emerge network
ELECTRONIC MEDICAL RECORDS & GENOMICS

Summary of Steering Committee Meeting: Winter 2018

January 25-26th, 2018 in Bethesda, MD

TABLE OF CONTENTS

DAY 1: THURSDAY

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

- Welcomed the group to the 32nd Steering Committee meeting
- eMERGE & Beyond Workshop materials are posted online and can be accessed via https://www.genome.gov/27569445/emerge--beyond-the-future-of-electronic-medical-records-and-genomics/. Potential future directions for eMERGE listed below:
    - Electronic Phenotyping for Genomic Research
    - Evidence Generation for Genomic Medicine
    - EMR Integration of Genomic Results and Automated Decision Support
    - Novel and Disruptive Opportunities in Genomic Medicine
- Two sites have completed sequencing (Harvard and Geisinger) for the eMERGEseq cohort. Columbia has shipped additional samples to be sequencing which will be paid for by other resources.
    - **ACTION ITEM**: Site PIs are asked to complete ROR timelines by February 5th, 2018, prior to the NIH Council Meeting
- To date, only 10 phenotypes have been implemented. With only 16 months left in eIII, and a total goal of 42 phenotyping algorithms completed by October 2018, sites are reminded to focus on increasing the phenotyping process rate.
- **ACTION ITEM**: The SC is asked to suggest topics for panel discussion on 'Best Practices' for future eMERGE III SC meetings.
- Meeting goals:
    - Complete ROR timelines, propose possible topics on best practices panel for future eMERGE meetings
        - Focus on the planned methods for ROR and learn from the sites who will share their initial ROR experiences
        - The group should also think about the science being conducted with the samples
    - Share results of ongoing science projects
    - Report on workgroup activities, results, products, and timelines in the remaining 16 months of eIII
- Site feedback suggests that there is eIII participant withdrawal due to lack of timely ROR. As widespread ROR at all sites is not in progress yet, it may be difficult to develop ROR best practices for both the provider and participant until results are more fully disseminated across the Network.

Network status update, opening remarks | Rex Chisholm *(SC Chair, Northwestern)*

- Suggestions and feedback from both the ESP and eMERGE & Beyond Workshop summary included:
    - Publish lessons learned
    - Increase phenotyping and integrate into CDS, publish on phenotyping challenges
    - Collect metrics on change in variant classification and downstream effects, changes in classification by site
    - Catalog rationale behind differences in ROR goals, procedures, and effects on providers and participants.
    - eMERGE projects are quite divergent, and in the future should may need to follow a single protocol. Need to determine the single 'best' approach.
    - A four year timeline for eIII was aggressive, and NHGRI is considering a Council proposal to extend eIII for an extra year in order to bring a more concept forward for eIV to Council in Winter of 2019.
- Since October 2017, eMERGE III has been registered on ClinicalTrials.gov as an observational study and PGRNseq study has been made available on dbGaP: phs000906.v1.p1. In addition, a data freeze on

eMERGEseq samples was set at 15,754; this will comprise the dbGaP interim dataset submission. The CC is working to QC submitted data files for these participants.
- 10/27 phenotypes deployed across Network. Record Counter has been updated (November 2017) and contains common data variables, these variables can be viewed on the eMERGE website: https://emerge.mc.vanderbilt.edu/phenotyping-projects/

Genomic data update | David Crosslin *(CC/UW)*
- eMERGEseq dataset: Aggregate calling to begin on 15,754/25,000 participants, BAMs have been received from Baylor and are available for download via Aspera.
- eMERGE I-III genotype chip array imputation dataset: 5,000 additional eIDs from Harvard have been added to original 83,717. A Manuscript Concept Sheet (MCS) will be circulated shortly to the authorship group for review concerning the GWAS set.
    - Ian presented the overall principle component analysis of eMERGE I-III genotype chip array imputation dataset
    - As a quality control measure an analysis of the newly imputed GWAS showed a new herpes zoster association identified through PheWAS codes compared to original analyses of the prior GWAS set (NT238)
        - Chromosome 3 transcription binding site; implicated with Kaposi's sarcoma-associated herpesvirus (KSHV),HIV, Epstein-Barr similar association.
        - May conclude this is associated with a general loci for a viral response and regulatory interaction
- SPHINX
    - Searching by variant, chromosome number and Asian ancestry added.
    - Annotation of prior clinical associated from NHGRI GWAS catalog.
    - SPHINX was designed specifically to showcase the PGRNseq panel, and only pointed to the 82 genes on that panel. The CC is enhancing its functionality to search and display the full genome, allowing for the incorporation of the eMERGEseq panel and other data in the future

Clinical sequencing centers update | Hana Zouk *(Partners/Broad)*; Eric Venner *(BCM/HGSC)*
- *Partners/Broad*
    - 10,265 samples of out 10,500 goal are sequenced
    - Completed 3,215 cases: 290 positive reports and 988 negative (UW)
    - 5.71% positive non-indication based returnable results
    - Only returning CNVs with three or greated of exons reported
- Baylor
    - 10,180 samples sequenced
    - 3.2% for non indication based returnable results, the majority falling within cancer
    - Report customizations for sites
    - Comparison of variant curation from the first 40 sequencing batches of uncategorized vs present variants in the at start of eMERGE.
    - Continuing to push into DCR, currently contains 1997 negative, 111 positive
    - New features for eDAP variant curation and analysis include access to the VIP database, report explorer, and report validation.
    - Access is available on an Android app and hopefully rolled out to other systems shortly

Science presentation: Estimating the efficacy of pharmacogenomics over a lifetime | Scott Hebbring *(Marshfield)*
- Gathered genetic information for 1700 deceased individuals and birth to death medical data.
- Examined impact of testing for pharmacogenomics on adverse drug events over a lifetime

    - If you have the variant you have an increased association of having an adverse drug event, significant for CYP2D6 (0.36 risk)
    - Broke down impact of PGx testing on age, specifically 40% would be prevented if test by age 50
- Study limits: sample population was primarily white from a single medical institution, medication data from EHR "messy" and extracting adverse drug events from EHR can be difficult, with few measures of clinical impact of alternative treatments.
- Conclusions: 80% of population carry PGx variants that may be influenced by two or three separate drugs over a lifetime. PGx testing by age 50 had the greatest benefit to participants, however may still be worth testing after age 50. 17% of ADE may be prevented by PGx testing by age 50.
- Discussion:
    - To investigate other age event associations over 300 participants would be needed to achieve appropriate power.
    - A deeper dive by conducting a cost analysis related to treatment adjustment in response to PGx testing would be a good next step, but Marshfield may not have the expertise to leverage. 300 participants at $250/participant.
    - Next steps include conducting a prospective study on this topic
        - Possibility for NIH funding for this, Marshfield should consider investigator initiated studies
        - Adverse events for patients over 70 years old center around drug-drug interactions, so a study of this type may be very impactful
    - Question regarding if there is predictive power, if individual has the variant and drug but had not had an adverse event yet
    - Marshfield leveraged CPIC guidelines to identify which single variant describes the particular star (*) allele.

Panel: Variability among sites in the process of returning CLIA validated results to participants & providers | Christie Hoell *(Northwestern)*; John Connolly *(CHOP)*

- Goal: To understand ROR processes at two sites, and have other sites provide comparative insights on the ROR processes similarities and differences. This is an open discussion to establish the best practices for ROR to participants from receipt of findings.
- Northwestern reviewed their ROR process flow and diagram. Negative results are mailed to participants, positive results undergo the following steps
    - Genetic counselor assigned and complete a chart review to determine if variants are known prior to eMERGE participation. If not, family history is analyzed. Case conference focused on clinical care recommendations and family implications..
    - Feature of Northwestern's ROR process: Censoring part of the report that a patient choose to not view and does not want access to (patient choice related to content viewed/returned to them).
        - Questions include: How is this tracked? How many censored sections are available? Which sections are typically censored? What are the metrics on this information?
    - Two groups in the study: one group received all results and the other receives choices on type result: treatment, condition type (known or uncertainty). The group tracks the choices and Ancillary Genomics System (AGS) automatically knows what parts of report to remove
    - Northwestern uses a PDF file which is difficult for providers to search & utilize.
    - Time invested.to build out the elements to streamline reports is estimated at 50% of a developer for the course of the year, in addition to Luke Rasmussen's (project and content lead) time. The hope is they can utilize this method for the return of other results in the future.
- CHOP presented an overview of their ROR process diagram, focusing on each main component.

    - Selected 3,000 of the n=80,000 CHOP patients for eMERGEseq. Piped into REDCap due to customizable functionality and generally designed for this field of research. Leverages centralized database and survey functionality.
    - Samples were both prospective and retrospective, similar to NU. CHOP returns results through a genetic counselor, however PGx return does not require in person return.
    - Automated the extraction of positive and negative reports (flagged) and then the genetic counseling team re-reviews the results and examines the medical record to confirm if the results have previously been received by the participant. Result is disclosed to the patient via a genetic counselor and then uploaded into the EHR via a PDF. PGx results are returned by a letter and uploaded to the EHR.
- Timing of result return to participants and entry into EHR
    - NU: Controls when the results are pushed into the medical record.
    - CHOP: Does not store records in EHR until results have been disclosed to participant. CHOP cannot post results into the EHR until the participant has been counseled, as per their IRB, genetic results are not available in the patient portal.
    - Geisinger: Results are imported to the patient portal quickly. They built in a 5 day delay to allow participants to be notified.
    - Vanderbilt: Plans to have the Primary Care Provider (PCP) return result to patient. Group meets weekly to decide next steps for results received. A letter is sent to patient with result and advice to communicate with PCP and uploaded to the Patient Portal. PCP is responsible for patient's health recommendations and next steps. System is set up to reflect real-world interactions, as the lack of genetic counselors would bottleneck a return on a large scale. The level of urgency of the result determines if the notification of result is given via phone or letter.
    - Harvard: Results start as research results and then transfer into clinical care to reflect likely system going forward for all research results return. They were able to begin returning earlier as they were working off the GWAS chip and ACMG59 on the research data before receiving back the 2500 results from the eMERGEseq dataset. Use a committee which includes geneticist and genetic counselors, the results are validated in a CLIA lab. Genetic counselor contacts participants that they have a result and meets with the participant.
    - KPW: Result is not released into the EHR until it has been returned to the participant.
    - CCHMC: Developed an app that checks GeneInsight every 24 hours and transfers the secure data. App flags positive results that are not selected by the participant and reviewed by a team. The file is a modified text format which allows for customization of what the adolescent diad (13-18 and parent) has selected on what they want to receive. Negative results are being either returned by MyChart or phone, positive results are returned by phone by a genetic counselor.
    - Debate surrounding the burden on the PCP, is it more efficient than using a genetic counselor or does it add undue stress. PCPs feel they should manage the results as they do with other aspects of care, however sometimes feel unprepared due to the complexity of the results. eMERGEseq will allow us to test different workflow streams across the sites and provide some feedback on how each of the various approaches is received.
    - Need to establish ROR workflows for updating variant interpretations. Need to develop processes for managing clinical annotation updates and ROR after eIII is completed.
    - There has been a push to allow patients to access all their information in their medical record, within 30 days regardless of whether a clinician has reviewed results. Thus, there's some urgency to disclose a result to a patient before they access it via their EMR. Providing the result initially rather than only informing that a result is available gives less anxiety to patients, and they arrive better informed with questions to ask the providers.

- At Harvard, experiences by one of their genetic counselors has indicated every patient wants something a little different returned. Some patients are considered "the worriers"? Those who will have an adverse reaction when results are returned. Tailoring the ROR process to address these complexities may be worthwhile.
- Need to determine the most ethically appropriate manner of returning eMERGEseq results to participants
- Best practices: Results should be returned quickly to participants, though it should not be posted in the patient portal until participant has been notified.
- Best practices: Assign participant to 'tier' status, depending on urgency of results.
    - Determine type of patient: wants or needs genetic counseling to return result, wants PCP to return result, etc
    - Determine type of result: urgent return, not urgent return, etc
- **ACTION ITEM:** Sites are reminded to send tools they have developed in response to eMERGEseq to CC to make widely available to the Network on the eMERGE website.

Network discussion: RE-AIM framework: Utilization at the site and network level | Alanna Rahm *(Geisinger)*

- Implementation science focuses on translating evidence into practice. Goal is to utilize the RE-AIM framework and apply implementation science to the eMERGE ROR process.
- RE-AIM Framework: Originally developed to encourage consistent reporting of results. Aimed to improve adoption and sustainability of effective, generalizable, evidence-based practices
    - Reach – to the target population
    - Effectiveness - Efficacy
    - Adoption – by target settings/staff
    - Implementation – consistency, costs
    - Maintenance – of program and effects over time
- Not all domains (for RE-AIM or other frameworks) need to be used for every site/study.
  A Network-wide framework guides collection and reporting of outcomes, development of new research questions, and fits with bigger picture of genomic medicine and how we are providing evidence for integration into care. Site-by-site differences become a strength in guiding the overall evidence-base and real-world utilization of genomic medicine.
- Different ROR processes data already collected can be analyzed to show elements necessary for ROR, adaptable processes, and context specific The framework allows for translation of research into practice more efficiently and effectively. By utilizing tools from implementation science, translation can happen yet be adaptable to new evidence and other changes.
- The group discussed if there are there other implementation models/frameworks to cross-compare RE-AIM with in order to make sure the group will select the best fit for eMERGE. Other frameworks exist with varying complexity. Most of the other models incorporate core foundational elements of RE-AIM. RE-AIM was chosen as it works with Geisinger's interest in developing a sustainable model as well as within the context of progress at their site.

Invited speaker: Multimodal automated phenotyping with application to PheWAS | Tianxi Cai *(Harvard)*

- This particular approach is a bottom-up approach to assist retrieval of phenotype information from a large number of phenotypes quickly.
- PheNorm
    - Normalization - removal of interference of healthcare utilization can improve predictive performance of ICD codes
    - Gather information from supporting evidence to predict the disease

- Presented an AUC (area under the ROC curve) comparison to prove the improvement in prediction performance.
- Limitations of PheNorm: does not provide predicted probability of disease, and does not provide a threshold value for classifying subjects as cases
- MAP (multi-modal automated phenotyping): is a refinement of PheNorm. Aimed to fit a sequence of mixture models to predicted probabilities for all patients and estimates of disease prevalence from each fitting. Synthesizes information via model averaging and classifies as a case if predicted probabilities exceed a given threshold.
- Discussed the risk, if any, of overfitting between all of the methods presented
    - Unsupervised method did not have labels, therefore no risk of overfitting.
    - Supervised method does have overfitting but conducts cross-validation to minimize risk
- Compared phenome study to Josh Denny's 2013 study to establish proof of concept to find the corresponding PheCODE and NLP term, which validated their approach. The group replicated Dr. Denny's results. (Denny et al. Nat Biotechnol; 2013. PMID: 24270849)
- Potential to consider top down and bottom up phenotyping methods, can use unsupervised methods for high throughput phenotyping.
- Dr. Cai clarified family history is not included in the NLP approach.

Science presentation: EHR-based computable phenotype for CKD diagnosis and staging with direct applications to kidenty precision medicine | Ning Sunny Shang & Krzysztof Kiryluk *(Columbia)*

- Motivation for focusing on phenotyping Chronic Kidney Disease (CKD) includes: prevalence, high morbidity, high cost, heterogeneous etiology, correlation to fhx/race/ethnicity, and that early CKD is frequently underdiagnosed.
- The Kidney Disease improving global outcomes (KDIGO) staging risk grid was leveraged by Columbia. The grid provided flexibility in selection of cases for genetic, observational, and interventional studies. Anticipated challenges included defining chronicity and steady state for valid GFR estimation, and harmonizing different methods of proteinuria quantification.
    - G-stage classifier: simple rule-based algorithm for serum Cr-based eGFR determination and staging
    - A-stage classifier: more challenge to defining. At least five different methods for quantification of proteinuria, including A24, UACR, P24, UPCR, UA protein dipstick test. Necessitates a machine learning approach (Training on N=68,617 patients across three health care systems).
        - Used urine protein/creatinine ratio (UPCR) based data as an A-stage classifier, "Columbia UPCR-based Ordinal Classifier"
- Higher heritabilities for the moderate and advanced CKD, the group is unsure of reasons behind this result. A similar pattern was seen with quantitative EGFR. Mild cases may be transient or environmental, leading to lower heritability. Age, sex, & zip code are adjusted for to help control environmental factors.
- Future research includes examining the cause and effect of comorbidities by life events and genetic correlation, further genetic studies using eMERGEseq and GWAS, and integrating a clinical decision support system.
- Targeting glomerular therapies are difficult because ICD-9 coding ignores glomerular disease subtypes, ICD-10 is better, however Natural Language Processing (NLP) is needed to define the disease itself. Group suggested examining controlled vocabularies or problem lists.

Science presentation: Family network approach to assess the trickle-down effect of genetic testing | Nora Henrikson *(KPW/UW)*

- KPW/UW conducted semi-structured phone interviews to identify at-risk relatives.
- Participants were asked if family members (blood relatives) would be open to receiving genetic information for three different scenarios:

    - From a genetic counselor as opposed to the family member.
    - From their PCP as provided by the genetic counselor.
    - A note placed in their EMR.
- Two variants were selected from the ACMG list Marfan Syndrome (FBN1) and Malignant hyperthermia (RYR1).
- The genetic results were simulated and not actually returned to family members.
- The group concluded that there was broad support for Kaiser involvement in risk notification and some conditional on assurance of relatives' permission.
- Group suggested creating a control group scenario (i.e. non-actionable finding vs actionable finding) in future work.

Phenotyping workgroup & OMOP supplement progress report | George Hripcsak *(Columbia)*

- Progress to date: Eight completed algorithms and 14 in development.
- Site staffing needs for phenotyping have evolved. eIII undertook more complex phenotypes and scientific questions than previous eMERGE phases. Would like to assess how broadly the phenotypes are implemented, how often, and science produced by these phenotypes across the research community.
- Autoimmunity sites are working together to determine a set of common variables to capture that may be incorporated into the Common Variables list in the future.
- Choice of common variables are based on overlapping scientific needs of studies. General variables, like top classes of medications will not be incorporated until there is a scientific need.
- OMOP conversion is on track. All sites should meet milestone #1's March 31st deadline.
    - Columbia has shared CKD OMOP phenotype definition with KPW/UW in order to begin studying the process of deploying OMOP and outcomes of this experiment.
    - Group has a concept sheet in development for the OMOP group's effort
        - Marc Williams suggested to explore tools or frameworks for assessing how to select phenotypes, instead of operating in a more ad-hoc manner.
        - George Hripcsak cautioned consideration of what experts have to say about areas of uncertainty.
    - Complexity of phenotype implementation is used to prioritize phenotypes
    - Complexity of development and validation is used to estimate time to complete. Complex phenotypes are difficult for even clinicians to determine disease from EHR.
        - Adjusted Clinical Group: will not pursue that pilot in the near future as the work exceeds the benefit for eMERGE.
- Future efforts
    - Publish lessons learned and continue phenotype algorithm development
- Discussion concerned how common variables are selected
    - When the scientific need arises, Network can bring these variables for considerations at the common variables subgroup meetings.
- **ACTION ITEM:** The Phenotyping group will recommend autoimmune variables to the common variable list.

Closing remarks | Rex Chisholm *(Northwestern)*

- NHGRI is considering the possibility of a one-year extension for eIII. The purpose of this would be to allow the Network to have an additional 12-months to help fully develop a concept to bring forth the NHGRI's Council Meeting in January-February 2019.
- In order for an RFA to be released for eIV it would be taken to council in January of 2019, NHGRI would present the best practices, etc, from the Network by December of 2018. In order to collect six months of outcomes, the Network would need to return results by May of 2018. In the October 2018 meeting, the

Network will discuss and refine the 'best practices' found by returning sequencing data to participants. A report will be created detailing these practices by the end of November.
- If Council is open to the 12-month extension in February 2018, NHGRI would request supplemental budgets/application and then take that to council in May of 2018.
- **ACTION ITEM:** Sites should plan to return the majority of Phase I eMERGEseq results to participants by May of 2018.

DAY 2: FRIDAY

Network discussion: Data harmonization and common measures across networks | Rex Chisholm *(Northwestern)*

- NIH Genomic Medicine Working Group (GMWG) provides guidance to NHGRI concerning their genomics medicine programs. GMWG is currently discerning best practices for common measures and outcomes to facilitate harmonization of data through application of common measures across networks. Previously, the group had circulated a spreadsheet with instructions for site PIs to complete and return to their Network's NHGRI Program Director in order to assess potential common measures. This worksheet no longer needs to be completed by the PIs.
- Marc Williams (Geisinger), a member of both CSER and eMERGE, presented on CSER's responses to the spreadsheet and potential overlaps with eMERGE.
- CSER's prioritized outcomes include clinical utility (top-ranked), followed by family utility (cascade testing), psychosocial (patient) utility and provider utility domains. CSER and eMERGE overlap in several areas related to the above outcome measures, largely facilitated through the eMERGE ROR/ELSI (Participant and HCP subgroups), Outcomes, and Familial Implications of ROR groups.
- Suggested moving from process outcomes which are easy to capture to health outcomes, which are more difficult but impactful. Local difference impacting implementation need to be considered as well.
- Group suggested limiting to focusing on clinical utility (yellow boxes), if other outcomes already exist (light blue) then can complete. Progress is need by the October 2018 SC meeting to facilitate best practices discussion.
- Mayo and Northwestern are interviewing participants with negative results which could fit in the psychosocial utility domain of the CSER prioritization.
- Healthcare provider (HCP) survey incorporates many of these outcomes, including return of negative results to participants.
- Group needs to be mindful of how they define a process outcome.
- Implementation of genomics in clinical care is equivalent to the return of result process.
- **ACTION ITEM**: ROR/ELSI workgroup will create a subgroup to apply RE/AIM framework to ROR process to identify common methods, patient and provider survey, familial implications of RO.
- **ACTION ITEM**: Outcomes workgroup will use Outcomes Protocol Database to collect specific outcomes for all disorders, assess ease and accuracy of outcome collection, and work with CSER to determine if behavior/outcomes is related to return of result.

Science presentation: Probing the virtual proteome to identify novel disease biomarkers | Jonathan Mosley *(VUMC)*

- Dr. Mosley presented research in leveraging the utility of biomarkers, measurable indicator of a biological state, to deliver and improve personalized healthcare.
- The researched used eMERGE EHR dataset to identify cases and controls.
- The original research model used genetic proxies of the protein levels to predict health outcomes, as opposed to standard epidemiological studies that measured protein level and associate with health outcomes, however this did not work well.

    - Strong associations, for example with missense variants, or common variation in general, can skew the associations, hence controlling for this negates the predictive model.
- Utilized SOMAscan Platform technology to determine if the proteins were potential biomarkers.
- Computed a predicted level of a protein for each sample and compared to the clinical diagnoses within the eMERGE group; essentially running a PheWAS on the proteins.
    - PheWAS can illuminate the biology of the protein but may lead to the 'right' answer for the 'wrong' reasons.
- Searching for virtual biomarkers to identify candidate associations worked, but needs additional validation to confirm association with the disease and not driven by alternative signals.
- There is potential for eMERGE to collaborate with this research and the validation of the findings.
    - Gail Jarvik (KPW/UW) was interested in further discussion.

Clinical annotation workgroup progress report | Gail Jarvik *(KPW/UW)* & Heidi Rehm *(Partners/Broad)*

- eMERGEseq sequencing results to date presented for all sites individually, broken down in bar-graphical form by indication, consensus and site-specific.
- Reclassification from VUS (variant of unknown significance) to likely pathogenic presented..
    - Partners/Broad reclassification of KPW/UW's variant to likely pathogenic was decided after following up with three labs in ClinVar that had provided the VUS finding and discovering additional Colorectal Cancer evidence.
    - Mayo is currently working with Baylor to reassess VUSs in lipid genes. Baylor has reclassified two VUSs to Likely Pathogenic (LP); four others have been reassessed and remained VUS, however the center is still waiting for Mayo's additional data.
    - KPW/UW could provide a control group for Mayo's lipid gene interpretations.
- Heidi presented ways eMERGE could become a resource to large consortiums and contribute to enhanced understanding of variant pathogenicity.
    - Partners is developing a system of classifying variants into three categories based on probability of shifting from VUS to LP to Pathogenic (P).
        - The system will determine how to select the highest yield and focus on the ones that are most likely to change and the proportion of those that make up the total VUS set.
        - Defining phenotypes associated with eMERGEseq panel genes would alleviate the need to classify probability.
    - Discussed the possibility of making the eMERGE set available for other labs examining the same gene.
    - Suggestion was given to modify SPHINX to incorporate the case and control status.
    - Future possibility of submitting case level data to ClinVar if eMERGE consortium submitted the records but as the CSG are currently submitting to ClinVar it is in an aggregate data form.
        - This could also possibly be completed in gnomAD.
    - It may be beneficial to complete the classification process in reverse and gain evidence for likely pathogenic and pathogenic classifications.
- **ACTION ITEM:** The Clinical Annotation workgroup will create a small group devoted to the development re-phenotyping pathogenic and likely pathogenic VUSs and propose options at the next Clinical Annotation workgroup call Tuesday, February 13th.

EHR Integration workgroup progress report | Casey Overby Taylor *(Geisinger/JHU)* & Sandy Aronson *(Harvard)*

- Continuing monthly milestone tracking across the sites with a focus on three main domains: retrieving and storing structured results, return of results to clinicians, and return of results to participants.
- To address the ESP's recommendation, file format and tool sharing has increased.

- There has been a rise in confidence with the XML format the group generated as it becomes more validated Network-wide.
- The Lynch Syndrome CDS guide was shared which was developed jointly with CSER and DIGITizE.
- The workgroup has publications completed and in progress. Luke Rasmussen has represented the workgroup well on various panels at conferences.
- Next steps for the workgroup were detailed:
    - Lessons learned surrounding processing batched reports should be documented.
    - Generating a simple framework to capture a range of decision support interventions, including passive and active intervention types, as well as pre-test and post-test timing. In this framework, workgroup is including non-technical approaches with potential to be enabled by technology in the future.
    - Delivering eMERGE community content using electronic decision support and creating avenues to explore the purpose of intervention (e.g. screening) and genetic test result types (e.g. risk actionable).
- **ACTION ITEM**: Sites should update and cite key next steps in the EHRI tracking grid
- **ACTION ITEM**: Sites should indicate passive or active approaches taken for clinical decision support to pre and post genetic testing
- **ACTION ITEM**: The EHRI workgroup should document lessons learned surrounding integration of batched reporting of results

Science presentation: Strategies and tools for combining research discovery with N-of-1 reporting | Richard Gibbs *(BCM/HGSC)*

- Richard presented the contrast of clinical and research data tools, examples of synergism between the tools, and opportunities.
- There were over 28,000 manual report reviews done off the 25,000 participants and 109 genes.
- Neptune data tool components extract the PHI data from the data to be made public. eCAP contains BAMs, VCF, and metadata for research in a de-identified space. eDAP contains support for clinical decision management.
- Mosaicism examines read ratios, homozygotes read 100% one allele, heterozygous have 50/50 reads typically, but ultimately mosaics skew the reads.
- ANTON Data Lake was tested with the genotype data leading to result that adding the phenotype data accelerates discovery.
- In conclusion, the study found that the rate of variant interpretation is limiting, eMERGE tool sets provide research and clinical interoperability, phenotypes are needed beyond the indications for testing. There is a proposal to add in phenotypes at diagnostic stage across the program.
- The group discussed that data on mosaicism is observed in about 2000 of the participants which may indicate that these variations exist in a large percentage of the population. There a question of whether the WGS data can be leveraged with this approach.

Genomics workgroup progress report | Megan Roy-Puckelwartz *(Northwestern)*, Patrick Sleiman *(CHOP)*, & David Crosslin *(CC/UW)*

- Data freeze of eMERGEseq will be multisample called and will create a research data set for analyses. BAMs can be downloaded from Aspera currently.
- The 5000 extra participants from Harvard for GWAS set have been imputed and they will re-merge a V2 eI-III imputed GWAS dataset.
    - The group depicted the genetically determined ancestry in eMERGE imputed array dataset

    - Relatedness as a confounder in this dataset was highlighted. Group is evaluating different methods to include/exclude relatedness in the dataset. Currently, keeping only one member from each family (consequently losing ~8000 participants)
- SPHINX has been enhanced to incorporate search via chromosome position or rsID. Prior clinical associations from GWAS catalog have been annotated. Additional discussion needed for incorporating other datasets into SPHINX, including the clinical variant data
- Group has elected to pursue three projects DNAnexus collaborative projects that leverage the eMERGEseq data.
    - CCHMC: a multisample calling of all eMERGEseq subjects using resources in DNAnexus. This provides a resource for both eMERGE and the community at large.
    - CHOP: proposal to look at copy number variation in the eMERGEseq data. This will also provide a resource for both eMERGE and the community at large.
    - Baylor: candidate sites of mosaic alleles project detailed in Richard Gibbs' science presentation.
- Best practices for future panel discussion:
    - Determine effectiveness of several variant callers
    - Efficacy of multisample calls and the differential effects of multisample calling in different ancestry groups
    - Need for Sanger sequencing for confirmation of high quality next-gen sequencing
    - Best practices for dealing with relatedness in large samples
    - Most effective uses of SPHINX
    - Best practices for CNV calling, both eMERGEseq and WGS cross-comparison to see which tools work best.
- **ACTION ITEM**: The Genomics workgroup will define datasets to be contained within SPHINX, and best approaches for utilization.
- **ACTION ITEM:** Adam Gordon and Heidi Rehm will draft an MCS to address if sanger confirmation of sequencing data is still required
- **ACTION ITEM:** The three joint eMERGEseq DNAnexus project leaders will submit accompanying MCS.
- **ACTION ITEM**: The CC will add an option to choose which workgroup is affiliated with the MCS in the publication tracking REDCap survey to facilitate workgroup paper updates.

PGx workgroup progress report | Laura Rasmussen-Torvik *(Northwestern)* & Cindy Prows *(CCHMC)*

- The group presented current site plans for returning PGx results
- Six MCS have been published, six more moving towards publication over the next several months Future publications include PGx SNP PheWAS and additional knowledge generation for SNPs near CPIC guidelines.
- eMERGE III SNp list
    - Over 100 PGx SNVs on the eMERGEseq platform, even though there are only a few being returned. This is a valuable resource for future research. Expands the sample size 25,000 participants compared to the 9000 in the PGRNseq dataset.
    - Most recent list has been circulated to both the Genomics Workgroup and PGx Workgroup.
- Developing phenotypes on the PGx data has proven challenging, as phenotypes are more complex, however the PGx phenotypes were not prioritized (Metaformin response and response to HF medication).
- Our current data on the PGRNseq cohort were collected one time in the spring of 2015, and have not been updated. Date of return of results have not been reported and may also be useful.
    - One possible solution is to refresh the data.
    - The date of return of result data would be a new element that has never been collected.
        - For the eMERGEseq dataset the return of the core variants may differ from the return of PGx variants. The Network is collecting the date of the core variants only.

    - In addition to site costs, the coordinating center would have to handle the collation of this new data which could be burdensome. Additionally, Mt. Sinai and Boston Children's are two eII sites that are no longer funded and would likely not send data.
- Sites need to determine how much effort is involved in pulling these data. If we refreshed just the common variables data, sites could incorporate that into the twice yearly pulls for the GWAS and eMERGEseq set. The issue surrounds the 'date of return' and if and how sites would be able to pull this data for both the PGRNseq data set and then also the eMERGEseq data set.
- Group is recommended to pursue R01 opportunities for return of CYP2D6.
    - Opioid response is another R01 opportunity to explore.
- **ACTION ITEM:** Laura Rasmussen-Torvik will develop a survey to determine the burden and benefit of a data refresh of the PGx cohort, including common variables, a medication data dump, and date of return into the EHR.

ROR/ELSI workgroup progress report | Ingrid Holm *(BCH)* & Iftikhar Kullo *(Mayo)*

- The goal is to return results by June 1st, 2018 and about 20% progress has been completed.
    - KPW/UW, Mayo, and Northwestern have returned results, other sites plan to return shortly.
- The heterogeneity of ROR at the sites is actually a strength, as we have many models being tried at one time. This will allow for more data collection and help better inform on best practices.
- The workgroup has completed two data collection instruments to use ROR process data: Participant survey and Healthcare Provider survey.
- There was discussion of ways to address and access what has been learned from the return of result process using the RE-AIM framework and data collection tools.
- Future possibilities of the ROR/ELSI workgroup in eMERGE IV:
    - Collaborate with the Outcomes workgroup to identify common ROR outcome measures across other NHGRI funded projects (CSER & IGNITE), and convene a ROR subgroup for this project.
        - Many of the CSER harmonizing measures are being completed by the ROR workgroup.
    - The versioning of results will be an important element to consider, as new data is incorporated it may change the impact of the results on the participants.
- EHRI-ROR/ELSI joint project, *Preferences for research updates among eMERGE biobank participant*s, will be circulated for authorship soon.
- IRB Perspectives project, *Institutional Review Board Responses to Targeted Genome Sequencing Projects in the eMERGE Network,* was accepted for publication.

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

- The group completed development of a distributed outcome measurement tool using REDCap including 17 data collection instruments corresponding to the process of return of result and disease/phenotypes.
- Sites are working through the beta test phase, plan to release final form by February 2018.
- Future work on variant interpretation cases and with the ROR/ELSI workgroup to address common outcome measures across other NHGRI funded projects (CSER & IGNITE).
- There is an opportunity to use the eMERGE-III extension to complete a follow up EHR abstraction.
- **ACTION ITEM**: The CC will compile all site Outcomes data and review/summarize for interim analysis to present at the October 2018 meeting.
- **ACTION ITEM:** The CC will begin creating an "abstracter's guide" for the Outcomes workgroup forms detailing answers to questions and defining the ROR process.

Closing remarks | Rex Chisholm *(SC Chair, Northwestern)*

- Excellent meeting and discussion, especially due to the deadlines arriving in the next year.

Action Items

PIs/Sites

- Site PIs are asked to complete ROR timelines by February 5th, 2018, prior to the NIH Council Meeting
- The SC is asked to suggest topics for panel discussion on 'Best Practices' for future eMERGE III SC meetings.
- Sites are reminded to send tools they have developed in response to eMERGEseq to CC to make widely available to the Network on eMERGE website.
- Sites should plan to return the majority of Phase I eMERGEseq results to participants by May of 2018.

Coordinating Center

- The CC will add an option to choose which workgroup is affiliated with the MCS in the publication tracking REDCap survey to facilitate workgroup paper updates.
- The CC will compile all site outcome data and review/summarize for interim analysis to present at the October 2018 meeting.
- The CC will begin creating an "abstracter's guide" for the Outcomes Workgroup forms detailing answers to questions and defining the ROR process

Phenotyping

- The Phenotyping group will recommend autoimmune variables to the common variable list

Clinical Annotation

- The Clinical Annotation workgroup will create a small group devoted to the development re-phenotyping P and LP VUSs and propose options at the next Clinical Annotation workgroup call Tuesday, February 13th.

EHRI

- Sites should update and cite key next steps in the EHRI tracking grid
- Sites should indicate passive or active approaches taken for clinical decision support to pre and post genetic testing
- The EHRI workgroup should document lessons learned surrounding integration of batched reporting of results

Genomics

- The Genomics workgroup will define datasets to be contained within SPHINX, and best approaches for utilization.
- Adam Gordon and Heidi Rehm will draft an MCS to address if sanger confirmation of sequencing data is still required
- The three joint eMERGEseq DNAnexus project leaders will submit accompanying MCS

PGx

- Laura Rasmussen-Torvik will develop a survey to determine the burden and benefit of a data refresh of the PGx cohort, including common variables, a medication data dump, and date of return into the EHR.

ROR/ELSI

- ROR/ELSI workgroup will create a subgroup to apply RE/AIM framework to ROR process to identify common methods, patient and provider survey, familial implications of ROR

Outcomes

- Outcomes workgroup will use Outcomes Protocol Database to collect specific outcomes for all disorders, assess ease and accuracy of outcome collection, and work with CSER to determine if behavior/outcomes is related to return of result