emerge network
ELECTRONIC MEDICAL RECORDS & GENOMICS

External Scientific Panel

Packet

October 4th, 2019


National Human
Genome Research
Institute

TABLE *of* CONTENTS

DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service

National Institutes of Health
Bethesda, Maryland 20892

www.nih.gov

October 4, 2019

Dear eMERGE External Scientific Panel members,

The eMERGE Network is currently making progress on the April 2019 ESP recommendations. The Network is collecting 6-month outcome data and assessing penetrance, finalizing efforts to coordinate the delivery of genomic data to the Network, and developing and validating e-phenotypes that use natural language processing.

The Network has obtained ~68% (796/1170) of 6-month outcome data. A second data freeze, anticipated in January 2020, will complete collection of 6-month outcome results. A few sites have started collecting 12-month outcome data. A summary of results to date can be found on the "RoR tracker" and the "Outcomes and Penetrance Progress Tracker." In parallel to this effort, the Network is finalizing collection of additional phenotype and clinical data for penetrance analysis. Data collection is expected to be completed by the end of 2019.

The Network continues to address challenges related to coordinating and delivering genetic data into an EMR as reflected in "Harmonizing Clinical Sequencing and Interpretation for the eMERGE III Network" in *The American Journal of Human Genetics*. This Network-wide paper describes lessons learned about overcoming significant challenges associated with collecting, analyzing, and reporting genetic results across a disparate network. An additional key accomplishment is eMERGE's development of an XML file structure for genetic test results. The conversion of data from a PDF file to a structured format enables incorporation of genetic test results into clinical decision support. How these results impact clinical decision making was published in *Frontiers in Genetics*, "Genomic Information for Clinicians in the Electronic Health Record: Lessons Learned from ClinGen and eMERGE." Because of this success, the XML structure is being used to help define a FHIR standard for genetic test results.

The Network is also developing and validating e-phenotypes (including ones that include natural language processing) across multiple sites and disparate datasets. They leveraged the Observational Medical Outcomes Partnership (OMOP) model to translate disparate data into a consistent format. This work was highlighted in "Facilitating phenotype transfer using a common data model," in the *Journal of Biomedical Informatics*."

We appreciate the time and effort you devote to eMERGE, and we look forward to your continued input, especially at the eMERGE ESP meeting on October 4, 2019.

The eMERGE Coordinating Center (CC) has prepared this booklet in collaboration with eMERGE investigators to ensure a productive meeting. Please review these materials prior to the meeting.

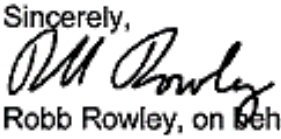
Within the booklet you will find the following important materials:

- ❑ Agenda for ESP Meeting
- ❑ Network Overview
- ❑ ESP Recommendations and Responses from April 2019
- ❑ eMERGEseq Clinical Sequencing and Return of Results
- ❑ Network Data Resources and Management
- ❑ eMERGE Workgroup Progress

- ❑ eMERGE Lessons learned panels
- ❑ Resources & Tools

If you have any questions, or would like more information, please do not hesitate to contact us or CC program staff. We look forward to seeing you October 4th, 2019.

Sincerely,



Robb Rowley, on behalf of the NHGRI eMERGE team

Robb Rowley, MD
Project Director, eMERGE
Division of Genomic Medicine, NHGRI, NIH
robb.rowley@nih.gov

CC:
Jyoti Dayal
Teri Manolio
Baergen Schultz
Ken Wiley

ESP Agenda

Friday, October 4th, 2019

Venue: **The Hilton**, 1750 Rockville Pike, Rockville, MD, 20852
Meeting Room: Roosevelt/Madison

8:00-8:30 a.m.	Breakfast (open) | Executive session with ESP (Closed session; **Library**)
8:30-8:45 a.m.	Opening remarks | Robb Rowley (NIH/NHGRI) & Teri Manolio (NIH/NHGRI)
8:45-8:55 a.m.	Comments from ESP Chair | Howard McLeod (Moffitt Cancer Center)
8:55-9:15 a.m.	eMERGE Network overview: Priorities, goals, progress and ESP recommendations | Rex Chisholm (SC Chair, Northwestern)
9:15-10:20 a.m.	Panel: Discovery in eMERGE data • GWAS, Polygenic Risk Scores & PheWAS demonstrate a polygenic determination of vesicoureteral reflux | Miguel Verbitsky (Columbia) • Phenotype Associations of LPA variants differ by race: A phenome wide association study | Benjamin Satterfield (Mayo) • The clinical utility of predicted family histories for Mendelian and genetically complex forms of disease | Scott Hebbring (Marshfield)
10:20-10:40 a.m.	**Networking Break**
10:40-11:50 a.m.	Lessons learned panel: Impact of ROR on downstream analyses • Informing and harmonizing variant interpretation| Heidi Rehm (Partners/Broad), Iftikhar Kullo (Mayo), & Adam Gordon (Northwestern) • Return of results pathways, barriers, and harmonizing across sites |Ingrid Holm (BCH) & Iftikhar Kullo (Mayo) • Impact of ROR process on Outcomes assessment | Josh Peterson (VUMC/CC) & Marc Williams (Geisinger)
11:50-12:30 p.m.	**Working Lunch**
12:30-1:10 p.m.	Interim Data Analysis: Clinical Outcomes | Hakon Hakonarson (CHOP), Josh Peterson (VUMC/CC), & Marc Williams (Geisinger)
1:10-1:35 p.m.	Status of eMERGE FHIR Specification and Implementation Project | Larry Babb (Partners/Broad)
1:35-2:25 p.m.	Input/Feedback from the ESP, general discussion
2:25-2:30 p.m.	Closing remarks | Rex Chisholm (SC Chair, Northwestern)
2:30 p.m.	**Adjourn**
2:30-3:15 p.m.	Executive session with ESP (Closed session; **Roosevelt/Madison**)

NETWORK OVERVIEW

eMERGE is a national consortium, organized by the NHGRI, that conducts discovery and clinical implementation research in genomics and genomic medicine at 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.

Over the last three phases, eMERGE has developed and validated 68 electronic phenotypic algorithms and published over 685 manuscripts. Genetic data generated and maintained by the eMERGE Network includes GWAS array, exome sequencing, whole genome sequencing, and pharmacogenomics panels. The Network has over 135,000 unique participants, and over 150,000 samples. Over 25,000 individual's DNA were sequenced during Phase III on a new sequencing panel, eMERGESeq, which contains 109 genes and 1551 SNVs of interest, and includes 68 actionable genes and 14 actionable SNVs that are being returned to participants across the Network. Clinical return to participants is complete on all but a few remaining participants. Six-month outcomes measures are being collected across sites, and initial outcomes and penetrance data are being examined. The coordinating center has focused its efforts over the last year on creating large analysis ready datasets to enable Network research as well as to collaborate externally. The 25,000 eMERGEseq dataset, the 104,000 HRC imputed GWAS dataset, and an imputed merged set of structural variant genotype data were all released in July of 2019. The coordinating center collects phenotypic information including: demographics, statin meds, lipid & autoimmune labs, BMI, ICD and CPT codes for all participants. These files are available for all our genetic data sets.

As the Network moves through its final year the focus is to document and disseminate lessons learned so as to enrich the genomic medicine community. The Network has hosted 'lessons learned' panels on Return of Results, Phenotyping, EHR integration, Harmonization of Sequencing Data, Outcomes, and Genomics. In our October meeting we will focus on the specific learnings relevant to the impact of return of results on collection of Outcomes data and analyses. The Network is continuing to publish these lessons learned throughout the remainder of Phase III. The Network is advancing knowledge surrounding the utilization and return of genetic data to providers and participants and helping to shape the landscape of clinical genetics research.

emerge Workgroups

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

ESP RECOMMENDATIONS *from* April 2019

Recommendation #1: The Network should research the discrepancy in RoR between the sites to ensure that the significant effort to date can capture 6-month and 12-month outcome data.

Discrepancies in ROR across sites were due in part to processes that varied across sites as well as differences in definitions in what constituted a given type of non-return. To resolve these discrepancies, the ROR/ELSI workgroup identified subgroups of participants that led to the inability of sites to return results and developed strict definitions of these subgroups to ensure that they were consistent across sites. Based on reasons for lack of return the following groups were defined across sites and the number in each of these categories is being tracked across the Network. (https://docs.google.com/spreadsheets/d/1e2TnRcWrweebK1AfoGbeyiLnAk0TpO2WSqIEhYVLMe4/edit#gid=0).

- Decliners: Previously consented participants who chose not to receive or consent to receive return of results.
- Non-responders/Passive withdrawal:
  - Participants previously consented to receive results however, the team could not contact the participant, the team lost contact with the participant part way through the return process, or attempted return failed (phone disconnected, letter undeliverable).
  - Participants previously consented to eMERGE sequencing but not specifically to receive results through eMERGE and the eMERGE team could not recontact the participant to consent them for return of results.
- Active withdrawal: Participant contacted the eMERGE team specifically to withdraw from the study.

The workgroup recognizes that non-responders will impact the ability to collect data to assess outcomes and penetrance and perform these analyses in the Outcomes and Clinical Annotation workgroups, respectively. Outcomes data cannot be collected for participants where RoR could not be completed. However, regardless of return of results status, for consented eMERGE patients, the group has the ability to collect retrospective EHR data relevant to the phenotype to assess penetrance. To support penetrance data collection within the subgroups of patients defined above, new 'penetrance only' data collection forms were created and these data collection efforts are running in parallel with outcomes data collection. These forms also contain data entry questions identifying the reason a given result was not returned which will help inform overall trends both within and across sites in the Network.

ESP RECOMMENDATIONS *from* April 2019

Recommendation #2: The Network members should consider applying for additional grants or supplements to analyze the data and study penetrance.

The Clinical Annotations workgroup is conducting an initial analysis on preliminary penetrance data and aims to have the final dataset completed by the end of 2019. The workgroup has added data collection forms to obtain penetrance data from both participants with results returned as well as those for which return did not take place. The group will continue to look for funding opportunities for penetrance and one potential avenue may be having junior faculty apply for K awards examining penetrance. One caveat is that penetrance studies require large populations in order to be informative, especially with phenotypes associated with lower allele frequencies. A future opportunity to collaborate with the *All of Us* research program could allow for penetrance research at scale using the eMERGE methods. Additionally, there are several manuscripts proposed to investigate penetrance that may assist with future funding opportunities:

- NT 245: Penetrance, cancer types, and outcomes of cancers associated with germline mutations in hereditary breast cancer genes and the impact of return of results of mutations for hereditary breast cancer on medical utilization and health outcomes
- NT 311: Penetrance and outcome of pulmonary hypertension associated with germline BMPR2 mutations in a population based cohort
- NT 313: Assess penetrance of cancer among mutations carriers for hereditary breast cancer genes

ESP RECOMMENDATIONS *from* April 2019

<u>Recommendation #3</u>: Regarding the healthcare provider survey, the Network should attempt to connect with providers personally, such as conducting town halls or educational sessions to engage providers and create awareness of the eMERGE Network.

In order to maximize Health Care Provider (HCP) response to the surveys, members of the HCP group reach out to HCPs on a more personal basis through emails and phone calls. The HCP group is also personally connecting with HCP by conducting interviews with a subset of providers who have received positive and negative results in eMERGE. Connecting with providers through town halls or educational sessions to raise awareness is an excellent idea but would be difficult to achieve on a Network-wide basis for a variety of logistical and feasibility reasons. However, a number of sites have engagement and educational activities for HCPs. For example, Geisinger established a Clinical Advisory Board that meets quarterly with the study team. Through this group they have presented at Grand Rounds, have newsletter posts that are distributed broadly through their daily online update, present at departmental meetings for primary care (internal medicine, women's health, community medicine), and hosted condition specific symposia for BRCA and FH. At Northwestern they met with departments throughout their recruitment period to educate them about eMERGE and see if physicians would be willing to allow recruitment in their clinics. They also had case conferences to discuss some of their results and management recommendations. At Columbia they had sessions to engage providers and create awareness of the eMERGE study when they began the recruitment. They have given talks at the Nephrology grand rounds, Liver Transplant meeting, and for the general CUMC community about the study.

ESP RECOMMENDATIONS *from* April 2019

Recommendation #4: The EHRI workgroup should clarify what sites have integrated the XML-based genetic testing results and CDSS into the EMR. The workgroup should also work with other consortia and commercial entities that are focused on developing CDSS and use existing frameworks to help reduce effort and ensure adoption.
The EHRI Workgroup has surveyed sites to understand their processes regarding using XML files to integrate genetic results into the EMR. They also determined what types of clinical decision support are being performed at each site. Several sites have successfully implemented CDS into the EMR for eMERGE return. The group is also tracking which sites are providing CDS for ACMG59 genes. Individual sites are working with either technology vendors (including at least two sites working with Epic) to optimize their handling of genetic results. Three sites have successfully integrated the XML based CDS into the EMR and one additional site has integrated it for research purposes. The EHRI FHIR sub-group also is documenting and addressing issues experienced in the efforts to transform the eMERGE XML file to an FHIR standard.

Recommendation #5: The Network should prioritize the publication of the lessons learned manuscripts before the end of the current phase of eMERGE.
Currently we have nine published lessons learned articles, one accepted, one under review, and eleven manuscript concept sheets in development, listed below. The CC will continue to track publications, quarterly updates, and disseminate data releases to the Network in a timely manner to ensure publications can be completed prior to the end of eMERGE III. The CC will also review any manuscript concept sheets that have not made significant progress with leadership in order to determine if action should be taken. Our current lessons learned MCS span Clinical Annotation, Return of Results, Outcomes, PGx, and Phenotyping.

Clinical Annotation & Sequencing Centers

- NT244: Harmonizing clinical sequencing and interpretation for the eMERGE-3 Network (published)

EHRI

- NT 184: Empowering genomic medicine by establishing critical sequencing result data flows: the eMERGE example (published).
- NT341: Genomic information for clinicians in the Electronic Health Record: Lessons learned form ClinGen and eMERGE (minor revisions address & resubmitted)

PGx

- NT168: Healthcare provider education to support integration of pharmacogenomics in practice: the eMERGE Network experience (published).
- NT155: Pharmacogenomic clinical decision support design and multi-site process outcomes analysis in the eMERGE Network (published).

Continued on next page…

ESP RECOMMENDATIONS *from* April 2019

Recommendation #5: *Continued*

ROR

- NT181: Parental attitudes toward consent and data sharing in biobanks: a multi-site experimental survey (published)
- NT224: Ethical considerations related to return of results from genomic medicine projects: the eMERGE Network (Phase III) experience (published)
- NT273: Returning genomic results to eMERGE participants: The who, what, where, and how of disclosure (accepted)
- NT277: Operationalizing participant choice about genomic results: Beyond all or none ACMG recommended genes (in development)
- NT322: The Reckoning: What we found after return of results for 25,000 eMERGE-3 participants (in development)
- NT323: Challenges in returning results in the eMERGE consortium (in development)
- NT330: Approaches to the return of actionable adult-onset conditions in pediatric research: Lessons learned from eMERGE-III (in development)
- NT332: Network-wide lessons learned from the reporting of negative test results (in development)

Outcomes

- NT274: Harmonizing outcomes for genomic medicine: comparison of eMERGE outcomes to ClinGen outcome/intervention pairs (published)
- NT296: Collection and analysis of large-scale outcome measures following targeted next generation sequencing (in development)
- NT352: Lessons from eMERGE on readiness for genomic clinical decision support implementation (in development)

Genomics

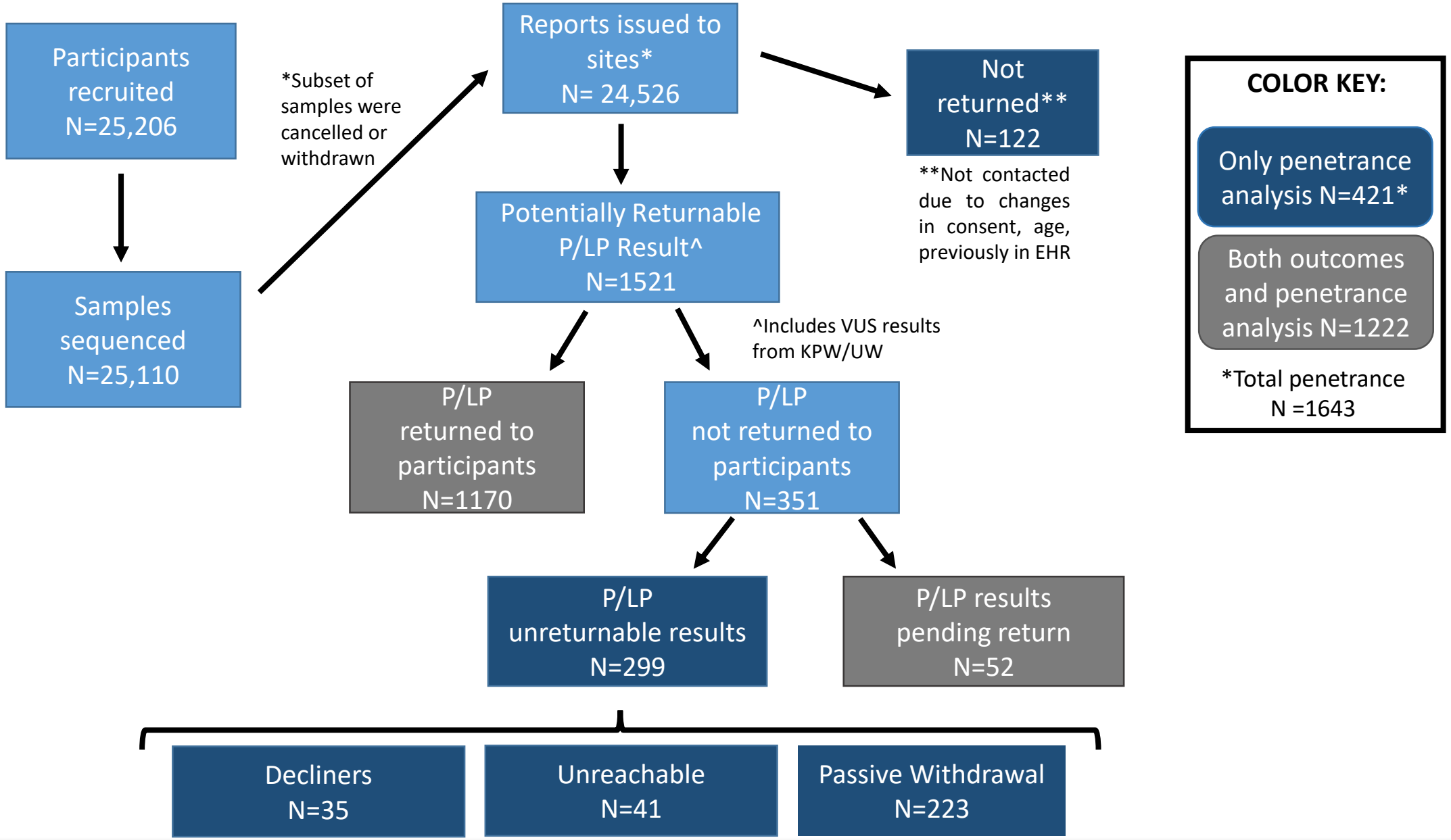
- NT357: Lessons from the eMERGE Network: Balancing genomic discovery and implementation science (in development)

Phenotyping A subset of the lessons learned publications are also being included in a special issue on Phenotyping Methodology in the *Journal of Biomedical Informatics*.

- NT268: OMOP Information Model for Phenotyping (published)
- NT327: A Study of Phenotype Algorithm Portability (in development)
- Site specific: Detecting time-evolving phenotype topics via tensor-factorization on EHRs: Cardiovascular disease case study (published)
- Site specific: Pathway analyses of genomic pathology tests for prognostic cancer subtyping (in development)
- Site specific Ensembles of NLP systems for portable phenotyping solutions (in development)

*e*MERGESEQ SEQUENCING & RETURN of RESULTS

Enrollment, sequencing, & return of results in *e*MERGEseq



eMERGE CSGs : Major Accomplishments from Phase III

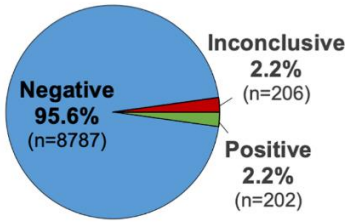
Co-Chairs: Richard Gibbs (Baylor), Heidi Rehm (Partners/Broad), & Niall Lennon (Partners/Broad)

1. Developed and validated the eMERGEseq panel for generating and interpreting genomic sequence data for over 10,500 eMERGE III participants, with over 1,519 positive and 206 inconclusive reports issued to clinical sites
2. Harmonized various components of the sequencing and reporting workflow with Baylor-HGSC including data receipt from the clinical sites, assay development, test validation, primary analysis, pre-test variant harmonization, variant classification, report content, data delivery to sites.
3. Established on-going process for variant harmonization between the two CSGs by sharing variant interpretations monthly to resolve any discordances. Classified variants are submitted to ClinVar.
4. Developed processes for notification of sites when reported variants were reclassified. A systematic approach to variant re-analysis is underway enabling the establishment of more robust pipelines for the return of updated results (see next slides).

Publications

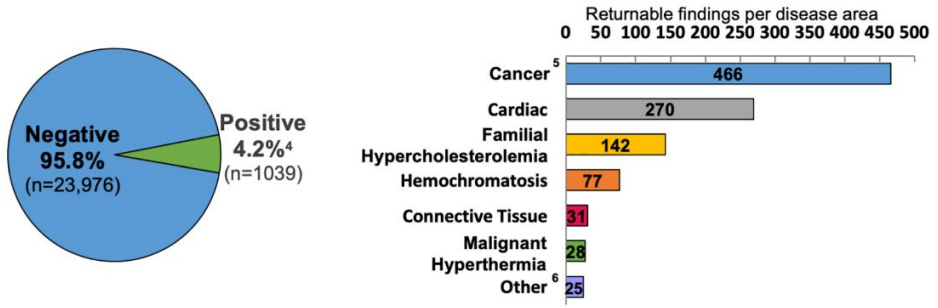
Marker paper: Harmonizing Clinical Sequencing And Interpretation For The Emerge III Network; The eMERGE Consortium; Published in the *American Journal of Human Genetics (AJHG)*, Aug 22nd, 2019

A Indication-based returnable results (n=9,195)

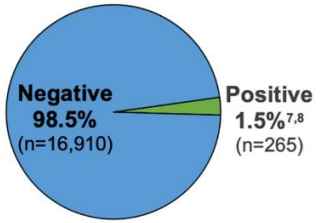


Indication[1]	Total	Positive	Negative	Inconclusive
Colorectal cancer/Polyps	2593	39	2348	206
Arrhythmia	188	5	183	n/a
Asthma	135	0	135	n/a
Cardiomyopathy	16	1	15	n/a
Chronic Kidney Disease	1069	7	1062	n/a
Obesity	33	0	33	n/a
Pulmonary Hypertension	28	0	28	n/a
Tubular Sclerosis Complex	12	11	1	n/a
Abnormality of pain sensation	678	0	678	n/a
Autistic Behavior	102	1	101	n/a
Ehlers-Danlos Syndrome	89	1	88	n/a
Hyperlipidemia[2]	3057	56	3001	n/a
Pediatric Migraine	473	0	473	n/a
Breast/Ovarian Cancer[3]	427	79	348	n/a
at risk for colorectal cancer	126	0	126	n/a
at risk for Breast/Ovarian Cancer	38	0	38	n/a
At risk for prostate cancer	131	2	129	n/a
TOTAL	9195	202	8787	206

B Non indication-based consensus returnable results (n=25,015)



C Non indication-based site-specific returnable results (n=17,175)



Disease/Disorder	P/LPVariants in site specific genes	Total
Cancer	*CHEK2*	96
	JAK2	3
Cardiac	*KCNE1 (risk allele)*	62
	KCNE2	4
Clotting	*F11*	3
Endocrine	*M4CR*	12
Neuromuscular	*TOR1A*	4
	CACNA1A	3
Neurological	*TTR*	1
	ATM	36
	APT1A2	1
Respiratory/ Hypertension	*SERPINA1*[8]	27
	CFTR[8]	7
	BMPR2	1
Dermatological	*FLG*	2
Renal	*CFH*	1
	TOTAL	265

Develop harmonized structured genetic test report standards compliant with FHIR/HL7

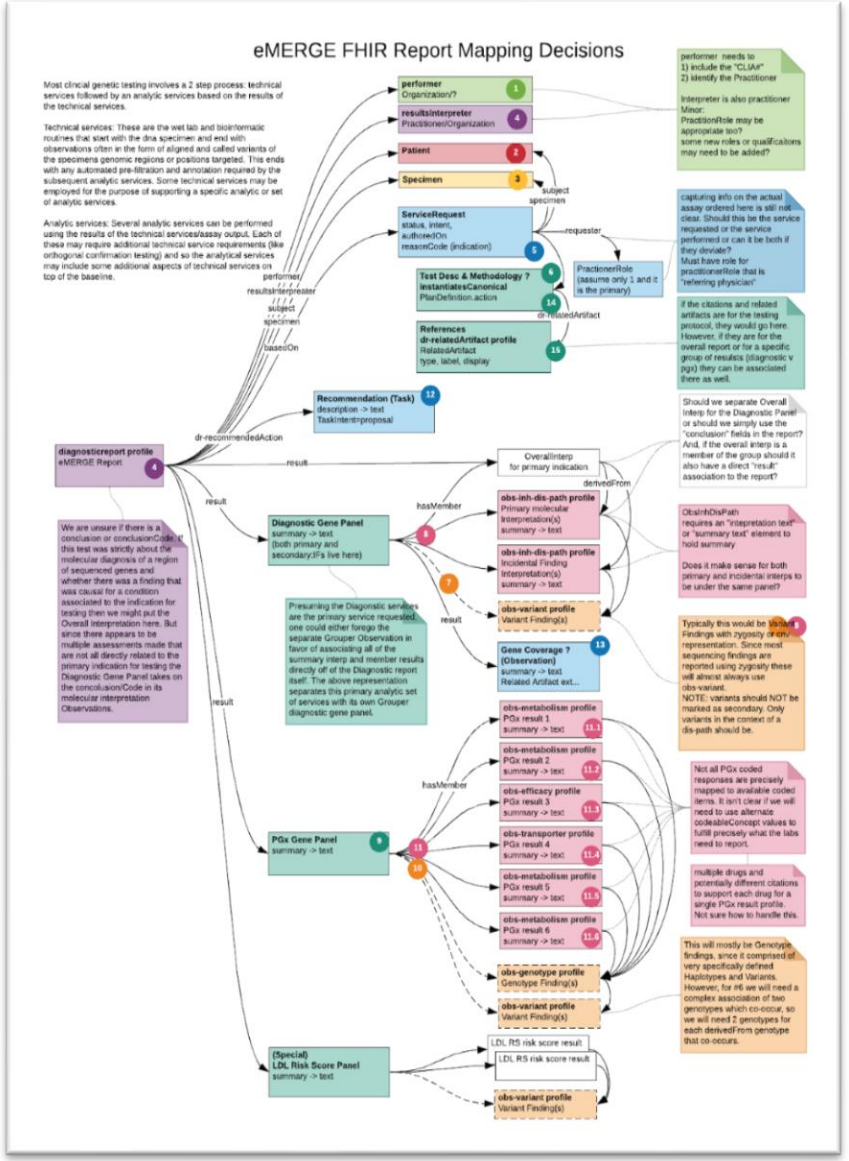
Project Leads: Mullai Murugan (Baylor) & Larry Babb (P/B)

Goals:

- Develop a computable and standardized clinical reporting specification for eMERGE clinical genetic reports capitalizing on the HL7 Clinical Genomics FHIR national standard,
- Use the FHIR specification and create a pilot implementation that generates a set of pre-identified existing eMERGE reports in FHIR format
- Work with a clinical site to identify feasibility of ingesting the FHIR formatted eMERGE reports into the EMR.

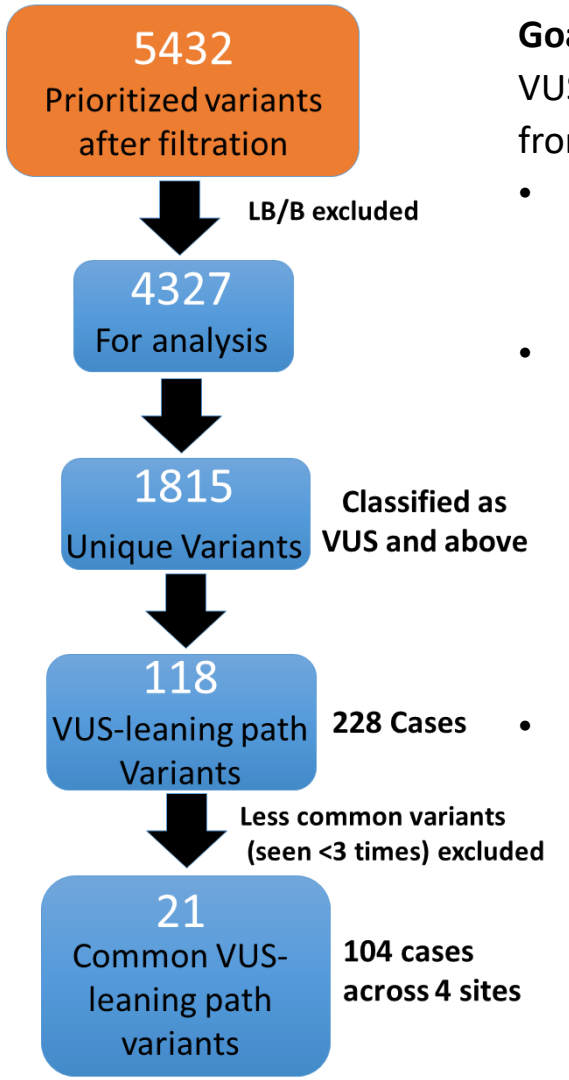
Progress:

- Mapping of the eMERGE report to draft Clinical Genomics FHIR IG Spec nearly complete (see image).
- Identified design for creating eMERGE FHIR specification.
- Ongoing effort to collaborate and reconcile differences with HL7 FHIR CG Workgroup re specification.
- Ongoing effort towards completion of the specification.
- Pilot implementation of specification started. Infrastructure setup on AWS and API work is in currently in progress.
- Working with Northwestern to extract and load FHIR reports generated for pilot and to identify a potential use case for CDS.



Partners-Broad CSG: ongoing work

CSG variant Reanalysis of VUS-Leaning Pathogenic variants using eMERGE EHR Data



Goal: Identify variants with higher probability of reclassification from VUS to LP/P (VUS-Leaning pathogenic) and integrate phenotypes data from EHRs for improved interpretation.

- 118 unique VUS-leaning path variants have been identified affecting 228 participants (diagram left), vast majority are in cardiac genes (graph, right).
- 21 deemed common (seen in <3 participants, diagram, left) and became the basis of a manageable, prioritized list of variants sent to clinical sites with impacted participants for retrieval of additional EMR phenotype data (affected status, family history etc.) for incorporation into variant reassessment.
  - These 21 variants were found in 104 participants and span 13 genes.
- Chart review summaries received from 1 site: No strong correlation with expected phenotype, 3 participants with possible correlation:
  - *APOB:* 1 patient with high LDL but does not meet familial hypercholesterolemia dx; 1 patient with coronary heart disease and elevated triglycerides.
  - *MYL2:* patient has atrial fibrillation but not HCM.

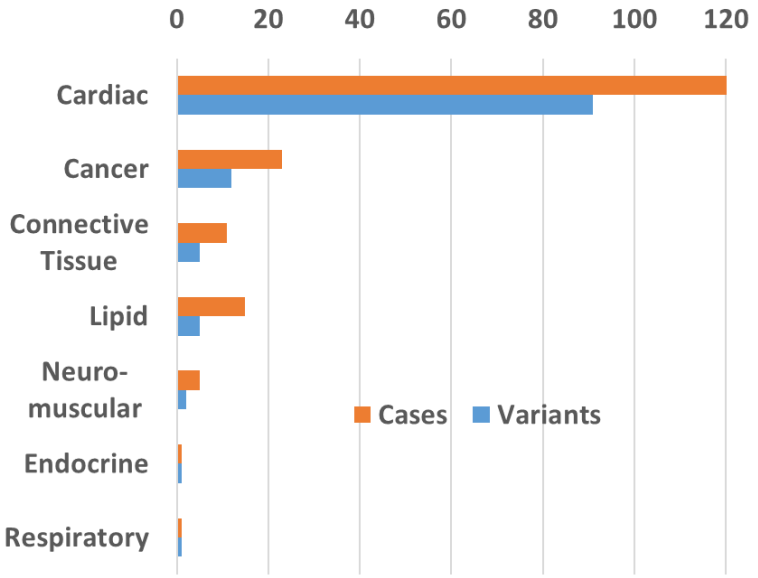


Table of 118 Unique VUS-Leaning Pathogenic Variants in eMERGE III Cohort (40 genes, 228 patients) (above)

Next steps:

- Combine Partners-Broad and Baylor-HGSC VUS-Leaning path list, identify additional common VUS-leaning path variants to send to sites for EHR review for relevant clinical phenotypes.
- Incorporate any relevant data from sites into variant reassessments and determine whether reclassification is warranted

Partners-Broad CSG: ongoing work

CSG re-assessment of variation identified in eMERGE participants

Goal: Identify variants most likely to undergo reclassification by looking at interpretation differences between CSG classifications and those from ClinGen-approved Expert Panels and high-quality clinical labs in ClinVar that occurred after initial analysis in eMERGE .

- 21 variants have been newly classified by another lab or expert panel as pathogenic/likely-pathogenic (P/LP)

Next steps: Review updated evidence and determine whether reclassification is warranted. Updated classifications will be shared with Baylor-HGSC for variant harmonization to occur as needed, prior to contacting clinical sites with impacted participants.

In a subset of reported LP/P variants, reclassifications have already been happening. Ex: Variant reassessment by the CSG during the course of eMERGE III prompted by its identification in either an eMERGE or non-eMERGE individual.

- 11/565 reported unique variants, affecting 23/743 total reports have been reclassified (see table).

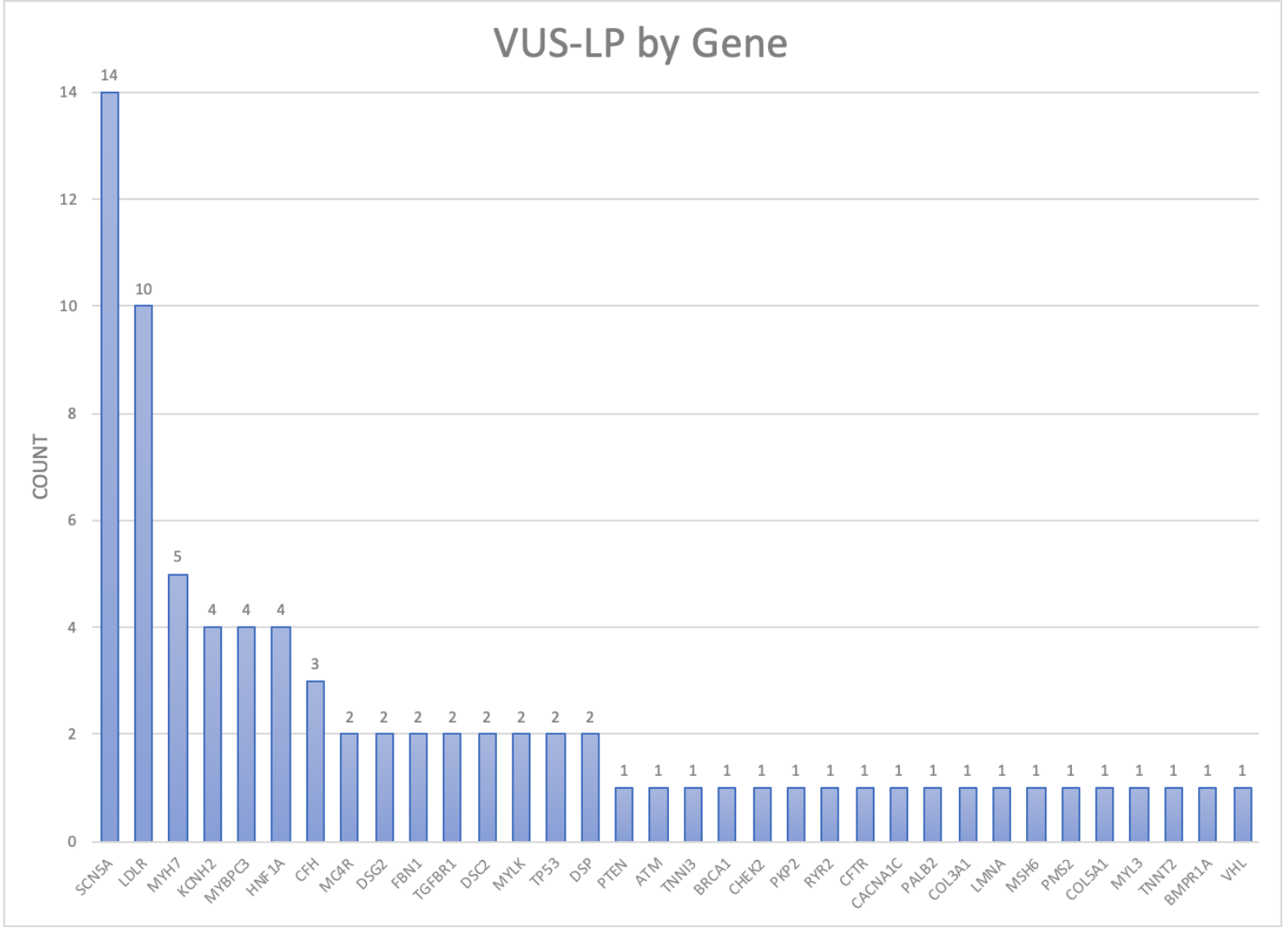
Next steps: contact clinical sites for preferences on issuing amended reports for those updates not yet sent to sites.

Variant	Original	Update	Site	Time Elapsed	Reason for Update
RYR1 c.1840C>T (p.Arg614Cys)	LP	P	UW/KP, CCHMC, Harvard, Geisingerx3	4 months	New Evidence
MYBPC3 c.(?_26)_(1090_?)del	VUS	LP	UW/KP	5 months	Rules change
SCN5A c.1603C>T (p.Arg535X)	LP	P	UW/KP	9 months	New Evidence
BRCA2 c.9777delT (p.Ile3259fs)	LP	P	Geisinger	3 months	New Evidence
LDLR c.1027G>A (p.Gly343Ser)	LP	P	UW/KP, Geisinger, Harvard	4 months	New Evidence
BRCA1 c.594-1G>T	P	VUS	UW/KP	4 months	Expert Panel - transcript evidence
MYL3 c.170C>G (p.Ala57Gly)	LP	VUS	UW/KP and Geisinger	5 months	New Evidence
DSC2 c.631-2A>G	LP	VUS	UW/KP	9 months	Expert Panel – LOF mutation mechanism unclear
SCN5A c.3911C>T (p.Thr1304Met)	LP	VUS	UW/KP, CCHMC, Harvardx2	5 months	New evidence and more stringent review of old evidence
MYBPC3 c.2671C>T (p.Arg891Trp)	LP	VUS	Geisinger	2 months	Rules change (ACMG/ClinVar EP)
SCN5A c.393-1C>T	LP	VUS	UW/KP, CCHMC	18 months	Rules change (ACMG/ClinVar EP)

Baylor CSG: ongoing work

VUS Leans Pathogenic

- Total: 79 Variants
- 50% are cardiomyopathy or arrhythmia genes
- Common findings:
  - PM1 (hotspot or functional domain)
  - PM2 (Absent from population databases)
  - PP3: Predicted damaging *in-silico*
- Contacted clinical sites with Variant lists
  - Received results from 4 sites (35 variants)
- 1 case has strong overlap with clinical indication (variant in VHL)
  - 4 more with possible overlap (e.g. dx of hyperlipidemia, variant in LDLR)
  - Assessing



Baylor CSG: ongoing work

Variant reanalysis

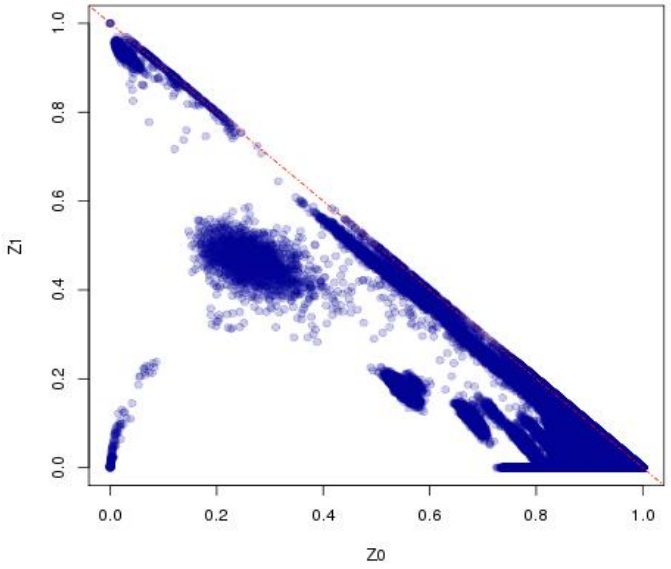
- Supplement Aims 1a and 3c propose to re-issue reports when variant classifications change
- We developed the ReVU (Reanalysis of Variants and Updater) tool to help with ClinVar reanalysis
- In initial analysis, from ClinVar, we identified 158 variants with a new Pathogenic classification
- All VUS per our initial review
- 148 remain VUS, 8 still under final review
- Examples:
  - COL3A1 p.Gly532Ser. No ClinVar or literature evidence at the time of review. Now Invitae submitted in 2018 and classified it as LP.
  - BRCA2, p.Gly2609Val. No ClinVar or literature evidence at the time of review. Now Ambry (LP), Invitae (VUS) and GeneDx (VUS) all submitted to ClinVar
- Starting assessment of a set of a set of an additional 290 pulled from ClinVar in August.

NETWORK DATA RESOURCES & MANAGEMENT

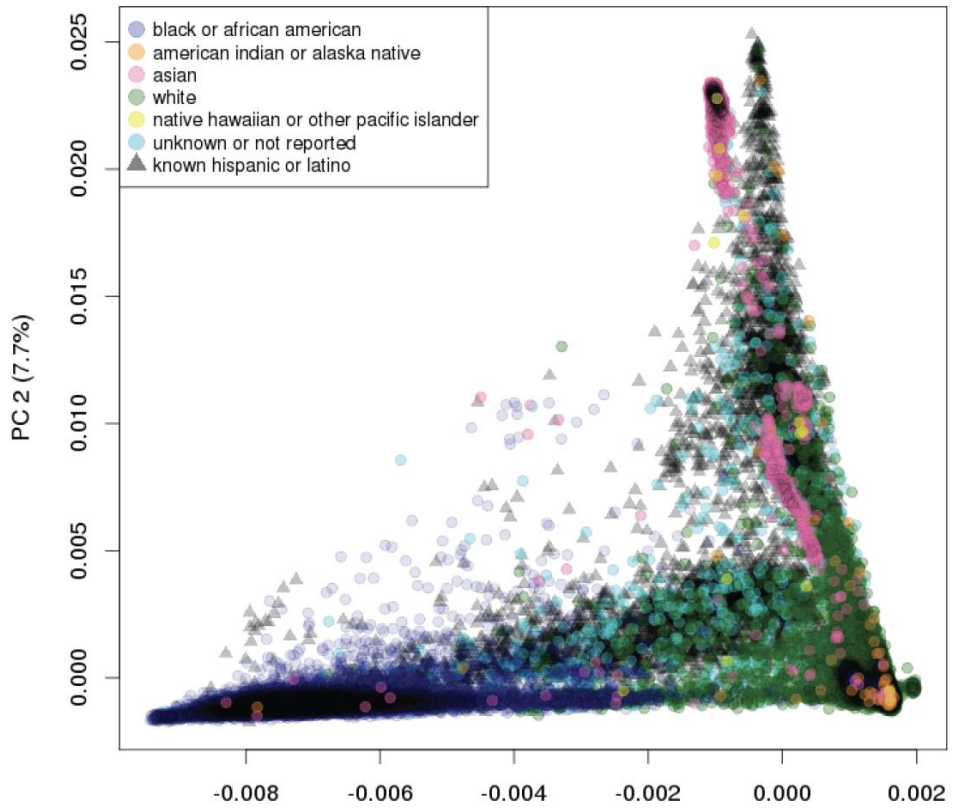
DATASET: *e*MERGE I-III array imputed & merged SNVs

Medical Center	Participants	Gender Female	Gender Male	Gender Unknown	African / Black	Native American	Asian	White	Hispanic / Latino	Pacific Islander	Ancestry Unknown
bsch	1019	423	596	0	66	2	21	676	125	0	129
ccmc	6505	2759	3729	17	615	6	71	5442	157	5	209
chop	10465	4835	5630	0	4666	7	161	4890	321	3	417
colu	2065	1008	1057	0	162	6	62	619	297	1	918
geis	3105	1470	1634	1	9	2	0	3080	13	0	1
harv	30714	16780	13934	0	1525	0	677	25517	1652	0	1343
kpuw	3316	1888	1428	0	109	12	89	2929	69	6	102
mayo	10247	5185	5058	4	39	0	21	8812	1042	0	333
mrsh	4756	2878	1878	0	3	31	13	4690	14	0	5
mtsi	6255	3700	2555	0	4046	33	3	679	1297	0	197
nwun	4847	4029	818	0	598	0	0	4208	35	0	6
vand	21814	11947	9867	0	3836	17	99	17312	205	0	345
Total	105108	56902	48184	22	15674	116	1217	78854	5227	15	4005

Self reported demographics eMERGE GWAS



*Qualifies for International 100K Cohort Consortium (IHCC)



Genet Epidemiol. 2019 Feb;43(1):63-81. doi: 10.1002/gepi.22167. Epub 2018 Oct 8. Sign in

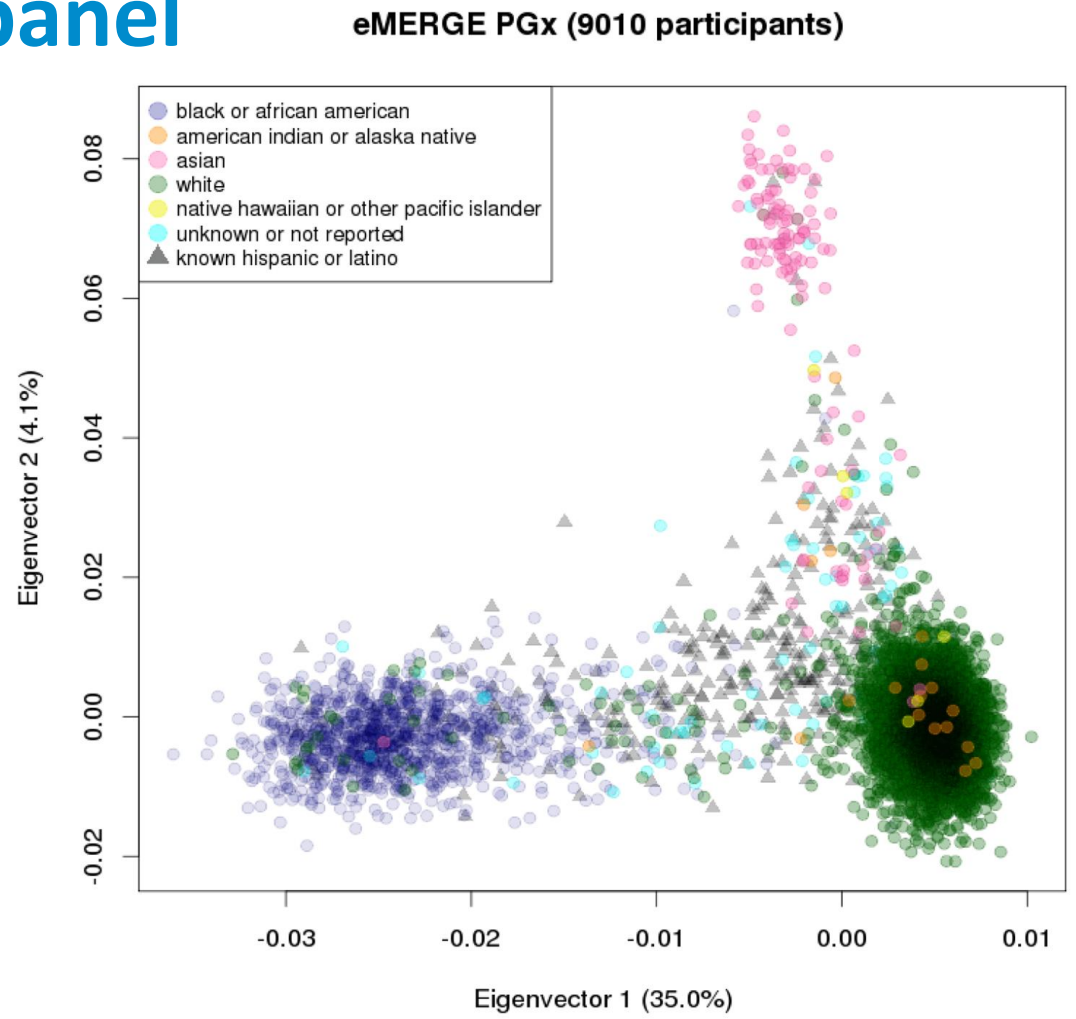
The eMERGE genotype set of 83,717 subjects imputed to ~40 million variants genome wide and association with the herpes zoster medical record phenotype.

Stanaway IB[1], Hall TO[1], Rosenthal EA[2], Palmer M[2], Naranbhai V[1,3], Knevel R[3], Namjou-Khales B[4], Carroll RJ[5], Kiryluk K[6], Gordon AS[2], Linder J[7], Howell KM[7], Mapes BM[7], Lin FTJ[8], Joo YY[8], Hayes MG[8], Gharavi AG[6], Pendergrass SA[9], Ritchie MD[10], de Andrade M[11], Croteau-Chonka DC[3], Raychaudhuri S[3,12], Weiss ST[3], Lebo M[3], Amr SS[3], Carrell D[13], Larson EB[13], Chute CG[14], Rasmussen-Torvik LJ[8], Roy-Puckelwartz MJ[8], Sleiman P[15], Hakonarson H[15], Li R[16], Karlson EW[10], Peterson JF[5], Kullo IJ[11], Chisholm R[8], Denny JC[5], Jarvik GP[2]; eMERGE Network[16], Crosslin DR[1].

DATASET: PGRNseq (PGx) targeted panel

Site	Number of Files
Boston Children's Hospital	109
CCHMC	765
CHOP	1784
Geisinger	1086
GHC/UW	989
Marshfield Clinic	747
Mayo Clinic	1012
Mt. Sinai	884
Northwestern University	731
Vanderbilt University	903
Total	**9010**

Table 1: Number of PGRNseq BAMs and gVCFs per eMERGE site

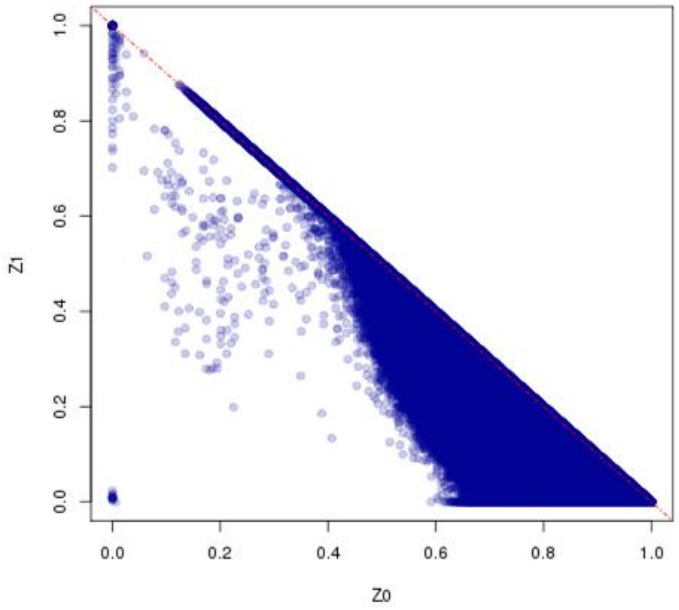


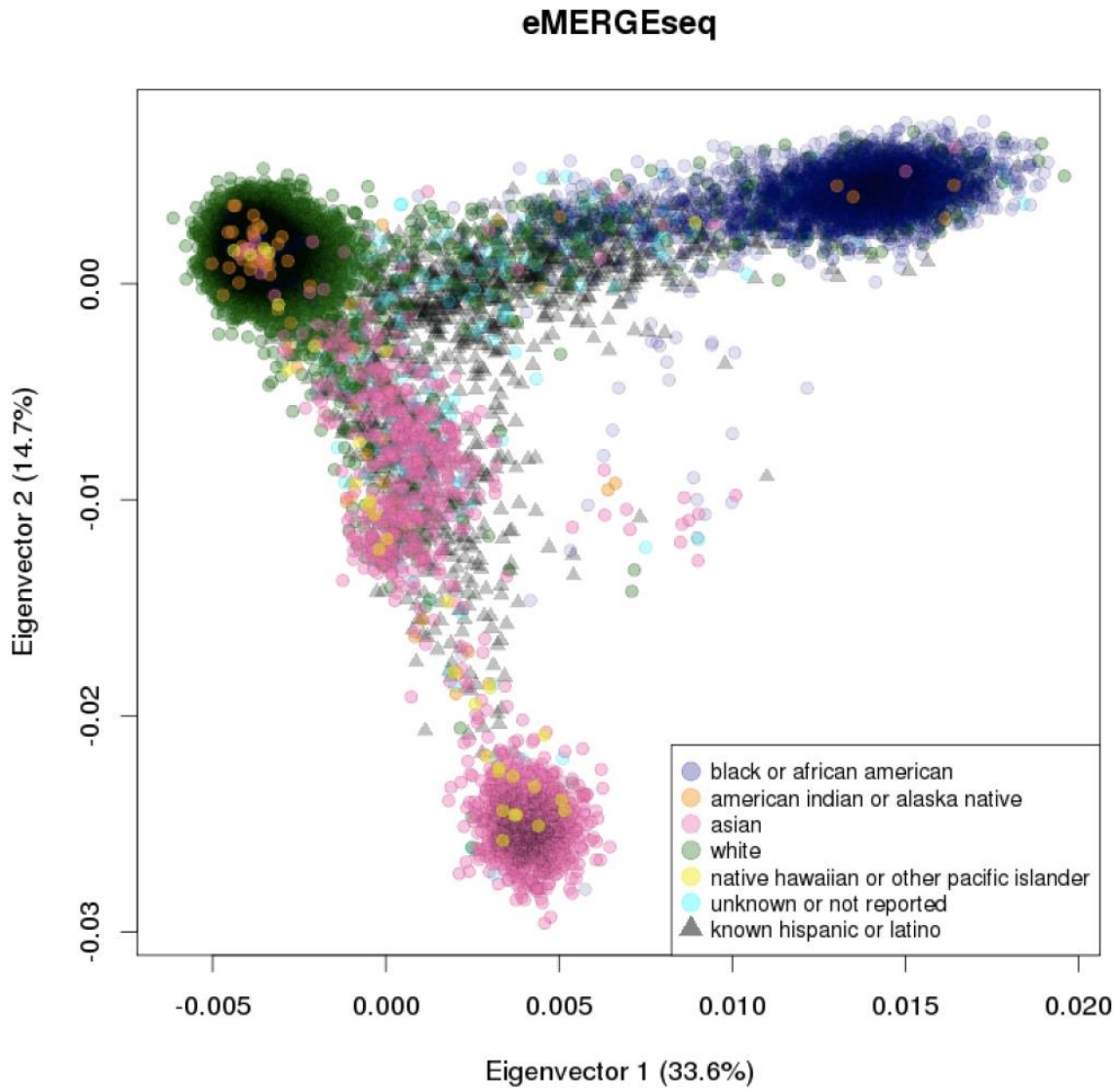
Gordon A.S., Stanaway I.B., ... Jarvik G.P., **Crosslin D.R.** (2019) "Pharmacogenetic variation identified via targeted next-generation sequencing among 9010 eMERGE Network participants", target is *Pharmacogenetics and Genomics*, invitation to revise and resubmit.

DATASET: *e*MERGEseq (eIII) ACMG+ targeted panel

Medical Center Site	Number of Participants
Columbia	2,580
CCHMC	2,956
CHOP	2,991
Geisinger	2,498
KP/UW	2,499
Harvard	2,493
Meharry	495
Mayo Clinic	3,020
Northwestern University	2,984
Vanderbilt University	2,440
Total	**24,956**

Table 1: Number of participants per eMERGE site





Phasing and imputation of structural variation using the ~100,000 *e*MERGE array samples

Medical Center	Participants	Gender Female	Gender Male	Gender Unknown	African / Black	Native American	Asian	White	Hispanic / Latino	Pacific Islander	Ancestry Unknown
bsch	1019	423	596	0	66	2	21	676	125	0	129
ccmc	6505	2759	3729	17	615	6	71	5442	157	5	209
chop	10465	4835	5630	0	4666	7	161	4890	321	3	417
colu	2065	1008	1057	0	162	6	62	619	297	1	918
geis	3105	1470	1634	1	9	2	0	3080	13	0	1
harv	30714	16780	13934	0	1525	0	677	25517	1652	0	1343
kpuw	3316	1888	1428	0	109	12	89	2929	69	6	102
mayo	10247	5185	5058	4	39	0	21	8812	1042	0	333
mrsh	4756	2878	1878	0	3	31	13	4690	14	0	5
mtsi	6255	3700	2555	0	4046	33	3	679	1297	0	197
nwun	4847	4029	818	0	598	0	0	4208	35	0	6
vand	21814	11947	9867	0	3836	17	99	17312	205	0	345
Total	105108	56902	48184	22	15674	116	1217	78854	5227	15	4005

Table 1: eMERGE Demographics

- **Goal:** Impute common structural variation (both indels and larger events) for 100,000 eMERGE array samples, and link to phenotype data derived from EHR
- Imputation finished, QCed, and data made available to the Network
- Includes ABO blood type calls, match known US distribution almost exactly (validated against known serological data)

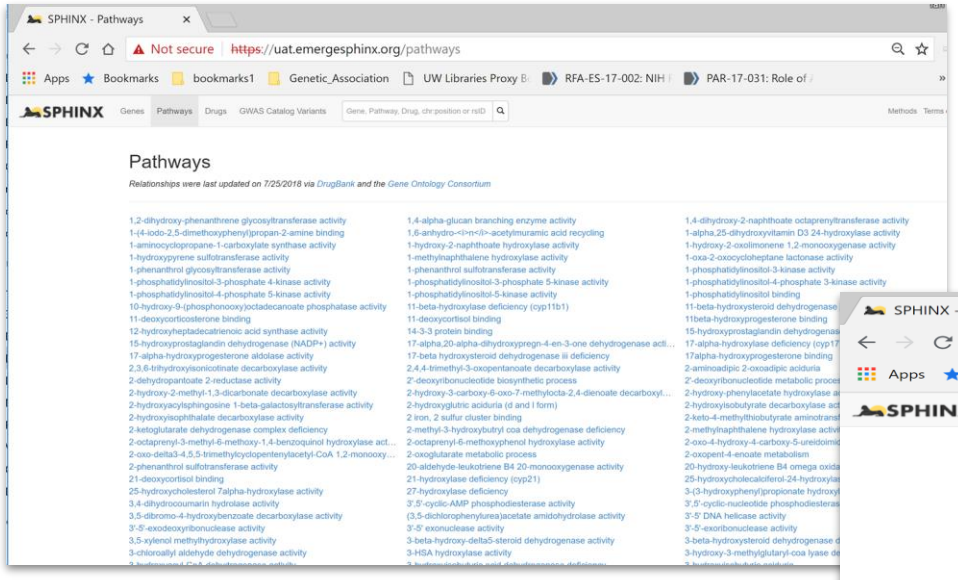
Copy Number State	Concordance
Total All	0.801 = 127138 / 158748
0	0.744 = 40675 / 54641
1	0.884 = 67114 / 75954
3	0.918 = 19227 / 20952
4	0.012 = 122 / 6451
5	0 = 0 / 345
6	0 = 0 / 373
7	0 = 0 / 8
8	0 = 0 / 17
10	0 = 0 / 7

Table 2: Imputed Copy Number Concordance of CNVs and Duplications with PennCNV Calls

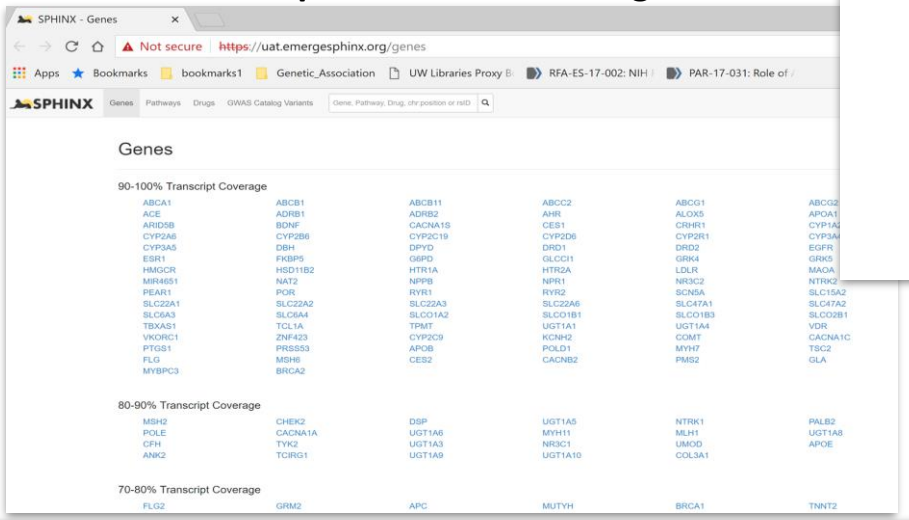
Continued enhancement of SPHINX

Plan to rebrand SPHINX as Electronic MEdical Records and GENomics Toolkit (*e*MERGENT)

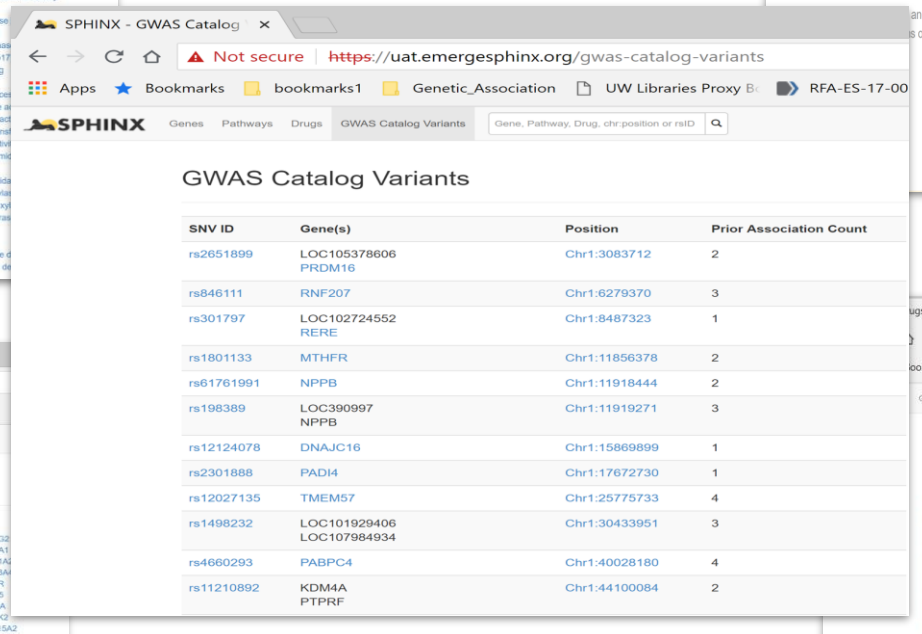
Gene ontology pathways



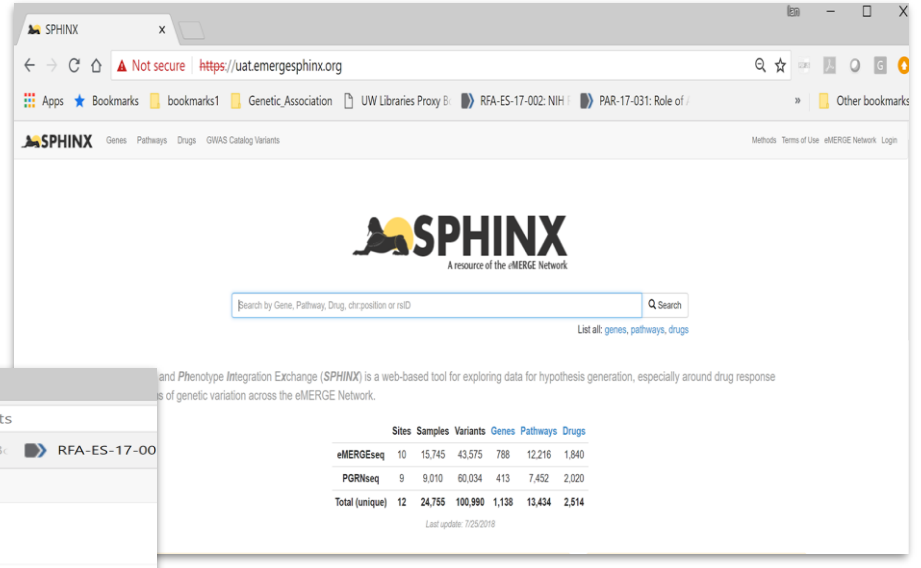
Gene percent exon coverage



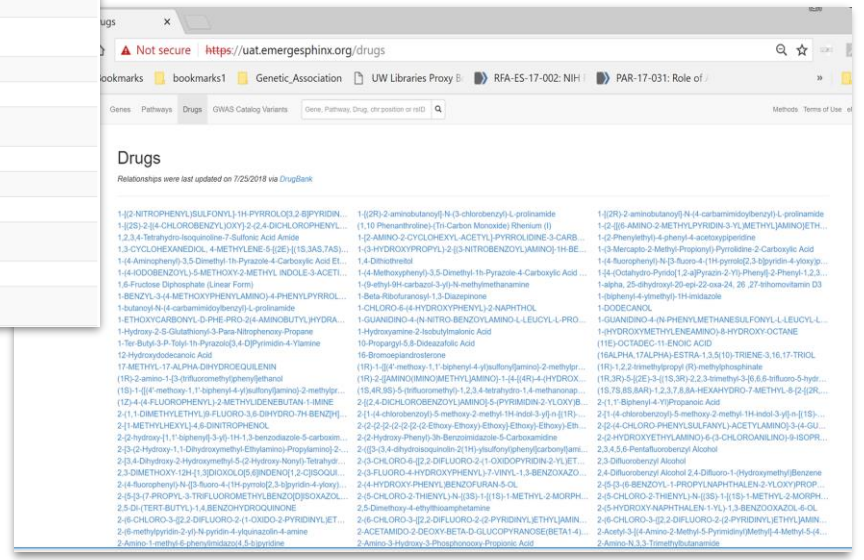
NHGRI GWAS catalog variants in PGRNseq and eMERGEseq



Addition of eMERGEseq data



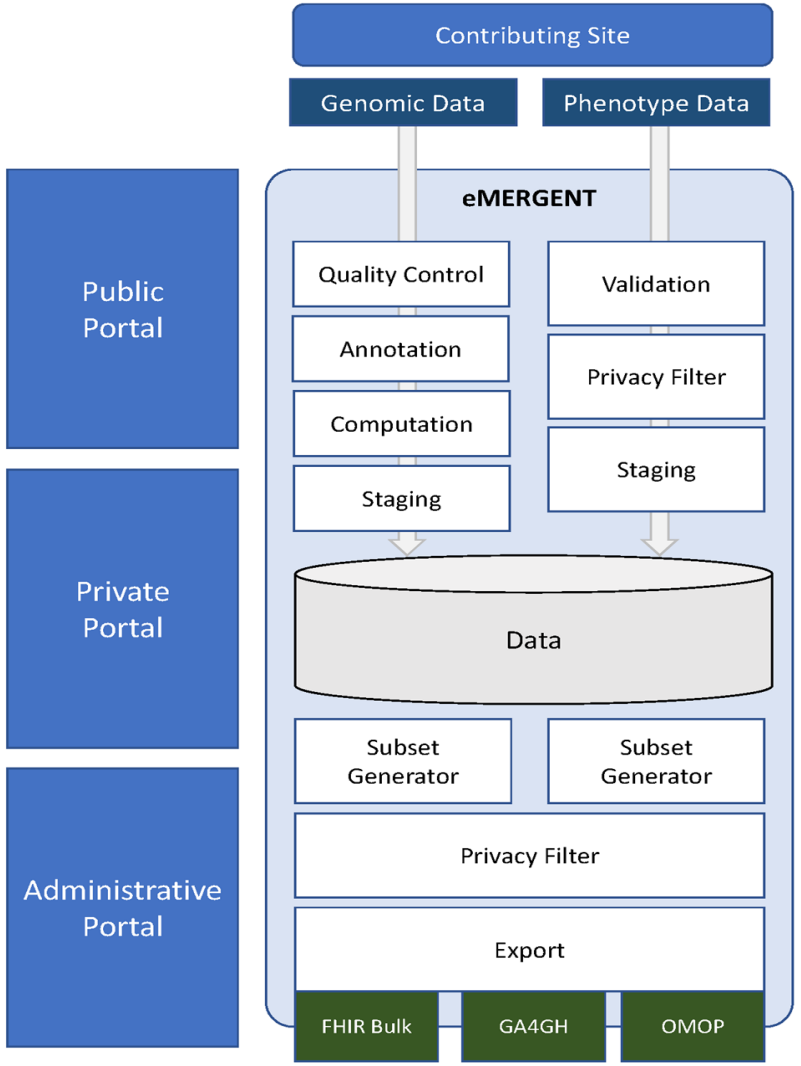
Gene-drug interactions from DrugBank



Automatic refreshes of pathway & drug data on a weekly basis

The *e*lectronic *ME*dical *R*ecords and *GEN*omics *T*oolkit (*e*MERGENT)

- Applied for a Genomic Community Resources (U24) grant to build tool to enhance research using the rich eMERGE phenotype and genetic data
- Received 100% support across NHGRI programs and eMERGE PIs
- Proposing we build in the NHGRI Genomic Data Science Analysis, Visualization, and Informatics Lab-Space (AnVIL) ecosystem
  - Being considered *clinical conduit* for AnVIL
  - Will support research activities not directly planned for implementation by AnVIL, both eMERGE and general research community.
- Modular implementation activities in the future (U24, R01)
  - EHR integration
  - CDS (populating data needed to make decisions)
  - Outcomes
  - ELSI



eMERGE III Network Datasets & Phenotype Variables

Common variable refresh completed (July 2019)

- eMERGEseq V2
- eMERGE I-III Merged Imputed GWAS V3
- PGRNseq
- Exome Chip
- Whole Genome Sequencing (WGS; NU samples only)

eIII case/control status files to be submitted to dbGaP

Phenotype
Adult Familial Hypercholesterolemia _Stage 1
Adult Familial Hypercholesterolemia_Stage 2
Colorectal Cancer
Epilepsy
Chronic Rhinosinusitis
Chronic Kidney Disease
Contrast induced nephropathy
Hearing Loss
Rheumatoid arthritis
FattyLiver_Case 1
FattyLiver_Case 2
FattyLiver_Case 3
FattyLiver_Case 4
Intellectual Disability
Ovarian Uterine Cancer
Autoimmunity
Anxiety

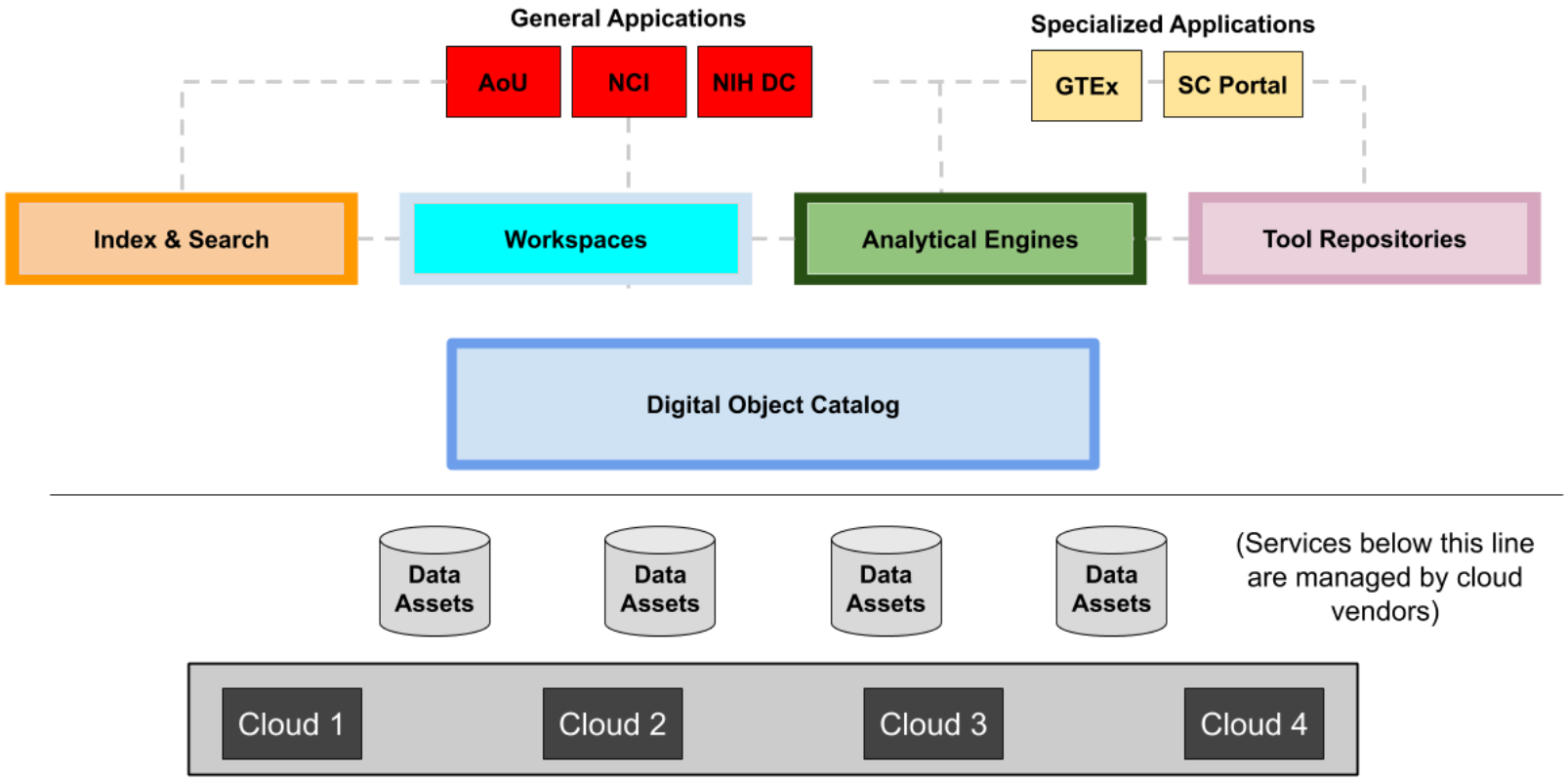
eMERGE network-wide data sets

Set Name	Platform	Count
eI-eIII imputed*	GWAS	105,108
Exome chip	Exome	12,555
Whole exome	Sequencing	3,745
PGRNseq	Sequencing	9,010
Whole genome	Sequencing	1,800
eMERGEseq	Sequencing	24,956
Total Current		**158,174**

Common variable files available for Network datasets

Data Type	Common Variable
Demographics	Sex, year of birth, decade of birth, race, ethnicity
Codes	ICD, CPT
BMI	Height, weight, BMI
Labs	Serum total cholesterol, LDL, HDL, Triglycerides, Glucose fasting/non-fasting/unknown, White Blood Cell, & Autoimmunity labs
Medications	Cerivastatin sodium, Rosuvastatin, Simvastatin, Fluvastatin, Pravastatin, Lovastatin, Atorvastatin, Pitavastatin
Other	Case/Control Status on Phase I / Phase II Phenotypes

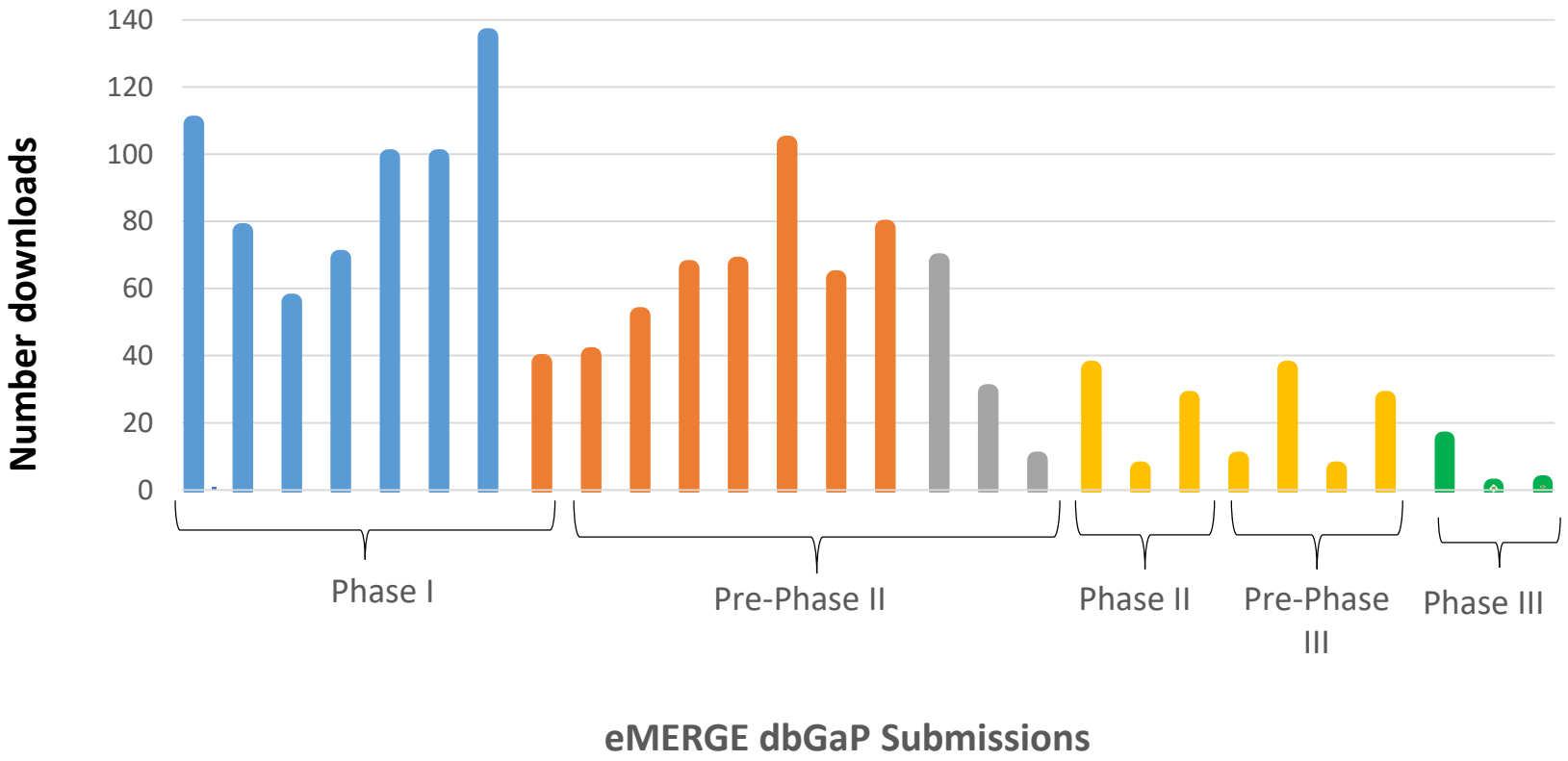
Migration of *e*MERGE data assets to AnVIL



AnVIL Ecosystem (figure courtesy of Dr. Anthony Philippakis)

Impact: dbGaP & Website Analytics *2007-2019*

Downloads *of* *e*MERGE dbGaP Submissions as of September 2019



> 1422 external downloads as of September 2019

*e*MERGE Website

Average usage past 6 months

- 87.3% new visitors
- 1723 sessions/month
- 1298 users/month
- Views from 80 countries

PheKB Website

Average usage past 6 months

- 82.1% new visitors
- 1455 sessions/month
- 842 users/month
- Views from 76 countries

Network Collaborations

- **CSER (Clinical Sequencing Evidence-Generating Research)**
  - eMERGE and CSER: The Convergence of Genomics and Medicine meeting was held in February 2017 and another joint meeting was held in January 2019.
  - Clinical Annotation workgroup collaboration on variant interpretation and reanalysis.
  - EHR Integration workgroup collaborated with CSER on Lynch Syndrome CDS Guide.
  - "*Return of Genomic Results to Research Participants: The Floor, the Ceiling, and the Choices in Between*" published in AJHG.
- **IGNITE (Implementing Genomics in Practice)**
  - Joint meeting held was held in January 2016.
  - Manuscript in development "*IGNITE Clinical Informatics Working Group: Genetic Data Pipeline Project*" led by Paul Dexter will collect responses on CPIC common practices and challenges from members of IGNITE, eMERGE, and CSER.
- **AoU (*All of Us* Research Program)**
  - Many shared investigators overlap in eMERGE and AoU including sequencing and data coordinating centers.
  - ROR/ELSI workgroup lessons learned manuscripts (NT273, NT322, and NT332) will help other consortiums understand the impact of returning positive, negative, and PGx results to participants.
- **MeTree Supplement Grant**
  - Collaboration between eMERGE (VUMC, Geisinger, and Northwestern) and Duke's MeTree family health history (FHH) collection tool.
- **CPIC (Clinical Pharmacogenetics Implementation Consortium)**
  - Joint eMERGE and CPIC call in January 2019 to discuss existing eMERGE PGx and EHR data to build evidence base for guidelines.
  - eMERGE PGx prominently featured (in area of PGx implementation) at June 2019 CPIC conference.
- **ClinGen**
  - eMERGE EHRI workgroup members work along side ClinGen's EHR group to inform EHRI developments and research in both Networks

*e*MERGE WORKGROUP PROGRESS

Clinical Annotation

Co-Chairs: Gail Jarvik (KPW/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 (KPW/UW)

Outcomes

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

PGx

Co-Chairs: Laura Rasmussen-Torvik (NU) & Cindy Prows (CCMHC)

Phenotyping

Co-Chairs: Chunhua Weng (Columbia) & Wei-Qi Wei (VUMC)

RoR/ELSI

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

*e*MERGE CLINICAL ANNOTATION WG: Major Accomplishments from Phase III

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

The Clinical Annotation work group has focused on activities that built consistency of approaches to the gene and variant interpretation across the eMERGE sequencing centers and study sites as well as supported contribution to public knowledge bases. We have:

- Applied the ClinGen approach to gene-disease validity assessment to all genes on the eMERGE gene panel (including genes where single or a few variants are associated with disease), defined each associated condition and the strength of evidence
- Developed consistency in variant interpretation approaches across CSGs
- Developed consensus on clinically reportable variants in the eMERGE panel and whether to recommend return to patients
- Worked jointly with the ROR/ELSI WG to gather feedback and develop consensus on standard language used in clinical reports
- Supported discussions and decisions on whether to return challenging genes, cases and variants to participants
- Facilitated regular ClinVar submissions for all variants interpreted for the eMERGE program
- Summarized incidental findings across the network (pending publication)

Publications

- Marker paper: Harmonizing Clinical Sequencing And Interpretation For The eMERGE III Network; The eMERGE Consortium; Published in AJHG Aug 22nd, 2019
- Incidental finding paper, Gordon et al, in draft

Race/ Ethnic.	Participants	IF Rate %
Native Amer	68	1.47
Hispanic	772	2.07
Black	2544	2.83
Asian	1391	2.80
White	13393	3.35
Unknown	3747	2.45

Breakdown of incidental findings by race

eMERGE CLINICAL ANNOTATION WORKGROUP: Final Efforts & Milestone Completion

Penetrance Project

Collaboration with Outcomes, Genomics, & ROR/ELSI Workgroups

Network year 5 milestone: Expand the understanding of penetrance and the impact a variant has on clinical outcomes, using a variety of approaches including cascade testing and family history analyses.

Evaluate in participants NOT ascertained for a relevant phenotype.

Aligned chart review for penetrance with that for outcomes using the same forms. Data are being collected along side Outcomes forms and additionally in participants without completed return of results.

Challenges: Aligning outcomes forms for penetrance, timeline with form completion

Progress: 856 participant records of outcomes and penetrance data

Initial Focus on Tier 1 conditions: Breast Cancer, Familial Hypercholesterolemia, Colorectal Cancer

Table of Penetrance Analysis Order (below)

Outcomes Form	Site	Order- Comments
Familial Hypercholesterolemia (Adult/Peds)	Mayo/Geisinger	1
CRC/Polyps / Lynch Syndrome	KPW/UW	2
Cardiomyopathy	Northwestern	3
Ornithine transcarbamylase deficiency	Geisinger	4
Tuberous Sclerosis	Geisinger	4
Breast Cancer	Columbia	5 – Start with non-BRCA1/2 genes (PTEN, PALB2, CHEK, ATM; BRCA1/2 for functional
Arrhythmias	Vanderbilt	6
Ehlers Danlos Syndrome – Classical	CCHMC	7
Ehlers Danlos Syndrome – Vascular	CCHMC	7
Aortic Dilatation / Aortopathy	Mayo	9
Chronic Kidney Disease	Columbia	10
Hemochromatosis	KPW/UW	11 C282Y hom and C282Y/H63D
22q Deletion/Duplication	CHOP	Need genotype data called as path/LP

eMERGE CLINICAL ANNOTATION WORKGROUP: Final Efforts & Milestone Completion

Variant Interpretation and Reanalysis

Collaborations with CSER, ASHG and CSGs

CSER collaboration: CSER-eMERGE Variant Bake-off 2 (labs; in progress)

ASHG collaboration: *"The Responsibility to Recontact Research Participants after Reinterpretation of Genetic and Genomic Research Results."* Am J Hum Genet. 2019 Apr 4;104(4):578-595. PMID: 30951675

The workgroup included representatives from the National Society of Genetic Counselors, the Canadian College of Medical Genetics, and the Canadian Association of Genetic Counsellors. The final statement includes twelve position statements that were endorsed or supported by the following organizations: Genetic Alliance, European Society of Human Genetics, Canadian Association of Genetic Counsellors, American Association of Anthropological Genetics, Executive Committee of the American Association of Physical Anthropologists, Canadian College of Medical Genetics, Human Genetics Society of Australasia, and National Society of Genetic Counselors.

Guidance to CSGs: Provide guidance to the clinical and functional analysis of VUS-Leaning Pathogenic variants as well as CSG Re-interpretation Project (see details on CSG Plans)

*e*MERGE EHRI WORKGROUP: Major Accomplishments from Phase III

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

- **Engineering**
  - We established what is to the best of our knowledge the first interinstitutional network capable of transmitting structured genetic results from a heterogenous set of laboratories to a heterogeneous set of providers using a common data format
  - **We have consistently shared information and experiences across our sites as we work to integrate these results into our clinical systems and learn from work to establish clinical decision support**
- **Science**
  - We have published our experience establish the above network and open sourced the underlying XML format
  - **We have consistently tracked and surveyed progress and challenges across this network**
  - **Recent publications to disseminate lessons learned**
    - Aronson S, Babb L, Ames D, Gibbs RA, et al. 2018. Empowering Genomic Medicine by Establishing Critical Sequencing Result Data Flows: The eMERGE Example. *Journal of the American Medical Informatics Association*. 25(10), pp.1375-1381.
    - Williams MS, Taylor CO, Walton NA, et al. Genomic Information for Clinicians in the Electronic Health Record: Lessons Learned from ClinGen and eMERGE. *Frontiers (minor revisions addressed; resubmitted to journal)*
    - Under development: NT213, NT270, NT272, NT352, NT342, NT277.1
- **Community**
  - **Working with HL7 Clinical Genomics group as development proceeds on a FHIR Genetic Result Resource based on our XML format**
  - We have consistently participated in panels and significant conferences to articulate our experience
    - Rasmussen L, Ames D, Aronson S, Babb L, Overby C. Design and Implementation of a Structured Sequencing Report Format: A Multi-Stakeholder Perspective from eMERGE. Panel: 2017 AMIA Annual Symposium.
    - Weng C, Murugan M, Freimuth RR, Aaronson S, Taylor CO. Panel: eMERGE EHR Integration Workgroup Panel on Lessons Learned and Future Directions. eMERGE Steering Committee Meeting. September 26, 2018.
    - Williams M, Freimuth R, Fiol G, Rasmussen L, Patel R, Dwight S. Panel: Moving Genomics into the Clinic: Informatic approaches from eMERGE, ClinGen, HL7 and GA4GH. 2018 AMIA Translational Informatics Summit.
    - Rasmussen L, Murugan M, Aronson S, Nestor J, Walton N. Panel: Operationalizing Innovation for the Return of Genetic and Genomic Results. 2019 AMIA Informatics Summit.
    - Murugan M, Taylor CO. Genomics Rendering in FHIR. *Federated Data Query Workshop.* Hosted by Johns Hopkins University in Baltimore, MD. May 20, 2019.
    - Zayas-Caban T, Pratap S, Taylor CO, Aronson S. Panel: Implementation of EHRs and Genomic Medicine in Diverse Communities. *Equity, Diversity, and Data Science in Genomics Workshop.* Hosted by the Carl R. Woese Institute for Genomic Biology at University of Illinois in Champaign, IL. September 4, 2019.

eMERGE EHRI WORKGROUP: Final Efforts & Milestone Completion

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

- **Status**
  - FHIR based version of the HL7 Clinical Genomics Resource will be open sourced
  - Workgroup continuing to track progress and share lessons across the group through:
    - Site Presentations
    - Milestone Tracker Updates
    - EHR and CDS Survey (preliminary results on last slide)

- **Key Challenges**
  - Support for Genetics/Genomics in the EHR remains highly underdeveloped and often inappropriately constructed
  - Genetics results transmitted outside of the eMERGE network are rarely transmitted in structured form
  - Numerous challenges remain including adequately managing knowledge on genetic results over time and integrating the combination of genetic results and knowledge into mainline clinical process flows

eMERGE EHRI WORKGROUP: Cross-Workgroup Collaboration Efforts

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

EHRI and ROR/ELSI workgroup

- Collaborative work in progress (NT270: Preferences for research updates among biobank participants)
  - Data analyses underway with data from two eMERGE institutions/affiliates (N>600)
- Collaborative work in progress (NT277.1: Operationalizing participant choices about genomic results: Beyond all or non ACMG recommended genes)

EHRI and Phenotyping workgroup

- Collaborative work in progress (NT342: Comorbidity Clusters in Clinical Conditions: An Analysis of Electronic Health Record Data) - Connection between clinical decision support and identifying disease severity.
- Publication from work to date:
  - Taylor CO, Lemke KW, Richards TM, Roe KD, He T, Arruda-Olson A, Carrell D, Denny JC, Hripcsak G, Kiryluk K, Kullo I. Comorbidity Characterization Among eMERGE Institutions: A Pilot Evaluation with the Johns Hopkins Adjusted Clinical Groups® System. AMIA Summits on Translational Science Proceedings. 2019;2019:145.

EHRI and PGx workgroup

- Collaborative work in progress (NT352: Lessons from eMERGE on readiness for genomic clinical decision support implementation)
- Common need to address the challenge of implementing genomic results in the EHR
- Presented during EHRI WG monthly meeting

*e*MERGE EHRI WORKGROUP: Phase III EHR Implementation Overview

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

- CDS based on genetic test results established or under development at all 10 sites
  - Possible gap in CDS used to identify patients that are eligible for genomic testing
- Successes with establishing optimized CDS for ACMG genes at 3 of 10 sites
- In many cases both the PDF and XML format is being stored in the EHR (5 of 10 sites).
  - Steps to link structured data from XML to CDS needs to be clarified

CDS Implementation results	Respondents (N=10*) N (%)	Comments
When *any* genetic testing is available		
Alert-based CDS and Infobutton	4 (40%)	Screening and CDS for FH and PGx, Infobuttons for FH and CRC
Alert-based CDS only	2, +3 under dev (50%)	Screening and CDS for FH, Inherited heart disease, PGx, Opioids; Long QT/Abacavir sensitivity/Carabemazepine sensitivity
Infobutton only	1 (10%)	Patient-level & cohort-level screening; Infobutton for summary review screen
No response/Not applicable	2 (20%)	
When results for ACMG genes available (highest maturity level indicated)		
Optimized	3 (30%)	e.g., studying usage and impact; Note: different maturity level indicated for different types of CDS
No response/Not applicable	7 (70%)	
eMERGE reports and/or XML stored in EHR?		
Yes	6 (60%)	
No response/Not applicable	4 (40%)	
Ingesting XML format into EHR?		
Yes	5 (50%)	
No	3 (30%)	Being used for research separate from EHR
No response/Not applicable	2 (20%)	

*Mt. Sinai, CCHMC, NU, KPW/UW, CHOP, Harvard, Mayo, Geisinger, Columbia, VUMC

eMERGE Genomics WORKGROUP: Major Accomplishments from Phase III

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

- Manuscript detailing the genotyping and imputation of ~84,000 eMERGE subjects has been published in Genetic Epidemiology (Stanaway et al; Genetic Epidemiology, 2019).
- The Genomics Workgroup (Sleiman) led supplement-funded efforts to link Geocoding data to eMERGE participants.
- The genomics workgroup provided guidance to the eMERGE CC regarding genetic data activities in order to produce four large multiple use discovery-based datasets
- The Genomics Workgroup established and conducted focus groups to help design the user interface and experience with the Electronic MEdical Records and GENomics Toolkit (eMERGENT), which will be built off the ideas surrounding the current SPHINX design. A U24 Genomics Community Resource proposal was submitted Spring 2019 (Crosslin – multi-PI).
    - The focus groups included collaboration with the EHR Integration and Clinical Annotation Workgroups to provide FHIR CDS Hooks via the eMERGENT resource.
    - Privacy security safeguards were also discussed.

eMERGE Genomics WORKGROUP: Final Efforts & Milestone Completion

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

- The Genomics Workgroup will continue to provide guidance to the eMERGE Coordinating Center regarding genetic data activities.
- Working with the eMERGE Coordinating Center, the Genomics Workgroup will continue to provide guidance and feedback as the eMERGE genetic and phenotype data are migrated over to the NHGRI Genomic Data Science Analysis, Visualization, and Informatics Lab-space (AnVIL).
- The Genomics Workgroup will continue to lead workshops to discuss the enhancement of SPHINX and ultimately eMERGENT should the U24 get successfully funded.
- The Genomics workgroup will provide feedback and guidance for polygenic risk score analyses and interpretation across the Network to support evolving research strategies in translational medicine

*e*MERGE Genomics WORKGROUP: Cross-Workgroup Collaboration Efforts

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

- Working with investigators Network-wide, the Genomics Workgroup is leading a manuscript to provide guidance with lessons learned through eMERGE with a central theme of balancing genomic discovery with implementation science. The title of the manuscript is *"Lessons from the eMERGE Network: Balancing genomic discovery and implementation science."*

- The Genomics Workgroup will leverage the eMERGE datasets and EHR-derived phenotypes to collaborate with other consortia, such as CSER, GIANT, and TOPMed. An example would be the creation and validation of polygenic risk scores.

- Once data are migrated to the AnVIL, the Genomics Workgroup will provide guidance and feedback regarding user experience, especially with analyses.

*e*MERGE OUTCOMES WORKGROUP: Major Accomplishments from Phase III

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

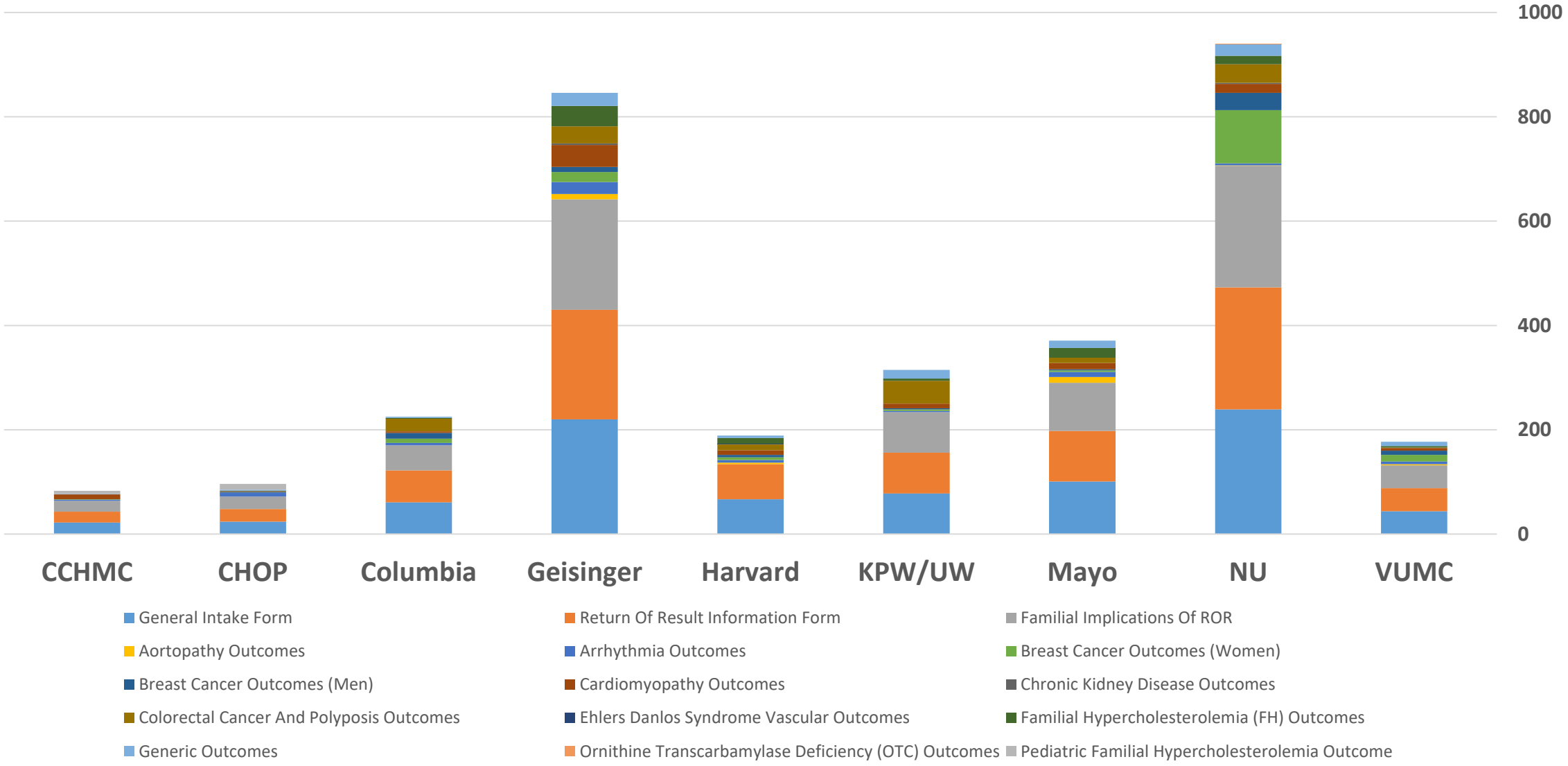
- The Outcomes workgroup developed an approach to collect discrete and harmonized process and clinical outcomes for all phenotypes associated with the returnable eMERGEseq variants including 15 forms for specific phenotypes and a generic outcome form for rare phenotypes.
- Developed 11 implementation guides for disease specific Outcomes data collection
- In collaboration with the ROR Workgroup, the Workgroup collected objective information about the ROR process and familial implications of ROR.
- As of September 2019, 796 participants have 6-month outcomes collected representing approximately 68% of the total returned cohort.
- A provisional data set consisting of all outcomes collected to date has been assembled and distributed to sites and MCS authors for data quality checks and preliminary analysis.
- Harmonization of Outcomes approaches between eMERGE and ClinGen explored in workgroup paper “Harmonizing Outcomes for Genomic Medicine”

Outcomes Collection Instruments

- General Intake Form
- Return of Result Information Form
- Familial Implications of ROR Form
- Aortopathy Outcomes
- Arrhythmia Outcomes
- Breast Cancer Outcomes – Men
- Breast Cancer Outcomes - Women
- Cardiomyopathy Outcomes
- Chronic Kidney Disease Outcomes
- Colorectal Cancer and Polyposis Outcomes
- Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) Outcomes
- Ehlers Danlos Syndrome – Classical Outcomes
- Ehlers Danlos Syndrome – Vascular Outcomes
- Familial Hypercholesterolemia (FH) Outcomes
- Ornithine Transcarbamylase Deficiency (OTC) Outcomes
- Pediatric Familial Hypercholesterolemia Outcomes
- Tuberous Sclerosis Complex (TSC) Outcomes
- 22q Deletion/Duplication Syndrome Outcomes
- Generic Outcomes (for conditions not listed above)

eMERGE OUTCOMES WORKGROUP: Data collection progress by site

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



eMERGE OUTCOMES WORKGROUP: Final Efforts & Milestone Completion

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

- A second data freeze is planned for January 2020 representing a complete 6-month data set and potentially final 12-month data set
- MCS and manuscripts for individual phenotypes: CDC Tier 1 conditions, selected conditions such as cardiomyopathy with sufficiently high N to be finalized
- Harmonizing process and clinical outcomes across individual outcomes forms to report a summarized eMERGEseq impact on participants.
- Collaborative efforts with ROR and Clinical Annotations

Number of Outcomes Forms across 796 participants	
Aortopathy	27
Arrhythmia	63
Breast Cancer – Women	154
Breast Cancer – Men	72
Cardiomyopathy	106
Chronic Kidney Disease	7
Colorectal Cancer and Polyposis	159
Ehlers Danlos Syndrome Vascular	1
Familial Hypercholesterolemia	94
Generic Phenotype	96
Ornithine Transcarbamylase Deficiency	1
Pediatric Familial Hypercholesterolemia	16

eMERGE PGx WORKGROUP: Major Accomplishments from Phase III

Co-Chairs: Laura Rasmussen-Torvik (Northwestern) & Cindy Prows (CCHMC)

Determine e3 PGx return of results processes

- Collaborated with central labs to determine which genes & star alleles to return and adaptations for different reporting strategies (included in individual reports vs. batched reports)
  - Included in Harmonization paper MCS NT244, https://doi.org/10.1016/j.ajhg.2019.07.018

Examine cross-site PGx implementation process

- Monitored ROR progress & gathered data about specific challenges of PGx implementation into EHR https://docs.google.com/spreadsheets/d/1wJG_HqELiIsZSCZ3VEQanFDOYrgiNFb97CmSXMJyt1g/edit#gid=0
- eIII PGx results returned to participants to date at Vanderbilt, Marshfield & planned at Northwestern
- eIII PGx CDS activated at Geisinger, Marshfield, Vanderbilt and planned at Northwestern
- PGx plans vs. PGx actual ROR included in ROR manuscripts NT273, parts 1 (Georgia Weisner lead) and 2 (Kathy Leppig lead)

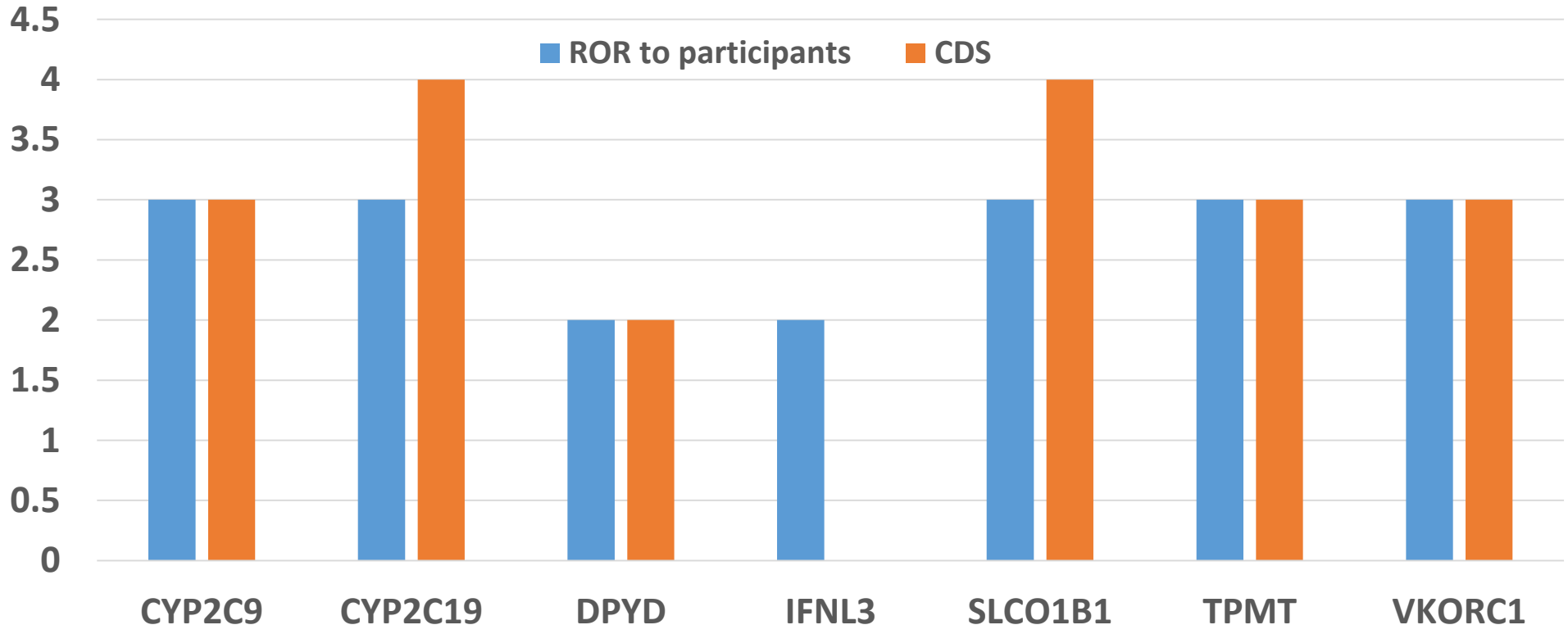
Coordinate & promote pharmacogenomic discovery

- Types of variation across genes on the PGRNSeq platform by A. Gordon under review at Pharmacogenetics and Genomics
- Numerous papers in process
  - Details here: https://docs.google.com/spreadsheets/d/1AKRY-RDWngzyAZn0UhCNXHAo1Ozq9I69S9FZnhoxpok/edit#gid=1606396516

*e*MERGE PGx WORKGROUP: Major Accomplishments from Phase III

Co-Chairs: Laura Rasmussen-Torvik (Northwestern) & Cindy Prows (CCHMC)

Graph: Number of sites returning or planning to return PGx results (below)



eMERGE PGx WORKGROUP: Final Efforts & Milestone Completion

Co-Chairs: Laura Rasmussen-Torvik (Northwestern) & Cindy Prows (CCHMC)

- Initiated collaborations with CPIC
  - eMERGE prominently featured at CPIC meeting in June 2019
- Responded to challenges with PGx implementation
  - Met with EHRI workgroup to discuss best approaches to communicate lessons learned
  - Multiple workgroup members now participating in EPIC's "Genomic Braintrust" to offer feedback about new genomics indicator in EPIC
    - Sites with Genomic Indicators module are largely relying on a manual process to enter results
- Explored feasibility of project(s) that could contribute evidence for CPIC guideline(s) that contain recommendations as "optional" or "not enough data"
- Supporting NT 335, Evaluating the 'Star Allele' PGx nomenclature standard in the context of automated interpretation of panel, exome and genome sequencing results
  - Genomic data obtained and phasing completed
  - Responsive to most recent CPIC meeting (and PharmVar)
- Determine if any additional PGx outcomes are feasible
  - Outcomes forms focused on disease risk per network wide prioritization.

Continuing Efforts

- Continued discussions about best ways to communicate lessons learned with respect to EHRI for PGx results
  - Scientific papers?
  - Interaction with specific companies?
  - Presentations in other forums
  - *Collaborating with EHRI in these discussions*
- Continue to monitor and update
  - Which genes returned at which site on what timeline
  - When results are returned and to whom (patient, EHR, provider)
  - When CDS is turned on
  - Unanticipated challenges

eMERGE PHENOTYPING WORKGROUP: Major Accomplishments from Phase III

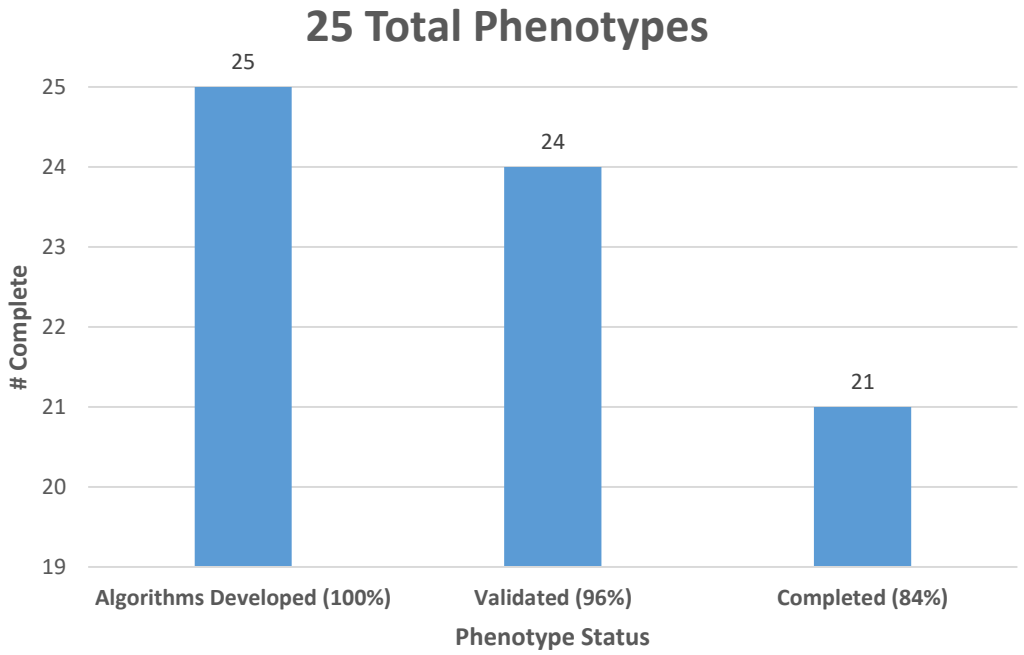
Co-Chairs: Wei-Qi Wei (Vanderbilt) & Chunhua Weng (Columbia)

Primary goals

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

Phenotype Development & Implementation

- *e I & II phenotypes (March 2017) – 13 (100%) complete*
- *e III phenotypes (September 2019) - see graph on the right*
- *Natural Language Processing (NLP; September 2019) – 2 (40%) developed*
- *Common Variables collected by CC to use with all studies*
  - ICD/CPT codes
  - BMI/Weight/Height
  - Phecodes
  - Medications (Statins)
  - Labs (HDL, LDL, Total Cholesterol, Glucose, Triglycerides, WBC differentials, Autoimmunity)
- *Common Variables refreshed in August 2019*



Data Standardization Efforts

- *OMOP: Converted the EHR data of eMERGE cohort to OMOP*
- *Phecode: Expanded into both ICD 9 and 10 versions*

New Publications:

- Zhao et al. "Detecting Time-Evolving Phenotypic Topics via Tensor Factorization on Electronic Health Records: Cardiovascular Disease Case Study." *Journal of Biomedical Informatics*.

Top Publications:

- Hripcsak et all. "Facilitating phenotype transfer using a common data model." *Journal of Biomedical Informatics.*
- Hripcsak et all. "Effect of vocabulary mapping for conditions on phenotype cohorts." *Journal of the Medical Informatics Association.*
- Namjou et al. "GWAS and enrichment analyses of non-alcoholic fatty liver disease identify new trait-associated genes and pathways across eMERGE Network". *BMC Medicine*
- Carrell et al. "The machine giveth and the machine taketh away: a parrot attach on clinical text deidentified with hiding in plain sight." *Journal of the Medical Informatics Association.*
- Levine ME et al. "Methodological variations in lagged regression for detecting physiologic drug effects in EHR data". *Journal of Biomedical Informatics .*
- Zhao J et all. "Learning from Longitudinal Data in Electronic Health Record and Genetic Data to Improve Cardiovascular Event Prediction." *Nature Scientific reports.*
- Schuemie MJ et al. "Improving reproducibility by using high-throughput observational studies with empirical calibration." *Philosophical Transactions of the Royal Society A.*

*e*MERGE PHENOTYPING WORKGROUP: Final Efforts & Milestone Completion

Co-Chairs: Wei-Qi Wei (Vanderbilt) & Chunhua Weng (Columbia)

Future Efforts:

- Finalizing validation and implementation of eIII algorithms
- Developing and implementing the selected five NLP phenotypes
  - Plan for release
    - Long QT- Arrythmias & Lupus & (Autumn 2019)
    - ACO- Asthma/COPD Overlap & Familial Hypercholesterolemia *or* Chronic Rhinosinusitis (Winter 2019)
    - Familial Hypercholesterolemia *or* Chronic Rhinosinusitis (Spring 2020)
- Developing NLP-based eIII lessons learned manuscript concept sheet (Spring 2020)

Challenges:

- Data quality issues (e.g., BMI data plausibility)
- Lack of shared representative data for generalizable algorithm development
  - PPV from the second-round validation remains low
- Lack of contextual knowledge of text across institutions to achieve the portability of NLP algorithms

*e*MERGE ROR-ELSI WORKGROUP: Major Accomplishments from Phase III

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

- Summary of the primary goals of the Workgroup
  1. Describe challenges in returning results (positive and negative) in large scale sequencing projects.
  2. Explore the challenges involved in identifying at-risk family members and informing them of their potential risk and collect responses of the family members.
  3. Estimate the institutional impact of RoR by developing guidelines for ROR and for IRB and consent language.
  4. Disseminate lessons learned on the various aspects of genomic medicine implementation by activities such as publishing articles that propose the key elements for effectively returning genomic results to providers and patients and comparing the impact different methods of RoR have on patient and physician care across all sites.
- Major accomplishments of the Workgroup from the last four years (Phase III)
  - RoR completed at all sites
  - ROR Processes study – manuscript submitted *"Returning genomic results to eMERGE participants: The who, what, where, and how of disclosure"*
  - ROR lessons learned study – manuscript in process *"The Reckoning: What We Found After Return of Results for 25,000 eMERGE3 participants"*
  - Participant surveys – coordinated across sites, site-specific efforts (interviews, surveys). Data dictionaries reconciled and data being placed in one REDCap database at CC. 4 concept sheets proposed
  - HCP surveys and interviews – R01 funded. Surveys of 141 HCP completed. Concept sheets proposed. Qualitative interviews of HCP receiving positive results (8) and negative results.
- Top/relevant publications, including those on Workgroup lessons learned and special issues
  - Fossey, et. al., Ethical Considerations Related to Return of Results from Genomic Medicine Projects: The eMERGE Network (Phase III) Experience, J Pers Med. 2018
  - Pet, et. al Physicians' perspectives on receiving unsolicited genomic results, *Genet Med*. 2018

eMERGE ROR-ELSI WORKGROUP: Final Efforts & Milestone Completion

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

- Updates on remaining Workgroup efforts and projects

Lead	Title
Georgia Wiesner	Returning genomic results to eMERGE participants: The who, what, where, and how of disclosure
Christin Hoell & Cindy Prows	Operationalizing participant choices about genomic results: Beyond all or none ACMG recommended genes (patient results)
Luke Rasmussen	Operationalizing participant choices about genomic results: Beyond all or none ACMG recommended genes (technical)
John Lynch	Understanding the return of results process: Content review of patient summary letters
Kathleen Leppig	The Reckoning: What We Found After Return of Results for 25,000 eMERGE3 participants
Colin Halverson	Challenges in Returning Results in the eMERGE consortium
Ingrid Holm	Approaches to the return of actionable adult-onset conditions in pediatric research: Lessons learned from eMERGE 3
Richard Sharp & Maureen Smith	Network-wide lessons learned from the reporting of negative test results
Iftikhar Kullo and David Kochan	Sequencing Centers and eMERGE Site Interactions related to Return of Genomic Results in Phase III of the eMERGE Network
Hila Milo Rasouly and Julia Wynn	Family communication following return of positive results
Ingrid Holm	Impact of results on participants - Partiicpant survey workgroup
Ingrid Holm	Utility of results - Partiicpant survey workgroup
Ellen Clayton and Ingrid Holm	Privacy and Confidentiality - Participant survey workgroup

- Plans for completion of outstanding projects
  - Work groups formed
- Updates on challenges that the Workgroup has/is facing in nearing end of eMERGE III
  - Assessing impact of different processes across sites on ROR
  - Non responders, Decliners, deceased participants, transition to adulthood
  - Lab: Gender mismatch, mosaicism, genotype-phenotype mismatch, variant reclassification,
  - Non-responder definition, total #, plan for RoR, assessing outcomes and penetrance in non-responders, Lessons learned

*e*MERGE ROR-ELSI WORKGROUP: Cross-Workgroup Collaboration Efforts

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

- Determine the impact of RoR on patients' outcomes immediately, 6 months and/or 12 months after RoR
  - Processes of care, clinical utility, family utility, provider utility, psychosocial factors.
- ROR Information form and Familial Implications of ROR form of the Outcomes Protocol Database collects ROR and family implications before collecting outcomes and penetrance: collaboration with Outcomes and Clinical Annotation workgroups
- MeTree pre-implementation: collaboration with EHRI and Outcomes workgroups
- "Operationalizing participant choices about genomic results: Beyond all or none ACMG recommended genes" collaboration with EHRI workgroup

*e*MERGE Lessons Learned Panels

*e*MERGE LESSONS LEARNED: Impact of ROR on downstream analysis

Workgroups: Clinical Annotation, Return of Results, & Outcomes

ESP Meeting: October 4th 10:40-11:50 a.m.

- Informing and harmonizing variant interpretation| Heidi Rehm (Partners/Broad), Iftikhar Kullo (Mayo), & Adam Gordon (Northwestern),
- Return of results pathways, barriers, and harmonizing across sites |Ingrid Holm (BCH) & Iftikhar Kullo (Mayo)
- Impact of ROR process on Outcomes assessment | Josh Peterson (VUMC/CC), & Marc Williams (Geisinger)

eMERGE EHRI WORKGROUP: Lessons Learned

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

Panel: Lessons learned panel conducted during the October 25, 2018 Steering Committee meeting in Bethesda, MD.

Goals: Discuss implementing clinical decision support. Discuss communication of risk information to family members and cascade screening

Summary: The EHRI group discussed obstacles and lessons learned during the integration of clinical data into a variety of EHRs.

Obstacles:

- The EHR teams at local sites have competing projects and time allocations, EHR integration of eMERGE data needed to be woven into the queue.
- Transitions to new EHRs occurred at several sites, which caused delays and even more intense competition for resources.
- Large teams with asynchronous communication and changing personnel caused setbacks.
- Compliance regulations from some states caused issues with data usage and return.

eMERGE EHRI WORKGROUP: Lessons Learned

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

eMERGE Experience: It was feasible to build a unified clinical network linking heterogeneous laboratories and provider systems in the context of an NIH consortium, although was not simple. The group will work to inform the genetic HIT standards by developing a FHIR profile that codifies all of the combined experience.

Overarching Lessons Learned:

- Data standardization and harmonization is key when returning genomic test results to a variety of clinical sites. Mechanisms to track, analysis tools, and manage data (including for genetic variant reclassifications) are needed for effective integration of results in the EHR.
- Genetic aware clinical decision support should drive off of a variant knowledge base and requires access to structured data and knowledge. This process should not be hand coded. Designing and maintaining such a knowledgebase also requires tight collaboration between clinicians, laboratories, and IT professionals.
- EHR integration of genomic test results at each site requires an oversight process for what and how content is presented to clinicians, including understanding where in the healthcare setting to make data interpretations, and clinical and patient guidance accessible.
- Creating a standard data flow pipeline is key to integrating genomic test results into the EHR. This pipeline will differ depending on site regulations, study design, and requirements.

eMERGE GENOMICS WORKGROUP: Lessons Learned

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

Panel: Lessons learned panel conducted during the June 21st, 2019 Steering Committee meeting in Seattle, WA.

Goals: To examine lessons learned in the Genomics group specifically surrounding creation and compilation of large data sets, analyses, and timing of data release in a large consortium.

Summary: The eMERGE Network produced several large, rich datasets including array sets focused on genomic discovery and sequencing datasets focused on implementation science. However, this production requires significant time and money. Networks should clearly outline the analysis and product goals prior to compiling a large dataset, including diversity and phenotypic status of samples if that is an important component in analysis. Analysis and computation costs are significant in large datasets and this should be considered when considering hosting data on cloud computing services. Cloud computing can be beneficial for analyses pipelines as it can be used to optimize processes and management.

Obstacles:

- Adding additional samples to datasets after data freezes have been released, caused delays in analysis and came at a significant cost to resources
- Though the size of eMERGE's datasets are a strength, working with these multi-terabyte files requires time and resources, and this will be come increasingly important as data analyses are moved onto the cloud computing environment.
- Early analyses were postponed due to lack of phenotypic and case/control data early on in the Network cycle, and the recurring promise of a 'new dataset' to be released in the future.

eMERGE GENOMICS WORKGROUP: Lessons Learned

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

Obstacles (continued):

- With the focus on the eMERGEseq dataset, the Network did not prioritize analyses on past datasets, like PGRNseq, that are still a rich source of discovery.
- As data were added over the course of many phases of eMERGE, naming convention of individual site GWAS and even PGRNseq files were not consistent, which required additional time and effort in order to combine and clean the files.

Overarching Lessons Learned:

- Demographic files should be collected prior to genetic data compilation and datasets frozen at that time.
- Adding and removing participants once a large dataset is compiled costs significant amounts of time and resources.
- Clearly defined data freezes which take into account diversity, phenotypic data, and discovery & implementation goals should be outlined at the beginning of the network to maximize data delivery and analysis time at the sites.
- Standard naming conventions are necessary when trying to combine files from multiple sequencing centers will maximize efficiency and turn around time.
- Cloud computing can be used to set up standard pipelines for analysis, saving time, resources, and improving consistency.

Going forward: Future networks should be clear about their goals, and when decisions have to be made, they should limit the amount of time for comments in order make a final ruling on a given decision. Likewise, the workgroups should maximize participation in decision making by having clear question and surveys. The group will also work through how to move genomic & phenotypic data to the AnVIL platform, including data on DNA Commons.

eMERGE OUTCOMES WORKGROUP: Lessons Learned

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

Panel: Lessons learned panel conducted during January 18, 2019 Steering Committee meeting in Bethesda, MD.

Goals: To examine lessons learned during the creation and implementation of Outcomes forms across sites after return of actionable variants in the eMERGEseq panel. The PDFs of the final forms can be found here.

Summary: The Outcomes panel discussed general lessons learned from creation of the Outcomes forms, initial findings and considerations from a subset of adult participants (Mayo) and pediatric participants (CCHMC & CHOP). In general, to capture the breath of outcomes across the eMERGEseq panel, deciding what data elements for a given disease was a difficult task, especially across diverse populations. For the pediatric participants, the child's preference to receive actionable results sometimes differed from their parents or guardians, this in addition to having to re-consent the participant if they turned 18 years old during the course was a main difference when comparing to other cohorts in which only adults were enrolled.

Obstacles:

- Harmonizing the Outcomes forms across the sites was difficult, as some issues did not present until data entry commenced. Early development and use of abstraction guides are an important element to determine how sites should interpret Outcomes forms questions.
- Adding in 'penetrance' related questions to the Outcomes forms was inefficient as the Outcomes forms were not originally designed to ascertain penetrance related data elements. This also caused a delay in launching the forms and data entry across the Network.
- Penetrance data elements were required for all actionable results, however initially Outcomes forms were only to be completed on participants where return of results took place. Future studies may consider creating penetrance only forms to be filled out in parallel.
- Site-hosted Outcomes forms caused too much variation in data elements and would have made data compilation very difficult during the initial Network-wide analysis. It was necessary to move the Outcomes forms to a central location, hosting by the Coordinating Center.
- Sites had differing IRB requirements when it came to limited versus de-identified data entry, de-identification of the dates by shifting all dates for a given record a set number of days (date shifting) was required when filling out the forms for some sites.

eMERGE OUTCOMES WORKGROUP: Lessons Learned

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

Going Forward: The Outcomes group will use the Coordinating Center (CC) hosted REDCap instances of the approved and deployed Outcomes forms to complete the six-months outcomes data. Sites will enter additional penetrance data as needed for appropriate forms. An interim six-month outcomes data analysis is scheduled for October, 2019, which should be more streamlined with the consolidation of the forms to a single REDCap instance. A general Outcomes lessons learned paper on process and intermediate health related outcomes framework was published in the BMC Journal collection (Williams et al., 2018).

Overarching Lessons Learned:

- The Network achieved broad coverage of applicable phenotypes with some sacrifice in the depth of outcome phenotyping
- Context is important to understand the changes (or lack thereof) in health services delivered
- Context is difficult to uniformly capture across a large cohort; complementary qualitative assessments may be critical
- Pediatric cohorts offer the potential for longitudinal studies in the future.

*e*MERGE PHENOTYPING WORKGROUP: Lessons Learned

Co-Chairs: Chunhua Weng (Columbia) & Wei-Qi Wei (VUMC)

Panel: Lessons learned panel conducted during the June 25th, 2018 Steering Committee meeting in Cincinnati, OH.

Goals: To demonstrate the challenges faced during phenotype development and how eMERGE solved the issues or plans to solve the challenges. To discuss how has the nature of phenotyping evolved during the eMERGE program. To describe what has worked well (machine-learning, NLP, etc) during phenotype development.

Summary: The Phenotyping group catalogued issues that cause delays and difficulty during algorithm development and implementation as well as potential solutions or 'lessons learned' to these obstacles.

Obstacles:

- Logic, complexity of logic, number of data elements, and modalities of data all alter complexity of the phenotype.
- Complexity of the algorithm and scientific question calls upon a select set of individuals to develop and validate, some of which are hard to schedule due to clinical commitments.
- Data Dictionaries can also add complexity, time, and effort.

eMERGE PHENOTYPING WORKGROUP: Lessons Learned

Co-Chairs: Chunhua Weng (Columbia) & Wei-Qi Wei (VUMC)

Going Forward: The Phenotyping group will continue to catalogue complexities of algorithm development and implementation as well as publish lessons learned. Moving forward, incorporation and streamlining of natural language processing and transition to the OMOP common data model will act as 'experiments of nature' to compare to previous implementation methods.

Overarching Lessons Learned:

- Strong project management is needed to keep queues organized, projects assigned, and issues resolved at both the Network and Site level.
- Algorithms as flowcharts are most effective, direct codes do not port well currently.
- Better understanding and cataloguing the complexity of an algorithm allows for better planning.
- Local experts are needed to implement and review depending on complexity of the science.
- Adopting a common data model and common vocabularies can facilitate implementation and data transfer.
- Centralizing commonly used data elements saves programmer time.

eMERGE ROR WORKGROUP: Lessons Learned

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

Panel: Lessons learned panel conducted during the January 25th, 2018 Steering Committee meeting in Bethesda, MD.

Goals: To highlight each site's return of results process from receipt of clinical reports to delivery to the participant. To discuss variability among sites in the process of returning CLIA validated results to participants & providers

Summary: eMERGE sites represent a spectrum of return of results, this allows for a well rounded ability to understand how differences affect overall return. Allows for overarching analysis of how methods affect patient comprehension, engagement, and outcomes.

Obstacles:

- Differences in the ROR process across sites, although allowing for experiments of nature, provides challenges to studying the impact of ROR.
- Coordinating the impact of ROR on Health Care Providers across sites as sites have different processes for who returns the results and to whom results are sent.
- Coordinating the participant survey across sites is challenging, given the different populations at each site and the different site priorities regarding their focus of research.

eMERGE ROR WORKGROUP: Lessons Learned

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

Going Forward: The ROR group will continue to collect and analyze data in collaboration with Outcomes and EHRI working groups and publish on how differences in participant, provider, and institutional involvement effects the overall return process.

Experiments of 'nature':

- Cohorts un-selected vs selected for a particular trait - One site selecting based on genotype.
- Choice vs no choice for return of secondary findings.
- Negative results returned vs not returned
- Variation in timing of placement of results in EHR.
- Randomization vs observational
- Pediatric vs adult.

Overarching Lessons Learned:

- The ROR process is dependent on institutional cultures and priorities and thus it is difficult to create standard guideline for practice.
- IRBs vary significantly in their requirements, processes, and views towards ROR from genomic sequencing.

*e*MERGE SEQUENCING CENTER HARMONIZATION: Lessons Learned

Co-Chairs: Richard Gibbs (Baylor), Heidi Rehm (Partners/Broad), & Niall Lennon (Partners/Broad)

Panel: Lessons learned panel conducted during the October 26, 2018 Steering Committee meeting in Bethesda, MD.

Goals: To examine lessons learned during the harmonization of sequencing centers across a network for variant return to participants as well as data usage for research. The paper ''*Harmonizing Clinical Sequencing and Interpretation for the eMERGE III Network*" has been published by the *American Journal of Human Genetics (AJHG) as of August 2019 (PMID 31447099)*.

Summary: The CSGs discussed issues and lessons from the harmonization process the Network conducted during the creation and development of the eMERGEseq panel.

Obstacles:

- Items to be harmonized included data from the collection sites, assay development, test validation, primary analysis, variant classification, report content, data delivery to sites, and progress reporting to the Network including detection rates for different content and tested populations.
- Harmonization during the first year or so caused the rate of sequencing and reporting to be slow. As details were agreed upon, the rate of sequencing and reporting increased rapidly.
- Creating the panel itself slowed progress as well. Sites submitted their top six requested genes and those were combined with the ACMG 56 list. The site genes and SNVs required clinical reporting criteria to be assigned. A consensus list of returnable content was agreed upon but most sites added or subtracted content based on site-specific consent and protocols for return of results.
- A process for notification of sites when a reported variant was reclassified was developed. A more systematic approach to reanalysis is now underway.

eMERGE SEQUENCING CENTER HARMONIZATION: Lessons Learned

Co-Chairs: Richard Gibbs (Baylor), Heidi Rehm (Partners/Broad), & Niall Lennon (Partners/Broad)

- **Going Forward:** The CSGs will plan to develop harmonized structured genetic test report standards compliant with FHIR/HL7. The CSGs will also examine the triggers and frequency of reanalysis and reinterpretation issues. Establish pipelines for return of updated results going forward is an important next step.

Overarching Lessons Learned:

- The CSGs worked together and with the Clinical Annotation group to come up with a consensus set of clinically actionable genes and SNVs that would be reported on. Sites were allowed to request additional genes or SNVs that were included on the panel to also be reported on in site-specific reports if clinical validity was met.
- The CSGs worked together to ensure the development and validation of the eMERGEseq panel was concordant, this required two rounds of probe design and validation resulting in 99.8% (Partners/Broad) and 99.9% (Baylor) coverage of bases.
- CSGs communicate share variant interpretations monthly to resolve any discordant variant interpretations Classified variants are submitted to ClinVar.
- A class of non-reportable variants (VUS-leaning pathogenic) were identified as targets for follow-up in case eMERGE clinical data may be able to move the variant into the Likely Pathogenic reportable range.

eMERGE Resources & Tools

Resources

- eMERGE website: www.gwas.org
- eMERGE manuscripts to date: https://emerge.mc.vanderbilt.edu/publications/
- eMERGE studies currently submitted to, and/or accessible in, dbGaP: https://emerge.mc.vanderbilt.edu/dbgap/
- eMERGE data platform information: https://emerge.mc.vanderbilt.edu/wp-content/uploads/2015/02/Platform-Information-eMERGE.docx

Tools

- PheWAS Catalog: Functions as a platform for analysis of phenotypes against a single gene variant, https://phewascatalog.org
- Phenotype Knowledgebase (PheKB): Offers a collaborative environment to build and validate electronic algorithms to identify characteristics of patients within health data, https://phekb.org/
- eMERGE SPHINX: Operates as a tool for exploring data for hypothesis generation, especially around drug response implications of genetic variation across the eMERGE PGx cohort, https://www.emergesphinx.org/
- CDS Knowledgebase (CDSKB): Catalogs and shares clinical decision support implementation artifacts and designs consideration for genomic medicine programs, https://cdskb.org/
- DocUBuild: web application for creating and sharing documents that can be accessed electronically from Electronic Health Record (EHR) systems, https://docubuild.fsm.northwestern.edu/