packet

External Scientific Panel

emerge network
ELECTRONIC MEDICAL RECORDS & GENOMICS

ESP Conference Call
February 25, 2016

TABLE *of* CONTENTS

DEPARTMENT OF HEALTH & HUMAN SERVICES Public Health Service

National Institutes of Health
National Human Genome
Research Institute
31 Center Drive MSC 2152
Building 31, Room 4B09
Bethesda, MD 20982-2152

February 8, 2016

Dear eMERGE External Scientific Panel members,

We are happy to let you know that eMERGE Phase III, the next four-year award, started on September 1st, 2015. eMERGE Phase III aims to: (1) sequence and assess the phenotypic implication of rare variants in the eMERGEseq panel with 109 clinically relevant genes presumed to affect gene function in about 25,000 individuals; (2) integrate genetic variants into EMRs for improvement of genetic risk assessment, prevention, diagnosis, treatment and/or accessibility of genomic medicine; (3) create community resources such as phenotyping/genotyping tools; and (4) conduct research on best practices for informed consent, protection of human subjects for data sharing, and return of genomic results.

We appreciate the expertise and effort you devoted to prior phases of eMERGE, and we look forward to your continued input in Phase III. The first External Scientific Panel (ESP) meeting for eMERGE III will be held on February 25, 2016 at 3:30-5:00pm, via teleconference/webinar.

The eMERGE Coordinating Center (CC) has prepared this booklet in collaboration with the eMERGE investigators to ensure a productive meeting. We would like to ask that you review these materials prior to the meeting.

Within the booklet you will find the following important materials:

- Agenda for ESP Conference Call on Feb. 25
- Introduction to Phase III:
  - Sequencing and clinical reporting
  - Network data management
  - Network working groups
  - Minutes of the two recent eMERGE III SC meetings

Please note that these same materials will also be made available to you on the eMERGE website at https://emerge.mc.vanderbilt.edu/. If you have any questions, or would like more information, please do not hesitate to contact us or the CC program staff.

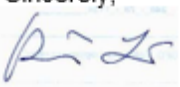
We welcome your input to make the eMERGE Network as successful as possible, especially input on sequencing (pipeline, dataflow, clinical reporting, timeline, and return of results). We also would appreciate your answers to the following questions:

1. How important would it be for the sequencing centers to report all classes of genetic variants (pathogenic, likely pathogenic, benign, likely benign, and uncertain significance as described in the "ACMG Standards and Guidelines") to all of the study sites, given that most of the study sites will only return pathogenic and likely pathogenic variants to patients and clinicians?

2. Do you have any concerns about dataflow and data management after reviewing the ESP packet?
3. In terms of the eMERGE III organization structure, are additional workgroup(s) needed for eMERGE III?

We look forward to your recommendations at the teleconference.

Sincerely,



Rongling Li, on behalf of the NHGRI eMERGE team

Rongling Li, MD, PhD, MPH,
Project Director, eMERGE
Division of Genomic Medicine
NHGRI, NIH
lir2@mail.nih.gov

AGENDA

External Scientific Panel (ESP): Conference Call

3:30 p.m. (EST) | February 25, 2016

Toll-Free: 1-888-936-7423

Long-Distance: +1 (510) 365-3331

Access Code: 662-506-819

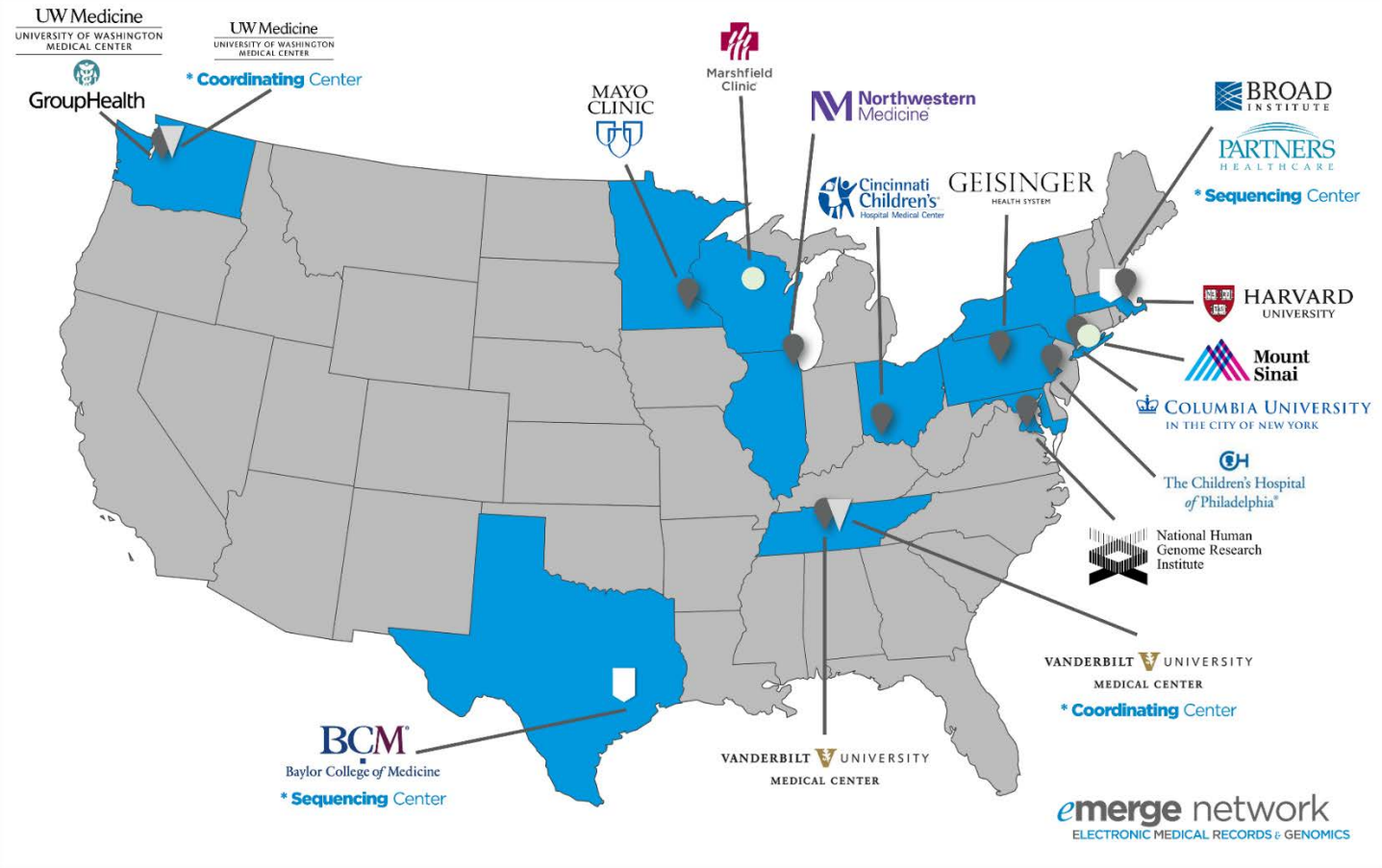
Meeting Link: https://attendee.gototraining.com/r/2111359563458610178

- Welcome, Opening Remarks, General Updates – Rongling Li & Howard McLeod ***2 minutes***
- Network Introduction
    - Summary – Rex Chisholm ***5 minutes***
    - Feedback from ESP Members ***5 minutes***
- DNA Sequence & Analysis Pipeline –Richard Gibbs & Niall Lennon ***15 minutes***
- Clinical Annotation Workgroup – Birgit Funke & Heidi Rehm ***15 minutes***
- Discussion and Suggestions from ESP ***20 minutes***
- Executive Session – Rongling Li ***30 minutes***

INTRODUCTION *to* PHASE III

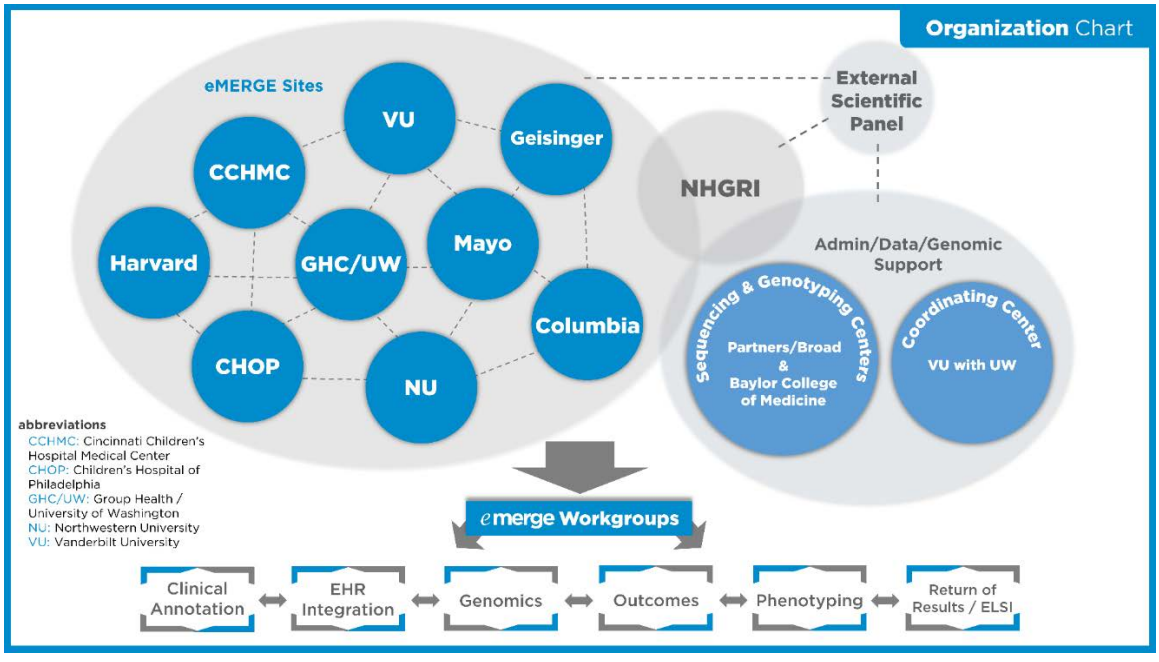
eMERGE is a national consortium, organized by NHGRI, that conducts discovery and clinical implementation research in genomics and genomic medicine at medical research institutions across the country. eMERGE research combines DNA biorepositories with electronic health record (EHR) systems for large-scale, high-throughput genetics research with the ultimate goal of returning genomic testing results to patients in a clinical care setting. eMERGE researchers are experts in the diverse fields of genomics, statistics, ethics, informatics, and clinical medicine

eMERGE studies and pilots **Genomic Medicine Translation** through **Discovery, Implementation, Tool Development, and Health Care and Social Impact Assessment**. During Phases I and II, the Network deployed 37 electronic phenotype algorithms across more than 58,000 subjects with dense genomic data, and more than 40 new phenotypes are prioritized for genomic and targeted sequencing data during eMERGE III. A large-scale survey of patient attitudes regarding data sharing was completed, contributing to rule making for biobanks. Sites across the network have implemented institution-specific models of pharmacogenomics, returning drug metabolism information in the clinic. Implementation in eMERGE III will represent a broader indication set, including ostensibly healthy subjects. Themes of bioinformatics, genomic medicine, privacy, community engagement, and human subjects protections are of particular relevance to eMERGE.



INTRODUCTION *to* PHASE III (cont.)

Site	Principal Investigator(s)
Children's Hospital of Pennsylvania (CHOP)	Hakon Hakonarson, MD, PhD
Cincinnati Children's Hospital Medical Center (CCHMC)	John Harley, MD, PhD
Columbia University	Chunhua Weng, PhD; Ali Gharavi, MD; & George Hripcsak, MD
Geisinger Health System	Marylyn Ritchie, PhD & Marc Williams, MD
Group Health Cooperative & University of Washington (GHC/UW)	Eric Larson, MD, MPH (GHC) & Gail Jarvik, MD, PhD (UW)
Harvard	Scott Weiss, MD; Elizabeth Karlson, MD; Shawn Murphy, MD; Jordan Smoller, MD
Mayo Clinic	Iftikhar Kullo, MD & Stephen Thibodeau, PhD
Northwestern University	Rex Chisholm, PhD & Maureen Smith, MS
Vanderbilt University	Dan Roden, MD & Joshua Denny, MD
Coordinating Center	Paul Harris, PhD (Vanderbilt)
Central Sequencing & Genotyping Centers (CSGs)	
Baylor College of Medicine	Richard Gibbs, PhD
Partners/Broad	Birgit Funke, PhD; Stacey Gabriel, PhD; & Heidi Rehm, PhD



GOALS *and* SPECIFIC AIMS

eMERGE III aims to continue to develop and validate electronic phenotyping algorithms for large-scale, high-throughput genomics research; to discover genetic variants related to complex traits; to disseminate results and lessons learned to the scientific community; and to deliver state-of-the-art genomic knowledge, methods, and approaches to clinical decision support and clinical care.

Specific Aims:

1. Sequence and assess the phenotypic implication of rare variants in ~100 clinically relevant genes presumed to affect gene function in about 25,000 individuals
2. Assess the phenotypic implications of these variants
3. Integrate genetic variants into EMRs for clinical care
4. Create community resources (RFA-HG-14-025, RFA-HG-14-026, RFA-HG-14-027)

Significant effort will be devoted to expanding utilization of the eMERGE PGx data generated in eMERGE II as well as the extensive GWAS data that has been aggregated.

In addition, eMERGE III will continue to assess health impact, cost effectiveness, and ethical, legal and social implications of reporting genetic variants on a broader population scale for patients, clinicians and healthcare institutions.

eMERGE-SEQ OVERVIEW *and* CLINICAL REPORTING

*e*MERGE III: GENE LIST DESCRIPTION

ACMG56

SITES' TOP 6

Inclusion Criteria:

- phenotype or indication
- clinically actionable or discovery
- sequencing performance
- size

SNPs

Inclusion Criteria:

- PGx SNPs
- known loci for clinical actionability or phenotypes
- ancestry panels
- fingerprinting

FINAL DESIGN

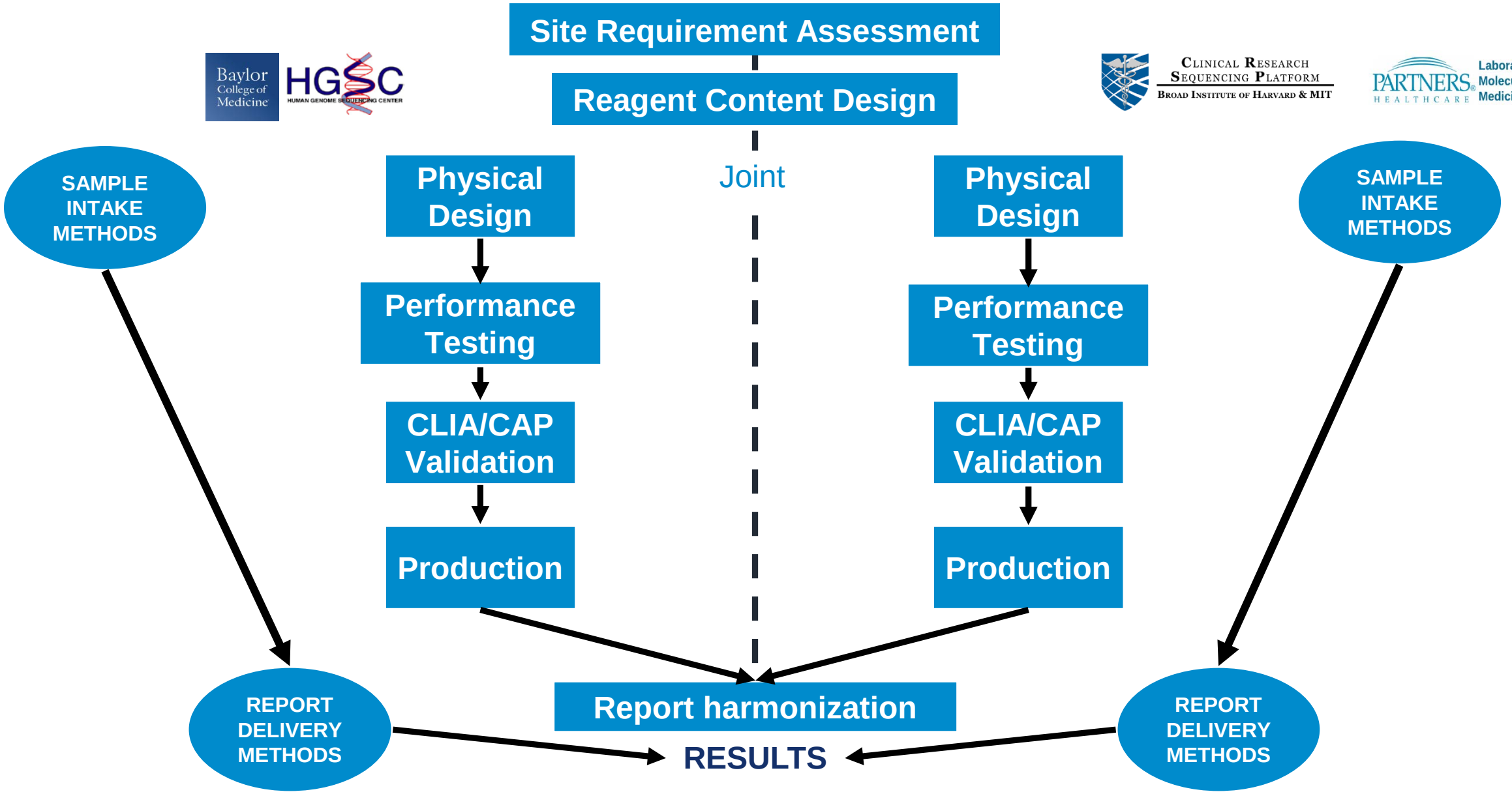
~ 110 Genes

~ 1500 SNPs

uniform coverage between sequencing centers

See Slide 35 for comprehensive gene list

eMERGE III: SEQUENCING WORKFLOW



*e*MERGE III: TECHNICAL DATA *on* *e*MERGE- seq PANEL

Test data from 83 samples:

GENERAL COVERAGE

Average Coverage	Median Coverage	Bases > 1X	Bases > 10X	Bases > 20X	Bases > 40X
335X	331X	99.95%	99.78%	99.62%	99.20%

GENE/TARGET COVERAGE (Total 109 Genes)

100% >20 X	97-100% > 20 x	90-97% > 20 x
102	5	2

SNP COVERAGE

1540/1551 Coverage > 20 x

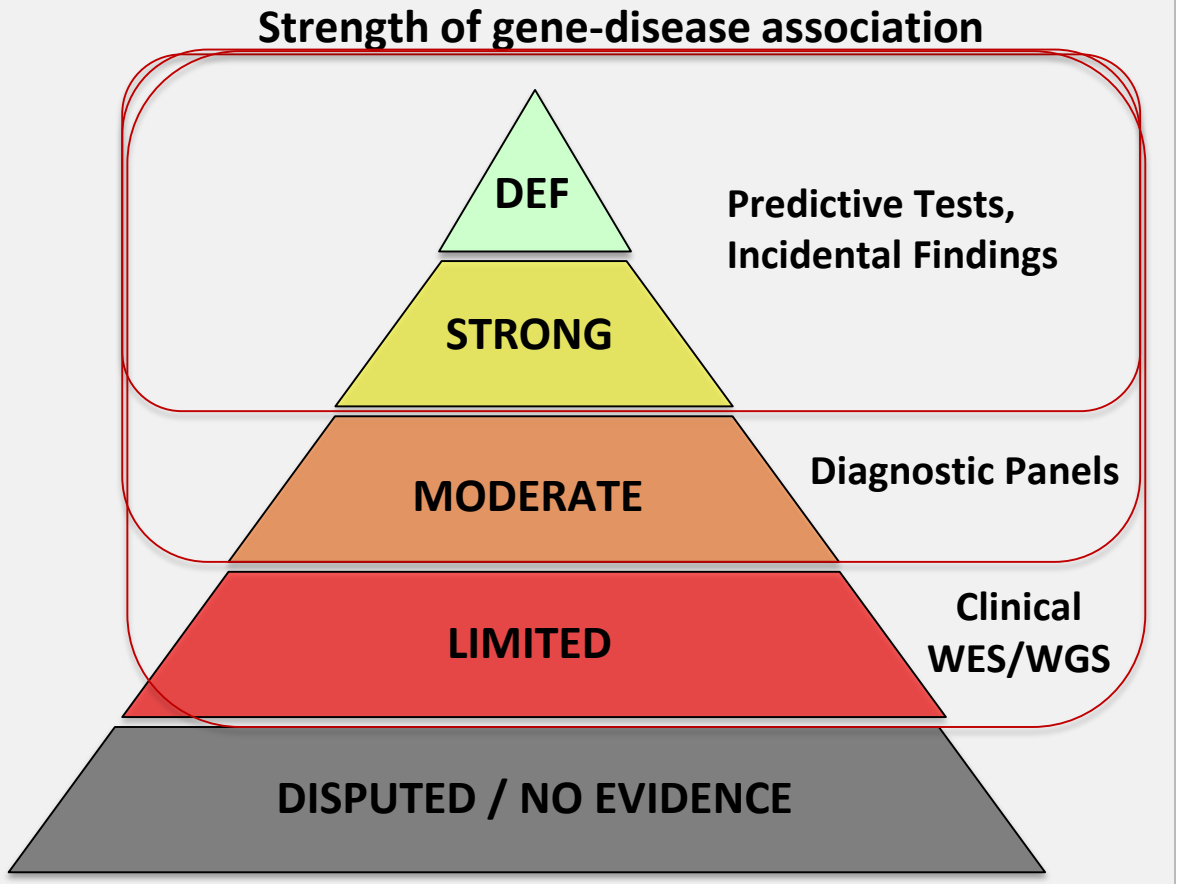
Conclusion: Outstanding reagent performance

*e*MERGE III: CLINICAL REPORTING

WHICH RESULTS SHOULD BE RETURNED TO PATIENTS?

Two step process

1. Clinical validity of genes/variants: is the disease association backed up by sufficient evidence?
2. Clinical actionability: would knowledge of a clinically valid variant impact management or treatment?



- No universal / mandatory guidelines (yet)

→ eMERGEIII: needs discussion

*e*MERGE III: CLINICAL REPORTING

eMERGE III panel = 109 genes + ~1500 SNPs (operational + submitted by sites)

eMERGE network perspective: What should be reported clinically?

- 56 genes ("ACMG56") - Clinical validity/actionability well defined
- 53 genes submitted by sites - Need to establish/confirm validity
- SNPs - Need to establish/confirm validity

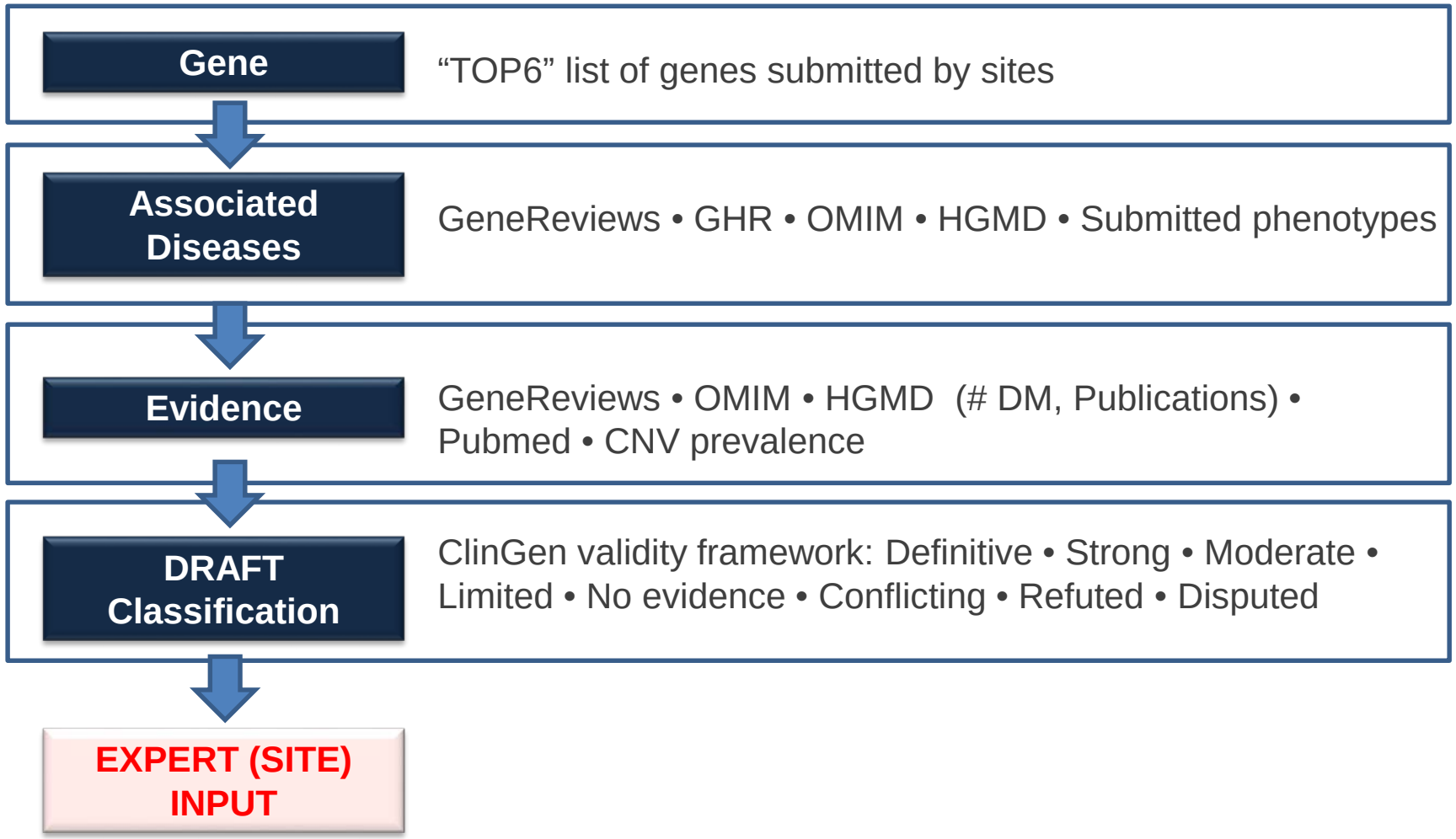
Sequencing Centers' perspectives:

- CSGs are maintaining clinically curated gene/variant databases for clinical ops
- We are adding eMERGE III to our routine clinical operations
- **Whatever we do for eMERGE needs to be in sync!**

Additional complexity:

- Minimize discrepancies between the **2 CSGs**
- Harmonize between **CSGs** and **sites** as to what is clinically valid

*e*MERGE III: CURATION PROCESS "TOP 6" GENES



*e*MERGE III: GENE CURATION

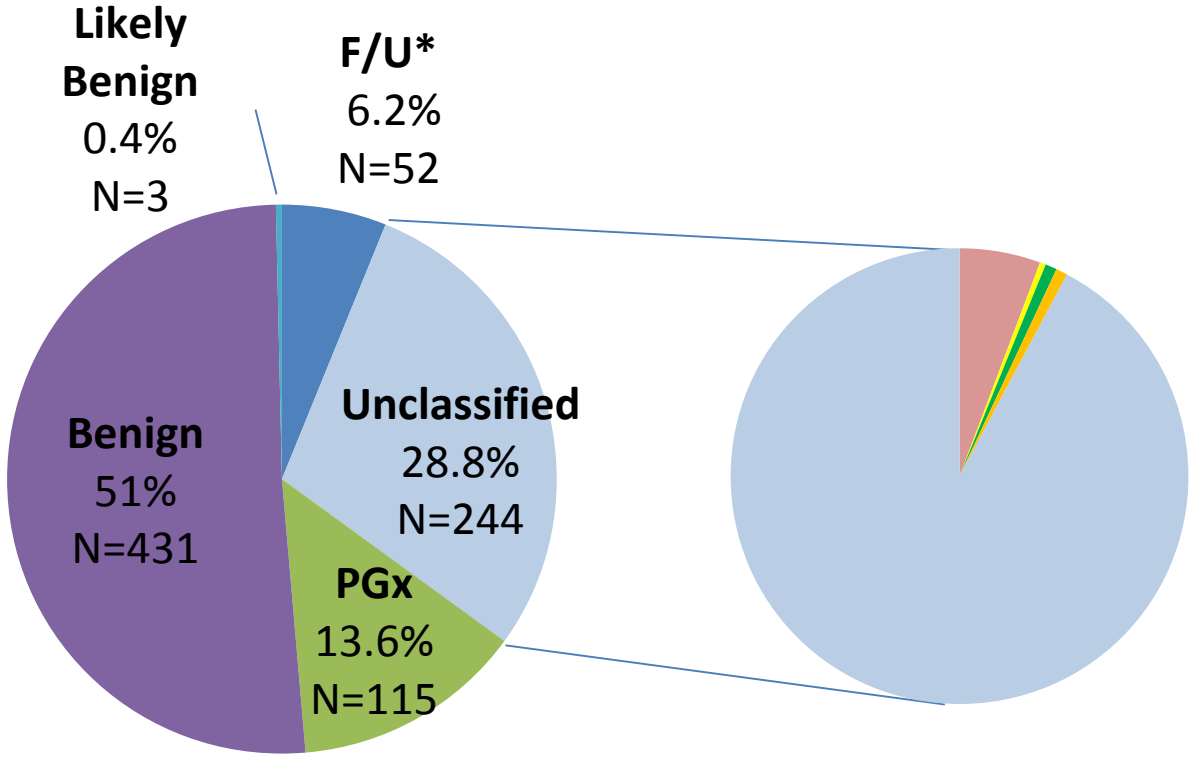
COL5A1: Well-established disease association

Site Requested	Site-Specific Phenotypes	Disease	ClinGen Classification	Summary of Evidence
CCHMC	**CCHMC**: EDS w/hypermobility **Mayo**: EDS	Ehlers-Danlos syndrome (EDS)	Definitive	• 46% of individuals with classic EDS have an identifiable pathogenic variant in *COL5A1* • >125 families with pathogenic variants • Segregation in affected individuals • Mouse knock-out model with similar phenotype

*e*MERGE III: SNP LIST CONTENTS CLINICALLY ASSESSED

Informatics based SNP triage variants (custom script)

- Benign/likely benign (based on MAF + absence in clinical databases)
- All else: "unclassified" → those need in depth assessment



*likely intronic SNPs (scripts used only ExAC and 1000 Genomes)

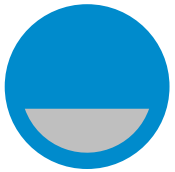
- Path, 5.8%, 14 SNPs
- LP, 0.4%, 1 SNP
- VUS, 0.8%, 2 SNPs
- Benign, 0.8%, 2 SNPs
- Not seen at Partners/Broad or Baylor, 92%, 225 SNPs

NEXT: Understand why these variants were submitted by sites

Not yet fully cross-referenced with BCM data

*e*MERGE III: GENE CURATION

Next Steps



1) Partners/Broad to curated all TOP6 genes

163 gene-disease pairs (53 TOP6 genes on eMERGE panel)

- Some >1 associated disease/phenotype
- Many are clinical variations of the same disease

72 gene-disease pairs done

29 DEFINITIVE

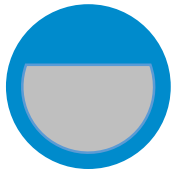
13 STRONG

11 MODERATE

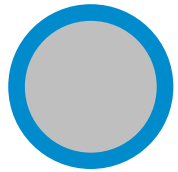
15 LIMITED

4 RISK

5 with a second association of MOD/LIM significance



2) Sites to review and comment



3) Network-wide consensus

Consensus on what is regarded actionable and should be included on clinical reports generated by the CSGs

*e*MERGE III: GENE CURATION

Harmonization of Variant Interpretation (LMM and Baylor)

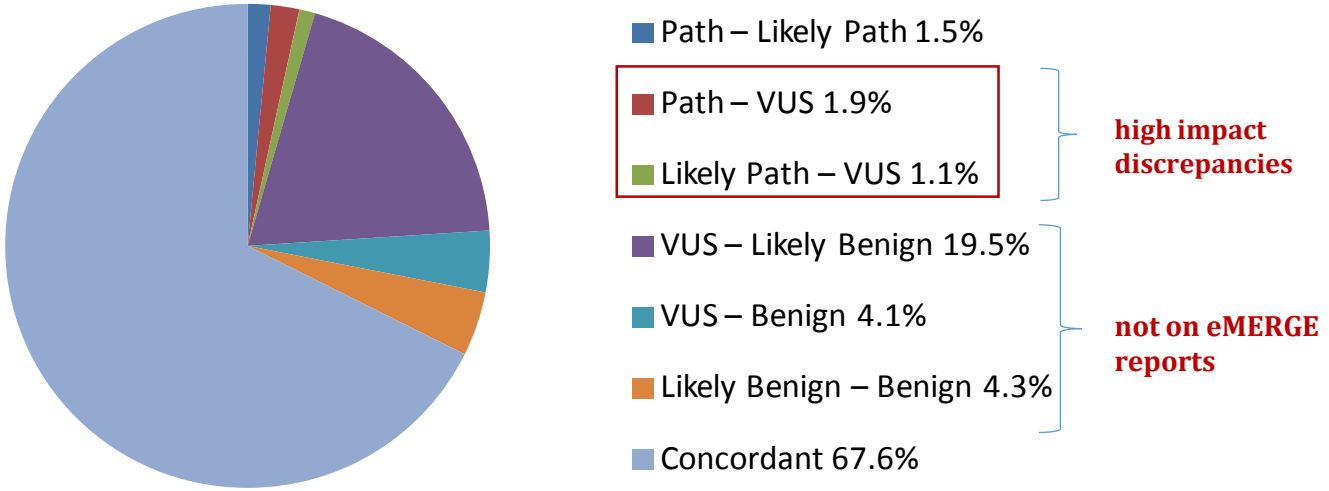
Status Updates

CSGs exchanged all previously reported variants and their corresponding clinical classifications (Pathogenic, Likely Pathogenic, VUS, Likely Benign, Benign)

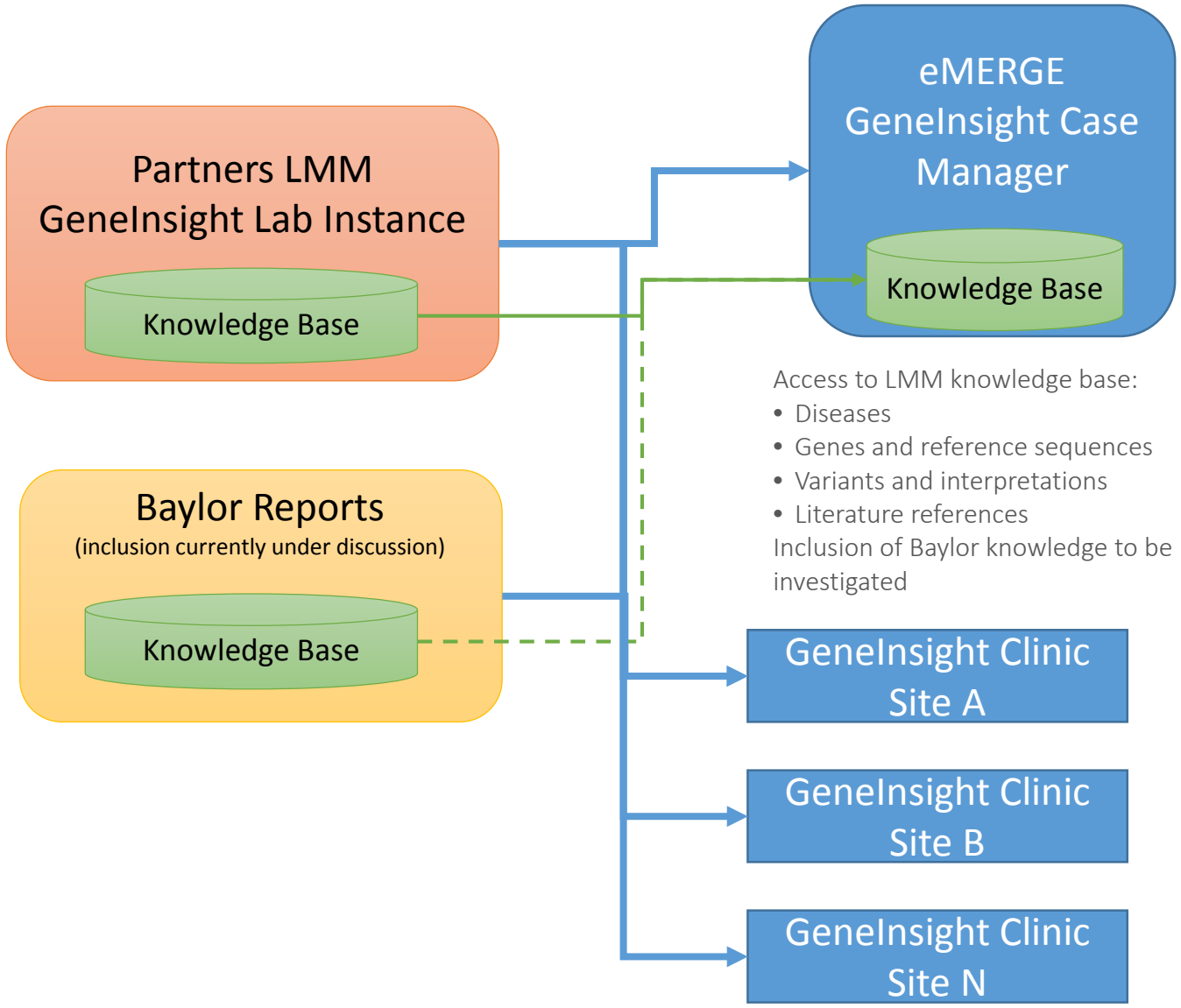
- LMM: n= 3,878 (880 seen >= 3x)
- BCM: n= 18,016 (3,104 seen >= 3x)

Discrepancy analysis (1,047 variants shared)

- 90% concordant (P, LP, VUS only)
- 67.5% concordant (all variant classifications) BCM only recently added Lik Ben + Lik Path



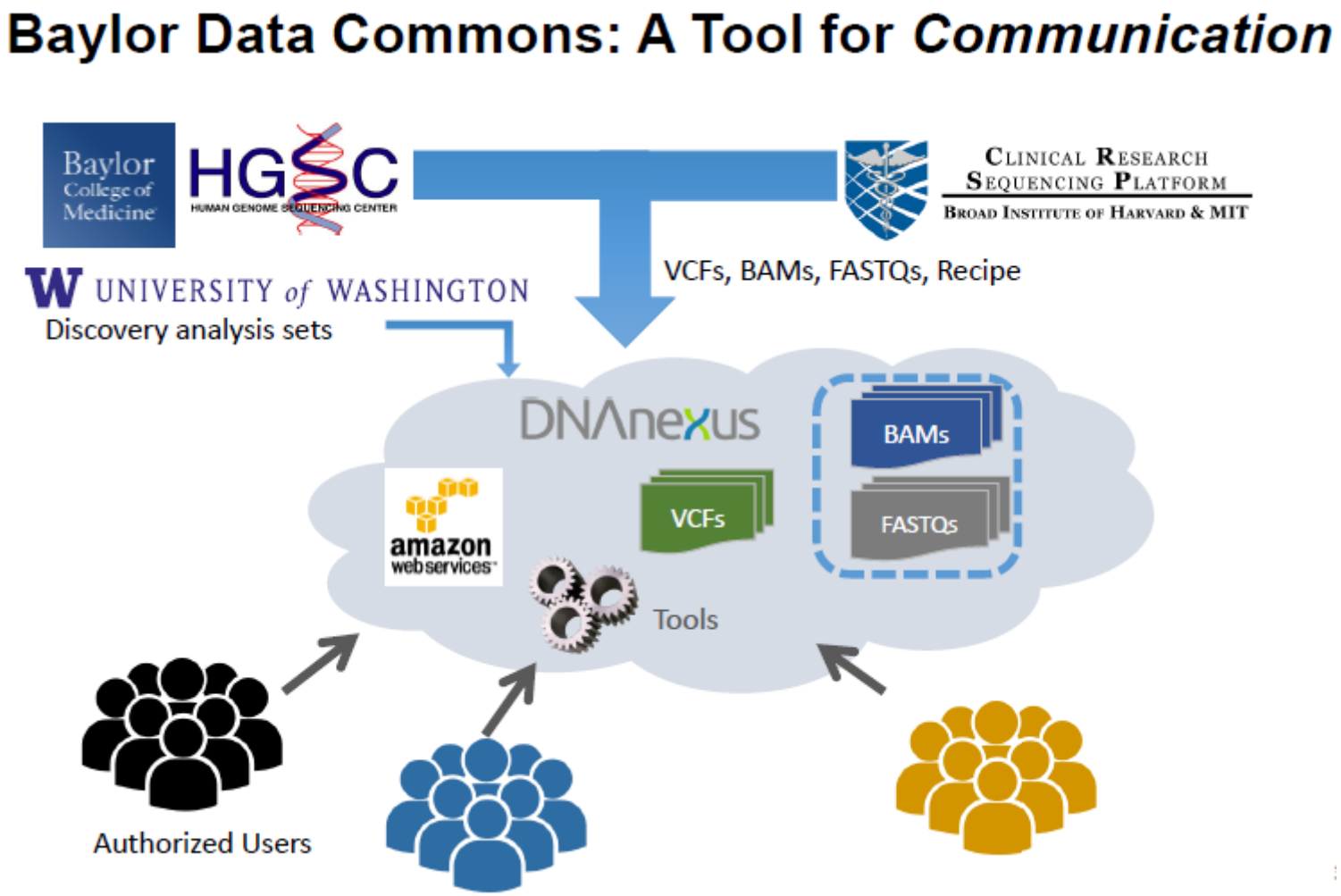
*e*MERGE III: ARCHITECTURE – *GeneInsight*



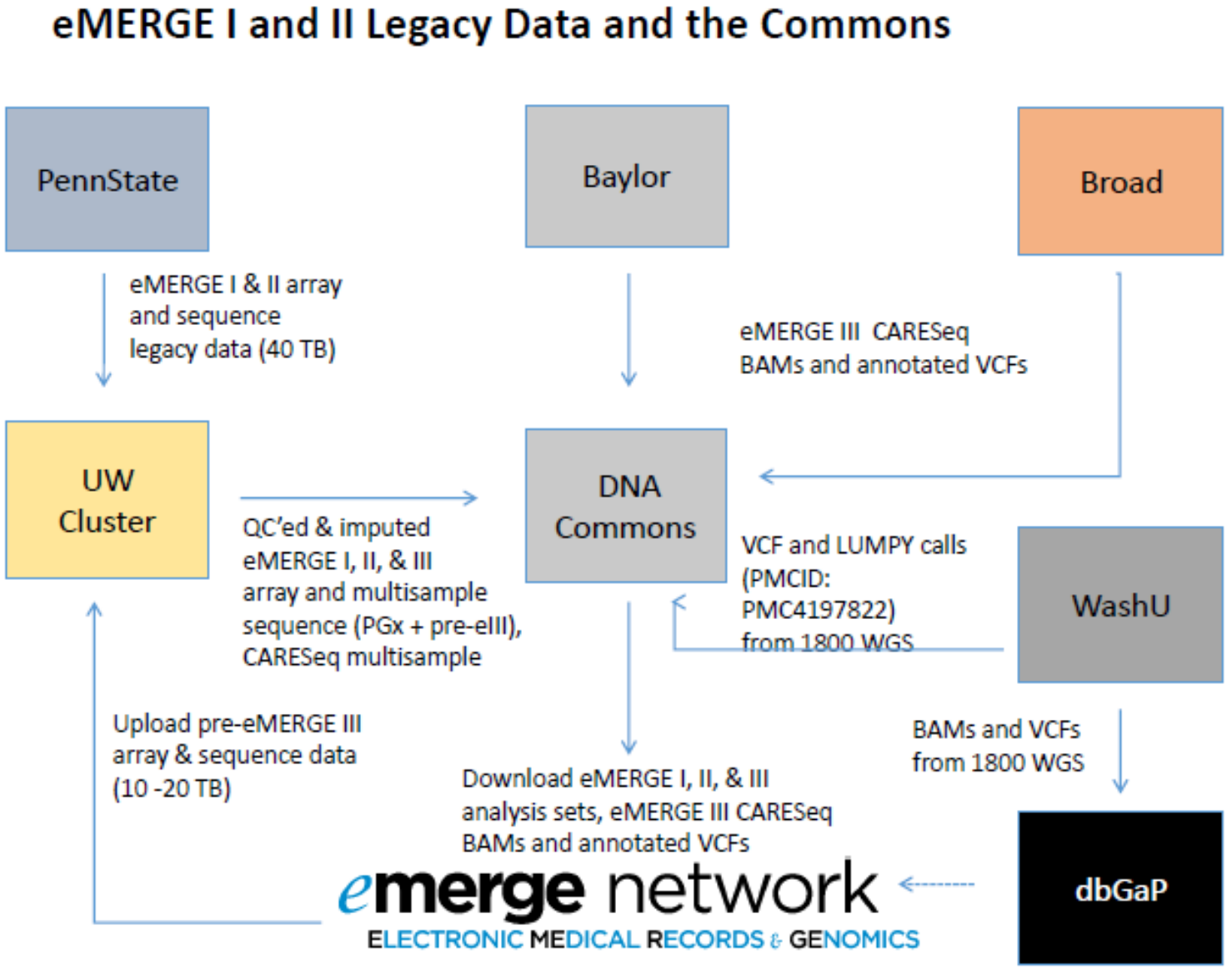
- Store de-identified reported eMERGE cases
- Provide query capabilities for searching across cases using different parameters
- All eMERGE sites, CC, CSGs, NHGRI log into the same instance and have access to all cases across network

- Simple query interface for finding specific cases, designed for physician use
- Designated study staff and/or physicians receive case specific variant alerts when variant interpretations change in their patients
- GICs are site specific and therefore can contain PHI

*e*MERGE III: ARCHITECTURE – *DNAnexus*



*e*MERGE III: ARCHITECTURE – *DNAnexus*



NETWORK DATA MANAGEMENT

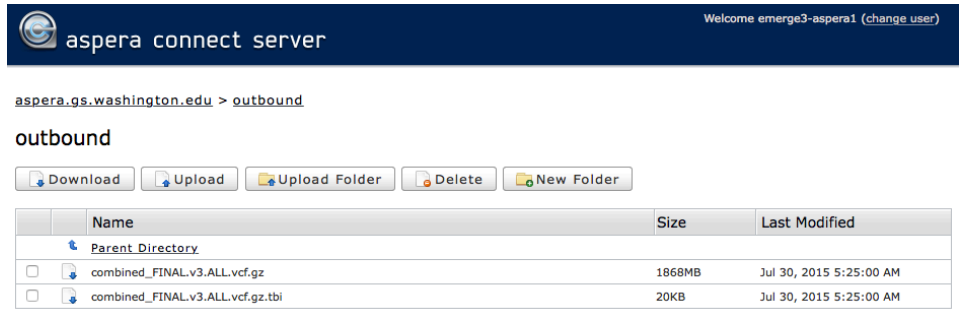
*e*MERGE III Genomic Data: Incoming Genotype Array Data Estimates

	CCHMC	CHOP	CU	Harvard	Mayo	GHC-UW	Geisinger	NU	VU
Ethnicity					**n=**				
His/L	52		671	269	1000	47	36	253	106
NonHis/L	1330		2416	4447		1906	9176	1154	10454
Unknown	5		0	215		0	0	0	87
Self-Report Race									
AI/AN	0	0	2	4		28	4	0	6
Asian	22	~60	180	79		69	19	17	29
NH/PI	2	0	0	0		2	5	1	0
Black/AA	152	~585	644	266		49	31	351	1444
White/EA	1100	~855	1930	4367		1768	9139	918	9101
Unk/NR	111	0	331	215		37	14	120	67
Total	1387	~1500	3087	4931	1000	1953	9212	1566	10,647

*e*MERGE III: CC – Genetic Data Activities

All eMERGE I & II legacy data (array and sequence) stored at the University of Washington

- Genotype array data (~55,000)
- Imputed data (~55,000)
- Sequence data (PGx target ~9000)



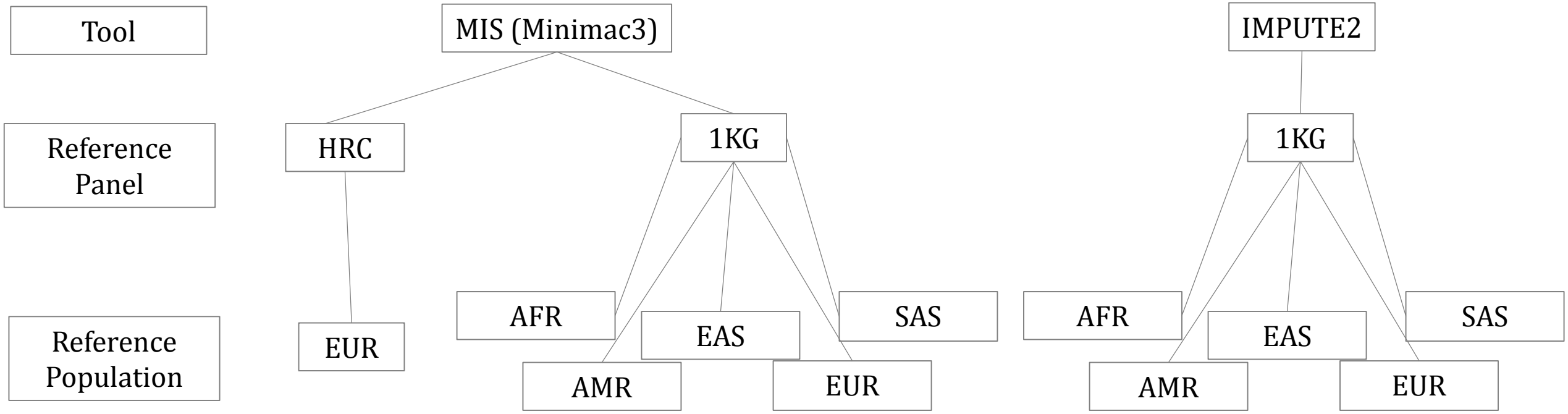
Network dissemination and acquisition through our Aspera server

- Dedicated 10Gb/sec Science DMZ/I2 network link for data
- Both analyses sets and raw data are available
  - E.g. Annotated PGx multisample and ~9000 BAMs

We will harvest eMERGE III BAMs (~25,000) from the CSGs and store for the duration of eMERGE III for future dissemination

- Will also manage the eMERGE III target multisample with current reference and calling algorithms

Imputation Comparisons: Michigan Imputation Server (MIS) vs IMPUTE2



- The CC (UW) plans to impute all pre-eMERGE III array data that is suitable for imputation.
- Randomly selected 1000 European ancestry and 1000 African ancestry eMERGE I participants
- Compared two reference panels (Haplotype Research Consortium – HRC and 1000 Genome -1KG)
- Compared MIS (Minimac3) and IMPUTE2 for imputation speed and metrics

*e*MERGE III Genomic Data: Proposed Samples for Sequencing

	CU	CCHMC	CHOP	Geisinger	GH-UW	Harvard	Mayo	NU	VU	TOTAL	
Ethnicity					n=						%
Hisp/L	502	56	180	36	94	120	510	386	132	2016	7.8
Non-Hisp/L	1998	2444	2820	2964	2473	2380	2490	2614	2868	23051	90.2
Race											
AI/AN		2		4	25	6	8	6	5	56	0.2
Asian	176	28	120	9	1204	50	2	59	26	1674	6.5
NH/PI		2		3	9			2	1	17	0.1
Bl/AA	250	1056	1170	35	62	142	9	486	782	3992	15.6
White	2040	1912	1710	2949	1260	2288	2471	2062	2186	18,878	73.9
Unk/MR	34				7	14	510	385		950	3.7
Total	2500	3000	3000	3000	2567	2500	3000	3000	3000	25567	

*e*MERGE III: PHENOTYPES

Lead Site	Phenotype	Intended Cohort
CCHMC	Pediatric Pain Perception, Pain Sensitivity, Migraine	*e3* Sequencing
	Primary Pulmonary Hypertension	*e3* Sequencing
	Hypermobility, EDS	*e3* Sequencing
	Pain Management, Opioid Dependence, Neonatal Abstinence	*e3* Sequencing
CHOP	Epilepsy, AED Response	*e3* Sequencing/GWAS
	Intellectual Disability	*e3* Sequencing/GWAS
	Obesity	*e3* Sequencing/GWAS
Columbia	Chronic Kidney Disease	*e3* Sequencing/GWAS
	Breast Cancer	*e3* Sequencing/GWAS
	Heart Failure / Cardiomyopathy	*e3* Sequencing/GWAS
	Liver Disease/Cirrhosis	*e3* Sequencing/GWAS
	Autoimmunity	*e3* Sequencing/GWAS
	Stroke / Cerebrovascular Disease	*e3* Sequencing/GWAS
Geisinger	Pediatric Familial Hypercholesterolemia	*e3* Sequencing
	Ornithine Transcarbamylase (OTC) Deficiency-non-classic Presentation	*e3* Sequencing
	Tuberous Sclerosis Complex	*e3* Sequencing
GHC/UW	Polyps / Familial Colorectal Cancer	*e3* Sequencing
	Endometrial and Ovarian Cancer	*e3* Sequencing
	Sexual Dysfunction	*e3* Sequencing
	Depression	*e3* Sequencing
Harvard	CAD	*e3* Sequencing/GWAS
	Hyperlipidemia	*e3* Sequencing/GWAS
	Bipolar	*e3* Sequencing/GWAS
	Schizophrenia	*e3* Sequencing/GWAS
	Asthma	*e3* Sequencing/GWAS
	Rheumatoid Arthritis	*e3* Sequencing/GWAS
Mayo	Contrast Nephropathy	GWAS/PGRN-Seq (PGx)
	Heparin-induced Thrombocytopenia	GWAS/PGRN-Seq (PGx)
	Metformin Response	GWAS/PGRN-Seq (PGx)
	Response to Heart Failure Medication	GWAS/PGRN-Seq (PGx)
NU	Valvular Disease	*e3* Sequencing/GWAS/PGRN-Seq (PGx)
	Atopic Dermatitis	*e3* Sequencing/GWAS
	Chronic Rhinosinusitis	*e3* Sequencing/GWAS
	Adult Headaches, Migraine	Sequencing, PGRN-Seq & GWAS
VU	Arrhythmias, (Atrial Fibrillation, QT Prolongation, Conduction System Disease, Brugada Syndrome)	*e3* Sequencing
	Cancer Susceptibility (plus Cancer PheWAS)	*e3* Sequencing
	Hereditary Amyloidosis	*e3* Sequencing
	Pneumonia	GWAS
	Urinary Tract Infections	GWAS
	Dry Eye	GWAS
	Hearing Loss	GWAS

INTRODUCTION *to* eMERGE WORKGROUPS

Clinical Annotation

Co-Chairs: Gail Jarvik (GHC/UW) & Heidi Rehm (Partners/Broad)

EHR Integration

Co-Chairs: Sandy Aronson (Harvard) & Casey Overby (Geisinger/UMD)

Genomics

Co-Chairs: Sekar Kathiresan (Harvard) & Megan Roy-Puckelwartz (NU)

Outcomes

Co-Chairs: Hakon Hakonarson (CHOP), Josh Peterson (Vanderbilt), & Marc Williams (Geisinger)

Phenotyping

Co-Chairs: Josh Denny (Vanderbilt) & George Hripcsak (Columbia)

RoR/ELSI

Co-Chairs: Ingrid Holm & Iftikhar Kullo

Clinical Annotation Workgroup

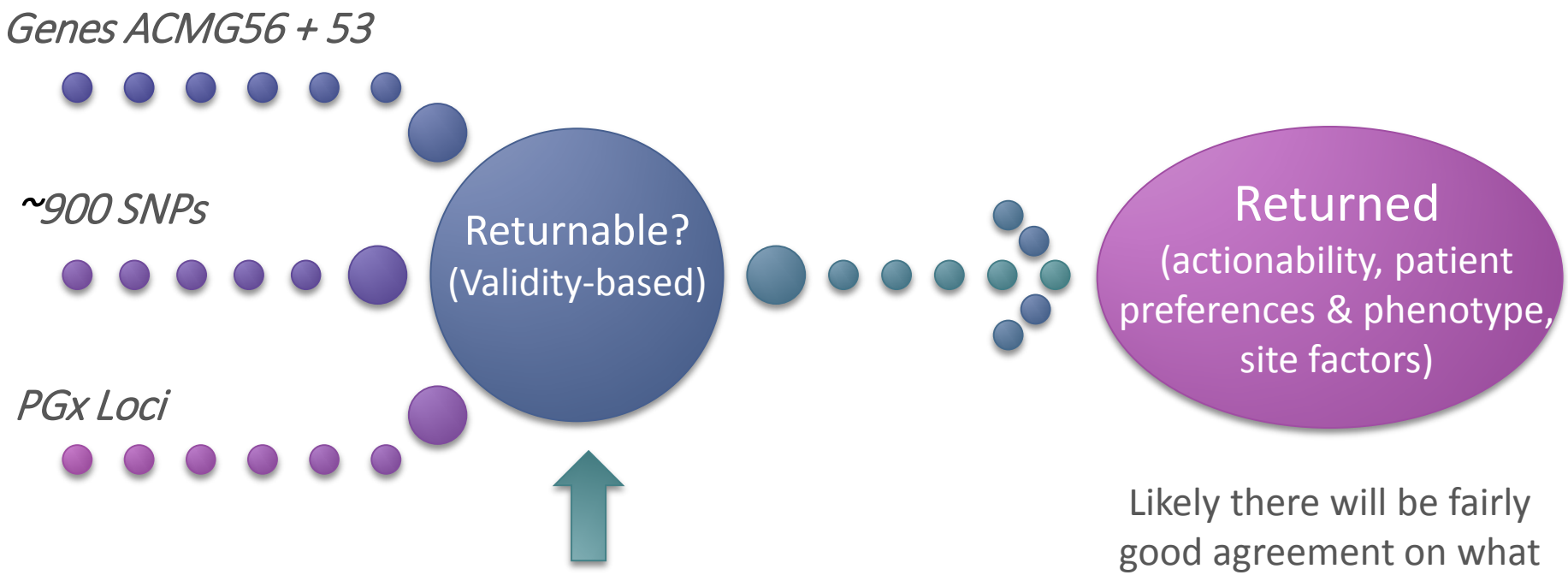
Co-Chairs: Gail Jarvik (GHC/UW) & Heidi Rehm (Partners/Broad)

Clinical Annotation Workgroup Charter

In eMERGE III, the Clinical Annotation workgroup will focus on activities that build consistency of approaches to the gene and variant interpretation across the eMERGE sequencing centers and study sites as well as support contribution to public knowledge bases.

1. **Apply the ClinGen approach to gene-disease validity assessment to all genes on the eMERGE gene panel (including SNP genes), defining each associated condition and the strength of evidence**
2. **Develop consistency in variant interpretation approaches**
    a. Compare variant interpretations from CSGs and eMERGE sites on all previously classified variants in genes in the eMERGE gene panel via comparison of ClinVar submissions
    b. Identify and resolve differences (prioritize most common and most different)
3. **Develop consensus on the most common clinically reportable variants in the eMERGE panel and whether to recommend return to patients**
    a. Evaluate evidence for pathogenicity (monogenic disease) or contribution to phenotype (PGx, risk alleles)
    b. Work jointly with the ROR/ELSI WG to decide categories of variants to return (by phenotype/condition, gene-disease validity level, actionability, penetrance, diagnostic vs SFs, etc.)
4. **Facilitate regular ClinVar submissions for all variants interpreted for the eMERGE program**
5. **Work with the ROR/ELSI WG to develop an environment for ongoing discussion and sharing of challenging genes, cases and variants considered for return (prospective or retrospective)**
6. **Work jointly with the ROR/ELSI WG to gather feedback and develop consensus on standard language used in clinical reports**

*e*MERGE III: FRAMEWORK *for* REPORT CONTENT



Goal to be in agreement across network

Likely there will be fairly good agreement on what goes into the report from the 109 genes but more distinction on the SNP list

*e*MERGE III: SITES EVALUATING PRELIMINARY GENE CURATION

Definitive or strong evidence

LIKELY ACTIONABLE GENE-DISEASE PAIRS				Additonal reported phenotypic manifestations of gene variants + Site phenotpes							
Gene	Disease 1	Disease 2	Disease 3	1	2	3	4	5			
ATM	Ataxia-telangiectasia (DEF)	Breast cancer susceptibility (STR)		Gastric, colorectal, ovarian, prostate, general cancer (LIM)	Lymphoma, mantle cell (LIM)	Hodgkin disease (LIM-NO EV)	VU cancer susceptibility				
ATP1A2	Hemiplegic migraine (DEF) Juvenile										

- Confirm draft validity classification
- Distinct disorders vs. phenotypic spectrum of single disease
- Define actionable gene-disease pairs

A	B	C	D	E	F	G	H	I	J	K	L	M
LIKELY ACTIONABLE GENE-DISEASE PAIRS					Additonal reported phenotypic manifestations of gene variants + Site phenotpes							
Gene	Disease 1	Disease 2	Disease 3		1	2	3	4	5	6	7	8
CFH					Hemolytic-uremic syndrome (HUS)	Complement factor H (CFH) deficiency (Columbia: autoimmunity)	Columbia: chronic kidney disease	Dense deposit disease (DDD)/membranoproliferative glomerulonephritis type II (MPGNII)	Age-related macular degeneration (AMD)			
CORIN					Pre-eclampsia, eclampsia	Harvard: CAD						

- Need expert guidance on which gene-disease pairs are likely "definitive" or "strong"

Courtesy of: Birgit Funke

EHR Integration Workgroup

Co-Chairs: Sandy Aronson (Harvard) & Casey Overby (Geisinger/UMD)

EHR Integration Workgroup Charter

Serve the eMERGE Network in three key areas:

Engineering → Establish, document and seek to continuously improve process flows for delivery of eMERGE reports and data

Science → Experiment with Innovative Approaches that Go Beyond Core Requirements and Evaluate Their Effectiveness

Community → Liaise with other groups, engage in collaborative projects, disseminate learning and best practices

EHR Integration Workgroup – Progress and Timelines

Co-Chairs: Sandy Aronson (Harvard) & Casey Overby (Geisinger/UMD)

Initial Focus is on Establishing Required Foundational Infrastructure

- Determine the **clinical report data flows** need within the eMERGE network to enable the sites to meet their objectives
- Specify **clinical system connections and enhancements** required to deliver data to sites in the manner they require

Activities

- Each site has produced a **high level clinical report data flow diagram** – ***Completed***
- Interviewing each site about **clinical system connections/enhancements** – ***In progress***
- Establishing a subgroup to specify a **common file exchange format** – ***In progress***
- **Identifying groups to engage** in collaboration and dissemination activities – ***In progress***
- Planning **network-wide scientific projects** – ***In progress***

Genomics Workgroup

Co-Chairs: Sekar Kathiresan (Harvard) & Megan Roy-Puckelwartz (NU)

Genomics Workgroup Charter

MISSION: The Genomics workgroup will identify best practices and facilitate analyses to assess the phenotypic impact of common and rare variant data arising from eMERGE II and III.

GOALS *for* eMERGE III:

1. **eMERGE has produced a number of GWAS with nearly significant hits or significant hits that require validation/replication. The Genomics workgroup will:**
    a. Coordinate further analysis of these datasets utilizing imputation with HRC of the eMERGE II data
    b. Coordinate integration of GWAS from two new sites
    c. Identify datasets that can either be bolstered or replicated by existing data at new eMERGE sites and facilitate exchange of data
2. **Interact with the CC and SC to identify and test possible QC and analysis pipelines for rare variant association testing**
3. **Determine if preexisting sequencing standards are appropriate for the genes sequenced in the eMERGE III cohort.**
4. **In conjunction with the Phenotyping working group the Genomics workgroup will:**
    a. Identify/compile existing phenotype data
        i. Create/maximize a central, highly detailed database for what data exists
    b. Systematically evaluate where data can be enhanced
    c. Prioritize data points that would be most powerful for both eMERGE II and eMERGE III data
    d. Implement processes to procure highest priority data and hasten experimental progress
5. **Update/Overhaul SPHINX to meet the broader needs of eMERGE III**
6. **Identify tools that need to be built for or included in DNA Nexus**
7. **Determine tools/metrics for functional annotation of variants**
8. **Include Structural Variants in final output**

Outcomes Workgroup

Co-Chairs: Hakon Hakonarson (CHOP), Josh Peterson (Vanderbilt), & Marc Williams (Geisinger)

Outcomes Workgroup Charter

MISSION: The Outcomes workgroup will develop cross-site outcomes to track implementation and impact of eMERGE III sequencing. The workgroup will focus on answering the overarching question of whether eMERGE III-generated genomic results impact health care utilization and outcomes of importance to patients and families.

Outcomes will consist of specific process measures or health outcomes to determine impact:

Process outcomes

a) Changes in health care utilization associated with reported genomic variation

b) Return of results process measures *(in collaboration with ROR workgroup)*

c) Clinician response measures *(in collaboration with ROR workgroup)*

Health outcomes

a) Intermediate outcomes *(a biomarker or finding indicating future benefit or harm is more likely)*

b) Clinical outcomes *(the benefits or harms to a patient who receives an intervention)*

c) Patient reported outcomes related to genetic susceptibility *(in collaboration with ROR workgroup)*

d) Family reported outcomes related to genetic susceptibility *(in collaboration with ROR workgroup)*

Objectives for the workgroup:

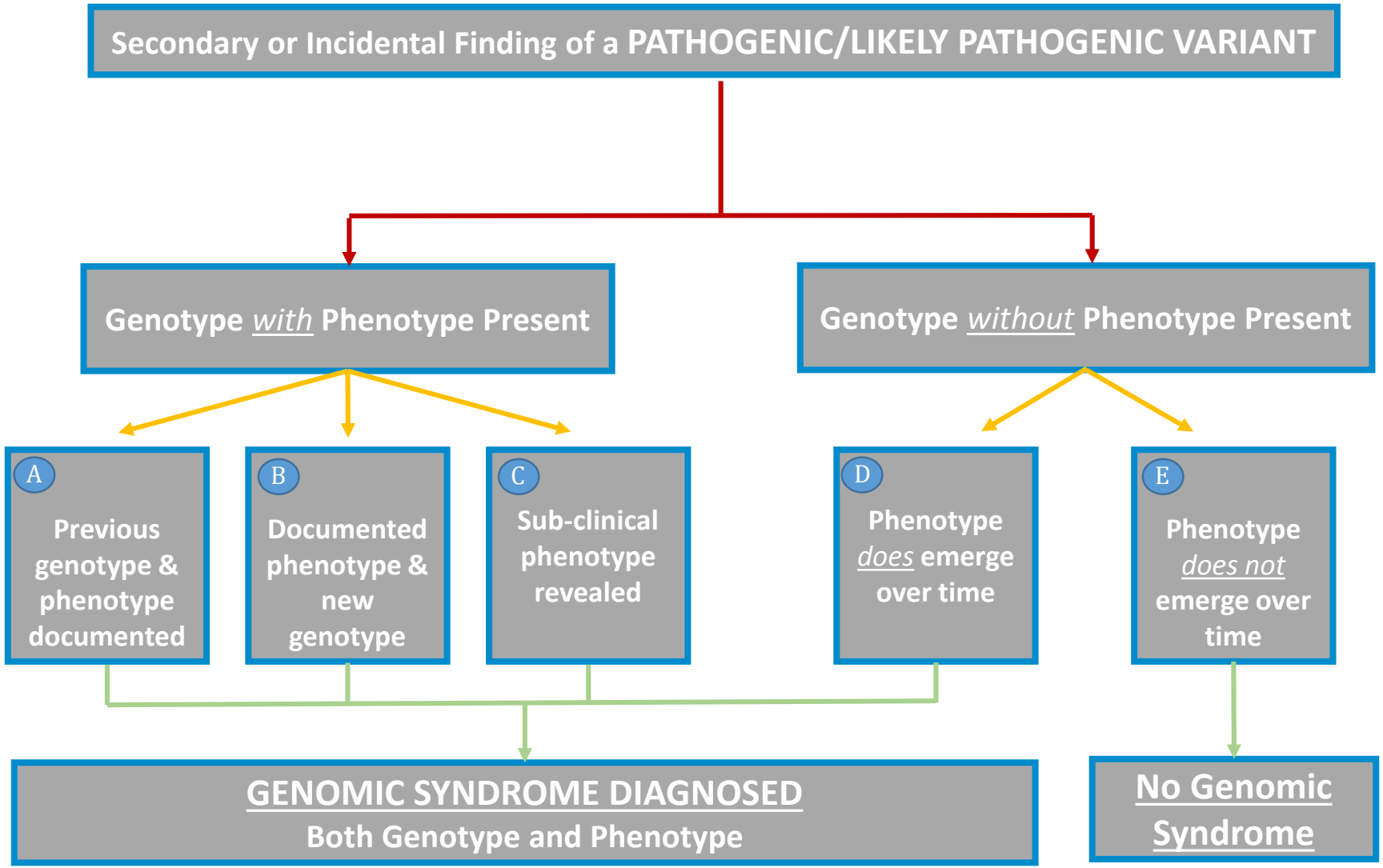
1. Define and prioritize eMerge III outcomes and impact
2. Develop a framework to guide outcome assessment at all sites
3. Designate mandatory vs optional outcomes
4. Develop a reporting mechanism and schedule
5. Follow through on eMERGE PGx evaluation plans

The Outcomes workgroup will create two *subgroups*:

1. Economic outcomes: the impact of outcome differences on economic measures
2. Pediatric outcomes: the distinct outcomes pertinent to pediatric enrollees

Outcomes Workgroup

Outcomes Framework: Classification and Definitions



Process Outcomes
potential changes in health care utilization related to returning genetic information

Intermediate or Surrogate Outcomes
e.g. a biomarker indicating benefit or harm is more likely

Clinical Outcomes
e.g. the benefits or harms to a patient who receives an intervention

Outcomes Workgroup

Outcomes Map: Structured by Site Phenotypes and Associated Genes

Site	Phenotype	Associated Genes
CHOP	Epilipsy, AED response	*GABRD, SCN1A, SCN2A*
	Intellectual disability	*CHRNA7, DPP6, GRM1, KIF1A, MAPT, PAFAH1B1, PPP2R1A, TCF4*
	ASD	*CACNA1B, GRM1, GRM2, GRM3, GRM4, GRM5, GRM6, GRM7, GRM8*
	Obesity	*FTO, LEP, MC4R, PCSK1, POMC*
CCHMC	Pediatric Pain Perception, Pain Sensitivity, Migraine	*SCN9A, NTRK1, COMT, MTHFR, ESR1, ESR2, GCH1, DRD1, DRD2, DRD3, SLC6A4*
	Primary pulmonary hypertension	*BMPR2*
	Hypermobility, EDS	*COL5A1, COL5A2, COL3A1*
	Pain Management, Opiod Dependance, Neonatal Abstinence	*CYP2D6, OPRM1, FAAH, ABCB1, SLC22A1*
Columbia	Chronic Kidney Disease	*HNF1B, TTR, GLA, CFH, C5, MEFV*
	Breast Cancer	*CHEK2, PALB2, JAK2, ATM*
	Heart Failure / Cardiomyopathy	*HFE, TTR, GLA, MEFV*
	Cirrhosis	*HFE, PNPLA3, TM6SF2*
	Autoimmunity	*CFH, C2, C5, BANK1, IFIH1*
	Stroke / Cerebrovascular Disease	*GLA*
Geisinger	Familial Hypercholesterolemia	*LDLR, APOB, PCSK9, ANGPTL3, ANGPTL4, APOA5, LPL, PLTP, PON1, SLC25A40*
	Chronic Rhinosinusitis	*CFTR, NOS1, TNF, IL33, SERPINA1*
	Ornithine transcarbamylase (OTC) deficiency-non-classic presentation	*OTC*
	Tuberous Scleroisis Complex	*TSC1, TSC2*

Group Health/UW	Colorectal cancer/polyps	*APC, MSH2, MLH1, PMS2, MSH6, TP53, STK11, MUTYH, PTEN, POLE, POLD1, BMPR1A, SMAD3, SMAD4, TGFBR1, TGFBR2, SNPs, GWAS*
	Ovarian Cancer	*SLC25A40, PON1, PLTP, APOE, ANGPTL3, ANGPTL4, APOA5, LPL, APOB, APOE, LDLR, PCSK9*
	Sexual Dysfunction	*APOB, APOE, LDLR, PCSK9, SLC25A40, PON1, PLTP, LPL*
	Depression	*GRM8, VDR, CACNA1C, various SNPs, GWAS*
Harvard	CAD	*CORIN*
	Hyperlipidemia	*ANGPTL3*
	Bipolar	*CACNA1C*
	Schizophrenia	*TCF4*
	Asthma	*VDR*
	Rheumatoid Arthritis	*TYK2*
Mayo	Familial Hypercholesterolemia	*LDLR, APOB, PCSK9 - actionable ; LDLRAP1 ; APOA5, APOC3, LPL, APOE – potentially actionable (studies favoring benefit of targeted TG-centered intervention are still underway, e.g. –APOCIIIRx antisense*
	Polyps / Familial Colorectal Cancer	*MLH1*, *MSH2*, *MSH6*, *PMS2*, *EPCAM* (Lynch syndrome); *APC* (Familial adenomatous polyposis); *MYH/MutYH* (MYH-associated polyposis); *STK11* (Peutz-Jeghers syndrome); *PTEN* (PTEN hamartoma tumor syndrome (i.e., Cowden syndrome)); *TP53* (Li-Fraumeni syndrome); *BMPR1A*, *SMAD4* (Juvenile polyposis syndrome); *GREM1* (Hereditary mixed polyposis syndrome); *AXIN2 (Oligodontia-colorectal cancer syndrome), JAK2, EPCAM, SDHA, POLD1*
	Ascending Aortic Dilatation/Aneurysm	*FBN1, ACTA2, TGFBR2, MYH11, TGFBR1, SMAD3, MYLK – actionable, SLC2A10 (arterial tortuosity syndrome), COL5A1 (Ehlers-Danlos syndrome) – actionable, discovery, SMAD4, FBN2 (Beals-Hecht syndrome), COLSA2, COL3A1 (Ehlers-Danlos syndrome) – actionable, NOTCH1 (aortic valve disease) – discovery, TGFB2 – discovery, actionable*
	Triglycerides	
Northwestern	Atrial Fibrillation	*SCN5A, LMNA, RYR2, KCNQ1, KCNH2*
	Valvular disease	*FBN1, TGFBR1, TGFBR2, SMAD3, ACTA2, MYLK, MYH11*
	atopic dermatitis	*FLG1, FLG2, TSLP, IL4, IL4R, IL13, SPINK5, IL1, CCL17*
	Chronic Rhinosinusitis	*CFTR, NOS1, IL33, SERPINA1, TNF*
	Adult Headaches, migraine	*CACNA1A, ATP1A2, SCN1A*
Vanderbilt	Cirrhosis	*HFE, SERPINA1*
	Arrhythmias, (Atrial fibrillation, QT Prolongation, conduction system disease, Brugada Syndrome)	SCN5A, KCNQ1, KCNH2, RYR2, KCNJ2, ANK2, KCNE1, CACNA1C, LMNA
	Cancer Susceptibility (plus Cancer PheWAS)	CHEK2, PALB2, JAK2, ATM, Breast: BRCA1, BRCA2, PTEN, TP53; Colon/GI: APC, MLH1, MSH2, MSH6, MUTYH, PMS2, STK11; Endocrine: MEN1, NTRK1, RET, SDHAF2, SDHB, SDHD; Neuro: NF2, TSC1, TSC2; Ovarian: BRCA1, BRCA2, MLH1, MSH2, MSH6, PMS2; Pancreatic: MLH1, MSH2, MSH6, PMS2
	Hereditary Amyloidosis	TTR
	CFTR PheWAS	CFTR

Phenotyping Workgroup

Co-Chairs: Josh Denny (Vanderbilt) & George Hripcsak (Columbia)

Phenotyping Workgroup Charter

The Phenotyping Workgroup carries out core functions in eMERGE III phenotyping and advances the science of phenotype development.

Phenotyping is defined broadly, including not only case and control identification, but also cohort identification—with probability estimation and subtype determination—and the extraction of continuous features.

The workgroup:

- defines the process for generating phenotypes,
- manages phenotype development, validation, and evaluation,
- facilitates research into symbolic and numeric techniques like knowledge engineering and machine learning,
- adopts or develops standards for phenotyping,
- collaborates with other workgroups and outside stakeholders, and
- disseminates the algorithms, tools, and results.

Phenotyping Workgroup – Progress and Timeline

Co-Chairs: Josh Denny (Vanderbilt) & George Hripcsak (Columbia)

1) Prioritize Phase I and Phase II phenotypes

- Based on priorities set by primary sites (e.g., is study completed) and network-wide feasibility assessments
- Distinguished which sites (only new or all) and whether new subjects or rerun all
- First four phenotypes: Atopic Dermatitis, BPH, ADHD, and Appendicitis

2) Prioritize Phase III Phenotypes

- Site survey pending

3) Update

- Initial 3,000 genotype cohort

4) Common Data

- Criteria: freely available, open definition process, deep information model, broad coverage, extensible
- Purposes: consistent phenotype implementation, computable representation, sharing, actual storage data (latter is for the future)
- Currently surveying sites for current information models
- Columbia to pilot OHDSI-OMOP Common Data Model

ROR/ELSI Workgroup

Co-Chairs: Ingrid Holm (BCH), Iftikhar Kullo (Mayo)

ROR/ELSI Workgroup Charter

1. **Actionability.** Develop and identify categories and thresholds of actionability. The Clinical Annotation Workgroup will initially assign these categories and thresholds to individual variants.
2. **Versioning**. Assess ways to address the dynamic nature of genetic knowledge, i.e. potential change in risk as additional susceptibility variants are identified (with Clinical Annotation Workgroup).
3. Review **methods of governance** including informed consent at sites and the role of participant and patient decision making in return of results
4. Evaluate **mechanisms of ROR at sites**. Review commonalities and differences and establish standards, e.g. around how and when results are returned and what information is provided with the results
5. **Develop and assess interpretive reports**, clinical decision support logic, and provider education (with EHRI WG)
6. Review **patient education** and patient portals
7. Assess the **ethical, legal, and social implications** of returning results in eMERGE III, in particular incorporation into the EHR
8. Assess the impact of ROR on patients' relationship with their health care providers
9. Evaluate the psychosocial responses to the ROR, including the impact on participants and patients and their families (with Outcomes WG)
10. Study the impact of data sharing on participant and patient **privacy and confidentiality**

ROR/ELSI Workgroup – Progress & Timelines

Co-Chairs: Ingrid Holm (BCH), Iftikhar Kullo (Mayo)

- **Monthly conference calls**
- **Collected data data from all sites on return of results projects and plans at each site, as well as outcome measures**
  - Timeline: Completed
- **Developing projects to study the impact of return of results on patients across the eMERGE sites.**
  - Timeline: Define projects in year 1
- **Develop and publish standards for ROR for eMERGE.**
  - Timeline: First 12 months
- **Studies on the ELSI issues of ROR on patients: Develop surveys or other data collection tools to implement across the sites.**
  - Psychosocial impact
  - Impact on families
  - Parent/child relationships
  - Timeline: Develop project over the first 12 months.
- **Coordinating efforts with the CSER consortium on outcomes and measures**
- **Develop surveys or other data collection tools to study impact on return of results on health care providers across site – to submit for supplemental funding**
  - First year
- **Joint meetings with the Outcomes WG to coordinate efforts across the WG.**
  - Ongoing
- **Joint publication with Clinical Annotations group – eMERGE process and criteria for actionablilty of variants for return.**
  - Timeline: First 1-2 years

PGx Status *from* eMERGE II

PGx Dataset:
PGRNseq data & EHR data for 9015 Subjects

Return of Results implemented at all 10 sites

Sequencing at 5 different sites showed > 99% concordance

SPHINX
public variant search + PGx population maf (global, ea, aa)
emergephinx.org
9 Sites
82 Genes
38112 variants
60 pathways
515 drugs

SPHINX
cohort definition: login required
demographics
ICD / CPT codes
meds
variants
pathways

publications

Association of Arrhythmia-Related Genetic Variants With Phenotypes Documented in Electronic Medical Records. JAMA. 2016 Jan 5;315(1):47-57.

Practical considerations in genomic decision support: The eMERGE experience. J Pathol Inform. 2015 Sep 28; 6(50).

Design and anticipated outcomes of the eMERGE-PGx project: a multicenter pilot for preemptive pharmacogenomics in electronic health record systems. Clin Pharmacol Ther. 2014 Oct;96(4):482-9.

Genetic Variation among 84 Pharmacogenes: the PGRN-Seq data from the eMERGE Network. Clin Pharmacol. Ther. (accepted with minor revisions)

phenotype data collected – not in SPHINX

Lipids
WBC
RBC
platelets

process outcomes studies

Provider Education
Patient Education
CLIA Concordance
CDS Comparison
Incidental Findings

phenotypes

MACE on Clopidogrel
Methylphenidate
Intractable Epilepsy
Lipids Levels
Adverse Events

CERC SURVEY *from* eMERGE II

Patient Perspectives on Broad Consent in Biobank Research in the eMERGE Network: Brief Overview & Findings

Background: A survey regarding participant willingness to enroll themselves, and their children <18 years of age, in a biobank and perspectives on broad consent and data sharing

- Participants randomized to 1 of 3 hypothetical biobanks: 1) Tiered consent, controlled data sharing; 2) Broad consent, controlled data sharing; 3) Broad consent, open data sharing.
- Oversampling of minorities, younger individuals, those with less education and from rural areas allowed us to obtain the opinions of these under-represented populations regarding data access and consent.

Brief Findings:

- Participant's willingness to enroll was highest for the broad controlled biobank, although the difference in willingness to participate was not large between the 3 biobank scenarios.
- Among adults with a child <18 years, willingness to participate in a biobank (for themselves) was much higher than their willingness to enroll their child<18 in a biobank for all 3 biobank scenarios.
- Among adults with a child <18 years, perceived benefits of biobank participation were lower, and concerns about biobank participation were higher, for their child vs. for themselves, which may explain the lower willingness to enroll their child in a biobank compared to themselves.

CERC SURVEY: RESULTS

Willingness to Participate: All Respondents

N (%, CI)	Tiered controlled N=4224	Broad controlled N=4405	Broad open N=4371
No	487 (12, 10-15)	513 (12, 11-15)	611 (15, 13-18)
Not sure	853 (22, 19-24)	853 (20, 17-22)	913 (21, 18-24)
Yes	2758 (66, 62-71)	2880 (68, 64-71)	2702 (64, 59-68)

Comparison	X^2_{df}, p-value
Tiered controlled vs Broad controlled vs Broad open	$X^2_2 = 7.1$, $p = 0.029$
Broad controlled vs Broad open	$X^2_1 = 6.9$, $p = 0.009$
Broad controlled vs Tiered controlled	$X^2_1 = 0.7$, $p = 0.406$

CERC SURVEY: RESULTS

Willingness to Participate: Parents compared to their child <18 years

PARENTS N (%, CI)	Tiered controlled N=1880	Broad controlled N=1897	Broad open N=1891
No	221 (13, 10-18)	219 (10, 8-13)	244 (12,10-15)
Not sure	408 (23,19-28)	388 (21, 17-25)	427 (23,19-28)
Yes	1206 (63, 57-69)	1234 (69, 63-73)	1171 (65, 59-70)

CHILD < 18yr N (%, CI)	Tiered controlled N=1880	Broad controlled N=1897	Broad open N=1891
No	358 (22, 17-27)	393 (22, 16-29)	409 (22, 18-26)
Not sure	496 (28, 24-32)	495 (26, 23-29)	525 (30, 27-33)
Yes	1014 (51, 44-57)	987 (52, 46-58)	944 (49, 43-54)

Comparison	Self (Parents) X^2_{df}, p-value	Child < 18 X^2_{df}, p-value
Tiered controlled vs Broad controlled vs Broad open	$X^2_2 = 7.6$, p = 0.022	$X^2_2 = 0.9$, p = 0.636
Broad controlled vs Broad open	$X^2_1 = 2.2$, p = 0.120	$X^2_1 = 0.7$, p = 0.397
Broad controlled vs Tiered controlled	$X^2_1 = 5.6$, p = 0.018	$X^2_1 = <0.1$, p = 0.953

CERC SURVEY: RESULTS

Information needs, concerns and benefits: Average scores for themselves (parents with a child <18) and for child

Mean (CI)	Tiered controlled	Broad controlled	Broad open	All
Informational Needs	4.03 (3.97, 4.09)	3.98 (3.93, 4.04)	4.02 (3.95, 4.08)	4.01 (3.96, 4.05)
Biobank Concerns – *Self*	3.19 (3.07, 3.30)	3.18 (3.08, 3.28)	3.19 (3.08, 3.31)	3.18 (3.10, 3.27)
Biobank Concerns – *Child*	3.66 (3.54, 3.79)	3.61 (3.51, 3.71)	3.62 (3.51, 3.72)	3.62 (3.53, 3.70)
Perceived Benefits – *Self*	3.83 (3.74, 3.91)	3.88 (3.80, 3.95)	3.86 (3.79, 3.93)	3.85 (3.79, 3.91)
Perceived Benefits – *Child*	3.67 (3.60, 3.75)	3.68 (3.59, 3.78)	3.67 (3.60, 3.74)	3.67 (3.61, 3.74)

Legend				
1 = Definitely Not	2 = Probably Not	3 = Not Sure	4 = Yes Probably	5 = Yes Definitely

CERC SURVEY *from* eMERGE II: Recently Completed & Currently In-Process Manuscript Projects

- A Systematic Literature Review of Individuals' Perspectives on Broad Consent and Data Sharing in the United States (Lead: Nanibaa' Garrison) *Published: Genet Med. 2015 Nov 19.*
- Literature Review Manuscript focusing on Privacy and Governance (Lead: Nanibaa' Garrison) *Draft in process*
- Developing a National Survey on Consent Across a National Network of Genomic Medicine Sites (Lead: Maureen Smith & Ingrid Holm) *In process; target submission date: 2/1/16*
- Cognitive Interviewing Methodology across six sites (Lead: Melanie Michelson) *Submitted to J Genetic Counseling*
- A consortium's experience with IRB approval and distribution of a large consortium survey (Lead: Jen McCormick) *In process; 1st draft under review*
- What are patients' views on consent and data sharing in biobank research? A Large Multisite Experimental Survey in the US (Lead: Saskia Sanderson) *Draft in process*
- Sampling strategy/Geocoding Manuscript (Lead: Nate Mercaldo/Jonathan Schildcrout) *Draft in process*

Collaborations: Affiliate Members

emerge network Affiliate Members

Approved

Cornell University

Member since: 2016

Membership: Clinical

U.S. Air Force

Member since: 2012

Membership: N/A

Washington University

Member since: 2016

Membership: Non-Clinical

In Progress

Marshfield Clinic

Membership: Clinical

Mount Sinai

Membership: TBD

Loyola University Chicago

Membership: Non-Clinical

MATERIALS *of* INTEREST

January 2016 Steering Committee Meeting Materials:

- https://emerge.mc.vanderbilt.edu/january-2016-steering-committee-meeting/

September 2015 Steering Committee Meeting Materials:

- https://emerge.mc.vanderbilt.edu/1694-2/

Other Previous eMERGE Meetings:

- https://emerge.mc.vanderbilt.edu/?page_id=968

eMERGE Tools:

- https://emerge.mc.vanderbilt.edu/?page_id=9

General Resources – New Investigator Manual **&** eMERGE III Master Contact List**:**

- https://emerge.mc.vanderbilt.edu/member-resources/

Project Areas:

- https://emerge.mc.vanderbilt.edu/?page_id=7

Workgroup Charters:

- https://emerge.mc.vanderbilt.edu/workgroups/

eMERGE III GENE & SNP List:

- http://emerge.mc.vanderbilt.edu/wp-content/uploads/2015/11/eMERGE_Design_102315_FINAL.xlsx

*e*MERGE III: GENE LIST

ACMG56		
ACTA2	*MSH2*	*SCN5A*
ACTC1	*MSH6*	*SDHAF2*
APC	*MUTYH*	*SDHB*
APOB	*MYBPC3*	*SDHC*
BRCA1	*MYH11*	*SDHD*
BRCA2	*MYH7*	*SMAD3*
CACNA1S	*MYL2*	*STK11*
COL3A1	*MYL3*	*TGFBR1*
DSC2	*MYLK*	*TGFBR2*
DSG2	*NF2*	*TMEM43*
DSP	*PCSK9*	*TNNI3*
FBN1	*PKP2*	*TNNT2*
GLA	*PMS2*	*TP53*
KCNH2	*PRKAG2*	*TPM1*
KCNQ1	*PTEN*	*TSC1*
LDLR	*RB1*	*TSC2*
LMNA	*RET*	*VHL*
MEN1	*RYR1*	*WT1*
MLH1	*RYR2*	

Complete SNP list can be found here.

TOP6		
ANGPTL3	*FLG*	*POLD1*
ANGPTL4	*GRM1*	*POLE*
ANK2	*GRM2*	*PON1*
APOA5	*GRM5*	*SCN1A*
APOC3	*GRM7*	*SCN9A*
APOE	*GRM8*	*SERPINA1*
ATM	*HNF1A*	*SLC25A40*
ATP1A2	*HNF1B*	*SLC2A10*
BMPR1A	*IL33*	*SMAD4*
BMPR2	*IL4*	*TCF4*
CACNA1A	*KCNE1*	*TCIRG1*
CACNA1B	*KCNJ2*	*TNF*
CACNA1C	*MC4R*	*TSLP*
CFH	*MTHFR*	*TTR*
CFTR	*NTRK1*	*TYK2*
CHEK2	*OTC*	*UMOD*
COL5A1	*PALB2*	*VDR*
CORIN	*PLTP*	