emerge network
ELECTRONIC MEDICAL RECORDS & GENOMICS

External Scientific Panel

Packet

October 7, 2016

TABLE *of* CONTENTS

LETTER *from* NHGRI

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

September 15, 2016

Dear eMERGE External Scientific Panel members,

We are now one year into eMERGE III and we are glad to let you know the eMERGE investigators have made significant progress. Specifically, they have: 1) developed the eMERGE sequencing panel which includes 109 candidate genes and 1,547 SNPs; 2) established the sequencing pipeline; 3) refined sequencing data transfer tools, software and process to transfer data for association studies and for clinical implementation; and 4) imputed the eMERGE III genome-wide genotyping data that eMERGE sites contributed to the network before the eMERGE III awards (pre-e3) and merged the imputed pre-e3 data with the previously imputed eMERGE I and II datasets. The eMERGE sequencing was launched in September 2016.

We appreciate the expertise and effort you have devoted to eMERGE in the past, and we look forward to your continued input, especially at the joint eMERGE III Steering Committee and External Scientific Panel meeting on October 6-7, 2016 at the Hilton Washington DC/Rockville Hotel & Executive Meeting Center, 1750 Rockville Pike, Rockville, MD 20852-1699.

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.

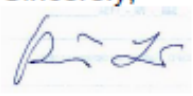
Within the booklet you will find the following important materials:

- Agenda for ESP Conference
- Investigators' response to ESP recommendations from February 2016
- eMERGE sequencing update
- Network data management update
- eMERGE Workgroup progress report

If you have any questions, or would like more information, please do not hesitate to contact us or the CC program staff.

We look forward to seeing you at the meeting.

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

*e*MERGE STEERING COMMITTEE MEETING AGENDA

for Thursday, October 6, 2016

Hilton Washington DC/Rockville Hotel & Executive Meeting Center *in the* Plaza Ballroom

Toll-Free: 1 877-309-2075 ● **Long Distance:** +1 (510) 365-3231 ● **Access Code:** 314-121-494
https://attendee.gototraining.com/r/6735906576759640322

7:30-8:30 a.m.	**Networking Breakfast – *Plaza Foyer***
8:30-8:40 a.m.	Welcome, Opening Remarks, General Updates – Rongling Li (NIH/NHGRI)
8:40-8:50 a.m.	Announcements, Opening Remarks – Rex Chisholm (SC Chair, Northwestern)
8:50-9:50 a.m.	The eMERGE sequencing panel: CSG updates on performance, content curation and sequencing progress – Richard Gibbs (BCM), Niall Lennon (Partners/Broad), Birgit Funke (Partners/Broad), and Larry Babb (GeneInsight/Sunquest)
9:50-10:00 a.m.	**Networking Break – *Plaza Foyer***
10:00-10:20 a.m.	Prevalence and Clinical Implications of Genetic Variants Associated with Familial Hypercholesterolemia in a Large Clinical Population – Marc Williams (Geisinger)
10:20-12:20 p.m.	Workgroup Breakout Session #1 • Outcomes – *Plaza Ballroom* • Clinical Annotation – *Jefferson* • Phenotyping – *Adams*
12:20-12:40 p.m.	**Working Lunch – *Plaza Foyer***
12:40-1:00 p.m.	Variability in Assigning Pathogenicity to Incidental Findings: Insights from LDLR Sequence Linked to the Electronic Health Record in 1013 Individuals – Maya Safarova (Mayo)
1:00-1:20 p.m.	PMI Network Update – Josh Denny (Vanderbilt)
1:20-1:40 p.m.	EHRI Infobutton Subgroup: DocUBuild Platform – Luke Rasmussen (Northwestern)
1:40-3:40 p.m.	Workgroup Breakout Session #2 • RoR/ELSI – *Plaza Ballroom* • Genomics – *Jefferson* • EHR Integration – *Adams*
3:40-3:50 p.m.	**Networking Break – *Plaza Foyer***
3:50-4:10 p.m.	Estimate of disease heritability using 4.7 million familial relationships inferred from electronic health records – Nicholas Tatonetti (Columbia)
4:10-4:30 p.m.	Adolescent and Parent Choices about Return of Genomics Research Results: Development of Tools to Facilitate Decision Making – Melanie Myers (CCHMC)
4:30-4:50 p.m.	Evidence of hybrid vigor in a human population from PheWAS – Todd Edwards (Vanderbilt)
4:50-5:15 p.m.	Facilitating Investigator-Initiated Grant Applications in Genomic Medicine – Teri Manolio (NIH/NHGRI)
5:15 p.m.	**Meeting Adjourned**
5:15-6:00 p.m.	Leadership and Workgroup Chair(s) Meeting – *Jefferson*
5:15-6:00 p.m.	PGx Meeting *(for members of the PGx Workgroup) – Adams*

*e*MERGE STEERING COMMITTEE MEETING AGENDA

for Friday, October 7, 2016

Hilton Washington DC/Rockville Hotel & Executive Meeting Center *in the* Plaza Ballroom

Toll-Free: 1 877-309-2075 ● **Long Distance:** +1 (510) 365-3231 ● **Access Code:** 314-121-494
https://attendee.gototraining.com/r/6735906576759640322

Joint Session with the External Scientific Panel (ESP)

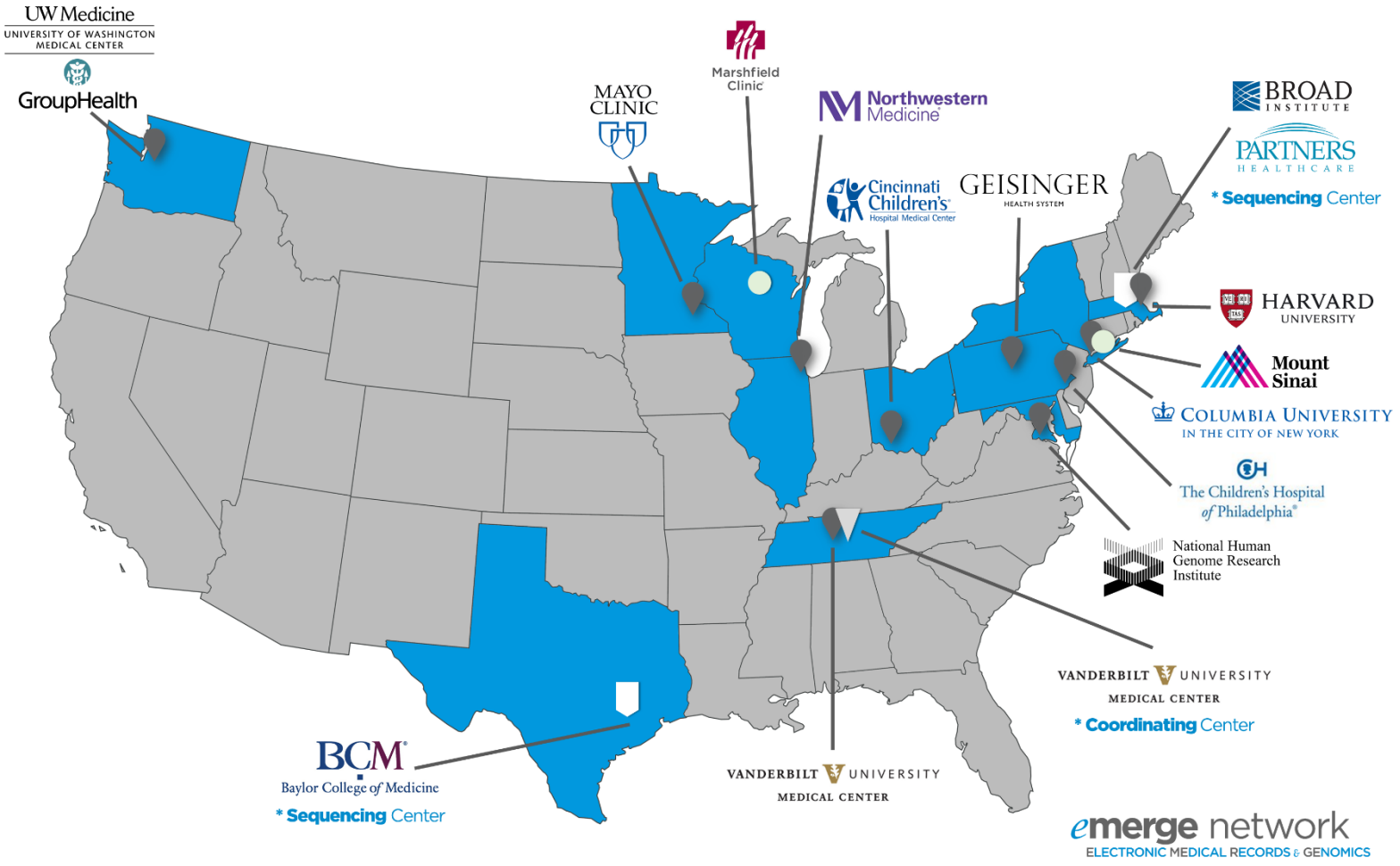
7:00-8:00 a.m.	**Networking Breakfast – *Plaza Foyer***
7:30-8:00 a.m.	Executive Session with ESP *(Location TBD)*
8:00-8:15 a.m.	Opening Remarks – Teri Manolio (NIH/NHGRI) & Rongling Li (NIH/NHGRI)
8:15-8:25 a.m.	Comments from ESP Interim Chair – Eta Berner (UAB)
8:25-8:45 a.m.	eMERGE Network Overview: Priorities and Goals; Review of Progress of Prior ESP Recommendations & Best Practices Topics – Rex Chisholm (SC Chair, Northwestern)
8:45-9:25 a.m.	The eMERGE sequencing panel: CSG updates on performance, content curation and sequencing progress – Richard Gibbs (BCM), Birgit Funke (Partners/Broad) and Sandy Aronson (Partners/Harvard)
9:25-9:45 a.m.	CERC Survey Project Update & Discussion – Ingrid Holm (BCH) & Maureen Smith (Northwestern)
9:45-10:00 a.m.	**Networking Break – *Plaza Foyer***
10:00-10:30 a.m.	Clinical Annotation Workgroup Report – Gail Jarvik (GHC/UW)
10:30-11:00 a.m.	Genomics Workgroup Report – Megan Roy-Puckelwartz (Northwestern), Patrick Sleiman (CHOP) & David Crosslin (GHC/UW/CC)
11:00-11:30 a.m.	Phenotyping Workgroup Report – Josh Denny (Vanderbilt) & George Hripcsak (Columbia)
11:30-12:00 p.m.	eMERGE PGx Project Update & Discussion – Laura Rasmussen-Torvik (Northwestern)
12:00-12:30 p.m.	**Working Lunch – *Plaza Foyer***
12:30-1:00 p.m.	EHR Integration Workgroup Report – Sandy Aronson (Partners/Harvard) & Casey Overby (Geisinger/JHU)
1:00-1:30 p.m.	RoR/ELSI Workgroup Report – Ingrid Holm (BCH) & Iftikhar Kullo (Mayo)
1:30-2:00 p.m.	Outcomes Workgroup Report – Hakon Hakonarson (CHOP), Josh Peterson (Vanderbilt) & Marc Williams (Geisinger)
2:00-2:15 p.m.	**Networking Break – *Plaza Foyer***
2:15-2:45 p.m.	Input/Feedback from ESP, General Discussion
2:45-3:00 p.m.	Closing Remarks
3:00 p.m.	**Meeting Adjourned**
3:00-3:30 p.m.	Executive Session with ESP *(Location TBD)*

Please address any inquiries to Kayla Howell at the eMERGE Coordinating Center - kayla.m.howell@Vanderbilt.edu

NETWORK OVERVIEW

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.



emerge Workgroups

- Clinical Annotation
- EHR Integration
- Genomics
- Outcomes
- Phenotyping
- Return of Results / ELSI

ESP RECOMMENDATIONS *from* FEBRUARY 2016

1) The ESP recommended that the network study the social and ethical issues involved in the process of making scientific decisions about variant annotation and reporting clinically reportable variants (e.g. who is involved when decisions are made about returning results, etc). This study could lead to a network publication.

- The RoR/ELSI Workgroup is addressing this recommendation through a series of Network-wide projects, including:
  - Participant Survey Project – Sites have collaborated and generated a set of shared survey metrics for the baseline (pre-disclosure) and post-disclosure surveys to address participant perspectives on RoR.
  - Impact of Return of Genomic Results on Health Care Providers – Sites are collaborating to develop and test a survey assessing the impact of returnable results on healthcare providers.
  - Developing an approach to the family members of a participant with returnable results: defining measurable outcomes of cascade screening of at risk family members – The RoR/ELSI Workgroup is collaborating with the Outcomes Workgroup to address optimal practices for family communication or regarding measures of successful outcomes of such return.
  - IRB Perspectives Project – gathering experiences at sites with IRB interactions around return of unsolicited genetic results

2) The ESP recommended that both sequencing centers issue clinical reports for all variants to the sites, pending resolution of budgetary issues.

- NHGRI staff and eMERGE Investigators acknowledge that this recommendation is outside of budgetary feasibility for the project, and that the consensus clinical reports coming from sequencing centers will only contain Pathogenic and Likely Pathogenic variants. Individual sites can request specific variants that do not fall into LP / P classification but are considered important for work at that site. Further, the sequencing data will contain all variants and can be analyzed outside the clinical reporting workflow. Some sites are planning to elevate these variants to clinical operations as new evidence changes the understanding of their role in human health.
- DNANexus, hosted by Baylor, will be used to streamline customized tool development and genomic data analysis and usage across the network with regarding sequencing data and analysis. Sequencing data for research use from both sequencing centers are uploaded to DNANexus. GeneInsight, hosted by Sunquest, is a system that allows transfer of meta-data, clinical information, and reporting associated with gene sequencing results. Each sequencing center will issue clinical reports to the sites they sequence.

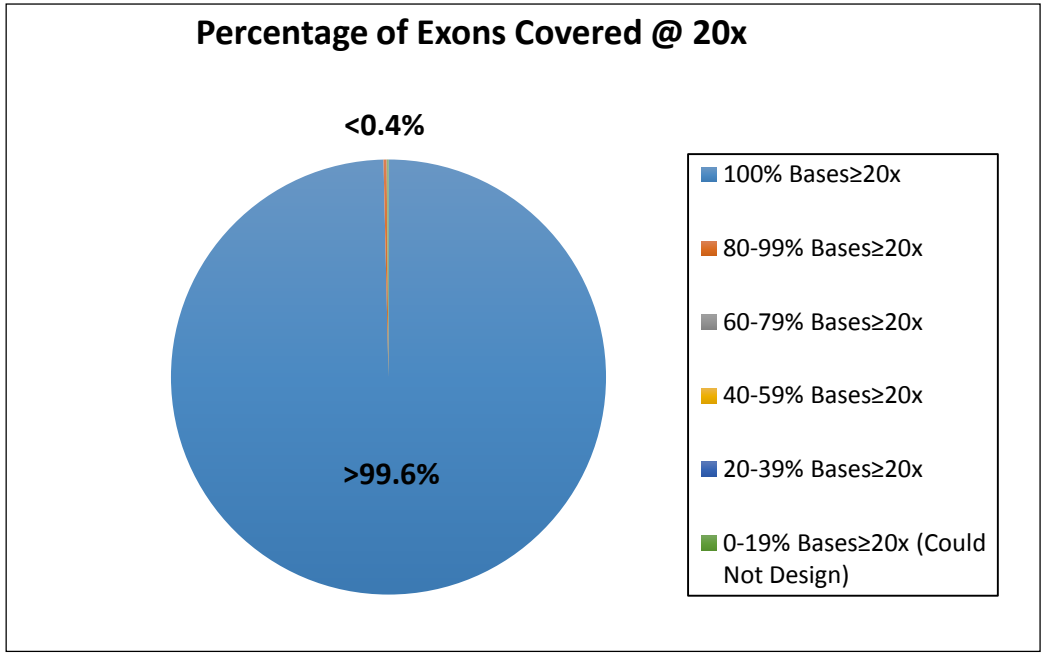
*e*MERGE-Seq OVERVIEW & CLINICAL REPORTING

SEQUENCING: Assay Performance, Baylor College Medicine

Version 2 Reagent Performance

Design Type	Genes 100% ≥ 20x (n = 109)	SNPs 100% ≥ 20x (n = 1551)	# of bps < 20X in >10% of samples
eMERGE 1 Only	99	1541	1332
1:1 (eMERGE 1:Spike In)	103	1545	720
1:2 (eMERGE 1:Spike In)	103	1546	475

eMERGE + Spike In – Genes With Regions < 20x	
PKP2	NM_004572_cds_8
COL5A1	NM_001278074_cds_0, NM_000093_cds_0
CACNA1B	NM_001243812_cds_18, NM_000718_cds_18
TGFBR1	NM_001130916_cds_0 , NM_001306210_cds_0, NM_004612_cds_0
RYR1	NM_000540_cds_90, NM_001042723_cds_89
APOB	NM_000384_cds_28
SDHD*	NM_001276506_cds_3
CHEK2*	NM_001005735_cds_13



*Regions Not Able To Design

Genes in **Red** are Low Covered at Both Sequencing Centers

Exon numbering (e.g.: NM_001278074_cds_0): first exons in RefSeq transcripts are annotated as “0”

SEQUENCING: Low Coverage Details, Baylor College Medicine

eMERGE + Spike In – Genes With Regions < 20x		1:2 (eMERGE 1:Spike In)				
Gene	**Annotation**	**Chr**	**Start**	**End**	**Length**	**ClinVar**
PKP2	NM_004572_cds_8	12	32996115	32996133	19	None
COL5A1	NM_001278074_cds_0, NM_000093_cds_0	9	137534130	137534142	13	None
CACNA1B	NM_001243812_cds_18, NM_000718_cds_18	9	140918177	140918185	9	None
		9	140917896	140917898	3	None
		9	140917899	140917937	39	None
TGFBR1	NM_001130916_cds_0 , NM_001306210_cds_0, NM_004612_cds_0	9	101867546	101867584	39	rs863223805 - Benign rs11466445 - Benign
RYR1	NM_000540_cds_90, NM_001042723_cds_89	19	39055819	39055904	86	rs539194350 - Uncertain significance rs193922855 - Uncertain significance rs587784372 - Uncertain significance rs193922854 - Uncertain significance rs796065337 - Uncertain significance rs193922853 - Uncertain significance rs794728692 - Uncertain significance
APOB	NM_000384_cds_28	2	21266811	21266817	7	None
SDHD	NM_001276506_cds_3	11	111963803	111963921	119	None
CHEK2	NM_001005735_cds_13	22	29126407	29126536	130	None

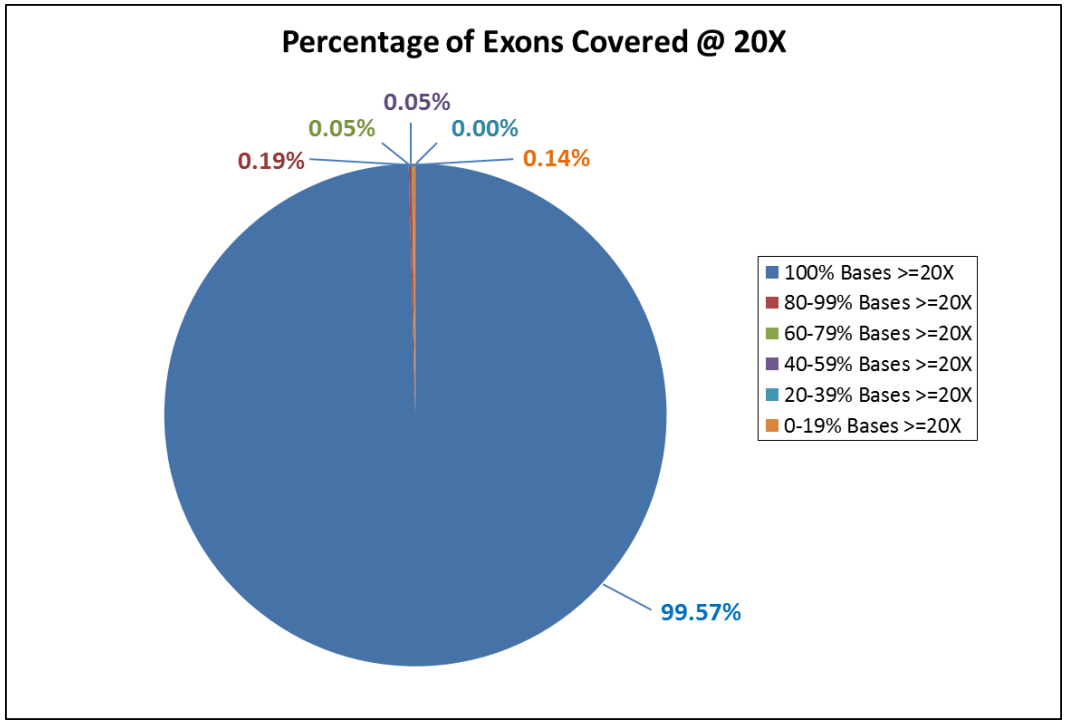
We were not able to design for the two regions in **green**.

SEQUENCING: Assay Performance, Partners/Broad

Version 2 Reagent Performance

Design Type	Genes 100% ≥ 20x (n = 109)	SNPs 100% ≥ 20x (n = 1551)	# of bps < 20X in >10% of samples
V1	86	1532	3364
V2 (eMERGE v1 +Spike In)	101	1547	710

eMERGE + Spike In – Genes With Regions < 20x	
CACNA1A	NM_023035_cds_6
CACNA1B	NM_000718_cds_0 and NM_001243812_cds_18
KCNH2	NM_000238_cds_11
KCNQ1	NM_000218_cds_0
PRKAG2	NM_024429_cds_3
PTEN	NM_001304717_cds_0
RYR1	NM_000540_cds_90
TGFBR1	NM_001130916_cds_0



Exon numbering (e.g. NM_001304717_cds_0): first exons in RefSeq transcripts are annotated as "0"

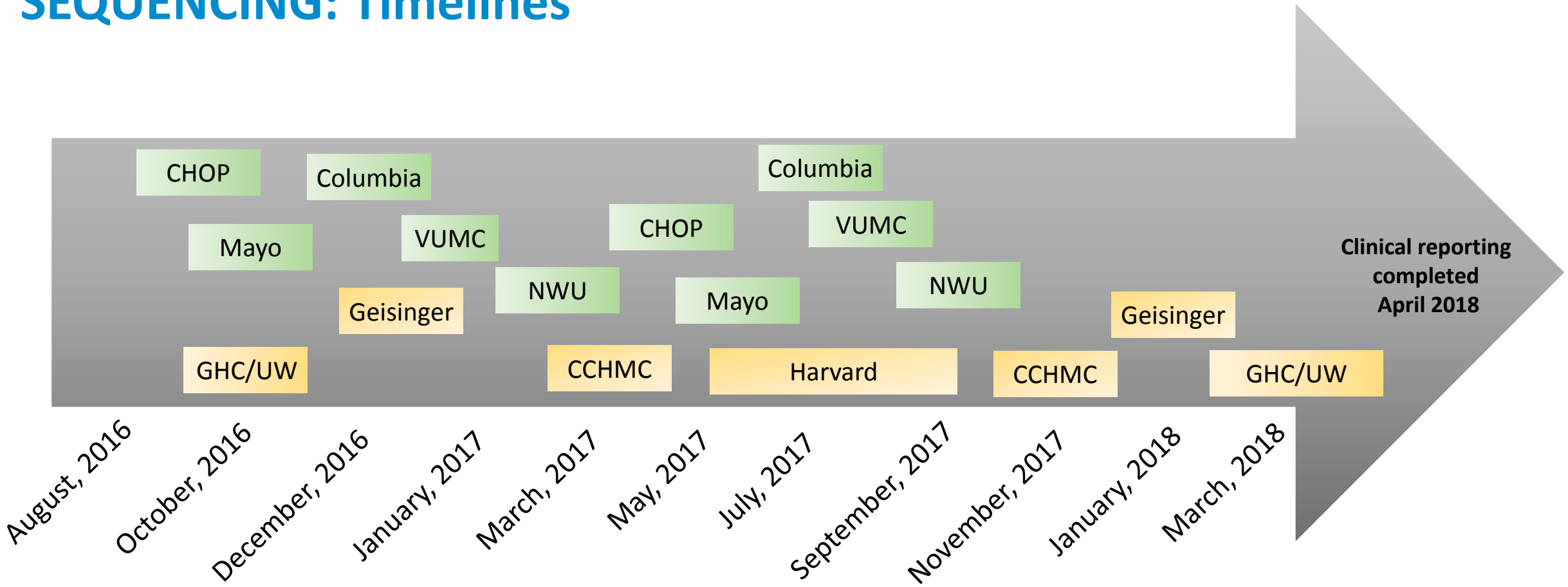
SEQUENCING: Low Coverage Details, Partners/Broad

eMERGE Version 2 Genes With Regions < 20x		eMERGE Version 2 Bases < 20X				ClinVar clinical significance and review status (see below for key) (Note: classifications not independently verified by Partners/Broad)
	Annotation	**Chr**	**Start**	**End**	**Length**	
CACNA1A	NM_023035_cds_6	19	13323455	13323554	100	None
CACNA1B	NM_000718_cds_0	9	140772386	140772438	53	None
	NM_001243812_cds_18	9	140917918	140917956	39	None
		9	140918057	140918064	8	None
KCNH2	NM_000238_cds_11	7	150655147	150655169	23	**rs199472884 – Pathogenic/Likely Pathogenic (2 stars)** 1 Pathogenic (0 stars)
KCNQ1	NM_000218_cds_0	11	2466329	2466496	168	**rs199473441, rs397508096, rs794728563, rs794728544 – Pathogenic (1 star)** 2 Pathogenic (0 stars) 3 Conflicting 4 Uncertain significance
PRKAG2	NM_024429_cds_3	7	151262432	151262453	22	None
		7	151262455	151262469	15	None
PTEN	NM_001304717_cds_0	10	89623707	89623712	6	None
RYR1	NM_000540_cds_90	19	39055829	39056003	175	2 Pathogenic (0 stars) 6 Uncertain significance 1 Likely Benign/Benign
TGFBR1	NM_001130916_cds_0	9	101867488	101867584	97	1 Uncertain significance 3 Likely Benign 3 Benign

ClinVar review status legend

0 stars = no assertion criteria provided or no interpretation
1 star = assertion criteria provided by one submitter or multiple conflicting submissions (with assertion criteria)
2 stars = 2 or more submissions (with assertion criteria) with the same interpretation
3 stars = Expert Panel approved

SEQUENCING: Timelines



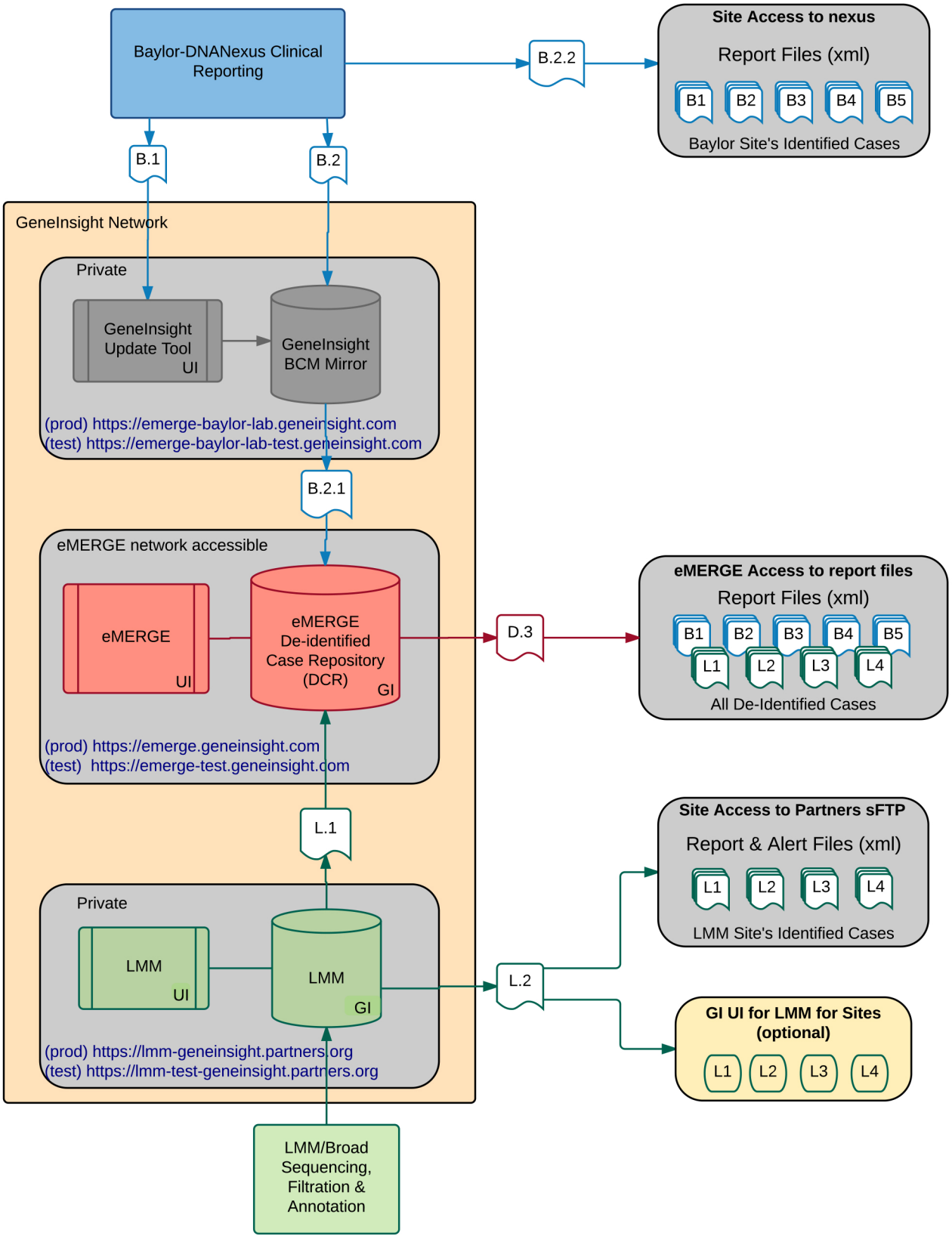
Batch One

Batch Two

Partners/Broad

Baylor College Medicine

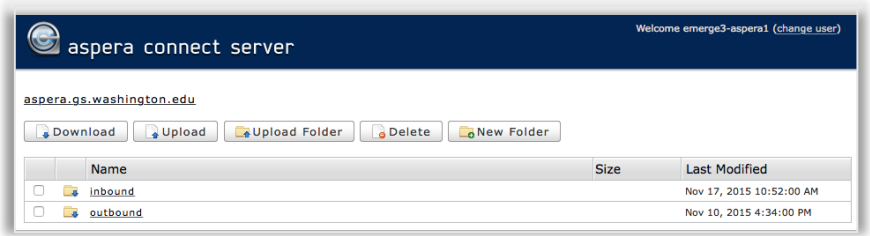
CLINICAL REPORTING: Sequencing Data Integration Overview



Legend
B.1 DNAnexus generates and manually uploads the GI variant upload file for all variants contained in the set of cases to be uploaded in step B.2.
B.2 The set of cases are uploaded (individually) using 2 web service operations to both load and finalize the cases, which trigger the push to the de-identified case repository.
B.2.1 Finalized reports are de-identified and sent to central repository along with all associated variant knowledge.
B.2.2 Generate identified report message XML document and copy to nexus site.
L.1 Finalized reports are de-identified and sent to central repository along with all associated variant knowledge.
L.2 Generate identified report message XML document and copy to site-specific sFTP location
(optionally) identified cases are sent to GI Clinic repositories for UI searching & accessibility.
D.3 As reports are sent to the DCR, generate and store de-identified report message XML document for all emerge member access.

NETWORK DATA RESOURCES & MANAGEMENT

Aspera High-Speed File Transfer Software *at the* University *of* Washington



- At the beginning of eMERGE 3, the Coordinating Center invested in an Aspera server in the Department of Genome Sciences.
- This Aspera server has a dedicated 10Gb/sec Science DMZ/I2 network link for data dissemination and acquisition
- To date, eMERGE sites have deposited **4.96 terabytes** of data and have downloaded **1.97 terabytes**.
- DNANexus staff also have an account, so all data available to the Commons.
- All legacy and analysis data are available for download including the following:
  - All eMERGE I, 2, and pre-3 array data
  - All imputed data, SHAPE-IT pre-phased files, individual files
  - Multisample PGRNseq VCF (will soon be updated with new file)
  - We will also store eMERGEseq data for the duration of the project

GENOMIC DATA: PGRNseq Multisample Calling

Re-called Multisample VCF for ~9000 PGRNseq Cohort

- Mapped to genome reference hs37d5.fa
- Realign INDELs using IndelRealigner (GATK)
- Calculate recalibration matrix using BaseRecalibrator (GATK)
- Generate gVCF files with HaplotypeCaller (GATK)
- Calling multisample VCF using GenotypeGVCFs (GATK)
- Plan to add additional annotation to SPHINX to provide more bioinformatic data to users
- Will also provide an additional summary of frequency in the Asian ancestry cohort in addition African and European ancestries
- Drs. Gordon and Crosslin will lead Network-wide project analyzing the data. A manuscript concept sheet has been circulated

GENOMIC DATA: Imputation

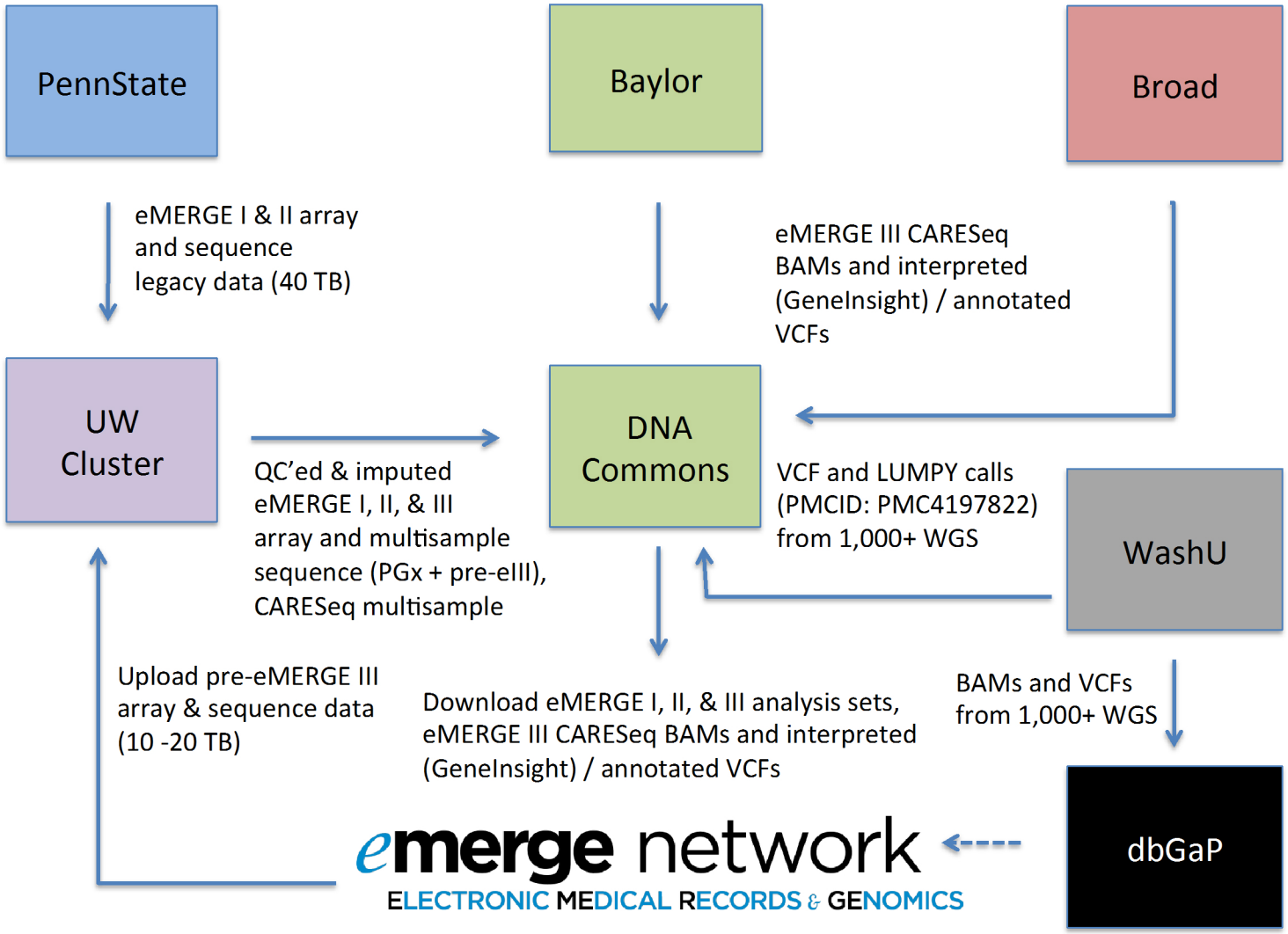
The pre-3 array data has been imputed against the 1000 Genome (Phase 3) and merged with the eMERGE legacy data, producing an analysis ready set of ~80,000 participants

We are in the process of QC'ing the merged set

Will produce a principal component and identity by descent matrix of the merged set

The individual imputed site data is available for download on DNANexus and Aspera, and the merged data will be available shortly

GENOMIC DATA: Assess New Datasets *from* eMERGE III Partners



PHENOTYPES: Development & Implementation

Phase 1

TOTAL PHENOTYPES

14

Implemented across:

GWAS Cohort ***(n = 18,663)***

Phase II

TOTAL PHENOTYPES

28

Implemented across:

GWAS Cohort ***(n = 55,029)*** **OR**

PGx Cohort ***(n = 9,015)***

Phase III

TOTAL PHENOTYPES

42

To be **Implemented across:**

GWAS Cohort ***(n = ~80k)*** **AND/OR**

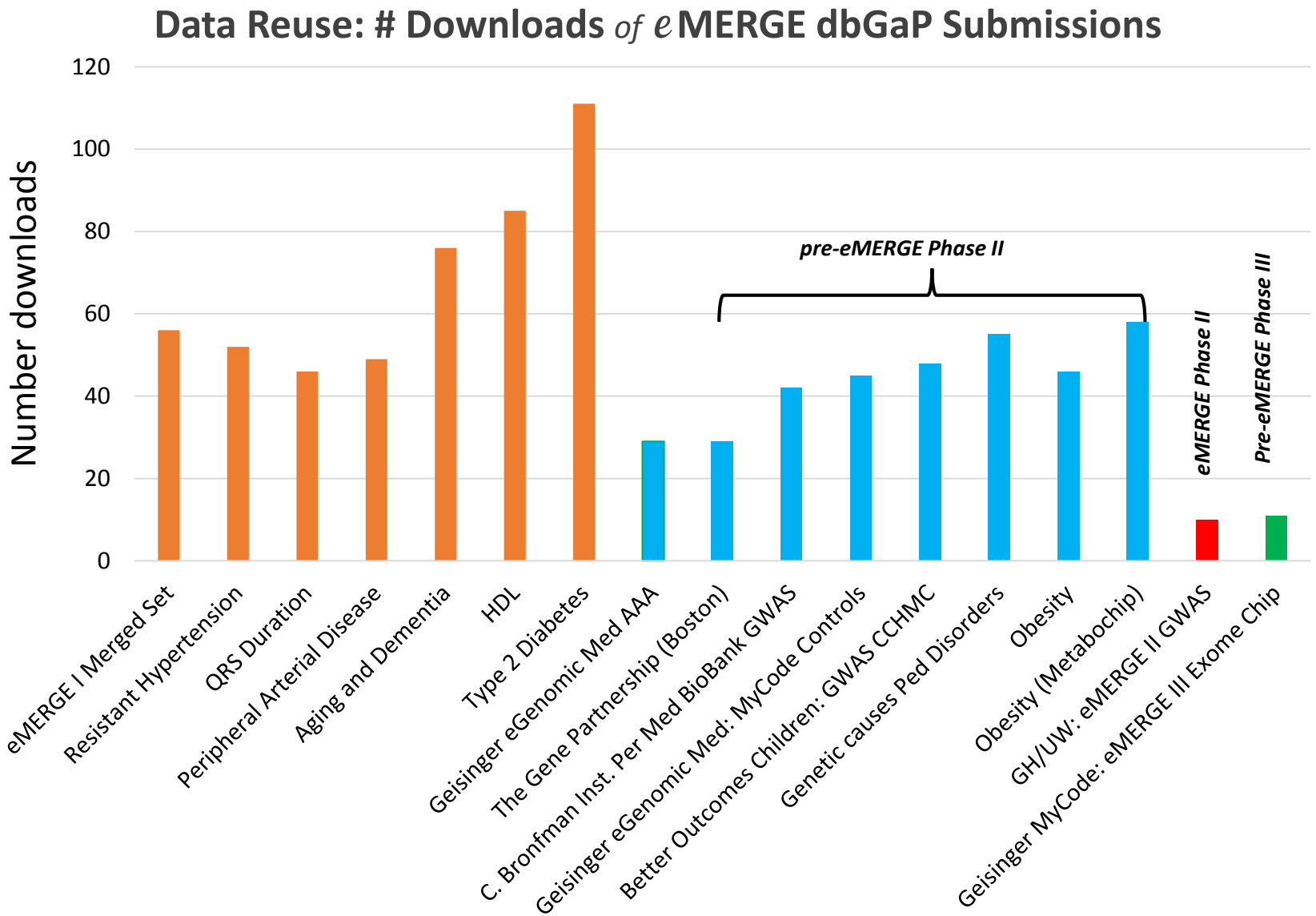
PGx Cohort ***(n = 9,015)*** **AND/OR**

e III Sequencing Cohort ***(n = ~25k)***

84

Network Phenotypes Developed & Implemented Across ~114k Participants by **Summer 2018**

IMPACT: dbGaP & Website Analytics



eMERGE Website

Average usage past 6 months

- 63% new visitors
- 1920 sessions/month
- 1246 users/month
- Views from 97 countries

PheKB Website

Average usage past 6 months

- 54% new visitors
- 830 sessions/month
- 516 users/month
- Views from 37 countries

dbGaP Update: eMERGE Merged II set now accessible via the dbGap site (Cohort size: 55,029)

*e*MERGE WORKGROUP PROGRESS

Clinical Annotation

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

EHR Integration

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

Genomics

Co-Chairs: Megan Roy-Puckelwartz (NU), Patrick Sleiman (CHOP) & David Crosslin (GHC/UW)

Outcomes

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

Phenotyping

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

PGx

Chair: Laura Rasmussen-Torvik (Northwestern)

RoR/ELSI

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

CERC Survey

Co-Chairs: Ingrid Holm (BCH) & Maureen Smith (Northwestern)

*e*MERGE CLINICAL ANNOTATION WORKGROUP: Status & Accomplishments

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

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.

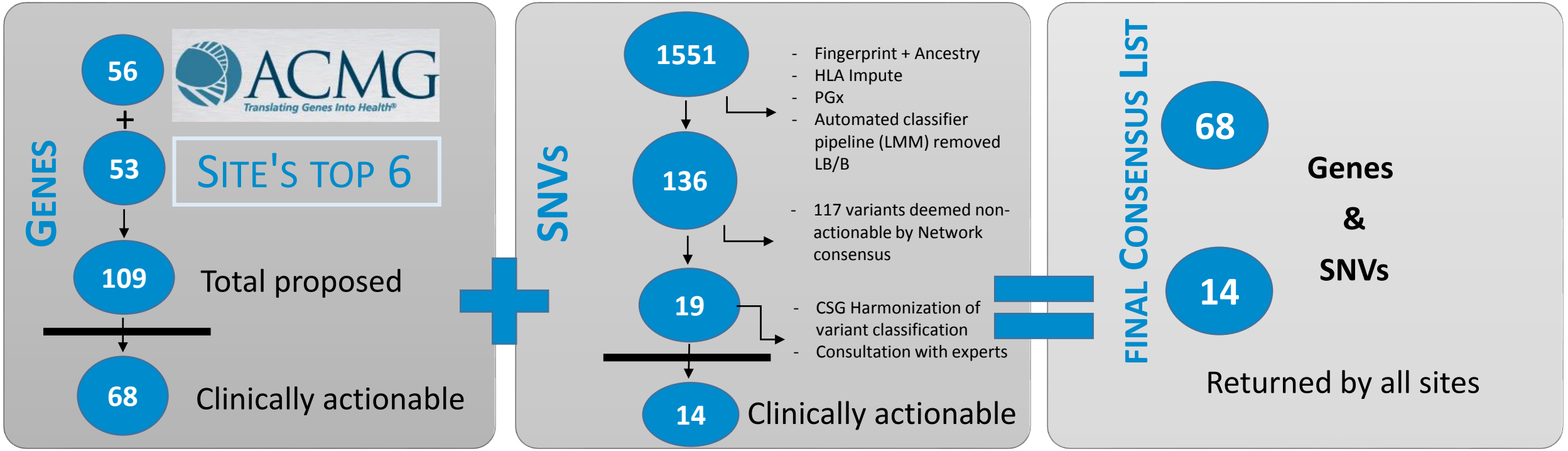
1. Apply the ClinGen approach to gene-disease validity assessment to all genes on the eMERGE gene panel (including SNP genes), defining each associated condition and the strength of evidence
2. Develop consistency in variant interpretation approaches
    a. Compare variant interpretations from CSGs and 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

eMERGE CLINICAL ANNOTATION WORKGROUP: Status & Accomplishments

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

	'15		2016												2017												
	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J
Apply ClinGen Approach to Gene-Disease validity	•	•	•	•	•	Complete																					
Develop Variant Interpretation Consistency																											
• **Collect and compare existing variant interpretations from sites**	•	•	•	•		Complete																					
• **Build consensus on existing variants and process for novel variants**	Existing variants complete. Interim knowledge exchange plan complete. More efficient solutions under discussion.														•	•	•	•	•	•	•	•	•	•	•	•	•
Consensus on Returnability																											
• **Required gene and variant evidence levels**	•	•	•	•	•		Complete																				
• **Other factors**	•	•	•	•	•		Complete																				
ClinVar Submissions	LMM up to date, Baylor plans to submit Jan 2017		•						•						•						•						•
Interface with ROR/ELSI on Return results and barriers								•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•

CLINICAL REPORTING: Overview *of* Consensus Lists *for* eMERGE III



Comprehensive List of Genes and SNVs on Next Two Slides

CLINICAL REPORTING: Gene Consensus List *for* *e*MERGE III

Consensus List *for which* **Pathogenic** *or* **Likely Pathogenic Variants** *will be* **Returned**

Phenotype	Gene‡
Cancer susceptibility and tumor diseases	APC, BMPR1A, BRCA1, BRCA2, MEN1, MLH1, MSH2, MSH6, MUTYH[#], NF2, PALB2, PMS2, POLD1, POLE, PTEN, RB1, RET, SDHAF2, SDHB, SDHC, SDHD, SMAD4, STK11, TSC1, TSC2, TP53, VHL, WT1
Cardiac Diseases	ACTA2, ACTC1, COL3A1, COL5A1, DSC2, DSG2, DSP, FBN1, GLA[+], KCNE1[§], KCNH2, KCNJ2, KCNQ1, LMNA, MYBPC3, MYH7, MYH11, MYL2, MYL3, MYLK, PKP2, PRKAG2, RYR2, SCN5A, SMAD3, TGFBR1, TGFBR2, TMEM43, TNNI3, TNNT2, TPM1
Hypercholesterolemia	APOB*, LDLR*, PCSK9
Diabetes & Kidney Disease	HNF1A, HNF1B
Ehlers-Danlos Syndrome	COL3A1, COL5A1
Neuromuscular Diseases	CACNA1A, CACNA1S, RYR1
Ornithine Transcarbamylase (OTC) Deficiency	OTC[+]

‡ Site TOP-6 genes are indicated in blue

*semi (incomplete) dominant, [+]x-linked, [#]recessive, [§] dominant or recessive

CLINICAL REPORTING: SNV Consensus List *for e*MERGE III

Consensus List *of* Actionable Pathogenic *or* Likely Pathogenic Variants *to be* Returned*

rs#	Gene	Molecular Consequence	Associated Disease	Mode *of* Inheritance	Disease Category
rs77931234	*ACADM*	c.985A>C (p.Lys329Gln)	Medium-chain acyl-CoA dehydrogenase (MCAD) deficiency	AR	Inborn error of metabolism
rs387906225	*ALDOB*	c.360_363delCAAA (p.Asn120Lysfs)	Hereditary fructose intolerance	AR	Inborn error of metabolism
rs386834233	*BCKDHB*	c.832G>A (p.Gly278Ser)	Maple syrup urine disease	AR	Inborn error of metabolism
rs79761867	*BCKDHB*	c.548G>C (p.Arg183Pro)	Maple syrup urine disease	AR	Inborn error of metabolism
rs80338898	*FAH*	c.782C>T (p.Pro261Leu)	Tyrosinemia type I	AR	Inborn error of metabolism
rs1801175	*G6PC*	c.247C>T (p.Arg83Cys)	Glycogen storage disease type I	AR	Inborn error of metabolism
rs397509431	*CPT2*	c.1239_1240delGA (p.Lys414Thrfs)	Carnitine palmitoyltransferase II (CPT II) deficiency	AR	Inborn error of metabolism
rs113993962	*BLM*	c.2207_2212delATCTGAinsTAG ATTC (p.Tyr736Leufs)	Bloom Syndrome	AR	Cancer susceptibility
rs193922376	*MSH2*	c.942+3A>T	Lynch syndrome	AD	Cancer susceptibility
rs6467	*CYP21A2*	c.293-13C>G	21-hydroxylase deficiency	AR	Endocrinology
rs6025	*F5*	c.1601G>A (p.Arg534Gln)	factor V Leiden thrombophilia	Risk	Clotting disorder
rs1800562	*HFE*	c.845G>A (p.Cys282Tyr)	Hereditary hemochromatosis	AR	Iron storage
rs28940579	*MEFV*	c.2177T>C (p.Val726Ala)	Familial Mediterranean fever	AR	Inflammatory
rs61752717	*MEFV*	c.2080A>G (p.Met694Val)	Familial Mediterranean fever	AR	Inflammatory

*** AR and risk variants: Only bi-allelic (homozygous, or if applicable compound heterozygous) variants will be returned**

eMERGE CLINICAL ANNOTATION WORKGROUP: Status & Accomplishments

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

Harmonization of previously reported variants in eMERGE genes (Baylor – Partners/Broad)

CSGs exchanged all previously reported variants

(Path, Likely Path, VUS, Likely Benign, Benign)

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

44

- Discrepancies involving variants classified as VUS, Likely Pathogenic or Pathogenic

28

- High impact discrepancies affecting report inclusion (Path-VUS/Likely Path-VUS)

7

- ≥5 occurrences in previously sequenced probands at BCM+ LMM families in "actionable" genes

- Reassessed variants (ACMG guidelines)
- Incorporated internal CSG data/evidence

CONSENSUS

eMERGE EHRI WORKGROUP: Status & Accomplishments

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

- Primary goal is to facilitate establishment of network data flows (work in progress)
  - High level Data Flow Diagrams (including Interviews and Surveys)
  - Uniform Structured Report Data Delivery Format
  - Flow of data and knowledge into the de-identified case repository

- Infobutton subgroup established
  - Goal 1: Coordinate infobutton and information resource activities across eMERGE workgroups that are complementary to efforts of other NHGRI-funded projects
  - Goal 2: Expand the development and evaluation of a content management system to optimize the reuse of quality information on genetic screening and testing
  - 3 presenters to date:
    - June - Dr. Kristin Weizel (IGNITE)
    - July - Dr. Guilherme del Fiol (OpenInfobutton)
    - Aug - Dr. Bret Heale (ClinGen)

eMERGE EHRI WORKGROUP: Future Efforts

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

- Complete work required to establish network data flows

- Infobutton subgroup efforts:
  - Development of the content management system
  - Recruiting beta testers

- Planned papers
  - Discussion of Network Data Flows (concept sheet approved)
  - Evaluation of contextual elements for IB retrieval (concept sheet in preparation)
  - Potential Collaboration with CSER on Cost of Establishing Genetic Aware Clinical Decision Support

eMERGE GENOMICS WORKGROUP: Status & Accomplishments

Co-Chairs: Megan Roy-Puckelwartz (NU), Patrick Sleiman (CHOP) & David Crosslin (GHC/UW)

The following projects were implemented or facilitated by the Genomics Workgroup

DNANexus

- Basic infrastructure including access/permissions has been determined (implementation ongoing)
- Pipeline for data analysis determined
- Necessary tools identified (a number of tools are already available on DNANexus)
  - For tools not yet available, Genomics WG is facilitating tool development
- In-person training complete
  - Additional web-based training will be offered as needed

DataSet Availability

- Imputation complete for eMERGE III data
  - eMERGE III imputed data merged with legacy data (available to group)
  - PCA & IBD analysis complete on these data
- PGRNSeq
  - New multisample call complete and available
- Phenotype Data
  - CC gathering/collating phenotype data for imputed data set and additional legacy data
    - These data will include basic demographic information, available to all members

*e*MERGE OUTCOMES WORKGROUP: Status & Accomplishments

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

Overall Progress Preparing *for* Outcomes Studies

Completed — Map Universe of Possible Outcomes for Returned Genes

Outcomes Map
https://goo.gl/xE6o7Y

Completed — Prioritize Gene(s)-Outcomes Pairs

Phenotype Assignments
https://goo.gl/2m33nw

In-Process — Define Specific Outcomes Projects

PHENOTYPE	SITE LEAD*	# SITES CONTRIBUTING DATA
Arrhythmias	Vanderbilt	7
Breast Cancer	Columbia	5
Chronic Kidney Disease	Columbia	5
CFTR	Geisinger	4
EDS, Classical	CCHMC	5
EDS, Vascular	CCHMC	5
Familial Hypercholesterolemia*	Mayo (adult); Geisinger (peds)	8
HF / Cardiomyopathy	Northwestern	8
OTC	Geisinger	5
Polyps	UW	7
Tuberous Sclerosis	Geisinger	6
Aortic Dilatation	Mayo	8

*Harvard also leading outcome study associated with hyperlipidemia variants

eMERGE OUTCOMES WORKGROUP: Future Efforts

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

Outcomes WG Manuscript

- Publish targeted population, phenotypes, outcomes (Refinement of Outcomes Map)
- Library of outcome assessment algorithms for top phenotypes
- Categorization framework for key population stratifications

Cohort profiling to estimate outcome event rates

- Expected variant rates
- Baseline rate of primary outcomes
- Estimate size of stratification

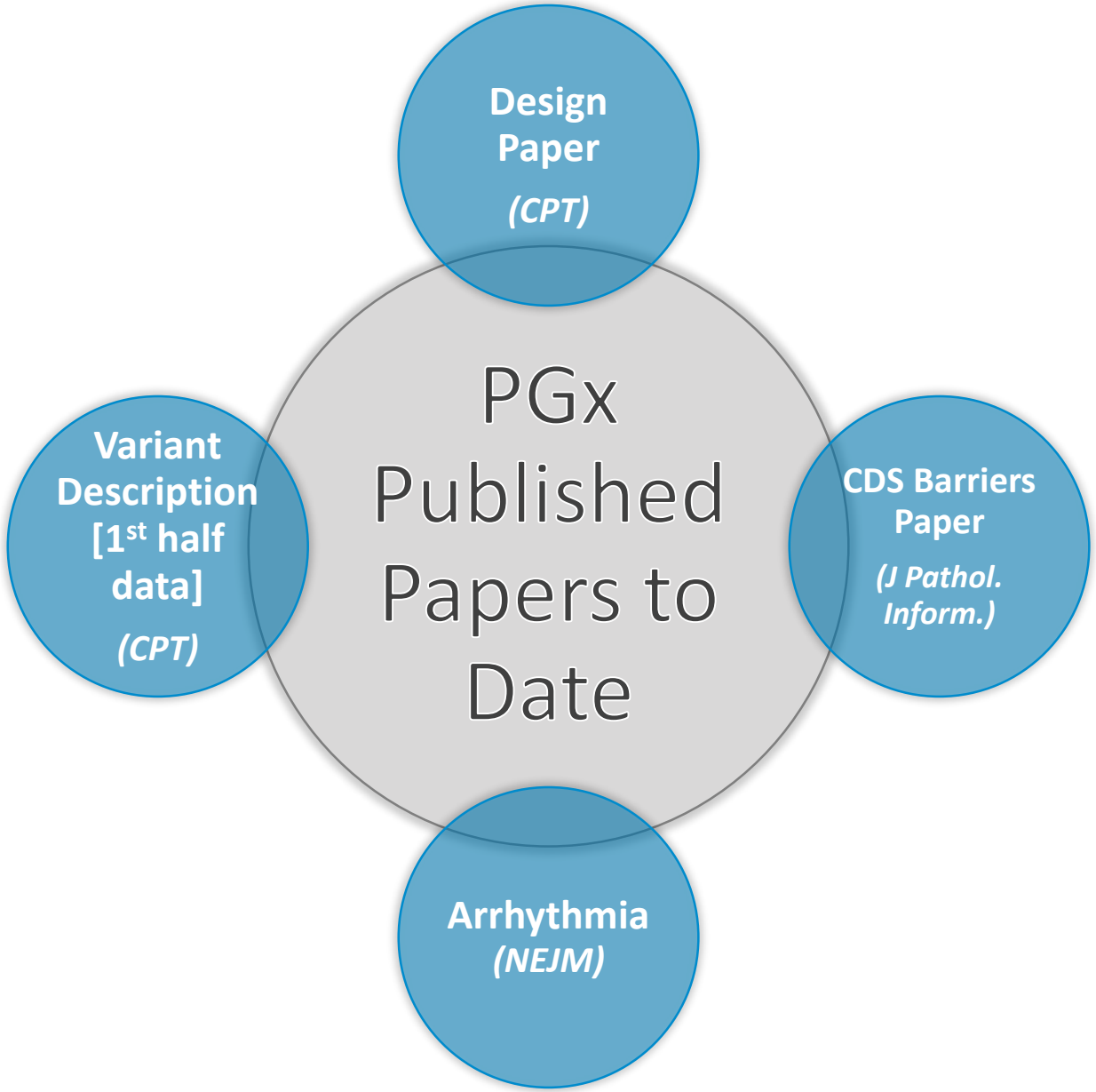
Other

- Established **Familial Implications of ROR Subgroup** (Janet Williams from Geisinger will lead)
  - Purpose is to evaluate outcomes related to cascade genetic testing that results from identification of a proband from eMERGE-seq platform
  - Will present some challenges given that most family members will not be enrolled in eMERGE
- Organized Economic evaluation
  - Josh Peterson, David Veenstra and Marc Williams leading

eMERGE PGx WORKGRUP: Update

Chair: Laura Rasmussen-Torvik (Northwestern)

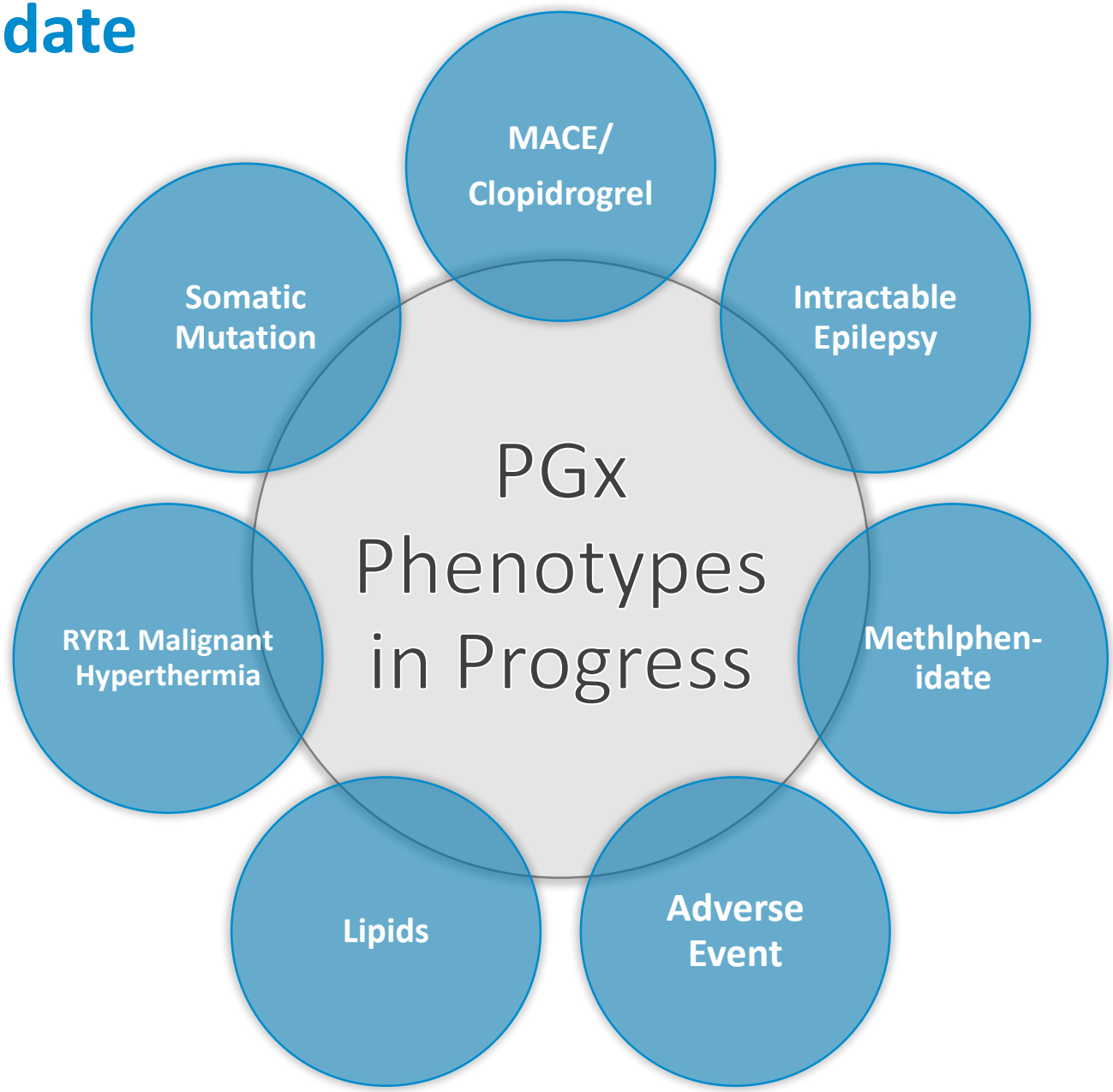
- Promoted Integration of PGx phenotypes into the Phenotyping WG pipeline.
  - Process developed to review and schedule phenotypes.
- Promoted and furthered PGx-specific papers
  - Concordance paper (in draft)
  - ROR paper (data collection form released)
  - Somatic mutation paper (additional data request disseminated)
  - CDS response paper (JAMIA abstract)
- Helped to initiate HLA-B calling on platform
- Coordinated with all other working groups to efficiently continue project which overlaps many e3 domains



eMERGE PGx WORKGRUP: Update

Chair: Laura Rasmussen-Torvik (Northwestern)

- Coordinate summary analyses of network-wide PGRNseq data (full set)
  - Comparing PGRNseq-derived haplotypes with results from orthogonal clinical platforms
  - Generating precise allele frequency information for PGx variants of known effect
  - PCA-based ancestry analysis to refine association testing among these data
- Discuss best practices for use of network-wide recalled and annotated PGRNseq dataset
  - Provide feedback on proposed PGx projects utilizing this dataset
- Further characterize outcomes captured outside the EHR as part of PGx (i.e. local surveys, interviews)
- Working with phenotyping group to refine and prioritize phenotypes for PGx analyses in the context of other e3 priorities



*e*MERGE PHENOTYPING WORKGROUP: Status & Accomplishments

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

Phenotype Development & Implementation

42 **Phase III Network Phenotypes** proposed – 5 ready for Network implementation and/or secondary validation, others in development

Expanded implementation of **select Phase I and II Network Phenotypes** – 5 re-implemented across Phase III cohort, 7 additional in process

Data Standardization Efforts

Medical Home Definition: Defined & adopted first Network-wide definition, discussed caveats/special cases

Common Data Model Subgroup: Adopted a 2-level phenotype definition; OMOP and OMOP-on-i2b2 data model

Cardio Data Core Repository: Extracting a set of core ECG and echo variables across sites for use in developing cardio-focused phenotypes, central processing of cardiac reports encouraging but privacy challenges need to be addressed

*e*MERGE PHENOTYPING WORKGROUP: Future Efforts

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

Phenotype Development & Implementation

- Continue developing and implementing Phase III Network Phenotypes across their intended cohorts (*e*III sequencing, *e*III GWAS, + PGx)
- Complete the expanded implementation of select Phase I and II Network Phenotypes across

Data Standardization Efforts

- Medical Home Definition: Implement new definition across Phase II phenotypes
- Common Data Model Subgroup: Pilot new 2-level phenotype definition and pilot use of adopted data model
- Cardio Data Core Repository: Continue work around standardizing and centralizing the extraction and storage of key ECG and echo variables across sites

eMERGE ROR/ELSI WORKGROUP: Status & Accomplishments

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

- Monthly conference calls
- Collected data from all sites on return of results projects and plans at each site, as well as outcome measures – Completed
- Project to study the ELSI impact of ROR on patients across the eMERGE sites: Develop data collection tools to implement across sites.
  - Participant survey subgroup – weekly calls
  - Domains: Baseline – decisional conflict; Post-disclosure – Decisional regret, Privacy, Intent to share with family members, Impact of genetic findings
  - Baseline survey and post-disclosure survey questions to be included on most sites' surveys – Completed
- Project to study the ELSI impact of ROR on health care providers across the eMERGE sites – eMERGE 3 Ancillary Study Pilot Project funded by the NHGRI ELSI Branch
  - Goals: to assess the impact of disclosure of unsolicited genetic results on provider perceptions of appropriate clinical management, including both HCPs' perceptions of clinical benefit/utility, and their perception of their responsibilities in relationship to the role of other HCPs
- IRB Perspectives Project – gather experiences at sites with IRB interactions around return of unsolicited genetic results
  - Concept sheet completed; data being collected
- Approaches to Returning Clinically Actionable Results from Next Generation Sequencing Panel in a Healthy Population
  - Joint project with Vanderbilt clinical site (lead)
- Family history project - family communication supplement designed to understand how to contact family members
  - Geisinger Lead; Joint Outcomes/RoR project
- Joint meetings with the Outcomes WG to coordinate efforts across the WG – Ongoing
- Joint publication with Clinical Annotations group – eMERGE process and criteria for actionability of variants for return – Formulating concept sheet

eMERGE ROR/ELSI WORKGROUP: Future Efforts

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

- Project to study the ELSI impact of ROR on patients across the eMERGE sites
  - Complete questions for follow-up surveys
  - Sites to implement surveys and include cross-site questions
  - Analysis of data across sites to address hypotheses
- Project to study the ELSI impact of ROR on health care providers across the eMERGE sites – eMERGE 3 Ancillary Study Pilot Project funded by the NHGRI ELSI Branch
  - Subgroup (BCH, CC, Geisinger, CCHMC) to develop and pilot survey with input from RoR-ELSI work group for dissemination across eMERGE – goal is to implement across eMERGE sites
- IRB Perspectives Project – gather experiences at sites with IRB interactions around return of unsolicited genetic results
  - Gather data across sites, submit publication
- Approaches to Returning Clinically Actionable Results from Next Generation Sequencing Panel in a Healthy Population
  - Continue data gathering, submit publication – Vanderbilt lead
- Family history project - family communication supplement designed to understand how to contact family members
  - Broaden to all sites – Geisinger lead
- Develop and publish standards for ROR for eMERGE.
  - Timeline: 1-2 years
- Joint meetings with the Outcomes WG to coordinate efforts across the WG.
  - Continue to coordinate
- Joint publication with Clinical Annotations group – eMERGE process and criteria for actionability of variants for return.
  - Convene a smaller group from both Clinical Annotations and RoR-ELSI work groups to collect data and submit publication

*e*MERGE CERC SURVEY WORKGRUP: Update

Co-Chairs: Maureen Smith (Northwestern) & Ingrid Holm (BCH)

"Patient Perspectives on Broad Consent in Biobank Research in the eMERGE Network"

- 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 Hispanics, Asians, other races, and those with low education resulted in far greater representation of minorities than would have been possible with random sampling.
- Participant's willingness to participate in biobank randomized to did not significantly differ between the 3 scenarios.
- Participant's willingness to enroll their child <18 y in biobank randomized to was significantly greater for the biobank with broad consent and controlled data sharing than either of the other two scenarios.
- Willingness to participate associated with white race, higher education, high income, not being religious, trust in researchers, and little concern about privacy

*e*MERGE CERC SURVEY WORKGRUP: Update

Co-Chairs: Maureen Smith (Northwestern) & Ingrid Holm (BCH)

TITLE	FIRST AUTHOR	MANUSCRIPT #	STATUS
A Systematic Literature Review of Individuals' Perspectives on Broad Consent and Data Sharing in the United States	Nanibaa' Garrison	NT125	*Genetics in Medicine* 2016 July
Response to Patryn and Zagaja.	Nanibaa' Garrison		*Genetics in Medicine* 2016 July
Developing a National Survey on Consent Across a National Network of Genomic Medicine Sites	Maureen Smith	NT146	Under review, *BMC Medical Research Methodology*
Patients' attitudes towards consent and data sharing in biobank research: a large multi-site experimental survey in the US	Saskia Sanderson	NT167	Under review, *Science Translational Medicine*
Central vs. Local IRB Oversight: Lessons learned from an eMERGE Network multi-site survey study	Jennifer McCormick	NT153	Reformulating to respond to NIH Central IRB requirement. To be submitted to *Clinical and Translational Science*
Literature Review paper 2 – privacy & governance	Nanibaa' Garrison	NT144/145	In finishing stages. To submit to *Genetics in Medicine*
Sampling strategy/geocoding	Nate Mercaldo	NT169	Final draft in progress. To submit to *Epidemiology*
Broad consent and data sharing in biobank research: An eMERGE Network Study of Parent Perspectives	Armand Antommaria	NT181	In progress. To submit to *Pediatrics, The Journal of Pediatrics,* or *JAMA Pediatrics*
Conducting cognitive interviews to inform the development of a survey on broad consent and data sharing. A multi-site study with participants from diverse backgrounds	Melanie Myers	NT147	Reviewed by *J of Genetic Counseling*-would require extensive rewrite, selecting an alternative journal
Impact of type of institution on patients' views on consent and data sharing in biobank research	Ingrid Holm	-	Under discussion

MATERIALS *of* INTEREST

February 2016 Conference Call Meeting Materials

https://emerge.mc.vanderbilt.edu/february-2016-esp-conference-call-2/

May 2016 Steering Committee Meeting Materials

https://emerge.mc.vanderbilt.edu/past-meetings/may-2016-steering-committee-meeting/

Manuscripts *(to date)*

https://emerge.mc.vanderbilt.edu/publications/

Data Resources *(used to date)*

https://emerge.mc.vanderbilt.edu/tools/phenotype-data/

*e*MERGE Tools

eRC

https://biovu.vanderbilt.edu/EmergeRC/

PheKB

https://phekb.org/

CDSKB

https://cdskb.org/

SPHINX

https://www.emergesphinx.org/