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. 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 100 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 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. 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:
14. Sites will send CC contacts for Variant Calling Decision Making.
15. Sites will send CC Variant Calling Pipeline and QC Descriptions (if determined) Note: File headers from BAM and vcf files would be fine!
16. Review and Establish feasibility of using CIDR's batch control methods (for cross-site concordance validation).
17. Review and Establish feasibility of using CIDR's sequence control ("barcode" SNPs and Golden Gate).
18. Compare Site's Calling Pipeline results to CIDR's data – Genomics workgroup/CC.
19. Choose a consensus variant calling pipeline and QC protocol for site-based calling that will be used by all sequencing sites.