VANDERBILT UNIVERSITY
MEDICAL CENTER

Balancing discovery and implementation in eMERGE

Dan M. Roden, MD
Vanderbilt University School of Medicine

Discovery

Studying cohorts

- in purpose-generated research datasets
- in the EMR

Implementation

Using a highly interactive electronic medical record

- to provide real-time clinical advice
- to track outcomes

VANDERBILT UNIVERSITY MEDICAL CENTER

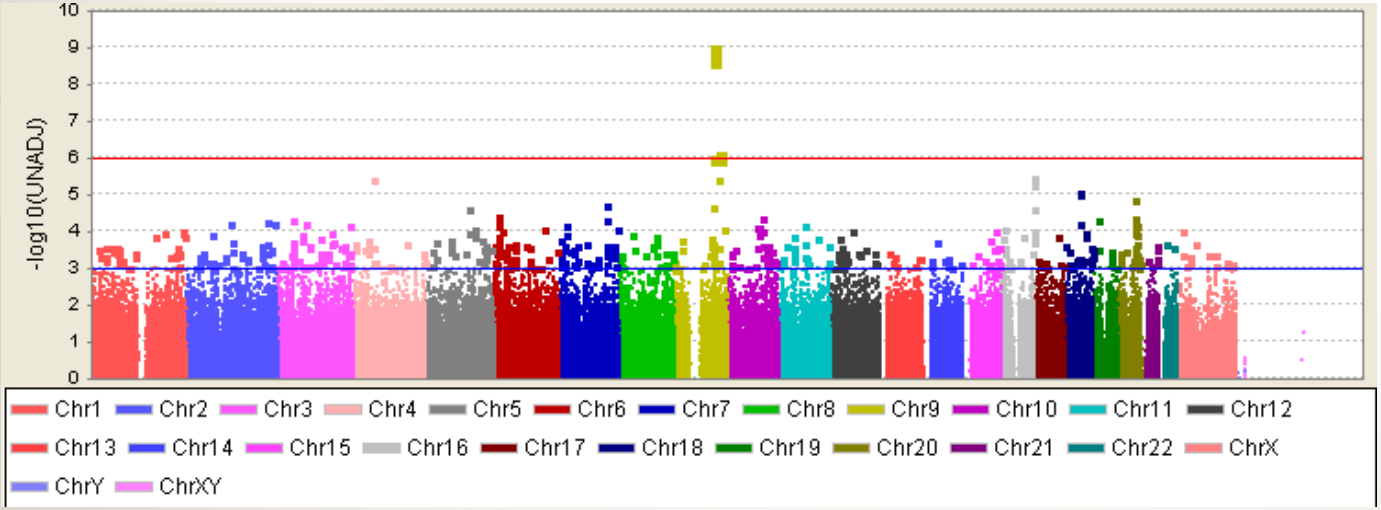
Discovery science in eMERGE

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

Site	Primary Phenotype	Total Genotyped Subjects	Primary Hypothyroidism			
			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.

Algorithms can be deployed across multiple EMRs

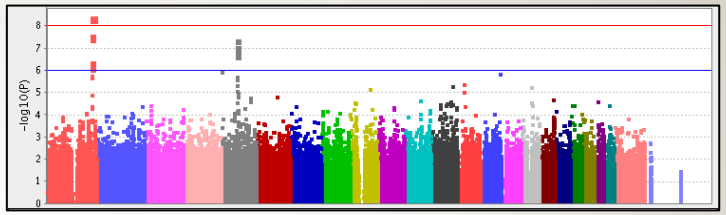


Analyses can be performed using extant data

VANDERBILT UNIVERSITY MEDICAL CENTER

Denny et al., 2011

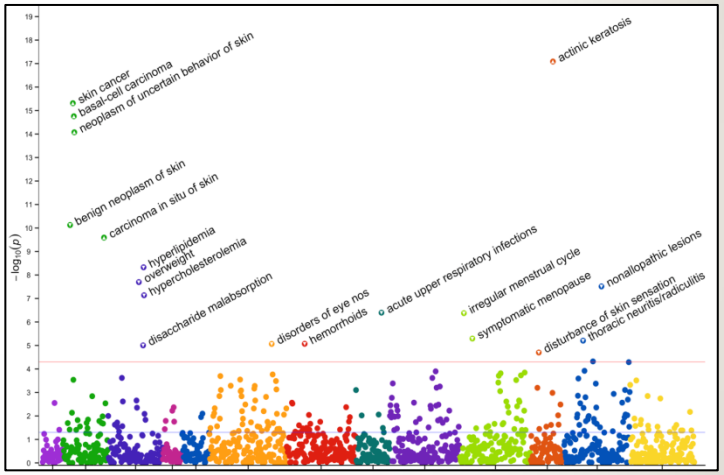
GWAS: Target phenotype → association P value



chromosomal location

The phenome-wide association study

PheWAS: Target genotype → association P value

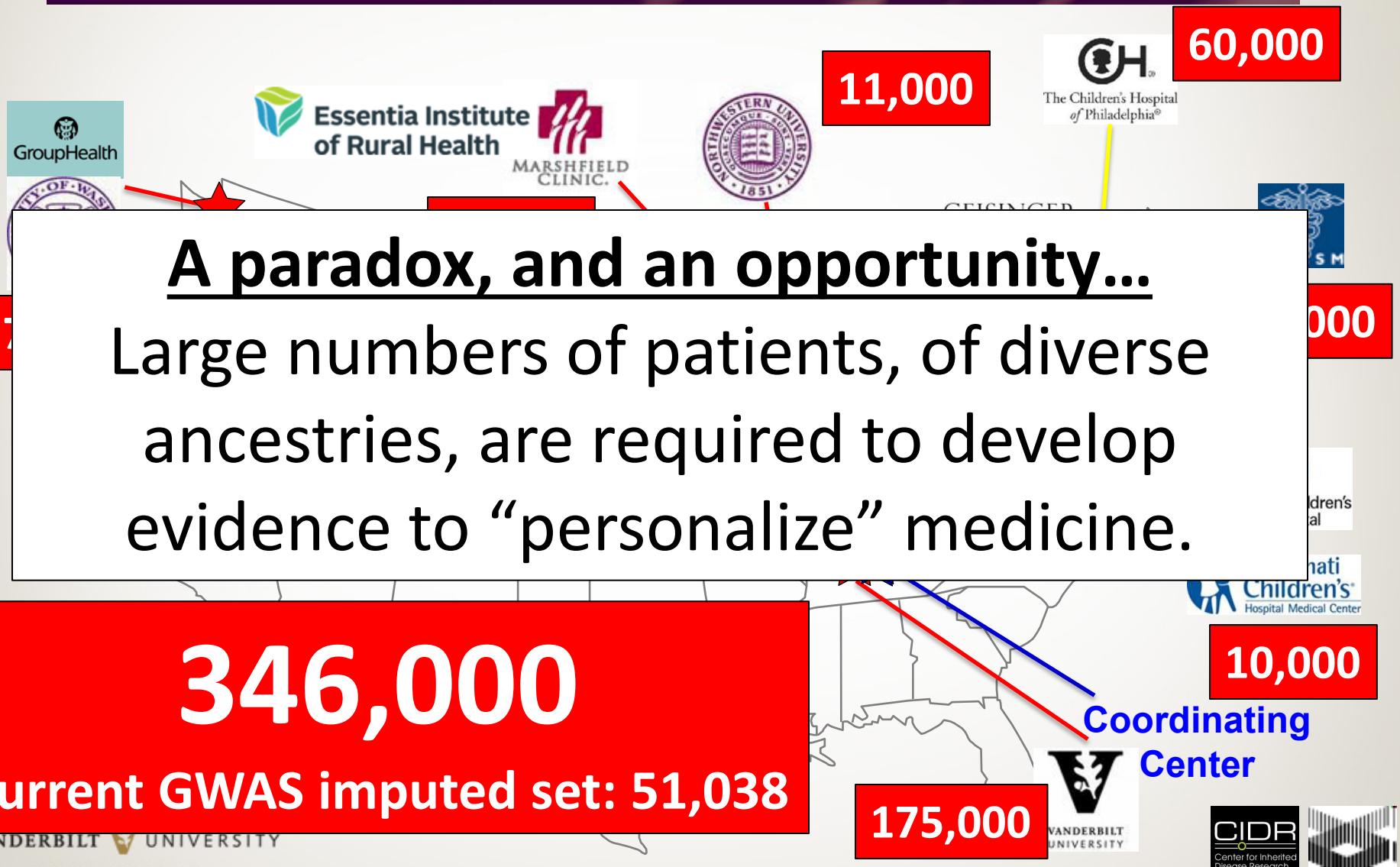


diagnosis code

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

eMERGE Network

electronic medical records & genomics



A paradox, and an opportunity…

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

346,000

Current GWAS imputed set: 51,038

Balancing the discovery and implementation missions

- What can eMERGE contribute to discovery…
  - …in which others also engaged?
  - …for which eMERGE is near-uniquely positioned?
- What can eMERGE contribute to implementation…
  - …in which others also engaged?
  - …for which eMERGE is near-uniquely positioned?

Discovery versus Implementation

The "easiest" examples

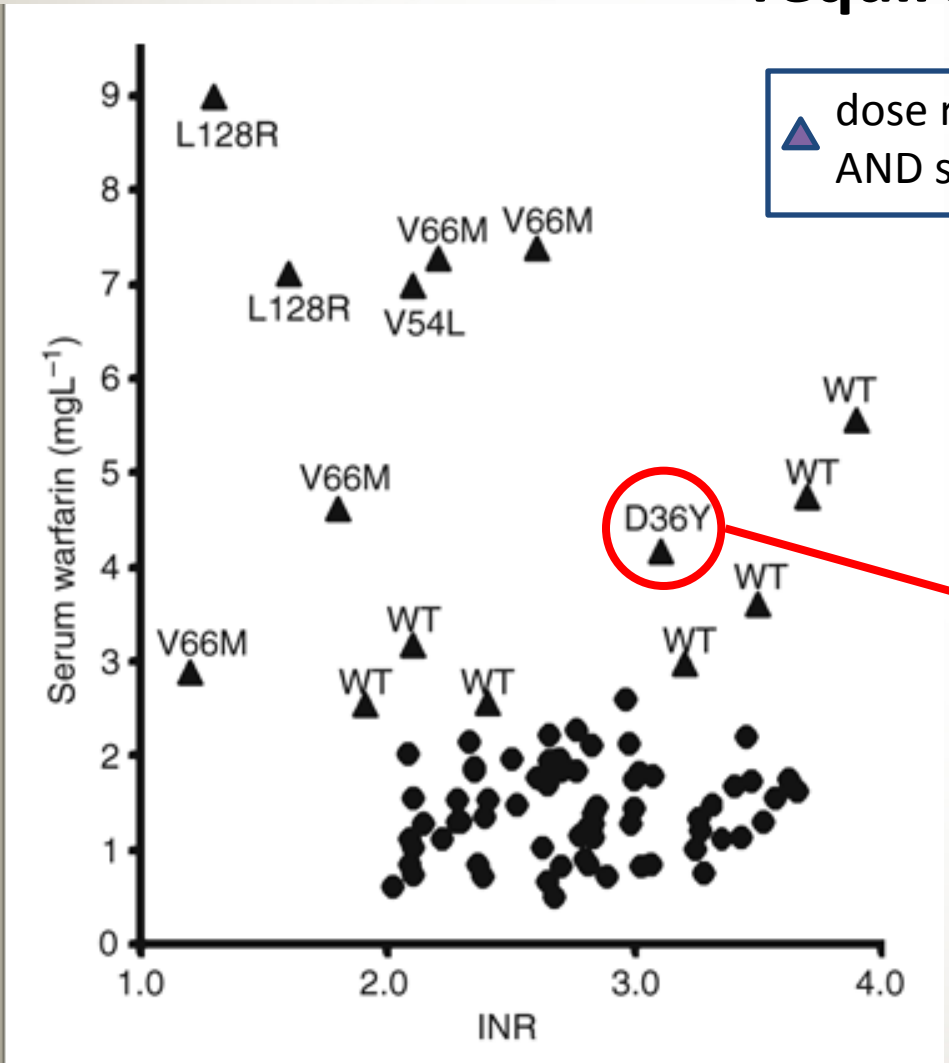
- Some drug responses
- Some cancer susceptibility

Do we really know all there is to know about variable responses to commonly used drugs?

- Rare variants
- Ancestry

Warfarin: not so simple….

Rare variants in VKORC1 associated with high dose requirements



▲ dose requirement >20 mg/day
AND serum warfarin >2.3

identified in 8/15
Ashkenazi patients
requiring >11 mg/day

Harrington et al., 2008

Warfarin: not so simple….

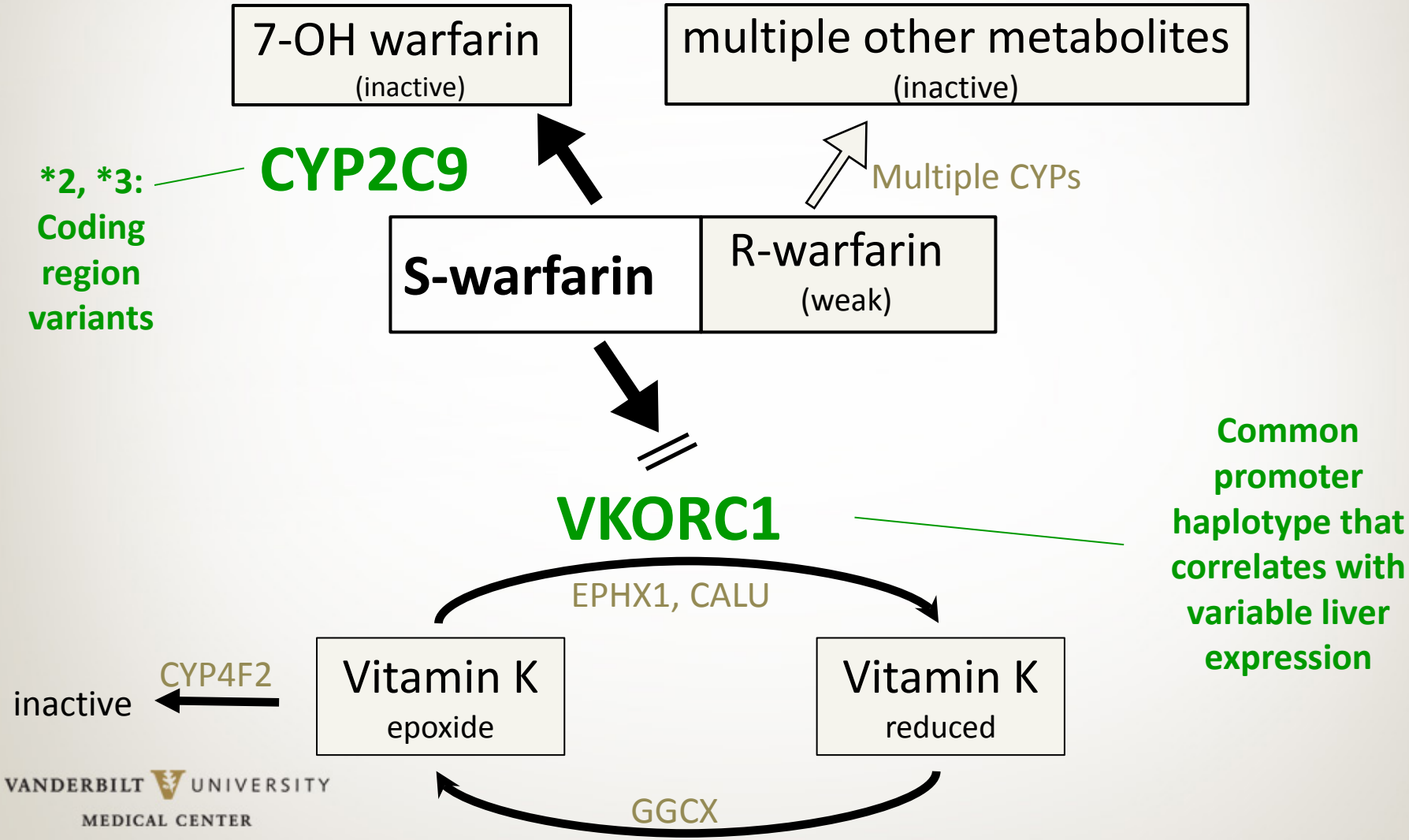
Gene	SNP	Minor Allele Frequency			All		
CYP2C9*2	rs1799853		12.86%		8.48E-12		
CYP2C9*3	rs1057910		5.72%		3.32E-25		
VKORC1	rs2359612		38.47%		6.38E-55		
VKORC1	rs9934438		38.11%		1.07E-60		
VKORC1	rs9923231		38.14%		3.40E-60		

Warfarin: not so simple….

Gene	SNP	Minor Allele Frequency Overall	Caucasian	AA	All n = 1,170	EA n = 1,025	AA n = 145
CYP2C9*2	rs1799853	11.53%	12.86%	2.45%	8.48E-12	1.45E-11	0.5047
CYP2C9*3	rs1057910	5.22%	5.72%	1.74%	3.32E-25	9.06E-24	0.01556
VKORC1	rs2359612	36.56%	38.47%	23.26%	6.38E-55	1.30E-58	0.3112
VKORC1	rs9934438	34.67%	38.11%	10.76%	1.07E-60	1.50E-58	0.002842
VKORC1	rs9923231	34.69%	38.14%	10.76%	3.40E-60	4.80E-58	0.002842

Multiple gene effect

The warfarin pathway



VANDERBILT UNIVERSITY
MEDICAL CENTER

Warfarin: not so simple….

Gene	SNP	Minor Allele Frequency Overall	Caucasian	AA	All n = 1,170	EA n = 1,025	AA n = 145
CYP2C9*2	rs1799853	11.53%	12.86%	2.45%	8.48E-12	1.45E-11	0.5047
CYP2C9*3	rs1057910	5.22%	5.72%	1.74%	3.32E-25	9.06E-24	0.01556
VKORC1	rs2359612	36.56%	38.47%	23.26%	6.38E-55	1.30E-58	0.3112
VKORC1	rs9934438	34.67%	38.11%	10.76%	1.07E-60	1.50E-58	0.002842
VKORC1	rs9923231	34.69%	38.14%	10.76%	3.40E-60	4.80E-58	0.002842
CYP4F2	rs2108622	28.10%	30.53%	10.84%	9.00E-07	1.85E-06	0.3671
EPHX1	rs2292566	14.22%	14.09%	15.14%	0.9372	0.5237	0.132
GGCX	rs11676382	9.04%	9.97%	2.45%	0.2755	0.3374	0.5976
GGCX	rs699664	37.81%	34.35%	37.93%	0.04851	0.05031	0.7907
CALU	rs339097	1.34%	0.05%	10.42%	0.06144	NA	0.04574
CYP2C9*6	rs9332131	0.31%	0.10%	1.74%	0.0008942	NA	0.001348
CYP2C9*8	rs7900194	NA	NA	6.94%	NA	NA	0.00701
CYP2C9*11	rs28371685	0.48%	0.25%	2.08%	0.6528	NA	0.427

Discovery versus Implementation

Some other "easy" examples

- Factor V Leiden
- HFE
- APOL1

The poster children:
Are these the only ones?
Deploy? How? How to measure impact?

Discovery versus Implementation

Getting harder

- Complex combinations of markers (e.g. risk scores): genomic and other

- Development and validation
- How to deploy
- How to measure impact and outcome

Discovery science that 346,000 DNA samples coupled to EMRs can enable

- PheWAS
- Complex outcomes:
  - Longitudinal over time
  - Disease x drug x response
  - Variable outcomes by disease subtypes

Discovery science that 346,000 DNA samples coupled to EMRs can enable

- PheWAS
- Complex outcomes:
  - Gene x Longitudinal over time
  - Gene x Disease x drug x response
  - Gene x Variable outcomes by disease subtypes

Discovery science that 346,000 DNA samples coupled to EMRs can enable

- PheWAS
- Complex outcomes:
  - Gene x Longitudinal over time
  - Gene x Disease x drug x response
  - Gene x Variable outcomes by disease subtypes
- Consideration of ancestry issues
- To what extent can data be deidentified and retain discovery value?

Implementation science that 346,000 DNA samples coupled to EMRs can enable

- What? What evidence matters?
- How?
- In who?
- Educating providers and patients
- Decision support
- Outcomes

Discovery

Implementation

Studying cohorts

- in purpose-generated research datasets
- in the EMR

Using a highly interactive electronic medical record

- to provide real-time clinical advice
- to track outcomes

VANDERBILT UNIVERSITY MEDICAL CENTER