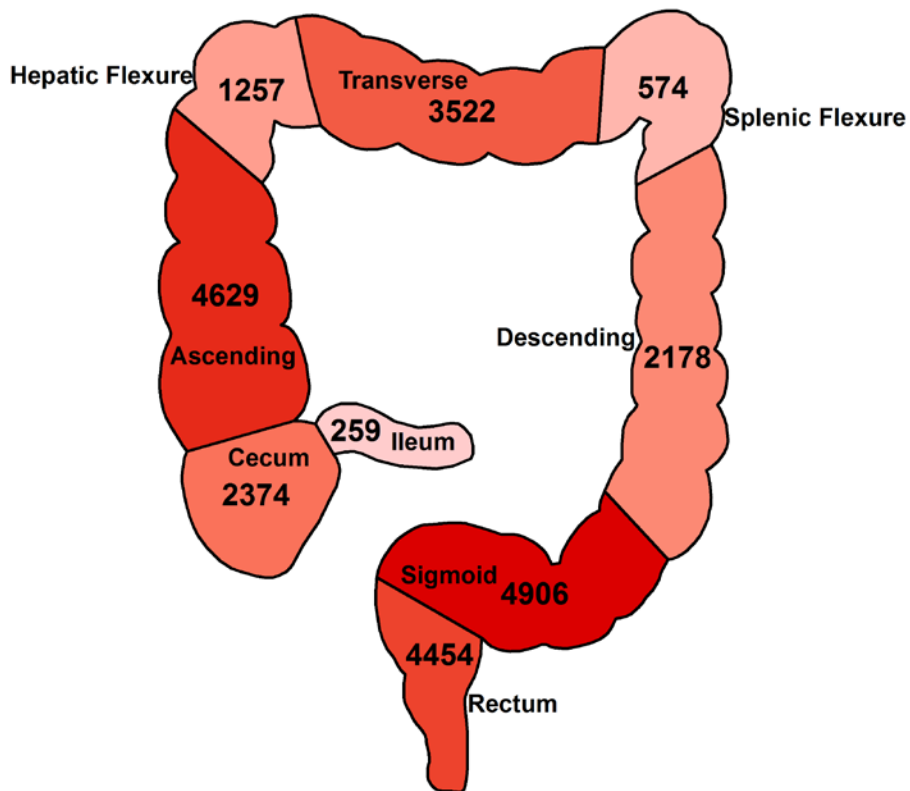
Colon Polyp NLP Algorithm

1. **±Colon Polyps**: checks EHR for colonoscopy with linked path report w. polyp mention

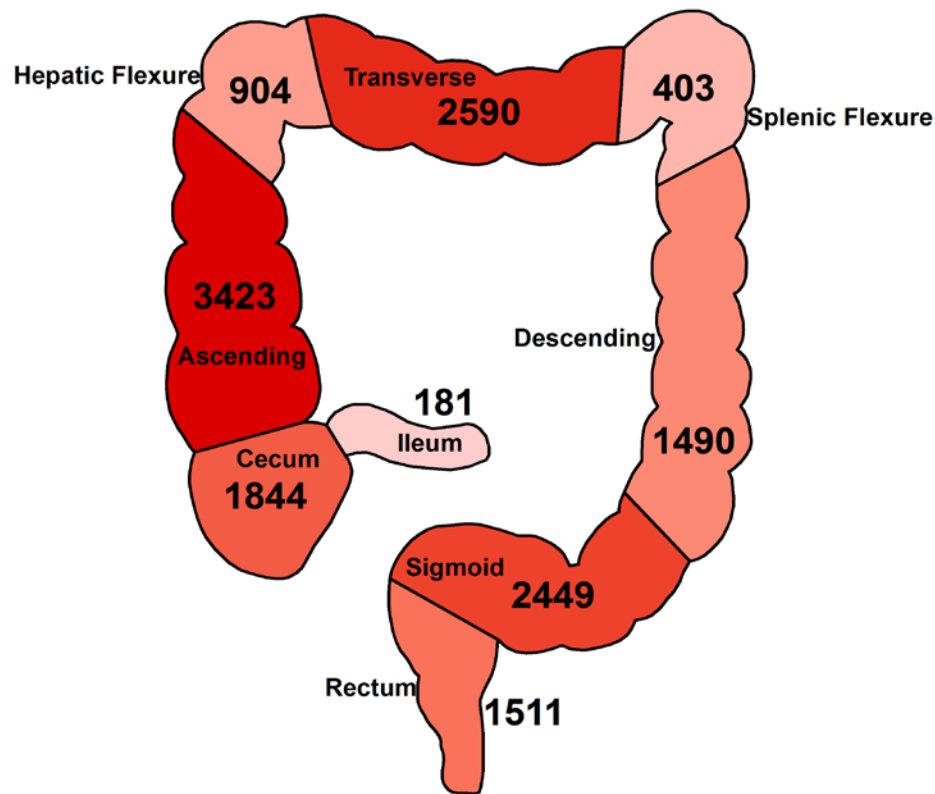2. **Type + Location**: NLP on path reports to extract type and location

Colonoscpy Reports → **NLP (Sectioning, RegEx)** → Indication {screening, other} / Completion {complete, incomplete} / Quality of Prep. {adequate, poor}

Pathology Reports → **NLP (Sectioning, RegEx, Concept Linking)** → Type {tubular adenoma, serrated sessile adenoma, carcinoma, inflammatory, hyperplastic} / Location {rectum, sigmoid, descending, splenic flexure, transverse, hepatic flexure, ascending, cecum, ileum}

	Sensitivity	PPV	Specificity	NPV
±Colon Polyps	98%	94%	94%	98%
Type + Location	96%	98%	NA	NA

Northwestern Medicine®

Colon Polyp Findings By Location

All Colon Polyps: Count per Location

- Hepatic Flexure 1257
- Transverse 3522
- Splenic Flexure 574
- Ascending 4629
- Descending 2178
- Ileum 259
- Cecum 2374
- Sigmoid 4906
- Rectum 4454

Adenoma Count Per Location

- Hepatic Flexure 904
- Transverse 2590
- Splenic Flexure 403
- Ascending 3423
- Descending 1490
- Ileum 181
- Cecum 1844
- Sigmoid 2449
- Rectum 1511

Northwestern Medicine®

Returning Genomic Results to the EHR

- How to store…PDF inadequate
- What to store
  - Only variants of medical significance
  - Actionable
  - Evidence changes over time
- Clinical Decision support
  - Sharable Decision Logic
  - Patient facing supporting material

Northwestern Medicine®

System Architecture
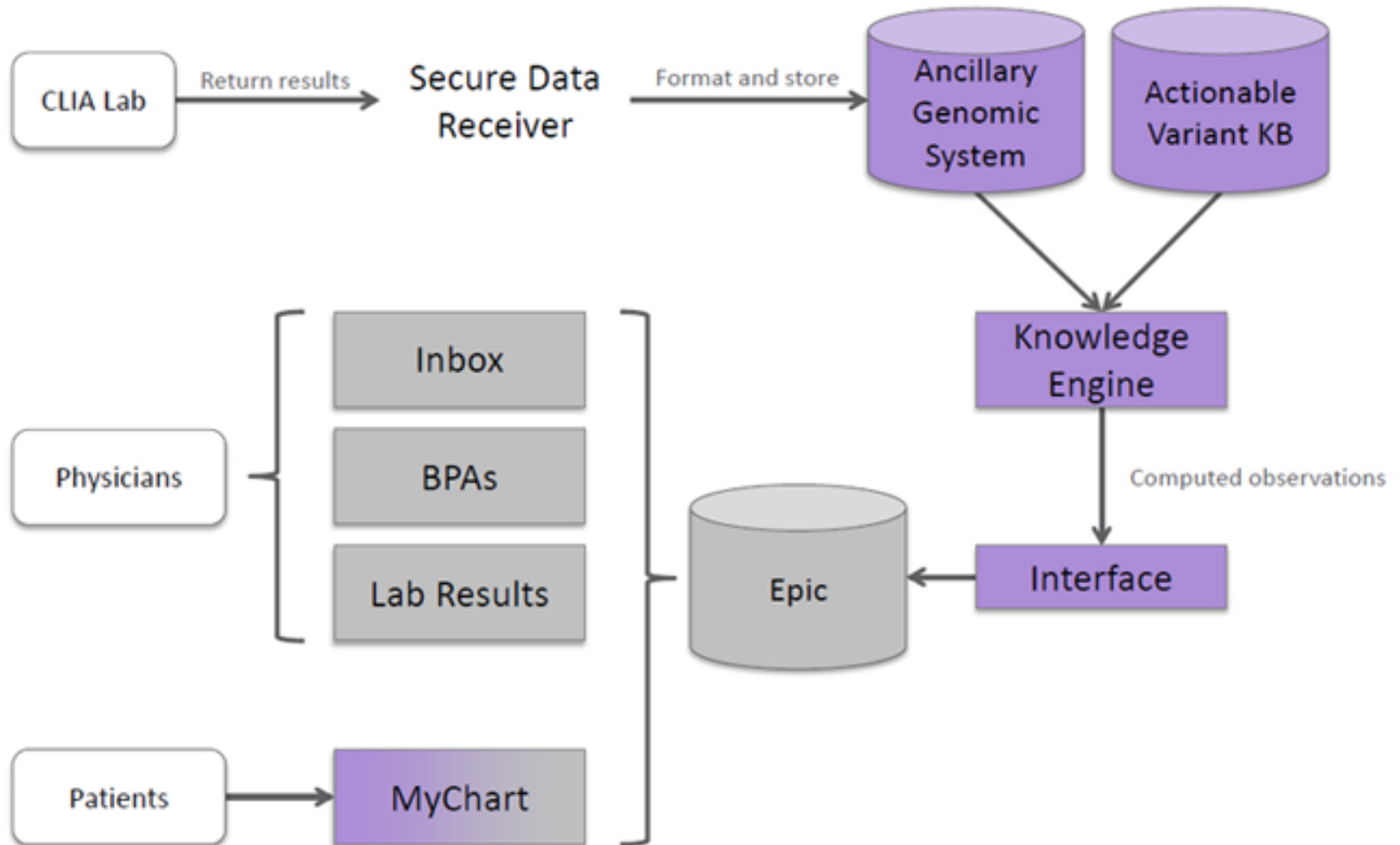
Data from CLIA Labs

MRN	CYP2C9	SLCO1B1	VKORC1	CYP2C19
12345	*1/*2	T/C	A/A	*1/*1
12346	*2/*3	T/T	A/G	*1/*17
...				

- Want to preserve the original data
- Want to let physicians make sense of it
- Want to let patients make sense of it

Ancillary Genomic System

- Entity-Attribute-Value (EAV) schema

ID	PatientID	ParentID	AttrID	AttrName	Value1
1	H0150629	NULL	5	CYP2C9 Diplotype	*1/*1
2	H0150629	NULL	13	CYP2C19 Diplotype	*1/*2
3	H0150629	NULL	9	VKORC1 Genotype	A/A
4	H0150629	NULL	14	rs12248560	C
5	H0150629	NULL	15	rs28399504	AG
6	H0150629	NULL	16	rs41291556	T
7	H0150629	1	18	Value Observation	Inferred
8	H0150629	2	18	Value Observation	Direct

- Create computed observations

  Clopidogrel Metabolism: Poor Metabolizer (Predicted)

  Simvastatin Metabolism: Normal Metabolizer (Predicted)

| Date Created: | | | | MRN: | |

Date Created: [   ] [📅] [   ] [📅]   MRN: [   ]

Patient Last Name: [   ]   Patient First Name: [   ]

DOB: [   ] [📅]

[Filter Results] [Clear Filter]

| Need Review | Approved | Released | All | Result Files | **3 results displayed** |

Creation Date ⬍	Patient Name ⬍	DOB ⬍	MRN ⬍	Test Name ⬍	Results ⬍	
12/31/2013 10:44:20 AM	Zztest, Hamish	12/26/1961	10000084	Simvastatin Metabolism	Normal Activity (Predicted)	[Details]
12/31/2013 10:44:21 AM	Zztest, Hamish	12/26/1961	10000084	Clopidogrel Metabolism	Ultrarapid Metabolizer (Predicted)	[Details]
1/26/2014 10:01:56 PM	Zztest, Hamish	12/26/1961	10000084	Warfarin Dosing		[Details]

AGS Interpretation - Report Preview

Warfarin Dosing   [More Details] [Edit]

Value: Patient: Zztest, Hamish
Created: 01/26/2014 22:01:56   MRN: 10000084
                                Birth Date: 12/26/1961

Report:
RESULT
CYP2C9 *1/*2
VKORC1 (rs9923231) G/G

Use specific information below to determine the appropriate starting dose at http://www.warfarindosing.org.

VKORC1-1639/3673: GG (warfarin insensitive)
CYP2C9*2: CT (heterozygous)
CYP2C9*3: AA (wildtype)
CYP2C9*5: CC (wildtype)
CYP2C9*6: AA (wildtype)

INTERPRETATION
Patient carries one active and one reduced activity CYP2C9 allele and, therefore, is expected to be able to metabolize medications via CYP2C9 less effectively. Intermediate metabolizers may require non-conventional doses of medications whose major metabolic pathway is CYP2C9 or use of another drug that is not processed by CYP2C9.

Clinical Decision Support

🚩 **Patient on clopidogrel, but genetic results indicate patient may be a <u>poor metabolizer</u>ⓘ   Medication may be ineffective – consider alternative.**

Acknowledge Reason: [   ] 🔍 ▤

Discussed results with patient

☐ Open SmartSet: Patient education (After Visit Summary)

↰ Click to review medications
↰ View clinical references related to results
↰ View patient materials related to results

Refresh

eMERGE PGx
Pharmacogenomics

MDConsult
Drug information

MyResults.org
Patient education

MyResults.org

Home — Results — Resources — FAQs — About Us
Start here!   Understand your results   Glossary, videos, and links   Frequently asked questions   Independent and non-p

Clopidogrel (Plavix)
Simvastatin
Tegretol (Carbamazepine)
Warfarin (Coumaden)

Clopidogrel (Plavix)

Overview | The Test | Common Questions | The Science | Media and Recom

- **What is Clopidogrel?**
- Clopidogrel is also called Plavix. It is a drug used by doctors to treat or prevent strokes and
preventing the blood from clotting so that it flows easier through the body. Clopidogrel is som

- **What is being tested?**
- People react differently to medicine and some of those different reactions can be related to
in their genes might not respond to particular medications as well as other people. The gene
called, *CYP2C19*. This test will look for some of the genetic differences in the CYP2C19 gen
clopidogrel.

- **How will this affect my health care?**
- If testing shows that you might be less responsive to clopidogrel, you may be prescribed a

Genetic Test Results

About These Results

Many things can explain why a person has a medical condition, or why different people respond to the same medication in different ways. Genetic testing, which looks for changes (also called polymorphisms, or mutations) in your DNA, can help your doctor determine if you are at risk for getting a condition, or which medication you might need.

It's important to know that being at risk for a condition doesn't mean you will necessarily get it. The results of these tests should be used with other pieces of evidence collected by your doctor to make any type of medical decision.

Hemochromatosis 🛈

Results

C282Y Homozygote
Tested on 3/15/2012

What does this mean?

You **may** be at risk for developing a condition known as hemochromatosis. Your doctor can perform additional tests to see if you currently have this condition

---

Clopidogrel (Plavix) Metabolism 🛈

Results

CYP2C19*2 Homozygote
Tested on 5/17/2011

What does this mean?

Your body **may not** get any benefit from the drug clopidogrel (also known as Plavix). If you are currently on this drug, you should contact your doctor to discuss if an alternative drug may be right for you.
If you are not currently taking Plavix, it is important to know about this in case your doctor needs to prescribe it in the future.

Send a message to your doctor's office

eMERGE OnLine Resources

eMERGE is a national network organized and funded by the National Human Genome Research Institute (NHGRI) that combines DNA biorepositories with electronic medical record (EMR) systems for large scale, high-throughput genetic research in support of implementing genomic medicine.

```
CAGACAGTAATC
TAAATTCGCCGT
GAAATGATCATC
```

POPULAR TOOLS (CLICK ON A BUTTON BELOW)

PheKB
A **knowledgebase** for discovering phenotypes from electronic medical records

MyResults.org
An **informational tool** for educating patients about genetic test results

SPHINX
A **data exploration tool** for genetics-related drug response hypothesis generation

Infobutton Project template

*e*merge **Model Consent Language**

PheWAS catalog

Summary

- Biobanks and EHRs are increasingly playing a critical role in identifying associations between genetic variation, disease risk, drug efficacy and clinical outcomes

- Longitudinal mining of electronic medical records can be used to provide the most up to date phenotype associated with human biospecimens

- Research use can be an important driver of EHR quality

- Networks of EHR-linked biobanks that share samples and data have the potential to increase statistical power to detect genetic associations, population diversity in these studies, and overall research efficiency

- Methods to store genomic variation in EHR will enable personalized medicine

Northwestern Medicine®

How Can EHR-linked Biobanks be used to discover new gene-disease associations and Improve Disease Models?

How well does this work?

- Pretty well, but could be better with more standardization

What are the issues with this approach?

- Heavy reliance on NLP
- Poor consistency between different clinical sites, even for coded data

NM **Northwestern Medicine®**

How Can EHR-linked Biobanks be used to discover new gene-disease associations and improve Disease Models?

- *What can be done to improve the quality of phenotypes?*
    - Broader adoption of standards
    - Include research uses in meaningful use criteria

- *How can this inform disease models?*
    - Phenotype algorithms should also help inform disease models—integrate lessons learned from the "interative" phenotyping process.

Acknowledgements

- NUgene Governance Committee
  & Community Advisory Committee

- NUgene team:
  - Rex Chisholm, PhD (PI)
  - Warren Kibbe, PhD (Co-founder)
  - William Lowe, MD (Medical Dir.)
  - Wendy Wolf, PhD
  - Maureen Smith, MS, CGC
  - Jennifer A. Pacheco
  - Tony Miqueli
  - Sharon Aufox, MS, CGC
  - Oana Popescu
  - Nicole Sheehan
  - Noah Goss
  - Maribeth Miceli

- More information about NUgene:
  http://www.nugene.org

- NU eMERGE team:
  - Rex Chisholm
  - Phil Greenland
  - Abel Kho
  - Bill Lowe
  - Wendy Wolf
  - Maureen Smith
  - Geoff Hayes
  - Loren Armstrong
  - Doug Scheftner
  - Laura Rasmussen-Torvik
  - Pedro Avila
  - Amy Lemke
  - May Law
  - Jen A. Pacheco
  - Joel Humowiecki
  - Will Thompson
  - Arun Muthalagu
  - Krishna Dornadula

Northwestern Medicine®