eMERGE Network Steering Committee Meeting
June 4-5, 2012
Minneapolis, MN

Attendance

Network Members in Attendance

CCHMC/BCH	Beth Cobb
CCHMC/BCH	John Harley
CCHMC/BCH	Ingrid Holm
CCHMC/BCH	Isaac Kohane
CCHMC/BCH	Bahran Namjou (phone)
CCHMC/BCH	Cassandra Perry
CCHMC/BCH	Wendy Wolf
CHOP	John Connolly
CHOP	Hakon Hakonarson
CHOP	Brendan Keating
Geisinger	David Carey
Geisinger	Andy Faucett
Geisinger	Samantha Fetterolf
Geisinger	Helena Kuivaniemi
Geisinger	David Ledbetter
Geisinger	Gerard Tromp
Geisinger	Janet Williams
Geisinger	Marc Williams
GH/U WA	David Carrell
GH/U WA	David Crosslin
GH/U WA	Gail Jarvik
GH/U WA	Eric Larson
Marshfield	Richard Berg
Marshfield	Murray Brilliant
Marshfield	Terrie Kitchner
Marshfield	Simon Lin
Marshfield	Jim Linneman
Marshfield	Cathy McCarty
Marshfield	Peggy Peissig
Marshfield	Carol Waudby
Mayo	Sue Bielinski
Mayo	Chris Chute
Mayo	Mariza de Andrade
Mayo	Matt Durski
Mayo	John Heit
Mayo	Hayan Jouni
Mayo	Barbara Koenig
Mayo	Iftikhar Kullo
Mayo	Jennifer McCormick
Mayo	Jyoti Pathak
Mt. Sinai/Columbia	Erwin Bottinger
Mt. Sinai/Columbia	Steve Ellis
Mt. Sinai/Columbia	Omri Gottesman
Mt. Sinai/Columbia	Casey Overby
Mt. Sinai/Columbia	Kash Patel
Mt. Sinai/Columbia	Nicholas Tatonetti
Mt. Sinai/Columbia	Chunhua Weng
Northwestern	Rex Chisholm
Northwestern	Geoff Hayes (phone)
Northwestern	Abel Kho
Northwestern	Jennifer Pacheco
Northwestern	Laura Rasmussen-Torvik
Northwestern	Anna Roberts (phone)
Northwestern	Maureen Smith
Northwestern	Justin Starren
Northwestern	Will Thompson (phone)
NHGRI	Rongling Li
NHGRI	Teri Manolio
NHGRI	Ian Marpuri (phone)
NHGRI	Eugene Passamani
Vanderbilt	Ellen Clayton
Vanderbilt	Dana Crawford
Vanderbilt	Josh Denny
Vanderbilt	Janina Jeff
Vanderbilt	Dan Roden
Vanderbilt	Sarah Stallings
Vanderbilt	Sara Van Driest
CC	Melissa Basford
CC	Jonathan Haines
CC	Brad Malin
CC	Lauren Melancon
CC	Marylyn Ritchie

Network Invitees and Guests

Aurora Healthcare	Michael Michalkiewicz
Baylor College of Medicine	Steve Scherer
UK BioBank	Cathie Sudlow
U. of Arizona	Bonnie Hurwitz (phone)
U. of Minnesota	Susan Wolf
USAF	Ronald Miller

Decisions and discussion

NHGRI Program Office Report – Rongling Li

Rongling provided the group with a NIH appropriation update and outlined three upcoming funding opportunities:

- Genomic Medicine Pilot Demonstration Projects (U01)
- Genomic Medicine Pilot Project Demonstration Projects Coordinating Center (U01)
- Center for Genomic Studies on Mental Disorders (U24)

Gary Gibbons was named the new director of the National Heart, Lung, and Blood Institute, and Teri Manolio was mentioned for her recent honors. Rongling continued by highlighting eMERGE news including: Brad Malin's testimony to the National Committee of Vital and Health Statistics on behalf of eMERGE and the official notice of the pediatric sites, CCHMC/BCH and CHOP joining the eMERGE Network.

Rongling also outlined the number of participants and genotyped patients for Phase I and Phase II, including the Pediatric sites, for a total of 87,469 genotyped patients to date. NHGRI consortia workgroups were presented to the group with a focus on the overlap. This overlap is being addressed by the eMERGE Network as collaborations with external groups continues to grow.

Goals for this meeting include:

- Identify the best way to integrate pediatric sites
- Identify specific approaches to collaborate with VA, Air Force, UK BioBank and other possible collaborators
- Determine the milestones and timelines for eMERGE PGx and site specific genomic medicine pilot studies
- Plan for dissemination of the lessons learned and research results to the scientific community

Pediatric Site Presentation: The Children's Hospital of Philadelphia (CHOP) – Brendan Keating

In addition to the Center for Applied Genomics (CAG) at CHOP there are nine other centers of emphasis within the CHOP system along with a collaboration with the University of Pennsylvania. CAG was founded in 2006 and currently has >60 active disease areas and >150,000 samples genotyped. The mission of the Center for Applied Genomics is to use genomics approaches to develop new and better ways to diagnose and treat children affected by complex medical disorders, both common and rare. The CAG BioBank continues to grow at ~15,000 samples per year. Currently, the CAG Biobank contains ~150,000 samples most of which are of European American ancestry (51%). Other ancestry groups represented include: African American (38%), Asian American (5%), Native American <2%), and Hispanic (5%). Disease coverage includes all major common pediatric diseases along with a large collection of rare diseases. In additional to pediatric samples CAG has ~15,000 adult samples, mostly parents of children who have been genotyped. Genotyping platforms utilized are Illumina & Affymetrix for GWAS data and Illumina and Life Technologies for sequencing data. The average record in

CHOP's EMR (EpicCare) dates back 5.5 years per patient. The EMR records receive a soft update every 3 months and a hard update every 12. CHOP outlined their institutional goals:

- Mine specific disease phenotypes to establish a phenotype/genotype database for future clinical development.
- Extend their Clinical Laboratory Improvement Act (CLIA)
- Establish guidelines and governance rules for the CAG biorepository and databases, and generate informed consent procedures

Pediatric Site Presentation: Cincinnati Children's Hospital Medical Center/Boston Children's Hospital (CCHMC/BCH) – John Harley & Zak Kohane

John and Zak gave a brief summary of both CCHMC and BCH. A total of 5,562 GWAS were submitted to dbGAP for pediatric patients at these two sites. Their Biobank is composed of a combination of tissue and DNA with about 20,000 frozen tissue samples from operations in pathology (waste tissue) and about 10,000 DNA samples. The Biobank continues to grow at a rate of about 2,500 new DNA samples per month. Only a small fraction of consented patients do not wish to have results returned to them. Genetic pharmacology has been integrated into the EMR with PGx information linked to inpatient medication ordering since 2004. The group outlined their proposed phenotypes, some of which will overlap nicely with the adult population currently in eMERGE. It was also stated that the pediatric community is ahead in accepting PGx into clinical care and CCHMC/BCH looks forward to integrating into the eMERGE-PGx project. Zak briefly outlined their i2b2 system and SHRINE aggregator and the 73 centers nationally and internationally it connects.

Pediatric Site Discussion

Strategies to effectively integrate the two new pediatric sites into the existing eMERGE Network were discussed. The Network agreed that the new pediatric sites should be fully integrated into the existing workgroups and that new pediatric specific workgroups were not necessary and would only create a divide. The pediatric sites were strongly encouraged to name site members to participate in each of the 6 existing workgroups if they have not already done so.

eMERGE Phase II Site: Mount Sinai – Erwin Bottinger

Erwin reviewed the Mount Sinai Biobank enrollment, demographic and genotyping numbers showing the diversity of their subjects. In addition to adult patients, Mount Sinai began enrolling pediatric patients in January 2012. The exome chip has been performed on 3,238 African American patients to identify significantly associated rare variants in known triglycerides gene APOB, and preliminary results were shared. Mount Sinai is currently collaborating with almost every Network site in some capacity and continues their ongoing collaboration with Columbia. An update was provided for phenotype development and genomic data submitted to the CC for imputation. Erwin reviewed their drug-induced liver injury algorithm in detail, walking the group through their case definition, merged methods, and topic modeling plan and results. Erwin also reviewed CLIPMERGE, a Clinical Implementation of Personalized Medicine through Electronic health Records and Genomics tool. This tool is designed to sit alongside the EMR to aide with clinical decision support. This tool will support Aim 2 in the eMERGE-PGx project which states that actionable variants will be deposited into the EMR by displaying the result and providing decision support. The architecture of CLIPMERGE was outlined along with how Mount Sinai would be utilizing this tool for non-diabetic kidney disease in patients of African ancestry.

eMERGE Phase II Site: Geisinger – David Carey & Marc Williams

The Geisinger team briefly described to the group the importance of studying abdominal aortic aneurysm (AAA) and their interest in the project. The group outlined the clinical features of AAA stating that many patients are undiagnosed due to it being asymptomatic and not easily identifiable in a simple lab test. The procedure to fix AAA is safe and effective, but if this disease goes undetected it can be life threatening. Geisinger has received a PA-CURE Grant for Translational Genomics for AAA. Through this grant Geisinger plans to create a novel AAA risk scoring tool, prospectively validate the genomically-informed risk model in an outpatient population, and develop and evaluate a clinical implementation plan utilization of genomic data in outpatient clinics. Another project being worked on at Geisinger focuses on the implementation and modeling of IL28B genotyping in Hepatitis C infection. Implementation will be executed through three avenues: standardized order sets, clinical notes, and reporting. Clinical decision support will include clinical, laboratory elements and genotypes. Currently there is a moratorium on implementation until the Geisinger EPIC update is completed. Decision analysis and modeling has shown that protease inhibitors added to standard therapy increase response rate in treatment resistant patients. Modeling will be done for genotype-guided therapy along with cost-comparison analysis.

Steve Scherer – Baylor College of Medicine, representing the PGRN

Steve provided the Network with a brief overview of PGRN, outlining their mission, research groups, Network resources and leadership team. The primary cost driver in capture sequencing is the library prep. The basic Illumina workflow was outlined for the group starting with sample intake through library QC. Once QC is complete samples go through liquid-phase hybridization capture, cluster generation, and reversible dye terminator sequencing. The analysis pipeline and CASSANDRA components were discussed in addition to orthogonal platforms for variant validation. The VIP Capture Gene List was selected from nominations from PGRN Research Groups. Out of the 239 nominated genes, 84 were selected to be included. Once the chip was completed the PGRN-Seq platform was tested with a HGSC HapMap test set and a HapMap/TGP Trio set. Preliminary test data was run, and the platform is performing well with the exception of CYP2D6. PGRN is still working to overcome the current CYP2D6 issue that has been identified. In addition to the PGRN-Seq platform a smaller platform is being created - the PGx-Seq. This platform is composed of CPIC genes or near-CPIC genes and can be run on either Illumina MySeq or Ion Torrent. The process for utilizing the Ion Torrent was reviewed and the MySeq and Ion Torrent platforms were compared. Steve closed his presentation with a brief overview of the ENCODE Project (The Encyclopedia of DNA Elements), an international collaboration of research groups funded by NHGRI. The goal on ENCODE is to build a comprehensive parts list of functional elements in the human genome.

Cathie Sudlow – UK BioBank

Cathie presented an outline of the UK BioBank. This large prospective cohort study includes 500,000 UK adults age 40-69 (at time of recruitment). Information collected includes baseline data on lifestyle, personal and family medical history, physical measures and biological samples. Follow up for disease outcomes will occur over the next 10-20+ years. Goals of the UK BioBank include establishing genetic and environmental determinants of common diseases of middle and old age and improving prevention, diagnosis and treatment of cancer, heart disease, stroke, arthritis, dementia and other diseases. The size of the data set will generate sufficient numbers of cases of diseases to allow adequately powered nested case-control and case-cohort

studies. The UK BioBank recruits at 22 different centers based in large cities located in England, Scotland and Wales. Participants go through a series of stations at each center, answering questions and giving samples of blood, urine and saliva. Cathie shared the characteristics of UK BioBank participants which she explained was an accurate portrayal of the UK population, although this group is less socioeconomically deprived than the UK average. There are plans for enhanced phenotyping in the way of web-based questionnaires, repeat assessments, re-contacting participants, and wrist-worn accelerometers. If additional funding were to become available 100,000 patients would have an imaging visit and a standard panel of assays on samples from all participants would be run. Long term follow-up and advantages come from the fact that NHS provides the majority of healthcare in the UK and that many routine procedures are coded, including cancer registrations, death registrations, and hospital episode data. Cathie showed the group the list of prevalent conditions identified at recruitment, incident outcomes during follow up and the projected incidents through 2022. Access to the UK BioBank data is available for health-related research that is in the public good. Applications are reviewed by the Coordinating Center, access sub-committee and the Ethic and Governance Council. The only cost incurred by the researcher is the cost of application and provision of the data, assay results or samples the scientists require. All results are to be shared with UK BioBank so that advances can be built on by others.

eMERGE-PGx Collaboration – Dan Roden

The overall goal of the eMERGE-PGx project is to initiate a multi-site test of the concept that sequence information can be coupled to electronic medical records for use in healthcare. Dan reviewed with the group the project plan for the eMERGE-PGx project as outlined in the proposal sent to NHGRI. PGRN and eMERGE capabilities were also outlined showing how these two consortium complement each other's work. The proposed project has three aims:

- deploy the PGRN-Seq platform across the eMERGE Network;
- integrate validated genotypes into the EMR and assess uptake, acceptance, and clinical impact;
- develop a repository of variants of unknown significance.

Dan briefly outlined Vanderbilt's PREDICT model and current decision support tools as an example of possible avenues to take when thinking about implementation.

The group discussed some pending logistical questions brought up by NHGRI reviewers and anticipated site timelines.

- Sites discussed their site specific timelines for (re)consenting patients; some sites are ready now while others will not be ready to begin this process until early-late fall.
- The group discussed the two potential platforms being created by the PGRN: PGRN-Seq (84 variants) and PGx-Seq (24 variants). After discussing the pros and cons of each platform it was decided that the larger PGRN-Seq platform would allow the groups a larger discovery set in addition to the returnable actionable variants.
- A single Network-wide subject consent was briefly discussed. Many expressed their anticipated difficulty with this task but believed that similar elements could be used Network-wide. Mayo can send a draft of their consent form to the Network for review.

Site Specific Presentation: PR Interval Fine-Mapping Project in African Americans – Janina Jeff (Vanderbilt)

Janina presented to the group her efforts on the collaborative PR Interval Fine-Mapping Project in AAs from Vanderbilt and Northwestern biobanks. GWAS was performed on 455 AAs looking at 930K SNPS and 4 ECG traits. The specifics of the project were reviewed including those excluded and the NLP algorithm utilized. SNPS for fine-mapping were selected by utilizing two approaches: fine-mapping of GWAS-identified disease loci (European) and fine-mapped regions identified by admixture mapping. These two approaches were outlined and results from each were shared with the group. This study has replicated previous associations and has identified potential novel regions from fine mapping signals identified by GWAS and admixture mapping for PR interval. Next steps include testing for pleiotropy and looking at the *in silico* replication of results. An abstract has been submitted to ASHG and a manuscript is currently under review at Circulation: Cardiovascular Genetics.

Michael Michalkiewicz – Aurora Health Care

Michael outlined the Aurora Heath Care system. The Aurora EMR has been system wide since 1998 and contains 4.7 million patients, making it the largest not-for-profit health care provider in the state of Wisconsin. Areas of focus include cardiovascular diseases, cancer, obstetrics, gynecologic surgery and neurological procedures. ORBIT, Aurora's biorepository, is linked to their EMR system and continues to add ~900 samples per month. ORBIT is an opt-in participation model that asks patients if any leftover blood sample can be used for research. This population can also easily be studied longitudinally as most patients do not move from the area. Aurora is working to build an EMR mining team and is working to implement some eMERGE algorithms (PAD, AAA, RH). In addition to implementation eMERGE algorithms they are construction their own for pre-eclampsia and pulmonary hypertension. They are also working to establish the Aurora Center for Medical Bioinformatics to provide expertise in large volume data retrieval, transformation, analysis, and data presentation dedicated to clinical and research programs.

EHR Integration Workgroup Report – Erwin Bottinger & Justin Starren

Workgroup activities were outlined and include two manuscript proposals with the possibility of creating a special journal edition around the "So you want genomics in your EHR" concept. The workgroup invited representatives from HL7 and CDSC to join and present at workgroup meetings and are working to further collaborate with these two groups. Health eDecision is another group with whom the EHRI workgroup is interested in starting discussions. Justin went on to further explain the Crossing the Omic Chasm concept, expressing the current hurdles and showing how the group proposes to handle the integration into the EHR and will look to the PGx project to push this implementation along. It was noted that there is a commercially available database with an effective database model/design that can be used to store all of the "Ancillary Omics" that are to be returned through the EMR for Clinical Decision Support. The EHRI workgroup will also collaborate with the Genomics workgroup to obtain a greater understand of the VCF files that are currently the standard for genomic data.

Genomics Workgroup Report – Dana Crawford

The main activities of the genomics workgroup, to date, include imputation of Phase I and Phase II data, the network-wide Resistant Hypertension project, genetic risk scores and collaborations. Imputation for Phase I genotypes has been completed and is now available for download. The Phase II data is currently being imputed and Marylyn anticipates this set will be completed by

the end of July. Both of these sets have been imputed using Beagle and have used the 1000 Genomes October 2011 Cosmopolitan set as a reference panel. Dana provided a Resistant Hypertension update outlining the phenotype definitions and current project results. Next steps are in place and include incorporating data from Mount Sinai, updating the GWAS catalog analysis, composing a manuscript and a possible collaboration with PGRN. A second project was outlined, Genetic Risk Scores and eMERGE, and progress update was given. Next steps to move this study forward are in process and include calculating African American weighted GRS using African American data, collecting clinical data for comparison, using imputed data for missing relevant SNPS, and expanding the study beyond T2D. The workgroup plans to continue working on imputation, resistant hypertension, genetic risk scores, and collaborative projects (AAA, Resistant Hypertension with PGRN, and the exome chip project.)

Return of Results Workgroup Report – Gail Jarvik & Iftikhar Kullo

The workgroup defined their charter and group priorities. The workgroup plans to focus on genetic risk scores, the eMERGE PGx project and collaborations within eMERGE and non-eMERGE groups. Iftikhar outlined the clinical utility of risk scores, examples of diseases and the potential actionability, along with the limitations associated with risk scores. Gail shared each site's plans for the PGx project regarding implementing specific PGx variants. All sites will be focusing on warfarin, clopidogrel, and statins. Most sites will also explore additional targets. Issues that were addressed by the RoR workgroup included consent and difficulties with implementation for certain variants. It was suggested that the sites work together to create common language for their consent forms. The workgroup has requested a representative from the CCHMC/BCH site speak to the group at the next meeting to discuss their neuropsych study. One additional project the group is working on is a Network wide paper on hemochromatosis. The concept sheet for this paper has been circulated and Gail anticipates calls to discuss this project will begin in late June. The workgroup plans to continue its collaborations with other eMERGE workgroups as well as external consortia such as the CSER and RoR consortia.

Phenotyping Workgroup Report – Josh Denny & Peggy Peissig

The workgroup briefly reviewed the number of months a selection of Phase I phentoypes took to complete. These phenotypes ranged from taking a total of 8 months to 24 months depending upon the algorithm. Peggy went on to show the PPV progress from Phase II phenotypes. The workgroup has separated the list of 21 phenotypes into 3 categories. Most primary phenotypes have been run and validated at the primary and secondary sites. Most secondary phentoypes have run and validated at the primary site with some being moved to the secondary sites. Most tertiary phentoypes are still in production at the primary site. Due to the large number of phenotypes at varying stages of production the workgroup has decided to select "Fast Track" phenotypes to move along more quickly if they meet the following criteria:

- Readiness of algorithm (must be validated and implemented at multiple sites)
- Technical feasibility
- Validation feasibility
- Novelty/new information contribution
- Impact
- External pressure (consortia, etc.)

The workgroup outlined the currently nominated phenotypes with the intention of having one phenotype nominated from each site. Sites should send the workgroup co-chairs a few slides

addressing the above criteria to be reviewed by the full Phenotyping workgroup for prioritization. The goal of this fast tracking is to have 4 phenotypes completed by the end of October 2012. Josh presented early C.Diff results and shared with the group screen shots from PheKB.org. The workgroup has prepared a plan for incorporating the new pediatric sites and plans to begin by building a collection of pediatric phenotypes and status, identifying overlapping phenotypes between the adult and pediatric sites, and working to include common phenotypes across all sites.

CERC Workgroup Report – Maureen Smith

CERC activities include:

- eMERGE PGx
- Liaisons
- Physician/Patient Consultation
- Pediatric sites/issues
- CLIA/CAP

The group discussed drafting a master consent that can be modified as needed by each site. Some Network sites have a PGx consent already drafted and are ready to go to the IRB. Mayo will plan to send their current consent to the CERC workgroup as a starting point, and the workgroup was encouraged to meet within the next week to start this process. Education will also be a topic addressed by the CERC workgroup. This education should be integrated into clinical decision support systems, and the workgroup will need to call on outside experts (content experts and education and communications experts) to properly facilitate effective education tools. The workgroup is currently participating or plans to participate in the following consortia:

- Return of Results Consortium (co-chairs actively participate on these calls/meetings)
- CSER – Pediatric workgroup
- CTSA Biobanking workgroup

The workgroup decided to form a Physician/Patient Consultation sub-group that has been meeting monthly to discuss strategies, issues, and survey results. The workgroup determined that there was no need for an additional pediatric workgroup as many of the sites were planning to start or have already begun a pediatric biobanking project. Pediatricians may be more receptive to genetic testing but there are many ethical, legal and psychosocial implications that must be considered. These will be topics for discussion at upcoming workgroup calls. Early in Phase II the workgroup felt that there was a need to post the CLIA/CAP guidelines but it seems that most of those early issues have been resolved. Instead of posting the guidelines the workgroup plans to query the Network for remaining questions and produce a Q&A/information page on the eMERGE wiki as a resource.

Closing Remarks, Final Discussion – Rex Chisholm

Rex thanked the Steering Committee for a productive meeting and good discussion had by the group. Great progress has been made and Rex encouraged everyone to continue working on the topics discussed during the general sessions and breakout time.

Action Items

1. Networks sites will submit slides to the Phenotyping co-chairs outlining their proposed fast track phenotypes.

2. CC will coordinate a call for CERC/PGx workgroup members to discuss consent forms for the PGx project.
3. Mayo and Marshfield will send their currently drafted consent forms to the CC to be reviewed and used during the call being scheduled.
4. CC will create a list of members who are involved in other consortia in addition to eMERGE, such as the CSER and Return of Results consortia, to identify the overlap between Networks.
5. CC will e-mail all workgroup co-chairs with an updated list of workgroup participants.
6. The Return of Results workgroup will invite a representative from CCHMC to their next workgroup call to give a brief presentation concerning pharmacogenetics/pharmacology service ROR.
7. EHRI and Genomics workgroup will plan to attend the other workgroup's monthly calls to discuss genetic file formats.
8. The Genomics Workgroup will share a VCF file format with the EHRI Workgroup for the initiation of data filtering approaches.
9. Sites will send the number of additional samples they plan to submit to Marylyn Ritchie.
10. CC will work to schedule a PGx call to discuss outstanding topics from the SC meeting: platform to be utilized, project timeline, Network-wide consent form and a data repository.
11. CERC workgroup will need to call on outside experts (content experts and education and communications experts) to properly facilitate effective education tools.
12. Rongling will follow up with dbGaP to determine the pediatric data QC status.
13. The pediatric sites will submit their datasets to the CC after dbGaP completes the data QC.
14. Phenotyping workgroup will work with the pediatric sites to identify the relevant network phenotypes such as C. Diff.