Background Information Table of Contents

I. Network Documents

eMERGE Network Collaborations

Cincinnati Children's Hospital Medical Center/Boston Children's Hospital

1. Joris A. Veltman, Nijmegen Centre for Molecular Life Sciences, **Radboud University Nijmegen Medical Centre**: Trio Exome Sequencing
2. Children's Hospital of Philadelphia: Asthma and ADHD (both led by CHOP
3. Vanderbilt/CHOP: Early Childhood Obesity (led by CCHMC/BCH)
4. **The International Consortium for the Genetics of Systemic Lupus Erythemat**osus (SLEGEN): Harley, CCHMC member site
5. Dr. Hutton leads the consortium to federate studies based on the EMR across 30 centers that treat children with Inflammatory Bowel Disease (R01 HS020024).
6. Co-PIs Dr. Louis Kunkel and Dr. Christopher Walsh at BCH with co-investigators Dr. Isaac Kohane and Dr. Ingrid Holm: **Phenotypic and Genetic Factors in Autism Spectrum Disorders**. This study led to BCH designation as a Simons Simplex Collection site and a member of the Boston-area Autism Consortium.
7. PI Dr. Amy Roberts at BCH: **Cardiac Gene Project**.
8. Director Dr. Alan Beggs at BCH: **Manton Center for Orphan Disease Research** (Dr. Ingrid Holm and Dr. Isaac Kohane - Gene Discovery Core scientists). This center frequently collaborates with the Gene Partnership program, from which BCH's eMERGE cohort originates.
9. Dr. Ingrid Holm at BCH leads **Informed Cohort Oversight Board** – multidisciplinary team from multiple institutions to oversee research functions and communication of results back to patients; Return of Results Consortium – funded by, and part of, the NHGRI ELSI Research Program.
10. Dr. Isaac Kohane: Led the **collaborative development of i2b2** (piloted across hospitals in the Harvard Catalyst) which has been adopted by over 60 academic health centers internationally. Supports an Academic Users' Group of over 250 members from over 65 independent institutions that meets biannually for code workshops, discussion of application issues, preview of coming software, and networking; New efforts directly related to i2b2 have resulted in at least 27 new collaborative grants. Established Summer Institute in Bioinformatics and Integrative Genomics – in partnership with the Harvard-MIT Division of Health Sciences and Technology (HST).
11. Dr. Ingrid Holm at BCH is a co-investigator on a newly-awarded U19 cooperative agreement – "Genome Sequence-Based Screening for Newborn Illness and Childhood Risk". The Principal Investigators for this study are Dr. Robert Green of Brigham and Women's Hospital and Dr. Alan Beggs of BCH.

Children's Hospital of Philadelphia

1. **Neurodevelopmental Genomics: Trajectories of Complex Phenotypes**: The goal of this study is to perform comprehensive neurocognitive phenotyping on 10,000 children who are already genotyped and perform methylation profiling, imaging and establish EBV cell lines on a subset of them. This will be made available as a public resource. (PIs: H. Hakonarson, CHOP and R. Gur,UPenn)
2. **CHOP/PENN Center for Excellence for Autism Research:** This is a comprehensive study of neuroanatomical features of children with ASD. It aims to characterize brain structure, connectivity (using diffusion tensor imaging), and brain function to investigate possible causal mechanisms for the heterogeneity in ASD. (Co-PI H. Hakonarson, CHOP)
3. **Cholestatic Liver Disease Consortium:** An ancillary study, Genetic Modifiers of Liver Disease Severity in Alagille Syndrome is designed to identify genetic factors associated with liver disease severity in Alagille syndrome. (PI, N. Spinner, CHOP)
4. **Brain, Behavior and Genetic Studies of 22q11.2 Deletion Studies**: The objective of this collaboration between CHOP and Penn is to combine genetic and neurobiologic paradigms for understanding pathogenesis and for detection of genes that modulate susceptibility to psychosis with phenotypic features of schizophrenia and related disorders (SCZ). (PI, B. Emanuel, CHOP)
5. **Alzheimer's Disease Genetics Consortium:** The goal of this study is to perform a whole genome scan to test for association of AD to a high-throughput tag-SNP arrays. (PI, G. Schellenberg, UPenn)
6. **Pediatric Cardiac Genomics Consortium:** We are studying the genetic basis of conotruncal defects. Genetic risk factors for disease are identified using genome wide association studies and

copy number variant analyses. An inception cohort with tetralogy of Fallot would be ascertained. (PI, E. Goldmuntz, CHOP; PI on genotyping network facility: H. Hakonarson)

7. **NICHD multi-center, multi-ethnic longitudinal Bone Mineral Density in Childhood Study:** We are conducting a genome-wide association study of bone mineral accretion during childhood. Our goal is to identify the genetic determinants of bone health early in life with the ultimate goal of identifying new pathways for osteoporosis prevention. (Co-PI, S. Grant, CHOP and B. Zemel, CHOP)
8. **Clinical Sequencing Exploratory Research**: The Applying Genomic Sequencing in Pediatrics group is bringing genomic sequencing into pediatric clinical settings. Researchers will work with families, scientists, and ethicists to determine how patients should be counseled and educated before testing, what data should be provided back to individual families, and what educational tools will help in understanding the implications of the testing. (Co-PIs, I. Krantz, CHOP and N. Spinner, CHOP)
9. **NGS:** Whole genome scan to test for association to autism utilizing high-throughput tag-SNP arrays. The project used genotyping and sequencing data generated by CAG and CAG-collaborators for association studies of both common and rare variants in autsm. (PI H. Hakonarson)
10. **NGS:** Type 1 diabetes, targeted sequencing. (PI, H. Hakonarson)
11. **NGS:** Crohn's disease, targeted sequencing. (PI, H. Hakonarson, CHOP)
12. **NGS:** Exome sequencing, early forms of schizophrenia. (PI, R. Gur, UPenn)
13. **GWAS:** Childhood obesity. (PI, S. Grant, CHOP)
14. **GWAS:** Gene-Environment Interactions in Asthma. (Co-PI, H. Hakonarson, CHOP)
15. **GWAS:** Neuroblastoma. (PI, J. Maris, CHOP)
16. **GWAS:** High density lipoprotein cholesterol. (PI, D. Rader, Upenn)
17. **GWAS:** Latent autoimmune diabetes in adults. (PI, S. Grant, CHOP)
18. **GWAS:** Schizophrenia. (PI, H. Hakonarson, CHOP)
19. **PGx:** Ventilation in pediatric patients with respiratory failure. (PI: A. Zuppa, CHOP)

Geisinger

1. **HMO Research Network:** member site; collaborating with investigators at Group Health on the pharmacogenomics of extreme weight gain due to 2nd/3rd generation anti-psychotic use funded by the International Severe Adverse Events Consortium (iSAEC).
2. **Pharmacogenomics Research Network (PGRN):** associate member site; clopidogrel genetics collaborations with investigators from the University of Maryland funded by the NIH as part of the Pharmacogenomics of Antiplatelet Intervention 2 (PAPI-2) Study, a multi-center prospective double-blind randomized comparative effective trial comparing personalized anti-platelet therapy to standard of care in patients undergoing percutaneous coronary interventions.
3. **Clinical Decision Support Consortium:** The goal of the CDSC is to assess, define, demonstrate, and evaluate best practices for knowledge management and clinical decision support in healthcare information technology (IT) at scale –across multiple ambulatory care settings and EHR technology platforms.
4. **AAA Meta-GWAS Consortium:** Collaborating with WTCCC, University of Utrecht, deCODE Genetics, and New Zealand AAA Study.
5. **Aneurysm Global Epidemiology Study (AGES):** Collaborating with Edward Choke from University of Leicester, UK.
6. **Mid-Atlantic Nutrition Obesity Research Center (NORC):** The Mid-Atlantic NORC brings together resources at the University of Maryland, Johns Hopkins University, The US Department of Agriculture and Geisinger Health System; and cross-cutting expertise in various fields that can address the basic mechanisms that determine individual responses to nutrient intake and energy imbalance.
7. **PA CURE:** A grant awarded by the PA-DOH and lead by Geisinger in collaboration with The University of Pittsburgh and Temple University to create a risk scoring tool and test its utility for population screening of AAA.

Group Health/University of Washington

1. Alzheimer's Disease Genetics Consortium (**ADGC**)
2. Strategic Health IT Advanced Research Projects (**SHARP**) : member site; an effort to develop open-source tools for transforming EHR data into standards-conforming comparable information suitable for large-scale analyses, inferencing and integration of disparate health data

3. (GENE enVironment Association) (**GENEVA):** to test and validate an automated algorithmatic method to identify mosaic regions
4. Northwest Institute of Genetic Medicine (**NWIGM**): to use the ongoing NWIGM biorepository to expand subject pool for eMERGE Phase 2. There are approximately 1700 subjects in NWIGM that have not been GWAS'd for eMERGE to date and which remain as an expansion cohort. All 2024 subjects in NWIGM will have exome chips available to eMERGE.
5. HMO Research Network: member site; GHRI recently completed a CommonFund Collaboratory project to help develop resources and capacity to support large-scale epidemiological studies(megaepidemiology) with potential to link to HMO based biobanks to address critical questions about the burden, causes, course, treatment, and prevention of common and rare human diseases.
6. **Scalable and Robust Clinical Text De-Identification Tools:** New R01 funded by National Library of Medicine that is a collaboration between Group Health and Vanderbilt to develop novel text de-identification methods. This work may facilitate improved access within an institution to its clinical text (e.g. for phenotype development purposes) as well as sharing of clinical text across institutions.
7. **NWIGM:** to study markers of Erectile Dysfunction (*potential collaboration*) in collaboration with Dr. Hunter Wessels, University of Washington Medical Center. Several eMERGE sites have expressed interest in participating in this collaboration.
8. **CARE**: WBC (*potential collaboration*)

Marshfield/Essentia/Penn State

1. **AMD** consortium Retinal specialists from Vanderbilt confirmed case status. Data will be shared in the near future with the network which is attempting to identify novel genetic variants associated with AMD.
2. **Glaucoma** consortium: additional samples were GWAS'd. Several papers have been accepted for publication for Phase 1. (1R01EY022305-01)
3. **PhenX** received admin supplement for year 1 of eMERGE II. Seven sites were funded in the PhenX RISING network. The network modeled their data sharing agreement on the eMERGE DUA. Supplement funding is complete. PhenX looked at Ecologic Stressors, Post-Traumatic Stress Disorder, and Drug Use in Detroit, University of Michigan, Ann Arbor, Allison Aiello, PhD and National Institute on Drug Abuse. PhenX has 1 paper published and 3 more papers under review.
4. **NDPBRN National Dental Practice Based Research Network** to support dentists in individual practices or CHC to conduct research.
5. **"Oral-Systemic Health Consortium"** with Mt. Sinai, Univ. of Pittsburgh, UMDNJ, and UNC. The Oral-Systemic Health Consortium is in the process of enrolling study subjects along with microbiome samples
6. **HMORN** member site for Marshfield and Essentia; The Collaboratory will seek to enable large-scale epidemiological studies to address critical questions about the burden, causes, course, treatment, and prevention of common and rare human diseases. Development/enhancement of web-based patient portals and creation of megacohorts seem to be achievable goals of an eMERGE/HMO RN Collaboratory partnership.
7. **International Health Terminology Standards Development Organization** (IHTSDO): standards for safe, precise and effective exchange of clinical and health information.
8. **Clinical and Translational Science Awards** (CTSA): The goal of the University of Wisconsin Institute for Clinical and Translational Research (UW ICTR) is to create an environment that transforms research into a continuum from investigation through discovery and to translation into real-life community practice, thereby linking even the most basic research to practical improvements in human health.
9. **Wisconsin Genomics Initiative** (WGI): The Wisconsin Genomics Initiative advances personalized health care research. Scientists will be able to predict an individual's risk of developing a disease, precisely target a personalized treatment and ultimately prevent disease before it occurs.
10. **PGRN-PGPOP** - PharmacoGenomic discovery and replication in very large patient POPulations. PGPop was conceived as a network resource to provide to PGRN an opportunity to identify large groups of real world patients with known drug exposures and outcomes for pharmacogenomic study in a clinical setting.
11. **Iron Metabolism** (Christopher Vulpe, PI) "An integrated molecular approach to understand variation in iron metabolism" (1R01GM083198-01A1).
12. **A Traumatic Brain Injury (TBI) Patient Registry to Examine Variations in Diagnosis, Treatment and Outcomes** (Allen Heinemann and Abel Kho) As one specific Aim we propose to: Develop and validate a traumatic brain injury phenotype, linked to existing genetic data, using EHR data within the

eMERGE consortium. In addition to defining the phenotype, we would also try and document certain objective measures of TBI measures and outcomes.

13. **GEMS Consortium** A number of pharmacogenomics research centers (the "consortium") are interested in discovering new genetic variants that are important for predicting myopathy in the context of statin therapy using candidate gene association studies (CGAS), genome wide association studies (GWAS), exome scanning, whole genome sequencing, and other methods. Many members of the consortium have collected cohorts of patients who have been treated with lipid modifying drugs and have obtained a variety of response phenotypes, including both efficacy and toxicity endpoints, and have genotype data and/or DNA available from these participants. Members of the consortium are willing to share these data for the purposes of replicating the findings of others, and for combined (meta) analysis. Based on the success of similar consortia, other disease genetics consortia, and other pharmacogenetic consortia GEMS seeks to utilize this same approach to advance pharmacogenomics for statin therapy.

Mayo Clinic

1. **AT&T** Foundation
2. Multi-source Integrated Platform for Answering Clinical Questions (**MiPAQ**)
3. The Environmental Determinants of Diabetes in the Young Study (**TEDDY**)
4. **World Health Organization (WHO):** International Classification of Disease (ICD)-11 Revision
5. Strategic Health IT Advanced Research Projects (**SHARP**): lead site; an effort to develop open-source tools for transforming EHR data into standards-conforming comparable information suitable for large-scale analyses, inferencing and integration of disparate health data
6. cancer Biomedical Informatics Grid (**caBIG**®)
7. Genome-wide Association Study (GWAS) of **Venous Thromboembolism**
8. Genetic Epidemiology Network of Arteriopathy (**GENOA**)
9. Consensus Measures for Phenotypes and Exposures (**PhenX**): Network collaboration on eleMAP and other data standards projects
10. Pharmacogenomics Research Network (PGRN): lead site, an effort to lead discovery and advance translation in genomics, in order to enable safer and more effective drug therapies.
11. Pharmacogenomics Ontology Network Resource (PHONT within PGRN): PGRN network resource, an effort to support meta-analyses, achieve translational goals, and facilitate the messaging of pharmacogenomics-related data from and into clinical environments such as EMRs
12. National Center for Biomedical Ontology (NCBO): collaborating site, an effort to leverage ontologies to increase access to and understanding of defined terms and relationships in the biomedical domain.
13. Beacon Community Program within the ONC (Beacon/ONC): lead site, an effort to demonstrate the ability of health IT to transform local health care systems.
14. Clinical and Translational Science Awards (CTSA): lead site, an effort to speed discovery and advance science aimed at improving our nation's health
15. Office of National Coordinator for Health Information Technology (ONC) is a resource to the entire health system to support the adoption of health information technology and the promotion of nationwide health information exchange to improve health care.

Mount Sinai School of Medicine

1. **GIANT** (Genetic Investigation of Anthropometric Traits) – GWAS data contributed for discovery analysis for anthropometric traits from all BioMe participants and workgroup participation
2. **COGENT BP** (Continental Origins and Genetic Epidemiology Network) – GWAS data for BP from African American BioMe participants contributed for discovery analysis
3. **GHBP** (Genomics in Hispanics for Blood Pressure) – GWAS data for BP from Hispanics contributed for discovery analysis
4. **Massachusetts Institute of Technology** Computer Science and Artificial Intelligence Laboratory (John Guttag): Predictive Modelling and Personalized Health Decision Support Tools
5. **Genetics of Obesity and related traits in African Americans** – GWAS data of BMI from African Americans BioMe participants contributed for discovery and follow up analysis and workgroup participation
6. **African American Type 2 Diabetes Genetics Consortium** – GWAS data of T2D from African American BioMe participants contributed for analysis
7. **CHARGE (Cohorts for Heart and Aging Research In Genomic Epidemiology)** Exome chip date contributed for analysis of BP from all BioMe participants

8. **CKDGen (CKD Genetics Consortium)** – GWAS data contributed for discovery analysis and workgroup participation
9. **CKDGen (CKD Genetics Consortium)** Exome chip data contributed for analysis of BP from all BioMe participants
10. **GLGC** (Global Lipids Genetics Consortium) Exome chip data contributed for discovery analysis of all lipids from all BioMe participants
11. **GLGC** (Global Lipids Genetics Consortium) Exome chip data contributed for follow-up analysis of CAD from all African American BioMe participants
12. **ESP-LDL (Exome Sequencing Projects LDL Cholesterol)** Exome chip data contributed for follow-up analysis of LDL from all BioMe participants
13. **MAGIC (Meta-Analyses of Glucose and Insulin-related traits Consortium) Exome chip data** contributed for discovery analysis of HbA1c from all BioMe participants
14. **CHARGE (Cohorts for Heart and Aging Research In Genomic Epidemiology)** Exome chip data contributed for discovery analysis of Glycaemic traits from all BioMe participants
15. **CHARGE (Cohorts for Heart and Aging Research In Genomic Epidemiology)** Exome chip data contributed for follow-up analysis of Amyloidoses from all BioMe participants
16. **MEDIA MEta-analysis of type 2 DIabetes in African Americans (MEDIA) Consortium** GWAS data contributed for discovery analysis of T2D from African American BioMe participants.
17. TranscenD (TRANS-ethnic Evaluation of vitamin D) GWAS data contributed for discovery analysis of Vitamin D from all BioMe participants.
18. **Lipids in HA** GWAS data contributed for discovery analysis of Lipids from all Hispanic American BioMe participants.
19. **AAGILE** (African American Glucose and Insulin Genetic Epidemiology (AAGILE) Consortium) GWAS data contributed for discovery analysis of HbA1c and glucose from all African American BioMe participants.
20. **DIAGRAM+ and GOT2D (Genetics of Type 2 Diabetes)** GWAS data contributed for follow-up analysis of T2D from all BioMe participants.
21. **BP in HA** GWAS data contributed for discovery analysis of BP from all Hispanic American BioMe participants.
22. **ICBP** (International Consortium for Blood Pressure) GWAS data contributed for discovery analysis of BP from all European American BioMe participants.
23. **Anthropometric Traits in HA** GWAS data contributed for discovery analysis of Anthropometric Traits from all Hispanic American BioMe participants.
24. **T2D Genes** Targeted sequencing for follow-up analysis of T2D from all European American BioMe participants.

Northwestern University

1. **African American Type 2 Diabetes Genetics Consortium (MEDIA):** Don Bowden (Wake Forest) leads a collaborative group of ~12 sites including Northwestern and Vanderbilt. This consortium conducted a meta-analysis of genomic determinants of Type 2 Diabetes in African Americans. Paper is submitted to Nature Genetics, with multiple eMERGE investigators and the eMERGE consortium represented as authors.
2. **COGENT** (Continental Origins of Genetic Traits): New meta-analysis consortium focused on African-American and Latino populations; NU participates in trait-specific working group (for height & RBC phenotypes)
3. **GIANT** Consortium: Meta-analysis of height from >100,000 subjects across multiple studies. Awaiting draft of manuscript from consortium lead analysts.
4. Vitamin D Related Innate Immunity in Influenza (with VU)
5. **PAGE** (Population Architecture using Genomics and Epidemiology) : inc. a PheWAS replication in eMERGE; VU also wanted blood count data for AA pts (was provided)
6. **CTSA with Mayo** (neutropenia and thrombocytopenia thus far)
7. **CAGE?** (QRS in African Americans with VU)
8. eMERGE Urological GWAS RO1
9. **UCSF: Metformin Project**
10. **AAA Genotyping**– collaborating with Geisinger
11. **De-Id Project (Anonymization of clinical codes in support of genome-phenome association studies)**
12. **AAGILE/MEDIA Fasting Glucose meta-analysis**

Vanderbilt University

1. **QRS GWAS Consortium**: across eMERGE; replication planned with CHARGE
2. QT Interval GWAS Consortium **(QT-IGC)**
3. **PRIMA** (CHARGE-led mega meta-analysis of PR interval)
4. Pharmacogenomics Research Network (**PGRN**): member site
5. **Kaiser Permanente and Marshfield Clinic** – statins & MI
6. **UCSF and RIKEN** – Metformin-related glycemic response
7. **Marshfield Clinic, Harvard Crimson, and Harvard Pilgrim** – Asthma response to inhaled steroids; methotrexate-induced liver injury
8. **RIKEN** – ACE inhibitor-associated angioedema
9. **Baylor** – Amio-induced pulmonary toxicity and thyrotoxicosis
10. **St Jude Children's Research Hospital** – Steroid-induced osteonecrosis
11. **Children's Hospital Oakland Research Institute (CHORI) and Marshfield Clinic –** Statin effects in asthma
12. **University of Florida –** Cerebrovascular disease and clopidogrel
13. **University of Colorado Boulder** – Seborrheic keratosis
14. **University of Maryland** – Association of CES1 G143E with bleeding and decreased events during clopidogrel therapy
15. **ICPC (International Clopidogrel Pharmacogenomics Consortium)**
16. **CPIC (Clinical Pharmacogenetics Implementation Consortium)**
17. Consensus Measures for Phenotypes and Exposures (**PhenX**): Network collaboration on eleMAP and other data standards projects
18. GANI_MED biorepository in Greifswald, Germany: biobank ELSI collaboration
19. Children's Hospital of Eastern Ontario (CHEO) / University of Ottawa to study re-identification risks in electronic medical records data tied to genomic records.
20. Sabanci University and Zirve University in Turkey to develop anonymization strategies for longitudinal records, which we evaluated with the Vanderbilt's QRS GWAS cohort.

eMERGE Phase II Publications from April 2013 -October 2013

Published/Accepted Phase II Network Manuscripts

1. Integration of Genomics into the Electronic Health Record Special Issue. GENET MED. 2013 Oct 15(10).
    a. Williams M, Kannry J. Introduction to the Special Issue: Integration of Genomics into the Electronic Health Record.
    b. Gottesman O, Kuivaniemi H, Tromp G, Faucett WA, Li R, Manolio TA, et al. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future; **PMID: 23743551**
    c. Kho AN, Rasmussen LV, Connolly JJ, Peissig PL, Starren JB, et al. Practical challenges integrating genomic data into the electronic health record.
    d. Ury A. Storing and Interpreting Genomic Information in Widely Deployed Electronic Health Records Systems; **PMID: 23949573**
    e. Marsolo K, Spooner A. Clinical Genomics in the World of the Electronic Health Record; **PMID: 23846403**
    f. Hartzler A, McCarty C, Brilliant M, Clayton E, Faucett A, et al. Stakeholder engagement: a key component of integrating genomic information into electronic health records; **PMID: 24030437**
    g. Chute C, Wood G, Ullman-Cullere M. Some experiences and opportunities for big data in translational research; **PMID: 24008998**
    h. Iftikhar K, Hazin R, Clayton E, Brothers K, Williams M. Ethical, Legal and Social Implications of Incorporating Genomic Information into Electronic Health Records; **PMID: 24030434**
    i. Overby C. Opportunities for Genomic Clinical Decision Support Interventions. Tarczy-Hornoch P. A survey of informatics approaches to whole exome and whole genome clinical reporting in the electronic medical record.
    j. Peterson J, Denny J. Electronic Health Record Design and Implementation for Pharmacogenomics: a Local Perspective; **PMID: 24009000**
    k. Kannry J, Williams M. The Undiscovered Country: The Future of Integrating Genomic Information into the EHR.

2. Gottesman O, Kuivaniemi H, Tromp G, Faucett WA, Li R, Manolio TA, et al. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. Genet. Med. 2013 Jun 6; **PMID: 23743551**

3. Chute CG, Kohane IS. Genomic medicine, health information technology, and patient care. JAMA. 2013 Apr 10;309(14):1467–1468. **PMID: 23571583**

4. Wei W-Q, Leibson CL, Ransom JE, Kho AN, Chute CG. The absence of longitudinal data limits the accuracy of high-throughput clinical phenotyping for identifying type 2 diabetes mellitus subjects. Int J Med Inform. 2013 Apr;82(4):239–247. **PMID: 22762862; PMCID: PMC3478423**

In Process Phase II Network Manuscripts

1. Developing Consents for Returning Pharmacogenomics Results: The eMERGE Experience. Lead Investigator: Maureen Smith (NU)

2. Genomics Workgroup's Frontiers in Genetics Special Issue
    a. Imputation and QC for combining genome-wide datasets. Lead Investigators: Marylyn Ritchie (MC/EIRH/PSU & CC)
    b. Evaluation of population stratification in large biobanks linked to EHR. Lead Investigator: David Crosslin (GroupHealth/UW)
    c. State of the state in returning genomics research results. Lead Investigator: Iftikhar Kullo (Mayo)
    d. EMR-linked GWAS study: Investigation of variation landscape of loci for Body Mass Index in children. Lead Investigator: Bahram Namjou (CCHMC/BCH)
    e. PheWAS in EHR datasets. Lead Investigator: Josh Denny (VU)

f. Using publicly available controls for GWAS studiesVendor Perspectives. Lead Investigator: Brackie Mitchell (External Collaborator)
g. Review of eMERGE progress in genomics – first 6 years. Lead Investigator: Dana Crawford (VU)
h. Replication of metabolic phenotypes from EHR data using the CardioMetabochip. Lead Investigator: Glenn Gerhard (Geisinger)
i. Analysis pipeline for the epistasis search – statistical versus biological filtering. Lead Investigator: Robert Elston (Geisinger)
j. Genetic risk prediction. Lead Investigator: Steve Schrodi (MC/EIRH/PSU)
k. The struggle to find reliable results in exome sequencing data. Lead Investigator: Zubin Patel (External)
l. EMR-linked CNV: Meta analysis of copy number variants across the emerge network. Lead Investigator: Patrick Sleiman (CHOP)
m. EMR-linked LoF: Assessing the functional consequence of loss of function variants using the electronic medical record. Lead Investigator: Patrick Sleiman (CHOP)
n. EMR-linked framework for assessing drug-genome interactions. Lead Investigator: Berta Castillo (CHOP)

3. Big Data Needs in Clinical Genomics: Use Cases Defined by Medical Providers. Lead Investigator: Bryan Weichelt (Marshfield)

4. Practical Approaches to the Omic Chasm. Lead Investigator: Justin Starren (NU)

5. PheKB.org: An Online Collaboration Tool for Phenotype Algorithm Development and Sharing. Lead Investigator: Jacqueline Kirby (CC)

6. A Rigorous Algorithm to Detect and Remove Inaccurate Height Measures within EHRs. Lead Investigator: Arun Muthalagu (NU)

7. Replication of SCN5A Associations with Electrocardiographic Traits in African Americans from Clinical and Epidemiologic Studies. Lead Investigator: Janina Jeff (VU)

Published/Accepted Phase II Site Specific Manuscripts

CHOP

1. Hamshere ML, Langley K, Martin J, Agha SS, Stergiakouli E, Anney RJL, et al. High loading of polygenic risk for ADHD in children with comorbid aggression. Am J Psychiatry. 2013 Aug 1;170(8):909–916. **PMID: 23599091**

2. Monda KL, Chen GK, Taylor KC, Palmer C, Edwards TL, Lange LA, et al. A meta-analysis identifies new loci associated with body mass index in individuals of African ancestry. Nat. Genet. 2013 Jun;45(6):690–696. **PMID: 23583978; PMCID: PMC3694490**

Geisinger

1. Sun X, Elston R, Morris N, Zhu X. What Is the Significance of Difference in Phenotypic Variability across SNP Genotypes? Am. J. Hum. Genet. 2013 Jul 31; **PMID: 23910463; PMCID: PMC3738833**

2. Jones GT, Bown MJ, Gretarsdottir S, Romaine SPR, Helgadottir A, Yu G, et al. A sequence variant associated with sortilin-1 (SORT1) on 1p13.3 is independently associated with abdominal aortic aneurysm. Hum. Mol. Genet. 2013 Apr 9; **PMID: 23535823; PMCID: PMC3690970**

GroupHealth/UW

1. Lambert J-C, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Extended meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nature Genetics. 2013

(Accepted)

2. Larson EB. Building trust in the power of “big data” research to serve the public good. JAMA. 2013 Jun 19;309(23):2443–2444. **PMID: 23780455**

3. Reitz C, Tosto G, Vardarajan B, Rogaeva E, Ghani M, Rogers RS, et al. Independent and epistatic effects of variants in VPS10-d receptors on Alzheimer disease risk and processing of the amyloid precursor protein (APP). Transl Psychiatry. 2013;3:e256. **PMID: 23673467; PMCID: PMC3669917**

4. Miyashita A, Koike A, Jun G, Wang L-S, Takahashi S, Matsubara E, et al. SORL1 is genetically associated with late-onset Alzheimer’s disease in Japanese, Koreans and Caucasians. PLoS ONE. 2013;8(4):e58618. **PMID: 23565137; PMCID: PMC3614978**

Marshfield/Essentia/PSU

1. Pendergrass SA, Verma SS, Holzinger ER, Moore CB, Wallace J, Dudek SM, et al. Next-generation analysis of cataracts: determining knowledge driven gene-gene interactions using biofilter, and gene-environment interactions using the phenx toolkit. Pac Symp Biocomput. 2013;147–158. **PMID: 23424120; PMCID: PMC3615413**

Mayo

1. Jouni H, Shameer K, Hazin R, Asmann YW, de Andrade M, Jull IJ. Clinical Correlates of Autosomal Chromosal Abnormalities in an Electronic Medical Record-Linked Genome-wide Association Study. J. Invest Med HICR. 2013 (Accepted)

2. Ludvigsson JF, Pathak J, Murphy S, Durski M, Kirsch PS, Chute CG, et al. Use of computerized algorithm to identify individuals in need of testing for celiac disease. J Am Med Inform Assoc. 2013 Aug 16; **PMID: 23956016**

3. Ridgeway JL, Han LC, Olson JE, Lackore KA, Koenig BA, Beebe TJ, et al. Potential Bias in the Bank: What Distinguishes Refusers, Nonresponders and Participants in a Clinic-Based Biobank? Public Health Genomics. 2013;16(3):118–126. **PMID: 23595106**

Northwestern

1. Smith ME, Aufox S. Biobanking: The Melding of Research with Clinical Care. Curr Genet Med Rep. 2013 Jun 1];1(2):122–128.

2. Chisholm RL. The opportunities and challenges of implementing genomics-informed personalized medicine. Clin. Pharmacol. Ther. 2013 Aug;94(2):181–182. **PMID: 23872829**

Vanderbilt

1. Wei W-Q, Cronin RM, Xu H, Lasko TA, Bastarache L, Denny JC. Development and evaluation of an ensemble resource linking medications to their indications. J Am Med Inform Assoc. 2013 Sep 1;20(5):954–961. **PMID: 23576672**

2. Rosenbloom ST, Madison JL, Brothers KB, Bowton EA, Clayton EW, Malin BA, et al. Ethical and practical challenges to studying patients who opt out of large-scale biorepository research. J Am Med Inform Assoc. 2013 Jul 25; **PMID: 23886923**

3. Brothers KB, Westbrook MJ, Wright MF, Myers JA, Morrison DR, Madison JL, et al. Patient awareness and approval for an opt-out genomic biorepository. Personalized Medicine [Internet]. 2013 Jun;10(4):349–359.

4. Oetjens MT, Denny JC, Ritchie MD, Gillani NB, Richardson DM, Restrepo NA, et al. Assessment of a pharmacogenomic marker panel in a polypharmacy population identified from electronic medical records.

Pharmacogenomics. 2013 May;14(7):735–744. **PMID: 23651022; PMCID: PMC3725600**

eMERGE Phase II Publications from June 2011 -October 2013

*denotes manuscripts published or accepted between April –October 2013
#denotes manuscripts developed between April –October 2013

Published/Accepted Phase II Network Manuscripts

1. *Integration of Genomics into the Electronic Health Record Special Issue. GENET MED. 2013 Oct 15(10).
    a. Williams M, Kannry J. Introduction to the Special Issue: Integration of Genomics into the Electronic Health Record.
    b. Gottesman O, Kuivaniemi H, Tromp G, Faucett WA, Li R, Manolio TA, et al. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future; **PMID: 23743551**
    c. Kho AN, Rasmussen LV, Connolly JJ, Peissig PL, Starren JB, et al. Practical challenges integrating genomic data into the electronic health record.
    d. Ury A. Storing and Interpreting Genomic Information in Widely Deployed Electronic Health Records Systems; **PMID: 23949573**
    e. Marsolo K, Spooner A. Clinical Genomics in the World of the Electronic Health Record; **PMID: 23846403**
    f. Hartzler A, McCarty C, Brilliant M, Clayton E, Faucett A, et al. Stakeholder engagement: a key component of integrating genomic information into electronic health records; **PMID: 24030437**
    g. Chute C, Wood G, Ullman-Cullere M. Some experiences and opportunities for big data in translational research; **PMID: 24008998**
    h. Iftikhar K, Hazin R, Clayton E, Brothers K, Williams M. Ethical, Legal and Social Implications of Incorporating Genomic Information into Electronic Health Records; **PMID: 24030434**
    i. Overby C. Opportunities for Genomic Clinical Decision Support Interventions.
    j. Tarczy-Hornoch P. A survey of informatics approaches to whole exome and whole genome clinical reporting in the electronic medical record.
    k. Peterson J, Denny J. Electronic Health Record Design and Implementation for Pharmacogenomics: a Local Perspective; **PMID: 24009000**
    l. Kannry J, Williams M. The Undiscovered Country: The Future of Integrating Genomic Information into the EHR.

2. *Gottesman O, Kuivaniemi H, Tromp G, Faucett WA, Li R, Manolio TA, et al. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. Genet. Med. 2013 Jun 6; **PMID: 23743551**

3. *Chute CG, Kohane IS. Genomic medicine, health information technology, and patient care. JAMA. 2013 Apr 10;309(14):1467–1468. **PMID: 23571583**

4. *Wei W-Q, Leibson CL, Ransom JE, Kho AN, Chute CG. The absence of longitudinal data limits the accuracy of high-throughput clinical phenotyping for identifying type 2 diabetes mellitus subjects. Int J Med Inform. 2013 Apr;82(4):239–247. **PMID: 22762862; PMCID: PMC3478423**

5. Jeff JM, Ritchie MD, Denny JC, Kho AN, Ramirez AH, Crosslin D, et al. Generalization of Variants Identified by Genome-Wide Association Studies for Electrocardiographic Traits in African Americans. Ann. Hum. Genet. 2013 Mar 28; **PMID: 23534349; PMCID: PMC3743946**

6. Starren J, Williams MS, Bottinger EP. Crossing the Omic Chasm: A Time for Omic Ancillary Systems. JAMA. 2013 Mar 14;1–2. **PMID: 23494000**

7. Crosslin DR, McDavid A, Weston N, Zheng X, Hart E, de Andrade M, et al. Genetic variation associated with circulating monocyte count in the eMERGE Network. Hum. Mol. Genet. 2013 Jan 12; **PMID: 23314186; PMCID: PMC3633369**

8. McDavid A, Crane PK, Newton KM, Crosslin DR, McCormick W, Weston N, et al. Enhancing the Power of Genetic Association Studies through the Use of Silver Standard Cases Derived from Electronic Medical Records. PLoS ONE. 2013;8(6):e63481. **PMID: 23762230; PMCID: PMC3677889**

9. Kullo IJ, Jarvik GP, Manolio TA, Williams MS, Roden DM. Leveraging the electronic health record to implement genomic medicine. Genet. Med. 2012 Sep 27; **PMID: 23018749**

In Process Phase II Network Manuscripts

1. # Developing Consents for Returning Pharmacogenomics Results: The eMERGE Experience. Lead Investigator:

Maureen Smith (NU)

2. #Genomics Workgroup's Frontiers in Genetics Special Issue
    a. Imputation and QC for combining genome-wide datasets. Lead Investigators: Marylyn Ritchie (MC/EIRH/PSU & CC)
    b. Evaluation of population stratification in large biobanks linked to EHR. Lead Investigator: David Crosslin (GroupHealth/UW)
    c. State of the state in returning genomics research results. Lead Investigator: Iftikhar Kullo (Mayo)
    d. EMR-linked GWAS study: Investigation of variation landscape of loci for Body Mass Index in children. Lead Investigator: Bahram Namjou (CCHMC/BCH)
    e. PheWAS in EHR datasets. Lead Investigator: Josh Denny (VU)
    f. Using publicly available controls for GWAS studiesVendor Perspectives. Lead Investigator: Brackie Mitchell (External Collaborator)
    g. Review of eMERGE progress in genomics – first 6 years. Lead Investigator: Dana Crawford (VU)
    h. Replication of metabolic phenotypes from EHR data using the CardioMetabochip. Lead Investigator: Glenn Gerhard (Geisinger)
    i. Analysis pipeline for the epistasis search – statistical versus biological filtering. Lead Investigator: Robert Elston (Geisinger)
    j. Genetic risk prediction. Lead Investigator: Steve Schrodi (MC/EIRH/PSU)
    k. The struggle to find reliable results in exome sequencing data. Lead Investigator: Zubin Patel (External)
    l. EMR-linked CNV: Meta analysis of copy number variants across the emerge network. Lead Investigator: Patrick Sleiman (CHOP)
    m. EMR-linked LoF: Assessing the functional consequence of loss of function variants using the electronic medical record. Lead Investigator: Patrick Sleiman (CHOP)
    n. EMR-linked framework for assessing drug-genome interactions. Lead Investigator: Berta Castillo (CHOP)

3. #Big Data Needs in Clinical Genomics: Use Cases Defined by Medical Providers. Lead Investigator: Bryan Weichelt (Marshfield)

4. #Practical Approaches to the Omic Chasm. Lead Investigator: Justin Starren (NU)

5. #PheKB.org: An Online Collaboration Tool for Phenotype Algorithm Development and Sharing. Lead Investigator: Jacqueline Kirby (CC)

6. #A Rigorous Algorithm to Detect and Remove Inaccurate Height Measures within EHRs. Lead Investigator: Arun Muthalagu (NU)

7. #Replication of SCN5A Associations with Electrocardiographic Traits in African Americans from Clinical and Epidemiologic Studies. Lead Investigator: Janina Jeff (VU)

8. Seeking Informed Consent for the Inclusion of Samples from Children and Adolescents in Biorepositories: Practical Approaches and Model Language. Lead Investigator: Kyle Brothers (VU)

9. Mechanistic Phenotypes: An Aggregative Phenotyping Strategy to Identify Disease Mechanisms Using GWAS Data. Lead Investigator: Jonathan Mosley (VU)

10. PCA Loadings. Lead Investigator: Gerard Tromp (Geisinger) and David Crosslin (UW)

11. Portable Applications for Implementing Multi-Site Clinical NLP Algorithms. Lead Investigator: David Carrell (GroupHealth)

12. Evaluation of a Secure Multiparty Computation Protocol to Enable Genome-Phenome Meta-Analysis in the Cloud. Lead Investigator: Wei Xie (VU)

13. Admixture Mapping and Subsequent Fine Mapping Reveals Novel Loci for Type 2 Diabetes in African Americans. Lead Investigator: Janina Jeff (VU)

14. Knowledge driven search for gene-gene interactions associated with cataract in the eMERGE network. Lead Investigator: Marylyn Ritchie (CC)

15. Evaluation of a Differentially Private Top-k SNP Publication Strategy. Lead Investigator: Mehmet Kuzu (External

Collaborator); Brad Malin (VU)

16. Effective Use of Electronic Health Records to Identify Venous Thromboembolism: Results from the eMERGE Network. Lead Investigator: Jyoti Pathak (Mayo)

17. PheWAS Analysis of the GWAS Catalog: Finding New Disease Relationships for Previously Published Genetic Associations. Lead Investigator: Josh Denny (VU)

18. GWAS of Infection or Colonization with Community Associated Methicillin-Resistant Staphylococcus Aureus (CA-MRSA). Lead Investigator: Abel Kho (NU)

19. MEDIA Study Collaboration. Lead Investigator: Geoff Hayes (NU)

20. A Trial of Pre-emptive Pharmacogenetic Genotyping: The PGx Project of the eMERGE Network. Lead Investigator: Laura Rasmussen-Torvik (NU)

21. Genetic Risk Factors for Development of Diverticulitis. Lead Investigator: Abel Kho (NU)

22. Diverticulosis. Lead Investigator: Will Thompson (NU)

23. The Geographic Distribution of Colon Polyps. Lead Investigator: Will Thompson (NU)

24. Colon Polyps. Lead Investigator: Abel Kho (NU)

25. Genome-wide Association Study of Extreme Obesity Defined by Electronic Medical Record Phenotyping. Lead Investigator: Glenn Gerhard (Geisinger)

26. Design Patterns for the Development and Validation of Phenotype Algorithms. Lead Investigator: Luke Rasmussen (NU)

27. A Collaborative Approach to Develop an Electronic Health Record Phenotyping Algorithm for Drug-Induced Liver Injury. Lead Investigator: Casey Overby and Chunhua Weng (Columbia)

28. Genetic Risk Scores for Complex Diseases in the eMERGE Network: Characterization and Predicative Abilities in Clinical Settings. Lead Investigator: Logan Dumitrescu (VU)

29. Genetic Variants Associated with Response to Heart Failure Treatment: The Electronic Medical Records and Genomics (eMERGE) Network. Lead Investigator: Sue Bielinski (Mayo)

30. Genome Wide Association of Risk of Heart Failure: The Electronic Medical Records and Genomics (eMERGE) Network. Lead Investigator: Sue Bielinski (Mayo)

31. Using Electronic Health Records to Identify Heart Failure Cohorts with Differentiation for Preserved and Reduced Ejection Fraction and Assessment of Treatment Response. Lead Investigator: Sue Bielinski (Mayo)

32. Penetrance of Hemochromatosis (HFE) RS1799945 (H63D) and RS1800562 (C282Y) Homozygosity and Compound Heterozygosity. Lead Investigators: Carlos Gallego (UW), Daniel Kim (UW), Josh Denny (VU), and Maureen Smith (NU)

33. Meta-Analysis of Genome-Wide Association Studies for Abdominal Aortic Aneurysms. Lead Investigator: Greg Jones (External Collaborator) and Helena Kuivaniemi (Geisinger)

34. Genetic Variation that Predicts Susceptibility to Clostridium. Lead Investigator: Josh Denny (VU)

35. Genetic Variation that Predicts Susceptibility to Herpes Zoster. Lead Investigator: David Crosslin (UW)

36. GWAS of Venous Thromboembolism (VTE) among White Americans. Lead Investigator: John Heit (Mayo)

37. GWAS of Venous Thromboembolism (VTE) among African-Americans. Lead Investigator: John Heit (Mayo)

38. Genome-Wide Association Study of Abdominal Aortic Aneurysms with Electronic Medical Record Phenotyping. Lead Investigators: Helena Kuivaniemi and Gerard Tromp (Geisinger)

Published/Accepted Phase II Site Specific Manuscripts

CHOP

1. *Hamshere ML, Langley K, Martin J, Agha SS, Stergiakouli E, Anney RJL, et al. High loading of polygenic risk for ADHD in children with comorbid aggression. Am J Psychiatry. 2013 Aug 1;170(8):909–916. **PMID: 23599091**

2. *Monda KL, Chen GK, Taylor KC, Palmer C, Edwards TL, Lange LA, et al. A meta-analysis identifies new loci associated with body mass index in individuals of African ancestry. Nat. Genet. 2013 Jun;45(6):690–696. **PMID: 23583978; PMCID: PMC3694490**

3. Mitchell JA, Hakonarson H, Rebbeck TR, Grant SFA. Obesity-susceptibility loci and the tails of the pediatric BMI distribution. Obesity (Silver Spring). 2013 Feb 14; **PMID: 23408508; PMCID: PMC3661695**

4. Shi L, Zhang X, Golhar R, Otieno FG, He M, Hou C, et al. Whole-genome sequencing in an autism multiplex family. Mol Autism. 2013;4(1):8. **PMID: 23597238; PMCID: PMC3642023**

5. March ME, Sleiman PM, Hakonarson H. Genetic polymorphisms and associated susceptibility to asthma. Int J Gen Med. 2013;6:253–265. **PMID: 23637549; PMCID: PMC3636804**

6. Glessner JT, Li J, Hakonarson H. ParseCNV integrative copy number variation association software with quality tracking. Nucleic Acids Res. 2013 Jan 4; **PMID: 23293001; PMCID: PMC3597648**

7. Guo Y, Lanktree MB, Taylor KC, Hakonarson H, Lange LA, Keating BJ. Gene-centric meta-analyses of 108 912 individuals confirm known body mass index loci and reveal three novel signals. Hum. Mol. Genet. 2013 Jan 1;22(1):184–201. **PMID: 23001569; PMCID: PMC3522401**

8. Asselbergs FW, Guo Y, van Iperen EPA, Sivapalaratnam S, Tragante V, Lanktree MB, et al. Large-scale gene-centric meta-analysis across 32 studies identifies multiple lipid loci. Am. J. Hum. Genet. 2012 Nov 2;91(5):823–838. **PMID: 23063622; PMCID: PMC3487124**

Geisinger

1. *Sun X, Elston R, Morris N, Zhu X. What Is the Significance of Difference in Phenotypic Variability across SNP Genotypes? Am. J. Hum. Genet. 2013 Jul 31; **PMID: 23910463; PMCID: PMC3738833**

2. *Jones GT, Bown MJ, Gretarsdottir S, Romaine SPR, Helgadottir A, Yu G, et al. A sequence variant associated with sortilin-1 (SORT1) on 1p13.3 is independently associated with abdominal aortic aneurysm. Hum. Mol. Genet. 2013 Apr 9; **PMID: 23535823; PMCID: PMC3690970**

3. Wang X, Morris NJ, Zhu X, Elston RC. A variance component based multi-marker association test using family and unrelated data. BMC Genet. 2013;14:17. **PMID: 23497289; PMCID: PMC3614458**

4. Zhu X, Feng T, Elston RC. Linkage-disequilibrium-based binning misleads the interpretation of genome-wide association studies. Am. J. Hum. Genet. 2012 Nov 2;91(5):965–968; author reply 969–970. **PMID: 23122590; PMCID: PMC3487138**

5. Lu Q, Wei C, Ye C, Li M, Elston RC. A likelihood ratio-based Mann-Whitney approach finds novel replicable joint gene action for type 2 diabetes. Genet. Epidemiol. 2012 Sep;36(6):583–593. **PMID: 22760990; PMCID: PMC3634342**

6. Wang X, Morris NJ, Schaid DJ, Elston RC. Power of single- vs. multi-marker tests of association. Genet. Epidemiol. 2012 Jul;36(5):480–487. **PMID: 22648939; PMCID: PMC3708310**

GroupHealth/UW

1. *Lambert J-C, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Extended meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nature Genetics. 2013 (Accepted)

2. *Larson EB. Building trust in the power of “big data” research to serve the public good. JAMA. 2013 Jun 19;309(23):2443–2444. **PMID: 23780455**

3. *Reitz C, Tosto G, Vardarajan B, Rogaeva E, Ghani M, Rogers RS, et al. Independent and epistatic effects of variants in VPS10-d receptors on Alzheimer disease risk and processing of the amyloid precursor protein (APP). Transl Psychiatry. 2013;3:e256. **PMID: 23673467; PMCID: PMC3669917**

4. *Miyashita A, Koike A, Jun G, Wang L-S, Takahashi S, Matsubara E, et al. SORL1 is genetically associated with late-onset Alzheimer’s disease in Japanese, Koreans and Caucasians. PLoS ONE. 2013;8(4):e58618. **PMID: 23565137; PMCID: PMC3614978**

5. Whitcomb DC, LaRusch J, Krasinskas AM, Klei L, Smith JP, Brand RE, et al. Common genetic variants in the CLDN2 and PRSS1-PRSS2 loci alter risk for alcohol-related and sporadic pancreatitis. Nat. Genet. 2012 Dec;44(12):1349–1354. **PMID: 23143602; PMCID: PMC3510344**

6. Coppola G, Chinnathambi S, Lee JJ, Dombroski BA, Baker MC, Soto-Ortolaza AI, et al. Evidence for a role of the rare p.A152T variant in MAPT in increasing the risk for FTD-spectrum and Alzheimer’s diseases. Hum. Mol. Genet. 2012 Aug 1;21(15):3500–3512. **PMID: 22556362; PMCID: PMC3392107**

7. Allen M, Zou F, Chai HS, Younkin CS, Crook J, Pankratz VS, et al. Novel late-onset Alzheimer disease loci variants associate with brain gene expression. Neurology. 2012 Jul 17;79(3):221–228. **PMID: 22722634; PMCID: PMC3398432**

8. Floyd JS, Heckbert SR, Weiss NS, Carrell DS, Psaty BM. Use of administrative data to estimate the incidence of statin-related rhabdomyolysis. JAMA. 2012 Apr 18;307(15):1580–1582. **PMID: 22511681; PMCID: PMC3483067**

9. Jun G, Naj AC, Beecham GW, Wang L-S, Buros J, Gallins PJ, et al. Meta-analysis confirms CR1, CLU, and PICALM as alzheimer disease risk loci and reveals interactions with APOE genotypes. Arch. Neurol. 2010 Dec;67(12):1473–1484. **PMID: 20697030; PMCID: PMC3048805**

Marshfield/Essentia/PSU

1. *Pendergrass SA, Verma SS, Holzinger ER, Moore CB, Wallace J, Dudek SM, et al. Next-generation analysis of cataracts: determining knowledge driven gene-gene interactions using biofilter, and gene-environment interactions using the phenx toolkit. Pac Symp Biocomput. 2013;147–158. **PMID: 23424120; PMCID: PMC3615413**

2. Foth W, Waudby C, Brilliant MH. Certificates of confidentiality and the Marshfield Clinic’s Personalized Medicine Research Project. Virtual Mentor. 2012;14(8):653–656. **PMID: 23351322**

Mayo

1. *Jouni H, Shameer K, Hazin R, Asmann YW, de Andrade M, Jull IJ. Clinical Correlates of Autosomal Chromosal Abnormalities in an Electronic Medical Record-Linked Genome-wide Association Study. J. Invest Med HICR. 2013 (Accepted)

2. *Ludvigsson JF, Pathak J, Murphy S, Durski M, Kirsch PS, Chute CG, et al. Use of computerized algorithm to identify individuals in need of testing for celiac disease. J Am Med Inform Assoc. 2013 Aug 16; **PMID: 23956016**

3. *Ridgeway JL, Han LC, Olson JE, Lackore KA, Koenig BA, Beebe TJ, et al. Potential Bias in the Bank: What Distinguishes Refusers, Nonresponders and Participants in a Clinic-Based Biobank? Public Health Genomics. 2013;16(3):118–126. **PMID: 23595106**

4. Wooten EC, Hebl VB, Wolf MJ, Greytak SR, Orr NM, Draper I, et al. Formin homology 2 domain containing 3 variants associated with hypertrophic cardiomyopathy. Circ Cardiovasc Genet. 2013 Feb;6(1):10–18. PMID: 23255317; PMCID: PMC3578062

5. Seedorff M, Peterson KJ, Nelsen LA, Cocos C, McCormick JB, Chute CG, et al. Incorporating expert terminology and disease risk factors into consumer health vocabularies. Pac Symp Biocomput. 2013;421–432. **PMID: 23424146; PMCID: PMC3587774**

6. McCormick J. Whole Genome Sequencing. Lahey Clinic Medical Ethics Journal. 2013; 20(1): 1-2.

6. Rea S, Bailey KR, Pathak J, Haug PJ. Bias in Recording of Body Mass Index Data in the Electronic Health Record. American Medical Informatics Association (AMIA) Clinical Research Informatics Summit, 2013.

7. Zhu Q, Freimuth RR, Pathak J, Chute CG. Using Clinical Element Models for Pharmacogenomic Study Data Standardization. American Medical Informatics Association (AMIA) Clinical Research Informatics Summit, 2013.

8. Li D, Simon G, Chute CG, Pathak J. Using Association Rule Mining for Phenotype Extraction from Electronic Health Records. American Medical Informatics Association (AMIA) Clinical Research Informatics Summit, 2013.

10. Kiefer RC, Freimuth RR, Chute CG, Pathak J. Mining Genotype-Phenotype Associations from Public Knowledge Sources via Semantic Web Querying. American Medical Informatics Association (AMIA) Clinical Research Informatics Summit, 2013

11. Tao C, Jiang G, Oniki TA, Freimuth RR, Zhu Q, Sharma D, et al. A semantic-web oriented representation of the clinical element model for secondary use of electronic health records data. J Am Med Inform Assoc. 2012 Dec 25; **PMID: 23268487; PMCID: PMC3628064**

12. Zhu Q, Freimuth RR, Lian Z, Bauer S, Pathak J, Tao C, et al. Harmonization and semantic annotation of data dictionaries from the Pharmacogenomics Research Network: A case study. J Biomed Inform. 2012 Nov 29; **PMID: 23201637; PMCID: PMC3606279**

13. Tao C, Pathak J, Solbrig HR, Wei W-Q, Chute CG. Terminology representation guidelines for biomedical ontologies in the semantic web notations. J Biomed Inform. 2012 Sep 28; **PMID: 23026232; PMCID: PMC3563768**

14. Clifford L, Singh A, Wilson GA, Toy P, Gajic O, Malinchoc M, et al. Electronic health record surveillance algorithms facilitate the detection of transfusion-related pulmonary complications. Transfusion. 2012 Aug 31; **PMID: 22934792**

15. Pathak J, Kiefer RC, Chute CG. Using Semantic Web Technologies for Cohort Identification from Electronic Health Records for Clinical Research. AMIA Summits Transl Sci Proc [Internet]. 2012 Mar 19 [cited 2012 Oct 1];2012:10–19. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3392057/ **PMID: 22779040; PMCID: PMC3392057**

16. Pathak J, Kiefer RC, Bielinski SJ, Chute CG. Mining the Human Phenome using Semantic Web Technologies: A Case Study for Type 2 Diabetes. AMIA Annu Symp Proc. 2012;2012:699–708. **PMID: 23304343; PMCID: PMC3540447**

17. Pathak J, Kiefer RC, Bielinski SJ, Chute CG. Applying semantic web technologies for phenome-wide scan using an electronic health record linked Biobank. J Biomed Semantics. 2012;3(1):10. **PMID: 23244446; PMCID: PMC3554594**

18. Pathak J, Kiefer R, Freimuth R, Chute C. Validation and discovery of genotype-phenotype associations in chronic diseases using linked data. Stud Health Technol Inform. 2012;180:549–553. **PMID: 22874251**

19. Li D, Endle CM, Murthy S, Stancl C, Suesse D, Sottara D, et al. Modeling and Executing Electronic Health Records Driven Phenotyping Algorithms using the NQF Quality Data Model and JBoss® Drools Engine. AMIA Annu Symp Proc. 2012;2012:532–541. **PMID: 23304325; PMCID: PMC3540464**

Northwestern

1. *Smith ME, Aufox S. Biobanking: The Melding of Research with Clinical Care. Curr Genet Med Rep. 2013 Jun 1];1(2):122–128.

2. *Chisholm RL. The opportunities and challenges of implementing genomics-informed personalized medicine. Clin. Pharmacol. Ther. 2013 Aug;94(2):181–182. **PMID: 23872829**

3. Thompson WK, Rasmussen LV, Pacheco JA, Peissig PL, Denny JC, Kho AN, et al. An Evaluation of the NQF Quality Data Model for Representing Electronic Health Record Driven Phenotyping Algorithms. AMIA Annu Symp Proc. 2012;2012:911–920. **PMID: 23304366; PMCID: PMC3540514**

Vanderbilt

1. *Wei W-Q, Cronin RM, Xu H, Lasko TA, Bastarache L, Denny JC. Development and evaluation of an ensemble resource linking medications to their indications. J Am Med Inform Assoc. 2013 Sep 1;20(5):954–961. **PMID:**

23576672

2. *Rosenbloom ST, Madison JL, Brothers KB, Bowton EA, Clayton EW, Malin BA, et al. Ethical and practical challenges to studying patients who opt out of large-scale biorepository research. J Am Med Inform Assoc. 2013 Jul 25; **PMID: 23886923**

3. *Brothers KB, Westbrook MJ, Wright MF, Myers JA, Morrison DR, Madison JL, et al. Patient awareness and approval for an opt-out genomic biorepository. Personalized Medicine [Internet]. 2013 Jun;10(4):349–359.

4. *Oetjens MT, Denny JC, Ritchie MD, Gillani NB, Richardson DM, Restrepo NA, et al. Assessment of a pharmacogenomic marker panel in a polypharmacy population identified from electronic medical records. Pharmacogenomics. 2013 May;14(7):735–744. **PMID: 23651022; PMCID: PMC3725600**

5. Altman RB, Clayton EW, Kohane IS, Malin BA, Roden DM. Data re-identification: societal safeguards. Science. 2013 Mar 1;339(6123):1032–1033. PMID: 23449577; PMCID: PMC3740512

6. Atreya RV, Smith JC, McCoy AB, Malin B, Miller RA. Reducing patient re-identification risk for laboratory results within research datasets. J Am Med Inform Assoc. 2013 Jan 1;20(1):95–101. **PMID: 22822040; PMCID: PMC3555327**

7. Xia W, Heatherly R, Ding X, Li J, Malin B. Efficient discovery of de-identification policy options through a risk-utility frontier. Proceedings of the third ACM conference on Data and application security and privacy [Internet]. New York, NY, USA: ACM; 2013 [cited 2013 Jul 31]. p. 59–70. Available from: http://doi.acm.org/10.1145/2435349.2435357

8. Heatherly RD, Loukides G, Denny JC, Haines JL, Roden DM, Malin BA. Enabling genomic-phenomic association discovery without sacrificing anonymity. PLoS ONE. 2013;8(2):e53875. **PMID: 23405076; PMCID: PMC3566194**

9. Westbrook MJ, Wright MF, Van Driest SL, McGregor TL, Denny JC, Zuvich RL, et al. Mapping the incidentalome: estimating incidental findings generated through clinical pharmacogenomics testing. Genet. Med. 2012 Nov 29; **PMID: 23196672; PMCID: PMC3648626**

10. Wilke RA, Ramsey LB, Johnson SG, Maxwell WD, McLeod HL, Voora D, et al. The clinical pharmacogenomics implementation consortium: CPIC guideline for SLCO1B1 and simvastatin-induced myopathy. Clin. Pharmacol. Ther. 2012 Jul;92(1):112–117. **PMID: 22617227; PMCID: PMC3384438**

11. Tamersoy A, Loukides G, Nergiz ME, Saygin Y, Malin B. Anonymization of Longitudinal Electronic Medical Records. IEEE Transactions on Information Technology in Biomedicine [Internet]. 2012 May;16(3):413 –423. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22287248 **PMID: 22287248**

12. Clayton EW. Sharing individual research results with biospecimen contributors: counterpoint. Cancer Epidemiol. Biomarkers Prev. 2012 Feb;21(2):260–261. **PMID: 22313940**

13. Schildcrout JS, Denny JC, Bowton E, Gregg W, Pulley JM, Basford MA, et al. Optimizing Drug Outcomes Through Pharmacogenetics: A Case for Preemptive Genotyping. Clinical Pharmacology & Therapeutics [Internet]. 2012 [cited 2012 Oct 1];92(2):235–242. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22739144 **PMID: 22739144**

14. Pulley JM, Denny JC, Peterson JF, Bernard GR, Vnencak-Jones CL, Ramirez AH, et al. Operational Implementation of Prospective Genotyping for Personalized Medicine: The Design of the Vanderbilt PREDICT Project. Clinical Pharmacology & Therapeutics [Internet]. 2012 [cited 2012 Oct 1];92(1):87–95. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22588608 **PMID: 22588608; PMCID: PMC3581305**

Meeting Summary

eMERGE Network – ESP Teleconference

Executive Session – 4/19/2013

ESP	Eta Berner (UAB)	**NHGRI**	Rongling Li
	Gerardo Heiss (UNC)		Ian Marpuri
	Stan Huff (Intermountain Healthcare)		Teri Manolio
	Howard McLeod (UNC, Chair)		
	Lisa Parker (Pittsburgh)		

The External Scientific Panel (ESP) met with members of NHGRI staff in Executive Session after the ESP teleconference held on April 19, 2013.

- The ESP was pleased with how enthusiastically the Network responded to the ESP's recommendations from the last joint ESP-Steering Committee meeting.
- The ESP appreciated the quality of the packet for this teleconference.
- The ESP recognized the Network's increased productivity on publications and recommended the Network to emphasize disseminating products such as methods, tools and software for the public. It also wanted the Network to evaluate the utility of these products by tracking the number of users and soliciting comments from these users. This information could subsequently help improve Network tools. The ESP also wanted the CC to track the developmental stages of eMERGE tools.
- For the eMERGE-PGx project, the ESP noted that the Network seems to have a very large number of goals. NHGRI staff thought this list of goals would be refined once implementation began. The ESP acknowledged that clinical outcome metrics would be useful for the field; however, the Network will likely have the greatest impact at this early stage if it ensures it that the projects provide better understanding of the implementation process. Useful data could be collected to cover the following aspects of implementation: clinician training methods, time and resources expended on training, methods and staff involved for communication of information to patients, patient recruitment and eligibility, and physician and patient feedback. The guidelines, protocols, measures of performance, logs, and evaluation tools used in the implementation of the various goals could be consolidated into a searchable database to help investigators identify features and indicators associated with successful implementation.
- The ESP recommended that the sites should try to build on the commonalities among the different genomic medicine pilot projects. Although having a wide range of projects allows the Network to learn more, the investigators should try to leverage the power of the consortium to get more out of the projects. For example, the Network investigators could collaborate to develop improved instruments to capture patient and physician reactions regarding return of results. Use of the same instrument enables comparison across sites, helps avoid duplicative efforts to draft and validate instruments at each site, and captures the true feelings of patients and physicians at each site.

- The ESP thought that eMERGE could respond to the ACMG guidelines on returning results based on its experiences.
- The ESP suggested that the working group reports need to better outline all of the activities and accomplishments that each group is involved in. The reports seemed uneven when compared to one another. This needn't result in longer reports, but concise information on why, when, how, and where can be detailed.

ESP Recommendations:

1) The Network should continue to disseminate products, especially methods, tools and software for the public.
2) The Network should better track usage of its tools and solicit comments from users.
3) The CC should track the developmental stages of eMERGE tools.
4) The eMERGE-PGx project should focus more on better understanding the implementation process than on measuring clinical outcomes.
5) The Network should find ways to use common instruments or measurements among the genomic medicine projects.
6) The working group reports should better detail progress on activities and accomplishments within the group.

Meeting Summary
eMERGE Network External Scientific Panel
Conference Call – 4/19/2013

Attendance

ESP Attendees:
University of Alabama-Birmingham: Eta Berner; **Univeristy of North Carolina**: Gerardo Heiss, Howard McLeod; **University of Pittsburg**h: Lisa Parker; **InterMountain Healthcare**: Stan Huff
Network Attendees:
CCHMC/BCH: John Harley**; CHOP:** Hakon Hakonarson; **Geisinger:** Marc Williams**; GroupHealth/UW:** Gail Jarvik; **Marshfield/Essentia/PSU**: Murray Brilliant, Cathy McCarty; **Mayo**: Iftikhar Kullo; **Northwestern**: Rex Chisholm; **Vanderbilt**: Dan Roden; **NHGRI**: Rongling Li, Ian Marpuri, Teri Manolio; **CC**: Melissa Basford, Jonathan Haines, Brandy Mapes, Lauren Melancon

Absent: **University of Utah**: Jeff Botkin; **Cleveland Clinic**: Charis Eng; **Mount Sinai: Erwin Bottinger**

Decisions and discussion

Opening Remarks/Welcome (Rongling Li & Howard McLeod)

Rongling welcomed the group and thanked the ESP for their service to the Network. She continued to thank the eMERGE investigators for their hard work and progress. Howard also thanked the ESP members and the Network sites for their serious response to previous ESP recommendations along with great progress.

Network Overview and Response to ESP recommendations (Rex Chisholm)

An overview of past ESP recommendations and the Network's response were presented. Highlights included publications, including a Network citation analysis and two upcoming special issues by the EHRI and Genomics workgroups. The Network continues to publish a large number of papers each year, and the number of publications for 2013 is promising due to the many manuscripts in development. The ESP had recommended that the Network track tools and their implementation by other groups in an effort to share their collective knowledge with the greater scientific community. These tools were presented to the ESP; the tracking of their implementation is still in process and will be reported on in October. The United States Air Force is now an official Affiliate Network member. Other institutions have inquired about obtaining affiliate membership but none have completed the process yet. The eMERGE Network is also actively collaborating with the Return of Results and CSER Consortia and hope to schedule a joint in-person meeting in late 2013.

ESP Response and Discussion

The ESP was pleased to see how seriously their recommendations were being taken, evaluated, and executed by the Network. The ESP encouraged the Network to find effective ways to track people who are using Network-created phenotyping algorithms, genomics tools and pipelines and informatics tools. The CC is tracking website hits and is in the beginning stages of tracking algorithm implementations on PheKB as well as users from various external sites who are posting their own phenotypes. The CC will work to capture a profile of tools from the various eMERGE sites.

eMERGE PGx Outcomes and Discovery

As part of the eMERGE PGx project a subgroup has formed to look specifically at outcomes of interest as defined by Network sites. The objective is to finalize a list of both mandatory and optional outcomes so that data can be pooled to identify trends to inform future studies. Collecting these outcomes will also facilitate a dialog between sites concerning successes and challenges. A preliminary set of metrics and process and clinical outcomes, still under refinement, was presented. The Network will be surveyed to gather information about metrics sites are already collecting, outcomes the sites plan to begin gathering, or are interested in developing, and the anticipated difficulty associated with measuring each outcome. The preliminary list of metric areas includes:

- Predictive Algorithm
- Indication Testing
- Recruitment & Retention
- Lab & Data Processing Metrics
- Assay Quality
- Clinical Validation Metrics
- Pharamcogenomics Test Metrics
- Pharmacogentic Result Reporting – EMR Integration and CDS
- CDS Metrics
- Predictive Algorithm (optional)
- Survey-requiring outcome metrics (optional)
- Clinical Outcomes (optional)

The ESP inquired about collecting clinical outcomes and expressed their concern that the group may be spreading themselves too thin. Marc clarified that after the survey was complete the list would be modified to better suit the needs of the Network/site goals and those listed as mandatory and optional were likely to change. It was also mentioned that some sites were already collecting clinical outcomes and were looking for feedback and suggestions from others. Marc also noted that he is an active member of the Clinical Decision Support Consortium (CDSC). The CDSC is currently working on a pharmacogenomics project of their own and has taken one of eMERGE's PGx use cases to expand the impact of the CPIC guidelines, specifically to translate these guidelines into implementable decision support tools.

Genomic Medicine Projects – ESP Led

Geisinger answered the ESP's question concerning their Genomic Medicine Pilot project with regard to creating synergy between other sites and the psychosocial impact and ethical issues around returning their results. These topics have been discussed internally at Geisinger and focus groups have also been used to gain a better understanding of how these returned results will be received. The ACMG incidental findings report was released, and with that new information Gesinger plans to further engage their providers and program oversight committee along with reengaging focus groups. Members of the Geisinger team are actively involved in the CERC workgroup. This involvement will facilitate cross site collaborations, discussions, and sharing.

Geisinger has found that most individuals participating in their focus groups are interested in having their results returned which is contradictory to other sites in the Network. The ESP agreed that common instruments would be the best way to determine similarities and differences among site populations. Geisinger stated that there are not a large number of validated instruments available hence the election to utilize a qualitative and exploratory approach. The ESP encouraged the group to continue to work on using common instruments to ensure that all sites were comparing similar information. It was mentioned that the CERC workgroup is interested in the topic of patient and

physician responses and is creating a set of questions that can be used by all sites for comparison. Currently, this effort is targeting the PGx project since all sites have this in common and this is an eMERGE funded implementation project.

Mayo discussed their Genomic Medicine pilot project focusing on risk scores. Two arms of the Mayo project lead to a meeting with a genetic counselor. This was proposed in an effort to create a standard format, allowing the site to look at reactions to this information by surveys. Mayo also believes that regardless of patient's results, there will be the opportunity to discuss conventional risk and family history. Mayo is still working through this process in an effort to form an effective plan for patients. Discussions include exploration of whether the person returning the results needs to be a genetic counselor or if an education study coordinator would be appropriate. Either way this interaction would be brief. Risk will be portrayed to patients by a simple pictogram showing 100 people as icons and the risk is depicted by colored icons. The ESP was pleased to hear that this was Mayo's approach to depicting risk as many studies have shown this to be one of the most effective ways to express probability to patients.

Not depicted on the table is a screening arm that will hopefully maximize the number of individuals who will have a meaningful risk change.

General Discussion and Recommendations from the ESP

The ESP had no further questions for the eMERGE investigators.

eMERGE Network Steering Committee Meeting
June 3-4, 2013
Philadelphia, PA

Attendance

Network Members in Attendance

CCHMC/BCH	Beth Cobb
CCHMC/BCH	John Harley
CCHMC/BCH	Ingrid Holm
CCHMC/BCH	Zak Kohane
CCHMC/BCH	Bahram Namjou-Khales
CCHMC/BCH	Cassandra Perry
CCHMC/BCH	Imre Solti
CCHMC/BCH	Wendy Wolf
CHOP	Berta Almoguera
CHOP	John Connolly
CHOP	Hakon Hakonarson
CHOP	Brendan Keating
CHOP	Lyam Vazquez
Geisinger	Ken Borthwick
Geisinger	Andy Faucett
Geisinger	Helena Kuivaniemi
Geisinger	Diane Smelser
Geisinger	Gerard Tromp
Geisinger	Janet Williams
Geisinger	Marc Williams
GHC/U W	David Carrell
GHC/U W	David Crosslin
GHC/U W	Malia Fullerton
GHC/U W	Andrea Hartzler
GHC/U W	Eric Larson
MC/Essentia/PSU	Murray Brilliant
MC/Essentia/PSU	Molly Hall
MC/Essentia/PSU	Terrie Kitchner
MC/Essentia/PSU	Cathy McCarty
MC/Essentia/PSU	Peggy Peissig
MC/Essentia/PSU-CC	Marylyn Ritchie
Mayo	Mariza de Andrade
Mayo	Barbara Koenig
Mayo	Iftikhar Kullo
Mayo	Jennifer McCormick
Mt. Sinai/Columbia	Erwin Bottinger
Mt. Sinai/Columbia	Casey Overby
Mt. Sinai/Columbia	Chunhua Weng
Northwestern	Rex Chisholm
Northwestern	Maureen Smith
Northwestern	Justin Starren
Northwestern	Laura Rasmussen-Torvik
NHGRI	Rongling Li
NHGRI	Teri Manolio
NHGRI	Simona Volpi
Vanderbilt	Ellen Clayton
Vanderbilt	Dana Crawford
Vanderbilt-CC	Josh Denny
Vanderbilt-CC	Brad Malin
Vanderbilt	Josh Peterson
Vanderbilt	Dan Roden
Vanderbilt-CC	Sarah Stallings
CC-MC/Essentia/PSU	Gretta Armstrong
CC	Melissa Basford
CC	Jonathan Haines
CC	Brandy Mapes
CC	Lauren Melancon
CC-MC/Essentia/PSU	Shefali Setia

Affiliate Members

Air Force	Ronald Miller

Network Invitees and Guests

Mayo	John Black
CIDR	Kim Doheny
CIDR	Kurt Hetrick
CIDR	Jane Romm
UC San Diego/iDASH	Lucila Ohno-Machado

Monday, June 3

Full Session

Welcome, Opening Remarks, General Updates – Rongling Li

An NIH update for fiscal year 2013 was provided:

- The federal government is funded through continuing resolution for Fiscal Year 2013. Reductions in NIH's budget have been reduced and the 2013 NHGRI budget will be $483M.

Jon Lorsch was named the new director of the National Institute of General Medical Sciences and NHGRI is actively recruiting Extramural Bioinformatics Program Directors.

Rongling discussed a new NHGRI initiative: Big Data to Knowledge (BD2K). This is a trans-NIH effort with the goal that by the end of the decade, enable a quantum leap in the ability of the research community to maximize the value of the growing volume and complexity of biomedical data. This project will have four programmatic areas:

- Facilitating Broad Use of Biomedical Big Data
- Developing and disseminating analysis methods and software for biomedical big data
- Enhancing training for biomedical big data
- Establishing centers of excellence for biomedical big data

BD2K will be holding a series of workshops, beginning this summer. Funding for this project begins in fiscal year 2014.

Rongling presented the genotyped and imputed sample counts for the Network in Phase I and Phase II to date. Phase I and Phase II phenotypes were also briefly discussed.

Goals for the meeting include:

- Update achievement in the past 22 months of eMERGE-II including: (1) Phenotyping, (2) genotype imputation, (3) Genome-wide association analyses, (4) return of genomic/genetic results, (5) EMR integration, and (6) clinical decision support.
- Identify obstacles and plan to overcome them
- Develop plan for addressing the ESP recommendations at the October's meeting
- Share experience on genomic medicine pilot studies
- Update on PGx projects

Takeaways for the meeting:

- Respond to the ESP recommendations
- Genomic medicine pilots and PGx projects
- Refine the eMERGE publication policy
- Update plan for dissemination of the network-wide lessons learned and research results to the scientific community
- Propose future directions for eMERGE

Resistant Hypertension – Dana Crawford

The Resistant Hypertension project was started in Phase I as a Network-wide project that utilized existing genotype data (like hypothyroidism) that was to be supplemented with new samples and genotyping. The algorithm was created in 2010 and a manuscript was developed as recently as 2012. Issues that remain outstanding for the project include:

- Phenotype definition – not considering controlled hypertensives on two medications as controls and normotensives as controls.
- African American data – sample size was small and a large age difference between cases and controls.

- Imputed data
- Incorporating eMERGE II samples and data.

There were originally two case and control definitions for this phenotype. The original goal was to identify extreme cases and controls. The sample size for Resistant Hypertension was too small to consider control groups with 0, 1, 2, 3, or 4 medications in analysis. The definition of concurrent drug use, medication classes utilized, and the phenotype definition were reviewed. As were original stats, results, and study site contributions for European Americans and African Americans.

The Resistant Hypertension project has been improved and has the following characteristics: RH cases (I/II) vs. controlled hypertensives only, European American only (only 22 AA controls), incorporate new eMERGE II samples, and use imputed data. New site counts reveal that Geisinger only has cases and Mount Sinai is diverse. Imputation on eMERGE I/II was available last week on PSU servers and Dana's group ran tests of association for EA RH cases (I/II) and controlled hypertensives. The preliminary Q-Q plot is abnormal and may be showing a bias, possibly a platform bias. Moving forward, the group favors going back and looking more closely at the controlled hypertensive definition to get more controls, especially for African Americans.

eMERGE PGx Session

Process Outcomes- Survey Results and Discussion - Josh Peterson walked the group through the results of the survey distributed by the Process outcomes subgroup. The group was tasked with developing consensus around process outcomes and assign priority - mandatory vs. optional. Domains included: Recruitment Performance, Sequencing and Validation Performance, Test result metrics, CDS metrics clinical training/education and patient education. Eleven outcomes were identified as optional outcomes while fourteen outcomes were identified as mandatory. This list can be found on the eMERGE website (insert the link here). Some members were concerned about the extra work this will take while others expressed the importance of these measures for the field. Teri suggested that eMERGE partner with CSER concerning the definition of some of these proposed measures. Many agreed that the low hanging fruit would be the quantitative traits. The Genomics and Phenotyping workgroups were encouraged to collaborate and propose specific targets for a data set by year end.

Possible High Impact Use Cases were identified as clopidogrel, warfarin, allopurinol, carbamazepine, and thiopurines. Other potential use cases include genes highlighted in the ACMG "actionable" gene list along with other genes Network members deem interesting. It was noted that eMERGE would be the first to systematically generate data on 6 genes on the ACMG list. This will be a good opportunity. It was suggested that Les Biesecker be invited as a speaker at an upcoming Steering Committee meeting.

End of Year deliverables were also discussed. Suggested deliverables included:

- Select process outcomes
- Genotyping
    - Each site will provide 300 samples to CIDR (2700 total)
    - Potentially site-specific sequencing
    - Initial validation of PGRNSeq for specific variants at CIDR validated by Sequenom
    - Coriell samples will also be sequenced at all sites running the PGRNSeq platform
    - Population information including, demographics of patients enrolled and distribution of patients enrolled

The structure of the variant/phenotype server was reviewed. Discussions are ongoing concerning the structure and information to be included in the repository. Initial phenotype components may include: billing codes, medications, vitals and labs. Other goals were: decision support algorithms, provider education materials, patient education website, and predictive algorithm performance.

John Harley and Keijan Zhang presented information about CCHMC's CYP2D6 Analysis. CCHMC utilizes CYP2D6 Taqman Low Density Array genotyping panel and CYP2D6 PCR Deletion/Duplication Assay. They are interested in variants that change enzyme activity – this set includes about 20 variants – 8 of which are indels *1 and *2A are normal and the rest disrupt function. John walked the group through their genotyping process, QC, assessments for the CYP2D6 region, and results. 28 CYP2D6 mutant variants were detected and were validated with official controls, patient controls and Coriell samples. Teri suggested that CCHMC compare their internal results to the PGRNSeq platform results.

Jane Romm from CIDR gave a brief project update. To date PGx reagent has been ordered and received from Nimblegen/Roche, CIDR has run the 32 HapMap Trios to validate the reagent and concordance was evaluated to UW 32 HapMap Trio dataset, Barcodes were developed on the GoldenGate assay for internal sample tacking, the first set of PGRNSeq samples was received in May from UW, the barcode was run and samples began production on 5/28/13. The coverage of the HapMap trios between CIDR and UW was 968,004 targeted based in PGRNSeq capture in 96 samples. Additional coverage, call and concordance data was shared to show the high concordance between UW and CIDR. A timeline moving forward broken out by site was discussed to give the Network an indication of when CIDR is expecting samples from various sites and when sites should expect their data back from CIDR. CIDR is using 2 of the 6 clinical control samples on each plate of 96 as a clinical validation cross-checking measure.

CYP2D6 Discussion – John Black

John Black reviewed CYP2D6 analysis as it relates to the frequency of hybrid genes in clinical samples and its impact on phenotype prediction. Since May 2013, Mayo has performed CYP2D6 analysis for 723 eMERGE samples. John briefly detailed current results and suggested considerations for the group as they prepare for CYP2D6 analysis via PGx. Beginning August 2013, Mayo's Personalized Genomics Lab will begin offering complete CYP2D6 testing in a CLIA approved, CAP inspected, and NYS inspected environment. With an EDTA blood sample and a Luminex assay, the lab will determine copy number, and if needed, CYP2D6-2D7, CYP2D7-2D6, phase of duplication and/or CYP2D6 full gene Sanger sequencing. As part of their PGx project, Mayo will compare their 300 CIDR samples to CYP2D6 analysis performed by the Personalized Genomics Lab.

Imputation/QC "Bootcamp" – Marylyn Ritchie

Marylyn presented the QC and imputation strategy being utilized by the CC. The eMERGE data is being imputed using the October 2011 1000 Genomes cosmopolitan panel. The group started imputation with BEAGLE and eMERGE v1.0 was imputed on this platform. BEAGLE was incredibly slow and the genomics workgroup decided that eMERGE Phase II data would be imputed utilizing IMPUTE2. The eMERGE v2.0 data was all imputed on the IMPUTE2 platform. Marylyn and her team have imputed all of the eMERGE Phase I and Phase II data and are currently working to clean this data. Multiple files are available for download and it is critical that sites fully understand what data they are downloading before trying to utilize the data. Files are labeled as either CLEAN or DIRTY and it is recommended that sites only work with the clean data.

The eMERGE imputed data are binary PLINK files that are provided in three formats *.bed, *.bim and *.fam. Genotypes with probability of 51% or higher were called in the PLINK files and others were set to missing.

Marylyn walked the group through many plots showing related samples, principle components, and coverage after QC. Additional QC and filtering will be required prior to use. Marylyn demonstrated the size of this job by sharing the imputation hours and CPU time utilized. A total

of 1,643,570.34 hours of computation time has been utilized for imputation of eMERGE Phase-I and -II data.

The Network was encouraged to:
- read all emails and QC documents
- be aware of what data you are downloading - there are currently 7 datasets available for download - some are clean but others are dirty, all are labeled
- QC the subset that you are working with

External Speaker – Lucila Ohno-Machado – iDASH
iDASH is a National Center for Biomedical Computing that develops new algorithms, open-source tools, computational infrastructure, and services that will enable biomedical and behavioral researchers nationwide to integrate data for analysis, anonymization, and sharing. iDASH will eventually serve as a data broker to provide the platform, software, and infrastructure needed to share data, tools, and policies while ensuring security, scalability, and flexibility. iDASH is currently made up of the integrated clinical data warehouses from five University of California medical centers and affiliation institutions that serve over 12 million patients combined. The Center's current objectives are to monitor patient safety, improve outcomes, and promote research. Another priority is to develop a "more informed" consent so that patients know what they are sharing and can choose to share more or less than what is typically shared. In order to achieve this goal, the Center is working on improving transparency in the use of data and biospecimens, creating simpler consent language, and developing a tiered consent mechanism.

Site Specific Presentation: Vanderbilt – Dan Roden
The Vanderbilt DNA bank, BioVU, has 166,142 samples as of May 27, 2013. Dan reviewed the Vanderbilt record counter and its capabilities for searching all BioVU samples, the BioVU Record Counters works similarly to the eMERGE record counter searching ICD codes, CPT codes, labs, and nutrition. BioVU has been genotyped by multiple projects including eMERGE-I, VESPA and now by the high density chips utilized for eMERGE-II. Vanderbilt will execute their four specific aims by coupling their discovery (BioVU) and clinical implementation (StarChart/PREDICT) platforms. Genetic signals are identified in BioVU and other very large research databases and provided there is an appropriate evidence base, relevant genotypes are embedded in clinical records. These actions serve as the platform for Vanderbilt to achieve their specific aims:
- Phenotyping
- Genomic Signals
- Patient Engagement
- Optimizing data access and patient privacy

Josh Denny is leading Vanderbilt's phenotyping effort. Vanderbilt is actively developing, implementing and validating phenotypes both developed by Vanderbilt and by other eMERGE sites. To date, Vanderbilt has led the effort on the C. Difficile (with GHC/UW) and ACE inhibitor cough algorithms. Both PheWAS and genetic risk score studies are also being implemented. The effort around genetic risk scores is being led by Dana Crawford.

Vanderbilt's Pharmacogenomic Resource for Enhanced Decisions in Care and Treatment (PREDICT) project serves as Vanderbilt's program for genotyping and tracking patient outcomes. Project work flow includes:
- A population of patients is selected who are "at high risk" for receiving a drug with an actionable "pharmacogenetic" story.
- The individuals are genotypes on a platform that assays genotypes important for variable actions of many drugs preemptively.

- These genotypes are stored and the informatics tools to provide point-of-care advice are developed and outcomes can be tracked.

Vanderbilt started with clopidogrel in 2010 when the FDA black box label identified a high risk group. Views from both PREDICT and StarPanel were displayed to highlight the architecture of the systems and Vanderbilt's point of care decision support model. Josh Peterson works with outcomes. PREDICT results are also displayed in the Vanderbilt Patient Portal under a "Your Risk" section. Ellen Clayton has been facilitating interviews about PREDICT, currently 21 semi-structured interviews have been conducted with 21 patients who have been seen in interventional cardiology clinic prior to catheterization. General findings from these interviews include:

- Patient did not remember being told about PREDICT
- Most have an accurate memory of medication change
- About half would have liked more information, particularly those who were tested but had no medication change
- Almost all would participate in PREDICT again, participating in a specific genetic test, and participate in a general genetic test.

Brad Malin continues to work on privacy. One recent project is Anonymized Clinical Data in Big Groups that works to create a system to successfully anonymize individuals in big groups of data. This project is ongoing and is seeking participation from the other eMERGE sites.

Site Specific Presentation: Group Health/University of Washington – Eric Larson

Group Health and the University of Washington reviewed their Specific Aims with the group:

- Extension of GWAS discovery analysis in EMR - currently GH/UW is the primary site for C. Difficile, varicella zoster, and carotid artery atherosclerotic disease. In addition to these phenotypes they are acting as the secondary side for numerous other phenotypes.
- Appropriate and effective integration of genomic information into the clinical care and EMR. GH/UWash is working to assess the feasibility of integrating genomic information into the GroupHealth patient-centered Medical Home (PCMH) using a prototype EMR user interface.
- Collaborate within and extend eMERGE. GH/UW is working as part of many external projects and consortia and has recently submitted an R01 focused on “Genetic architecture of memory and executive functioning in Alzheimer's Disease".

The group's C. Difficile and herpes zoster phenotypes' case/control definition was reviewed along with preliminary results. Preliminary analysis data was also discussed. Unfiltered imputation results were reviewed by David Crosslin and additional analyses with filters applied are planned. Analysis goals include : (1) to add additional pending data from eMERGE sites into the analysis, (2) survival analysis, (3) joint vs. meta analysis, (4) control for ancestry, (5) refine clinical model, (6) understand the combined imputed data.

Completed or in process activities at GH/UW to satisfy Aim 2 include: participant-observations of GHC Genomics Improvement Project (completed), Qualitative needs assessment with stakeholders (complete), and designing and testing functional prototypes for specific drug gene pairs through decision-making scenarios. GH/UW is currently in the process of designing and testing their functional prototypes. This will be executed through three primary studies:

- Model development and contextual inquiry specifically prescribing workflow models and interviews with patients and providers.
- Wireframe mock ups and participatory design with providers
- Functional prototypes and usability testing with providers - planned for 2014

GH/UW is working to finalize their algorithm for Carotid Artery Atherosclerosis Disease (CAAD). The group anticipates this phenotype will be ready for Network-wide implementation in fall 2013. Phenotype implementation will rely heavily on natural language processing (NLP).

Tuesday, June 4th

Workgroup Updates

EHRI Integration –Justin Starren & Marc Williams
The co-chairs briefly updated the status of current projects and discussed future plans. The workgroup has contributed to the following recent projects:

- "Genomic Medicine, Health Information Technology, and Patient Care." Published via *JAMA*, April 2013.
- "Crossing the Omic Chasm: A Time for Omic Ancillary Systems." Published via *JAMA*, March 2013.
- "The Electronic Medical Records and Genomics (eMERGE) Network: Past, Present and Future." Published via *Genetics in Medicine*, June 2013.

The Genetics in Medicine Special issue will be submitted in July 2013. Site specific updates regarding EHR implementation was overviewed along with the joint EHRI/CERC workgroup Infobutton project, to which many of the sites are contributing. The group plans to collaborate with the CSER Consortium during the joint October eMERGE/CSER meeting and is preparing a formal response to the ESP recommendations.

Return of Results –Iftikhar Kullo
Iftikhar briefly reviewed site specific plans for returning incidental findings related to the PGx projects and Genomic Medicine Pilots as well as outlined site specific impact of the ACMG recommendations. While not all sites plan to return incidental findings, many sites plan to allow for patient choice when returning results. Workgroup projects include:

- HFE –Abstraction form under revision, sites will begin Network –wide abstraction using revised, automated form.
- Chromosomal Abnormalities –The group will identify phenotypic correlates of autosomal abnormalities in an effort to standardize the nomenclature of these abnormalities.

Moving forward, the group plans to address the following issues:

- Monitor site specific response to the ACMG recommendations relevant to PGRNSeq
- Collaborate with CERC to consider a Network-wide assessment of patient provider responses to incidental findings from PGRNSeq
- Continue work on Network projects related to chromosomal abnormalities and penetrance
- Continue engagement with CSER/RoR and CRVR
- Develop a manuscript related to RoR activities

Publication Policy Review –Jonathan Haines
Jonathan outlined the goals and principles of the eMERGE publication policy and briefly reviewed current publication metrics. To date, there are 85 Network projects and 191 total projects either published or in development. As of June 2013, eMERGE publications have been cited over 1,000 times. Over 110 projects have been published and more than 75 are in development. The group was reminded that single site projects and abstracts can be published without approval, though a citation and copy of the publication must be sent to the CC. A collaborative project involving two or more sites constitutes a Network paper and is defined by any of the following criteria:

- Involves data from all sites
- Makes extensive use of Network-generated meta-data or procedures/protocols
- Focuses on topics relevant to eMERGE
- Involves policy/guidelines applicable to more than one site

Network members were reminded to cite eMERGE grant support with appropriate grant numbers and acknowledge the eMERGE Network. The group also discussed solutions for approving and tracking special journal issue contributions as these projects are gaining momentum within the workgroups. Workgroups should create one main manuscript concept sheet for the project and abridged manuscript concept sheets for each of the articles so that the

project can be shared Network-wide as a single packet. Sites have also agreed to forward the publication/abstraction portion of their quarterly progress reports to the CC in an effort to assist with publication tracking.

CERC WG Update–Maureen Smith & Ingrid Holm
The co-chairs provided a brief overview of their ongoing projects, which include:

- Pediatric Biobanking Consent Manuscript–Lead: Kyle Brothers
- PGx Consent Manuscript –Lead: Maureen Smith
- Patient Education Website –Lead: John Connolly
- Genomics Project Implementation Across Diverse Healthcare Settings –Lead: Malia Fullerton

The workgroup's other education-related activities include collaborating with the EHRI workgroup to develop content for the eMERGE InfoButton project and teaming up as much as possible to develop additional educational materials for the PGx project. The group has contributed the article, "Engaging Stakeholders and Setting Goals" as part of the EHRI workgroup's Genetics in Medicine Special Issue project and is actively pursuing collaborations with the CSER and RoR Consortiums. The group also plans to contribute to the new eMERGE Pediatric Workgroup on issues related to consent and return of results. Moving forward, the group has proposed working with the Return of Results workgroup and CSER Consortium to craft an official response to the ACMG Recommendations. This project could center on patient/physician interviews in relation to the PGx project as a means of understanding perceptions of incidental findings, returnable results, and actionability.

eMERGE PGx Variant Repository Subgroup Update –Josh Denny & Marylyn Ritchie
The subgroup is tasked with constructing a centralized repository for storing variant information and data analysis that results across the six PGx sequencing sites –each of which have their own variant calling pipelines and QC filters. The group has begun surveying each of the sites' variant calling pipelines in order to determine similarities and differences among the sites' variant calling methods. Based on a comparison of the UW and CIDR pipelines, the current repository plan is as follows: 1)The CC will receive VCF and BAM files from six eMERGE sequencing sites; 2)Variant repository version 1.0 will include site generated variants; and 3)The CC will consider recalling variants from BAM files once all data has been received. The group plans to model the PGx variant repository off UW's Exome Variant Server design.

The following action items are planned:

- Each site will send contact information for sample preparation, sequencing, and alignment/variant calling to the CC (Sarah Stallings).
- The CC will collect details on reference being used, alignment software, and variant calling software and coordinate this information with the group so that these items are uniform among the sequencing sites when possible
- CIDR will send information on positive control trios and clinical samples to the sequencing sites
- CIDR will send 96 barcode SNP list to the CC to disseminate among the sites
- Sites will not perform structural variant calls right now
- All sequencing sites should get the 96 HapMap sample BAM and VCF files from CIDR and run BAM files through their variant calling pipeline and compare with CIDR VCF files.

Privacy Update –Brad Malin
Brad reviewed recent identifiability issues within the scientific community. Several recent publications, including several authored by eMERGE members, focus on re-identification of research volunteers from data linked to genome-wide association studies. Topics being addressed most recently include:

- Protecting individuals whose data are re-identified

- Interplay with data sharing
- Making re-identification results public
- Publication and reproducibility of re-identification results and scientific value
- Re-identifiability and confidentiality
- Direct contact with re-identified subjects
- Legal maneuvers
- Validity and scoring methods

Brad also discussed a developing project related to clinical profile anonymization within eMERGE. VU investigator, Raymond Healtherly, is currently leading a collaborative project to anonymize eMERGE cohorts in context of their biorepository and EMR populations. Finally, Brad briefly outlined the benefits of using secure meta-analysis within the context of multi-site studies. Secure meta-analysis will allow sites to obscure their raw data, allow for the combination of data without revealing any particular site contributions, and scales according to the number of participating sites.

Genomics WG Update–Dana Crawford & David Crosslin
The co-chairs briefly overviewed recent workgroup discussions and developing plans, including their goals for imputation and QC documentation moving forward. Current workgroup projects include:

- Frontiers in Genetics Special Issue
- LogR and B Allele
- Genetic Risk Scores
- Gene-Gene Collaborations

In terms of imputation validation, the group is exploring PCA as a tool for deciphering between meta-analysis and joint analysis as a means to creating a combined data set prior to imputation. The group is also surveying variant calling methods across the PGx sequencing sites to ensure that sequencing data can be merged into one standardized, quality set. The co-chairs also reviewed UW's NEXT Medicine Study variant database, a tool constructed via REDCap and used for annotating study-specific variants. This method for documenting and transferring variant data may be useful when developing the PGx repository.

Phenotyping WG Update–Josh Denny & Peggy Peissig
The co-chairs briefly reviewed the status of Phase II phenotypes. Current Phenotyping challenges and proposed solutions were also outlined. These include:

- Challenge 1: Data Standardization
  - Proposal: Secondary sites must validate case/control accuracy and follow the proposed data dictionary when sending data extract (including covariates) to the primary site. Primary sites may reject data submissions that are incorrectly formatted.
- Challenge 2: Basic Demographic Information
  - Proposal: The workgroup aims to provide basic demographic information on nearly 100% of eMERGE subjects using a standard, Network-wide definition.
- Challenge 3: Rare Phenotypes (Ex. DILI)
  - Proposal: Because PPVs are low, these phenotypes require manual review of all possible cases –execution should be optional if a difficult manual review is required. Primary sites can engage other sites with extra funds or staff to help with execution if they wish.

Other updates included:

- eMERGE Record Counter now includes real time updating, pediatric data, and CPT codes.

- With regards to the eMERGE Record Counter and PGx Repository, the group is working to map medication data to RXNorm ingredients using tools such as MedEx-UIMA and cTAKES.
- JAMIA Special Issue on Electronic Health Records: Driven Phenotyping –Jyoti Pathak, Josh Denny, and Abel Kho are serving as guest editors for this developing special issue.

Closing Remarks, Final Discussion –Rex Chisholm

Rex discussed the leadership team's goals for the Network as it prepares for a possible eMERGE Phase III. The Network is encouraged to maintain the discovery and implementation components of eMERGE I and II, but is also expected to generate Network-wide evidence for personalized medicine and EMR implementation. With regards to eMERGE Phase III phenotypes, the Network will focus on variants that are uncommon, but not extremely rare. The Genomics and Phenotyping workgroups are tasked with identifying subjects with functional null genotypes of interest and combining them with phenotypic data from the EMR. Rex thanked the Steering Committee for a productive meeting.

June Steering Committee Meeting Action Items

1. A pediatric working group will be established, led by John Harley. For the new pediatric workgroup, the CC will coordinate with the pediatric sites as well as additional sites, particularly those with pediatric samples and an interest in contributing, to identify membership and schedule a regularly occurring call.
2. As part of the PGx project, Mayo will compare analysis of their 300 CIDR samples to CYP2D6 analysis performed by Mayo's Personalized Genomics Lab.
3. Sites are encouraged to compare the analysis of their PGx samples to CCHMC's CYP2D6 validation chart (pg. 17 of CCHMC's CYP2D6 presentation).
4. Sites will continue to read QC documentation and emails before downloading data from the CC.
5. Sites will begin regularly submitting the publication/abstract portion of their quarterly progress reports to the CC.
6. The CC will distribute the Variant Repository Action Items to the group for review and execution.
7. As part of the process for preparing for a possible eMERGE Phase III, the Genomics and Phenotyping workgroups are tasked with identifying subjects with functional null genotypes of interest and combining them with phenotypic data from the EMR. The Genomics workgroup will begin this process and report back on an upcoming PI call.
8. CC will revise publications policy to reflect the process for Special Issue publications. The process will be for the organizers to submit a single Manuscript Concept Sheet for the entire publication that details each article and proposed authorship. As the publication matures, abbreviated MCS will be submitted for each article.
9. Network ResHTN project team will go back and revisit the phenotype algorithm, focusing on controlled hypertensive definition, especially in the African American population
10. Coordinating Center will remap End of Year deliverable timeline to reflect new sequencing and data processing time expectations along with cross-site platform comparison and orthogonal validation
11. ACMG returnable variants list – suggested that group work with CSER on the question of which variants to return. Group should invite Les Biesecker to upcoming eMERGE meeting
12. Address stratification artifacts in imputed data
13. eMERGE PGx workgroup to develop data QC and variant calling plan for use across the network for PGRNSeq data (plan and timeline completed) – specific tasks for sites doing PGRNSeq sequencing (CHOP, CIDR, Geisinger, GHC / UW, Mayo, Mount Sinai) are:

13.1 Sites will send CC Variant Calling Pipeline and QC Descriptions (if determined) Note: File headers from BAM and vcf files.

13.2 Review and Establish feasibility of using CIDR's batch control methods (for cross-site concordance validation).

13.3 Review and Establish feasibility of using CIDR's sequence control ("barcode" SNPs and Golden Gate).

13.4 Compare Site's Calling Pipeline results to CIDR's data – Genomics workgroup/CC.

13.5 Choose a consensus variant calling pipeline and QC protocol for site-based calling that will be used by all sequencing sites.