emerge network
ELECTRONIC MEDICAL RECORDS & GENOMICS

External Scientific Panel

Background Material

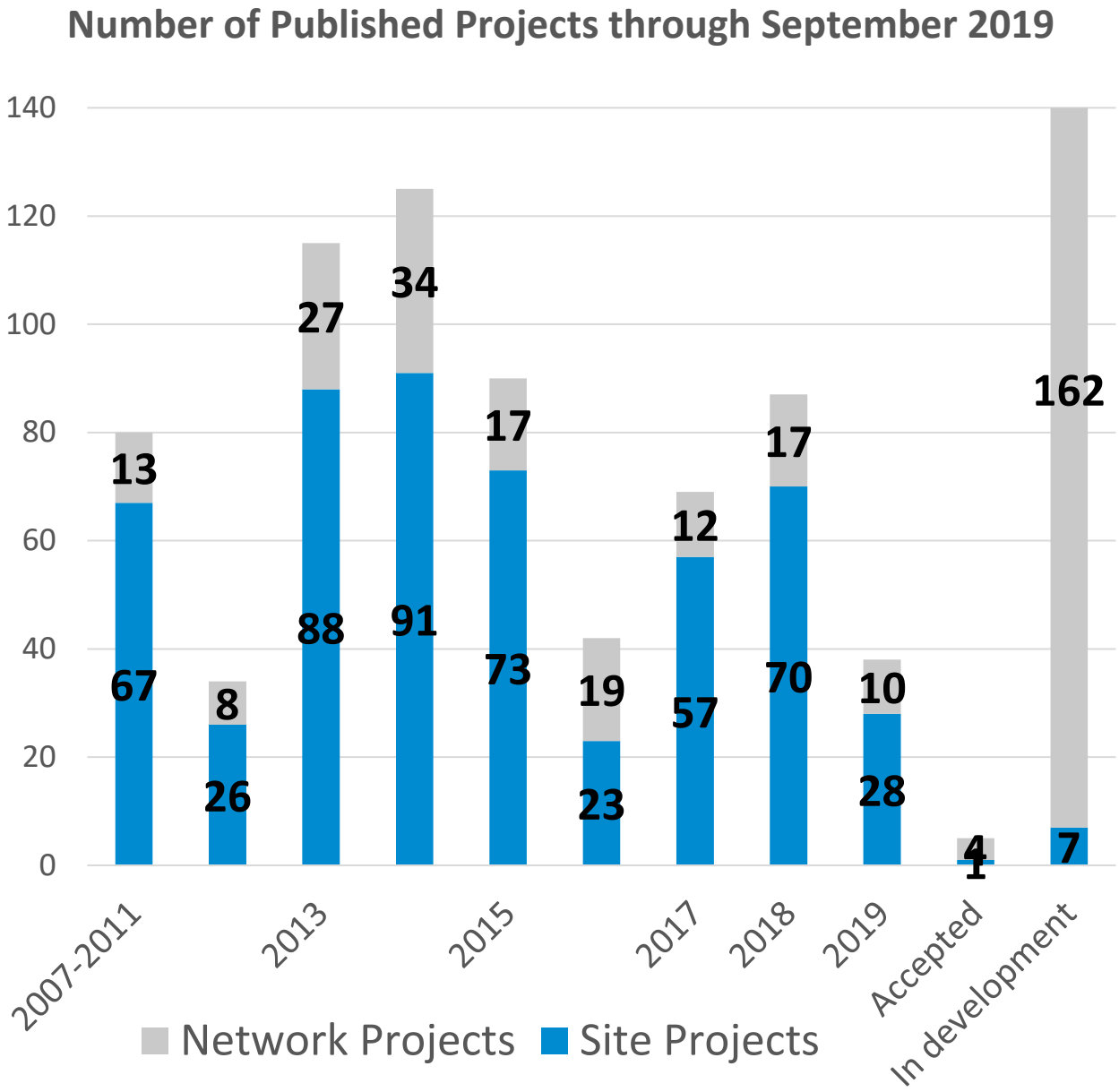
October 4th, 2019

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
HARVARD UNIVERSITY
KAISER PERMANENTE
MAYO CLINIC
MEHARRY MEDICAL COLLEGE
Northwestern Medicine
PARTNERS HEALTHCARE
VANDERBILT UNIVERSITY MEDICAL CENTER
UW Medicine UNIVERSITY OF WASHINGTON MEDICAL CENTER

National Human Genome Research Institute

TABLE *of* CONTENTS

*e*MERGE PUBLICATIONS



= 854 Total Projects

e MERGE PUBLICATIONS
from April 2019-September 2019

Submitted/Accepted and Published Network Manuscripts

1. Kho A, Gawron A, Thompson W, Rasmussen L, Pacheco J, Muthalagu A, Roberts A, Rasmussen-Torvik L, Smith M, Hayes G, Scheftner D, Armstrong L, Denny J, Larson E, Carrell D, Ralston J, Jarvik G, Li R, Ritchie M. Genetic risk factors for development of Diverticulitis/Diverticulosis. (Submitted)
2. Edwards T, Torstenson ES, Gilbertson J, Tsosie K, Giri A, Li C. MVtest: a method to flexibly model the genetic determinants of trait variability. (Submitted)
3. Edwards T, Tsosie K, Bray M, Hartmann K, Stewart E, Wise L, Jeff J, Kenny E, Loos R, Peissig P, McCarty C, Newton K, Reeds S, Scholes D, Kho A, Jones S, Torstenson E, Setiawan VW, Wellons M, Ruiz-Narvaez E, Edwards D, 23andMe analytic team. Examining gene variants in eMERGE samples for association with uterine fibroids. (Submitted)
4. Almoguera B, Vazquez L, Mentch F, Connolly J, Sleiman P, Almoguera B, Green C, Linneman J, Brilliant M, Borthwick KM, Sundaresan A, Williams M, Hakonarson H. Genome-wide Association Study of Atopic Dermatitis in Adult and Pediatric Populations. (Submitted)
5. Keating B, Ritchie M, Rasmussen L, Denny J, Roden D, Berg D, Brilliant M, Kullo I, de Andrade M, Namjou B, Cobb B, Harley, J, Crosslin D, Carrell D, Larson E, Jarvik G. Investigation of CETP SNPs with LDL-C, BMI and risk of T2D. (Submitted)
6. Hakonarson H, Xiao C, Ritchie MD, Connolly J, Crosslin D, Sleiman P, Almogeura B, Keating B. Quantitative and discreet trait analysis across eMERGE-I phenotypes. (Submitted)
7. Castillo B, Robbins E, Vazquez L, Abrams D, Mentch F, Connolly J, Sleiman P, Heysinger E, Gordon A, Manolio T, Manzi S, Wolf W, Holm I, Hakonarson H. An investigation into the genetics of Intractable Epilepsy in the pediatric population. (Submitted)
8. Smoller J, Cox N, Ledbetter D, Martin C, Ritchie M, Denny J, Namjou B, Lingren T, Chung W. PsycheMERGE. (Submitted)
9. Namjou B, Kottyan L, Rothenberg M, Namjou B, Lingren T, Marsolo K, Cobb B, Kaufman K, Harley J, Jarvik G, Crosslin D, Gordon A, Williams M, Kuhn B, Kirchner L, Borthwich K, Karlson E, Lee Y. NLP algorithm development and GWAS study for Eosinophilic Esophagitis (EoE) using eMERGE subjects. (Submitted)
10. Edwards T, Edwards T, Velez Edwards D, Denny J, Roden D, Hellwege J, Stallings S, Carroll R Jarvik G, Crosslin D, Gordon A, Devi P, Setia S, Ritchie M, Borthwick K, Shang N, Hripcsak G, Peissig P, Linneman J, Briliant M. A genetic association study of benign prostatic hyperplasia (BPH) in the eMERGE network. (Submitted)
11. Mosley J, Roden D, Wells Q, Denny J, Kullo I, Brittain E, Larson E, Murad A, Williams M, Chung W, Jarvik G, Borthwick K, Olson A, Rasmussen-Torvik L. Identifying clinical phenotypes associated with genetic predictors of cardiac structure. (Submitted)
12. Rehm H, Gibbs R, Smollrt J, Rasmussen L, Crosslin D, Smith M, Jarvik G, Almoguera B, Stanaway I, Gordon A, Sleiman P, Hakonarson H, Hayes G, Larson E, Leppig K, Harley J, Prows C, Namjou B, Hoell C, Kullo I, Peterson J, Person T, Fan X, Lindor L, Thibodeau S, Pendergrass S, Pulk R, Levy B, Overby C, Williams M, Murray J, Chung W, Aronson S, Weng C, Caraballo P, Freimuth B, Holm I. Harmonizing Clinical Sequencing And Interpretation For The eMERGE III Network. (Submitted)
13. Robinson J, Roden D, Bastarache L, Carroll R, Chen C, Jackson G, Mou T, Connolly J, Mentch F, Hayes G, Crane P, Hebbring S, Crosslin D, Gordon A, Rosenthal E, Stanaway I, Wei W, Petukhova L, Namjou B, Zhang G, Walton N, Jarvik G, Larson E, Weng C. Burden of Disease Associated with Extremes of Body Mass Index. (Submitted)
14. Venturi Y, Ritchie M, Denny J, Stanaway I, Larson E, Hebbring S, Smoller J, Jarvik G, Pendergrass S, Williams M, Hakonarson H, Rasmussen-Torvik L. Identifying pleiotropic effects in cardiovascular and neurological diseases using I data. (Submitted)
15. Knevel R, Raychaudhuri S, Karlson E, Slowikowski K, Cessie S, Terao C, Huizinga T, Lui J. Using genetic data in a Bayesian precision medicine framework to prioritize rheumatic disease diagnoses. (Submitted)
16. Venner E, Murugan M, Gibbs R. ARBOR: An Identity and Security Solution for Clinical Reporting. (Submitted)
17. Hripscak G, Shang N, Peissig P, Rasmussen L, Liu C, Benoit B, Carroll R, Carrell D, Denny J, Dikilitas O, Gainer V, Jose S, Joss M, Klann J, Mentch F, Marsolo K, Murphy S, Natarajan K, Pacheco J, Wei W, Weng C. OMOP Information Model for eMERGE Phenotyping. 2019;1-5. PMID: 31241152
18. Namjou B, Harley J, Lingren T, Marsolo K, Ritchie M, Verma S, Deshmukh D, Cobb B, Li R, Carey D, Still C, Mirshahi T, Wood C, Crosslin D, Carrell D, Jarvik G, Larson E, Gharavi A. 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. (Submitted)
19. Pucklewartz M, McNally E, Chisholm R, Pottinger T, Dellefave-Castillo L, Robinson A, Pesce L, Smith M, Pacheco J, Rasmussen-Torvik L. Whole genome sequencing analysis of 900 biobank participants to identify genotypes that predict cardiomyopathy. (Submitted)
20. Lingren T, Thaker V, Lingren T, Kennebeck S, Namjou B, Bickel J, Patibandla N, Savova G, Solti I, Holm IA, Harley J, Kohane IS, Crimmins

N. Pediatric Providers are Poor at Identifying Severe Obesity in Young Children at Two Tertiary Pediatric Medical Centers. (Accepted)
21. Velez Edwards D, Zhu X, Franceschini N, Denny JC, Edwards TL. The COGENT consortium meta-analysis of blood pressure African ancestry cohorts. (Accepted)
22. Mosley J, Denny J, Roden D, Bastarache L, Wells Q, Edwards T, McCarty C, Thompson W, Chute C, Jarvik G, Crosslin D, Larson E, Kullo I, Pacheco J, Peissig P, Brilliant M, Linneman J, Namjou B, Prows C, Ritchie M, Borthwick K, Verma SS. Gordon A, Palmer M, Williams M. Linking biomarkers to clinical phenotypes based on underlying genetic risk. (Accepted)
23. Mosley J, Denny J, Roden D, Wang T, Bastarache L, Shaffer C, Edwards T, McCarty C, Thompson W, Chute C, Jarvik G, Crosslin D, Larson E, Kullo I, Pacheco J, Peissig P, Brilliant M, Linneman J, Namjou Khales B, Ritchie M, Borthwick K, Verma SS, Carrell D, Prows C, Kiryluk K, Gordon A, Devi P, Mentch F, Sleiman P, Pendergrass S, and Kitchner T Framingham collaborators. Identifying clinical phenotypes associated with serum protein and metabolite levels. (Accepted)
24. Sood N, Hysinger E, Sleiman P, Mentch F, Vazquez L, Connolly J, Almoguera B, Hakonarson H, Jarvik G, Crosslin D, Gordon A, Devi P, Larson E, Carrell D, Manolio T, Borthwick K, Sundaresan A, Weiss S, Karlson B, Lasky-Su J. Influence of Guideline Adherence and ADRB2 SNPs in Predicting Exacerbation Frequency in Asthma Patients. (Accepted)
25. Kullo I, Fan X, Olson J, Jose M, Safarova M, Radecki Breitkopf C, Winkler E, Snipes S, Kohan DC, Carney M, Pacyna J, Chute CG, Gupta J, Jose S, Venner E, Murugan M, Jiang Y, Zordok M, Farwati M, Philogene M, Smith E, Shaibi GQ, Caraballo P, Freimuth R, Lindor NM, Sharp R, Thibodeau SN. The Return of Actionable Variants Empiric (RAVE) Study, a Mayo Clinic Genomic Medicine Implementation Study: Design and Initial Results. Mayo Clinic Proc. 2018; 1600-1610. PMID: 30392543
26. Williams JL, Chung WK, Fedotov A, Kiryluk K, Weng C, Connolly JJ, Harr M, Hakonarson H, Leppig KA, Larson EB, Jarvik GP, Veenstra DL, Hoell C, Smith ME, Holm IA, Peterson JF, Williams MS. Harmonizing Outcomes for Genomic Medicine:Comparison of eMERGE Outcomes to ClinGen Outcome/Intervention Pairs. Healthcare (Basel). 2018 Jul 13;6(3)
27. Overby Taylor C, Weiner J, Hripscak G, Walton N, Borthwick K, Denny J, Kiryluk K, Carrell D, Wei W, Larson E, Peissig P, Ye Z, Kullo I. Pilot evaluation of ACGs for characterizing co-morbidities in eMERGE cohorts. (Accepted)
28. Chiang T, Gibbs R, Wu K, Muzny D, Wu T, Bi W, Leduc M, Meng L, Wen S, Yaping Y, Eng C, Fritz S, Salerno W, Beorwrinkle E, Venner E, Garvik G, Schaid D, Leppig K, Carrell D, Andrade M, Connolly J, Hakonarson H. Atlas-CNV: Calling single and multi-exon CNVs with low FDR from the eMERGE clinical gene targeted data. 2019; 2135-2144. PMID: 30890783
29. Liu J, Finkelstein J. Towards Pharmacogenomics-Driven Mediciation Risk Assessment in People with Polypharmacy. Stud Health Technol Inform. 2018; 247:880-884. PMID: 29678087
30. Stanaway I, Hall TO, Rosenthal EA, Palmer M, Naranbhai V, Knevel R, Namjou-Khales B, Carroll RJ, Kiryluk K, Gordon AS, Linder J, Howell KM, Mapes BM, Lin FTJ, Joo YY, Hayes MG, Gharavi AG, Pendergrass SA, Ritchie MD, de Andrade M, Croteau-Chonka DC, Raychaudhuri S, Weiss ST, Lebo M, Amr SS, Carrell D, Larson EB, Chute CG, Rasmussen-Torvik LJ, Roy-Puckelwartz MJ, Sleiman P, Hakonarson H, Li R, Karlson EW, Peterson JF, Kullo IJ, Chisholm R, Denny JC, Jarvik GP, eMERGE Network, Crosslin DR. The eMERGE genotype set of 83,717 subjects imputed to ~40 million variants genome wide and association with the herpes zoster medical record phenotype. Genet Epidemiol. 2018 Oct 8. PMID: 30298529
31. Wei W, Denny J, Krauss R, Laarson E, Jarvik G, Elkind M, Shang N, Hripcsak G, Weng C, Fasel D, Peissig P, Ritchie M, Lee M, Cronkite D. LPA variants are associated with residual cardiovascular risk in patients receiving statins. Circulation. 2018 Oct 23;138(17):1839-1849. Doi: 10.1161/CIRCULATIONAHA.117.031356. PMID: 29703846
32. Almoguera B, Vazquez L, Mentch F, Connolly J, Sleiman P, Almoguera B, Green C, Linneman J, Brilliant M, Borthwick KM, Sundaresan A, Williams M, Hakonarson H. Novel locus for atopic dermatitis in African Americans and replication in European Americans. J Allergy Clin Immunol. 2019 Mar;143(3):1229-1231. Doi: 10.1016/j.jaci.2018.10.038. Epub 2018 Nov 9. PMID: 30414857
33. Zhong Y, Luo Y, Rasmussen L, Starren J. Classifying design patterns of I-driven phenotyping algorithms. 2018 Nov 15 [cited 2019 Jan 3]; Available from: http://arxiv.org/abs/1811.06183
34. Hall T, Jarvik G, Stanaway I, Crosslin D, Rosenthal E, Hakonarson H, Mentch F, Pendergrass S. Unfolding of hidden white blood cell count phenotypes for gene discovery using latent class mixed modeling. Genes Immun. 2018 Nov 21. PMID: 30459343
35. Salem J, Roden D, Mosley J, Denny J, Shoemaker B, Ellinor P, Jarvik G, Jose S, Larson E, Velez Edwards D, Ramirez A, Davis L, Hayes G, Hakonarson H, Weng C, Edwards T, Sleiman P, Fasel D. Identifying clinical phenotypes associated with genetically predicted TSH levels. JAMA Cardiol. 2019 Jan 23. Doi: 10.1001/jamacardio.2018.4615. [Epub ahead of print].
36. Herr T, Peterson J, Rasmussen LV, Caraballo P, Peissig P, Starren J. Pharmacogenomic clinical decision support design and multi-site process outcomes analysis in the eMERGE Network. J Am Med Inform Assoc. 2019 Feb 1;26(2):143-148. Doi: 10.1093/jamia/ocy156. PMID: 30590574
37. Mercaldo N, Schildcrout J, Behrens J, Pacheco J, Horowitz C, Hitz P, Ziniel S, DeWalle J, Williams J, Shrubsole M. Enrichment sampling for a multi-site patient survey using electronic health records and census data. J Am Med Inform Assoc. 2019 Mar 1;26(3):219-227. Doi: 10.1093/jamia/ocy164. PMID: 30590688
38. Ko D, Crosslin D, Denny J, Jarvik G, Pittman K, Gibbs K, Barker J, Gopalakrishnan A, Salinas R, Antonia A, Glover L, Balmat T, Ingham A, Delong M, Cao Y, Lee S, Heitman J, Valdivia R, Harley J, Namjou B, Larson E, Kiryluk K, Stanaway I, Jarvik G, Wiesner, G, Connolly J. An atlas of genetic variation connecting cell biology to human disease. Cell Host & Microbe. 2018 Aug 8;24(2):308-323.

In Process Network Manuscripts

1. The Association between Variants in Ion Channel Genes and Arrhythmia Phenotypes. Lead Investigator: Ben Shoemaker
2. The Association between Variants in Long QT Syndrome Susceptibility Genes and ECG features. Lead Investigator: Ben Shoemaker
3. Does clinical context expand the clinical implications of actionable genetic findings and high PRS scores? Lead Investigator: Hila Rasouly
4. Development, refinement and validation of polygenic risk score incorporating family and environmental data for colorectal cancer in multiple ancestry data. Lead Investigator: Elisabeth A. Rosenthal
5. Lessons from eMERGE on readiness for genomic clinical decision support implementation. Lead Investigator: Casey Overby Taylor
6. eMERGE PheWAS catalog of GWAS SNPs. Lead Investigator: Todd Edwards
7. Phenotypic signatures of genome-wide polygenic risk scores for complex traits. Lead Investigator: Atlas Khan
8. Family communication following return of positive results. Lead Investigator: Lead Investigator: Hila Milo Rasouly and Julia Wynn
9. Mendelian randomization study of the association of lipid risk SNPs with the development of breast, prostate and other cancers. Lead Investigator: Daniel Lee & Shefali Verma
10. SIM1 Rare Variation Association with Erectile Dysfunction. Lead Investigator: Gail Jarvik & Ian Byrell Stanaway
11. Genetics of Osteoarthritis Consortium. Yanfei Zhang & Ming Ta Michael Lee
12. CCHMC top 6 gene- MC4R; a brief-report—evaluation and prevalence of MC4R mutations among emerge-seq population. Lead Investigator: Bahram Namjou & John Harley
13. Pleiotropic Associations Between Predicted JAK Pathway Genes' Expression and the Clinical Phenome. Lead Investigator: Jacklyn N. Hellwege
14. Associations Between Sex-Specific Renal Function Polygenic Risk Scores and the Clinical Phenome by Sex. Lead Investigator: Jacklyn N. Hellwege
15. Comorbidity Clusters in Clinical Conditions: An Analysis of Electronic Health Record Data. Lead Investigator: Ting He
16. Genomic Information for Clinicians in the Electronic Health Record: Lessons Learned from ClinGen and eMERGE. Lead Investigator: Marc Williams
17. Sequencing Centers and eMERGE Site Interactions related to Return of Genomic Results in Phase III of the eMERGE Network. Lead Investigator: David Kochan
18. Phenotype risk scores identify patients at a high risk of hereditary cancer syndromes and improve variant interpretation. Lead Investigator: Chenjie Zeng & Lisa Bastarache
19. Effect of genomic regulation on Clopidogrel Response in African Americans. Lead Investigator: Tanima De
20. Gender Genotype Analysis for Clinical Quality Assessment. Lead Investigator: Jianhong Hu
21. Evaluating the 'Star Allele' PGx nomenclature standard in the context of automated interpretation of panel, exome, and genome sequencing results. Lead Investigator: Adam Gordon (KPW/UW)
22. Network-wide lessons learned from the reporting of negative test results. Lead Investigator: Richard Sharp (Mayo) & Maureen Smith (NU)
23. GWAS for Lupus Identified with a Classification Criteria-Based Phenotyping Algorithm. Lead Investigator: Theresa Walunas (NU)
24. Approaches to the return of actionable adult-onset conditions in pediatric research: Lessons learned from eMERGE 3. Lead Investigator: Ingrid Holm (BCH)
25. Framework for Assessing NLP-based I Phenotype Algorithm Complexity. Lead Investigator: Yuan Luo (NU)
26. Genomic Data: Building a Path from Lab to Clinic. Lead Investigator: Nephi Walton (Geisinger)
27. A Study of Phenotype Algorithm Portability. Lead Investigator: Ning Shang (Columbia)
28. Psychiatric Manifestations of Variations in ACMG59 Genes. Lead Investigator: Y-C Feng (Harvard)
29. Early prediction of risk for Alzheimer's disease and related dementia using data-driven, scalable analysis of electronic health records and genetic data. Lead Investigator: Ji-Hwan Park (External)
30. Replication of Million Veterans Program GWAS on Abdominal Aortic Aneurysm. Lead Investigator: Derek Klarin (External)
31. Challenges in Returning Results in the eMERGE consortium. Lead Investigator: Colin Halverson (VUMC)
32. The Reckoning: What We Found After Return of Results for 25,000 eMERGE3 participants. Lead Investigator: Kathy Leppig (KPW/UW)
33. Pleiotropic Associations Between a Uterine Leiomyoma Polygenic Risk Score (PRS) and the Clinical Phenome. Lead Investigator: Jacklyn N. Hellwege (VUMC)
34. Identify potential risk genes for common calcium kidney stones. Lead Investigator: Krzysztof Kiryluk (Columbia)
35. Use of Infobuttons to Find Answers to Clinician's Questions in Clinical Genomics. Lead Investigator: Michael Watkins (External)
36. Novel encoding method EDGE offers enhanced ability to identify genetic interactions. Lead Investigator: Molly Hall (External)

37. Eliminating Genomic Medicine Research Barriers through Informatics: the eMERGE Network's Resource Library and its Impact on Multi-Site Projects. Lead Investigator: Brandy Mapes (CC) & Luke Rasmussen (NU)
38. The Development of an Imputed Structural Variant Genomic Dataset and Association to Neurological and Alcohol Use Disorder Electronic Medical Record Phenotypes with Biobank Scale Subject Ascertainment. Lead Investigator: Ian Stanaway (KP/UW)
39. Assess penetrance of cancer among mutations carriers for hereditary breast cancer genes. Lead Investigator: Katherine Crew (Columbia)
40. Genetic variants associated with Osteoarthritis (Pan-OA analysis). Yanfei Zhang (Geisinger)
41. Penetrance and outcome of pulmonary hypertension associated with germline BMPR2 mutations in a population-based cohort. Lead Investigator: Na Zhu (Columbia) & Carrie Welch (Columbia)
42. Evaluate in silico prediction of pathogenicity of missense variants using sequence data and I phenotypes of the eMERGE III sequencing cohort. Lead Investigator: Haicang Zhang (Columbia) & Xiao Fan (Columbia)
43. An efficient algorithm for distributed analysis of pleiotropy in Electronic Health Record data. Lead Investigator: Ruowang Li (External)
44. Genome-wide association studies of spinal disorders and pain. Lead Investigator: Pradeep Suri (External)
45. Risk of colorectal cancer and age of CRC onset associated with HFE C282Y/C282Y homozygosity. Lead Investigator: Gail Jarvik (KPW/UW)
46. Evaluating the utility of high-throughput functional scoring of BRCA1 missense variants in eMERGE-III participants. Lead Investigator: Adam Gordon (NU)
47. Genetic variants associated with metformin response and intolerance. Lead Investigator: Suzette Bielinski (Mayo)
48. Predictive Utility of Polygenic Risk Scores for Coronary Heart Disease in the eMERGE Network. Lead Investigator: Ozan Dikilitas (Mayo)
49. eMERGE Harvard Site Top 6 Genes. Lead Investigator: Samira Asgari (Harvard)
50. IGNITE Clinical Informatics Working Group: Genetic Data Pipeline Project. Lead Investigator: Paul Dexter (External)
51. Understanding the return of results process: Content review of patient summary letters. Lead Investigator: John Lynch (CCHMC)
52. Genetic Loci for Polycystic Ovary Syndrome (PCOS). Lead Investigator: Yanfei Zhang and Kevin Ho (Geisinger)
53. Association between variants in FBN1 and human height. Lead Investigator: Samira Asgari (Partners/Broad)
54. Creation of an OMOP-Based Phenotype Algorithm for Systemic Lupus Erythematosus (SLE). Lead Investigator: Theresa Walunas (NU)
55. Collection and Analysis of Large-Scale Outcome Measures following Targeted Next Generation Sequencing. Lead Investigator: Josh Peterson (CC)
56. Exploring the vagueness of definitions in a phenotype algorithm. Lead Investigator: Anika Ghosh (Northwestern)
57. Framework for Assessing I-Based Phenotype Algorithm Complexity. Lead Investigator: Luke Rasmussen (NU)
58. Discovery of candidate SNP biomarkers for predicting the risk for periprosthetic osteolysis and other adverse outcomes in joint arthroplasties. Lead Investigator: Yelizaveta Torosyan (External)
59. Genome-wide association study of carotid artery atherosclerosis disease in eMERGE. Lead Investigator: Melody Palmer (KPW/UW)
60. How genomic patterns are linked to phenotypic patterns. Lead Investigator: Shefali Setia Verma (External)
61. Genetic risk factors for Intracerebral Hemorrhage and Subarachnoid Hemorrhage: A Meta-Analysis of Genome-Wide Association Studies. Lead Investigator: Sarah Pendergrass (Geisinger)
62. Detecting pleiotropy across neurological disorders and cardiovascular diseases via multi-trait joint association analysis. Lead Investigator: Xinyuan Zhang (External)
63. Concordance between pharmacogenomic results extracted from research generated whole exome sequencing data and a CLIA generated next generation sequencing platform. Lead Investigator: Rebecca Pulk (Geisinger)
64. Phenotype risk score (PRS) to evaluate effects of TTR variants and phenotypic spectrum of transthyretin-related hereditary amyloidosis. Lead Investigator: Quinn Wells (VUMC)
65. Validation of PharmCAT annotations. Lead Investigator: Aurage Verma (External)
66. Genetic Architecture of Anxiety Disorders. Lead Investigator: John Connolly (CHOP)
67. Discovery-based CNV Analyses of eMERGE-seq data. Lead Investigator: Patrick Sleiman (CHOP)
68. GWAS-PheWAS Approach to Infection-Associated Stroke. Lead Investigator: Neal Parikh (Columbia)
69. Genetic resistance to common life-threatening infections and its evolutionary effects on susceptibility to complex immune traits. Lead Investigator: Ning Shang (Columbia)
70. Operationalizing participant choices about genomic results: Beyond all or none ACMG recommended genes. Lead Investigator: Christin Hoell (NU)
71. A Phenome-Wide Association Study of Obstructive Sleep Apnea Candidate Genes. Lead Investigator: Sarah Pendergrass (Geisinger)
72. New Genetic Loci for Obstructive Sleep Apnea. Lead Investigator: Sarah Pendergrass (Geisinger)

73. Returning genomic results to eMERGE participants: The who, what, where, and how of disclosure. Lead Investigator: Georgia Wiesner (VUMC)
74. Design and testing of clinical decision support tools for 8 ACMG genes in 3 CDC Tier 1 Genetic Conditions included in the eMERGE-seq panel. Lead Investigator: Jordan Nestor (Columbia)
75. Preferences for research updates among biobank participants. Lead Investigator: Casey Overby Taylor (Geisinger)
76. Common and rare variant association of Opioid Addiction using the network-wide eMERGE 3 cohort. Lead Investigator: Shawn Murphy (Harvard)
77. Development of ancestry specific polygenic risk scores and exploration of genetic-environmental interactions affecting vitamin D concentrations. Lead Investigator: Kathryn E. Hatchell (External)
78. Shared and distinct genetics of childhood asthma and adult obstructive lung disease. Lead Investigator: Su H. Chu (Harvard)
79. A PheWAS approach to identify clinical correlates of rare variation within genes linked to Hereditary Spastic Paraplesia. Lead Investigator: Adam Gordon (KPW/UW)
80. Genetic risk for gout in patients with hyperuricemia. Lead Investigator: Rachel Knevel (Harvard)
81. Comparison of clinical interpretation of genetic variation in the eMERGEseq cohort using multiple methods. Lead Investigator: David Crosslin (KPW/UW)
82. Evaluating global and local ancestry and admixture across PheWAS phenotypes in eMERGE-III. Lead Investigators: Digna R Velez Edwards and Soumya Raychaudhuri (VUMC)
83. Association of CYP2D6 genotypes in the eMERGE PGx Cohort with I-derived phenotypes. Lead Investigator: Iftikhar Kullo (Mayo)
84. Replication of novel lipids-associated loci in eMERGE. Lead Investigator: Yao Hu (External)
85. Phenotype risk scores to estimate phenotypic effects of rare variants. Lead Investigator: Lisa Bastarache (VUMC)
86. Replication of GWAS of Cisplatin-induced Ototoxicity. Lead Investigator: Eileen Dolan (External)
87. Deep Phenotyping in electronic health records for facilitating diagnosis of genetic disorders. Lead Investigator: Chunhua Weng (Columbia)
88. Association of Genes and Variants in the eMERGEseq Panel with LDL cholesterol and Triglyceride Levels. Lead Investigator: Xiao Fan (Mayo)
89. Epistasis of HLA with trans loci alleles. Lead Investigator: Ian Stanaway (KPW/UW)
90. PheWAS of ABO Blood Groups. Lead Investigator: Ian Stanaway (KPW/UW)
91. Association Studies of Variants in KCNQ1, KCNH2, RYR2, SCN5A, ANK2, CACNA1C, and KCNE1 with arrhythmia and ECG phenotypes in 25,000 eMERGE 3 participants. Lead Investigator: Ben Shoemaker (VUMC)
92. Association of Obesity with Postoperative Complications Using Phenome-wide Association Studies and Mendelian Randomization. Lead Investigator: Jamie Robinson (VUMC)
93. Mendelian Randomization to Identify Phenotypes and Procedures Associated with BMI and Obesity. Lead Investigator: Jamie Robinson (VUMC)
94. Association Studies of Rare Variants in HNF1B, UMOD, WT1 and CFH and Known Risk Alleles with Chronic Kidney Disease in 25,000 eMERGE 3 participants. Lead Investigator: Miguel Verbitsky (Columbia)
95. 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. Lead Investigator: Katherine Crew (Columbia)
96. A comparison of genetic effects for migraine in children versus adults using eMERGE participant. Lead Investigator: Bahram Namjou (CCHMC)
97. Clinical outcomes after screening for cardiomyopathy genes. Lead Investigator: Christin Hoell (NU)
98. Genome-wide Association Study of Chronic Rhinosinusitis in Adult population. Lead Investigator: Agnes Sundaresan (Geisinger)
99. Identifying clinical phenotypes associated with echocardiographic indices of right heart structure and function. Lead Investigator: Jonathan Mosley (VUMC)
100. Association analysis of triglycerides with SLC25A40 sequence data in 25,000 eMERGE participants. Lead Investigator: Elisabeth A. Rosenthal (KPW/UW)
101. Association analysis of neutrophil count with TCIRG1 sequence data in 25,000 eMERGE participants. Lead Investigator: Elisabeth A. Rosenthal (KPW/UW)
102. Establishing a patient-specific disease model of an inherited arrhythmia reveals a general arrhythmogenic signaling pathway. Lead Investigator: Kevin Bersell (VUMC)
103. GWAS of Medicare Risk Adjustment Model scores. Lead Investigator: Scott Hebbring (Marshfield)
104. PheWAS of Polycystic Ovary Syndrome (PCOS) GWAS loci in the electronic health records of 38,000 adult subjects in the eMERGE Consortium. Lead Investigator: Yoonie Joo (NU)
105. The Genetic Architecture of Severe and Familial Hypercholesterolemia. Lead Investigator: Xiao Fan (Mayo)
106. Age-specific effects of common genetic variants and human leukocyte antigen types. Lead Investigator: Jeff Goldstein (VUMC)

107. APOB, PLTP, and PON1 eMERGEseq association with lipid and vascular phenotypes. Lead Investigator: Melody Palmer (KPW/UW)
108. eMERGE GWAS for GIANT meta-analysis. Lead Investigator: Sarah Pendergrass (Geisinger)
109. Genome-Wide Association Studies for Lipid Traits in the eMERGE network. Lead Investigator: Sarah Pendergrass (Geisinger)
110. GWAS and PheWAS for acne gene discovery and comorbidity assessment. Lead Investigator: Lynn Petukhova (Columbia)
111. Portable Precision Phenotype Algorithm for Chronic Rhinosinusitis with and without Nasal Polyps using the I. Lead Investigator: Jen Pacheco (NU)
112. Genetic studies of central centrifugal cicatricial alopecia. Lead Investigator: Lynn Petukhova (Columbia)
113. Genetic assessment of neuropsychiatric and metabolic comorbidities among autoimmune disease patients. Lead Investigator: Lynn Petukhova (Columbia)
114. In Silico Integration of Epidemiologic and Genetic Evidence on the Sex/Race-Related Modifying Effects on Hip Arthroplasty Outcomes. Lead Investigator: Yelizaveta Torosyan (External)
115. Characterizing the costs of implementing genomic clinical decision support. Lead Investigator: Patrick Mathias (KPW/UW)
116. The relative impact of environmental and genetic factors on phenotypic expression of disease. Lead Investigator: Kathryn Jackson (NU)
117. I Phenotyping of Atopic Dermatitis in Adults using more detailed data including from NLP. Lead Investigator: Alona Furmanchuk (NU)
118. Machine learning-based discovery of Atopic Dermatitis (AD) Sub-Populations in Adults. Lead Investigator: Alona Furmanchuk (NU)
119. Infobutton Genomic Medicine Initiatives Survey. Lead Investigator: Luke Rasmussen (NU)
120. The Genetic Architecture of Auto-Inflammatory and Auto-Immune Diseases. Lead Investigator: Patrick Sleiman (CHOP)
121. Identifying relationship between obesity and post-operative complications through Mendelian randomization. Lead Investigator: Tom Mou (VUMC)
122. 22Q11.2 Deletion Syndrome, Leveraging Copy Number Variation to Examine Health Outcomes. Lead Investigation: Patrick Sleiman (CHOP)
123. Genomewide association studies of developmental disorders of speech, language, and scholastic skills. Lead Investigator: Reyna Gordon (VUMC)
124. New approaches to the genetic basis of developmental language disorder. Lead Investigator: Reyna Gordon (VUMC)
125. CNV Association of quantitative and discreet traits across eMERGE-II/III array and PGRNSeq datasets. Lead Investigator: Hakon Hakonarson (CHOP)
126. Incidental and secondary Findings (Ifs) in 10,000 eMERGE participants. Lead Investigator: Adam Gordon (KPW/UW)
127. AutoImmuneMERGE – PheWAS for major Auto-Immune Diseases, whole genome sequencing. Lead Investigator: Rachel Knevel (Harvard)
128. Lipid distribution in Pediatric population. Lead Investigator: Agnes Sundaresan (Geisinger)
129. Common and rare variation associated with valvular disease in eMERGE 3. Lead Investigator: Laura Rasmussen Torvik (NU)
130. Genomics of structural kidney and urinary tract defects. Lead Investigator: Miguel Verbitsky (Columbia)
131. A Phenome-wide Survey of the Phenotypic Effects of High- Frequency Human-Derived Alleles. Lead Investigator: Tony Capra (VUMC)
132. Pharmacogenetic variation identified via targeted next-generation sequencing among 9000 eMERGE subjects. Lead Investigator: Adam Gordon (KPW/UW)
133. Comprehensive genetic association study of kidney traits across the EMERGE network. Lead Investigator: Krzysztof Kiryluk (Columbia)
134. Detection of copy number variants (CNVs) and their kidney disease associations across the EMERGE network. Lead Investigator: Miguel Verbitsky (Columbia)
135. Combined GWAS-PheWAS Approach to Serologic Markers of Autoimmunity & Inflammation. Lead Investigator: Krzysztof Kiryluk (Columbia)
136. Knowledge driven rare variant PheWAS in eMERGE to identify regions associated with disease using collapsing based approach. Lead Investigator: Marylyn Ritchie (External)
137. Possible somatic mutation in targeted sequence data. Lead Investigator: Kenneth Kaufman (CCHMC)
138. 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 (KPW/UW)
139. An investigation of somatic mutations in PGx-eMERGE dataset. Lead Investigator Kenneth Kaufman (CCHMC)
140. Quantitative and discreet trait analysis across eMERGE-II phenotypes. Lead Investigator: Hakon Hakonarson (CHOP)
141. Genome-wide Association Study of Gastroesophageal Reflux Disease (GERD) in Adult and Pediatric Populations. Lead Investigator: Patrick Sleiman (CHOP)
142. Using PheWAS to assess disease comorbidity and potential pleiotropy of genetic risk scores for rheumatoid arthritis. Lead Investigator: Robert Carroll (VUMC)

143. Association of Variation in 84 Pharmacogenes with Low-density Lipoprotein Cholesterol Levels in the eMERGE-PGx Project Lead Investigator: Xiao Fan (Mayo)
144. Multiscale Analysis of Influenza Host-Pathogen Interactions: Fluomics. Lead Investigator: Eun-Young Kim (NU)
145. Burden of structural variation and PheWAS. Lead Investigator: Adam Gordon (KPW/UW)
146. Copy number variation burden analysis on a range of phenotypes in the eMERGE network. Lead Investigator: Dokyoon Kim (Geisinger)
147. Chromosomal anomalies that affect levels of white blood count (WBC) and its differential. Lead Investigator: Melody Palmer (KPW/UW)
148. Extracting the Quality of Prostate Cancer Care from Electronic Healthcare Records. Lead Investigator: Tina Hernandez-Boussard (External)
149. Colon Polyps – GWAS. Lead Investigator: Laura Rasmussen-Torvik (NU)
150. Genetic Variants Associated with Response to Heart Failure Treatment: The Electronic Medical Records and Genomics (eMERGE) Network. Lead Investigator: Suzette Bielinski (Mayo)
151. Genome-Wide Association of Risk of Heart Failure: The Electronic Medical Records and Genomics (eMERGE) Network. Lead Investigator: Suzette Bielinkski (Mayo)
152. Genetic variation that predicts susceptibility to Clostridium difficile. Lead Investigator: Taryn Hall (KPW/UW)
153. Enhancing Scientific Productivity through Collaboration: Best Practices for Development of a Network Data Sharing Agreement. Lead Investigator: Melissa Basford (CC)
154. Population structure of a metropolitan biobank sample and the clinical implications for research. Lead Investigator: Tess Pottinger (NU)
155. Structural Variation identified from whole genome sequencing of 1200 biobank participants. Lead Investigator: Megan Pucklewartz (NU)

Site-Specific Manuscripts

CCHMC

1. B. Namjou, T. Lingren, Y. Huang, S. Parameswaran, B.L. Cobb, I.B. Stanaway, J.J. Connolly, F.D. Mentch, B. Benoit, X. Niu, W.-Q. Wei, R.J. Carroll, J.A. Pacheco, I.T.W. Harley, S. Divanovic, D.S. Carrell, E.B. Larson, D.J. Carey, S. Verma, M.D. Ritchie, A.G. Gharavi, S. Murphy, M.S. Williams, D.R. Crosslin, G.P. Jarvik, I.J. Kullo, H. Hakonarson, R. Li, S.A. Xanthakos, J.B. Harley, and M.N. The eMERGE Network, GWAS and enrichment analyses of non-alcoholic fatty liver disease identify new trait-associated genes and pathways across eMERGE Network. 2019 Jul 17;17(1):135. PubMed PMID: 31311600
2. Ni Y, Bermudez M, Kennebeck S, Liddy-Hicks S, Dexheimer JW. A Real-Time Automated Patient Screening System for Clinical Trials Eligibility in an Emergency Department: Design and Evaluation. 2019 Jul 24;7(3). PubMed PMID: 31342909
3. Pervola J, Myers MF, McGowan ML, Prows CA. Giving adolescents a voice: the types of genetic information adolescents choose to learn and why. 019 Apr;21(4):965-971. PubMed PMID: 30369597

Harvard

4. Jorge A, Castro VM, Barnado A, Gainer V, Hong C, Cai T, Cai T, Carroll R, Denny JC, Crofford L, Costenbader KH, Liao KP, Karlson EW, Feldman CH. Identifying lupus patients in electronic health records: Development and validation of machine learning algorithms and application of rule-based algorithms. 2019 Aug;49(1):84-90. PubMed PMID: 30665626
5. Ge T, Chen C-Y, Ni Y, Feng Y-CA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nat Commun. 2019 April 6; 10(1):1-10; PubMed PMID: 30992449
6. Hripcsak G, Shang N, Peissig PL, Rasmussen LV, Liu C, Benoit B, Carroll RJ, Carrell DS, Denny JC, Dikilitas O, Gainer VS, Marie Howell K, Klann JG, Kullo IJ, Lingren T, Mentch FD, Murphy SN, Natarajan K, Pacheco JA, Wei WQ, Wiley K, Weng C. Facilitating phenotype transfer using a common data model. J Biomed Inform. 2019 August; 96: 1-7.
7. Zhang X, Veturi Y, Verma S, Bone W, Verma A, Lucas A, Hebbring S, Denny JC, Stanaway IB, Jarvik GP, Crosslin D, Larson EB, Rasmussen-Torvik L, Pendergrass SA, Smoller JW, Hakonarson H, Sleiman P, Weng C, Fasel D, Wei WQ, Kullo I, Schaid D, Chung WK, Ritchie MD. Detecting potential pleiotropy across cardiovascular and neurological diseases using univariate, bivariate, and multivariate methods on 43,870 individuals from the eMERGE network. 24:272-283. PubMed PMID: 30864329
8. Zheutlin AB, Dennis J, Restrepo N, Straub P, Ruderfer D, Castro VM, Chen C-Y, Kirchner HL, Chabris CF, Davis LK, Smoller JW. Penetrance and pleiotropy of polygenic risk scores for schizophrenia in 106,160 patients across four healthcare systems. Am J Psychiatry. 2019 Aug 16. PubMed PMID: 31416338

Geisinger

9. Snyder SR, Hao J, Cavallari LH, Geng Z, Elsey A, Johnson JA, Mohamed Z, Chaiyakunapruk N, Chong HY, Dahlui M, Shabaruddin FH, Patrinos GP, Mitropoulou C, Williams MS. Generic Cost Effectiveness Models: A Proof of Concept of a Tool for Informed Decision-making for Public Health Precision Medicine. Epub 2019 Jun 12. PubMed PMID: 31189173
10. Patel P, Hu Y, Kolinovsky A, Tripathi B, Geng Z, Ruhl J, Krishnamurthy S deRichemond C, Khan A, Kirchner L, Raghu Metpally R, Jones L, Schwartz M, Sturm A, Ledbetter DH, Carey DJ, Snyder S, Williams MS, Blankenship J, Mehra VC. The Hidden Burden of I-Identified Familial Hypercholesterolemia: Clinical Outcomes and Cost of Medical Care. Epub 2019 Jun 29. PubMed PMID: 31256702

KP/UW

11. Henrikson, N. B., P. R. Blasi, J. J. Corsmo, E. Sheffer Serdoz, A. Scrol, S. M. Greene, T. L. Matthews and J. D. Ralston. You Really Do Have to Know the Local Context": IRB Administrators and Researchers on the Implications of the NIH Single IRB Mandate for Multisite Genomic Studies. Epub 2019 May 14. PubMed PMID: 31113270.

VUMC

12. Carrell DS, Cronkite DJ, Li M, Nyemba S, Malin B, Aberdeen J, Hirschman L. The machine giveth and the machine taketh away: a parrot attack on clinical text de-identified with hiding in plain sight. 2019 Aug 7. 1-9. PubMed PMID: 31390016

OVERVIEW *of e* MERGE TOOLS

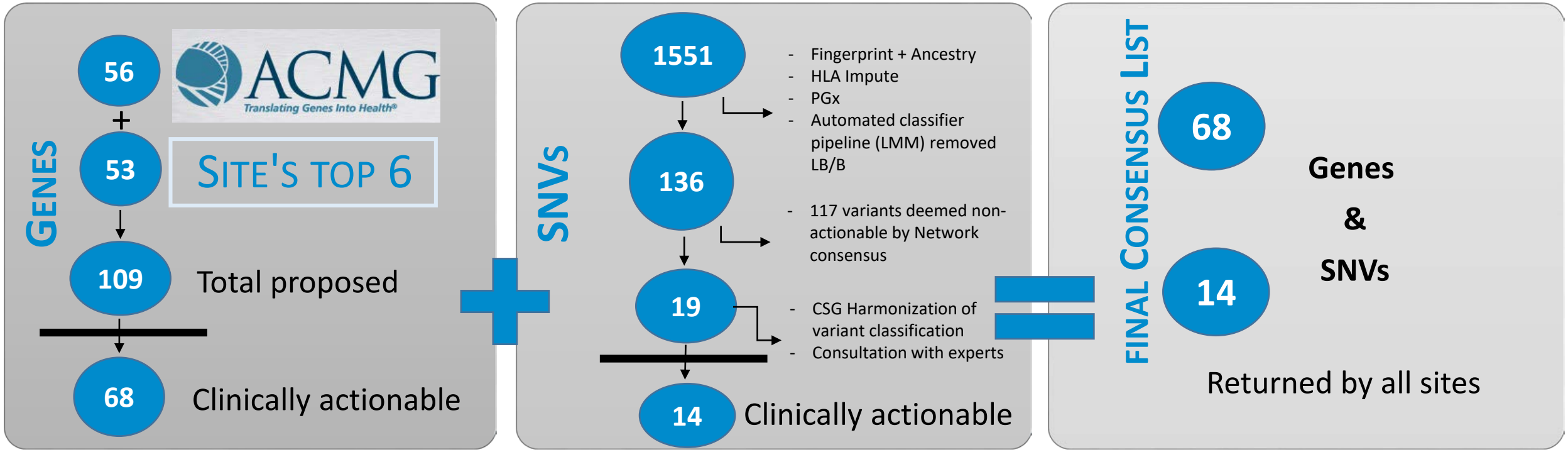
Main resources and tools produced and supported *(click on blue links):*

- ANNOVAR: A tool that aids in annotation of genetic variants from diverse genomes
- CDS KB: A repository of clinical decision support knowledge designed to support clinical processes from diagnosis and investigation through treatment and long-term care.
- DocUBuild: A tool that provides infrastructure to support sharing genomic medicine content in a way that is compliant with the HL7 infobutton standard.
- 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 of PheWAS results for single-nucleotide polymorphisms (SNPs) present in the NHGRI GWAS Catalog, HLA types, and introgressed Neanderthal variants.
- SPHINX: A data exploring tool for genetics related drug response hypothesis generation, especially around drug response implications of genetic variation across the eMERGE PGx cohort.
- Synthesis-View, PheWAS-View, PheWAS R Package and Phenogram: Visualization tools for genome & phenome-wide data.
- IMAGene: An educational website including both English and Spanish about the use of personalized medicine to improve health care for everyone.
- Learning Genetics: Learn about exome sequencing and secondary findings.

e MERGE PHASE III PHENOTYPES

Primary Site	Phenotype
CCHMC	Migraine
	Fatty Liver condition = (NAFLD/NASH-Alcoholic)
	Post-Event Pain
CHOP	Epilepsy
	Intellectual disability
	Anxiety
Columbia	Chronic Kidney Disease (eGFR, proteinuria)
	Autoimmunity
	Breast cancer
Geisinger	Pediatric Familial Hypercholesterolemia (FH)
KPW/UW	Colorectal Cancer (CRC)
	Ovarian and Uterine Cancer
	Depression
Harvard	Rheumatoid arthritis
	COPD/ACO
	Bipolar disorder
Mayo	Adult Familial Hypercholesterolemia (FH)
	Contrast induced nephropathy (PGx)
	Peripheral arterial disease
	Metformin response (PGx)
Northwestern	Chronic Rhinosinusitis
	Atopic Dermatitis
	Lupus
VUMC	Hearing Loss
	Arrhythmias
	Pneumonia

CLINICAL REPORTING: Overview *of* Consensus Lists *for* *e*MERGE III



Comprehensive List of Genes and SNVs on Next Two Slides

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

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

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

‡ Site TOP-6 genes are indicated in blue

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

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

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

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

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

CSG Status Update: Sequencing & Reporting

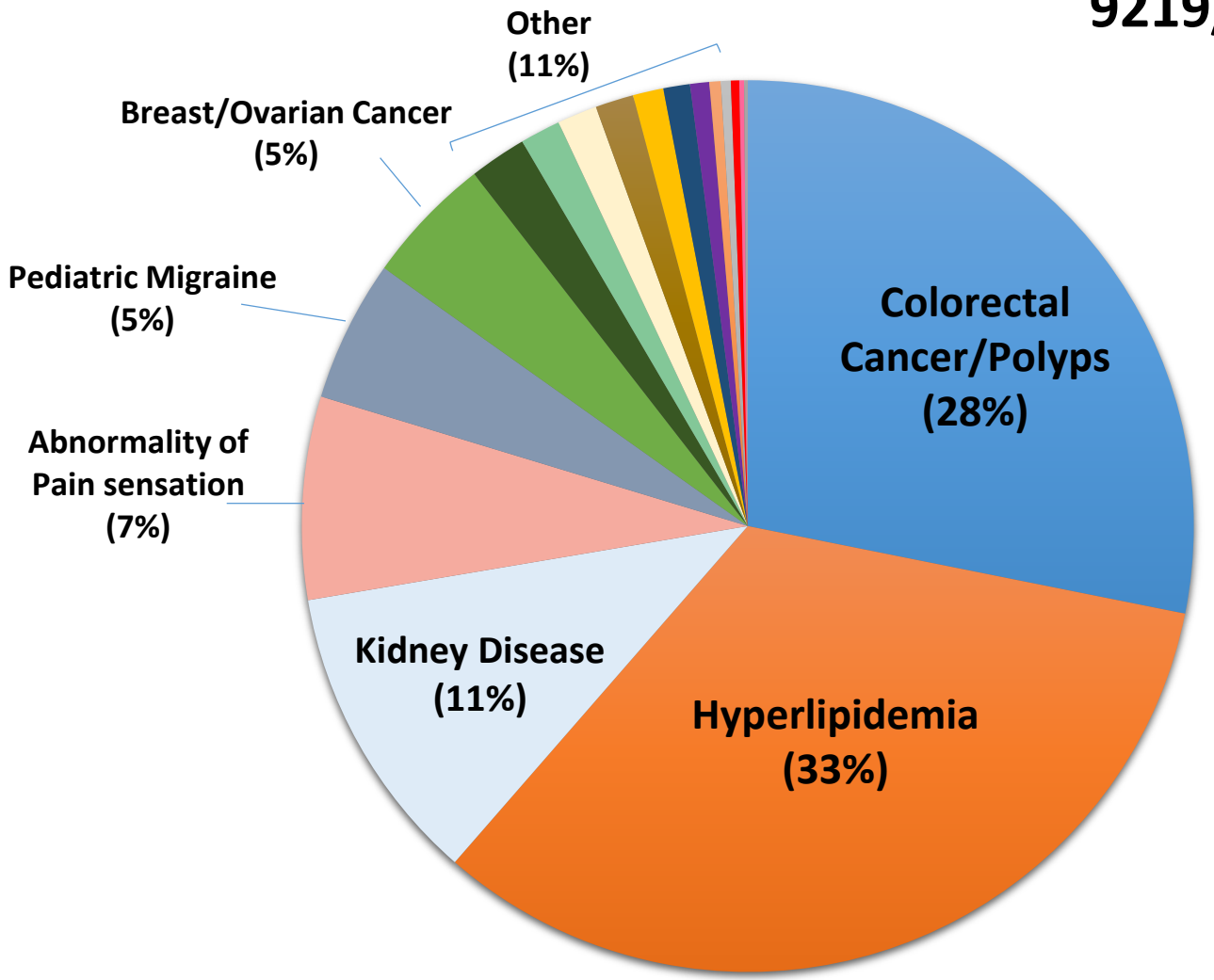
		Samples received			Sequencing status and performance				Reporting	
CSG	Site	Round 1	Round 2	**Total**	Completed Sequencing	Mean Coverage	Bases @ 20X (%)	Targets covered (%)	Positive Reports	Total reports***
Partners/Broad	UW/KPW	1221	1279	**2500**	2500	416	99.70	99.90	92	2408#
	Geisinger	1251	1249	**2500**	2500	391	99.60	99.90	263	2237
	CCHMC	1494	1506	**3000**	3000	417	99.70	99.80	115§	2885§
	Harvard	1269	1231	**2500**	2500	401	99.70	99.80	76	2424
Baylor HGSC	Northwestern	3000	8	**3008**	3008	340	99.70	99.91	280	2997
	Mayo	3000	36^	**3036**	3036	344	99.75	99.95	129	3036
	CHOP	3000	0	**3000**	3000	342	99.90	99.98	102	3000
	Vanderbilt	2490	0	**2490**	2475*	336	99.90	99.98	225	2454
	Columbia	2581	91^	**2672**	2591**	338	99.89	99.97	167	2585
	Meharry	500	0	**500**	500	310	99.82	99.97	19	500
	Marshfield	1222	0	**1222**	875	319	99.77	99.95	51	854
	Total	21,028	5,400	**26,428**	25,985				1,519	25,380

**15 samples rejected*
***81 samples rejected*
^ extra samples

**** For Partners/Broad sites, includes total cases analyzed, since negative reports were not issued to all participants*
#UW/KPW: 1813 reports issued: 92, positive, 1519 negative, 202 inconclusive
§CCHMC: 251 reports issued: 97 positive, 18 carrier, 136 negative

eMERGE III Cohort

9219/25,068 cases with indications (37%)



- Colorectal cancer/Polyps*
- Hyperlipidemia*
- Kidney Disease
- Abnormality of pain sensation
- Pediatric Migraine
- Breast/Ovarian Cancer
- Arrhythmia
- Asthma
- At risk for prostate cancer
- at risk for colorectal cancer
- Autistic Behavior
- Ehlers-Danlos Syndrome
- Chronic Kidney Disease
- at risk for Breast/Ovarian Cancer
- Obesity
- Pulmonary Hypertension
- Cardiomyopathy
- Tubular Sclerosis Complex

* 611 patients had colorectal cancer and hyperlipidemia

Partners Broad Site Reporting Preferences

Actionable genes on consensus list	SOURCE	SITE 1: UW	SITE 2: GEISINGER	SITE 3: CCHMC adolescent cohort	SITE 3: CCHMC biobank cohort	SITE 4: HARVARD
BMPR1A	TOP6	X**		X	X	X
CACNA1A	TOP6	X		X		X
COL5A1	TOP6	X		X		X
HNF1A	TOP6	X	X	X		X
HNF1B	TOP6	X		X		X
KCNE1	TOP6	X	X	X		X
KCNJ2	TOP6	X	X	X	X	X
OTC	TOP6	X	X	X*		X
PALB2	TOP6	X		X		X
POLD1	TOP6	X**		X		X
POLE	TOP6	X**		X		X
SMAD4	TOP6	X**	X	X	X	X
ACTA2	ACMG56	X	X	X	X	X
ACTC1	ACMG56	X	X	X	X	X
APC	ACMG56	X**	X	X	X	X
APOB	ACMG56	X	X	X	X	X
BRCA1	ACMG56	X	X	X		X
BRCA2	ACMG56	X	X	X		X
CACNA1S	ACMG56	X	X	X	X	X
COL3A1	ACMG56	X	X	X	X	X
DSC2	ACMG56	X	X	X	X	X
DSG2	ACMG56	X	X	X	X	X
DSP	ACMG56	X	X	X	X	X
FBN1	ACMG56	X	X	X	X	X

GLA	ACMG56	X	X	X*	X	X
KCNH2	ACMG56	X	X	X	X	X
KCNQ1	ACMG56	X	X	X	X	X
LDLR	ACMG56	X	X	X	X	X
LMNA	ACMG56	X	X	X	X	X
MEN1	ACMG56	X	X	X	X	X
MLH1	ACMG56	X**	X	X		X
MSH2	ACMG56	X**	X	X		X
MSH6	ACMG56	X**	X	X		X
MUTYH#	ACMG56	X**	X	X	X	X
MYBPC3	ACMG56	X	X	X	X	X
MYH11	ACMG56	X	X	X	X	X
MYH7	ACMG56	X	X	X	X	X
MYL2	ACMG56	X	X	X	X	X
MYL3	ACMG56	X	X	X	X	X
MYLK	ACMG56	X	X	X	X	X
NF2	ACMG56	X	X	X	X	X
PCSK9	ACMG56	X	X	X	X	X
PKP2	ACMG56	X	X	X	X	X
PMS2	ACMG56	X**	X	X		X
PRKAG2	ACMG56	X	X	X	X	X
PTEN	ACMG56	X**	X	X	X	X
RB1	ACMG56	X	X	X	X	X
RET	ACMG56	X	X	X	X	X
RYR1	ACMG56	X	X	X	X	X
RYR2	ACMG56	X	X	X	X	X
SCN5A	ACMG56	X	X	X	X	X
SDHAF2	ACMG56	X	X	X	X	X
SDHB	ACMG56	X	X	X	X	X

SDHC	ACMG56	X	X	X	X	X
Additional Non-consensus genes	**SOURCE**	**SITE 1: UW**	**SITE 2: GEISINGER**	**SITE 3: CCHMC Adolescent**	**SITE 3: CCHMC Biobank**	**SITE 4: HARVARD**
SDHD	ACMG56	X	X	X	X	X
SMAD3	ACMG56	X	X	X	X	X
STK11	ACMG56	X**	X	X	X	X
TGFBR1	ACMG56	X	X	X	X	X
TGFBR2	ACMG56	X	X	X	X	X
TMEM43	ACMG56	X	X	X	X	X
TNNI3	ACMG56	X	X	X	X	X
TNNT2	ACMG56	X	X	X	X	X
TP53	ACMG56	X**	X	X	X	X
TPM1	ACMG56	X	X	X	X	X
TSC1	ACMG56	X	X	X	X	X
TSC2	ACMG56	X	X	X	X	X
VHL	ACMG56	X	X	X	X	X
WT1	ACMG56	X	X	X	X	X
CACNA1C	TOP6		X			
SERPINA1	TOP6			X*		
CHEK2	TOP6			X		
CFTR	TOP6			X*		
TYK2	TOP6			X*		
BMPR2	TOP6			X		
TCIRG1	TOP6			X*		

Only bi-allelic (homozygous, compound heterozygous) variants will be returned

* carrier status will be returned

** VUS's will be returned for 13 colorectal cancer genes in this site

Gene	rs #	Molecular consequence	SITE 1: UW	SITE 2: GEISINGER	SITE 3: CCHMC Adolescent	SITE 3: CCHMC biobank	SITE 4: HARVARD
ACADM	rs77931234	c.985A>C (p.Lys329Gln)	X		X*		X
ALDOB	rs387906225	c.360_363delCAAA (p.Asn120Lysfs)	X		X*		X
BCKDHB	rs386834233	c.832G>A (p.Gly278Ser)	X		X*		X
BCKDHB	rs79761867	c.548G>C (p.Arg183Pro)	X		X*		X
FAH	rs80338898	c.782C>T (p.Pro261Leu)	X		X*		X
G6PC	rs1801175	c.247C>T (p.Arg83Cys)	X		X*		X
CPT2	rs397509431	c.1239_1240delGA (p.Lys414Thrfs)	X		X*		X
BLM	rs113993962	c.2207_2212delATCTGAinsTAGATTC (p.Tyr736Leufs)	X		X*		X
MSH2	rs193922376	c.942+3A>T	X		X		X
CYP21A2	rs6467	c.293-13C>G	X		X*		X
F5	rs6025	c.1601G>A (p.Arg534Gln)	X		X		X
HFE	rs1800562	c.845G>A (p.Cys282Tyr)	X	X	X		X
MEFV	rs28940579	c.2177T>C (p.Val726Ala)	X		X*		X
MEFV	rs61752717	c.2080A>G (p.Met694Val)	X		X*		X

Only bi-allelic (homozygous, or if applicable compound heterozygous) variants will be returned unless otherwise noted, except MSH2 due to AD mode of inheritance

* carrier status will be returned

Baylor Site Reporting Preferences

Actionable genes on consensus list	SOURCE	SITE 1: Northwestern	SITE 2: Mayo	SITE 3: CHOP	SITE 4: Columbia	SITE 5: Vanderbilt
BMPR1A	TOP6	X	X	X	X	X
CACNA1A	TOP6	X	X	X	X	X
COL5A1	TOP6	X	X	X	X	X
HNF1A	TOP6	X	X	X	X	X
HNF1B	TOP6	X	X	X	X	X
KCNE1	TOP6	X	X	X	X	X
KCNJ2	TOP6	X	X	X	X	X
OTC	TOP6	X	X	X	X	X
PALB2	TOP6	X	X	X	X	X
POLD1	TOP6	X	X	X	X	X
POLE	TOP6	X	X	X	X	X
SMAD4	TOP6	X	X	X	X	X
ACTA2	ACMG56	X	X	X	X	X
ACTC1	ACMG56	X	X	X	X	X
APC	ACMG56	X	X	X	X	X
APOB	ACMG56	X	X	X	X	X
BRCA1	ACMG56	X	X	X	X	X
BRCA2	ACMG56	X	X	X	X	X
CACNA1S	ACMG56	X	X	X	X	X
COL3A1	ACMG56	X	X	X	X	X
DSC2	ACMG56	X	X	X	X	X
DSG2	ACMG56	X	X	X	X	X
DSP	ACMG56	X	X	X	X	X
FBN1	ACMG56	X	X	X	X	X
GLA	ACMG56	X	X	X	X	X
KCNH2	ACMG56	X	X	X	X	X
KCNQ1	ACMG56	X	X	X	X	X
LDLR	ACMG56	X	X	X	X	X
LMNA	ACMG56	X	X	X	X	X
MEN1	ACMG56	X	X	X	X	X
MLH1	ACMG56	X	X	X	X	X
MSH2	ACMG56	X	X	X	X	X
MSH6	ACMG56	X	X	X	X	X
MUTYH*	ACMG56	X	X	X	X	X
MYBPC3	ACMG56	X	X	X	X	X

MYH11	ACMG56	X	X	X	X	X
MYH7	ACMG56	X	X	X	X	X
MYL2	ACMG56	X	X	X	X	X
MYL3	ACMG56	X	X	X	X	X
MYLK	ACMG56	X	X	X	X	X
NF2	ACMG56	X	X	X	X	X
PCSK9	ACMG56	X	X	X	X	X
PKP2	ACMG56	X	X	X	X	X
PMS2	ACMG56	X	X	X	X	X
PRKAG2	ACMG56	X	X	X	X	X
PTEN	ACMG56	X	X	X	X	X
RB1	ACMG56	X	X	X	X	X
RET	ACMG56	X	X	X	X	X
RYR1	ACMG56	X	X	X	X	X
RYR2	ACMG56	X	X	X	X	X
SCN5A	ACMG56	X	X	X	X	X
SDHAF2	ACMG56	X	X	X	X	X
SDHB	ACMG56	X	X	X	X	X
SDHC	ACMG56	X	X	X	X	X
SDHD	ACMG56	X	X	X	X	X
SMAD3	ACMG56	X	X	X	X	X
STK11	ACMG56	X	X	X	X	X
TGFBR1	ACMG56	X	X	X	X	X
TGFBR2	ACMG56	X	X	X	X	X
TMEM43	ACMG56	X	X	X	X	X
TNNI3	ACMG56	X	X	X	X	X
TNNT2	ACMG56	X	X	X	X	X
TP53	ACMG56	X	X	X	X	X
TPM1	ACMG56	X	X	X	X	X
TSC1	ACMG56	X	X	X	X	X
TSC2	ACMG56	X	X	X	X	X
VHL	ACMG56	X	X	X	X	X
WT1	ACMG56	X	X	X	X	X
Additional non-consensus genes	**SOURCE**	**SITE 1: Northwestern**	**SITE 2: Mayo**	**SITE 3: CHOP**	**SITE 4: Columbia**	**SITE 5: Vanderbilt**
ANK2	TOP6	X				X
ATM	TOP6	X				X
ATP1A2	TOP6	X				
BMPR2	TOP6	X				X
CACNA1C	TOP6	X				X

CFH	TOP6	X				
CFTR	TOP6	X				
CHEK2	TOP6	X				X
FLG	TOP6	X				
MC4R	TOP6	X				
*MTHFR**	TOP6	X				
*NTRK1**	TOP6	X				
SCN1A	TOP6	X				
SCN9A	TOP6	X				
*SERPINA1 **	TOP6	X				
*SLC2A10 **	TOP6	X				
TCF4	TOP6	X				
*TCIRG1**	TOP6	X				
TTR	TOP6	X				X
*TYK2**	TOP6	X				
UMOD	TOP6	X				
*VDR**	TOP6	X				

* Only biallelic or homozygous variants will be returned for disorders with AR inheritance

Gene	rs #	Molecular consequence	SITE 1: Northwestern	SITE 2: Mayo	SITE 3: CHOP	SITE 4: Columbia	SITE 5: Vanderbilt
ACADM *	rs77931234	c.985A>C (p.Lys329Gln)	X	X	X	X	X
ALDOB*	rs387906225	c.360_363delCAAA (p.Asn120Lysfs)	X	X	X	X	X
BCKDHB*	rs386834233	c.832G>A (p.Gly278Ser)	X	X	X	X	X
BCKDHB*	rs79761867	c.548G>C (p.Arg183Pro)	X	X	X	X	X
FAH*	rs80338898	c.782C>T (p.Pro261Leu)	X	X	X	X	X
G6PC*	rs1801175	c.247C>T (p.Arg83Cys)	X	X	X	X	X
CPT2*	rs397509431	c.1239_1240delGA (p.Lys414Thrfs)	X	X	X	X	X
BLM*	rs113993962	c.2207_2212delATCTGAinsTAGATTC (p.Tyr736Leufs)	X	X	X	X	X
MSH2	rs193922376	c.942+3A>T	X	X	X	X	X
CYP21A2*	rs6467	c.293-13C>G	X	X	X	X	X

F5*	rs6025	c.1601G>A (p.Arg534Gln)	X	X	X	X	X
HFE*	rs1800562	c.845G>A (p.Cys282Tyr)	X	X	X	X	X
MEFV*	rs28940579	c.2177T>C (p.Val726Ala)	X	X	X	X	X
MEFV*	rs61752717	c.2080A>G (p.Met694Val)	X	X	X	X	X

* Only biallelic or homozygous variants will be returned for disorders with AR inheritance

Additional non-consensus SNPs	rs #	Molecular consequence	SITE 1: Northwestern	SITE 2: Mayo	SITE 3: CHOP	SITE 4: Columbia	SITE 5: Vanderbilt
ABCC8*	rs151344623	c.3989-9G>A (p.?)	X				
ATP7B*	rs76151636	c.3207C>A (p.His1069Gln)	X				
CLRN1*	rs111033258	c.144T>G (p.Asn48Lys)	X				
COL5A2	rs786205104	c.2031+1G>T (p.?)	X				
COL5A2	rs786205103	c.1924-2_1928del (p.?)	X				
DHDDS*	rs147394623	c.124A>G (p.Lys42Glu)	X				
DLD*	rs121964990	c.685G>T (p.Gly229Cys)	X				
F11*	rs121965064	c.901T>C (p.Phe301Leu)	X				
F11*	rs121965063	c.403G>T (p.Glu135*)	X				
FANCC*	rs104886456	c.456+4A>T (p.?)	X				
HPS3*	rs201227603	c.1163+1G>A (p.?)	X				
KCNE2	rs74315447	c.161T>C (p.Met54Thr)	X				
SPG7*	rs61755320	c.1529C>T (p.Ala510Val)	X				
TOR1A	rs724159981	c.907_909del (p.Glu303del)	X				

JAK2	rs77375493	c.1849G>T (p.Val617Phe)					X
KCNE1**	rs1805128	c.253G>A (p.Asp85Asn)					X

* Only biallelic or homozygous variants will be returned for disorders with AR inheritance

** KCNE1 on the gene list and considered a risk allele

Additional non-consensus SNPs returned for all Mayo patients	rs #	Molecular consequence	SITE 1: Northwestern	SITE 2: Mayo	SITE 3: CHOP	SITE 4: Columbia	SITE 5: Vanderbilt
PCSK9	rs2479409	N/A		X			
CELRS2	rs629301	N/A		X			
APOB	rs1367117	N/A		X			
ABCG8	rs4299376	N/A		X			
SLC22A1	rs1564348	N/A		X			
HFE	rs1800562	N/A		X			
MYLIP	rs3757354	N/A		X			
ST3GAL4	rs11220462	N/A		X			
NYNRIN	rs8017377	N/A		X			
LDLR	rs6511720	N/A		X			
APOE	rs429358	N/A		X			
APOE	rs7412	N/A		X			

no clinical interpretation will be provided for these variants.

eMERGE Network: ESP Executive Session Minutes

Monday, April 29th, 2019 at 2:00 p.m. EST (1:00 PM CST; 11:00 a.m. PST)

ESP: Howard McLeod, Kim Doheny, Lisa Parker, Stan Huff, Eta Berner, Vandana Shashi; **Baylor:** Richard Gibbs, Mullai Murugan; **BCH:** Ingrid Holm; **CCHMC**: John Harley, Bahram Namjou, Cindy Prows; **CHOP:** John Connolly, Hakon Hakonarson, Patrick Sleiman; **Columbia:** Chunhua Weng, Aileen Espinal, Stephanie Tang, Alex Fedotov, George Hripcsak; **Geisinger:** Marc Williams, Nephi Walton **Harvard:** Beth Karlson, Scott Weiss; **John Hopkins University**: Casey Overby Taylor, **KPW/UW:** Gail Jarvik, David Crosslin (CC), Eric Larson; **Marshfield:** Murray Brilliant, Scott Hebbring, Aniwaa Obeng**; Meharry:** Rajbir Singh, Samuel Adunyah; **Mayo:** Iftikhar Kullo, Richard Sharp; **Mt. Sinai:** Aniwaa Obeng **Northwestern:** Maureen Smith; **Partners/Broad:** Hana Zouk, Sandy Aronson; **VUMC:** Sarah Bland, Josh Denny, Wei-Qi Wei; **NHGRI:** Jyoti Dayal Gupta; Ken Wiley, Sheethal Jose, Robb Rowley, Teri Manolio; **CC:** Josh Peterson, Melissa Basford, Jodell Jackson, Kayla Howell, Laura Allison Woods, Michelle Stone, Brittany City

Absent: Rex Chisholm, Funmi Olopade

NOTES:

- **Welcome, Opening Remarks, General Updates |** *Robb Rowley & Howard McLeod*
  - The NHGRI would like to thank everyone for joining the ESP conference call including the ESP members, and the CC for organizing the packet prior to the call. The packet of preparatory materials provided to the ESP before the call was helpful to the ESP for updates on the Network.
- **Network Introductions |** Josh Peterson *(slides here)*
  - eMERGE has sequenced 25,380 eMERGEseq participants. Non indication-based results have a ~4% positive rate and indication-based testing has a positive rate of ~2.2%.
  - Four sites have completed ROR, and several other sites are completing their RoR
  - The eMERGEseq Freeze V1 is to be released publicly on dbGaP by the end of this week.
  - Five Lessons Learned panels have been convened at steering committee meetings; three panels are scheduled between June 2019 and January 2020. There are also 11 lessons learned manuscripts that have been either published (four) or in development (seven).
  - There have been over 1280 external downloads from eMERGE dbGaP submissions, and 755 eMERGE Network and site-specific projects have been published as of March 2019.
  - Regarding how eMERGE can help inform the All of Us program, the VUMC site returns all results (negative, PGx, P/LP) directly to providers first through the EHR, then disclosing results to participants directly in order to simulate real world situations where genetic counselors may not be available.
  - Six eMERGE sites are returning negative results. Baseline and follow-up surveys are being administered to elicit participant understanding. Mayo & Northwestern are conducting interviews on a subset of participants to further investigate the impact of negative results.
  - The ESP made a recommendation to strive for consistency across outcomes collection forms, and the eMERGE Outcomes Workgroup has worked to centralize and harmonize outcomes forms. Abstraction guides are being developed on all forms;the following have been finalized:
    - Aortopathy (Mayo)
    - Breast Cancer - Women (Columbia)
    - Cardiomyopathy (Northwestern)
    - Colorectal Cancer and Polyposis (KPW/UW)
    - Familial Hypercholesterolemia-Adult (Mayo)
    - Familial Hypercholesterolemia-Pediatric (Geisinger)
    - OTC (Geisinger)
    - Tuberous Sclerosis (Geisinger)
    - 22Q (CHOP)

- Sites are actively working to augment the work of the Network. Three sites have obtained independent funding sources for related work and KPW is preparing a grant using existing eMERGE tools to allow eMERGE data to be accessed by the public and other researchers.
    - VUMC leads a 'MeTree' supplement to develop an implementation strategy for the use of Family Health History (FHH) collection tool into the EHRs of diverse institutions and worked with Duke University, Geisinger & Northwestern.
    - BCH obtained a RO1 in 2018 to deploy a Healthcare Provider Survey, and disseminated surveys to nine eMERGE sites in order to capture healthcare provider's perceived utility of genetic tests and results.
    - Marshfield used internal resources to initiate the sequencing of approximately 2,000 recruited individuals on the eMERGEseq platform evaluating genetic risk factors for de-novo/predicted family histories for cancer or Familial Hypercholesterolemia (FH) and incorporate genetic results? into the EHR for standard of care.

- **Return of Results: Empirical Data & lessons learned |** Iftikhar Kullo & Ingrid Holm *(slides here)*
    - The ROR Workgroup has several ongoing projects, including NT277, NT300, NT323, NT330, and NT332, as well as proposals that have recently been submitted, including NT322, and NT273.
    - The Workgroup has recently published manuscripts in *Genetics in Medicine*, *Journal of Personalized Medicine*, and *Mayo Clinic Proceedings*.
    - The sites differ greatly between the methods of return of result, and the group is utilizing these differences in order to determine the impact of these various methods.
    - Many sites are very close to completing the return of positive results, and two-thirds of the sites are returning negative results. Of the sites returning negative results, many are almost complete. All sites are expected to have this completed by the end of the summer.
    - The Health Care Provider Survey (HCP) is only sent to providers that received positive results. The group plans to send a second survey to providers with negative results.
    - The Network is contributing to best practices for ROR through several methods. Across the eMERGE sites are a variety of 'who', 'what', and 'how' on cohort recruitment and methods manuscript concept sheets regarding return of results. Cohorts were selected differently: some were selected based on phenotype, adolescent versus pediatric cohorts, and participant choice versus no choice for return of secondary findings. Different results were returned across sites. For example, P/LP were returned at most sites, KPW/UW is the only site returning VUS, and Marshfield and VUMC are the only sites returning PGx.
    - ROR projects are hybrids between research and clinical work. They can be used for determining clinical utility, familial and societal impact, and how these results are perceived by patients and providers.
    - Ongoing ROR studies in eMERGE can help address the knowledge gap in understanding the penetrance of the ACMG59. Clarification may be needed from the ACMG to the larger scientific population.
    - The ESP recommended that sites develop a plan to optimize the response rate of patient surveys through EHR reminders, medical appointments, and other methods, in order to improve the quality of results.
        - The participant survey subgroups' Participant Survey Tracker allows for the ROR group to monitor progress of participant surveys. Sites also utilize multiple email reminders, follow-up calls, and in-person reminders, as well as paper mail. REDCap is used for automatic reminders.
    - The ESP recommended that the ROR and Outcomes workgroups focus on understanding the misconceptions surrounding ROR on patient care, and identify sustainable solutions or strategies that can address these misconceptions.
        - Data from the ROR and Outcomes group were collected from semi-structured interviews, participant surveys, and analyzed specifically in MCS NT300 and NT322. These data can be analyzed during the remaining year of eMERGE to guide research on common misconceptions surrounding ROR on patient care.
    - The ESP recommended that the ROR workgroup study physicians' responses to variant reclassifications.

        - The CSGs are working to quantify the cost and psychosocial implications of ROR and responses to variant reclassifications, as well as the mechanisms of disclosing new interpretations and how participants and providers respond to changes in variant interpretation. This is collected via interviews of health care providers who received notification of a reclassified variant to explore their response.
- *Questions from ESPs:*
    - How has the diversity between the sites ROR process actually affected the study of ROR?
        - The lack of continuity between sites approach to RoR creates hurdles and reduces overall generalizability, however this allows the Network to study many different approaches to ROR (adults/children, PGX/ACMG, primary/secondary)..
    - Are you pursuing more qualitative interviews with providers and were there difficulties recruiting providers?
        - The RO1 HCP survey grant includes interviewing providers that have received results that are positive and negative. Recruitment is based done off existing participant referral. The second wave may be easier because physicians have received results. There is a 25-30% response rate for interviews, which is an expected range. It is unknown at this time if there is systematic bias in the non-responders.
- **EHRI: Integration of the CDS into the EHR |** Sandy Aronson & Casey Overby Taylor *(slides here)*
    - The EHRI Workgroup has several ongoing projects, including: NT301, NT272, NT328, and NT213, as well as ongoing projects in collaboration with other workgroups, including: NT265 (EHRI/Phenotyping/Genomics), NT310 (EHRI/Phenotyping/Genomics), NT289 (EHRI/Genomics), NT236 (EHRI/Genomics), NT237 (EHRI/Genomics), NT294 (EHRI/Phenotyping), NT295 (EHRI/Phenotyping), NT309 (EHRI/Phenotyping), and NT270 (EHRI/ROR).
    - eMERGE is the first and only Network that can transmit structured data lab results to a variety of providers and sites. This is accomplished with a custom XML format. However, reporting standards are needed, especially as most clinical sites (9/10) use non-structured data format. GI XML served this purpose, but the investigators wanted a standard that could grow beyond eMERGE to be a national standard.
    - Fast Healthcare Interoperability Resources (FHIR) is the newest (next) generation standards framework created by HL7. FHIR aims to simplify implementation without sacrificing information integrity. It leverages existing logical and theoretical models to provide a consistent, easy to implement, and rigorous mechanism for exchanging data between healthcare applications (see here).
    - FHIR standards could improve computation, as well foster the integration and interoperability of genomic testing data. FHIR combines features of HL7s v2, HL7 v3, and CDA product lines while leveraging the latest web standards and applying a focus on implementability.
    - Mullai Murugan (Baylor) spoke to the creation of a national-based HL7 FHIR specification for utilization, and Nephi Walton (Geisinger) has worked on sending results directly into the EHR without the ancillary genetic system.
    -
    - A pilot subgroup could be formed, but currently, the issues are being logged using HL7 FHIR's ZULIP process. Targeted meetings are setup to resolve the issues. The investigators have met with the CSGs to demonstrate the current and planned work and to gain recommendations on how to have efficient collaboration.
    - There are a few challenges, including that the clinical genomics workgroup.timeline is codependent with HL7 It can take time to develop resolutions, even though there is a need for quick decisions on critical path issues/changes. To ensure forward progress the group has to establish its own standards. As the investigators continue through development and mapping, they track questions, issues, and discrepancies that can be adjudicated at a later time with the FHIR workgroup.
    - Geisinger uses the genomic indicator function in EPIC to create a genetic phenotypes. Genomic indicators serve as the disease or metabolizer status providing a a point of reference for CDS Genomic

indicators are used to establish clinical decision support, health maintenance schedules, and development of care paths.
    - Investigators built and integrated genomic indicators for CDC tier one conditions,several PGx variants, and integrated the decision support into the test environment.
    - The vision is for labs to use FHIR resources to integrate genomic information into the EHR and use VAR format to drive genomic indicator for infobuttons and patient's CDS. Investigators recognize that there is still significant work to be done to integrate the genomic data into the EHR.
    - Lessons learned includes that there continues to be need work with definitions and standards surrounding genomic indicators and phenotypes. It remains unknown if genetic indicators can be passed from the laboratory to the EHR. The investigators recognize the importance to include the laboratory in the process of establishing standards. The investigators plan to develop information resources on genetic conditions for patient/physician information.
    - In addition to other EHRI work, the EHRI Workgroup is writing a manuscript that summarizes the October 2018 Steering Committee meeting lessons learned panel.
- *Questions from the ESP?*
    - How is the current work shared?
        - The current specifications for the FHIR standards are still in draft mode, so is not quite ready for dissemination. However they continue to work with the FHIR workgroup to help inform and refine the national standards for genetic testing results.
        - For AoU, the plan is to create a FHIR standard; the lessons learned from the the GC group will be a great resource.
        - Genomic Indicators are still in a developmental state with labs and EMR vendors not currently adopting the standard. It will be critical to demonstrate to vendors that this is useful to facilitate the adoption of the FHIR standard. The group is currently working with EPIC. If EPIC adopts the standards it may help shift the utilization.
        - Investigators also upload relevant code and applications it to GitHub, which is a publicly used site to help disseminate the tools needed to implement the FHIR standard.
    - eMERGE is uniquely positioned to research the penetrance of P/LP in ACMG 59. Has this been considered?
        - The Clinical Annotation Workgroup is going to examine this, and the Outcomes forms are assessing penetrance at baseline and six-months post return.
        - For the cohorts selected for a particular trait it would be excluded, however the other variants will be examined to determine if relevant traits are present.
- **Discussion and suggestions from the ESP**
    - The ESP appreciated the significant progress the EHRI workgroup has made towards establishing a FHIR standard for sharing genetic testing results., there is a lot of commonality in the challenges surrounding FHIR standards.
    - The ESP is impressed with the significant amount of hard work and they encouraged the Network to continue the effort.

- **Executive Session**
    - The External Scientific Panel (ESP) met with members of the NHGRI staff in Executive Session after the ESP teleconference held on April 29, 2019. The ESP members appreciated the Network's comprehensive responses to their comments and suggestions from the October 2018 ESP in-person meeting. They commended the extraordinary effort of coordinating such a large network. The workgroups have made a significant effort addressing the ESP's past meeting recommendations.
    - The ESP was impressed with the Return of Results (RoR)/Ethical Legal Social Implications (ELSI) workgroup's various studies that are currently in progress that address the practical ELSI issues related to clinical genomics. The ESP noted that the Network is on course to collect 6-month outcome data from

returning the ACMG variants. However, they pointed out that the twelve month and longer follow-up is critical to maximize the study of penetrance. The ESP recommended that the Network review the current RoR schedule to identify any delays that will impact the collection of 12-month outcome data. The ESP recognized the Network's success in obtaining additional funding but recommended considering applying for additional grants or supplements to capture outcome data beyond the twelve months. The eMERGE expertise applied to this effort will significantly help the field of genomic medicine.

- The ESP recognized that the 25-30% response rate to the healthcare provider (HCP) survey is a good response. They urged the Network to consider ways to obtain additional responses, especially to the pathogenic/likely pathogenic variant results. An approach that goes beyond e-mails and letters offers an opportunity to improve response rates and to increase awareness among providers of the eMERGE Network. Other programs have found that conducting town halls or educational sessions not only improves the number of responses but also increases patient referrals and provider engagement with the program. An example the ESP highlighted was an approach taken by the Undiagnosed Disease Network (UDN) to engage providers by speaking at departmental meetings. This increased referrals and survey response rates for the UDN. A similar approach could be considered for the eMERGE Network.
- The ESP was impressed with the Electronic Health Records Integration (EHRI) workgroup's efforts and acknowledged that the Network's efforts extend beyond eMERGE. The Network has made significant progress with establishing standards for incorporating XML-based genetic testing results and clinical decision support systems (CDSS) into the EMR. This includes being a critical driver for establishing FHIR standards in genomic medicine. However, the extent of deployment of these standards among the Network is not clear. What are the lessons learned with deploying these in the EMR? How many sites have already installed genomic based CDSS? The ESP noted that the Network has a manuscript concept sheet (NT272) listed that addresses the impact of CDSS in clinical care and encouraged this effort to continue. The ESP appreciated that these tools and techniques are not unique to genomics, but the Network should continue to be the leader in helping healthcare systems integrate genomic information by researching standards and sharing their lessons learned. As such, the Network should continue and expand its work with other consortia and commercial entities that are focused on developing CDSS and use existing frameworks, instead of creating a different infrastructure specifically for genetic data.
- Lastly, the ESP recognized that the Network finishes this phase of eMERGE in less than a year. They emphasized the importance of prioritizing the manuscript concept sheets of the lessons learned and network-wide efforts to ensure the outstanding work by eMERGE is shared among the greater scientific and healthcare community.

- **ESP Recommendations**

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.

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

3) 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.

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.

5) The Network should prioritize the publication of the lessons learned manuscripts before the end of the current phase of eMERGE.

Summary of Steering Committee Meeting: Summer 2019

June 20th-21st, 2019 in Seattle, WA

TABLE OF CONTENTS

eMERGE DAY 1: THURSDAY

- **NHGRI program official report |** Robb Rowley *(NIH/NHGRI)* | slides
    - NHGRI is developing a strategic plan for the future, and looking for input. There is a new eMERGE RFA; questions can be directed to emergerfa@mail.nih.gov
    - Several funding opportunities are available. The NHGRI is looking for great proposals from clinical genomics applications.
        - The Electronic Medical Records and Genomics (eMERGE) Genomic Risk Assessment and Management Network- Coordinating Center (U01 Clinical Trial Required)
            - Application Due Date: August 2, 2019. Expiration Date: August 3, 2019
        - Investigator-Initiated Genomic Medicine Research (R01 Clinical Trial Optional)
            - Application Due Date(s): October 21st, 2019; June 19, 2020; October 20, 2020 Expiration Date: January 23, 2021
        - Limited Competition Centers of Excellence in Ethical, Legal and Social Implications (ELSI) Research (CEER) (RM1 Clinical Trial Optional)

        - Application Due Date(s): July 23, 2019 Expiration Date: July 24, 2019
    - Genomic Innovator Award (R35 Clinical Trial Optional)
        - Application Due Dates: October 30, 2019; October 30, 2020; Expiration Date: October 31, 2020
        - eMERGE investigators are encouraged to submit to the innovation award.
    - Novel Approaches for Relating Genetic Variation to Function and Disease (R21 Clinical Trial Not Allowed)
        - Application Due Date(s): July 16, 2019; November 16, 2019; July 16, 2020; November 16, 2020; July 16, 2021; Expiration Date: September 8, 2021
    - Ethical, Legal, and Social Implications (ELSI) of Genomic Research Regular Research Program (R01)
        - Application Receipt Date(s): Standard dates apply. Expiration date: September 8, 2020
    - Genomic Community Resources (U24)
        - Application Due Date(s): First due date is July 13, 2017; Expiration Date: January 20, 2020
  - Specific research questions can become notices of special interests (NOSIs). NOSIs are ways to increase efficiency in expressing interest in a scientific area while reducing the administrative burden of traditional program announcements.
  - The eMERGEseq Freeze1 (Phs001616) was submitted to dbGaP on 5/1/2019.
  - There are several datasets committed to deposit into dbGaP before 2020, including the GWAS dataset (N= 83, 717) with the updated case/control file, the GWAS dataset with additional Harvard and CCHMC samples (N= 105,000) and the eMERGEseq Freeze V2 (N=25,000).
  - The Network should pursue how to explore discrepancies in ROR between sites and how to interpret non-responder data to help understand the impact that this will have on outcomes data
  - The Network should prioritize publishing lessons learned papers as Year 5 funding continues, specifically focusing on Network-wide lessons learned.
  - Establishing XML standards for genetic testing results will help implement genomics into medical care. The eMERGE Network has been instrumental in helping develop FHIR standards that will allow the broader adoption of structured genetic test data into the EMR.
- **Announcements, opening remarks |** Rex Chisholm *(SC Chair, Northwestern)* | slides
  - AnVIL was confirmed as an affiliate member for pilot testing of eMERGE data.
  - eMERGE will be one of the first Networks to share data using AnVIL and will help shape the AnVIL data repository for discovery moving forward.
  - The final stages of ROR is in progress and four sites have already returned their results.
  - Four sites have completed ROR. There have been 19 Outcomes forms harmonized and deployed for data collection and abstraction guides. There is a significant amount of Outcomes data available over 640 forms.
  - EHRI group is working with HL7 clinical genomics workgroup to map eMERGE scheme to FHIR standards.
  - The ESP would like for the Network to focus on lessons learned. The Network will address penetrance moving forward and specifically during the October 2019 lessons learned panel.
  - The Network should review in development projects and prioritize which projects are feasible to complete within this phase.
  - The Network has a good history of publications but there are concerns about the number of manuscripts in development. The Network should prioritize publishing existing manuscripts, and lessons learned papers.
  - There have been 1350 external downloads of dbGaP as of May 2019.
  - Goals of this meeting include: genomics lessons learned panel, update on AnVIL plans for eMERGE dataset utilization, demonstrate progress of the seven network wide milestones and lessons learned manuscripts, and discussion of any major barriers for work completion in Year 5, including return of results, loss of follow up of participants, and impact of penetrance studies.

- **Genomic data update |** David Crosslin & Ian Stanaway *(UW/CC)* | slides
    - The GWAS V3 SNV imputation is almost complete.. Network investigators can find an updated report here.
    - The structural variation imputation set QC metrics have been looked at to release. There is about a 90% concordance with CNV calling, and about 60% of them compute with high concordancy.
    - There are multiple MCS's for discovery analyses planned with the 105,108 eMERGE array samples.
    - NT314: Ian will be imputing structural variations from the GWAS array samples, and he is looking for potential collaborations.
    - The eMERGEseq V2 sample (N = 24,956) has been compiled. Ian is running PCA analyses.
- **CSG Updates |** Eric Venner (BCM/HGSC) & Hana Zouk *(Partners/Broad)* | slides
    - A new CNV method that improves the sensitivity and resolution of CNV detection has been recently published in the paper "Atlas-CNV: a validated approach to call single-exon CNVs in the eMERGESeq gene panel" (Chiang et al, 2019).
    - Second set of reports are coming from ClinVar, that are 'mashed' up clinical variations. In the past month, BCM has identified 158 variants with a new pathogenic classification. There are eight remaining.
    - The ReVU (reanalysis of variants and updater) tool was developed to help with ClinVar reanalysis.
    - 148 remain VUS, 10 passed on for second review (eight remaining), second review ongoing.
    - To manage updated reports, ARBoR (Authenticated Resources in Hashede Block Registry) was recently released. All reports have a barcode; barcodes are scanned and verified to check validity using ARBoR Client Python API and ARBoRScan iOS and Android app. Report histories are recorded in encrypted ledger. A paper titled "ARBoR: an identity and security solution for clinical reporting" was recently released. (Venner et al, 2019).
    - All of the VUS that are likely pathogenic are pulled out of the sample (n= 79). 50% are cardiomyopathy or arrhythmia genes.
    - There were sex discrepancies found in SNP data (n= 73) and metadata that was reported in January and most of the samples were removed.
    - As part of the QC for the full eMERGE dataset, the CC reviewed heterozygosity in X and Y variants and matching to the sample metadata (submitted sex). Initially over 2400 samples were flagged due to a CHEK2 duplication but when issues were resolved there were only 222 still discrepant.
    - FHIR implementation is led by Larry Babb and Mullai Murugan. The EHRI WG is mapping the eMERGE reports to the FHIR standards to create a pilot implementation at a clinical site.
    - There is a discrepancy resolution group within ClinGen that prioritizes medical data. It is the intent to prioritize anything eMERGE-relevant in the group.
    - Alerts have been in reporting system, and 11 out of 565 reported variants that have issued reports that have affected 23 reports.
    - BCM and Broad have not compared the specific criteria that they are using to reclassify the VUS. Comparison of the specific criteria used for each lab is hard to determine. In MRA7 there are very specific rules. For others as long as we use the same classification is fine but in others this may not be the case.
    - Participant and provider responses to reclassified variants is currently not being captured, but there is an opportunity to do this on the 12-month outcomes form. This could be discussed in the Outcomes workgroup.
        - This was not built into the ROR protocol and how this will be done is important.
- **Update on an investigation of somatic mutations in eMERGE datasets |** Ken Kaufman *(CCHMC)* | slides
    - Ken gave update on ongoing somatic mutation project at CCHMC. Particularly relevant given the update with P/B showing that our understanding of somatic mutation can affect health (Yizhah et al, 2019).
    - Unlike germline mutations, somatic mutations are spontaneous and can only be detected in a small subset of cells.
    - Throughout the data, there are places that do not match. This can be sequencing artification or somatic mutations.

- CCHMC processed two large datasets, including: eMERGE elll set A (16,170 samples) and PGx (10,000 samples screened) and found 6% somatic mutation out of both datasets. 8% of reads have a polymorphism.
- Cardiomyopathy genes appear on the list of most somatic mutations.
- Findings demonstrated somatic variation in 178 genes, with 61 severe polymorphisms. Some polymorphisms appear to be germline. 225 appear in the VCF file, however only 75 correspond to the sample.
- One of the underlying goals of the project is to identify how damaging the variant will be. Findings showed that cells with somatic mutation *are* more likely to have a biological impact. Looking at functional predictions of the variants (how damaging) could be a way to prioritize the variants.
- Alternate allele frequency is seen about 1-3% of the time indicating that they would be present in up to 6% of the cells. Some are present in up to 30% of the cells. A lot of alternate alleles have variants in very low frequencies.
- Investigators obtained DNA from VUMV and NU, and tested 11 samples for validation using digital droplet PCR in order to validate results and were able to validate nine out of the 11. The samples that had digital droplet PCR had results close to that of the sequencing data.
- It is important to collect as many samples that have been sequenced in order to validate. These include samples from PGx data and the eMERGE III eMERGEseq freeze data.
- The group has been working with iGENOMX Riptide to sequence quickly, they use primers that look specifically at SNPs that investigators are interested in.
- The problem with a lot of samples coming in for validation is that the cost becomes prohibitive. Using this technology they will be able to look at 96-960 candidates for about $15-$20.
- Currently in the process of obtaining samples for validation and finalizing the validation strategy.
- ACTION ITEM: Investigators with questions or samples for validation regarding the somatic mutation project, NT171: *Possible somatic mutation in targeted sequence data,* should contact Ken Kaufman (kenneth.kaufman@cchmc.org) and Paul Gecaine (paul.gecaine@cchmc.org).
- With regards to which samples or criterion was used to re-test, everything has been data driven. Primary goal is to develop a pipeline that identifies a cost-effective method for identifying somatic variation from NGS; criteria includes many samples with a range of variants, including a range of alternate alleles to determine limitations to identify calling mechanisms for different variants.
- The group discussed implications for identifying different somatic mutations in different alleles
    - All three top mutations are in conditions including cardiomyopathy, which was also discussed earlier today. Potential for looking at differences in age, or other biological impacts.

- **AnVIL Progress & Discussion** | James Taylor *(JHU)* | slides
    - The goal for AnVIL is to create a scalable and computable cloud-based resource that supports data access and storing of large Network datasets driven by NHGRI programs. AnVIL will promote greater cost-effectiveness, and be more secure and compliant than prior data repositories.
    - AnVIl will be collaborative for datasets and analysis workflows and a hub for data access.
    - Terra, which houses AnVIL, facilitates analyses for workflows and mapping. WDL, CDL, Jupyter Notebook, Python, and R will all be supported.
    - Dockstore is a hosting service for different tools and workflows that can hold the work of Terra.
    - Galaxy, a web-based analysis environment for running analysis tools and building workflows for users with little to no programming or bioinformatics experience. Galaxy Toolshed has over 6,000 tools.
    - AnVIL needs a complete system to query data. Gen3 is currently filling this need, as it creates data models, indexes and queries.
    - AnVIL is working on creating a complete bioconductor environment for storing data.
    - AnVIL's security configurations and controls for data management are equivalent to those of a Federal Information Security Management Act (FISMA) Moderate System environment and built on top of the Google Cloud platform.

- AnVIL will roll out data models incrementally; a soft launch is expected at the end of June. The eMERGE data is not going to be included in the initial June rollout.
- In 2019/2020, AnVIL will add additional analysis environments for researchers to expand and use.
- New 1000 Genomes High Coverage Data platform is open access; investigators can look directly at the data objects that are located in AnVILG. The data objects can be mapped into workflows to be used for notebooks.
- Other initial datasets include GTEx Versions 7& 8 and selected datasets from the Centers for Common Disease Genomics (CCDG) and Centers for Mendelian Genomics.
- eMERGE's adjacent datasets are different than the current GWAS data stored in AnVIL which will require a bit more work to prepare.
- eMERGE will be the driver in building models due to facets such as structured phenotype models. Data can be analyzed in the platform using a wide variety of tools supported by AnVIL.
- The hope is as additional projects, tools, and datasets are brought in, this can be coupled with training materials for those who will be coming in to use this data. By providing this platform, they can train investigators using the best next methods on data repositories.
- If you have approval in dbGap to use a project you will be able to use that project in the AnVIL.
- Broad has a system that the Network is piloting as part of the launch, called the Data Use Oversight System (DUOS).. This will check if data access requests are compatible with the data use restrictions to automate use of data.
- Workshops and tutorials are planned for upcoming scientific meetings. There is a workshop at ASHG onhow to use AnVIL.
- For some datasets, access will be free, but analysis will incur a charge.
- All data access and usage on AnVIL will be audited, via the GCP auditing system. Every person that accesses the AnVIL data layer is documented. The AnVIL team is aware of who accesses the data and users are required to go through the authorization process.
- There is an access control framework built into the platform through Google compute. Rights form dbGaP will transfer to AnVIL users. This will be accomplished with users using their eCommons ID.
- Regarding DNAcommons project migration to other cloud environments, investigators would need to build an extraction layer in order to push data into another cloud environment.
    - There are some google services AnVIL will take advantage of. The intent is to use big query for large multi-sample VCF calling.
- AnVIL storage needs to have identifiers from data model mapped into data file. Each graph has figures mapped to ontology terms. Multiple data models are possible, however, integrative approaches across projects require mapping. Single data models are more simplified. AnVIL would simplify data model to facilitate querying.
- AnVIL is funded by two NIH U24-grant currently, with five years of support. Storage for datasets that are made available to the public are currently paid for by the grant. Ingress fees (add data) are covered by Google. Egress fees (pull data down) are paid by investigators toGoogle, but hopefully the data will be computed on the platform and users only egress derived data.

- **Challenges in Implementing Genomic Medicine in a Federally Qualified Community Health Center: Insights from Mayo Clinic's RAVE-Phoenix Study** | Gabriel Shaibi & Valentina Hernandez *(Mayo)* | slides
    - Arizona Return of Actionable Variants Empirical (RAVE) study enrolled 500 Latino participants from SPS biobank with either hyperlipidemia or colon polyps. Participants consented to have their sample sequenced and the results, actionable and negative, to be returned to them...
    - Dr. Laney Lindor, program director spoke with participants explaining the implications of ROR.
    - Arizona's ROR process is detailed in the publication "Developing a process for returning medically actionable genomic variants to latino patients in a federally qualified health center" (Shaibi et al, 2018).
    - 486 letters were sent for negative/neutral results. There were 926 follow up phone calls, 57 participants reported they did not receive the initial letter and requested a second letter which was sent by certified mail.here were no requests for consultations with medical geneticists after receiving result.

- There were 50 individuals who were brought back in with neutral or negative findings about four months after the site received theirresults and were interviewed.
- Participant motivations for becoming involved with the study included personal, family, and resource benefit and contributing to their community or public good. Major concerns included fears or worries about a positive diagnosis, unfamiliarity with research studies, and a distrust of medical institutions and genomic screening.
- Although participants were recruited for hyperlipidemia and colorectal cancer, only one participant had an actionable result of a lipidemia gene; four results were shown as P/LP.
- The average time from enrollment to ROR was 549 days. There were 10 actionable results. There were significant logistical challenges with many individuals. It required 3.2 calls and letters to re-engage, 6/10 were uninsured, 3 certified letters were marked "unclaimed" or returned, and one participant never responded to calls or certified letters.
- Eight participants only spoke Spanish. To account for these socio lingual differences, the program ensured there was always an interpreter available.
- The researchers highlight that it is important to think about how to approach participants with positive results, and the emotional impact of those results will have on the patient.
- There is a need to meet the community where they are in regards to location, timing, cultural preferences, needs, and practices, as well as financial burdens and insurance status.
- Speeding up the turnaround time in returning results would be helpful in eMERGE IV.
- Primary care providers need as much support as can be given (resources/educational materials).
- This is a time where researchers can be forward-thinking and determine how the communities can be best help by genetic testing.

- **Meharry site experiences with recruitment and ROR** | Raj Singh *(Meharry)* | slides
  - Meharry's role in eMERGE is to enroll 500 African American (AA) participants who are high-risk for cancer or with selected cancer. Evidence shows that the incidence of cancer is highest among AA compared to all racial populations. This is particularly poignant for AA males. Participants were recruited for breast cancer, prostate cancer, colorectal cancer, and lung cancer.
  - Investigators obtained blood samples for DNA extraction for germ-line sequencing as well as RNA genomic analysis and proteomics, captured socio demographic survey info and past and future health outcomes from EMR access.
  - Recruitment was started in April 2017, and enrollment began in September of 2017. 500 individuals were enrolled by February 2018. Meharry began returning results in May 2019.
  - Participants were predominantly recruited as at-risk for colon and prostate cancer. Interestingly, the least number of patients recruited were patients who had lung and/or colon cancer.
  - Results showed that of the 500 total results, 496 were received. Of these, 19/496 were positive.
  - Only two patients out of the 19 were able to be reached to discuss the results; one scheduled with a genetic counselor and one declined.
  - The results in the BRCA group showed six positive results. There were five positive results in the lung cancer group, four at risk. In the CRC group, there were four positive results.
  - ROR plan for negative results includes contact up to three times until the participant is reached followed by a letter. Challenges included not being able to extractDNA , had to amend the protocol to allow for return of non-cancer related genes, managing the large turnover in staff, including lead investigator of the study and coordinators, and Meharry had a poor rate of reaching participants in the early ROR phase.
  - Several notes regarding patients with positive results who lack insurance and need further follow-up:
    - Meharry suggested they meet with a genetic counselor and follow up with their PCP.
    - There is an antigen program available for participants to get tested at no cost.
    - Other sites have noticed that these same challenges. There is a network of providers that sites work with, but many do this on a cash price. This is a challenge that eMERGE could address moving forward.

- **Exploring the Impact of Family and Personal History on the Perceived Value and Usefulness of Negative Genetic Test Results** | Sharon Aufox *(NU)* | slides
    - Northwestern has been interested in ROR of negative results. There are concerns about receiving negative genomic results, including participants not understanding negative results or misunderstanding results, and participants not engaging with follow-up preventative screening.
    - Questions were focused on gathering data on usefulness and value on negative results.
    - 178/336 (53%) participants completed the survey..
    - First question that participants were asked was related to family history. Most participants (~60%) were interested in how their family history affected their results.
    - Reactions to the negative results were collected in the qualitative results. Most participants (~65%) felt that they were less worried about their families reactions.
    - Most participants (~70%) felt a 'peace of mind' about receiving negative results.
    - How participants altered their life was also tested; it was found that the majority of participants (~90%) did not really change their lifestyles based on their negative results.
    - Most participants (~90%) said that they would agree to additional genetic testing in the future even if it resulted innegative results.
    - Limitations included a small number of participants, minority populations were not well represented, and the population was highly educated.
    - Questions about concern of residual risks in negatives results were considered but were not included in the survey due to complexity.

eMERGE DAY 2: FRIDAY

- **Opening remarks |** Robb Rowley *(NIH/NHGRI)*
    - The Network is trying to wrap up ROR, the next 10 months will be focused on outcomes and penetrance.
- **Workgroup updates on milestones & discussion|** Moderated by Rex Chisholm *(SC Chair, Northwestern)*, led by Workgroup co-chairs **|** slides
    - Each workgroup was given two sides to summarize the current work that is ongoing in the respective workgroups, challenges encountered, and lessons learned.
    - ROR Workgroup: Iftikhar Kullo and Ingrid Holm
        - Many sites are in the process of returning negative results. ROR has been completed at four sites, and anticipate completion in other sites within the next few months.
        - The workgroup has different manuscripts with a focus on ongoing work and lessons learned, including Sharon Aufox's project presented yesterday.
            - Leadership has noted that the importance of focusing on completing the projects in development as well as lessons learned manuscripts. These prioritized projects can be decided upon in the next nine months.
            - The Participant Survey and Healthcare Provider Survey subgroups also have manuscripts in progress.
        - The group is currently working to define the subsets of Non-Responders (e.g. decliners, deceased participants, transition to adulthood). Prioritization has been placed on understanding non-responder status and the impact of non-responders on outcomes and penetrance analysis work.
        - To help consider geographic and environmental factors of non-response, the group plans to partner with the Genomics group for a geocoding project on non-responders.
        - Initial work can be done on variants with three or more instances to give insight on the time it takes.
        - ACTION ITEM: BCM and LMM will share variant list in order to determine the quantity of P/LP variants for a given disease and help to prioritize efforts.
        - BCM plans to have combined list of VUSs that might require reclassification by late-June, and the combined list from LMM and Baylor can be planned to be available by early-July.

        - ACTION ITEM: The Clinical Annotation workgroup and CSGs will provide details of the flow from a laboratory standpoint for VUS's that are reclassified.
        - ACTION ITEM: The ROR/ELSI workgroup will discuss common themes among sites regarding providing reclassification information to participants.
    - Clinical Annotation Workgroup: Gail Jarvik and Heidi Rehm
        - Clinical Annotation is focused on incidental findings across the network.
        - The preferred order for obtaining the outcomes data for penetrance analysis has been defined.
        - The group is waiting for the 6-months outcomes data to be entered for the Tier 1 diseases.
        - There are several VUS that have been upgraded to L/LP and reports are being generated for return.
        -
        - The group is discussing ways to provide feedback on the outcome forms penetrance questions.Collecting "non-responders" and "non-consenters" data should still be captured for penetrance analysis.
        - Communication between the leaders of Clinical Annotation, ROR, and Outcomes will continue to ensure the Network can maximize the data collected and analyses produced.
    - Outcomes Workgroup: Josh Peterson, Hakon Hakonarson and Marc Williams
    - As of June 2019, the group has completed outcomes data for over 500 participants, many that have more than one phenotype form completed.
    - There are approximately 1430 total outcomes forms to be completed.
        - The group plans to work with the ROR/ELSI workgroup to determine how this total number is affected by "non-responders" and "non-consenters." In addition, there are subsets of eMERGE patients without returned results.
        - According to the ROR Progress Tracker, sites received 1427 positive results and had 267 non-responders (some including PGx results). 1006 positive results were returned to participants.
    - To ensure the correct forms are being used for outcomes analysis (returned results) and penetrance analysis (returned and not returned results), a separate penetrance analysis form will be created for results not returned.
    - ACTION ITEM: The CC will work with the Clinical Annotation workgroup to create penetrance analysis forms for participants with results not returned.
    - For data analysis and QC, two data freezes are anticipated: the first will be released in September 2019 for the October SC/ESP meeting and a second in January for the Winter SC meeting.
        - The first freeze analyses are solely for the steering committee meeting presentations and initial QC of the data.
    - A general consensus is that the date of return should be defined as the date the result was added to the EHR.
    - ACTION ITEM: The CC will add the break down of returned results, non-responders, and non-consented individuals regarding Outcomes & Penetrance analysis topic to the July 2019 PI call for discussion.
    - ACTION ITEM: The CC will circulate the ROR tracker to the Network for reference.
    - EHRI Workgroup: Sandy Aronson and Casey Overby Taylor
        - There is progress being made on the transformation of work from XML format to FHIR.
        - A subgroup has been created for FHIR integration.
        - ACTION ITEM: Investigators interested in joining the FHIR Integration EHRI Subgroup should contact Michelle Stone (timoethia.m.stone@vumc.org) at the CC.
        - .
        - The FHIR integration subgroup will continue to identify use cases and technical implementations.
        - Two main papers under review:
            - NT352: Lessons from eMERGE on readiness for genomic clinical decision support implementation

        - NT301: IGNITE Clinical Informatics Working Group: Genetic Data Pipeline Project
    - The EHRI workgroup is in-process of performing a hazard analysis. The plan to use Hazard analysis to mitigate some of the hazards with integrating genomic information into the EMR that were identified.
- Genomics Workgroup: Megan Roy-Puckelwartz, David Crosslin, and Patrick Sleiman
    - The 105,000 GWAS and eMERGEseq V2 datasets are almost released.
    - All datasets are moving along quickly.
    - The SPHINX focus group discusses the new tool 'eMERGENT', David Crosslin has submitted a U24 for this work.
    - Collecting and compiling data is a challenge for this workgroup.
    - The CC will release the eMERGEseq V2 dataset shortly, there is just one more issue regarding 222 sex mismatches, that likely occured when the site sent demographic data to the CC, not necessarily contamination or sample swap issues. This will be described fully in the QC report. The CSGs and CC also realized there was duplicate mapping to a CHEK2 region and this caused some variants to map improperly to the Y chromosome, this region has been excluded from analysis. .
- Phenotyping Workgroup: Chunhua Weng and Wei-Qi Wei
    - Development of 23 of the 25 main algorithms have been completed. Many of them have already been implemented across the Network. Of the five NLP algorithms, two have already been developed and are currently being validated.
    - Small sample sizes and low PPV results has led to issues in algorithm development.
    - Lessons learned papers are the focus of the workgroup as well as learning how to implement NLP to improve portability of workgroup.
    - The Goal is for all original algorithms to be released by the next SC meeting, the group will continue to work on the NLP algorithms throughout Year 5.
- PGx Workgroup: Laura Rasmussen-Torvik and Cindy Prows
    - eMERGE PGx was prominently featured at CPIC in June.
    - Implementation of PGx is particularly active. The group is looking for Network-wide projects to continue PGx analysis.
    - ROR of PGx is being examined and completed at VUMC and CHOP.
    - NT260 is in-progress and will be submitted for publication soon.
    - One area for feedback would be what is the best way to influence EPIC leadership to make it more user-friendly. PGx is interested in publishing implementation strategies for EPIC as it is difficult to work with, and relay feedback in literature.
    - There is a small eMERGE group that works on implementing family health history into the EMR.
    - Family history or genetic information should be able to be inputted into EPIC in a standardized fashion in order to be used in the EHR.

- **Outcomes Workgroup Science Presentation: Case Studies |** Josh Peterson *(VUMC/CC)*, Maureen Smith *(Northwestern), Alanna Rahm (Geisinger), & Cindy Prows (CCHMC)* | slides
    - The Outcomes workgroup has collected over 500 records of outcomes data from the EHR including several interesting cases of diagnosis and treatment.
    - Outcomes data collection tells the story of return of results to participants and how participants are affected by the result.
    - Maureen presented a case study of a Northwestern participant.
        - The participant's primary care provider was also part of the Northwestern healthcare system.
        - When enrolled, the participant stated that he had no family history to suggest genetic testing. The results came back with two positive mutations of two forms of cystic fibrosis.
            - Results were returned over the phone.
            - He was thrilled to receive result and began relaying information based on his health history including chronic sinusitis, digestive issues, and infertility.

        - Participant was referred to the adult cystic fibrosis clinic located in the children's hospital and received full series of tests: chest x-ray, vitamin-d test, sweat chloride. Histherapy, resulting from his genetic testing, has made a big difference in his care and his outlook.
    - Alanna described case findings of a 61-year-old MyCode participant from Geisinger with multiple comorbidities: type 2 diabetes, kidney disease, hypothyroidism, PAD, anemia, and obesity.
        - Recently before receiving result, the patient had undergone an amputation of her right foot which led to ischemic stroke as a complication. An LDLR mutation result was returned in August 2017 while she was in the nursing home. She did not keep the appointment with the genetic counselor due to recovery issues. Shehad another amputation after return of result which led to another stroke. The patient also has extensive coronary artery disease. The participant is unable to tolerate statins and now has a BRCA diagnosis. Her son had a heart attack at 39 and was unable to take care of himself.
    - Cindy presented the case report of an adolescent participant from CCHMC. The adolescent and parent independently chose to learn all possible results.
        - Participant's father and sister had a history of migraine and temporary muscle weakness/paralysis. They were evaluated in the genetics clinic. The mother reported periodic numbness, falling, and word choice issues, however previous neurology evaluations were found to be negative. The father had hemiplegic migraines and engaged in post strenuous exercise when younger.
    - These three cases indicate the challenges for outcomes research but for the healthcare system in general in synthesizing information.
- **Genomics Workgroup lessons learned panel & discussion |** David Crosslin *(UW/CC)*, Megan Roy-Puckelwartz *(Northwestern)*, Ian Stanaway (UW/CC), & Mullai Murugan (Baylor) | slides
    - Megan Roy-Puckelwartz "*How sample size, diversity, and target genes affect implementation and discovery of datasets*"
    - Genomics is widely successful when you have a very large, rich dataset. Genome-wide data genotyping creates a richer discovery set, but requires extensive phenotyping algorithms.
    - Panel-based sequencing create a richer implementation data set. Panels can select genes to understand how genes interact, to understand penetrance, CNV analysis, mosaicism. This approach is not likely to result in new gene discovery. When you have panel-based sequencing, you need to be more thoughtful of how to create discovery hypotheses. Need to be vigilant about goals, and clear hypothesis need to be made.
    - Adding additional data is costly and time-consuming. Need to weigh benefits and costs.
    - continuing to add data and recreating multisample VCFs results in delays in data analysis. Groups need to be mindful that there will always be another dataset that is bigger and better.. The challenge for Networks is to understand the importance of data freezes and establishingclear goals and objectives to maximize the science for a given amount of funding.
    - An additional challenge is when groups create new requirements resulting in delays of analysis. This has been evident when NHGRI pushed for diverse datasets; this additional requirement lead to delays while we were waiting on the data from these samples to be available.
    - eMERGE has very large datasets, and the group needs to understand how size affects the speed and efficiency of analysis and QC. There needs to be better prioritization vs. speed and efficiency.
    - Investigators should limit the time for comments and suggestions when consensus is required. Make decisions quickly, and learn to prioritize are important elements regarding compilation and analysis of large data sets.
    - An example of the importance of organizing and prioritizing dataset analysis was with the Genomics Workgroup. They did not have the necessary power to run certain data because the group did not know the phenotype status. This is important especially in the future if anydata is not present or available.
    - Ian Stanaway "*Lessons Learned and Earned: Compute Time with Big Genomic Data and Naming Conventions and IDs*"

- To manage the array data (83k), it takes 96 hours of compute time for the largest batches. It took about six months to actually impute and merge the 78 sets. This does not include the QC. The compute time is about a month and a half when a change is made to put everything back together again. Ian cannot finish the genetic QC until the genetic and demographic files have been uploaded to the CC. This would then require a nine day rewrite.
- Collecting and finalizing the demographics file at the beginning of the process and freezing the dataset at that point, would help with frequent withdrawals and additions which add significant time . Potentially could move the workflow to Github to help upload data.
- For the GWAS set, there are several different ways the files were titled, including some without eIDs, he had to write code to detect differences and standardize.
- Investigators should consider to enlist IT in streamlining heterogeneity of naming and organizing data prior to initiating a large sequencing or genotyping project.
- The Network needs to be more disciplined about data freezes. Even once the pipeline is built, adding and subtracting data take significant amounts of time and effort.
- It is unknown how much this kind of work would cost on AnVIL, or a cloud computing environment.
    - Each restart and test there will be computation cost, so it will be more important to plan out analyses and data freezes on the front end.
- Mullai Murugan: "*eMERGE Commons and Genomic Analysis*"
- eMERGE is a complex ecosystem and requires a storage and analysis platform. The eMERGE Data commons primary goal is storage, but gives a portal that also allows for analysis. As there is so much data, there is a need for an analysis platform that gives access for data mining and research.
- The PHI partition allows access to authenticated users of clinical reports and other related files. The non-PHI partition focus contains raw data, access to all clinical sites, the analysis tools, and the BCM/LMM data. eDAP portal serves for PHI partition storage.
- eCAP is the eMERGE commons access portal, and serves as one of the analysis tools.
- There is a structural variation project with parliament 2. Aims are to increase resolution and sensitivity of CNV calls, identify novel copy variants. This would aid in interpretation and pathogenicity assignment.
    - Once the BAMs were constructed, they were pushed into parliament 2. It has taken 115,400 hours to compute the data.
- Data are just now starting to come from this project. CNV analyses were completed and showed some variation across chromosomes. The group tried to identify the number of CNVs across of all of the sample including deletions and duplications.
- Lessons learned include a common need for a cloud platform that helps process optimization and management that gives a standard model for operations.
- Next steps include finishing of the SV calling project and dissemination of results, as well as consideration of the AnVIL plan and how data can move from one environment to another without wasting too many resources.
- There can be orthogonal validation between the 4000 in eMERGEseq and the GWAS cohort.
- ACTION ITEM: The Genomics Workgroup will write up a paper on the lessons learned from eIII.
- The commons lessons are part of the harmonization paper.

- **The impact of return of unsolicited results on health care providers (HCPs) in eMERGE III: Preliminary findings** | Ingrid Holm *(BCH)* | slides
    - The health care providers survey (HCP) supplement R01 was one of the first to identify the clinicians point of view as health care providers face different challenges as part of disseminating ROR.
    - The eMERGE network is one of the first to return unsolicited genomic results to providers.
    - The HCP RO1 aimed to survey participants' health care providers one month after receiving an unsolicited positive result, conduct qualitative interviews at four eMERGE sites, and identify approaches to returning unsolicited genomic sequencing results.
    - The group interviewed providers six months after results were given to determine their thoughts/actions/feelings on delivering a positive result and actionable items.

    - The group also interviewed providers when negative results were given to assess any actions in delivering these results, as well as any feedback in how this may have affected clinical care.
    - There were 306 enrollment invitations sent to HCPs with a 38% response rate for email, 8% response rate for mail with a29% overall response rate.
    - About 80% of respondents are in primary care and practice in a hospital (41%) or outpatient (44%) setting.
    - Generally, initial reactions to receiving genomic result were positive (33.3% strongly agree) and found it to be informative (35.7% as strongly agree).
    - Most (77%) HCPs planned to spend less than 15 minutes with patients discussing the results.
    - Most patients were referred to a specialist (55%) and 41.2% were referred to geneticists for testing.
    - Most HCPs found that the genetic testing was important to the patient's healthcare (46% strongly agree). However, 14% of participants did not find the results useful and 25% of the patients wanted a letter containing results.
    - Statements from HCPs showed that while the idea of genetic testing is exciting, it is difficult to recommend it when they are not sure who will be providing the counseling or follow-up.
- **EHRI Workgroup Science Presentation: Hazard analysis of CDS |** Sandy Aronson (Harvard) & Casey Overby Taylor *(Geisinger/JHU)* | slides
    - Hazard analysis must be completed before introducing a new medical device into clinical care. A hazard analysis identifies possible issues with the device, classifying potential hazards, identifying mitigations, and assessing whether the overall risk profile of the device is acceptable..
    - As eMERGE sites move more towards CDS of electronics, a new working group has been introduced to perform the hazard analysis. The scope of the analysis included focusing on two types of genomic decision support:
        - The first type of genetic decision support involves a geneticist constantly signing out cases and providing clinicians with knowledge updates
        - Second scenario is PGx alert. PGx status may be known and the CDS is triggered when a provider needs ordering guidance.
    - Because patient PGx status is not known, using an ancillary genomic status approach is useful.
    - During the workgroup breakout session, 25 hazards were identified, including inappropriate alert firing context (nine), technical issues (eight), user experience problems that included issues that were disseminated to the clinician but not the patient (five), and knowledge management that included issues that were disseminated to the laboratory, not the clinician/patient (three).
    - A risk index table was created measuring occurrence by severity and then sorted by type. Although there are challenges in standardizing hazard analysis, it can provide great value in implementation. Any group that is implementing the hazard analysis must take into account local variation in practice that might not be recognized in a generic hazard analysis. It would be helpful to have examples that can help assisting how to predict hazards and label appropriately.
- **Clinical Annotations Workgroup Science Presentation: Incidental Findings |** Adam Gordon *(Northwestern)* | slides
    - For Adam Gordon's PGRNseq project (NT179), an incidental finding is a L/LP variant on the consensus actionable list, and unrelated to site-submitted participant indication. A finding is still incidental if the participant is discovered to have a relevant but unsubmitted indication. If a person had no indication, then there are no incidental findings, they are just findings.
    - There is difficulty in harmonizing indications, and this needs to be clarified in future publications.
    - Adam is working on the incidental findings committee for ACMG so the eMERGE incidental findings are contributing to the ACMG gene list. There was about 3.02% positive variants that had Incidental Findings (IF). By site, rates are consistent at about 2%, with the exception of Geisinger which showed 10% IF.
    - Incidental finding rates by self-reported race and ethnicity showed that certain genes may be more typical in White populations and Ashkenazi Jews. This may be due to founder effects.

- Cancer was the highest disease domain to find a likely pathogenic and pathogenic variant. The pathogenic variant result was the second most common finding. However, LP BRCA2 and LP LDR2 results were also the most commonly found incidental findings.
- The most commonly found individual findings were BRCA2 pathogenic gene in Ashkenazi Jews (1.52% IF%), and the MYBPC3 variant that was found in 13 individuals (11 Asian, 2 Unknown). LDLR had eight CNV incidental findings which was the most common.
- RYR1 is a founder allele and common in the Midwest. Investigators would be interested to see if RYR1 is more common in midwestern sites.
- ATM is not on the consensus list and not returned by all sites.
- ACTION ITEM: Investigators should contact Adam Gordon (adam.gordon@northwestern.edu) by the end of July if they are interested in collaborating on the PGRNseq paper.

- **Closing remarks |** Rex Chisholm *(SC Chair, Northwestern)*
  - Rex would like to thank Gail and David for hosting the SC Meeting in Seattle and organizing the Harbor cruise. Thank you to the CC for the support. The Network is pleased with how smoothly the meeting went.

Meeting Action Items

- **Coordinating Center**
  - The CC will work with the Clinical Annotation workgroup to create penetrance analysis forms for participants with results not returned.
  - The CC will add the break down of returned results, non-responders, and non-consented individuals regarding Outcomes & Penetrance analysis topic to the August 2019 PI call for discussion.
  - The CC will create a rough estimate about how many results we expect to obtain for 6 month and 12 month Outcomes
  - The CC will circulate the ROR tracker to the Network for reference.
- **ROR Workgroup**
  - The ROR/ELSI workgroup will discuss common themes among sites regarding providing reclassification information to participants.
- **Clinical Annotation Workgroup**
  - The Clinical Annotation workgroup and CSGs will provide details of the flow from a laboratory standpoint for VUS's that are reclassified.
  - Investigators should contact Adam Gordon (adam.gordon@northwestern.edu) by the end of July if they are interested in collaborating on the PGRNseq paper.
- **CSG Operations Group**
  - BCM and LMM will share variant list in order to determine the quantity of P/LP variants for a given disease and help to prioritize efforts.
- **Network**
  - Investigators with questions or samples for validation regarding the somatic mutation project, NT171: *Possible somatic mutation in targeted sequence data,* should contact Ken Kaufman (kenneth.kaufman@cchmc.org) and Paul Gecaine (paul.gecaine@cchmc.org).
- **Genomics**
  - The Genomics Workgroup will write up a paper on the lessons learned from eIII. The commons lessons are part of the harmonization paper.
- **EHRI**

Investigators interested in joining the FHIR Integration EHRI Subgroup should contact Michelle Stone (timoethia.m.stone@vumc.org) at the CC.