emerge network
ELECTRONIC MEDICAL RECORDS & GENOMICS

External Scientific Panel

Background Materials

October 7, 2016

BCM Baylor College of Medicine
BROAD INSTITUTE
Cincinnati Children's Hospital Medical Center
The Children's Hospital of Philadelphia
COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK
GEISINGER HEALTH SYSTEM
GroupHealth
HARVARD UNIVERSITY
MAYO CLINIC
Northwestern Medicine
PARTNERS HEALTHCARE
VANDERBILT UNIVERSITY MEDICAL CENTER
UW Medicine UNIVERSITY OF WASHINGTON MEDICAL CENTER
National Human Genome Research Institute

TABLE *of* CONTENTS

eMERGE Citation Analysis..... ***Pg. 2***

eMERGE Publications..... ***Pg. 3***

Overview of eMERGE Tools..... ***Pg. 8***

eMERGE Phase III Phenotypes..... ***Pg. 9***

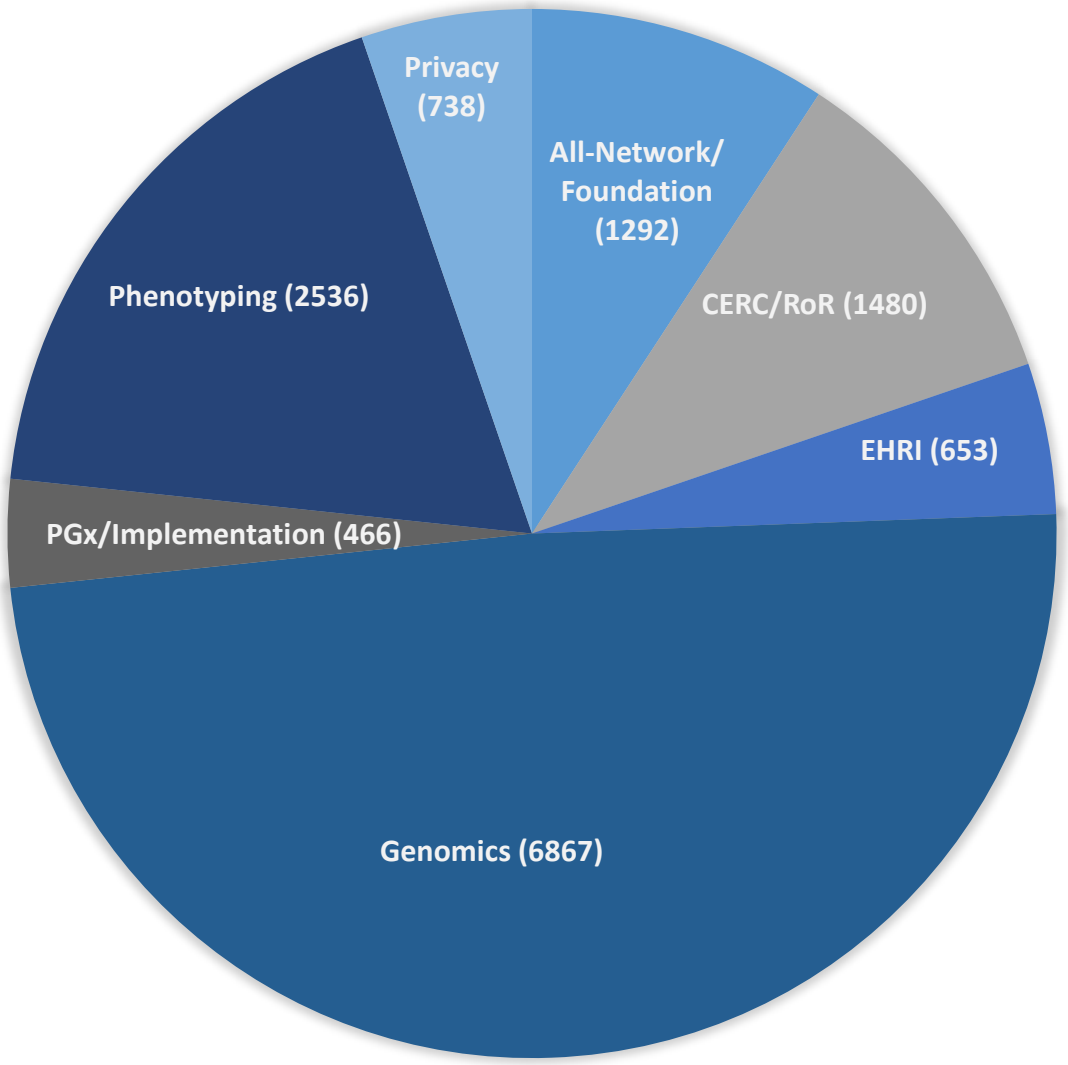
Minutes from ESP Meeting, Feb 2016..... ***Pg. 10***

Minutes from ESP Meeting, Feb 2016, Executive Session..... ***Pg. 11***

Minutes from Steering Committee Meeting, May 2016..... ***Pg. 13***

*e*MERGE CITATION ANAYLSIS

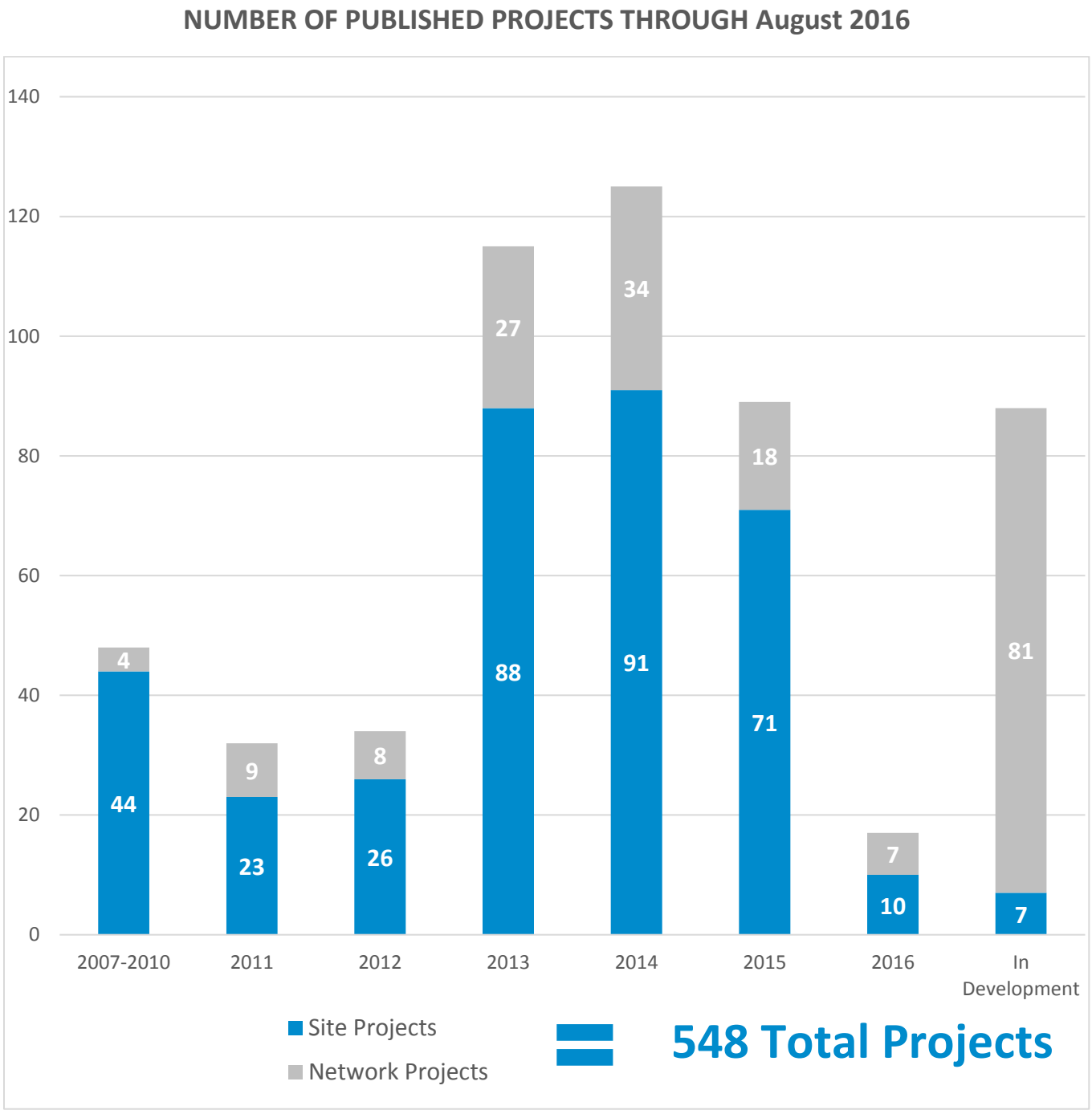
Citations of eMERGE Publications by Category



Cumulative Citation Count

2007- April 2016: 14,032

*e*MERGE PUBLICATIONS



*e*MERGE PUBLICATIONS

from February 2016 – October 2016

Digital Reference Library Available Here

Published/Accepted and Submitted Network Manuscripts

1. Kho A. Thompson W, Pacheco J, Muthalagu A, Roberts et al *GWAS of Infection or Colonization with Community Associated Methicillin-Resistant Staphylococcus Aureus (CA-MRSA)*. (Submitted)
2. Heit J, de Andrade M. *GWAS of Venous Thromboembolism (VTE) among African-Americans.* (Submitted)
3. Ritchie M, Holzinger E, Verma S, Farrall M, Drenos F et al. *Discovery and replication of genetic interactions for quantitative lipid traits.* (Submitted)
4. Peissig P, Schwei K, Kadolph C, Finamore J, McCarty C et al. *Development of a dynamic XML event-driven ophthalmologic data capture framework.* (Submitted)
5. Lingren T, Thaker V, Lingren T, Kennebeck S, Namjou B et al. *Pediatric Providers are Poor at Identifying Severe Obesity in Young Children at Two Tertiary Pediatric Medical Centers.* (Accepted)
6. Edwards T, Torstenson ES, Gilbertson J, Tsosie K, Giri A et al. *MVtest: a method to flexibly model the genetic determinants of trait variability* (Submitted)
7. Verma S, Bailey J, Lucas A, Bradford Y, Linneman J, et al. *Epistatic gene-based interaction analyses for glaucoma in eMERGE network and NEIGHBOR consortium.* (Accepted)
8. Verma A, Verma SS, Pendergrass SA, Crawford DC, Crosslin DR, et al. *eMERGE Phenome-Wide Association Study (PheWAS) identifies clinical associations and pleiotropy for stop-gain variants.* BMC Medical Genomics. 2016;9(1):19–25.
9. Jones G, Kuivaniemi H, Tromp G, Carey DJ, Smelser D et al. *Meta analysis of genome-wide association studies for abdominal aortic aneurysms*. (In Press).
10. Ritchie M, Lucas A, Verma S, Kim D, Crosslin D. et al. *Copy number variation burden analysis on a range of phenotypes in the eMERGE network.* (In Press).
11. Simonti CN, Vernot B, Bastarache L, Bottinger E, Carrell DS, et al. *The phenotypic legacy of admixture between modern humans and Neandertals.* Science. 2016 Feb 12;351(6274):737–741. PMID: 26912863
12. Kirby J, Speltz P, Basford M, Roden D, Haines J et al. *PheKB: A Catalog and Workflow for Creating Electronic Phenotype Algorithms for Transportability.* J Am Med Inform Assoc. 2016 Mar 28; PMID: 27026615
13. Bush WS, Crosslin DR, Owusu-Obeng A, Wallace J, Almoguera B, et al. Genetic variation among 82 pharmacogenes: The PGRNseq data from the eMERGE network. Clin Pharmacol Ther. 2016 Aug;100(2):160–169. PMID: 26857349Bielinski SJ, Carrell DS, Connolly JJ, Doheny KF, Gordon AS et al. *Genetic variation among 84 pharmacogenes from the PGRNSeq in the eMERGE Network.* Clin Pharmacol Ther. 2016 Feb 9. PMID: 26857349
14. Lingren T, Thaker V, Brady C, Namjou B, Kennebeck S, et al Developing an Algorithm to Detect Early Childhood Obesity in Two Tertiary Pediatric Medical Centers. Appl Clin Inform. 2016;7(3):693–706. PMID: 27452794
15. Lingren T, Chen P, Bochenek J, Doshi-Velez F, Manning-Courtney P, et al. Electronic Health Record Based Algorithm to Identify Patients with Autism Spectrum Disorder. PLoS ONE. 2016;11(7):e0159621. PMID: 27472449

In Process Network Manuscripts

**Manuscripts proposed in Phase III*

1. PhEMA phenotyping authoring tool validation with eMERGE BPH case algorithm as use/test case. Lead Investigator: Jen Pacheco (NU)*
2. Linking 25,000 eMERGE participants with highly-accurate imputed HLA regions to electronic health records. Lead Investigator: David Crosslin (GHC/UW)*
3. Establishing Electronic Genetic Report Flow Within the eMERGE Network to Enable Genomic Clinical Decision Support. Lead Investigator: Samuel Aronson (Partners/Broad)*
4. PsycheMERGE. Lead Investigator: Jordan Smoller (Harvard)*
5. Genomics of structural kidney and urinary tract defects. Lead Investigator: Miguel Verbitsky (Columbia)*
6. PheWAS for functional variants in the complement system. Lead Investigator: Krzysztof Kiryluk (Columbia)*
7. MEdicine Gene Annotation (MEGA): A REDCap based tool to support consensus variant interpretation. Lead Investigator: Wayne Liang (GHC/UW)*
8. Common and rare variation associated with valvular disease in eMERGE 3. Lead Investigator: Laura Rasmussen Torvik (NU)*
9. Common and rare variation associated with headache in eMERGE 3 Lead Investigator: Laura Rasmussen Torvik (NU)*
10. Linking biomarkers to clinical phenotypes based on underlying genetic risk. Lead Author: Jonathan Mosley (VU)*
11. Broad consent and data sharing in biobank research: An eMERGE Network Study of Parent Perspectives . Lead Investigator: Armand Antommaria (CCHMC)*
12. A Phenome-wide Survey of the Phenotypic Effects of High- Frequency Human-Derived Alleles. Lead Investigator: Corrine Simonti (VU)and Tony Capra (VU)*
13. Pharmacogenetic variation identified via targeted next-generation sequencing among 9000 eMERGE subjects. Lead Investigator: Adam Gordon (GHC/UW)*
14. Comprehensive genetic association study of kidney traits across the EMERGE network. Lead Investigator: Nikol Mladkova (Columbia)*
15. Detection of copy number variants (CNVs) and their kidney disease associations across the EMERGE network. Lead Investigator: Miguel Verbitsky (Columbia)*
16. Combined GWAS-PheWAS Approach to Serologic Markers of Autoimmunity & Inflammation. Lead Investigator: Nikol Mladkova (Columbia)*
17. Knowledge driven rare variant PheWAS in eMERGE to identify regions associated with disease using collapsing based approach. Lead Investigator: Anna O Basile (Geisinger)*
18. GWAS study on non-alcoholic fatty liver disease (NAFLD) in pediatric and adult population: comparison of size effect between adult and children using participants of the eMERGE Network. Lead Author: Bahram Namjou (CCHMC) *
19. The identification and reporting of actionable incidental genetic variants from large scale clinical sequencing of drug response genes. Lead Investigator Quinn Wells (VU).*
20. Feasibility of using geocoded US Census/ACS variables as proxy for socioeconomic status in genotype-phenotype interaction studies of T2DM and obesity. Lead Investigator: Kathryn Jackson (NU)*
21. Association of rare and common variants with insulin resistance: results from the eMERGE Network. Lead Investigator Daniel Kim (GHC/UW)
22. Investigation of CETP SNPs with LDL-C, BMI and risk of T2D. Lead Investigator: Brendan Keating (CHOP)
23. Quantitative and discreet trait analysis across eMERGE-I phenotypes. Lead Investigator: Joseph Glessner (CHOP)
24. Multi-site IRB review – experience of the eMERGE Network. Lead Investigator: Jen McCormick (Mayo)
25. Clinical Decision Support for Pharmacogenomics – Results from the eMERGE Network. Lead Investigator: Tim Herr
26. A phenome-wide association study to discover pleiotropic effects of PCSK9. Lead Investigator: Maya Saforova (Mayo)
27. An investigation of somatic mutations in PGx-eMERGE dataset. Lead Investigator Kenneth Kaufman (CCHMC)
28. The eMERGE Network: The practice of patient education in the return of genomic medicine results. Lead Investigator: Cassandra Perry (CCHMC/BCH)
29. Agreement between research-grade sequencing and CLIA validation genotyping in eMERGE-PGx. Lead Investigator: Laura R Torvik (NU)
30. The identification of adverse events in the eMERGE PGx cohort using the electronic health record, and assessing association with genetic variation in the 84 pharmacogenes. Lead Investigator: David Crosslin (GHC/UW)
31. An investigation into the genetics of Intractable Epilepsy in the pediatric population. Lead Investigator: Berta Castillo (CHOP)
32. Patients' views on consent and data sharing in biobank research: A large multisite experimental survey in the US. Lead Investigator: Saskia Sanderson (MSSM)
33. A targeted sampling scheme utilizing both EHR and census information. Lead Investigator: Nate Mercaldo (VU)

34. Developing a national survey on consent across a national network of genomic medicine sites. Lead Investigator: Ingrid Holm (BCH) & Maureen Smith (NU)
35. A Review of U.S. Individuals' Perspectives on Governance and Consent in Biobanking. Lead Investigator: Nanibaa' Garrison
36. Exploring the genetic architecture of Age-Related Macular Degeneration (AMD) in the eMERGE network. Lead Investigator: Molly Hall (Marshfield/PSU)
37. Genome-wide Association Study of Gastroesophageal Reflux Disease (GERD) in Adult and Pediatric Populations. Lead Investigator: Patrick Sleiman (CHOP)
38. Genome-wide Association Study of Atopic Dermatitis in Adult and Pediatric Populations. Lead Investigator: Berta Almoguera (CHOP)
39. Genome-wide Association Study of Asthma in Adult and Pediatric Populations. Lead Investigator: Berta Almoguera (CHOP)
40. Genome-wide Association Study of Attention Deficit Hyperactivity Disorder (ADHD). Lead Investigator: John Connolly (CHOP)
41. Investigation of PCSK9 SNPs with LDL-C, BMI and risk of T2D. Lead Investigator: Brendan Keating (CHOP)
42. Phenotype transportability across Electronic Health Records. Lead Investigator: Joshua Denny (VU)
43. The COGENT consortium meta-analysis of blood pressure African ancestry cohorts. Lead Investigator: Dinga Velez Edwards (VU)
44. Using PheWAS to assess disease comorbidity and potential pleiotropy of genetic risk scores for rheumatoid arthritis. Lead Investigator: Robert Carroll (VU)
45. Association of rare and common variants in LDLR, HMGCR, NAT2, ABCA1, and APOA1 with plasma lipid levels: results from 9000 participants of the eMERGE Network. Lead Investigators: Daniel Kim (Michigan), Erin Austin (Mayo)
46. PGRNseq and GWAS predictors of Methylphenidate (MPH) response. Lead Investigator: Tanya Froelich (CCHMC)
47. Discovery, Replication and Clinical Associations of Pathway-Based Trans-eQTL. Lead Investigator: Laura Wiley (VU)
48. Multiscale Analysis Of Influenza Host-Pathogen Interactions: Fluomics. Lead Investigator: Ellie Sang Sukerman (NU)
49. Rare RYR1, CACNA1S variant annotation, exposure history, observed phenotypes in cases and controls. Lead Investigator: Senthilkumar Sadhasivam (CCHMC)
50. Characterizing the individual and shared genetic components of pheWAS phenotypes. Lead Investigator: Jonathan Mosley (VU)
51. Examining gene variants in eMERGE samples for association with uterine fibroids. Lead Investigator: Todd Edwards (VU)
52. Association of APOL1 G1/G2 risk alleles with metabolic and cardiovascular traits. Lead Investigator: Girish Nadkarni (MSSM) & Miriam Udler (CCHMC/BCH)
53. Burden of structural variation and PheWAS. Lead Investigator: Adam Gordon(GHC/UW)
54. Genetic Risk Scores for Complex Diseases in the eMERGE Network: Characterization and Predicative Abilities in Clinical Settings. Lead Investigator: Logan Dumitrescu (VU)
55. Chromosomal Anomalies that Affect Levels of White Blood Count (WBC) and its Differential. Lead Investigator: David Crosslin (GHC/UW)
56. Extracting the Quality of Prostate Cancer Care from Electronic Healthcare Records. Lead Investigator: Tina Hernandez Boussard (External)
57. Portable Applications for Implementing Multi-Site Clinical NLP Algorithms. Lead Investigator: David Carrell (GHC/UW)
58. Effective Use of Electronic Health Records to Identify Venous Thromboembolism: Results from the eMERGE Network. Lead Investigator: Jyoti Pathak (Mayo)
59. Genetic Risk Factors for Development of Diverticulitis. Lead Investigator: Abel Kho (NU)
60. Colon Polyps. Lead Investigator: Abel Kho (NU)
61. Genetic Variants Associated with Response to Heart Failure Treatment: The Electronic Medical Records and Genomics (eMERGE) Network. Lead Investigator: Sue Bielinski (Mayo)
62. Genome-wide study of resistant hypertension using existing genomic data and electronic medical records. Lead Investigator: Logan Dumitrescu (VU).
63. Genetic variation that predicts susceptibility to Clostridium difficile. Lead Investigator: David Crosslin (GHC/UW)
64. Genome-Wide Association of Risk of Heart Failure: The Electronic Medical Records and Genomics (eMERGE) Network. Lead Investigator: Sue Bielinski (Mayo)
65. Genome-wide association study of extreme obesity defined by electronic medical record phenotyping. Lead Investigator: Glenn Gerhard (Geisinger)
66. Locus-specific PheWas study: An investigation of size effect of IRF5 and STAT4 in various autoimmune diseases and other phenotypes in all available samples in eMERGE. Lead Investigator: Bahram Namjou (CCHMC)
67. Quantitative and discreet trait analysis across eMERGE-II phenotypes. Lead Investigator: Joseph Glessner (Harvard)
68. A Highly Accurate Electronic Algorithm for the Classification of Asthma Severity in Children. Lead Investigator: Eric Hysinger (CHOP)
69. Cognitive Interviews associated with developing a national survey on consent across a national network of genomic medicine sites. Lead Investigator: Melanie Myers (CCHMC)
70. The eMERGE Network: Healthcare provider education to support genomic medicine in practice. Lead Investigator: Carolyn Vitek (Mayo)

Site-Specific Manuscripts

CHOP

1. Kim DS, Kim JH, Burt AA, Crosslin DR, Burnham N, et al. *Burden of potentially pathologic copy number variants is higher in children with isolated congenital heart disease and significantly impairs covariate-adjusted transplant-free survival.* J Thorac Cardiovasc Surg. 2016 Apr;151(4):1147–1151.e4. PMCID: PMC4801686
2. Desai A, Connolly JJ, March M, Hou C, Chiavacci Ret al. *Systematic data-querying of large pediatric biorepository identifies novel Ehlers-Danlos Syndrome variant*. BMC Musculoskelet Disord. 2016 Feb 16;17(1):80. PMID: 26879370 PMCID: 4754938
3. Finkel TH, Li J, Wei Z, Wang W, Zhang H, et al. *Variants in CXCR4 associate with juvenile idiopathic arthritis susceptibility.* BMC Med Genet. 2016 Mar 22;17:24. PMID: 27005825 PMCID: 4804485

Geisinger

1. Carey DJ, Fetterolf SN, Davis FD, Faucett WA, Kirchner HL et al. *The Geisinger MyCode community health initiative: an electronic health record-linked biobank for precision medicine research*. Genet Med. 2016 Feb 11. PMID: 26866580

Group Health/UW

1. Korngiebel DM, McMullen CK, Amendola LM, Berg JS, Davis JV, et al. *Generating a taxonomy for genetic conditions relevant to reproductive planning*. Am J Med Genet A. 2016 Mar;170(3):565-73. PMID: 26889673

Marshfield/Essentia/PSU

1. Li R, Dudek SM, Kim D, Hall MA, Bradford Y, et al. *Identification of genetic interaction networks via an evolutionary algorithm evolved Bayesian network*. BioData Min. 2016;9:18. PMCID: PMC4862166
2. Okula A, Wallace J, Peissig P, McCarty C, Brilliant M, et al. *Knowledge driven binning and PHEWAS analysis in Marshfield Personalized Medicine Research Project Using Biobin*. Pac Symp Biocomput. 2016;21:249-60. PMID: 26776191 PMCID: 4824557
3. Kim D, Lucas A, Glessner J, Verma SS, Bradford Y, et al. *Biofilter as a functional annotation pipeline for common and rare copy number burden*. Pac Symp Biocomput. 2016;21:357-68. PMID: 26776200 PMCID: 4722964

Mayo

1. Halverson CM, Clift KE, McCormick JB. *Was it worth it? Patients' perspectives on the perceived value of genomic-based individualized medicine.* J Community Genet. 2016 Feb 9. PMID: 26860291 PMCID 4796045

Vanderbilt

1. Schildcrout JS, Shi Y, Danciu I, Bowton E, Field JR, et al *A prognostic model based on readily available clinical data enriched a pre-emptive pharmacogenetic testing program.* J Clin Epidemiol. 2016 Apr;72:107–115. PMCID: PMC4779720

OVERVIEW *of e*MERGE TOOLS

Main resources and tools produced and supported:

- CDS KB: A repository of clinical decision support knowledge designed to support clinical processes from diagnosis and investigation through treatment and long-term care.
- eleMAP: A tool that allows researchers to harmonize their local phenotype data dictionaries to existing metadata and terminology standards such as the caDSR (Cancer Data Standards Registry and Repository), NCIT (NCI Thesaurus) and SNOMED-CT (Systematized Nomenclature of Medicine-Clinical Terms).
- eMERGE Infobutton Project: A template containing important content topics for genomic medicine.
- eMERGE Record Counter: A web-based research tool that provides exploratory data figures for research planning purposes and feasibility assessment.
- Model consent language: A publication representing the compiled work of eMERGE I investigators and consultants on consent language for the collection and storage of human biospecimens and data for future research, particularly those collections that have an electronic medical records component.
- MyResults.org: A website to educate providers and patients about genetic test results and the related medications.
- PheKB: An online collaborative environment for building and validating electronic algorithms to identify characteristics of patients within health data.
- PheWAS Catalog: A catalog that contains the PheWAS results for 3,144 single-nucleotide polymorphisms (SNPs) present in the NHGRI GWAS Catalog.
- SPHINX: A data exploring tool for genetics related drug response hypothesis generation.
- Synthesis-View, PheWAS-View and Phenogram: Visualization tools for genome & phenome-wide data.

*e*MERGE PHASE III PHENOTYPES

SITE LEAD	PHENOTYPE	INTENDED COHORT
CCHMC	Pediatric Pain Perception, Pain Sensitivity	*e3* Sequencing[1]
	Pediatric Migraine	*e3* Sequencing
	Primary Pulmonary Hypertension	*e3* Sequencing
	Hypermobility, EDS	*e3* Sequencing
	Pain Management, Opioid Dependence, Neonatal Abstinence	*e3* Sequencing
CHOP	Epilepsy Antiepileptic drug Response	*e3* Sequencing/GWAS[2]
	Intellectual Disability	*e3* Sequencing/GWAS
	Obesity	*e3* Sequencing/GWAS
Columbia	Autoimmunity	*e3* Sequencing/GWAS
	Breast Cancer	*e3* Sequencing/GWAS
	Liver Disease/Cirrhosis	*e3* Sequencing/GWAS
	Heart Failure / Cardiomyopathy	*e3* Sequencing/GWAS
	Stroke / Cerebrovascular Disease	*e3* Sequencing/GWAS
	Chronic Kidney Disease	*e3* Sequencing/GWAS
Geisinger	Pediatric Familial Hypercholesterolemia	*e3* Sequencing
	Ornithine Transcarbamylase (OTC) Deficiency-non-classic Presentation	*e3* Sequencing
	Tuberous Sclerosis Complex	*e3* Sequencing
GHC/UW	Polyps / Familial Colorectal Cancer	*e3* Sequencing
	Endometrial and Ovarian Cancer	*e3* Sequencing
	Sexual Dysfunction	*e3* Sequencing
	Depression	*e3* Sequencing
Harvard	Bipolar Disorder	*e3* Sequencing/GWAS
	Coronary Artery Disease	*e3* Sequencing/GWAS
	Schizophrenia	*e3* Sequencing/GWAS
	Rheumatoid Arthritis	*e3* Sequencing/GWAS
	Asthma	*e3* Sequencing/GWAS
	Hyperlipidemia	*e3* Sequencing/GWAS
Mayo	Adult Familial Hypercholesterolemia	GWAS/PGRN-Seq (PGx)[3]
	Contrast Nephropathy	GWAS/PGRN-Seq (PGx)
	Heparin-induced Thrombocytopenia	GWAS/PGRN-Seq (PGx)
	Response to Heart Failure Medication	GWAS/PGRN-Seq (PGx)
	Metformin Response	GWAS/PGRN-Seq (PGx)
NU	Chronic Rhinosinusitis	*e3* Sequencing/GWAS
	Valvular Disease	*e3* Sequencing/PGRN-Seq/GWAS
	Atopic Dermatitis	*e3* Sequencing/GWAS
	Adult Headaches, Migraine	Sequencing/PGRN-Seq/GWAS
VU	Pneumonia	GWAS
	Arrhythmias, (Atrial Fibrillation, QT Prolongation, Conduction System Disease, Brugada Syndrome)	*e3* Sequencing
	Hereditary Amyloidosis	*e3* Sequencing
	Cancer Susceptibility (plus Cancer PheWAS)	*e3* Sequencing
	Hearing Loss	GWAS
	Dry Eye	GWAS

[1]**e3 sequencing**: Sequencing data for ~25,000 participants across all nine Phase III Sites; [2]**GWAS**: Imputed GWAS data for ~80,000 participants across all 11 Phase I-III sites. [3]**PGRN-Seq (PGx) Cohort**: PGRNSeq sequences data for 9,015 participants across all nine Phase II sites.

eMERGE Network: ESP Conference Call Minutes

Thursday, February 25, 2016 | 3:30pm EST (2:30pm CST; 12:30pm PST)

Attendees:

BCH: Ingrid Holm; CCHMC: John Harley; CHOP: John Connolly, Hakon Hakonarson; Columbia: Alexander Fedotov, Ali Gharavi, George Hripcsak, Chunhua Weng; Geisinger: Marc Williams; GHC/UW: David Crosslin, Aaron Scrol; Harvard: Elizabeth Karlson, Shawn Murphy, Scott Weiss; Marshfield: Murray Brilliant; Mayo: Iftikhar Kullo, Stephen Thibodeau; Mt. Sinai: Northwestern: Rex Chisholm; Vanderbilt: Dan Roden; Baylor: Richard Gibbs; Partners/Broad: Birgit Funke, Stacey Gabriel, Niall Lennon, Heidi Rehm, NHGRI: Rongling Li, Teri Manolio, Ken Wiley, Jyoti Gupta, Kira Wong; CC: Paul Harris, Melissa Basford, Kayla Howell, Brianne Brucker Derveloy; ESP: Howard McLeod (Moffit Cancer Center), Eta Berner (University of Alabama – Birmingham), Kimberly Doheny (Johns Hopkins University), Gerardo Heiss (University of North Carolina), Stan Huff (InterMountain Healthcare), Lisa Parker (University of Pittsburgh), Vandana Shashi (Duke University)

Welcome, Opening Remarks, General Updates – *Rongling Li & Howard McLeod*

- Rongling welcomed the ESP members, and highlighted addition of two new eMERGE III ESP members: Kim Doheny (Johns Hopkins University) and Vandana Shashi (Duke University).
- Howard thanked the coordinating center for creation of the ESP conference documents, and expressed his excitement at the kickoff eMERGE III.

Network Introduction – *Rex Chisholm*

- In Phase III, the Network incorporated sequencing and a focus on return of results to participants.
  - The Network will have ~87,000 GWAS samples available to conduct common variant analysis.
  - The Network developed an eMERGE III sequencing panel, consisting of 109 genes (ACMG56 and 53 submitted by sites) and ~1500 SNPs.
  - This phase of eMERGE will focus on "detailed, deep sequencing" in order to discover novel, rare and new variants, as well as further develop the knowledgebase on previously identified variants.
- In Phase III, The Network has implemented the following organizational changes:
  - Sites: Columbia and Harvard were added. Mt Sinai and Marshfield are Network partners emeritus.
  - Sequencing Centers: Sequencing this phase will be performed by Partners/Broad and Baylor.
  - Workgroups: Clinical Annotation and Outcomes workgroups have been assembled. CERC and PGx workgroups have been absorbed into relevant Phase III workgroups. The Pediatrics workgroup has been discontinued, with the understanding that the work is represented in the newly formed workgroups and if pediatric specific opportunities should arise, they could be addressed based on the project. The CERC Survey workgroup continues and is wrapping up their Phase II work.
- The Network clarified that the 25,000 subjects selected for sequencing on the panel may be new subjects or part of the original cohort. The decision to enrich with Phase III phenotypes of interest is site-dependent, with mixed recruitment strategies.

DNA Sequence & Analysis Pipeline – *Richard Gibbs & Niall Lennon*

- Currently, the CSGs are validating sample preparation and sequencing processes, and expect to begin sequencing by May 2016. Technical conclusions indicate that both reagents are very high quality, and minor pipeline differences are being scrutinized.

- Differences in the reagents used by the CSGs were discussed. Partners/Broad is using the Illumina product, and Baylor is using the NimbleGen product. Both have similar performance in tests to date.
- ESP questioned if controls were being shared or sourced. Controls are indeed either shared or sourced, based upon practical issues such as the amount of DNA available and where it comes from.

Clinical Annotation Workgroup – *Birgit Funke & Heidi Rehm*

- The Clinical Annotation Workgroup was formed in response to the Network's need to build a solid foundation for consistent interpretation of results from the eMERGE-seq panel across sites. The workgroup has developed a pipeline to assess the clinical validity of gene/disease associations and assign draft ClinGen classifications to the eMERGE-seq panel of 109 genes, ~900 site proposed SNPs and pharmacogenomic loci. Further work will include finalizing ClinGen classifications and coming to a consensus on which variants are clinically actionable and therefore, returnable (site-specific policies and procedures will impact returnability as well).
- Participant consent was discussed, with particular interest in those participants consented before the creation of the sequencing panel. A broad range of return of results options has been discussed with participants. Options presented are site dependent and cohort-specific, ranging from broad consent with the full expectation to return results if care is affected (Geisinger) to a tier system designed to engage patients and providers in the return of results process (VU).
- The Network's plans for providing candidate gene information to patients and family members was discussed. Variant of unknown significance (VUS) results will not generally be returned to healthy patients, but GeneInsight does have a mechanism to update reports over the length of the study, so if a variant becomes implicated in disease it can be returned in the future. Further, raw data, including VUS data will be stored and available for analysis (ex: to study penetrance).

Discussion and Suggestions from ESP

- The ESP requested further information on the frequency of genetic testing data being returned to site EHRs, and if that was accomplished by a special pathway. The EHRI workgroup identified that a file-based transfer from GeneInsight is the optimal process for returning data to EHRs. The Network has identified local solutions, but not a generalizable solution to date.

Executive Session

- The ESP was pleased with the ESP packet and network presentations, and felt the investigators presented a good foundation for the future work that the Network will do. They felt that the presentations were aspirational due to the newness of eMERGE3, but the plans presented for future work were reasonable.
- The ESP expressed appreciation for how the eMERGE network had previously disseminated lessons learned from Ethical, Legal and Social Implications (ELSI) research to the scientific community and recommended that the network continue to study social and ethical issues, especially relating to the process of reporting clinically reportable variants (e.g. who is involved when decisions are made about returning results, etc). Therefore, other institutions and consortia can make use of the same practices and learn from any pitfalls. They expect to see publications on the process of gaining consistency in variant interpretation and clinical reportability.
- Regarding reporting of variants of all classes as described in the "ACMG Standards and Guidelines" to study sites (note: reporting variants of all classes to site investigators does not mean returning variants of all classes to patients' EMRs. Sites will decide, based on their study objectives and IRB approval, which variants they will return to patients), some ESP members recommended reporting of all classes of variants to the study sites because: 1) it would increase possibilities for scientific enquiry; and 2) it would also allow the network to compare the ethical, legal, and social implications across sites of having a clinical report including information that they do not plan to

report to patients. However, a few ESP members advocated the potential merits of permitting and studying different reporting practices, as well as the opportunity to study patient service seeking following reports and impact (including financial and service utilization) on institutions of different degrees of reporting.
NHGRI staff noted that reporting variants of all classes to sites likely has significant budgetary implications which the Institute will try to work through with the investigators.

- The ESP expressed some concerns about having one sequencing center appearing to be leading the sequencing efforts based on the presentations given. The NHGRI staff reassured them that the two sequencing centers are working together well and communicating to harmonize data flow. One center is leading on clinical reporting using GeneInsight, which was the focus of the presentation to the ESP, and the other center is leading on sequencing data generation and management using DNANexus.
- The ESP members discussed the potential value of reinstating the pediatric working group, and ultimately felt that it was best to continue without one given that it has not been judged to be productive in the past.

ESP Recommendations:

1. The ESP recommended that the network study the social and ethical issues involved in the process of making scientific decisions about variant annotation and reporting clinically reportable variants (e.g. who is involved when decisions are made about returning results, etc). This study could lead to a network publication.
2. The ESP recommended that both sequencing centers issue clinical reports for all variants to the sites, pending resolution of budgetary issues.

eMERGE Network

Summary of the eMERGE Steering Committee

May 5th & 6th, 2016 | Bethesda, MD

The third Phase III eMERGE Steering Committee Meeting was held on May 5-6th, 2016 in Bethesda, MD. In order to ensure that the Network continues on a productive note as we get further into our initial year, please find highlights from the Steering Committee Meeting below.

Presentation slides are available here (login required).

Day 1: Full-Day Session

Welcome, Opening Remarks, General Updates – *Rongling Li*

- President Obama increased the NIH's FY17 budget by $825M, +2.5% from FY16.
- Goals for the May 2016 eMERGE Steering Committee Meeting:
  - Genomic Sequencing – current status and the SC decisions. Specifically, timelines related to sample shipping to CSGS and data receiving from CSGs; and sequencing dataflow as it pertains to clinical reports (GeneInsight) and research data (DNAnexus)
  - eMERGE Data Management. Specifically, eMERGE 1-3 data, and eMERGE 3 sequencing data.
  - Scientific Projects.
  - Workgroup Activities.

Announcements, Opening Remarks – *Rex Chisholm*

ACTION ITEM: The Steering Committee is encouraged to consider how best to make use of data resources assembled over prior phases of eMERGE. These are valuable data resources, especially for common variant related studies. eMERGE Workgroup members are charged with proposing and accomplishing publications and deliverables using existing resources while eMERGE-Seq data are being generated.

Updates on the Imputed eMERGE 3 Data Set and new PGRNSeq Multisample Data Set – *David Crosslin, Amber Burt, Pari Devi, Adam Gordon & Gail Jarvik*

- eMERGE Pre-e3 array data and imputation:
  - As decided in the January Steering Committee Meeting, the CC is using the same pipeline used to impute phase 1 and 2 data to impute pre-3 data (Shape IT for pre-phasing, Impute 2 for the imputation, and 1000 Genomes (Phase I) for reference). Note: 5 sites provided imputable data (CCHMC, CHOP, Harvard, Mayo, VU). Others provided sequencing or exome chip data).
  - Final data set (to be available through Aspera and DNAnexus) includes:
    - IMPUTE2 probability and information files,
    - QC'd PLINK files,
    - PDF summary report of imputation and quality metrics, and
    - Complete (Phase 1,2,3) merged imputed dataset.
    - Prephasing output files (SHAPE-IT)
    - An Identity by Descent (IBD) matrix and principle component analysis (PCA) will be provided for each set.
  - Data should be available via the Aspera serve and on DNAnexus in the upcoming weeks and will eventually be put in dbGaP. There are roughly 80k participants in the study.

- The eMERGE RecordCounter now has a downloadable file of all eMERGE participants through phase II with their case or control status on each of the completed published eMERGE phenotypes. Site variability in cohorts and phenotype deployment makes this a sparse matrix.
- New data on pre-e3 subjects and re-running of prioritized phenotypes will be included in the eRC

ACTION ITEM: The CC will confirm new site data incorporation and data refresh schedule of the eRecordCounter.

- PGRNseq multisample calling and annotation:
  - The CC is currently ~75% finished calling the multisample VCF. The new files will be annotated using information from SeattleSeq, ClinVar, OMIM, HGMD, SnpEff, Variant Effect Predictor and prior clinical associations, and IBD and PCA will be provided. The dataset and will be available through Aspera and DNAnexus. SPHINX will also be updated. There are roughly 9100 participants in the study.

A CPIC-Translation Pipeline for Implementing Pharmacogenomics – *Marylyn Ritchie*

- Marylyn presented a pipeline (currently finalizing development) investigators can use to process research data to identify individuals who have a CPIC level A variant of interest. Data is passed through an algorithm that runs through all the possible allele combinations the individual could have, based on the coverage of the sequencing platform the data was produced by, and produces clinical recommendations. Next steps for the pipeline include finishing the website implementation and testing, disseminating the code, and developing the corresponding manuscript.

eMERGE PGx Updates – *Laura Rasmussen-Torvik*

- Status:
  - Enrollment complete: Approximately 9100 – will have final count with updated PGRNseq VCF being produced now.
  - The subgroup meets monthly to coordinate efforts with workgroups, assist with manuscripts in progress, and brainstorm next steps for studies
- On-going Projects:
  - CLIA validation agreement: This project is studying discordance in interpretation between original and sanger validated tests. Some sites couldn't place where the discordance happened, so a sub-set (those that could) may be used. The lead investigator is planning a letter to the editor.
  - Epilepsy: The study found NTRK2 is associated with drug resistant epilepsy in both single variant and gene based analyses. This Is a novel finding, as this gene has not been associated with epilepsy in the past, but is known to be involved in the CNS.
  - Provider education: This manuscript is in review. It concludes that there are some commonalities, but educational approaches varied significantly. Variations in approaches reflect different levels of institutional priority and culture, availability of resources, and the scope of the implementation of PGx in the organization. Education efforts are necessary, but not sufficient to ensure provider acceptance and adoption of pharmacogenomics.
  - CDS: An initial analysis on CDS design decisions and physician response has been performed and presented at AMIA as an abstract. A full manuscript is in development (lead investigator is following up with individual sites for clarification on data provided).
  - Somatic Mutations: Investigators have processed over 1700 samples and discovered 35 possible somatic mutations.
  - Lipids: Investigators are looking for genes associated with lipid lab variance. Additional new data are anticipated.
  - Adverse events: Investigators are looking for patterns in hematological counts. Additional labs and medications data are needed. The lead investigator will contact sites to obtain.

- Future directions:

- Outcomes outside of EHR: The group is moving forward with a homogenous attempt to capture what sites are already doing using surveys/interviews.
- Additional data capture from EHR: The PGx subgroup is working to identify other data and/or CDS information they can gather. They are also working on a schema to prioritize PGx related phenotypes with e3 phenotypes.

ACTION ITEM: Members interested in participating in the PGRNSeq multisample analysis project should contact Adam Gordon.

ACTION ITEM: Members interested in participating in the monthly PGx subgroup meeting should contact Brianne Brucker Derveloy and Laura Rasmussen Torvik

Pharmacogenetic Polymorphism as an Independent Risk Factor for Hospitalizations in Older Adults – *Joseph Finkelstein*

- Columbia investigators conducted a hypothesis-generating pilot study (nested case-control study design) in adults <70 years of age in order to determine if PGx polymorphisms are an independent risk factor for frequent hospitalizations in older adults with polypharmacy. Study results showed that participants with a PGx polymorphism indeed had higher instances of hospitalizations. In the study, hospitalization rates and costs were both lowered for patients with PGx polymorphisms whose medications were adjusted to account for their metabolism status.

DNA Sequencing Pipeline Update – *Richard Gibbs, Donna Muzny, Niall Lennon, Birgit Funke, Yaping Yang, Sandy Aronson & Heidi Rehm*

- eMERGE-Seq Design: 535KB (109 CDS genes, 1552 SNPs)
- Pipeline:
  - Baylor: The panel as designed has been developed and performance tested. Results: average coverage 347x with very high sensitivity, specificity, and PPV values. All gene variants (SNVs, Indels, CNVs) tested were ascertained. Baylor is currently winding up the CAP/CLIA validation process and is ready to go into production.
  - Partners/Broad: The panel as designed has been developed and performance tested. Results: Average coverage 256x, with very high sensitivity, specificity, and PPV values. Performance testing shows the assay missing bases in some indels and CNVs because of the small number of events assessed in the assay (did not test on full plate). Partners/Broad is working to increase mean coverage target to match Baylor. A reproducibility set is being run soon, and coverage analysis will be re-run after that. All Partners/Broad sites have been contacted and the 7 plates of UW samples are being stored at the Partners/Broad CSG awaiting final assay validation before sequencing begins. The second site will receive submission tubes and instructions shortly.
- Overlap Comparison: The group reviewed the list of exons and genes with low coverage for each CSG (this will change after Partners/Broad re-runs their coverage analysis). Sequencing centers expect only a small amount of missing data, and will ask site experts to comment on the clinical significance of the individual exons/genes that are not well covered. There are pathogenic variants in the non-covered and poorly-covered regions. Sequencing limitations will be listed in clinical reports. Options to address the coverage gaps issue revolve around assessing how far from the thresholds for acceptable coverage the regions are and potentially changing the thresholds if data would still be clinically reliable: 1) lower the accepted coverage threshold overall; 2) decrease the number of samples that required at the coverage threshold elected. To assess the impact of this, an assessment of actual coverage is needed rather than reporting regions not meeting a given threshold.
- Interpretation/Curation
  - Baylor's interpretation pipeline includes: bioinformatics prioritization, manual curation of new variants (which will be synchronized with Partners/Broad), primary and secondary review, Sanger sequencing confirmation if reporting (P, LP). Reports will include the following sections: demographics, the call/interpretation, a take home summary, a detailed information table, coverage data, methodology, and references.

- Partners/Broad's interpretation pipeline (after receiving the raw BAM data) includes: annotation/interpretation, filter non-reportable results, output, triage, approval, upload to GeneInsight/Reported (if P, LP- after Sanger confirmation). Reports will include structured data, interpretation analysis (what the call is) and if CNVs were analyzed, an interpretive summary (data behind why the call was made), coverage limitations, and detailed variant interpretations.

- Electronic Report Delivery and Knowledge Management Plans:
  - Plan: to deliver reports, maintain a common knowledgebase, synchronize curations, maintain a research repository, and provide an analysis environment.
  - Commonalities
    - Structured report and alert delivery: The eMERGE EHRI workgroup is defining the format for XML based file transfer (structured data) and PDF (clinical data) to be delivered to sites by sFTP.
    - Knowledge repository (GeneInsight): Partners/Broad and Baylor will both contribute to this evolving understanding of variant actionability. It will be available for investigators to search for information on variants. Data from other labs who use GeneInsight can possibly be made available through VariantWire, should the Network choose to submit an access application to VariantWire. It can also be used to search de-identified patient-level data from previously generated reports.
    - DNAnexus (AKA the Data Commons) is an analysis environment available to Network.
  - Differences: Note that both GeneInsight and DNAnexus are searchable, investigators can download search results to excel, and alerts will be generated as interpretive updates are made for as long as the systems are maintained. Data will also flow to ClinVar, so updates can also be found there.
    - Partners/Broad will deliver clinical reports through the GeneInsight Clinic application.
    - Baylor will deliver clinical reports through custom options on DNAnexus.

ACTION ITEM: CSGs will send draft clinical reports to CC for circulation.

ACTION ITEM: Partners/Broad will re-run their validation assay with a full sample plate and update the coverage information to Adam Gordon. Adam will reassess the coverage concordance again with this updated data from Partners / Broad.

ACTION ITEM: CSGs will consult with site experts to determine the clinical relevance of non-covered regions.

ACTION ITEM: Members with concerns regarding the data management of clinical report data should contact Rex Chisholm by 5/6/16.

Baylor/GeneInsight Harmonization – *Richard Gibbs & Sandy Aronson*

- Baylor will use GeneInsight's manual update and batch upload/download capabilities to harmonize their interpretations with Partners/Broad's with minimal time delay. This solution is already in place with good exchange between curators. Real time high tech coordination was evaluated and determined not to be valuable enough to justify its cost.

ACTION ITEM: Members who would like to be involved in the weekly Baylor/GeneInsight harmonization meetings should contact Ken Wiley.

Pathogenicity and actionability: Difficulty and opportunity – *Les Biesecker, Keynote Speaker*

- Pathogenicity: suggested approach to overcoming perceived difficulty. 1) Break down a highly dimensional problem into components, 2) address the uncertainties, 3) weight evidence objectively when presented with heterogeneous underlying data, 4) decouple implications from utility, and 5) preserve professional judgment where appropriate. The Sequence Variant Interpretation ClinGen Group has developed short (refine and clarify), medium (change) and long term (quantitative Bayesian framework) approaches to criteria.
- Actionability: ClinGen Actionability workgroup is working to 1) Standardize and unitize thinking of clinical utility of a variant, 2) Application of a semi-quantitative system to organize available knowledge to

enhance transparency and usefulness. This approach has avoided the nirvana fallacy. Actionability process follows a multistep approach: 1) Screening of variants, 2) Full evaluation including systematic identification of sources, determination of relevance, tiered ratings and data abstraction, 3) Rating and scoring of domains.

- Diagnosis: misconception that assigning pathogenicity is equivalent to making a diagnosis, and an issue with a desire to make clinical genomics "idiot proof." The speaker highlighted three separate functions to alleviate these issues: 1) the clinical lab should determine what is known or knowable about the variant, and the clinician should 2) use the variant to make a diagnosis (or not), 3) then use the diagnosis to change management. This could result in regulatory implications.
    - Bayesian Quantitative Genomics Approach which assigns variants a prior probability of pathogenicity, then modifies this prior based upon a piece of evidence. This approach offers many benefits and relatively few downsides that can be alleviated through education and additional data.
    - Statistical Decision Theory: assists in selection of errors that you would like to make, and moves away from unipolar approach of minimizing risks and harms of genomics by balancing them.
- Suggested reading: The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy by Sharon McGrayne.

The Monarch Initiative: Phenotype Ontologies for Data Integration and Discovery – *Melissa Haendel, External Speaker*

- Deep phenotyping within and across species can aid diagnosis, discovery, and translational matchmaking. Using the disease-phenotype data we already have, we can assess the quality of the phenotyping to improve diagnostic power. Text mining clinical notes for HPO terms can provide sufficient phenotyping for rare disease exome/genome analysis. An exchange standard is needed to facilitate distributed phenotype data sharing for clinics, labs, patients, and journals.
- The Monarch Initiative matches against known disease and models, while the Patient Archive matches against patients (100k genomes, Japan and Australia).

ADHD – from Phase 2 Algorithm to Clinical Trials – *Hakon Hakonarson, John Connolly & Berta Almoguera*

- The investigators provided a brief background of ADHD statistics, and then reviewed their analysis. An ADHD meta-analysis in European Americans identified a genome-wide significant relevant variant in Chromosome 11, as well as two other nominally significant variants. The genes identified are: *CNTN5* "Contactin 5," *LRRC58* "Leucine-rich repeat-containing protein 58," and *CLYBL* "Citrate lyase beta like." The ADHD GWAS in African Americans awaits replication.
- Discovery of Mutations (copy number variations/CNVs) in ADHD. A gene network analysis showed that genes interacting with the genes in the GRM family are enriched for CNVs in ~10% of the cases (610Q chip) ($P = 4.38 \times 10^{-10}$). eMERGE investigators have identified rare recurrent CNVs affecting glutamatergic neurotransmission genes that are highly overrepresented in multiple ADHD cohorts. Investigators discovered drug NFC-1 (fasoracetam) activates the mGluR pathway, is well tolerated, ameliorates cognitive impairment and hyperactivity in animal models and has structure-similar compounds with good safety protocols. This drug originally targeted vascular dementia, but through CHOP's 5-week, dose-escalation ADHD clinical trial, NFC-1 improved outcomes in over 80% of participants. Similar future trials are planned for autism and schizophrenia.

Day 2: Half-Day Session

Network Status Update – *Rex Chisholm*

- Publications Update: Emphasized to the Steering Committee members that outstanding publications should be addressed, and either moved along or withdrawn.

Whole Genome Sequencing Discussion – *Rex Chisholm*

- Thanks to excess capacity at the NHGRI-supported Washington University genome sequencing center, eMERGE members will have access to whole genome sequencing data for about 2,000 individuals from CHOP and Northwestern. In addition, whole exome sequencing data for 2700 individuals is available from Columbia. Some of Columbia's WES samples overlap with other eMERGE projects (e3 and PGx) and would therefore be orthogonal data for replication or validation.
- Potential projects for these data:
    - Add to PGx data for Marylyn's CPIC pharmacogenomics project.
    - Test the concordance of NU data (also processed on PGRNSeq platform) with PGx data.
    - Running eMERGE phenotype algorithms across WGS data.

ACTION ITEM: Members with WES/WGS data they would like to make available for Network analysis should contact David Crosslin.

ACTION ITEM: Members with ideas on projects utilizing WGS data should contact Rex Chisholm.

Clinical Annotation Workgroup Report-Out – *Heidi Rehm & Gail Jarvik*

- Final results of gene-disease association curation: 35 genes (out of the 53 genes submitted by sites) were found to have at least one definitive gene-disease association based on ClinGen curation with site review, and will be returned if the association is determined to be actionable. Results from the rest of the genes won't be included in clinical reports.
- Actionability assessment: The group decided that the 56 ACMG genes will be considered actionable with some caveats. The group generated a draft classification list of site-proposed actionable genes with disease association, and it will be circulated to the sites for review and consensus building.
- SNP validity assessment: Of SNPs in the final design, there are 91 that are reportable. 36 have already been classified by one or both of the CSGs. In parallel, all the SNPs were electronically triaged. Next steps: CSGs will reassess non-unique variants and resolve any discrepancies.
- Baylor/LMM Variant Interpretation Harmonization: CSGs exchanged all previously reported variants and are 90% concordant for Pathogenic, Likely Pathogenic, and Unknown Significance (VUS) categories. CSGs do not have same classification schemes, so those categories were the only ones adjudicated. The teams are now focused on resolving high-impact discrepancies.

ACTION ITEM: The Clinical Annotation workgroup will continue to harmonize variant classifications between CSGs and determine consensus variant actionability for clinical reporting.

ACTION ITEM: The Clinical Annotation workgroup will complete a VariantWire Application for expanded variant data access for the Network.

Autoimmunity PheWAS – *John Harley*

- The group reviewed the results of a PheWAS investigating autoimmune disease association in IRF5 & STAT4 genes. Results include:
    - Autoimmune Hemolytic Anemia and Pernicious Anemia results may be real and associated with IRF5. IRF5 and STAT4 haven't been studied as candidates yet (no GWAS done).
    - Type 2 Diabetes association is probably not real.
    - Autoimmune algorithms have not been developed in general
    - Sample size in this study is too small
    - New imputation data should confirm results and provide genotyping in the IRF5 promoter

EHR Integration Workgroup Report-Out – *Sandy Aronson & Casey Overby*

- The workgroup is currently focused on the engineering aspect of their charter: establish, document, and seek to continuously improve process flows for delivery of eMERGE reports and data. Members have nearly completed the common structured data exchange format, intend to submit a concept sheet regarding "Network Infrastructure for Lab Reporting and Knowledge Management," and are in the process of forming an Infobutton project subgroup.
- As a follow-up to the earlier Baylor/GeneInsight Harmonization discussion, the workgroup will work through the details as a common group and not as two lab-specific subgroups. EHRI members are invited to attend the Baylor/GeneInsight weekly meetings which are hosted on Fridays at 3:00pm EST (2:00pm CST; 12:00pm PST).
- The workgroup proposed creating a subgroup to define the regulatory and legal framework necessary to comply with HIPAA. Steering Committee members discussed various aspects of the regulatory environment. All members noted that their samples were collected under a research protocol, although HIPAA requirements remain site specific. It was unclear, however, what identified information would ever be shared among the Network.

ACTION ITEM: The EHRI workgroup will complete the common structured data exchange format.

ACTION ITEM: The EHRI workgroup will submit a concept sheet regarding "Network infrastructure for lab reporting and knowledge management."

ACTION ITEM: The EHRI workgroup will proceed with forming an Infobutton project subgroup.

Phenotyping Workgroup Report-Out – *Josh Denny & George Hripcsak*

- Progress:
    - The first 4 Phase I and II phenotypes are essentially complete, and the second 4 are in progress.
    - The group added Type 2 diabetes to those being run / re-run in e3.
    - Clarification: Secondary validation means implement, execute, and report PPV.
    - The group decided to investigate creating a cardiovascular core to process ancillary CV reports
- The group discussed how the eMERGE RecordCounter (eRC) and PheKB tools interact with one another and how to further integrate them (such as linking phenotype metadata in PheKB to eRC).
- The Common Data Model subgroup decided to develop a common analytical model for eMERGE phenotyping, and select (rather than generate) an information model. A supplement has been submitted to help sites convert to one of the models, allow investigators to perform a root cause evaluation of challenges, and work to increase NLP efficiency and consistency.
    - Schema: OHDSI/OMOP and i2b2
    - Terminologies: OHDSI/OMOP and maybe augment with others

ACTION ITEM: Phenotyping Workgroup will continue to implement the second group of four Phase I and II Phenotypes as prioritized by the group.

ACTION ITEM: Phenotyping workgroup chairs will circulate a survey exploring next steps of creating a Cardiovascular Core.

ACTION ITEM: Phenotyping workgroup (and Genomics workgroup) will develop proposal for a coordinated dataset for phenotypes rather than relying on PI to PI interactions to generate phenotyping datasets for each paper.

Discovery, Replication, and Clinical Associations for Pathway-Based Trans-eQTL – *Laura Wiley*

- Summary:
  - Investigators identified 15 replicating SNP-pathway associations and tested 9 for phenotypic associations across the eMERGE Network.
  - 1 SNP had significant phenotypic associations: rs425437, cancer of the lip.
  - Identification of biological hypothesis: SNP → Transcription Factor Binding → Increased Gene Expression → Altered Pathways → Phenotypic Association
- Investigators are currently planning molecular validation to demonstrate SNP does introduce HIF-1 binding and to demonstrate HIF-1 binding increases MARC2 expression. Investigators are also considering additional phenotyping validation at eMERGE III sites and confirming in tumor registry. If any eMERGE sites are interested in participating, please contact Laura Wiley.

CERC Survey Update – *Maureen Smith & Ingrid Holm*

- CERC Survey data collection was completed 9/15. Currently several publications are in progress/submitted/ accepted.
- The CERC Survey data (full survey, pilot survey, data dictionaries) will be open to the public through the eMERGE website after Network investigators have published their manuscripts. Users will be asked to complete a short REDCap survey before access. This will allow the group to capture data on how many and the types of investigators using the data. Grant acknowledgement/ public access compliance for external investigators using eMERGE data was also discussed.

Outcomes Workgroup Report-Out – *Hakon Hakonarson, Josh Peterson & Marc Williams*

- The workgroup has completed the outcomes map and is well into the process of prioritizing phenotype:gene pairs to measure cross-site outcomes. Future directions include cohort profiling, determining stratifications and estimating the number of patients within each stratum based upon expected variant rate. Approaches to outcome assessment include both cohort design and pragmatic RCT.
- Robert Green's LDLR study: 5 sites indicated interest in a multi-site RCT that examines change in both statin usage and LDLR as primary outcomes. This study would require external and/or private funding in order to sequence the LDLR in 100,000 people to get a sample size of 190 people to eventually randomize. This study could be transformed into an R01.
- Family outcomes: Janet Williams (Geisinger) will lead a joint effort on the topic of family outcomes with the ROR/ELSI workgroup that will focus on cascading effects of variant return. This project will measure additional health services received or otherwise ordered for a family member the can be attributed to the disclosure.

- Outcomes workgroup manuscript: disclose protocols and assessment algorithms prior to actual measurement of outcomes.

ACTION ITEM: The Outcomes workgroup will complete prioritization of phenotype:gene pairs with insight from the Phenotyping workgroup.

ACTION ITEM: The Outcomes workgroup will explore family outcomes with the ROR/ELSI workgroup, focusing on cascading effects of variant return.

ACTION ITEM: The Outcomes workgroup will develop a workgroup manuscript on the topic of disclosing protocols and assessment algorithms prior to actual measurement of outcomes.

A Phenome-Wide Association Study to Discover Pleiotropic Effects of PCSK9 – *Maya Safarova & Iftikhar Kullo*

- This project studied the pleiotropic effects of genes affecting LDL-C metabolism. A PheWAS analysis did not reveal evidence of pleiotropy for PCSK9 on a variant or gene level. The study revealed a novel paradigm of rapid ascertainment of pleiotropic effects of genes that are drug targets, with implications for identifying additional potential clinical applications of such drugs as well as off target effects.

Genomics Workgroup Report-Out – *Megan Roy-Puckelwartz & David Crosslin*

- The Genomics WG is analyzing how research is done in the eMERGE Network to inform the development of the permissions structure on DNAnexus. The group envisions 4 permission levels (network, site, collaborative group, and individual).
- On-site and webinar training for DNAnexus is available.

ACTION ITEM: Genomics workgroup (and Phenotyping workgroup) will develop proposal for a coordinated dataset for phenotypes rather than relying on PI to PI interactions to generate phenotyping datasets for each paper.

ACTION ITEM: Members should log on to DNAnexus to set up user accounts and explore the analysis tools.

ACTION ITEM: Genomics Workgroup members will use site-submitted analysis steps and tools to develop a consensus list for implementing on DNAnexus.

ACTION ITEM: Genomics Workgroup members will gather opinions about and needs for DNAnexus in-person and / or web-based DNAnexus training and identify the group needing in-person training from each site.

ACTION ITEM: Columbia's 2700 exomes of data will be added to the eMERGE organization on DNAnexus as soon as possible, either via transfer from UW or directly from Columbia's DNAnexus organization.

RoR/ELSI Workgroup Report-Out – *Ingrid Holm & Iftikhar Kullo*

- The workgroup provided updates on in-development projects.
    - Impact of Return of Genomic Results on Health Care Providers: The goals of the pilot project are to develop and test a survey of HCP that can be implemented across the eMERGE Network with future funding.
    - IRB perspectives around informed consent and return of results across the eMERGE III Network: Starting with a data collection for sites to describe their IRB process and experience.

- Patient Motivations for Seeking Elective Genetic Counseling in the Context of Clinical Genomic Sequencing: This study seeks to examine what motivates individuals (via qualitative interviews) to initiate a genetic counseling session in the context of clinical genomic sequencing, what individuals feel they gained as a result of these counseling sessions, and the extent to which participants' views about genomic sequencing results are affected by whether they receive genetic counseling. The workgroup will discuss how they can transform Mayo's site project into a Network-wide project
- Participant Survey Study: The pre-survey subgroup will identify domains and survey items to include on all site's survey by June 1st.
- Family History Project: The workgroup is discussing how to proceed with identifying ELSI issues around sharing genomic data with family members.
- ROCKET workspace repository: The workgroup will use this workspace to host patient surveys, provider surveys, qualitative interview guides and focus group topics.

ACTION ITEM: The ROR/ELSI workgroup members will begin data collection for sites to describe their IRB process and experience.

ACTION ITEM: The ROR/ELSI workgroup will discuss how they can transform Mayo's site project "Patient Motivations for Seeking Elective Genetic Counseling in the Context of Clinical Genomic Sequencing" into a Network-wide project

ACTION ITEM: The ROR/ELSI pre-survey subgroup will identify domains and survey items to include on all site's survey by June 1st.

ACTION ITEM: The ROR/ELSI workgroup is discussing how to proceed with identifying ELSI issues around sharing genomic data with family members.

ACTION ITEM: The ROR/ELSI workgroup will use the ROCKET workspace to host patient surveys, provider surveys, qualitative interview guides and focus group topics.

Summary of Action Items

1. The Steering Committee is encouraged to consider the following: using data resources already compiled in prior phases that can be very useful, especially for common variant related studies. eMERGE Workgroup members are charged with proposing and accomplishing publications and deliverables that could come from use of existing resources prior to the return of the eMERGE-Seq sequence data.
2. The CC will confirm new site data incorporation and data refresh schedule of the eRecordCounter. (complete)
3. CSGs will continue to assess their comparative coverage on the eMERGE-Seq platform:
    a. Partners/Broad will re-run their validation assay with a full sample plate and update the coverage information to Adam Gordon. Adam will reassess the coverage concordance again with this updated data from Partners / Broad.
    b. CSGs will consult with site experts to determine the clinical relevance of non-covered regions.
4. Network interaction opportunities:
    a. Members interested in participating in the PGRNSeq multisample analysis project should contact Adam Gordon.
    b. Members interested in participating in the monthly PGx subgroup meeting should contact Brianne Brucker Develoy and Laura Rasmussen Torvik
    c. Members who would like to be involved in the weekly Baylor/GeneInsight harmonization meetings should contact Ken Wiley.
5. eMERGE-Seq Data Workflow:
    a. CSGs will send draft clinical reports to CC for circulation (complete)

   b. Members with concerns regarding the data management of clinical report data should contact Rex Chisholm by 5/6/16. (complete)
6. Other EMERGE Sequencing Data Resources:
   a. Members with WES/WGS data they would like to make available for Network analysis should contact David Crosslin.
   b. Members with ideas on projects utilizing WGS data should contact Rex Chisholm.
7. The Clinical Annotation workgroup will continue to harmonize variant classifications between CSGs and determine consensus variant actionability for clinical reporting.
8. The Clinical Annotation workgroup will complete a VariantWire Application for expanded variant data access for the Network.
9. The EHRI workgroup will complete the common structured data exchange format.
10. The EHRI workgroup will submit a concept sheet regarding "Network infrastructure for lab reporting and knowledge management."
11. The EHRI workgroup will proceed with forming an Infobutton project subgroup.
12. Genomics workgroup (and Phenotyping workgroup) will develop proposal for a coordinated dataset for phenotypes rather than relying on PI to PI interactions to generate phenotyping datasets for each paper.
13. Phenotyping Workgroup will continue to implement the second group of four Phase I and II Phenotypes as prioritized by the group.
14. Phenotyping workgroup chairs will circulate a survey exploring next steps of creating a Cardiovascular Core.
15. Phenotyping workgroup (and Genomics workgroup) will develop proposal for a coordinated dataset for phenotypes rather than relying on PI to PI interactions to generate phenotyping datasets for each paper.
16. Access plus Association Analysis Data and Tools on the DNAnexus Platform:
   a. Members should log on to DNAnexus to set up user accounts and explore the analysis tools.
   b. Genomics Workgroup members will use site-submitted analysis steps and tools to develop a consensus list for implementing on DNAnexus.
   c. Genomics Workgroup members will gather opinions about and needs for DNAnexus in-person and / or web-based DNAnexus training and identify the group needing in-person training from each site.
17. Columbia's 2700 exomes of data will be added to the eMERGE organization on DNAnexus as soon as possible, either via transfer from UW or directly from Columbia's DNAnexus organization.
18. The Outcomes workgroup will complete prioritization and selection of phenotype:gene pairs for Network wide Outcomes data collection with insight from the Phenotyping workgroup.
19. The Outcomes workgroup will explore family outcomes with the ROR/ELSI workgroup, focusing on cascading testing following variant return.
20. The Outcomes workgroup will develop a manuscript on disclosing protocols and assessment algorithms prior to actual measurement of outcomes.
21. The ROR/ELSI workgroup members will begin data collection in a manuscript proposal to describe the inter-site variation in IRB process and experience.
22. The ROR/ELSI workgroup will discuss how they can transform Mayo's site project "Patient Motivations for Seeking Elective Genetic Counseling in the Context of Clinical Genomic Sequencing" into a Network-wide project.
23. The ROR/ELSI pre-survey subgroup will identify domains and survey items to include on all sites' surveys by June 1.
24. The ROR/ELSI workgroup will identify and investigate ELSI issues around sharing genomic data with family members.
25. The ROR/ELSI workgroup will use the ROCKET workspace to host patient surveys, provider surveys, qualitative interview guides and focus group topics.

Next Meeting: October 6-7, 2016 | Bethesda, MD