Summary of Steering Committee Meeting: Winter 2019

January 17-18th, 2018 in Bethesda, MD

TABLE OF CONTENTS

JOINT DAY 1: THURSDAY

Networking goals & accomplishments | Bruce Korf *(CSER)* & Rex Chisholm *(eMERGE)*

- Bruce Korf: Clinical Sequencing Evidence Research (CSER)
    - CSER studies the effectiveness of integrating genome sequencing into the clinical care of diverse and medically underserved individuals.
    - There are seven clinical project sites in addition to the coordinating center at the University of Washington, led by Gail Jarvik
    - Projected CSER enrollment counts are expected to peak to 7,000 by the end of the funding cycle.
    - Study organization and logistics, participant recruitment and engagement, sequencing and variant interpretation, and data sharing are the primary challenges.
    - Harmonization and merging of data sets, genomics and other types, continue to be an area of focus for the research community.
- Rex Chisholm: Electronic Medical Records and Genomics (eMERGE)
    - eMERGE is in its third cycle. In this phase, eMERGE is focusing on assessing clinically relevant genes in a 25,000 person cohort, including ACMG 59 genes, assessing the phenotypic implications of variants, integrating genetic variants into EMRs for clinical care, and creating community resources that can be used broadly as well as for anyone that is interested in addressing these challenges.
    - There are seven workgroups to focus on the goals of the Network. These workgroups address the Network-wide issues both in-person in meetings and via teleconferences held throughout the year.
    - The Network has sequenced a panel of 109 genes across all 10 sites. 68 genes and 14 SNVs are on a consensus list being returned by all sites.
    - All genetic samples and clinical reports have been returned to the 10 sites. A paper titled "*Harmonization Clinical Sequencing and Interpretation for the eMERGE III Network*" is in review at the *American Journal of Human Genetics* (The eMERGE Consortium, 2019).
    - Sites are in the process of returning results to participants and beginning six-month outcomes follow-ups.
    - 19 outcomes forms have been implemented with interim analysis targeted for October 2019.
    - The Network has established a genomic and phenotype dataset of over 148,000 participants with informatics tools to harness these data.
    - Approximately 70 phenotype algorithms will have been developed by the end of eMERGE III, and many will have been applied to the imputed GWAS and eMERGEseq datasets.
    - eMERGE moved to OMOP standards and is working to increase natural language processing (NLP) ability.
    - Between 2007-2019, the Network completed 698 total manuscripts and there are 27, 224 cumulative citations.
    - Areas for CSER collaboration could include outcomes, familial implications/cascade screening, ClinGen, and variant interpretation.

Harmonization of outcomes & measures: Selection & performance | Jessica Hunter *(CSER)* & Marc Williams *(eMERGE)*

- Jessica Hunter (CSER): Lessons Learned about Harmonizing Survey Measures for the CSER Consortium
    - Guiding principles for harmonizing survey measures for the CSER consortium include using existing, validated measures and minimizing change to these measures, protecting protected health information (PHI), and maintaining flexibility in the mode of administration.
    - There are site-specific differences in regards to population, setting, approach, and timing between that contribute to difficulties with harmonization.

    - Study population could be adults, children, or parental proxy. Similarly, setting could have ranged from NICU to adult outpatient.
    - Differences include the mode of survey administration, as differences in EMR/EHR and ability to collect information varied.
    - Timing of the harmonized measure depended on the site's study, although the majority were able to meet the harmonized timeline. Most surveys were administered from one week to over eight months.
    - Version control, specifically over email, made harmonization difficult. Subgroup versus main workgroup decisions led to miscommunications. Redoing work when a site began implementation prior to finalization was an issue, as was obtaining feedback and pushback.
    - Issues with cultural competency were a challenge, including acceptability, literacy levels, harmonizing sensitive questions, and shortening validated measures.
    - Workgroups noted that they had different intended uses for the same questions.
    - In the end, the majority of the variables and surveys were harmonized, however given site-specific goals there were still discrepancies.
    - Challenges post-harmonization include combining data across disparate sites, overlapping concepts or projects, needing supporting documentation for IRBs, and data collection, cleaning, and re-distribution.
      Kaiser is leading a validation admin supplement to validate a panel on the selection of harmonized measures for psychometric analyses important to interpreting results.
    - ACTION ITEM: Those interested in joining the expert panel on the selection of harmonized measures for CSER2 should RSVP to Frank Angelo (fana@uw.edu).
- Alanna Rahm (eMERGE): Outcomes Reporting across a Network: Experience & Lessons
    - Sites were assigned to create and fill out condition-specific outcomes forms. Sites added their own measures as needed. Outcomes Workgroup created common patient-reported outcomes forms that also related to chart outcomes.
    - Everyone is filling out a 6-month cohort, but there are differences between sites by how they are filling out, which adds to difficulty in harmonization (ie. mail, online, etc.).
    - Each site had autonomy to create process specific for their site and proposed project.
    - Creating a core set of agreed measures provides some data standard across sites, however it was a long process to agree upon these core measures.
    - Many sites had processes and began baseline collection before the common agreed measures were established. Therefore, backend variable names don't match, which contribute to difficulties in harmonization.
    - Additionally, there are other challenges for outcomes forms. The breast cancer form was originally women-only and Geisinger returned four BRCA results to men. For the original known mutation status pre-ROR, capturing information using binary variables was not efficient. Family outcomes are also difficult to collect.
    - ROR participant survey lessons learned include standardizing data entry, determining standards at beginning if possible or creating a crosswalk to standard variables, tracking sites, timing, common measures, and method of collection (ie. paper, electronic records, phone).
- Ingrid Holm (eMERGE): Health Care Provider (HCP) Survey Process
    - ROR/ELSI Workgroup felt that assessing impact of ROR on health care providers (HCP) was critical. Most eMERGE sites did not have an aim to study the impact of the ROR on HCPs, however the

ROR/ELSI workgroup mission is to assess ethical, legal, and social implications (ELSI) of reporting genetic variants for patients and HCPs.
    - There is little systematic knowledge concerning the views, concerns, and challenges for HCP regarding the desirability, utility, actionability, and meaning of incorporating results of genome sequencing into clinical care.
    - The workgroup applied for an administrative supplement to design a survey to interview HCP and later received an R01 to survey HCP within one month of receiving unsolicited positive genomic sequencing results.
    - Workgroup members found that there are differences in different methods in how results are returned, as was published in the paper "*Physicians' perspectives on receiving unsolicited genomic results*" (Pet et al, 2018).
    - There are differences in both the methods, timing, and the provider (who) returns the result to the participant, as this could include the primary care provider, genetic counselor, geneticist, or another specialist.
    - Timing of returning a result could differ, as they could be placed in the EHR before or after returning to the patient.
    - There are opportunities left to analyze how ROR methods affect impact on HCP, specifically with the responsibility and management of patient results, conversations with patients, benefits and concerns, and overall workflow.
- Group Discussion
    - The differences in ROR methods by sites can be assessed as a variable, which allows the Network to utilize the tools of implementation science to maximize the harmonization.
    - From a statistical perspective, teasing patterns out of the quality data can present challenges. There are statistical processes being developed to analyse the survey data, and use evaluation frameworks for implementation science to put the surveys and data in context.
    - The paper titled "*Ethical Considerations Related to Return of Results from Genomic Medicine Projects: The eMERGE Network (Phase III) Experience*" examines how the sites own IRB limited the ability to set up their project caused issues with harmonization (Fossey, 2018).
    - Sequencing results fall into many domains, and provider attitude may vary based on what is returned.
    - The power of statistics is limited due to the differences and variety of results. Variants can be generally grouped to add power to analyses, but still adds to the complexity of data analysis.
    - Qualitative interviews may help to address these issues, specifically the variability seen in analyses.
    - Even the concepts of 'positive' and 'negative' can cause issues, especially when something is clinically actionable with treatment pathways.
    - CSER is conducting qualitative interviews to address participant experience, however these are not harmonized as of yet. Data integration discussions are taking place. In eMERGE, the Health Care Provider survey has similar qualitative components. the Network has not begun the analyses regarding comparison of compiling themes across the sites.
    - Takeaways
        - While there is heterogeneity, identification of the tools being used between the Networks to examine the commonalities between the two Networks would be beneficial. Specifically regarding pooling data to increase sample size.

        - CSER's upcoming project on the validation of survey instruments may be an area for eMERGE collaboration.
        - The context of how a participant gets a result is different from eMERGE (biobank and research purposes) and CSER (patients with defined clinical problems). Both clinicians and patients would think about those results differently, comparing and contrasting these results could be very impactful.
            - eMERGE sites are primarily adult population and CSER is primarily pediatric.
        - Mayo has 6-month outcomes data for 60 participants. Creating an standard operating procedure (SOP) for initiation of Outcomes forms may be very helpful, both for deployment across a Network, or across many Networks.

Identify collaboration opportunities & projects | Rex Chisholm & Breakout Session Moderators
SUMMARY SLIDES

- Cascade testing of relatives, penetrance, and family communication
    - A CSER/eMERGE joint policy statement about changes to HIPAA is needed in order to acknowledge the familial nature of genetic disease.
    - Harmonization of the family communication survey could be looked at, specifically with CSER looking at eMERGE's survey and experience.
    - Potential projects could look at how family history information is structured in the EHR, penetrance data from cascade testing, and differences in family communication across cultural/ethnic groups.
- Stakeholder engagement
    - Both eMERGE and CSER participate in provider and participant engagement.
    - eMERGE & CSER are using surveys to gather feedback, however the groups have different relationships with HCPs. As stated previously, eMERGE has a primarily adult-based population, and CSERs is primarily pediatric. There are opportunities for CSER to incorporate primary care providers in their surveys.
    - CSER has a lot of community engagement boards; eMERGE does not have this aspect.
    - There is potential collaboration regarding dissemination of results and broad community engagement around the implementation of genomic medicine.
    - Both networks should discuss how to get stakeholder input on sustainability.
- Sequencing - tools for variant classification
    - Networks should share and compare all the classified variants from the CSER and eMERGE sequencing sites in order to determine if use of ACMG/AMP rules is more consistent and standardized across a spectrum of labs (ex. CSER/eMERGE), define where there are still differences in applying the rules, and evaluate consistency of case-level interpretation and effectiveness of CSER-harmonized overall results terms. This is done by defining the rate of interpretation concordance across the full dataset, and identifying and assessing variants with different classifications.
    - A second collaboration project proposal is to develop a structured genetic test result standard for use in eMERGE/CSER EHRs.
- Outcomes and healthcare utilization
    - The most alignment between the Networks comes between the eMERGE Outcomes Workgroup and the CSER Healthcare Utilization Workgroup (CUHEP).
    - The eMERGE and CSER Networks plan on sharing a paper on defining patient status framework.

    - ACTION ITEM: Marc Williams (mswilliams1@geisinger.edu) will share the eMERGE health outcomes forms PDFs with the CSER Network.
    - ACTION ITEM: The eMERGE CC and CSER CC will identify genes (ACMG 59, CDC Tier 1, Arrhythmia) shared across eMERGE and CSER to study outcome definitions and interventions.
    - ACTION ITEM: The eMERGE CC will organize joint calls between eMERGE Outcomes Workgroup and CSER's Healthcare Utilization Workgroup (CUHEP) for collaboration efforts.
    - The Network plans to take gene disease pairs where there are guidelines for intervention and inventory sites for what guidelines are being recommended and compile and compare and assess hypothetical impact on outcomes.
    - The Networks potentially will partner with the ACMG secondary findings committee to identify genes under consideration in which they can provide useful data.
- Scalable approaches to patient provider interactions
    - Networks should describe what 'standard of care' means among projects in the context of clinical settings, and conduct a needs assessment of barriers to uptake across sites.
    - A potential collaborative project could be documenting a model of genetic counseling for predictive genomics.
    - Both CSER and eMERGE had overlap between planned patient and provider surveys. Networks could collaborate in designing and harmonizing surveys for better outcomes.
- ELSI and perceived utility
    - In CSER, there are people undergoing clinical screening, where in eMERGE they are not.
    - CSER is developing measures of perceived utility, involving literature review, interviews, and validation.
    - There was not as much collaboration overlap between the two Networks, but there was a lot of discussion regarding the two networks in deciding what weight we have in giving to the patients' perceived utility.

Closing Remarks | Lucia Hindorff *(CSER)* & Rongling Li *(eMERGE)*

- From the 2017 CSER-eMERGE meeting several ideas were proposed:
    - Family cascade testing, including to develop tools and best practices for implementing successful family cascade projects.
    - Working with National Quality Forum (NQF) to develop quality measures for familial hypercholesterolemia and lynch syndrome, and collaboration on lynch syndrome CDS.
    - Developing clinical education tools for the management of specific results.
    - Collaboration in eMERGE ROR, Outcomes, and ELSI surveys, as well as other projects.
    - Opportunities not yet discussed include research challenges & opportunities in underserved populations, emerging consent standards, and ELSI issues relating to digital innovations.
- NHGRI is hopeful that the Networks can continue to work together to address the points of collaboration identified during this joint session.

eMERGE DAY 1: THURSDAY

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

- The sequencing and clinical reporting is completed for the 25,000 eMERGEseq dataset.

- The record counter refresh is underway for January 2019 including OMOP lab values for autoimmune and white blood cell differentials on all eMERGE datasets, including eMERGEseq, PGRNseq, GWAS, Exome Chip, WGS, and ICD 10 Phecodes.
- The ~15,000 additional Harvard samples for the GWAS set have been imputed and incorporated to produce a 99,000 sample GWAS multisample imputed set.
- As of January 2019, there have been 27,224 citations and over 1,300 external downloads.
- Goals for this meeting are to focus on ROR and Outcomes progress, to demonstrate progress of the seven Network wide milestones, to discuss the lessons learned journal submissions, and to have the Outcomes lessons learned panel.

NHGRI Program official report | Rongling Li *(NIH/NHGRI)*

- The Winter 2019 Meeting is the 35th eMERGE Steering Committee Meeting.
- Meeting evaluation results from the October 2018 eMERGE Steering Committee conference showed that 59% (n=22) found the program official report necessary.
- General consensus is that the official report can be brief but should include the overall program status and new developments, progress from last meeting, and information on funding opportunities.
- Director of NIH's National Institute of Arthritis and Musculoskeletal and Skin Diseases, Stephen I. Katz, MD PhD, passed in December 2018.
- There are several funding opportunities available at the NHGRI.
    - The NHGRI funding opportunities center around diversity in health-related research, mentorship and early career development awards, and computational genomics and data science awards for small businesses.
    - NOT-HG-19-010, which was released on December 19th, 2018. Titled: "*Research Supplements to Promote Diversity in Health-Related Research*."
    - PA-18-372 released on January 12th, 2018 and expires on January 8th, 2022. Titled: "*Mentored Clinical Clinical Scientist Research Career Development Award")* (Independent Clinical Trial Required).
    - PA-18-373 released on January 12th, 2018 and expires on January 8th, 2022. Titled: "*Mentored Clinical Clinical Scientist Research Career Development Award"* (Independent Clinical Trial Not Allowed).
    - PAR-19-061 expires on September 6th, 2021. Titled: "*Genomics and Data Science Opportunities for Small Businesses*."
- The eMERGE III Year 4 funding period is for 22 calendar months, lasting from 6/1/2018-3/31/2020. The "Year 5" supplement cannot really be called an extra year, as it is as supplement off of Year 4. The supplement runs from April 1, 2019 to March 31, 2020.
- An annual final year federal financial report (FFR) must be submitted for each budget period no later than 90 days after the end of the budget period.
- A final FFR shall be submitted at the completion of the award agreement for all awards. For final FFRs, the reporting period end date shall be the end date of the project or grant period. The official end date of Year 4 is now March 31, 2020.
- Annual RPPRs are submitted 45 days prior to the next budget period, however one is not needed April 2019 as the continuation of Year 4 is a supplement, not an official 'Year 5'. The final RPPR will be due 120 days from the period of performance end date.
- ACTION ITEM: NHGRI will provide further guidance on early FFR submission needed prior to the official end of the grant on March 31st, 2020.

- Rongling is stepping down as the NHGRI eMERGE Program Director. She will become the program director of the International 100k Cohort Consortium (IHCC). Robb Rowley will take over as NHGRI Program Director of the eMERGE Network.

Science presentation: Learning from longitudinal EHR and genetic data to better predict 10-Year cardiovascular event | Juan Zhao *(VUMC)*

- Cardiovascular disease (CVD) is the world's leading cause of death in the last 15 years. Roughly 15.2 million deaths and 27% of all global deaths in 2016 were attributable to CVD (Wei, 2018).
- Several models were proposed to predict CVD such as Framingham risk score from the Framingham study, ACC/AHA Pooled Cohort Risk Equations, and QRISK2 from the UK. However there are limitations to these models. They rely on limited risk factors, they use cross-sectional data, and they have lack of genetic information. These factors motivated the investigators to explore a new strategy for CVD prediction.
- The investigators use longitudinal EHR to understand changes in patient's medical conditions and genetic data combined for machine learning and deep learning models to predict 10-year CVD events.
- Investigators compared results to the gold standard of measurement, ACC/AHA, to identify outcomes.
- In addition to the traditional features, such as BMI and SNPS, the researchers included more lab features, including glucose, creatinine level, clinical events, and lab history, to statistically stratify by.
- Individuals with one or more CVD event within the 10-year prediction window were defined as cases, and individuals without any CVD events were defined as controls.
- The sample consisted of N= 109,490 (A: 47.4 +/- 14.7) Inclusion criterion was defined as age equal to or greater than 18, no prior CVD events, and as identifying from european or african ancestry.
- In experiment one, machine learning models all performed the baseline ACC/AHA equation by 30-40%. Incorporating more EHR features improved the performance. Among the two different models of the EHR data, machine learning using temporal features performed better.
- In experiment two, the late fusion model used 10,162 individuals with genotype data and EHRs. The predict score from both models were the input for late fusion model using logistic regression for the final prediction. The late fusion combined the genetics and EHRs improved results of the gold standard.
- Limitations include using BioVu data only, having a chart fragmentation issue with EHR, and the pre-selected SNPs.
- There were several advantages. Machine learning with longitudinal EHR data significantly improved CVD predictions. BMI, creatine and glucose are import predictors, and the researchers added these in additionally to the other traditional predictors. Adding genetic features offered extra benefits. Lastly, maximum, minimum, and standard deviation of lab values are more informative than median or mean.
- Group Discussion:
    - The comparisons may not be fair as ACC/AHA predicts adverse events.
    - NT304: Mayo is working on a manuscript for genetic risk score.
    - The seven-year prediction window is when features are captured. The 10-year prediction window is for outcomes.
    - The number of alleles are used as genetic features.
    - The controls do not have the diagnosis code in the prediction. The cohort was made up of in-patient patients.

Genomic data update | David Crosslin & Ian Stanaway *(UW/CC)*

- The version two of imputed data has been completed and uploaded. 15,494 samples from Harvard have just been added, with another 5,000 just uploaded, which will push the Network from n= 99,185 samples to over 100,000. Data is available on the Aspera server.
- A new publication titled "*Unfolding of hidden white blood cell count phenotypes for gene discovery using latent class mixed modeling*" is available (Hall, 2018).
- Imputation of BEAGLE using the ~100,000 eMERGE array samples is complete, batch based merging and QC is underway.
- There is an MCS NT314 titled "*The Development of an Imputed Structural Variant Genomic Dataset and Association to Neurological and Alcohol Use Disorder Electronic Medical Record Phenotypes with Biobank Scale Subject Ascertainment*" on structural variation imputation using the 100,000 eMERGE array samples. This is open for collaboration or insight from the Network.
- Baylor is currently uploading the last 8,000 of the 25,000 eMERGEseq samples to Aspera server for the second eMERGEseq data freeze.

Sequencing data & clinical reporting update | Richard Gibbs *(BCM/HGSC)* & Heidi Rehm *(Partners/Broad)*

- P/B presented an update on the completed Baylor and LMM/Broad interpretations.
- In terms of cohort sequenced, 9,219 of 25,068 cases had indications (37%), with colorectal cancer (28%), hyperlipidemia (33%), and kidney disease (11%) being the top three phenotypes.
- 201 (2.2%) indication-based returnable results, 1,042 (4.2%) of non-indication based consensus returnable results, and 253 (2.9%) non indication-based site-specific returnable results were found positive.
- eMERGEseq sequencing results of 25,068 cases showed that over 12 reports had two pathogenic/likely pathogenic variants.
- There are three GeneInsight user interfaces for identified and site-specific data, including the Deidentified Case Repository (DCR), GeneInsight Lab interface, and GeneInsight Clinic Interface.
    - The DCR currently contains all of LMM reports and 12,911/14,538 of Baylor reports, as well as variants reported by P/B and Broad and labs that support VariantWire.
    - The GeneInsight Clinic Interface is for identified data and it is the best system to receive alerts for variant reclassifications and report amendments.
    - Workgroup members seeking GeneInsight assistance can contact Heidi Rehm (hrehm@mgh.harvard.edu).
- Reviews were received back on January 7th, 2019 on the MCS NT244 "*Harmonizing Clinical Sequencing and Interpretation for the eMERGE III Network*" that was submitted in October 2018.
    - Reviewers commented favorably but raised points to address that require clarification or expansion.
    - Reviewers requested more HLA results that were out of the scope of the paper as well as information that will be included in other papers.
- Baylor has generated confirmed reports for three mosaic variants and waiting for Marshfield data. Data is now available on DNANexus data commons.
- The second data freeze is set and the data upload to Aspera is nearly complete. A few issues while defining the freeze set were sex discrepancy issues. Samples with errors will be removed prior to the final submission.
- There are 12,911 DCR reports (466 positive). Some efforts have been attributed to the difficulty in diagnosing PHS failures.

- The next ClinVar submission is in progress and will include the latest pathogenic/likely pathogenic variants. In addition, the CSGs have prioritized VUS where at least one ClinVar assertion says pathogenic or likely pathogenic (currently 456 variants in this set).
- The goal is to have every report available in FHIR format by May 2020.
- Working on creating variant harmonization tools by displaying variants where one site has path/lpath and transferring updated VIP file.
- Reanalysis activities in progress include obtaining additional phenotype data for VUS/LP borderline variants, new guidelines for MYH7, updating BRCA1 classification using new functional assay from KPW/UW, and single exon CNVs.

Science presentation: Chatbots as an innovative tool in the delivery of scalable genomic counseling | Tara Schmidlen *(Geisinger)*

- ClearGenetics combines the knowledge of experts in the field of genetics with artificial intelligence to create health-driven technology to provide modern end-to-end clinical care.
- Geisinger staff members work with Clear Genetics staff members to collaboratively develop chatbots for use with patients enrolled in MyCode Community Health Initiative.
- Chatbots are a technology based simulated conversation tool used in scaling communications; Chatbots can answer simple questions to Increase and maintain consumer engagement.
- Chatbots are easy to use, as they are deployed by a link and there is no app needed. They can be used on phones, tablets, and desktop PCs.
- Currently, chatbots are personalized to patients, and will eventually be used to have patient schedule appointments, schedule test kits.
- Geisinger's MyCode study is using chatbots.
- The first chatbot introduced as the Consent Chatbot, as the patient progresses through the chat, they can either get the 'deep dive' or highlights.
- Smart FAQ is available for patients' to ask questions. Chatbot uses NLP to generate a response, if unable the question is routed to Geisinger's team. The Chatbot gets smarter with the more responses used.
- The Family Sharing tool allows patients to share their information with family members easily and preview what they want to share with their family members prior to sharing it.
- Patients can choose whether they want to share on Facebook messenger vs. other methods.
- Cascade Chatbot to describe relative's results and likelihood of having the gene. The chatbot will help facilitate genetic counseling if family member would like to move forward.
- To gauge patient interest and improve chatbots prior to further testing/deployment, there were 3 In-Person focus groups for Consent Charbot and 3 In-Person focus groups for follow up.
- Focus group demographics of those already consented in the MyCode study: 70% women, 30% employed, and 94% caucasian.
- Focus group showed that participants found chatbots were more informative than in-person MyCode recruitment experience.
- Majority of participants would prefer to talk with relatives before sending Chatbot, but appreciate Chatbot for discussing difficult information.
- Iterations made to the Chatbot include the addition of a summary section, gratitude expression, information about blood draws, emphasis of HIPAA security at the beginning of Chatbot, and logo of HIPAA and Geisinger on the landing page.

- The team plans to implement EPIC EHR integration, track cascade uptake, track patient follow up and outcomes, interview patients and relatives that use Chatbots, along with further iteration and development of other chatbots.
- Patients began consenting for "electronic communications" in August 2018. 153 of 203 eligible patients were asked about electronic communication. 57.6% of eligible patients consented.
- The first chatbot was deployed in September 2018.
- 84 patients have been sent the "one month" chatbot; 35 have completed.
- 86 patients have been sent the "Family Sharing Tool"; 17/86 have shared and 22/25 invited family members have completed.
- Group Discussion
    - If a participant declines, the record is kept up to the point of dismissal, this is for tracking purposes and is compliant with the Geisinger IRB.
    - There are two identifiers required to enter into the Chatbot to assist with privacy related issues.
    - ClearGenetics has access to the choices, but not any of the patient identifiers or the identified information.
    - The Chatbot is quite scripted for the consent, and even the smart FAQ responses are scripted. Most of the time the smart FAQ answer is correct, there is a confidence interval associated with each NLP related question.
    - Real time physician interaction is not available yet, an email is kicked out if the answer cannot be generated.

eMERGE DAY 2: FRIDAY

Science presentation: Videos for scaling genetic education to large populations | Maureen Smith *(Northwestern)* & Julia Wynn *(Columbia)*

- Julia Wynn presented experience of pre-test genomic video education for Columbia eMERGE participants and lessons learned.
    - Columbia's IMAGene (Individualized Medicine through Application of Genomics) project targeted recruitment of individuals who self identify as Jewish or Latino by EHR, flyers, and community events.
    - Participants choose how to complete consent, surveys, and receive genetic education (paper, online web-portal, or mixed).
    - The web-portal was developed by a team of bioinformaticians, genetic counselors, physicians, translators, and community members over a period of 12 months at a cost of approximately $100,000.
        - Participants were able to self-navigate to consent form, study survey, educational materials, schedule their blood draw, and contact the IMAGene team.
        - Education materials included videos, PDFs of the videos, definitions of genetic terms, and the conditions included on the genomic screen.
    - There were 7 videos each ranging from 1.5 minutes to 3.5 minutes. The videos were in basic powerpoint animation format for most messages, and more complex animations for more complex concepts.

    - Out of the 471 Jewish or Latino people recruited, 81 did not have web-portal access and 390 had web-portal access.
        - 4/81 people without web-portal access completed consent and survey online.
        - 200/390 participants with web-portal access completed consent and survey online;
            - 303/390 participants had logged activity on self-guided learning via web-portal.
            - 291/303 participants completed the web survey and 136/291 accessed the videos.
    - 90% of participants strongly or moderately agreed that the videos helped them to understand how genetic variant affect your risk for a genetic condition.
    - Fewer participants endorsed that the videos helped them to make a decision about the results they wanted, the potential impacts on their family, and corrected misunderstandings.
    - Majority of participants elected access to the web portal but not all of these individuals completed the study through the portal.
    - Website access were more frequently used by men, participants under 45 years of age, non-Latino, participants with private insurance, education over a high school diploma, English speaking, born in the US, and those who did better on a baseline genetic knowledge Q.
- Maureen Smith presented experience of post-test negative results video education for Northwestern eMERGE participants and lessons learned.
    - Northwestern's general process of returning negative results include participants receiving a letter indicating their genetic test results are negative. The letter briefly explains result, genetic concepts defined, and a website is provided with further explanations about genes tested and study information.
    - Preliminary qualitative study on returning negative results was conducted by a genetic counseling graduate student to investigate patients' understanding and perceived utility of receiving negative genetic test results in a generally healthy population.
    - 17/52 participants received negative eMERGE test results and had consented to a follow-up interview. Participant demographics were 7 white, 5 African-American, and 5 Hispanic; all participants spoke english and had some college education.
    - Findings showed that participants were unfamiliar with genetic concepts, were unable to identify environment and behavioral disease-causing factors, and expressed a general lack of understanding of the meaning of a negative result.
        - All expressed some level of understanding about residual risk.
        - Several expressed confusion and/or disappointment that they had received no explanation for their current disease.
        - Many participants expected the genetic test report to be similar to other laboratory results and were not sure what genes were included in the analysis.
    - Video One: A Review of Genetic Testing (3:32 minutes long) provides a description of the study; addresses questions about genetic concepts including genes, variants, how variants related to health conditions; describes generally what was tested for in the study; explains genetic heterogeneity; and details how to contact the Northwestern team.
    - Video Two: Negative Test Results and Your Health (4:09 minutes long) explains what a negative test result means, details what genes were part of the test, defines residual risk, provides reinterpretation-changes in current knowledge; helps participant understand results if participant or family member has a condition; and guides how to share results with physician and family.

    - Northwestern is currently evaluating the effectiveness of the videos on comprehensive and perceived value of a negative result, with plans of conducting pre- and post- surveys in the future on videos.
- Group Discussion
    - Is there a standard method of measuring participant's understanding of negative result? NU plans to use surveys conducted at Mt sinai to measure baseline genetic knowledge.
    - NU places an emphasis on participants' benefit from the research.
    - 95% participants at Mayo had no disappointment in a similar study conducted formerly.
    - There could be differences of expectation versus understanding.
    - NU prefers to steer away from language that is negative, as they have noticed when people are told anything negative (ex. Negative result), they perceive it as bad news, and do not engage.

Outcomes Panel | Josh Peterson (VUMC/CC), Christopher Lee (Mayo), Margaret Harr (CHOP), & Cindy Prows (CCHMC)
A summary can be found on the website here.

- Josh Peterson presented the objectives of the panel as well as outcomes forms background, progress, and future efforts.
    - The objectives of the panel were to conceptualize outcome assessment, update on progress and goals, and present on Mayo and CCHMC experience conducting outcomes assessment.
    - eMERGE is a large cohort study which returns results to participants and conducts longitudinal follow-up.
        - eMERGE has completed recruitment and is the first group to collect outcomes from return of ACMG59 variants.
        - The heterogeneity in the eMERGE Network has a huge impact on collecting outcome, and there have been efforts for standardization in sequencing and reporting which will benefit outcomes
    - The three approach to outcomes assessment are return of results, application to clinical practice, and longitudinal follow up.
    - Outcomes forms are completed 6-months post-ROR and all sites should be finished by August 2019. As of December 2018, 17% of forms have been completed. Using the CSGs' reporting totals, there will be 1026 variant positive patients to collect outcomes on.
    - The Outcomes workgroup plans to develop timelines for 6-month and 12-month outcomes assessment and
    - Lessons Learned:
        - The Outcomes workgroup achieved broad coverage of applicable phenotypes with some sacrifice in the depth of outcome assessment on the phenotype.
        - The context is important to understand the changes in health services delivered, however context is difficult to uniformly capture.
        - These lessons learned provide an opportunity to inform national and international outcomes assessment efforts.
- Christopher Lee presented an analysis of 59 participants in the Mayo Clinic RAVE (Return of Actionable Variants Empirical) Study.
    - The RAVE study included 2538 patients: 57% identified as female and the average age was 63 years.

    - The objectives of the study were to ascertain clinical outcomes consequent to ROR at 6-months post-ROR and compare abstraction using eMERGE outcomes forms to “in-house” manual review.
        - For in-house review, each Mayo case was reviewed in detail by an MD and data was collected into an abstraction form.
    - There were 59 patients with 6 month follow up: 51 had their result disclosed by a genetic counselor in person and 8 had their result disclosed by a genetic counselor over telephone.
    - Variant categories for the patients included 26 with cardiovascular variant, 26 with cancer variant, and seven with other variant.
    - For cardiovascular variant finding, LDLR was the most common and for cancer variant finding, BRCA2 was the most common.
    - Out of 18 FH participants: 10 accepted referrals and seven were reviewed in a FH clinic. 3/10 participants have appointments pending.
    - Three FH participants were evaluated further with CT coronary calcium and one FH participant underwent exercise stress testing.
    - The 6-month FH Outcomes from the RAVE study include medication change (four participants), lifestyle intervention (two participants), and no change (one participant).
    - Outcomes for 59 participants at 6-month post-ROR include 31 referred to specialist, six with medication or therapy started or altered, two underwent prophylactic surgery, and 19 had surveillance initiated.
    - The eMERGE outcomes forms were released after the first iteration, can be highly granular, require data shifting, have minor bugs, and sites’ cannot make changes to fit local EHR system. Mayo’s “in house” abstraction was developed after multiple iterations with lower granularity and less emphasis on dates.
    - Challenges to collecting outcomes across eMERGE sites include the difficulty of harmonizing outcomes forms across multiple sites and balancing what data is necessary.
    - The advantages of the eMERGE outcomes forms is the comprehensive data capture.
- Margaret Harr and Cindy Prows presented the considerations and preliminary outcomes from returning genomic results in pediatrics.
    - CCHMC has two cohorts each with different pediatric ages and result choices.
    - Parents of children in the Prospective cohort, aged 13-17, made independent choices followed by dyad joint decisions and consented to ROR upon enrollment in biobank.
        - Parents had the choice to receive consensus list and additional, preventability, treatability, adult onset, and carrier status.
    - In the Retrospective biobank cohort, parents of children aged 1-17 consented to be re-contacted for ROR study and to re-link data and provided choices.
        - Parents were given consensus list minus genes for adult onset disorders and risk for diseases that may occur during childhood.
    - Parents choices included: results that are immediately actionable, results related to diseases that could be medically actionable, and results related to pharmacogenetics (drug processing).
        - The choices were given in a yes/no format.
        - Parents were informed that medically actionable adult onset disorders only returned to participants who consent as adults.
        - Choices were given at the time of enrollment and again immediately prior to ROR.

    - All of CCHMC's adolescent/parent dyads cohort (163 participants) were given choices and made choices: 66% of adolescents chose to receive all results and 76% of parents chose to receive all results.
    - 46% of CCHMC's biobank (91 participants) were given choices and 45% made choices: 18/19 parents chose to receive all results.
    - All of CHOP's prospective cohort were given choices (350 participants) were given choices and 100% of participants made choice to receive all results.
    - Challenges to ROR and Outcomes encountered with pediatric cohorts include pediatric participants "aging out" causing the need to re-consent as adults (33 at CHOP and 59 at CCHMC), patient-centered versus family-centered return of results process, difficulty in recontacting, and biobank issue of participant forgetting or unaware of initial study enrollment.
    - The pediatric cohorts will provide data for outcomes assessment and penetrance analysis as well as offer potential longitudinal studies in eMERGE IV.
        - 12 CHOP participants have a FH disease risk result and 5/5 of participants with previous screening had an abnormal result.
        - 11 CHOP participants have an Arrhythmia disease risk result and 1/10 participant with previous screening had an abnormal result.
- Group Discussion:
    - Genetic counselor in cardiovascular genetic has joined the team and the group plans to continue working with her on outcomes forms.
    - A CCHMC cardiologist is interested in the ROR process and outcomes study as most parents are opposed to starting children on statins. This could be connected to the familial cascade testing being done at Geisinger.
    - Participants over 18 can be re-consented to receive adult onset conditions.
        - CCHMC received de-identified results from Baylor and have flagged on re-consented list to appear when the time comes. The site plans to develop a plan of action for these cases.
        - Boston's IRB for BabySeq requires parents are given the option to return adult onset disease results, while VUMC.
        - VUMC and Geisinger's IRB does not require the return of adult onset disease risk results to pediatric patients.
    - Results related to the indication were given the choice to return to the patient, and most of the time the condition was already known.

ROR Workgroup Key Elements Report Out | Ingrid Holm *(BCH)* & Iftikhar Kullo *(Mayo)*

- Milestone #1: Determine the impact of ROR on patients' outcomes, immediately 6 months and/or 12 months after ROR
    - The group plans to accomplish this milestone by completing the analysis on participant surveys, HCP surveys, a new project on negative results.
- Milestone #2: Explore challenges identifying at-risk family members and informing them of risk and collect responses of the family members
    - Several sites have in-depth efforts on this topic include exploring HIPAA issue of returning to deceased patients at VUMC, CCHMC, and KPW/UW and the MeTree supplement.
- Current workgroup projects include participant surveys, healthcare provider surveys, IRB perspectives, Familial Implications of ROR, Optimizing sIRB Review for Genomic Research, and MeTree.

- MeTree is a one-year administrative supplement at Vanderbilt's eMERGE site (VGER) from June 2018 to May 2019.
    - The purpose of MeTree is to use implementation science methods to streamline integration of Family Health History (FHH) collection tool into the EHRs of diverse institutions.
    - The overall goal is to promote the adoption of FHH in genomic medicine by evaluating the facilitators and barriers of integrating MeTree FHH collection tool by leveraging the strengths of the eMERGE Network and Duke's MeTree FHH collection tool.
    - Currently there are three eMERGE sites working with Duke University: VUMC, Geisinger, Northwestern.
- Some challenges in the ROR process are refusals from participants to receive results, inability to contact participants, participants becoming incarcerated or adopted, and the need to re-consent at 18 years old.
- Workgroup publications in process:
    - NT277, *Operationalizing participant choices about genomic results: Beyond all or none ACMG recommended genes* is a joint project with the EHRI workgroup.
    - NT273, *Returning genomic results to eMERGE participants: The who, what, where, and how of disclosure* led by Kathy Leppig and Georgia Wiesner.
    - NT300, *Understanding the return of results process: Content review of patient summary letters* led by John Lynch and Janet Williams.
- Workgroup is considering potential projects on lessons learned and discrepancies in interpretation of sequencing results.

Phenotyping Workgroup Key Elements Report Out | George Hripcsak *(Columbia)* & Wei-Qi Wei *(VUMC)*

- Phenotyping Workgroup has a recent publication titled "*Learning from Longitudinal Data in Electronic Health Record and Genetic Data to Improve Cardiovascular Event Prediction*" (Zhao, 2018).
- For phase III of eMERGE, there are 26 phenotypes. Of these, 21 (81%) are in algorithm development, 20 (77%) are implemented, 21 (81%) have been validated, and 16 (62%) have been completed.
- To date, 33 phenotypes have been implemented since the start of eMERGE.
- Atopic Dermatitis is not ready to be run, but is close. Estimated projection is end of February 2019.
- The Common Variable Subgroup goal was to identify a common set of variables that can be collected and used for multiple studies. This subgroup was sunset in December 2018. The next data refresh is currently underway. There are 61 unique laboratory variables collected using the OMOP format, including 45 unique autoimmune disease labs.
- The goal of OMOP is to improve the sharing and consistency of phenotype definitions and data extraction across the Network. The adoption of the OMOP data model is ongoing. All sites have populated data in OMOP. They have tested phenotypes, and have returned a last few surveys. They have run common variables OMOP implementation. This supplemental group was sunset in December 2018.
- The NLP subgroup plans to develop a natural language processing component for a maximum of five high priority phenotypes, including ACO- Asthma/COPD, long QT/Arrhythmias, chronic rhinosinusitis, familial hypercholesterolemia, and systemic lupus erythematosus.
- Next steps are to finalize the order of implementation, finalize what and how to share with secondary sites, and polish evaluation metrics.
- ACTION ITEM: Sites should return the common variables refresh data to the CC by February 1st.
- ACTION ITEM: Sites should return the case/control files on phenotypes to the CC they are leading as they complete the phenotype implementation.

- ACTION ITEM: The Phenotyping Workgroup will submit the OMOP lessons learned paper to the *Journal of Biomedical Informatics* by March 1st.
- ACTION ITEM: The CC will collate names of those interested in participating in the main MCS *Journal of Biomedical Informatics* lessons learned paper so a MCS can be developed.
- ACTION ITEM: The CC will circulate clarification on the pediatric pain algorithm to include both adult/pediatric cohorts by February 15th, 2019.

Genomics Workgroup Key Elements Report Out | Megan Roy-Puckelwartz *(Northwestern)*, Patrick Sleiman *(CHOP)*, & David Crosslin *(KPW/UW)*

- The workgroup will add an additional 5,000 Harvard files to the GWAS imputed set. Currently, the 1800 Whole Genome Sequencing subjects will not be imputed and merged into the larger GWAS set yet, however if the Network is interested in utilizing them, the CC can impute them.
- Two Genomics Workgroup publications have been released recently, including: "*The eMERGE genotype set of 83,717 subjects imputed to ~40 million variants genome wide and association with the herpes zoster medical record phenotype*" (Stanaway et al, 2018), and "*Unfolding of hidden white blood cell count phenotypes for gene discovery using latent class mixed modeling*" (Hall et al, 2018).
- SPHINX is undergoing changes and a newly proposed tool called eMERGENT would replace the current form in the future. Once Genotypic and Phenotypic data is incorporated, it will be user-friendly and hold logs for individuals to download data on the private side. There would be private and public portals that would allow for tracking of utilization, both on the individual and population level analysis. Next focus group for the eMERGENT planning group is February 1st. It is likely that building eMERGENT would require supplemental funding.
- ACTION ITEM: Network members should contact Laura Allison Woods (laura.a.woods@vumc.org) at the CC if interested in joining the eMERGENT planning focus group.
- MCS NT179: The eMERGE-PGx paper with 9,010 participants titled "*Pharmacogenetic variation identified via targeted next-generation sequencing among 9010 eMERGE Network Participants*" has been submitted this week to *Pharmacogenetics & Genomics*.
- The Workgroup is interested in looking at the phase data with reference data with what is in the population.
- There are multiple structural variation (SV) efforts ongoing in eMERGE. CNV calls have been called on the eMERGE I dataset, and the Workgroup is leading efforts to secure intensity data from all sites and sample from eMERGE II and III. The group has already circulated an SOP to secure this data. Four sites have data for the CNV analysis, three sites are working on finding data for the CNV analysis, and two sites do not have the data for the CNV analysis.

General discussion & closing remarks | Rex Chisholm *(SC Chair, Northwestern)*

- Marc Williams proposed current eMERGE lessons learned papers to submit to BMC Medicine.
- Chunhua will be submitting to *Journal of Biomedical Informatics* special issue on Phenotype methodology.
    - ACTION ITEM: Workgroup members with phenotyping methods and results are invited to submit to the *Journal of Biomedical Informatics* special issue by March 1st.

ACTION ITEMS:

Please note some of the Action Items were gathered directly from the breakout sessions

PI/Network/CC:

- Those interested in joining the expert panel on the selection of harmonized measures for CSER2 should RSVP to Frank Angelo (fana@uw.edu).
- Workgroup members seeking GeneInsight assistance can contact Heidi Rehm (hrehm@mgh.harvard.edu).
- Marc Williams (mswilliams1@geisinger.edu) will share the eMERGE health outcomes forms PDFs with the CSER Network.
- NHGRI will provide further guidance on early FFR submission needed prior to the official end of the grant on March 31st, 2020.
- Workgroup members with phenotyping methods and results are invited to submit to the *Journal of Biomedical Informatics* special issue by March 1st.
- Network members should contact Laura Allison Woods (laura.a.woods@vumc.org) at the CC if interested in joining the eMERGENT planning focus group.
- The eMERGE CC will organize joint calls between eMERGE Outcomes Workgroup and CSER's Healthcare Utilization Workgroup (CUHEP) for collaboration efforts.
- The eMERGE CC will share the patient classification publication framework publication titled *"A collaborative translational research framework for evaluating and implementing the appropriate use of human genome sequencing to improve health*" with the CSER Network (Khoury, 2018).

Clinical Annotation

- The eMERGE and CSER Networks will identify genes (ACMG 59, CDC Tier 1, Arrhythmia) shared across eMERGE and CSER to study outcome definitions and interventions.

Phenotyping:

- Sites should return the common variables refresh data to the CC by February 1st.
- Sites should return the case/control files on phenotypes to the CC they are leading as they complete the phenotype implementation.
- The Phenotyping Workgroup will submit the OMOP lessons learned paper to the *Journal of Biomedical Informatics* by March 1st.
- The CC will collate names of those interested in participating in the main MCS *Journal of Biomedical Informatics* lessons learned paper so an MCS can be developed.
- The CC will circulate clarification on the pediatric pain algorithm to include both adult/pediatric cohorts by February 15th, 2019.

Outcomes:

- Sites should submit abstraction guides for the Outcomes form which they lead to the CC by February 15th.
- The CC will reach out to each site in to confirm if any Outcomes data needs to be imported into the CC housed REDCap instance.
- The CC will confirm all sites Outcomes REDCap instances will be switched to "in production" so data entry can commence.