eMERGE Network: Summary of the Steering Committee Meeting

October 9-10th, 2017 in Bethesda, MD

The Fall Year Three, Phase III eMERGE Steering Committee was held on October 9-10th, 2017 in Bethesda, 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).

Day One: Network Presentations

NHGRI Program Official Report | *Rongling Li (NIH/NHGRI)*

- Rongling Li, on behalf of the NHGRI eMERGE team, introduced Dr. Robb Rowley as the newest member of the team. He is a board-certified Internal Medicine physician with a background in cancer genetics and bioinformatics.
- The eMERGE and Beyond Workshop, organized by the NHGRI to discuss the future of Electronic Medical Records (EMR) and Genomics, will be held on October 30th, 2017 in Rockville, MD.
  - Goal of the meeting: To identify genomic medicine research areas that can fully use eMERGE infrastructures and resources and integrate into other NHGRI programs.
- Rongling provided updates on NIH Policy Awareness.
  - Notice of changes to NIH Policy for Issuing Certificates of Confidentiality effective on October 1, 2017. Certificates of Confidentiality are for NIH-funded research to protect the privacy of subjects by limiting the disclosure of identifiable information.
  - Clinical Trial Funding Opportunity Announcement Policy effective on January 25, 2018, to be detailed further by Lucia Hindorff (NIH/NHGRI).
  - Proposal to update Data Management of Genomic Summary Results under the NIH Genomic Data Sharing Policy investigators are encouraged to submit comments via a public comments page on the NIH webpage.
  - As the GWAS data sharing policy was revised, the standardized institutional certification form has four categories: General Research (GRU), Health/Medical/Biomedical (HMB), Disease-specific, and Other.
    - For data use limitation category Health/Medical/Biomedical (HMB), data is limited to health/medical/biomedical purposes, and does not include the study of population origins or ancestry. These data would generally be made available to any qualified investigator for research on any disease or health condition and these data would not be made available for research on non-disease traits (e.g. intelligence, personality traits) or population structure or ancestral origin without a clear relationship to disease.
- eMERGE III Data Generation Timeline goals include the completion of sequencing by March 2018, Network Phenotyping completion by December 2018, and dbGaP submission completed three months before the end of eMERGE III.
- Nine study sites and CC including Marshfield Clinic were awarded the FY17 OMOP Supplement Award.
- A CTSA supplement award was also given to three sites for single IRB studies.
- Goals of our current meeting included:
  - Update on genomic sequencing status, dataflow, and clinical report
  - Propose approaches for fully utilizing the eMERGE data
  - Share results of ongoing scientific projects, including workgroup reports and response to ESP
- Our next Steering Committee Meeting will be held on January 25-26th, 2017 in Bethesda, MD.

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

- Developments since the prior Steering Committee meeting (June 2017):

- 18,912 of 25,000 samples have been received; 15,182 sequenced; and 4,697 clinical reports issued.
- Eight of 27 phenotypes have been deployed across the Network and the Common variables list is final with data to be housed by the CC for download with approved MCS.
- Additional data elements are added to SPHINX including Asian ancestry allele data, linkage outs to dbSNP, and ability to search by rsID.

Genomic Data Update | *David Crosslin (UW/CC)*

- eMERGE I-III imputed dataset: 78 different genotype platform batches with 83,717 participants. Data has been imputed against the HRC reference and is available through Aspera.
  - Currently, the merged chromosome VCF set is 1.3 TB, containing both dosage and hard called variants.
- Harvard just submitted an additional 10,000 GWAS samples and another 5,000 will be expected shortly. The CC will incorporate these data into the imputed, merged GWAS set after working with the Genomics WG to discuss versioning.
- Imputation paper draft will be released soon and the poster will be presented at the ASHG Conference.
- PGRNseq: Sequencing 84 drug metabolic pathway genes with 9,010 participants. Approximately 60,000 variants have been observed.
  - Adam Gordon is completing the draft manuscript and will be available soon for author comment.
  - BAMs are being archived, to make room for eMERGEseq data.
  - The CC is adding prior clinical associations to the SPHINX website which will be updated on a regular basis
- eMERGEseq: The Coordinating Center (CC) will download the first 15,000 sequenced samples to prepare for dbGaP submission.

CSG Open Discussion | *Heidi Rehm (Partners/Broad) & Richard Gibbs (BCM/HGSC)*

- Heidi and Richard presented the CSG timeline and updates, and answered questions regarding reporting and receiving data.
- Both CSGs have delivered complete clinical reports to sites to begin their own local production development process.
- Questions were raised concerning the variant reclassification process.
  - The Clinical Annotation workgroup or data from other sites will help inform the CSGs variant reclassification.
  - Well phenotyped case/control cohorts may provide statistically significant evidence for changes in classification
  - Discussion of re-classification of variants may be a good topic at the eMERGE & Beyond meeting on October, 30th.
- Baylor's Copy Number Variant (CNV) module has been deployed, all reports will contain confirmed CNV's and addendums are being prepared for the first 2,500 reports.
- Lessons learned: Creating the eMERGEseq platform from scratch was an extended and thoughtful process with all sites contributing. This process, while valuable, delayed implementation of the platform during the initial funding period. Future funding cycles may want to focus on using pre-existing platforms in order to save time in the design and validation phase, allowing sites to receive their data more rapidly and provide time for clinical implementation aims.

Meharry Medical College: Site Update | *Samuel Adunyah (Meharry) & Philip Lammers (Meharry)*

- The study aims of the Meharry Translational Research Center (MeTRC) are to evaluate the genetic differences between those patients with either a history or who are at high risk for breast, prostate, colorectal, or lung cancers in African American populations.
- Enrollment: Currently 300/500 African American participants are enrolled, final enrollment on target for December 2017.
  - The patient population is on average 57 years of age, largely uninsured, 62% smokers, and a BMI of approximately 31.
- In addition to the eMERGEseq panel, RNA analysis and proteomics will also be analyzed.
- Clinical information will be obtained through sociodemographic survey information and past and future health outcomes from the EMR.
- Return of Results (ROR) is planned for mid-2018.

- The site will have a genetic counselor meet with participants with pathogenic mutations and the Primary Care Provider (PCP) will be notified via telephone and letter.
- "Not pathogenic" and "not likely pathogenic" mutations results will be returned via letter to participant and their primary care providers
  - An understanding that a negative result does not mean the participant cannot develop cancer and merely the participant does not have a strong association for the genes tested is essential to research protocol and recruitment plan at this site.
- Pathogenic results will be added to the EMR.
- Methods for the evaluation of ROR and the impact on future testing and treatment are under development.

- MeTRC was able to obtain approximately a 90% recruitment rate of subjects approached. This may be due to the research coordinator (an MD) clearly explaining the study and downstream consequences. Including returning negative results is a way in which to keep participants integrated and interested in participating.
- MeTRC noted that there is a mechanism in place to obtain feedback from community via the Vanderbilt-Meharry Alliance, led by Consuelo Wilkins.

Implementation of an Ancillary Genomics System at Northwestern | *Rex Chisholm (Northwestern)*

- An overview of Northwestern's eMERGE III study design and the implement of Ancillary Genomics System was presented.
- The Ancillary Genomics System stores data, creates a repository, completes analyses, and can exports results to the EHR.
- Northwest's eMERGE III study has a two-armed design. Arm A returns the suggested ACMG actionable genes as well as genes participants choose to have returned. Arm B returns all actionable results to patients.
- A survey with yes/no questions about participant's preference to receive information related to particular conditions was given at enrollment. Then, genes and/or individual variants were categorized into option categories.
- The role of an Axillary Genomics System at Northwestern:
  - Suppresses results that should not be shown to study team by default. Clinical provider request an exception to this if needed. This will require the clinician to obtain a log-in to process which is audited and available for review.
  - Creates PDF reports for the participant and the provider and uses HL7 to send these results to EHR.
  - Reprocesses results as knowledge changes after end of study period.
- DocUBuild integration has allowed for edits to be automatically passed to the Axillary Genomics System for both HL7 and PDF, reducing the need for multiple editing.
- Return of Result process for study:
  - Return of result is prioritized for LongQT and hypertrophic cardiomyopathy finding.
  - Northwestern is holding case conferences on pathogenic/likely pathogenic results with cancer, cardiac, familial hypocholesteremia, and endocrine cases (including EHR review).
  - Available XML results are planned to be returned over the next two months.

New Clinical Trials NIH Policy | *Rongling Li (NIH/NHGRI) & Lucia Hindorff (NIH/NHGRI)*

- Lucia Hindorff summarized changes related to NIH Clinical Trials policies, provided an overview of selected implementation details, and worked through selected case studies (grants.nih.gov/policy/clinical-trials/case-studies.htm).Policy guidance and resources for investigators are available here: ClinicalTrials.gov.
- NIH defines a clinical trial as a research study in which one or more human subjects are prospectively assigned to one or more interventions (which may include placebo or other control) to evaluate the effects of those interventions on health-related biomedical or behavioral outcomes.
- Effective on January 25, 2018, all grant applications and contract proposals involving one or more clinical trials must be submitted through Funding Opportunity Announcements (FOAs) or Request for Proposal (RFP) specifically designated for clinical trials.
- The program officer should be consulted if there are questions when filling out the FOA. The institution applying for the funding is responsible for determining if the project meets the definition of a clinical trial.

- The goal of these changes are to present key information to reviewers and staff in a consistent format and aligns with ClinicalTrials.gov for future data exchange.
- Return of genomic sequencing is considered an intervention.
- eMERGE IV would likely meet the Clinical Trial definition if it involves return of genomic result.

Analyses of EHR-derived clinical phenotypes to facilitate the diagnoses and characterizations of genetic disorders | *Chunhua Weng (Columbia)*

- A challenge in genomic medicine is that with a low diagnosis yield many patients receive negative results, but are actually positive and have the disease. 25%-30% of patients are diagnosed after sequencing.
- To increase yield and expedite clinical report generation, a phenotype first approach can be used instead of relying solely on sequencing technology.
- Phenolyzer combines a list of phenotype features with detailed characteristics of disease with public knowledge of phenotype and genotype to rank diseases. The resulting gene list is compared to prioritized variants to find the casual gene.
- The study's hypothesizes that given the EHR's high-level phenotypic information, a pipeline can be built to extract the phenotype and connect to Phenolyzer to accelerate diagnosis of genetic disorders.
- A pilot study was conducted to compare the performance of the EHR Phenolyzer system to expert derived diagnosis.
  - Using phenotype information alone group was able to prioritize gene for clinical decision support, and secondary site validation by Mayo Clinic yielded similar results.
  - Extracted all phenotype terms from published case report, imputed to Phenolyzer system and were about to identify casual gene and confirm the diagnosis.
  - Compared ICD-9 diagnosis codes with EHR notes used by Phenolyzer. ICD-9 codes were found to give less specific phenotypes and were not as effective to be used to prioritize genes.
- In conclusion, the study proves the feasibility of using NLP to extract deep phenotype from EHR narratives and the performance of EHR-Phenolyzer is comparable to that of expert-phenotyping followed by Phenolyzer.

Electronic health records elucidate complex relationships between genetics, the anatomy of the eye, and the disease | *Christopher Bauer (Geisinger)*

- The study connects the first batch of genetic data on 60,000 genotyped patients with 21 years of EHR eye exam data.
- Eye conditions are highly heritable and incurable, but therapies may slow progression if detected early.
- The goal of the study was to utilize EHR data to expand the phenotype scope of eye disease.
- EHR derived ocular traits were extracted (distance visual acuity, spherical refractive error, cylindrical refractive error, pupil dark size, pupil bright size, tonometry, cup-disc ratio, keratometry, and central corneal thickness).
- Disease analysis was based on ICD codes.
- In conclusion, the study found proof of principal high-throughput use of eye exam data and novel genetic loci were identified such as cup to disc ratio, corneal opacity, and macular puckering.
- Future directions of this research include further phenotypic refinement, consideration of additional EHR data, longitudinal measures, and rare variant analyses.

Phenotype risk scores from electronic health records discover patients with unrecognized Mendelian disease patterns | *Joshua Denny (VUMC)*

- The study addressed the idea of co-occurrence of Mendelian diagnoses in complex traits.
  - GWAS findings are over represented in Mendelian genes.
- With the use of EHR data Human Phenotype Ontology (HPO) can be mapped to PheWAS codes to provide clinical synopsis from OMIM.
- Phenotype Risk Score (PRS) calculation was developed from genotype risk score system and used to scale and weight the presence of Mendelian gene phenotypes.

  - The PRS were all created a priori
- PRS was validated on six clinically-diagnoses diseases in the EHR and a control group of non-diagnosed. The results found cases for five of six diseases. The sixth disease's negative finding was due to the lack of phenotypic pattern/symptoms for the condition.
- The study further analyzed the discovery of Mendelian genetic influences on complex disease using the Exome Array. Sequencing confirm the variants as potentially pathogenic and that the heterozygotes have disease risk.
- Grouping specific terminologies (for example from SNOMED) for PRS in phenotypic clusters can be difficult to establish but adds value
  - It is important to not be tied to one specific phenotypic procedure
- Component phenotypes were weighted based on ancestry in order to take into account racial and ethnic related phenotype effects
- In conclusion, PRS use may help reclassify variants as pathogenic or likely pathogenic, could have clinical impact, and potentially can assist with early diagnosis.
- Future plans for the study include propose running PRS tool in the eMERGEseq dataset.
  - There is potential for adding in information surrounding the pathogenicity of variants, however it is not clear how it will be incorporated into clinical tests.
  - Examining novel associations between clinical presentations and genetics is a critical component of diagnosis.

Implementation of precision medicine in a Latino-serving community health clinic | *Gabriel Shaibi (Mayo)*

- Gabriel Shaibi presented Mayo's work with Mountain Park Health Center (MPHC) to implement precision medicine in the Latino community.
- The Latino population is the fastest growing in the United States. Latinos are disproportionally impacted by chronic diseases such as obesity, type 2 disease, and cervical cancer. The high cases of cervical cancer are not due to genetic linkage but rather the lack of screening.
- Return of Actionable Variants Empirical (RAVE) project consent process includes:
  - First, participants from Sangre por Salud Biobank are selected that are phenotyped for both hyperlipidemia and colon polyps.
  - Second, selected individuals are brought back and asked to consent for sequencing and return of actionable results.
  - Third, a questionnaire is given with questions regarding health history and literacy, reasons for participations, knowledge of genomics, and expectations of participation.
- Participants were given a 15-minute video about the purpose of the project before meeting with research coordinator, prior to consenting.
- Return of Results (ROR):
  - Changed language to participant from negative result to non-actionable result.
  - Non-actionable result is given to participant via letter and phone call, also a copy of the negative result letter is uploaded to EMR.
  - Actionable result return process is determined by the RAVE review of the created participant profile.
  - If an actionable result is returned, a letter is sent to the patient and provider that an important result is available.
- Future plans for the study includes finalizing the actionable Return of Result processing (with Community Advisory Board and physicians at Mountain Park Health Center), returning initial eMERGEseq samples, analyzing baseline surveys, and developing interviews for ROR follow up.

Day Two: Network Presentations

Comments from ESP

- The ESP Chair thanks eMERGE Leadership and Coordinating Center for generating the October 2017 ESP Packet. The ESP commends the innovation still occurring within the Network, and reaffirms that the Network is used as a model for the field.

eMERGE Network Overview: Priorities and Goals, Review Progress of Prior ESP Recommendations and Best Practice Topics | *Rex Chisholm (SC Chair, Northwestern)*

- Specific aims for Phase III include: to sequence and assess clinically relevant genes presumed to affect gene function in ~25,000 individuals, assess the phenotypic implications of these variants, integrate genetic variants in EMRs for clinical care, and create community resources.
  - The Network features seven main working groups. eMERGE investigators applied for supplemental funding over the past year, per the ESP's recommendation, to embark on the OMOP informational model for eMERGE phenotyping, the Health care provider survey, geocoding the eMERGE GWAS dataset, and development of resources to facilitate single IRB review for multi-site research.
- The Network now has data on over 110,000 participants, and has developed informatics tools to harness this data.
- The Network has 633 total projects published through August 2016, as well as 119 ongoing Network-wide projects and 9 site-specific projects. There has been in total more than 1100 external downloads of eMERGE data from dbGaP as of August 2017.
- The Network is collaborating with the FDA on a project investigating hip arthroplasty data in the eMERGE GWAS dataset and related outcomes controlling for race and sex..
- The Network has addressed the ESP's recommendations from the April 2017 ESP Conference Call:
  - Capture the variability among the different sites and report differences to the scientific community. The ROR/ELSI Workgroup is implementing two Network-wide projects. The ROR methods survey capturing disclosure plans and placement in the EHR, how and why results are returned, and availability of genetic counselor. The Healthcare Provider survey is determining the type and role of, an HCP in understanding and communicating sequencing results. A Network-wide IRB project is also near completion, focusing on variability and logistical challenges site encountered working with their IRB review for genomic medicine studies.
    - ROR/ELSI and Outcomes Workgroups work together to document the process of return of results, specifically looking into the mechanisms of ROR and evaluating clinical, psychosocial and familial screening outcomes.
  - Use of eMERGE research data. There is historical data on ~110,000 participants, and the new sequencing data will feature ~25,000 participants. The Network is also developing 27 phenotypes. It is noted that there is much overlap in use of the datasets for each phenotype.
  - Surveys and studies comparing participants who receive or intend to receive results. The ROR/ELSI Workgroup has projects addressing impact of ROR to both Participants (baseline and 6-months) and Healthcare Providers, noting that these are focused on return of positive results, as only a few sites are returning negative results and only one is returning variants of unknown significance (VUS). This survey data will be integrated with data the Outcomes Workgroup is abstracting from the EHR.
  - Addressing changes in variant classification over time. Both CSGs reanalyze variants if seen in additional cases or new evidence arises leading to a variant reclassification. Baylor and Partners/Broad have processes in place to address reclassifications.
  - Phenotyping and EHR integration updates. The EHR Integration Workgroup is submitting a manuscript to JAMIA, which documents the flow of sequencing result data through the EHR. Sites have initiated progress on all aspects of EHRI, including report retrieval, finalizing EHR report storage strategy, and site requirements to return results to clinicians and participants. 8/27 electronic phenotype algorithms have been deployed Network-wide.

The Terms Associated with Genetics test (TAG): A new genetic literacy test piloted in the eMERGE III Columbia cohort | *Hila Milo Rasouly (Columbia)*

- Health literacy is critical to a patient's autonomy, ability to understand health-related information, doctor-patient relationships and overall health. Genetic literacy is expected to be similarly important, thus there is a need for a rapid, self-administered genetic literacy test.
- Columbia's objective with TAG is to develop a genetic literacy test, validate it in their site's eIII cohort, and assess the association between low genetic literacy and 1) patients' expectations form genomic medicine, and 2) patients' understanding of key points discussed during the informed consent process.
- Columbia met its objective by developing a 2-minute self-administered genetic literacy test, which was administered to 777 individuals. Low TAG was found to be a significant predictor of unrealistic expectations from genetic testing, and misunderstanding of the Genetic Information Non-Discrimination Act (GINA). These findings are independent of education, health literacy, and numeracy.
- Based on TAG, 34% of the participants have low genetic literacy. Significant predictors include: genetic knowledge (composite score based on four genetic knowledge questions), self-declared genetic knowledge, education, numeracy, and questionnaire language.

Clinical Sequencing Center Updates

Clinical Sequencing Center Update: Partners/Broad | *Heidi Rehm (Partners/Broad) & Hana Zouk (Partners/Broad)*

- P/B reviews site reporting preferences, and notes that depending on the site's needs, there are additions or exclusions from the accepted consensus list generated by the Clinical Annotation Workgroup in collaboration with the CSGs.
- Sequencing is on track. 9053 samples have been received from four sites (KPW/UW, Geisinger, CCHMC and Harvard), 6264 samples have completed sequencing, and 2087 reported cases have been completed (200 positive, 1712 negative, 175 inconclusive).
- Indication-based returnable results: 1.45% positive (n=38), 91.85% negative (n=2399), and 6.70% inconclusive (n=175). Non-indication-based consensus returnable results: 5.93% positive (n=317) although note that this percentage is slightly skewed due one site with sample selection based on suspicious genotype, 94.07% negative (n=5033).
- Pharmacogenomics Reporting: Seven genes and 20 SNPs. Report content in addition to lab and patient info will feature CPIC level A genes (variants for which guideline with dosing recommendations exists), minimum information for diplotype and category (e.g. metabolizer status), and accompanying supplementary tables (e.g. translation tables with genotype/coordinates, phenotype interpretation, dosing recommendations as per CPIC). To date, P/B has issued 1221 PGx reports for UW and 10 PGx reports for CCHMC's adolescent cohort.
- Variant Harmonization: CSGs exchanged all previously reported variants in eMERGE genes prior to pre-harmonization. The CSGs to date have encountered 23 variants with discrepancies (nearly half cardiovascular), and by reviewing and applying a few various processes, they have achieved consensus on 20 variants with three pending.
- ESP COMMENT: More public clarity and transparency on the rationale for the decisions made during the consensus process is needed. How and why is something defined as negative and actionable? Conceptual clarification of what is positive, what is negative.
  - P/B submits decisions to ClinVar in order to make knowledge public and uses ClinGen gene-disease validity framework.
- ESP COMMENT: What is the process to achieving consensus at the calling level? This should be in the public domain. Lessons learned are important to publish on as well.
  - Sites are also actively involved in this iterative harmonization of variant classification process.
  - Geisinger has data on how many participants' exomes were examined in order to populate their cohort. They had the opportunity to conduct selection bias for this phase as whole exomes were on file. These metrics could be useful when writing a manuscript on the harmonization process.

- ESP COMMENT: Why are negative results not being returned?
  - Due to site-specific decision based on their cohort of interest and funding.

Clinical Sequencing Center Update: Baylor | *Richard Gibbs (BCM/HGSC) & Eric Venner (BCM/HGSC)*

- Baylor has received ~9900 samples from sites, and sequenced ~9000 samples to date.
- Indication based returnable results: 3.4% positive (n=52), 96.6% negative (n=1481). Non-indication-based consensus returnable results: 3.4% positive (n=116), 96.6% negative (n=3331). Non-indication-based site-specific returnable results: 2.8% positive (n=44), 97.2% negative (n=1544).
- Beginning October 2017, CNV calls were confirmed, all reports moving forward will contain CNV calls, and will amend prior reports over next ~6 weeks.
- Baylor has built two new tools for eMERGE. Their goal was to design a tracking and metrics dashboard, develop a research tool for mining and analysis, and develop an access tool for reports and sequencing raw data. From this emerged eDAP (eMERGE Dashboard and Analysis Portal) and eCAP (eMERGE commons Access Portal).
- A manuscript is in development to document the creation of the eMERGEseq platform, NT244, "Harmonizing Sequencing and Clinical Interpretation for the eMERGE III Return of Results Program"
- ESP COMMENT: DNAnexus, how has this been as a tool in terms of data, data analysis, surprise costs?
  - As DNAnexus is not inherently designed for the novice computer user, DNAnexus in collaboration with the eMERGE Genomics Workgroup is providing quarterly training sessions for the Network hosted by the eMERGE CC and building tools and dashboards for the Network.

Workgroup Progress Reports

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

- Two manuscripts: 1) NT202: Incidental and secondary findings (Ifs) in 9,000 eMERGE participants; 2) NT244: eMERGEseq paper will combine work into a larger Network paper that spans the work of the CSGs and Clinical Annotation Workgroup to date.
- The group noticed that although sites sometimes appear discrepant, it is due to a difference in terminology between sequencing centers. Per a recent ClinVar lab survey, there is variance amongst labs as to what terminology is used for low penetrance variants in genes for Mendelian disorders. The Network can contribute the experience of clinicians and participants, specifically what these two groups take away from the different terms being used.
- The workgroup will take into consideration how often likely pathogenic (LP) variants change, as the return of LP variants goes against guidance from the ACMG. There seems to be new evidence being generated, as well as general re-review of LP variants, that will either cause significance to either decrease or increase. Workgroup may revisit decision to return LP variants in response to frequency of LP variant reclassifications.
- ESP COMMENT: Do you have the ability to look at CSG difference in variant calling over time?
  - Both CSGs are conducting proficiency testing. Variant calling is quite accurate, specifically on this panel, which is small with deep coverage. Both are doing Sanger confirmation as well.
- ESP COMMENT: The ESP asked the group to speak about structural variants
  - The CSGs are using tools to evaluate structural variants. This area is challenging as not all the tools provide the same call. This field is evolving, and thus presents challenges naturally. CNV detection is an area the CSGs will continue to monitor.

EHR Integration | Casey Overby (Geisinger/Johns Hopkins University) & *Sandy Aronson (Harvard; not present)*

- The EHRI workgroup focused on building consistant XML structure. There have been in-depth field content harmonization discussions and review, sites then provided feedback which has been incorporated into recent revisions (revising XML to include: PGx data, gene coverage, CNVs), and cases will be re-generated to include these new fields.
- The workgroup has written a manuscript on their work, entitled "Empowering Genomics Medicine by Establishing Critical Sequencing Result Data Flows: The eMERGE Example" which is under review at JAMIA.
- The workgroup is working with CSER to launch a Lynch Syndrome CDS Guide. Additionally the workgroup is promoting Infobutton efforts with ClinGen, discussing goals for a genetic data model in OHDSI, sharing XML format with HL7 Clinical Genomics group and the public, and in sharing XML tools publically.
- It will be critical for eMERGE IV to focus its resources on the scenarios and processes it would most like to support.
- ESP COMMENT: For the sites actively returning results, how many are using clinical decision support (CDS)?
  - Most are using some form of CDS, especially for PGx results. The group will survey sites to determine status.
- ESP COMMENT: To address the complexity of implementing phenotypes across sites, it may be worthwhile to try a "SMART on FHIR" program as a decision support tool.
- ESP COMMENT: Is there a role for eMERGE in the front-end decision support arena?
  - The Network's approach is to use genetic counselors.

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

- The Workgroup has completed re-implementation of 13 eI and eII phenotypes. Workgroup has developed 43% of its eIII algorithms, validated 32%, implemented across the Network 32%, and collected data on 7%.
- The workgroup developed a common variables project for the GWAS data to house variables commonly used by sites for phenotypic analyses. These data will be available through the CC, saving the programmers the effort of re-pulling for each data request.
- CC was able to combine collection of common variables with eRC refresh. The workgroup will discuss an extended version of variables on future workgroup calls.
- By leveraging the OMOP common data model (CDM) for remaining phenotypes, workflow will be improved.
- There are 28 Network-wide phenotyping manuscripts in process.
- Workgroup's future efforts including moving towards OMOP CDM, continuing development and implementation of phenotype algorithms, investigating Adjusted Clinical Groups (ACG) Systems which measures health status by grouping diagnoses (Geisinger is piloting), expanding common variables list wherein the use of OMOP may help with this, and continuing to identify workflow enhancements.
- ESP COMMENT: What are the limiting steps to the pace of developing algorithms? Do resources need to be reallocated?
  - The workgroup had to re-run eI and eII phenotypes initially before developing new algorithms. Co-chairs worked with the sites to develop a reasonable timeline and are now on schedule to complete algorithms by December 2018.
- ESP COMMENT: The success of phenotyping is critical to eMERGE IV, therefore the workgroup may want to finish as many phenotypes as possible, or earlier than possible, in order to ensure the likelihood of eIV.
  - If the workgroup were to use simpler or less accurate phenotypes, then it could move faster through its phenotype list. Unfortunately, there exists an iterative process problem. The workgroup attempts to maintain a balance between site-specific positive predictive values (PPV) versus the rest of the Network's PPV, in order to increase the generalizability. This is why the workgroup limited sites to one complex. OMOP may be able to increase the speed of implementation.

OMOP Supplement | *George Hripcsak (Columbia)*

- Supplement awarded to nine sites including Marshfield Clinic and the CC to convirt the network to the OMOP common data model, streamlining the phenotyping process.
- Presently, eMERGE cohorts are stored in disparate warehouses. Each phenotype must be translated by nine different models, adding to phenotype translational time and complexity.
- OHDSI has provided documentation, tutorials, links and other resources, which the Network can leverage for converting to the OMOP model.

- Network will implement OMOP CDM version 5.2 (includes nodes in NLP output), and at least the data model and vocabulary translations.
- Milestones: Sites are asked to convert EMR data to OMOP CDM v5.2 by March 2018. By June 2018, group will begin comparing OMOP CDM approach to site's previous data model, by identifiying test phenotypes that will be run on both the OMOP and standard data modeling across the sites. Additionally, group will participate in the application of the new approach to conduct phenotyping to additional eIII phenotypes that do not require NLP phenotype identifications.
- The CDM online forum is a valuable resource for the Network during the OMOP conversion process.

Genomics | *Megan Roy-Pucklewartz (Northwestern), Patrick Sleiman (CHOP) & David Crosslin (UW/CC)*

- eMERGE I-III merged imputed GWAS data set:
  - Merged chromosome VCF is ~1.3 TB. VCF by chromosome dosage (PLINK) and most-likely called (PLINK) for the 83,717 participants imputed off the Haplotype Reference Consortium using the Michigan Imputation Server.
- PGRNseq, PGx data set:
  - Features 9010 participants, sequenced 84 metabolic pathway genes, and ~60,000 variants were observed.
  - Data has been realigned, genetic variation recalled, and data will be annotated using SeattleSeq plus custom UW annotation. New multisample call is available to the Network via Aspera and DNAnexus.
- eMERGEseq data set:
  - Features ~25,000 participants, and aggregate calling will begin on ~15,000. This data will be added to SPHINX. An interim dbGaP submission is in the process.
- DNAnexus: Cloud computing is avalible to the Network. DNAnexus has developed additional tools for users to work with data and has helped facilitate informational tutorials. Currently, the group is identifying projects to run on DNAnexus.
  - Note that DNA downloads are restricted on DNAnexus, so sites should conduct larger downloads via Aspera.
- Whole Genome Sequencing (WGS): There are 900 adult genomes from Northwestern and 900 pediatric genomes from CHOP. Projects are being developed best utilize these data.
- Future efforts of the workgroup include:
  - Proposing three Network-wide genetic analysis manuscripts for the PGRNseq, HRC-imputed, and eMERGEseq data sets, respectively.
  - Leveraging these data sets and EHR-derived phenotypes to collaborate with other consortia, such as CSER, GIANT, and TOPMed.
  - Continue to provide guidance to the eMERGE CC regarding genetic data activities.
  - Continue to provide guidance to the DNAnexus group regarding genetic and phenotype data organization, and analysis tools.
  - Develop ideas for eMERGE IV in collaboration with the other Workgroups.
- ESP COMMENT: For the WGS comparison project idea, are you referring to exome array or exome sequencing data?
  - The workgroup has access to both data. For Northwestern, not many of the WGS participants were included in the current platforms.
- ESP COMMENT: What value metric does the workgroup intend to use for comparing WGS to Whole Exome Sequencing (WES)? What knowledge are you looking for?
  - Examinig if rare variation is functional; the group would have access to structural variation when compared to WES.
  - Examine the added value of imputation.
  - Having WGS data can also provide information outside the GWAS coding regions.
  - The group can use WGS to harvest SNPs and create risk scores.
- The Network should note: Shefali Setia Verma (Geisinger) gave a DNAnexus tutorial during the Genomics break out session on Day One, which was recorded and posted online here: https://emerge.mc.vanderbilt.edu/4771-2/ or direct Youtube link: https://youtu.be/kr9Ss0uqJKY

Pharmacogenomics (PGx) | *Laura Rasmussen-Torvik (Northwestern & Cindy Prows (CCHMC)*

- PGx has been a part of both eMERGE Phase II and III
  - eII PGx: Sites conducted sequencing on 84 genes. Each site was charged with implementing at least one gene-drug pair into their EHR.
  - eIII PGx: Many SNPs were added to eIII sequencing panel that are for PGx. Co-chairs surveyed sites on their plans for returning eIII PGx results. Only six of the nine sites are returning one or more PGx SNVs.
- Analysis of somatic mutation in PGRNseq data, project update: Led by Ken Kaufman at CCHMC, 5 new candidates were identified.
  - Will examine relationship between validated somatic mutation and phenotype. A manuscript is in progress for this project.
- CYP2D6 genotyping from PGRNseq v1 NGS data, project update: Led by John Black III at Mayo. CYP2D6 region has high homology pseudogenes, hybrid alleles and CNV issues that create obstacles for analyzing data from high throughput short read sequencing technology. Mayo's results are concordant with known results
  - CNVAR will be used to analyze all available eMERGE PGx samples sequenced on PGRNseq. Files have been downloaded and are at Mayo awaiting analysis. Results will be returned to eMERGE and will each feature a unique identifier, genotype (*allele), and phenotype data (metabolizer status).
- The workgroup has published six papers and there are eight manuscripts in progress.
- The workgroup continues to monitor ongoing PGx clinical decision support activity to aid in further analysis of implementation. Workgroup conducts joint projects with Genomics (HLA), Phenotyping (PGx phenotype prioritization) and ROR/ELSI (ROR for 6 ACMG genes on PGRNseq).
- eIII participants were generally not recruited for PGx outcomes, however eIII and eII data represents a huge sample size increase for extraction of relevant EHR-outcomes.
- Workgroup's future directions include collaborating with Genomics Workgroup on additional PGRNseq. VCF analysis (potentially with DNAnexus), and pursuing additional funding to examine eII and eIII projects through the lens of implementation science.
- ESP COMMENT: Replication studies would still be good/publishable while the workgroup is working on "harder" phenotypes.

Return of Results/Ethical, Legal and Social Implications (ROR/ELSI) | *Ingrid Holm (BCH), Iftikhar Kullo (Mayo)*

- The workgroup has many new projects and collaborations, as well as ongoing current projects.
  - Optimizing sIRB review for genomics research project.
  - "Deliberate Ignorance" wherein investigators try to understand why people do not want their genetic information returned after agreeing to have sequencing.
- Collaborations: Workgroup has an ongoing collaboration with the Outcomes, Clinical Annotations, and EHRI Workgroup.
  - Return of variants with low penetrance, specifically from the lens of clinicians and patients, and how they view results featuring these variants and what do they do.
  - Piloting a survey across the Network to ascertain biobank participants' preferences to learn more about the biobank and desire to be updated regularly about the biobank.
- Participant survey project update: Finalized items to ask at all sites (baseline, within 1-month of disclosure, 6-months post-disclosure).
- IRB Perspectives project: Describes IRB experience at each site. Manuscript is under review. Learned many lessons from site interactions with their own IRBs.
- HCP supplement project: Developed and piloted the survey. R01 is in review.
- ROR across sites project: Found that there is much site-specific variation across the Network. Cohorts unselected vs selected for a particular trait, genotype, secondary findings, or age.
- Lessons learned: How to address deceased patients, or those who refuse genetic counseling, addressing results revealing new disease, addressing impact of returning results to different age groups and implications for genomic screening.

- ROR to date:
  - KPW/UW has had the most experience returning results. Returned 20 cases of P/LP.
  - Northwestern: 2 results returned, both cardiac.
  - Mayo: Returned roughly 150 negative reports.
- ESP COMMENT: Can we apply the concept of "deliberate ignorance" to healthcare providers as well, particularly to their unwillingness to use the new genetic data?
- ESP COMMENT: When there is reclassification of variants, is the ROR/ELSI Workgroup studying how providers and patients are reacting to this information change?
- ESP COMMENT: Who is returning results to patients? It would be an interesting comparison to determine how patients understand these results depending on how and by who they are receiving the result.
  - Most sites are returning results through a genetic counselor.
- ESP COMMENT: Is the Workgroup making an effort to compare across sites those that are returning negative results in addition to the positive results? Why are some sites not returning negative results?
  - Georgia Wiesner and Kathy Leppig have already surveyed sites on what they intend to return, and sites will be tracking what they intended to do vs what they actually did. The Outcomes Workgroup will also provide information from the EHR. An R01 that the HCP survey group submitted includes a qualitative portion to help gather impact.

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

- The workgroup prioritizes gene-outcome pairs, then define specific outcomes projects. They will have custom data collection forms for each prioritized phenotype and consensus genes. Additionally they will collect data on phenotypes related to non-consensus genes and SNPs. Outcomes data on Pharmacogenomics is not being priortized due to the frequency of alleles and number of records that would require manual review.
- The workgroup has also been focused on:
  - Manually abstracted elements EHR: baseline (Pre-ROR) diagnoses, post-ROR diagnoses, pre-ROR testing, post-ROR testing, linkage of testing to ROR, consultations, therapeutics, clinical outcomes.
  - Automated EHR data pulls for demographics, general comorbidities.
  - A Mixed methods approach for family history.
- Economics subgroup update: can attach a standardized cost at the end of this project. Wrote an R01 was funded as a joint project with Geisinger (Susan Snyder) and University of Washington (Dave Veenstra) to determine cost-effectiveness of genomics sequencing using clinical scenarios defined by 59 ACMG disease risk genes and CPIC level A pharmacogenes.
- Pediatric subgroup: Looked at outcomes measures for children with asthma who were seen at pulmonary clinics.
- ESP COMMENT: Has the workgroup looked into spead of manual abstraction compared to building an algorithm to pull these outcomes, specifically as it differs for every variant?
  - Workgroup is planning to abstract by hand from the EMR as there are only ~1000 actionable variants.
- ESP COMMENT: Is this possible to automate?
  - Will likely be phenotype-specific, and due to funding constraints in eIII may be a possible for eMERGE IV.

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

- The group successfully geocoded data from the following sites (n=74,594): CHOP, CCHMC, Columbia, KPW/UW, Marshfield, Mayo, Northwestern, Vanderbilt. Data are located on DNAnexus.
- Group has 3 manuscripts in progress, and 1 manuscript describing DeGAUSS submitted to JAMIA.
- Distributed geocoding data using a shared pipeline: focused on the DeGAUSS approach, which allows for standardized pipeline to be run at each site without having to transport PHI.
- Extracted over 300 from the American Community Survey (ACS) and intersected these with all subjects from participating Network sites. A deprivation index was generated based on 5 of these ACS variables for all 73,056 census tracts.

- Adapted Environmental Protection Agency (EPA) gridded atmospheric mode known as the Community Multi-Scale Air Quality Model (CMAQ) with point air pollution measurements to estimate daily spatial surfaces of O3 (ppb) and PM2.5 (ug/m3) across all census tracts from 2002-2012.
- ESP COMMENT: There are many projects that could utilize this data.

Healthcare Provider Survey (HCP) | *Ingrid Holm (BCH)*

- Funded as a 12-month supplement project, and it's now complete except for pilot testing.
- Learned that many HCPs don't recognize any general ethical issues with returning genome sequencing results although there may be a few scenarios wherein there may be concern about the timing of ROR.
- Primary care providers (PCPs) view themselves as the go-to person for communication results and answering questions.
- Regarding clinician education, results can only be considered "actionable" if there are evidence-based interventions that can be offered. PCP education is key, PCPs want to be able to quickly understand what the result means and concrete management advice.
- ESP COMMENT: It would be beneficial to pull all these various data points together as this research has the ability to make an impact on practice.
- ESP COMMENT: Are there differences between pediatric vs adult providers?
  - In general, variance is small but currently the group is only using two adult onset conditions.

Action Items

NHGRI:

- Jyoti Gupta (NIH/NHGRI) will send a request to all eMERGE PIs to resubmit eMERGE III institutional certification for sequencing data.
- Network Leadership to confirm DNAnexus usage guidance for Genomics Workgroup and individual sites.

CC & Network:

- The CC will work with the CSGs organize a teleconference for an in depth discussion of the CSG tools available to the Network.
- The PIs will discuss the possibilities of registering the whole Network as an observational study versus registering individual sites.
- On the Return of Results Outcomes form, the CC will add "oncologist" to "type of specialist" question.
- CC will send request for new Phenotype workgroup publications.
- The CC will provide guidance to the Network on how to access the Common Variables phenotype data via the MCS process.

Meharry:

- Meharry investigators will explore opportunities to publish on their highly successful recruitment rate for their eMERGE study.

Genomics:

- The Genomics Workgroup will determine the scope and timeline for Version 2 of the imputed set, either moving forward now or wait until the spring to incorporate all extra 15,000 Harvard GWAS samples
- Investigators interested in using the Whole Genome Sequencing (WGS) data should contact Meg Roy-Pucklewartz and Patrick Sleiman directly for data transfer.
- The Genomics group will recommend storage options for the WGS data

Phenotyping:

- Phenotyping Workgroup Members are asked to update the eMERGE Phase III Phenotype Prioritization Implementation Grids. (https://docs.google.com/spreadsheets/d/13lWaavMpeSbAVc_agmLUcl1jlL9fuEzweobkRg1cubs/edit#gid=1748512012)
- The CC will work with the Common Variables Subgroup to draft rules for implementing standardized Data Dictionaries.
- MCS authors are requested to emphasize the complexity of the phenotypes the Network is producing in manuscript titles.
- Workgroup will develop an MCS that documents the complexity and lessons learned of implementing the great number of phenotypes.

Outcomes:

- Maureen Smith and Janet Williams will compare Family History form with family history section of Cardiomyopathy Outcomes form and reconcile.

ROR/ELSI:

- Workgroup members are requested to share "negative result" letters, especially for deceased participants.
- Ingrid Holm to explore incorporating "deliberate ignorance" into the HCP survey.

EHRI:

- The EHRI Workgroup will document site status regarding clinical decision support implementation.

OMOP:

- One member from each site is requested to join the OHDSI (www.OHDSI.org) and participate in the forums (http://forums.ohdsi.org)

ESP Executive Session Summary

ESP: Stanley Huff (IMH), Kimberly Doheny (JHU), Eta Berner (UAB), Vandana Shashi (Duke), Howard McLeod – Chair (Moffitt); **NHGRI:** Jyoti Gupta, Sheethal Jose, Rongling Li, Teri Manolio, Robb Rowley; **Absent*:** Gerardo Heiss (UNC), Lisa Parker (U Pittsburgh), Ken Wiley (NHGRI)

- The External Scientific Panel (ESP) met with members of the NHGRI Program Staff in an Executive Session before and after the Steering Committee (SC) meeting on October 10, 2016. Since some ESP members attended the first day of the SC meeting held on October 9, 2016, Rongling provided a brief update on Day 1 of the meeting.
- Overall, the ESP was impressed with the amount of progress eMERGE has made to date. They recognized that the Network is focusing on the impact of eMERGE genomic medicine research contribution to the broader scientific community. The members actively collaborate by publishing more Network-wide papers. They were also encouraged to see that sites are applying for outside funding to spend on eMERGE projects. The ESP recommended the eMERGE investigators document and publish the challenges faced and lessons learned to help the broader scientific community implement genomics.
- The ESP appreciated the sequencing centers' (CSGs) efforts to reach consensus on the gene list and the variant calling. However, they recommended that the CSGs should publish their best practices and rationale for decisions regarding variant classifications and clinical reports. In addition to submitting variant classification data to ClinVar, the CSGs should try to disseminate this information either through the website, other tools, or additional publications.
- The ESP would like to know how many of the sites are returning clinically actionable results through clinical decision support (CDS) tools integrated in to their Electronic Health Records (EHR) system. The ESP recommended that eMERGE

consider a pilot experiment where they can incorporate phenotyping algorithms into a clinical decision-support tool to support both clinical diagnosis and treatment. This will help to link the phenotype algorithm to the decision support.

- The ESP was concerned about the pace of the phenotyping process in eMERGE. They noted that completion of the phenotype algorithm development and implementation is a critical aspect of the overall success of eMERGE. The ESP recommended that the phenotyping workgroup should strive to finish the phenotype algorithms early and create a framework that will speed up this process. ESP suggested that the network establish ways to improve the dissemination and use of the phenotypes developed.
- The ESP had several recommendations regarding the content of the RoR surveys for patients and providers. First, they suggested that the Network should survey providers who are unwilling to use the data and understand their reluctance to incorporate genomic information. Second, the ESP would like to see separate surveys for patients who receive negative results, about their understanding of what their results mean. This information can help the broader scientific community decide whether or not to return negative results. Lastly, they recommended that these RoR surveys should include how providers and patients react to reclassification of variants.
- The ESP suggested that it would help to conduct a general assessment across sites on why some return negative results and others do not, recording reactions from both the patients and providers.
- Since there is site-specific variability in who is returning the results to the patients (genetic counselors, healthcare providers, etc…), the ESP recommended that it would be useful to compare how patients understand their results depending on who is communicating with them.
- The ESP recommended that the outcomes collection process should be automated in the long-term as manual extraction takes up lot of time and resources.
- The ESP stressed that the concept of DNAnexus is the future as it allows everyone to do analyses on the exact same dataset. However, there was a lot of confusion among the Network members related to the costs and datasets for analysis. The ESP expected eMERGE to resolve these issues as soon as possible.
- The ESP recommended that it would be beneficial to design educational programs for healthcare providers as they will be the intermediaries for the patients. If it is within the scope of eMERGE, they also recommended that the Network should become involved with various committees, such as American Academy of Pediatrics (AAP), to help set the curriculum for education of trainees in the medical field. Currently, there is no emphasis on teaching about genomes, exomes, big data, etc. in the medical curriculum.
- The ESP appreciated the efforts of the Meharry and Mountain Park Health Center (MPHC in Arizona) sites to recruit underrepresented populations to the Network. They recommended that the 2 sites should try to collaborate and publish on the similarities and differences between the two groups with respect to consent rate, recruitment and return of results.
- The ESP noted that eMERGE is a vast resource now and that might lead the external scientific community to be too intimidated to contribute data to the Network. They recommended that the coordinating center should advertise the affiliate membership policy and invite others to participate and contribute to the Network.

**Absent at the meeting, but reviewed the ESP packet and provided recommendations.*

ESP Recommendations

To Investigators:

1. The CSGs should publish their best practices and rationale for decisions regarding variant classifications and clinical reports.
2. eMERGE should consider a pilot project that incorporates phenotyping algorithms into clinical decision-support tools and determine impact on clinical diagnosis and treatment.

3. The phenotyping workgroup should strive to finish the development and implementation of the phenotype algorithms early and create a framework that will speed up this process.
4. Regarding the content of the RoR surveys for patients and providers:
    a. The Network should survey providers who are unwilling to use the data and understand their reluctance to incorporate genomic information.
    b. There should be separate surveys for patients who receive negative results about their understanding of what their results mean.
    c. These RoR surveys should include how providers and patients react to reclassification of variants.
5. eMERGE should conduct a general assessment across sites on why some return negative results and others do not, recording reactions from both the patients and providers.
6. It would be useful to compare how patients understand their results depending on who is communicating with them as this aspect differs across the eMERGE sites.
7. The outcomes collection process should be automated in the long-term as the manual extraction takes up lot of time and resources.
8. eMERGE should resolve the confusion related to the costs and datasets for analysis using DNAnexus as soon as possible.
9. It would be beneficial to design educational programs for healthcare providers as they will be the intermediaries for the patients.
10. Meharry and Mountain Park Health Center (MPHC in Arizona) should try to collaborate and publish on the similarities and differences between the two underrepresented groups with respect to consent rate, recruitment and return of results.
11. The Coordinating Center should advertise the affiliate membership policy and invite others to participate and contribute to the Network.

Next Meeting: January 25-26th, 2018 in Bethesda, MD

emerge network
ELECTRONIC MEDICAL RECORDS & GENOMICS