eMERGE Network: Summary of the Steering Committee Meeting

June 8-9, 2017 in Boston, MD

The Summer Year Three, Phase III eMERGE Steering Committee was held on June 8-9th, 2017 in Boston, MD. In order to ensure that the Network continues on a productive note as we begin our third year, please find highlights from the Meeting below and a collated list of action items at the end of the document.

Presentation slides are available here (login required).

Full Session

Welcome, Opening Remarks, General Updates | *Rongling Li (NIH/NHGRI)*

- Rongling Li, on behalf of the NHGRI eMERGE team, presented the Program Official Report noting the FY18 budget depends on congressional approval of the budget.
- The Network reviewed the eMERGE Phase III Timeline, with revised dates.
  - The 'Future of eMERGE' meeting is scheduled for October 30, 2017 in Rockville, MD. This meeting will bring together scientific experts outside of eMERGE, eMERGE ESP members, eMERGE PIs, and NIH representatives.
  - Goal of the meeting: To identify genomic medicine research areas that can fully use eMERGE infrastructures and resources and integrate into other NHGRI programs.
  - Opening presentations at the meeting will be given by Teri Manolio, Rex Chisholm (eMERGE overview), and the Precision Medicine Initiative.
- Goals of our current meeting included:
  - Update on genomic sequencing status, specifically on the clinical report, data integration, and dataflow.
  - Report issues related to the imputed eMERGE I-III dataset.
  - Prepare the response to the ESP comments.
  - Share scientific projects' results.
  - Suggest possible directions for a possible eMERGE IV.
- It is noted that Francis Collins has been re-appointed as NIH Director.
- Our next Steering Committee and ESP Meeting, October 9-10 in Bethesda, MD also marks the 10th anniversary of eMERGE.

Announcements, Opening Remarks | *Rex Chisholm (Northwestern/SC Chair)*

- Developments since the prior Steering Committee meeting (February 2017):
  - eMERGE received favorable ESP response after our April 2017 call. The ESP has an interest in variant classification changes and variability among sites. There are plans to emphasize phenotyping and EHR integration at the next meeting.
  - Imputation of array data on the Michigan Imputation Server (MIS) is complete and merged files have been released.
  - The CSGs have begun return of sequencing results and clinical reports to sites. To date, 15,589 samples have been received, 11,718 sequenced, and 2055 clinical reports issued.
  - PGx dataset is released and uploaded to dbGaP.
  - Rex noted that the PGx dataset is an incredible resource and the Network should work to maximize its usage. The Network needs to ramp up use of the existing data sets to increase the publications over the next year.

Sequencing and Genomics Data Update | *Heidi Rehm (Partners/Broad), Niall Lennon (Partners/Broad), Richard Gibbs (BCM/HGSC) & David Crosslin (UW/CC)*

- Baylor has sequenced 1641 samples from Northwestern, 2535 samples from Mayo, 1289 samples from CHOP, 1057 samples from Vanderbilt, and 1046 samples from Columbia.
  - Baylor is currently resolving a few issues receiving samples, but anticipates complete reporting by January 2019.
- Data are located on DNAnexus from Baylor and Partners/Broad samples. CNV calls will be on DNAnexus shortly and will be posted separately from VCF files.
- Partners/Broad has sequenced 1221 samples from the KPW/UW, 1250 samples from Geisinger, and 1432 samples from CCHMC.
- eMERGE I-III Imputed dataset: All eMERGE array data including legacy data have been imputed. Finished imputed batch and merged datasets have been posted to the Aspera outbound directories and will uploaded to DNAnexus.
  - Currently, the merged chromosome VCF set is 1.3 TB, containing both dosage and most-likely hard called variants. This is where the allele dose is the 0-2 sum of the two 0-1 phased haplotype allele's statistical dose, and the hard call is where the call of each 0-1 haplotype exceeds a 0.5 threshold, allowing for greater confidence in the hard called variants.
  - 250 withdrawn individuals were removed from the combined set.
- Files include:
  - VCF by chromosome dosage and most-likely called
  - PLINK by chromosome dosage files
  - PLINK by chromosome most-likely (hard) files
- For the *Genotype Chip Array Batch R^2* graph, R^2 represents the theoretical genotype correlated with the imputed genotype. All data are available, allowing users to choose their own R^2 for a particular analysis. David agreed to provide an analysis of the distributions of the various R^2 measures along with a correlation matrix showing how they relate to each other.
- Individuals can download the data from Aspera. DNAnexus is a powerful platform that can be used to analyze and process the data, as well as be used as a tool for cross-collaborative analysis.
  - Access to Aspera: https://aspera.gs.washington.edu/
  - Access to DNAnexus: https://www.dnanexus.com/
  - Webinars and tutorials for DNAnexus can be found on the Genomics workgroup page (login required): https://emerge.mc.vanderbilt.edu/workgroups/workgroup-genomics/
- PGRNseq multisample: Realigned BAMs and multisample VCF completed and posted to Aspera and DNAnexus. *CYP2D6* from Mayo and *HLA-B & HLA-DQB3* will be provided to the Network.
- ACTION ITEM: The Genomics workgroup will develop and release guidelines on best practices to manage data (e.g. VCFtools, PLINK common commands).
- ACTION ITEM: Correlations between the mean, median, etc. R^2 should be released by the CC to the Network for the eI-III imputed merged dataset.

Next Steps for Sequencing Data | *Megan Roy-Puckelwartz (Northwestern), Patrick Sieiman (CHOP), Larry Babb (GeneInsight) & Darren Ames (DNAnexus)*

- Whole Genome Sequencing (WGS) Data
  - Short read sequencing data are currently available for 1800 whole genomes: 900 from Northwestern and 900 from CHOP.

- The question was posed to the Network 'Whether these data are useful and what formats would be required to maximize their use?'
  - It was noted that BAM files are 80 TBs and VCFs may be more useful as they are smaller.
  - A full description of the data will be shared after which the Network will be surveyed to determine usefulness of the WGS data.
- It was noted that if there is not an eMERGE IV, there will be an attempt to maintain the data on DNAnexus, but this cannot be guaranteed.

- ACTION ITEM: CHOP and Northwestern will send out a formal description of the demographic and phenotype available on the participants with Whole Genome Sequencing data to the group in order to gauge usefulness.
- ACTION ITEM: The CC will formalize and collect metrics surrounding need and usefulness of WGS data to the Network. The PIs will be asked to suggest specific phenotypes they would like to study on the WGS data and the CC will collate these responses and return them to the Genomics group in September, 2017.
- Case Results Delivery Update
  - The XML file structure and content has been finalized.
  - Baylor is batch processing finalized reports to produce identified XML files at same time as uploading to the De-identified Case Repository (DCR).
  - DCR is generating a consistent form of all XML files.
  - LMM/Broad XML files are generated and uploaded to DCR as they are signed out (not in batch).
  - Baylor identified XML Files will have some extensions to the common XML format produced by the identified LMM/Broad & DCR files.

Overview of Kaiser Permanente Research Bank | *Eric Larson (KPW/UW) & Nazneen Aziz (KPW/UW)*

- Dr. Aziz presented that Kaiser Permanente Research Bank is the nation's largest not-for-profit health system. There are approximately 11.6 million members from the District of Columbia and eight regions: Washington, Northwest, Northern California, Southern California, Hawaii, Colorado, Georgia, and Mid-Atlantic States.
- Over 50% of Kaiser Permanente members are minority race-ethnicity, with each of the major race-ethnicity groups well represented. There is an increasing representation of lower income members with the Affordable Care Act.
- Participant population: general cohort, cancer cohort, and pregnancy cohort.
  - The distribution of cancer types among cancer cohort enrollees: 27% breast cancer, 12% prostate cancer, 7% colorectal cancer, 8% leukemia/lymphoma, 4% lung cancer, 3% thyroid cancer, 2% ovarian cancer, and 38% other.
- Kaiser Permanente Research Bank is composed of longitudinal and depth data, electronic health record data, genomic data, and a self-administered health survey to help researchers save time and money assembling cohort data.
- Kaiser Permanente Research Bank plans to strengthen connections between research and clinical care in order to improve the translation of evidence-based research into care delivery. They hope to accomplish this with a large survey to inform return of results and precision medicine strategy and by developing and testing a return of result pilot.

Ancestry, Admixture and Genetic Associates in a Cohort of Mexican Americans | *Iftikhar Kullo (Mayo)*

- The Sangre Por Salud (SPS) biobank comprises ~3000 Mexican American participants. As part of eMERGE III, a cohort of 999 individuals were genotyped. 500 with hyperlipidemia/colon polyps will undergo sequencing on the eMERGEseq panel.
- Associations (GWAS of lipid traits):

- Pre-Imputation: minor allele frequency = 0.05 and site missing rate = 0.05
- Imputation: 1000 Genomes Project phase III (Ad Mixed AMR) using Michigan Imputation server
- Post-imputation: minor allele frequency = 0.01 and INFO score = 0.7 indexes and sites with 3+ alleles were excluded
- After QC, 7.87 million SNVs remained.

- Identity-by-descent analysis discovered related samples (n=114): 54 families (size 2, 48; size 3, 6).
- The study identified loci associated with lipid traits and novel loci for triglycerides.
- Next Steps:
  - Ancestry: Infer the proportion of ancestry in the SPS cohort and Mexican-Americans in eMERGE III cohort using Native-Americans as a reference.
  - Associations: Replication in Mexican Americans in eMERGE III cohort.
  - Admixture mapping of traits that differ between Mexican Americans and European Americans (e.g. height, lipids, glucose, etc.).
  - AdWAS

SPHINX Data Visualization Options | *Brad Malin (VUMC/CC), Josh Denny (VUMC/CC) & David Crosslin (UW/CC)*

- SPHINX has updated search capabilities by rsID and chromosome position. There are plans to add Asian ancestry allele frequencies in cohort in addition to African and European ancestry. These changes will be rolled out in July.
- There has been discussion surrounding visualization of SPHINX genetic data, however a consensus has not been reached. There is a need to determine the most scientifically useful representation while keeping security and privacy in mind.
  - There is interest in individual level phenotypic data and drug response phenotypes
- The Network needs to address the need to preserve participant privacy when adding phenotypic data. This may include the ability to track attacks when users do not register.
- ACTION ITEM: The Genomics group will form a subgroup that will evaluate adding phenotype Record Counter-type functionality to the public-facing portion of the site and other data visualization options.

Clinical Impact of Results Across eMERGE | *Pane*l

- Return of Secondary Findings for HCM | *Mike Murray (Geisinger)*
  - Dr. Murray presented the early findings of returning results related to a Cardiac Hypertrophy.
  - Out of 14 participants, 11 were clinically evaluated prior to results, only three were without evidence of cardiac disease.
  - Two individuals were predicted to have their diagnosis of hypertensive heart disease changed to hypertrophic cardiomyopathy. One doctor did not change the record nor noted opinion of the contribution of the genetic change.
  - The study found that incidental findings do not indicate a conditional diagnosis for the individual. This led to the development of five diagnostic groups based on the presence and absence of genotype and phenotype for a particular condition.

- Investigating the expectations and intended utility of non-diagnostic genomic results for participants enrolled in eMERGE at Northwestern University | *Robyn Hylind (Northwestern)*
  - In order to improve implementation of consenting and returning secondary results from genomic sequencing, 14 participants were interviewed to determine their expectations for results of non-diagnostic genomic sequencing and how they intend to use these results.
  - People desired to receive total information back, however did not always consider the drawbacks.

- The uncertainty surrounding Genomics leads to individuals expecting absolute answers, seeking information without realizing implications, and needing more time to contemplate.
- The study concluded that improvements are needed in implementing consent and communicating of secondary results from genomic sequencing in order to manage patients' expectations.

- The Common Sense Model as a Framework for Exploring Participant Understanding and Use of Genomic Information in the Northwestern eMERGE III Study | *Courtney Scherr (Northwestern)*
  - The intention of the Common Sense Model of Self-Regulation is to explore the construction of risk perceptions and implications of such on coping behavior.
  - Measures of the study include socio-demographics, Early Adopter Scale, current engagement in health protective behaviors, thoughts about current or future illness, and the Common Sense Model Brief Illness Perception Questionnaire.
  - Study findings: Prevalence of illness was higher in early adopters. Participants perceived illnesses causes to be multifaceted. Those with a current condition were lower in emotion and perceived the condition to be less threatening.
  - Results suggest those living with an illness/disease adapt to the situation. This leads to the question of whether those identified as high-risk due to a genetic mutation experience a similar adaptation.
  - In the future, they plan to produce a longitudinal examination of the Common Sense Model constructs and behavior change over time. Plans also include exploration of how the CSM constructs interacts with other constructs and variables such as the Tripartite Risk Model, reactance and/or threats to identity, behavior, and communication.

- Returning genomic results in the eMERGE consortium: The how, where, and what of disclosure | *Georgia Wiesner (VUMC) & Kathy Leppig (KPW/UW)*
  - Dr. Wiesner and Dr. Jarvik presented sites process of returning results: what to return, to whom, and in what order.
  - The study protocol included an interview guide regarding planned return of results with representatives of each site.
  - Results show variability in the process of return among sites, specifically when information will be entered in the EHR and when results are given.
  - The processes of notification (e.g. letters, direct calls) and disclosure (e.g. genetic counselor, phone call) were found to differ among sites.
  - Overall there is significant variability in return of results between sites:
    - Institutionally (e.g. IT staff, EMR),
    - Participant population (some sites have nested case control studies or pediatric patients)
    - Phenotypes of participants
    - Receipt of results from labs

Electronic Health Record: an Untapped Re-source for Family-Based Genetic Research | *Scott Hebbring (Marshfield)*

- Dr. Hebbrig presented research using the electronic health record (EHR) to determine family prediction and heritability, Logistic Regression of Familial Relatedness (LRFR), and disease mapping.
- Families are easy to predict in EHR, but still face the limitations of a high false negative rate and the likeliness to capture younger families.
- EHR-linked families can be used to evaluate genetic influences: 1) measuring heritability can be a challenge, and 2) Logistic Regression for Familial Relatedness (LRFR) can separate genetic from non-genetic phenotypes (LRGR and h^2 are likely correlated)

- Using EHR family data, there is an improved statistical power for genetic mapping with additional recruitment or genotyping.

Genetic assessment of neuropsychiatric and metabolic comorbidities among autoimmune disease patients | *Lynn Petukhova (Columbia)*

- Dr. Petukhova progressed through three topics: 1) Comorbidities in autoimmune disease: immune system contributions to other health conditions; 2) Mendelian Randomization: leveraging genetic data to obtain unbiased effect estimates and insight into causal structure; 3) NT222 – Genetic assessment of neuropsychiatric and metabolic comorbidities among autoimmune disease patients: study design and preliminary data
- In NT222, Dr. Petukhova and co-investigators hypothesize that disease comorbidities detected with ICD co-occurrences can be biologically validated with Mendelian Randomization, specifically focusing on autoimmune diseases in cohorts of patients with metabolic and neuropsychiatric conditions.
- Future directions for this project include:
  - For significant associations, conditional analyses to look at patterns across autoimmune diseases
  - For comorbid conditions with plausible evidence for shared biology, identify cohort of comorbid patients to identify enriched alleles
  - Studies of causal order conducted with genetic predictors of autoimmune diseases in cohorts of patients with metabolic and neuropsychiatric conditions.
  - Other opportunities for Mendelian Randomization studies in eMERGE data: Additional comorbidities; Exposures; Estimate bias in medical records

eMERGE IV Brainstorming Discussion | *Rex Chisholm (Northwestern/SC Chair)*

- The Network needs to leverage the incredible investment already made in the sequencing and array data collected in the first III phases of eMERGE.
- eMERGE is unique due to EHR expertise; harmonization, scalability and very large merged data sets.
  - Clinical decision support to clinicians requires optimization of the whole environment
- Genomic methods beyond sequencing should be considered, i.e. methylation, software application building to streamline moving data across healthcare systems.
- Focus on high impact conditions where changes are more likely to be adopted in the longrun, and folded into health care.
  - Methods for enabling genomic information to follow patients as they move across the health system
  - Potentially creating a generalizable eMERGE SMART app, allowing data to be transferred and utilized across health systems and providers
  - Addressing physician fatigue with new technologies and data
- Lessons learned from eIII: Make data available as early as possible, use on the shelf panels, actionable variants make up a small percent of results, more time is needed to follow outcomes.
  - Move beyond analyzing a person as one data point, network and family interactions are important
    - Collaborations with the Genotype-Tissue Expressive database (GTEx) (Dr. Nancy Cox)
- Collaborations with large genetic data centers
  - Oak Ridge is analyzing the Million Veteran's program genetic data.

Value of genetics-informed drug dosing guidance in pregnant women: a potential eMERGE network-wide collaboration | Casey Overby *(Geisinger/JHU)*

- Dr. Overby presented a study detailing prescription drug use in pregnancy, importance of maternal genetics in drug prescribing, opinions for obstetric healthcare providers, and research opportunities to recognize a harmful or ineffective drug effects.
- A survey was designed to assess perceived needs of obstetric healthcare providers at John Hopkins University. Healthcare providers were majority female, 94% obstetricians and 2% gynecologists, and completed training 1-36 years ago.
  - Dose adjustments are common in pregnant women, with little outside consultation and use of genetics.
  - Few are unaware of resources for clinical guidance on genetics-informed drug dose use in pregnancy.
  - Most providers would find it valuable to have access to genetics-informed dosing information in the form of mobile app, embedded in the EHR, or desktop app.
- Next Steps:
  - Research opportunities in recognizing harmful or ineffective drugs.
  - Potential areas of collaboration using eMERGE participants and data.
  - Potential use of samples with pregnancy episodes (ICD9 codes 630-679.99, 650) include 254 samples from eMERGE SPHINX data, 582 samples from Hopkins Birth Cohort, and 27,700 samples from JHU CDW.

Workgroup Report-Outs

Phenotyping Workgroup | *George Hripcsak (Columbia) & Peggy Peissig (Marshfield)*

- OMOP funding is likely to be awarded, the Network is cautiously optimistic it will be received this fiscal year.
  - Once sites are on a common data model, the algorithms will be implemented more efficiently
  - This will allow the actual SQL coded program to be shared between sites and linkage to PheKB.
- ICD9-10 mapping:
  - The Network has taken the formal position that diagnoses should be coded as ICD9-CM, ICD10-CDM, and SNOMED.
  - The workgroup recognizes that many sites don't have SNOMED data, therefore are asked to, for now, code at least as ICD9-CM and ICD10-CDM.
- Common variables subgroup was formed to develop a coordinated set of variables that would be used for all algorithms, and potentially stored at the CC to reduce the burden on site from having to pull commonly used variables multiple times.
  - Discussion concerned use of lipid variables, PIs were concerned data would be shared prior to site specific analyses.
  - Concerns also centered around the fact that some suggested variables may not be as universally needed, e.g. White Blood Cell and specific lab that would not be applicable across studies
  - Interested parties are encourage to join the monthly subgroup calls which occur the first Monday of each month at 12pm EST (https://global.gotomeeting.com/join/682083997)

ACTION ITEM: CC will send out phenotype progress grid prior to each phenotyping meeting with the request site updates, as well as when phenotypes are announced.
ACTION ITEM: Phenotyping group will finalize list of common variables.

Prioritization of Unscheduled Phenotypes | *George Hripcsak (Columbia) & Peggy Peissig (Marshfield)*

- The goal is to focus on the phenotypes that are categorized as more simple to implement
- Eosinophilic esophagitis will be at the top of the unscheduled phenotype list, as it can be implemented in both adults and pediatrics and there is interest among sites, specifically Geisinger and CCHMC.
  - Eosinophilic esophagitis will still not be scheduled at this time, but as phenotyping progresses it will be the next phenotype incorporated into the schedule if there is effort available

Clinical Annotation | *Gail Jarvik (UW) & Heidi Rehm (Partners/Broad)*

- The workgroup discussed studies of mosaic variants in older patients, process of reporting copy number variants, possible manuscript proposals.
- Several large cohort studies of individuals non-selected for cancer showed increased number of mosaic variants in cancer related genes in older individuals. Clinically the data shows an absolute risk for hematological cancer of 0.5-1% per year.
- In order to determine whether to report results of mosaic variant findings the group chose to order a skin biopsy to confirm somatic or germline, then if it shows somatic, they will follow up with a CBC.
- Process of assessing CNVs and reporting
  - CNV calling using NGS-based algorithm (sequence coverage)
  - Additional visual review (manual)
  - Identify variants for confirmation by orthogonal methods (ddPCR)
  - Assess CNVs and report (likely pathogenic/pathogenic)
- The group is unable to identify a single variable to explain fail rates in regard to copy number.
- CNVs made up approximately 2% of returnable variables.
- Future manuscripts:
  - Harmonizing the Sequencing and Interpretation Approach for the eMERGE III Return of Results Program manuscript proposes to bring all steps sequencing center went through to develop to one paper.
  - Incidental and secondary findings in eMERGE-III participants study is of two sequencing centers collecting all reporting variants unrelated to an indication, and depositing into REDCap incidental findings database.

ACTION ITEM: The Clinical Annotation workgroup will develop and execute Manuscript Concept Sheets concerning harmonization of sequencing as well as incidental secondary findings.

Genomics Workgroup & Geocoding Supplement | *Megan Roy-Puckelwartz (Northwestern), Patrick Sleiman (CHOP) & David Crosslin (CC/UW)*

- Approximately 250 individuals who had withdrawn consent were removed from the imputed eMERGE I-III data set
- In addition to the full merged eI-III data set, site specific imputation data will be available on DNAnexus
- R^2 values are difficult to select in rare variant space (MAF<0.01)
  - R^2 ~0.4 may be appropriate for rare variants depending on ancestry
  - Cutoff thresholds are dependent upon platform and ancestry of subject
- PGx data set is available on both Aspera and DNAnexus
  - Two additional HLA genes were called, and those data will be available shortly
  - PGx is examining how people are returning the PGx SNPs as part of eMERGE III
  - A draft REDCap survey was created to capture these data and will be shared with the workgroup.

ACTION ITEM: If any Network member or site had downloaded the previous dataset, please either re-download or remove the ~250 subjects. The CC will provide lists of IDs of withdrawn subjects as needed.
ACTION ITEM: The Genomics workgroup will provide a "Starting Guide" document, which will cover filtering for general GWAS (MAF, IBD, R^2, platform/site, ancestry). Specifically concerning guidelines for R^2 usage.
ACTION ITEM: The Genomics workgroup will provide example code on DNAnexus for users to "start" analysis.
ACTION ITEM: PGx group will send draft survey to sites for review that focuses on examining how sites capture the return of PGx data in the eMERGEseq cohort.

Geocoding | *Patrick Sleiman (CHOP) & Abel Kho (Northwestern)*

- The group began work in September 2016 with the idea that where eMERGE participants live is a proxy for their exposure to environmental variables.
- Geocoding converts an address into latitude/longitude, and then links to variables that are related to a person's location on a map. The process involves the use of identifying information which must be conducted in a HIPAA and IRB compliant manner.
- Three current multi-site approaches were developed to compile data between sites:
  - Central site to conduct geocoding and analyses requires sharing confidential data upon individual sites' IRB approval.
  - Cloud-based computing approach requires BAA agreement with hosting provider and institution and expertise in cloud computing.
  - De-centralized approach in which sites independently geocode and derive geomarkers before removing identifiable information and provide de-identified dataset.

Decentralized and HIPAA Compliant Geocoding to Characterize Community and Environmental Exposures for Multi-Site Studies | *Patrick Ryan (CCHMC) & Richard Cole Brokamp (CCHMC)*

- Decentralized Geomarker Assessment for multi-site studies (DeGAUSS) is the fourth novel approach and software developed with the following objectives:
  - To create a standalone application to geocode and derive geomarkers
  - Ensure identifiable data never leave the local machine complying with IRB/HIPAA
  - For software to be reproducible and run the same on all devices at all sites
  - Docker is a software containerization platform that wraps software into a complete file system containing everything needed to run: code, runtime, system tools, and system libraries. The purpose of using containers is to ensure software runs the same regardless of environment by running directly on the system infrastructure rather than relying on guest OS or a virtual machine.
- Three containers were developed for eMERGE proof of concept

- Census tract
- Census tract level median household income
- Distance to nearest highway

- The DeGAUSS method was compared with other geotracking software and shown to have a median household income correlation of 0.97, median distance error of 40 meters, and a median distance to roadway error of 1 meter.
- The study concluded that geocoding and deriving geographic and environmental data for multi-site studies poses unique challenges and DeGAUSS software helps to overcome those hurdles.
- Next steps for the Geocoding Supplement:
  - Develop DeGAUSS into a resource for additional variables tied to local geography.
  - Explore network wide linkage to novel variables (e.g. air pollutants).
  - Currently the group is working on two manuscripts: A methods paper (DeGAUSS), expected to be submitted in July 2017; and a gene by environment manuscript. All sites are encouraged to participate.

ACTION ITEM: Run DeGAUSS software on remaining sites, where applicable.
ACTION ITEM: Input all Geocoding data in DNAnexus, and work with CC to determine how data could be made available to the Network.

Outcomes Workgroup | *Josh Peterson (CC/Vanderbilt), Hakon Hakonarson (CHOP) & Marc Williams (Geisinger)*

- The workgroup is finalizing outcomes data collection REDCap forms, which will be collated into one REDCap database and reviewed for consistency in structure and content when applicable across similar phenotypes.
- The Outcomes group is working with the CC and the ROR/ELSI workgroup to generate a REDCap version of the participant survey forms.
- The Outcomes group is collaborating with CSER to incorporate key questions related to healthcare and lifestyle actions taken based on result finding.
- Familial Implications of return of results subgroup: The group is collaborating with ROR and ancillary R01 studies to survey patients to assess cascade genetic testing, specifically investigating clear challenges related to consent and accessibility of data on family of proband.
- Economics: This subgroup is conducting analysis of standardized costs attached to differences in healthcare utilization and outcomes between variant positive and variant negative cohorts. The focus of the workgroup is to measure directly over timeline of grant, and simulated over a patient's lifetime is the focus of an investigator-initiated R01 under consideration.
- Pediatrics: This subgroup is assessing influence of guideline adherence and ADRB2 SNPs in predicting exacerbation of frequency in asthma patients. A draft paper is in its final stage and will be circulated shortly.

ACTION ITEM: The Outcomes Workgroup will finalize the Outcomes REDCap data collection form.
ACTION ITEM: Outcomes and ROR will work with the CC to generate REDCap versions of Participant Survey forms.

EHR Integration Workgroup | *Sandy Aronson (Harvard/Partners)*

- Data transfer format update:
  - XML structure has begun to be exercised
  - In-depth field content harmonization discussions and pragmatic optimization in progress
  - Initial feedback from sites has been garnered
- Manuscripts are in process related to development of the eMERGE Network EHRi infrastructure, clinical decision support (CDS) costs, and expanding the eII Infobutton project.
- The workgroup has shared XML format with the HL7 Clinical Genomics group, as well as publically (https://github.com/emerge-ehri/results-schema).
- The workgroup is collaborating with CSER on Lynch Syndrome CDS Guide. Members of the workgroup are represented across the community: ClinGen, CPIC, CSER, DIGITizE, GA4GH, IGNITE, and ISCC on recent evaluation of genomic infobutton.

ACTION ITEM: A manuscript concept sheet will be circulated detailing the development of the EHRi infrastructure in the eMERGE Network.

ROR/ELSI Workgroup | *Ingrid Holm (BCH) & Iftikhar Kullo (Mayo)*

- Geisinger's MyCodeTM community health initiative report on qualitative semi-structured interviews and deliberative engagement focus groups was reviewed.
  - Specific highlights include patient ROR experience interviews and survey, reasons patients chose not to tell certain family members, focus groups, and family member interviews to gather cascade screening perspectives.
  - Overall, participants reported positive feelings about ROR and cascade testing.
- The process of return of results at the various sites was highlighted and broken down into two main pipelines described below. A manuscript is in process.
  - 7/9 sites follow this protocol: 1) result to site; 2) review by committee; 3) contact participant 4) return by genetic counselor; 5) upload to EMR; 6) inform PCP
  - 2/9 sites follow this protocol: 1) result to site; 2) review by committee; 3) upload to EMR; 4) inform PCP; 5) contact participant; 6) return by genetic counselor or specialist
- An IRB perspectives paper, which describes the IRB experience at each site, is in process.
- Participant survey subgroup has finalized a list of items to ask at all sites (baseline, within 1-month of disclosure, and 6-months post-disclosure). The CC is in the process of developing a REDCap database for the common questions.

ACTION ITEM: The ROR workgroup will detail a manuscript surrounding the process of Return of Results across the Network
ACTION ITEM: The ROR workgroup will detail a manuscript surrounding IRB experiences of each site concerning return of results

Health Care provider supplement | *Ingrid Holm (BCH)*

- The HCP group has completed literature review and interview of HCPs to inform development of the survey (manuscript in progress).
  - A draft of the survey is completed and cognitive interviews are in progress.
  - Development of the REDCap database to house the survey has just initiated.
- Through initial HCP interviews, the group discovered that many providers do not see ethical problems with returning genome sequencing results in screening situation, and there are insurance concerns. The group also gathered data related to HCP perspectives regarding level of responsibility for results and request for further clinician education.
- An R01 was submitted for funding in Fall 2017 to conduct the survey network-wide.

ACTION ITEM: The HCP group will finalize their survey after data analysis of initial findings is completed.
ACTION ITEM: The HCP group will finalize their manuscript concerning their initial survey, interviews and findings.

Summary of Action Items

Coordinating Center/Network

1. CC will send out phenotype phenotype progress grid prior to each phenotyping meeting with the request site updates, as well as when phenotypes are announced. https://docs.google.com/spreadsheets/d/13lWaavMpeSbAVc_agmLUcl1jlL9fuEzweobkRg1cubs/edit#gid=754032392

2. If any Network member or site had downloaded the previous dataset, please either re-download or remove the ~250 subjects. The CC will provide lists of withdrawn subjects.
3. The CC will formalize and collect metrics surrounding need and usefulness of WGS data to the Network.

Phenotyping

4. Phenotyping group will finalize list of common variables.

Clinical Annotation

5. The Clinical Annotation workgroup will develop and execute Manuscript Concept Sheets concerning harmonization of sequencing as well as incidental secondary findings.

Genomics

6. The Genomics workgroup will release guidelines on best practices to manage data (e.g. VCFtools, PLINK common commands).
7. The Genomics group will form a subgroup that will evaluate adding phenotype Record Counter-type functionality to the public-facing portion of the site and other data visualization options.
8. The Genomics workgroup will provide a "Starting Guide" document, which will cover filtering for general GWAS (MAF, IBD, R^2, platform/site, ancestry). Specifically concerning guidelines for R^2 usage.
9. CHOP and Northwestern will send out a formal description of the demographic and phenotype Whole Genome Sequencing data to the group in order to gauge usefulness.
10. The Genomics workgroup will release guidelines for best practices on which R^2 to use for the eI-III imputed dataset.
11. The Genomics workgroup will provide example code on DNAnexus for users to "start" analysis.

PGx

12. The PGx group will send draft survey to sites for review that focuses on examining how sites capture the return of PGx data in the eMERGEseq cohort.

Geocoding

13. The Geocoding supplement will run DeGAUSS software on remaining sites, where applicable.
14. The Geocoding supplement will input all Geocoding data in DNAnexus, and work with CC to determine how data could be made available to the Network.

Outcomes

15. The Outcomes Workgroup will finalize the Outcomes REDCap data collection form
16. Outcomes and ROR will work with the CC to generate REDCap versions of Participant Survey forms. (See # 20)

EHRI

17. A manuscript concept sheet will be circulated detailing the development of the EHRi infrastructure in the eMERGE Network. (EHRI workgroup)

ROR/ELSI

18. The ROR workgroup will detail a manuscript surrounding the process of Return of Results across the Network
19. The ROR workgroup will detail a manuscript surrounding IRB experiences of each site concerning return of results
20. Outcomes and ROR will work with the CC to generate REDCap versions of Participant Survey forms. (See # 16)

Health Care Provider Survey

21. The HCP group will finalize their survey after data analysis of initial findings is completed.
22. The HCP group will finalize their manuscript concerning their initial survey, interviews and findings.

Next Meeting: October 9-10th, 2017 in Bethesda, MD