# Summary of Steering Committee Meeting: Fall 2018

October 25-26th, 2018 in Rockville, MD

## TABLE OF CONTENTS

- Panel: Lessons learned summary – ROR/ELSI, Phenotyping, EHRI | Ingrid Holm *(BCH)* Iftikhar Kullo *(Mayo)* Peggy Peissig *(Marshfield)* George Hripcsak *(Columbia)* Casey Overby *(Geisinger/JHU)* Sandy Aronson *(Harvard)*
- Genomics: Workgroup progress, future directions & data set updates | David Crosslin *(UW/CC)* Megan Roy Puckelwartz *(Northwestern)* Patrick Sleiman *(CHOP)*
- PGx: Workgroup progress & future directions | Cindy Prows *(CCHMC)* Laura Rasmussen-Torvik *(Northwestern)*
- ROR/ELSI: Workgroup progress & future directions | Ingrid Holm *(BCH)* Iftikhar Kullo *(Mayo)*
- EHR Integration: Workgroup progress & future direction | Sandy Aronson *(Harvard)* Casey Overby *(Geisinger/JHU)*
- Network discussion: Input/Feedback from the ESP | ESP Members & NHGRI Team
- Closing remarks | Rex Chisholm *(SC Chair, Northwestern)*

ACTION ITEMS

Presentation slides and meeting web page *(login required)*

ESP Executive Session notes & recommendations

## DAY 1: Thursday

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

- FY19 Budget: 39.1B (NIH); 556M (NHGRI) which is an increase from FY18 of approximately two billion.
- All of Us Research Program has announced that the new Chief Medical Officer is Kelly Gebo, MD, MPH from Johns Hopkins University. In addition, on September 25th, 2018, All of Us Research Program announced funding for genomics center totaling $28.6 million. Awardees include: Baylor College of Medicine w/ Johns Hopkins; Broad w/ Color and LMM; University of Washington.
- The "Genomics 2020" Strategic Planning Timeline was launched in February 2018. The timeline includes a final meeting in April 2020 and a manuscript submission in July 2020. Meetings will consist of workshops, townhalls and gatherings at existing meetings. In addition, there is a dedicated web page, social media tools and engagement of advisory groups (NHGRI Strategic Planning Process).
- The eMERGE III timeline estimates the dbGaP data submission to be completed by February 2019.
- NHGRI eMERGE Program Staff has requested that eMERGE proceed with dissemination of lessons learned in the following topic areas:
    - Implementation of OMOP + NLP to improve the portability of electronic phenotype algorithms.
    - Developing and implementing initial clinical decision alerts for genetic test results across multiple clinical sites.
    - Implementation and use of infobutton from health care providers and patients.
    - Development and implementation of a multi-sequencing center variant recalling pipeline.
    - Utilization and outcomes of geocoding data for eMERGE research.
    - Economic analysis of providing genetic testing.
- NHGRI Training Programs: Funding Opportunities are training programs to help the Network increase manpower for projects. One initiative is the Genomic Medicine T32, which is a postdoctoral training program focused on genomic medicine. The NIH is engaging diverse students via the Diversity Action Plan

(DAP) which includes educational activities and research experiences focused on exposing students to genomic medicine. The Clinical Investigator Development Award (K08) focuses on genomic medicine. The Research Scientist Development Award (K01) focuses on genome science and ELSI. Lastly, the NIH Loan Repayment Program is for US citizens, US national, or permanent residents with doctoral level degrees with qualifying educational debt equal to or in excess to 20% institutional base salary.

- NHGRI wants to know if the Network considers the program reports useful. Please provide feedback in the meeting follow up survey sent by the CC.

**Announcements, opening remarks** | Rex Chisholm *(SC Chair, Northwestern)*

- The Winter 2019 SC/CSER joint meeting will be held **January 17-18, 2019** in Bethesda, MD. The meeting will include a panel discussion on Outcomes lessons learned.
- The Spring 2019 ESP Conference Call will be held **Monday April 29th, 2019 at 2:00 PM EST**.
- The Summer 2019 SC meeting will be held in Seattle, WA on **June 20 & 21, 2019**. The Clinical Annotation workgroup will present lessons learned from penetrance analysis work.
- Developments since the June 2018 SC meeting include:
    - Network Workgroups have produced timelines and milestones for eMERGE Year Five.
    - Investigators have translated eMERGE cohort into OMOP common data model and completed the testing of Type II Diabetes electronic phenotype.
    - The community PheWAS Catalog tool was launched within PheKB to expand utilization of published PheWas studies and build a community resource.
- 692 eMERGE projects have been published since September 2007, including five Network projects that have been published since the June 2018 SC meeting. There have been 26,238 cumulative citations across eMERGE Network publications.
- There have been over 1,200 external eMERGE dbGaP downloads as of September 2018.
- The Network is strongly encouraged to focus their efforts on publishing their work and the lessons learned.
- Goals for Years 4 and 5 include publishing manuscripts, especially surrounding lessons learned, finalizing return of results and initiate outcomes data analysis, focusing on Natural Language Processing (NLP) phenotypes, and completing the final eMERGEseq dbGap data variables for submission.
- There was discussion surrounding if an infographic could be created specifically for the PheWAS Catalog, in order to represent the phenotypes we've generated against the genetic data.
    - This could be modeled from the GWAS graphic, which is displayed on the NHGRI website.
    - **ACTION ITEM:** The CC will produce an infographic for the phenotypes that have been generated on the genetic data.
- Harvard has another 20,000 GWAS samples forthcoming to the Network from the Illumina Mega Array, which will push the eI-III Merged Imputed Data Set well above the 100k threshold.
    - The diversity in the biobank include 15% African American and 15% Hispanic. This minority sample is reasonably representative of the greater Boston area population.

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

- The eMERGE imputation manuscript led by Ian Stanaway, *"The eMERGE genotype set of 83,717 subjects imputed to ~40 million variants genome wide and association with the herpes zoster medical record phenotype"*, has been accepted to *Genetic Epidemiology*
- Ian presented the methods behind the PGx PheWAS manuscript:

    - From 9,007 PGx participants, about 60,000 variants were sequenced in nearly 84 genes with pharmacogenomic significance. There were 652 PheWAS Codes with a Minimum Case Count of 200.
    - 106 variant positions were found in the PGx DNA sequencing which has a matching rsID in the GWAS catalog. 99/106 common GWAS Catalog SNPs with a minor allele frequency >0.01 in the PGx. ~33 SNPs do not have an EMR Phecode that makes sense and of these, ~1,021 regressio with EMR codes match a GWAS phenotype. ~66 SNPs have one or more EMR Phecode.
- BEAGLE phasing and imputation of structural variation using the ~100k eMERGE array samples is planned.
    - The goal is to impute common structural variation (both indels and larger events) for 100,000 eMERGE array samples, and link to phenotype data derived from EHR.
    - Multiple MCS's have been submitted for discovery analyses with these genetic event, and orthogonal data has been used to validate some events.
    - Brian Browning, the developer of BEAGLE, is interested in this project and will use the reference that is being created.
- SPHINX now has incorporated eMERGEseq data freeze samples, gene ontology pathways, gene percent exon coverage, NHGRI GWAS catalog variants in PGRNSeq and eMERGEseq, and drug-gene interactions from DrugBank. The CC is also working on a rebranding of SPHINX into eMERGENT (Electronic Medical Records and Genomics Toolkit); however, this is contingent upon receipt of external funding. A U24 supplement will be submitted in January 2019 by David Crosslin.
- The Network discussed how are drug-related phenotypes captured in the PheWAS catalog.
    - Pleiotropy has been noticed when seeking associations in the data (tacrolimus as an example).
    - There are discrepancies between EMR and what is being found in the catalog, such as with CYP2A6, which is for Epilepsy loci. CYP2A6 is a a caffeine metabolizer, where we know that valproate acid is metabolized by CYP2A6. This represents the fact that what we expect to see isn't always what is demonstrated.
- There was a discussion to determine if raw fluorescence data are needed for structural calls and experience harmonizing raw arrays and raw fluorescence.
    - In regard to the structural variant calling and how that is concerned with the use of arrays, Ian has not focused on that on the GWAS set as of yet, but has in the past with others. The GWAS dataset has different sensitivity based upon batch and Ian is focusing on the implications of this.
- When conducting the PGx PheWAS Ian mainly examined statins, specifically related to CYP2A4. This may be related to myalgias and pain in muscles. Though the analysis demonstrates an interaction and a significant effect, he can not speak to causality.
    - There are also an example of epilepsy phenotypes from CHOP in relationship to CYP2A6. Follow up and review of the drugs being prescribed for these individuals is needed, in order to determine any clinically significant relationship.
        - Due to the nature of GWAS, as a dataset, an investigator must adapt to the complexities. GWAS has one phenotype and a set of genes, so an investigator must know everything about that one phenotype. For PheWAS, you have to know about many different phenotypes. This makes the analysis a bit more difficult, as expertise is needed in a variety of areas.

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

- Partners-Broad
    - Reporting is complete. There are a handful of reports that Geisinger has to re-consent, but they will either be consented or withdrawn shortly. Sites that have received sample sequencing by P/B have received automated reports via GeneInsight.
    - There are 32% of cases with indications (n= 3,356/10,500); 2% cases with indication-based returnable results (n= 58/3,356); 4.5% with non-indication-based consensus returnable results (n= 477/10,500); 2% are copy-number variants.
    - The distribution of case reports by variant classes includes 596 unique variants across 750 case reports. Of note, there has been reclassification of 1.2% of variants affecting 16 patients since original reports (n= 7/596). Triggers for reclassification include new cases being observed or being contacted by an external investigator.
- Baylor
    - Only 22 reports remain left to be sequenced, not including Marshfield.
    - *"Atlas-CNV: a validated approach to call Single Exon CNVs in the eMERGESeq gene panel"* a manuscript on Baylor's improved CNV calls with method optimized for gene panels is under review at *Genetics in Medicine*
        - Authors have developed a new quality score and used MLPA confirmations to refine thresholds. There are fewer no-calls, a higher confirmation rate (lower FDR), and higher resolution (single exons). Single exons and previous no-calls will be part of reanalysis.
    - Out of 4,245 indication based returnable results, 3.2% were positive compared to 96.8% negatives. Of the 14,568 non-indication based site-specific returnable results, 3.8% were positive compared to 96.2% negative. 2.5% of the 8,536 of the non-indication based site-specific returnable were positive, compared to 97.5% negative.
    - Baylor plans to coordinate with multiple groups to adapt HL7 FHIR standards for structured data within the eMERGE Data Commons in Year 5. Baylor will also coordinate with GeneInsight and Health Level Seven International to transform the current eMERGE .xml tables into HL7 FHIR compliant format in this time. There is consideration of including Baylor-specific fields as well.
    - ClinVar submission is complete (SUB4151980), noting that 211 distinct Path/Likely Path variants were reviewed in eMERGE.
    - The VUMC ID issue has been worked through.
- The ClinVar submission does not include Benign/Likely Benign; however, the CSGs may reconsider. Need to define the criteria for benign, likely benign, pathogenic, likely pathogenic.
- The CSGs complete a harmonization every quarter to ensure cohesion between the two centers.
- UW has discussed clean up on unselected cases to solve issues that vary by site and sequencing center that impact the incidental findings. Adam Gordon has presented on this topic at the 2018 ASHG meeting.

**Network discussion: eMERGE Commons advantages, disadvantages, lessons learned** | Moderators: Will Salerno *(Baylor)*, Richard Gibbs *(Baylor)*, Eric Venner *(Baylor)*, John Didion *(DNAnexus)*

- The purpose of the eMERGE Data Commons is to support management of sensitive data for clinical reporting, maintain a data repository, provide an environment for large scale analyses, and act as a model for future programs
- The group conducted a survey on the effectiveness of eMERGE Commons aims:
    - To support management of sensitive data for clinical reporting scored: 8/10.

    - To maintain a de-identified data repository scored: 9/10. This refers only to the eMERGEseq dataset.
    - To provide an environment for large scale analyses scored: 6/10. This is represented by three large scale projects, including; multiple variant callers, SV calling, and mosaicism.
    - As a model for future programs scored: 8/10.
- Overall, a main lesson learned dealt with the issue of relabeling the VUMC IDs due to their IRB requirements. Infrastructure is needed to make the changes in a more streamlined and timely fashion going forward.
- Heterogeneity of the Network remains a challenge, as there are different types of data, users, ELSI, infrastructure, and analysis that must be addressed by eMERGE Commons (DNAnexus).
- NIH Data Commons Initiatives Model Organism Database (MOD), the NHGRI Genomic Data Science Analysis, Visualization, and Informatics Lab-space (AnVIL), and the NHLBI's Data, Storage, Toolspace, Access and analystics for biG data Emprowerment (DataSTAGE) brings large data and computations to many users by establishing security, federated access, consent management, scaling, standards, and utilizing FAIR (Findable, Accessible, Interoperable, and Reusable) principles.
- Focused Data Commons (e.g. eMERGE, CHARGE Commons, pFDA) serve specific communities through architecture for specific security, analysis, and data science needs. This enables rapid development and new feature cycles, drawing on federated commons resources, and allowing investigators and engineers to collaborate on actionable analysis.
- Eric Venner (Baylor) highlighted positive aspects of DNAnexus. For instance, the compliance is handled for DNAnexus, including authorization, authentication, encryption and HIPAA compliance. In addition, DNAnexus is extremely powerful, noting that their team has worked with Amazon to help hide some of the underlying complexity. Each small detail is individually complex, and when performed on the cloud level, there are multiple layers involved that contribute to the performance and power of the system.
    - One drawback is the steep learning curve which makes it difficult to utilize initially. Without a way to ask questions regarding how to build large systems of power, such as an onsite IT center, this makes the Commons more challenging. Cohort management is difficult when you have a large set of samples that can overlap from many states. Engineers are left to build their own design patterns that can be useful for launching jobs on the cloud.
- Ian Stanaway *(UW/CC)* utilized DNAnexus a few times, but ultimately had to use UW's system to complete analyses. UW's system has a set cost per space, and does not cost per run, however runs in DNAnexus have associated fees, even though consumers do not have to pay for the storage space itself. At this time, it is necessary to estimate the cost of the potential of this work, given the bulk of the data and due to the fact that the data is in transition.
    - Ian suggests creating a more standardized UNIX environment that teams can leverage. This can bridge the gap between the old system (i.e. where the imputation data is still housed) to the DNAnexus commons environment where the eMERGEseq data is housed. Investigation and identification of the cost-benefits of utilizing their own system versus Commons.
- Bahram Namjou-Khales *(CCHMC)* agreed that DNAnexus analyses are very time-efficient and can be conducted within a week. Applications are easy-to-use and straightforward, however can be corrupted, missing, or inactivated.

    - It is not easy to activate all BAM files at once across all subjects. It would be more efficient to put all clean/unique BAM files in a single folder to help minimize time spent searching and triaging files. The eMERGEseq data freeze 1 set is QC'ed, so it is not corrupted.
- Patrick Sleiman *(CHOP)* highlighted that the cloud useful when there is an established pipeline, and can analyze the ~12 million samples very quickly and efficiently.
    - Would be great to leverage with a fully operational pipeline designed for this tool.
    - Challenges: Commons are not flexible and cannot interact with other files and software easily.
    - DNAnexus Commons point of contact for user support was not clearly defined. The CC can help facilitate awareness, if changes are quickly communicated.
- John Didion *(DNANexus)* is the new point person for Baylor and eMERGE Commons and is now the principal scientist for DNAnexus.
    - Plan to provide more training solutions regarding DNAnexus.
    - DNANexus has designated a monitored support email for eMERGE users (eMERGEsupport@dnanexus.com). Responses will be provided within 1-2 business days depending upon the type of issue.
- Cost benefits depend on usage and deployment. If the investigator does not use it frequently the cloud is more efficient.
    - One of the goals of the AnVIL is for the NIH to provide free storage for large genomic datasets, including eMERGE, which can be harmonized with another program. The plan is to use commercial clouds as opposed to having them internally built. Phenotyping data harmonization across all NHGRI-funded programs is another forthcoming initiative.
    - AnVIL will be an external storage site, viewed as an extension of dbGaP. Data will remain available even if eMERGE as a Network does not continue and will be under controlled access.
    - **ACTION ITEM:** eMERGE NHGRI Program Staff, in coordination with the CC, will invite AnVIL team to present to the Network at an upcoming SC meeting or PI meeting.

**Science presentation: The cascade screening for hypercholesterolemia (CASH) trial: a progress report** | David Kochan *(Mayo)*

- To inform guidelines for screening family members of patients with Familial hypercholesterolemia (FH), the team designed a clinical trial to assess the yield of cascade screening for FH when Health Care Providers (HCPs) can directly contact family members and compare yield of cascade screening in FH patients with and without an identifiable pathogenic variant.
- Out of the total sample eligible for the CASH Study (n = 2,538), 358 total FDRs were FH negative and 97 total FDRs were FH positive.
- The conclusions from an interim analysis showed that the yield in the FH positive cohort with direct contact was 0.7. This is higher than yields reported with 'usual cascade' screening (~0.3). Yield in the FH negative cohort was much lower compared to FH positive members (0.1 vs 0.7, respectively). Cost analyses of screening in the two groups of FH patients remains an ongoing process. The yield in the FH negative cohort with direct contact was 0.1, compared to the MyCode *(Geisinger)* yield of 0.3.
- Next steps include completing enrollment of FH positive probands, cost analyses, and analyses of correlation between Polygenic Risk Score and yield. These data may lead to a larger trial as well.

- The group mailed an interest sheet to family members once the proband provided their contact info. If the individual was interested, the team would then email more additional information. The form did not specifically state a family member had been diagnosed with FH in the original mailed solicitation.
- David noted that the study had 12 positives, 22 negatives, and an overall was able to obtain approximately a rate 30% for cascade screening.
- *Comparisons to other studies:*
    - The Dutch implemented a proactive approach (knocking on doors, nurses taking blood) and there is a lack of stigma in the Netherlands. There is also a nationalized health care system, therefore this creates conditions that are not replicable within the US. This indicates that this comparative group is an outlier. Of note, there was funding previously but now that funding is gone the ascertainment is similar to other studies.
    - As a limitation, the majority of participants were also Caucasian; study is lacking in racial/ethnic diversity (recruitment from minority groups was not part of study design.)
- LDLs were not available for these relatives.

**Science presentation: Excess genetic variation in understudied populations confounds interpretation of medically actionable genes |** Megan Puckelwartz *(Northwestern)*

- Additional comments or questions about this work or following manuscripts can be sent to Tess Pottinger (tess.pottinger@northwestern.edu).
- NUgene Biobank Whole Genome Sequencing cohort (n=895) was sequenced at Washington University.
- Strong concordance (71%) found between self-reported race and genetic race in biobank participants: 95% African ancestry, 97% European ancestry, 70% Hispanic ancestry, and 22% Mixed ancestry.
- African ancestry subjects have more variation across genome than other groups. In the study, African ancestry includes African Americans.
- Both Hispanics and those of African ancestry are scattered across the concordance map, however on the plots for variants, those of mixed cohort have a broader range. This could possibly be due to two distinct Hispanic populations and when considering outliers. The analysis is not anchored within 1,000 Genomes as of yet.
- ACMG genes total about 0.5 ACMG VUS per subject with range 6.0. African ancestry NUgene biobank subjects still have more ACMG VUSs even after normalization and more cardiac ACMG VUS.
- In conclusion, African ancestry subjects have more total ACMG gene VUS's and more ACMG cardiac gene VUS's even when normalized for total number of variants. Both total and cardiac ACMG VUS's identified in African ancestry subjects have fewer ClinVar reports. These VUSs may be medically actionable, especially for the cardiac genes, and require further investigation.
- There was not an enrichment of those of African ancestry in their cardiomyopathy cohort.
- Discussion arose around nonsynonymous variants that are not in ACMG gene list, such as those that are specific for individuals from African ancestry. These may not be applicable to this population due to the fact that they are carrying more variation overall than non-minority races:
    - ACMG does not curate variants, they would be found in ClinVar.
- The group will further investigate these VUS's and the clinical implications they may have on the entire cohort.
- Allele frequencies in those of African ancestry vs. those of other descent were shown to be different, in addition to the global allele frequency of those of African ancestry vs. those of other descent.

- There are tools available to assess whether a VUS is associated with African ancestry.

**EHRI lessons learned and preparing for the future** | Moderators: Sandy Aronson *(EHRI Integration Chair, Harvard)* & Casey Overby *(Geisinger, Johns Hopkins University)* | Panel presentations can be downloaded from Casey Overby Taylor, Nephi Walton, David Fasel, Robert (Bob) Freimuth, Mullai Murugan, and Sandy Aronson.

- The workgroup conducted two surveys of EHR capabilities and implementation approaches.
- The EHR workgroup members and panelists include: Nephi Walton *(Geisinger)*, Mullai Murugan *(Baylor)*, Bob Freimuth *(Mayo)*, Sandy Aronson *(Harvard)*, & David Fasel *(Columbia)*.
- Nephi Walton posed the question of whether data has to go through a Laboratory Information Management System (LIMS) before the EHR. The genomic indicators are now available in Epic, which enables integration of discrete data into EHR systems. Discrete data is also viewed and utilized differently from scanned reports (PDFs), including that it:
    - Is more visible and potentially actionable in the sense that it provides discrete genetic data in EHR, which can facilitate the development of future clinical-decision support tools. There is also a need to define the type of information that should be displayed with the variant, such as shown with infobuttons.
        - A support structure is needed for physicians to understand what 'variant' means.
    - There are challenges with implementing Infobuttons within the EHR, including prioritization within the EPIC team. A system-wide integration effort is complex, requiring individuals overseeing content, understanding utility, and sourcing content from websites like PharmGKB.
        - Nephi offered possible solutions to these challenges for implementing Infobuttons within the EHR. These include employing a clinical informatics fellow, identifying an individual who will be responsible for content, demonstrating benefits outside of genetics, and applying these benefits to genomics.
    - Individually reviewing reports to determine inclusion/exclusion may be unsustainable. Nephi is developing an MCS for this process. There are certain criteria that variants have in place; however, others will require manual curation to determine if they should be included in the EHR for actionable and clinically relevant results.
- Mullai Murugan reviewed challenges the Baylor CSG encountered and attempted to overcome when transfering data for integration to the sites EHR systems, which include the complexity and heterogeneity of the Network and the difficulty harmonizing data and delivery.
    - The multi-institutional Network lends itself to complexity issues; however, by implementing variant harmonization, ontologies and coding systems, implementing structure and content of accessioning and reporting data in a HIPAA compliant manner (DNAnexus) some of these difficulties can be overcome.
        - The recently conducted EHRI Survey show that 7/10 eMERGE sites use structured data. Therefore, standardization would best assist with easier integration and communication. To do this, GeneInsight (GI) XML format was identified and the JSON2XML component was built. JSON to GeneInsight XML Converter Differing JSON and XML formats, needed to be converted and aligned. DNAnexus executes JSON2XML which conducts the mapping.
    - Mullai discusses future directions for EHRI, include identifying a need for adopting a national standard, possibly by leveraging the HL7 FHIR Standard for Clinical Genomics (Fast Healthcare Interoperability Resources).

        - The proposed plan for FHIR is to understand, consolidate, and finalize the FHIR genomics specification and genomics bundle for eMERGE. Next, to map to BCM/LMM existing report structure/content and reconcile differences with HL7 Clinical Genomics. From this, .xml(s) will continue to be generated. FHIR will be piloted for selected use cases around third quarter of Year 4.
        - There is a dedicated HL7 Workgroup for clinical genomics; Bob Freimuth is a chair.
    - Baylor identified a need to design tracking, dashboards, and analysis tools to facilitate monitoring and consumption of the data they were generating and developed eDAP (eMERGE Data and Analysis Portal) and eCAP (eMERGE Commons Access Portal).
- Bob Freimuth discusses the standards for sequencing results. The goal is to develop a common structure in order to increase interoperability within eMERGE, as there is a lack of standard computation representation for discrete variant data.
    - Bob addressed how eMERGE has developed their own XML schema, which was computationally friendly and easy to parse. For Mayo it was difficult to insert into their EHR due to EHR restructuring. There were also some difficulties obtaining institutional approval for building within the EHR, both on the previous and new EHR system.
    - Less than 5% of the project remained when the change to the new EPIC EHR system occurred. Therefore, the group had to design a new workflow and new ROR process. This project team workflow was more complex than envisioned. Additionally, changes in personnel have large effects on the process, and required the need for correct/complete process documentation and hand off procedures.
    - Future directions involve developing new CDS capabilities (including the management of FH vs PGx result/drug alert) getting discrete results into the new EHR, and building on highly successful PGx CDS implementations (eMERGE II and Center for I Medicine)
    - Stakeholders within this niche field of standards are coming together in order to address the need for rethinking genomic standards and push forth a few above the rest. They understand that creation of a unified standard may not be feasible at this point.
- Sandy Aronson created a first attempt at generating a genetics based clinical decision support rule, and discussed lessons learned from this task.
    - Variant by variant coding of genetic clinical decision support is infeasible. Support must be driven off a knowledge base. In order to address this, his team launched GeneInsight to bridge a gap between increasing genetic knowledge and EHR systems.
    - Another challenge was with interjecting functionality into a clinical workflow by introducing clinical decision support, which can derail the workflows. In addition, clinical decision support can alter this clinical workflow. Tight collaboration between clinicians, laboratories, and IT professionals is needed to do this well. This does not stop at launch, it continues throughout the lifespan of the CDS.
        - Specifically, Genetic Aware Clinical Decision Support is entirely dependent on access to structured data and structured knowledge. However, these data are highly fractured across the health care system.
        - "*Genetics Aware Clinical Decision Support*" chapter of book titled Foundations, Translations & Implementations (Aronson & Williams, 2016).

    - Reports are created through GeneInsight, which creates structured patient variants. Knowledge is curated within GeneInsight, which leads into this structured knowledge.
    - It is common for genetics to come from outside sources because knowledge is distributed.
- David Fasel reviewed that after Columbia receives the report, they are inputted into REDCap as a researcher interface. The advantage of parsing the HTML to XML is that it allows for more control over what authors want to display, as well as that it makes it easier to put in helpful links for the viewer. Results returned to EHR after counseling session
    - Challenges with this process include compliance restrictions with NY State on viewing genomic data, as NY law is strict about who can view genetic data. Second, although pushing the structured data into the EHR was fairly straightforward, there were still challenges in this process. For the positive reports, many documents and notes from genetic counselors had to be extracted, mapped, and pushed by hand and was associated with a large time and effort burden.
    - Identifying previous genetic diagnosis in the EHR can be troublesome, as searching for gene names in the EHR notes gives false positives.

**Science presentation: Detecting potential pleiotropy across cardiovascular and neurological diseases using univariate, bivariate, and multivariate methods on 43,870 individuals from the eMERGE Network** | Kogasudha (Sudha) Veturi *(University of Pennsylvania)* & Xinyuan (Blair) Zhang *(University of Pennsylvania)*

- In this study, Investigators address the link between myocardial disease and acute neurological diseases with cardiac failure, late-onset Parkinson's disease, Cardiovascular disease pathways, and Alzheimer's disease.
- The goal of this research is to identify the genetic contributions to the link between myocardial disease, acute neurological disease, and previously mentioned potential comorbid diseases.
- This study focused on eMERGE phase III imputed data. Authors used 533,878 SNPs; 65 phenotypes; n = 43,870, and conducted association analyses, adjusted by age, sex, eMERGE site, 6PCs. Investigators used hypertensive heart disease ICD9 (402) major groups and ICD9 code ranges and evaluated univariate, bivariate, and multivariate methods for detecting pleiotropy using eMERGE Phase III imputed genotype data and EHR data.
- Findings showed 31 Bonferroni significant variants overlapping all three methods using a very conservative threshold.
- Goal was to identify 200KB region around leap SNP, then find if the lead SNP is an eQTL. If these are identified, then the group located the most significant corresponding eGene using statistical colocalization.
- Of the Bonferroni significant variants, nine SNPs (associated with at least one cardiovascular and neurological disorder) were also found to colocalize with eQTLs across a host of tissue types. The authors integrative framework not only identified novel signals but also validated known pleiotropic signals in a unified way for cardiovascular and neurological diseases. Two known signals both had cardiac disorders, and findings also showed seven novel signals.
- Future directions include to replicate analyses on Penn Medicine Biobank (PMBB) data and UK Biobank data, and include ICD10 codes in analyses, examine signals at less conservative thresholds, and evaluate additional multivariate and Network-based methods to fine tune causal associations.
- Regarding the observation of eQTL changes that are involved in the mechanism in the phenotype the group can only provide suggestive associations with high probability from colocalization analyses.

However, in the future, author is open to leverage Mendelian randomization method approach to determine associations

- The phenotypes were mainly based off of Interest, centering on brain and heart phenotypes.
- The group is open to this investigating if animal models can add information regarding the importance of these loci. However, she will first attempt to replicate these findings leveraging other dataset cohorts and methodologies.

**Science presentation: EHR and in vitro functional approaches for improved classification of cardiac ion channel variants** | Andrew Glazer *(VUMC)*

A vision for personalized medicine includes individuals from population, genetic testing, identification of pathogenic variants, and alteration of care. Andrew's presentation focuses on the identification of pathogenic variants.

- CaseStudy: The investigator discussed an eMERGE VUMC patient with pathogenic arrhythmia variant, KCNH2. The patient is 63-year-old African American female with history of hypertension, diabetes, and high BMI; no known history or family history of LQTS. She has undiagnosed phenotype positive LQTS.
- There are 8 ACMG59 arrhythmia genes listed on the eMERGEseq-109 gene panel. For example, SCN5A (cardiac sodium channel) is responsible for loss of function variants (Brugada syndrome) or gain of function variants (long QT syndrome).
- New variants are interpreted by the type of variant, *in silico* tools, allele frequency, functional data, and disease associations.
- Investigators calculated the "literature penetrance" of SCN5A variants through 711 papers and gnomAD to find estimates of BrS and LQT risk for each variant. SCN5A R1644C was found in six papers with a total of eight carriers: five had BrS1, two had LQT3, and zero had another disease. R1644C present in one out of 252,432 alleles in gnomAD and has been functionally characterized in one paper.
- Variants of special interests at VUMC include loss of function (nonsense/frameshift/splice), P/LP variants, "Semi-common suspicious variants" such as SCN5A R1193Q and SCN6A 1103Y, KCNE1 D85N (common risk allele) that VUMC is returning, KCNE1 D76N (six in VUMC, two outside of VUMC) of which that VUMC is returning variant and cascade screening six families, and all variants in SCN5A (including VUS).
- In vitro functional measurements can help predict SCN5A disease risk. Investigators measured the peak current (maximum current in channel) and the late current; both measurements are important in conventional wisdom.
- VUMC is currently conducting high throughput patch clamp of 69 SCN5A variants to improve classification. The deep mutational scan separates mutants by functional class and identifies new loss and gain of function mutants.
- Conclusions are that SCN5A variants have incomplete penetrance and a lower penetrance than one might estimate from the literature or from ClinVar. The unselected cohorts are needed to better estimate true penetrance of these variants. Functional data can help aid in this classification.
- Regarding if there is a bias in the VUS data, approximately 5% of people have a VUS in the SCN5A gene, and the rate of disease is much lower. Most VUS will likely not be disease causing.
- Regarding the random sampling that has been conducted thus far; this study references the wild type and effect on channel function needs to be corrected for.

- Andrew developed an assay to determine if and how the drug sensitivity related to pathogenicity. There were five known variants of different degrees on the allelic spectrum and is working on an assay that distinguished these from wild type.

**Science presentation: Impact of a population genomic screening and counseling program on performance and recommended disease risk management** | Marc Williams *(Geisinger)*

- The MyCode Community Health Initiative was launched in 2007 with a sequencing goal of over 250,000. To date, over 220,000 (~70%) have consented in person and online through MyGeisinger, and over 92,000 exomes have been sequenced.
- The MyCode genomic screening and counseling program involves uploading a report into EHR to initiate the clinical program to facilitate genome-informed care for patient and family.
    - Eligible MyCode samples are sent for exome sequencing and those exome sequences undergo bioinformatic analysis of Geisinger genes. If there is a reportable result, variant confirmation is conducted in CAP or CLIA-certified clinical laboratories and a report is issued to Geisinger. If there is not a reportable result, exome sequences are saved for future bioinformatic analysis.
- To estimate the clinical utility of the genomic screening program, the genetic counselors and an OBGYN resident conducted a chart review among MyCode patients with CDC Tier 1 genomic result (Hereditary breast and ovarian cancer, Lynch syndrome, and Familial Hypercholesterolemia) and identified the primary outcomes: percentage not previously aware of variant via clinical GT, percentage who performed recommended disease risk management post-reporting of genomic result, and percentage with new diagnosis of relevant disease post-reporting of genomic result.
- In summation, the majority of patients enrolled in the MyCode program had not received a result of a new diagnosis of a relevant disease (89%) nor a previous genetic diagnosis (87%). Less than half (42%) of the participants had a personal history of relevant disease, whereas more than half of the participants (62% and 85%, respectively) were both risk management performed and eligible for risk management in the program.
- This supports the effectiveness of genomic screening programs in identifying previously undetected individuals at risk for preventable cancers and heart disease, and the ascertainment of genomic risk as it leads to relevant disease diagnoses.
- A focus for future studies on this area could be to explore the long-term assessment of adherence to recommended risk management, assessment of intermediate health outcomes, risk management performance and new disease diagnosis in HFE C282Y homozygotes, the uptake of cascade testing among first-degree relatives, and identify any biases regarding who consents for ROR.
- Considering the study location, there is a low Amish population, and the group is thinking about exploring it further in addition to examining the level of the Mennonite population.
- Future work includes exploring CDC tier 1 conditions and rare conditions (Cardiomyopathy, Long QT), re-phenotyping. Focusing on discovery research, most of which will transition to clinical care. Continuing development in this area as the research is investigator-driven and focused on the interests of the team. Research is also affected by funding, as well as the priorities of the system.
- Geisinger's IRB does not allow investigators to conduct randomized trials, but Geisinger will look at other trial designs.
- This work contributes to clinical exome sequencing program, and exomes are sequenced for non-indication individuals who are interested.

- Economic data assumptions can be formed from robust clinically relevant information on economic performance. The team hopes to prove data meets effectiveness threshold.

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

- The eMERGE project NT302 *eMERGE Harvard Site Top 6 Genes* has been put on hold due to the scope of genetic analysis. The lead author, Samira Asgari *(Harvard)* plans to conduct a PheWAS analysis across the site's top six genes. Marc Williams *(Geisinger)* raised the point that it would be effective to analyze all eMERGE genes (from interested sites) at one time.

## DAY 2: Friday

**eMERGE Network overview: Priorities and goals, review progress of prior ESP recommendations and best practices topics** | Rex Chisholm *(SC Chair, Northwestern)*

- Specific aims of eMERGE III include sequencing and assessing clinically relevant genes presumed to affect gene function in ~25,000 individuals, assessing the phenotypic implications of these variants, integrate genetic variants into EMRs for clinical care, and creating community resources
- Genetic and phenotypic data is currently on over 148,000 participants and informatics tools which harnesses the data (SPHINX, CDS_KB, PheKB, and the PheWAS catalog)
- There have been >1200 external dbGaP downloads as of September 2018.
- There have been/are numerous Network collaborations. This includes the FDA, which is ongoing, where the goal is to examine hip arthroplasty data in the eMERGE GWAS set and related outcomes controlling for race and sex. The PAGE consortium is an approved non-clinical affiliate and is ongoing.
- ESP Recommendation #1: The ROR workgroup should publish either a manuscript or a report to the Genomics community to create awareness of the initial ROR findings.
    - The ROR workgroup has many projects underway as well as recently published work on initial ROR findings. The manuscripts and concept sheets are outlined below. The group will continue to incorporate findings and lessons learned into manuscripts as eMERGE moves into year five.
- ESP Recommendation: The Network should explore collaborations with the UDN on topics related to variant classification and VUS.
    - Jyoti Dayal has reached out to the UDN sequencing core chair, Christine Eng, MD, at Baylor, to present approaches to variant interpretation used at the UDN.
- ESP Recommendation #3: The Network should address some of the reasons for the delay between receiving clinical reports from sequencing centers and returning these results to the patients.
    - There are several elements that contribute to the time elapsed between the return of clinical results to the site and the return to the participants. Some are needed to ensure proper return and adherence to IRB guidelines, others can be the focus of improvements in the future. The CORRELSI workgroup is planning a manuscript outlining reasons for delays in the return of results to the participants after sites receive the results, which will help investigators anticipate issues that need to be addressed.
- ESP Recommendation #4: The investigators should create an outline of their plans for the next 1.5 years of eMERGE and the goals they wish to achieve by the end of Year 5.
    - Over the last six months, the CC has worked with each workgroup to develop milestones and timelines associated with the goals for the next phase of eMERGE.

- The ESP requested to learn more about the FDA and the funding involved in the collaborations as part of the Year 5 Milestones. The FDA wanted to utilize eMERGE GWAS data and examine genetic risk factors affected surgery outcomes and provided a small amount of supplemental funding to the CC to pull these data and continues to collaborate with the Network.
- The ESP recommends sharing the delays and the strategies for addressing these delays. In addition, the group should categorize the site-specific delays. For example, several site-specific delays may be symptoms of a larger problem.
- eMERGE is creating a 'defacto' standard of how to integrate data into the EHR, by focusing on creating implementation guides.
- The Network is developing a system-wide analysis of how sites have approached the management of deceased participants.
- eMERGE has previously approached disseminating lessons via the white paper method. The ESP suggested another avenue for disseminating lessons learned was to post videos and tutorials on YouTube. This can also educate viewers on how to overcome any barriers encountered. This was supported by the NHGRI.
- John Connolly (*CHOP*) developed MyResults.org in eMERGE Phase II and launched it utilizing CHOPs biomedia team. They are in touch with the NHGRI education team about promoting eMERGE and DNA day (Spring 2019)
- Maureen Smith (*Northwestern*) has also worked on two separate four-minute videos to provide additional information to participants with negative results.

**Lessons learned for harmonizing sequencing centers for Network-wide products and initiatives** | Richard Gibbs *(BCM/HGSC)* & Heidi Rehm *(Partners/Broad)*

- The paper "*Harmonizing Clinical Sequencing and Interpretation for the eMERGE III Network*" has been submitted to the *American Journal of Human Genetics (AJHG)* and will also be posted on bioRxiv.
- Within the manuscript, the CSGs listed challenges encountered, broad categories to allocate these challenges to specific points in the harmonization/sequencing/interpretation process, as well as additional comments.
- Once harmonization issues were worked through, the rate of sequencing drastically increased.
- In the beginning of eIII, sites submitted nominations for top six genes. CSGs leveraged the ClinGen framework to assess validity of these genes. Clinical reporting thresholds were assigned to non-ACMG56 genes.
- Areas for further harmonization include the need for harmonizing initial variants, as well as to develop harmonized structured genetic test report standards compliant with FHIR/HL7.
- There is caution that low discordance between the CSGs regarding variant interpretation is sometimes attributed to the general lack of more-than-one interpretation available since the field is still in the discovery phase.
- Discordant variants are shared with ClinVar although not on a specific timeline. The ability to 'update' variants in real time on ClinVar would allow for more rapid communication of changes.
- CSGs and sites pipeline for variant reassessment as it pertains to ROR to the patients, specifically what is the appropriate timeline of the sequence of events and method of re-notification is a topic they are actively exploring and seeking to develop processes for. However updates into the EHR is still an avenue that needs to be developed with a more efficient pipeline.

    - The alerts are not burdensome; however, the issue is within tracking of the versioned reports within the EHR. Baylor has developed a blockchain based system that assigns a barcode to the reports which helps with identifying the most recent/accurate clinical report.
- Harmonization happens before reporting, both between the CSGs and with the sites to a limited extent. It is quite rare to have discordance. Most changes arrive via new knowledge reported within the field itself, and eMERGE CSGs seeking to reassess based on this external factor. However, because the new knowledge can come out at any point, this lends itself for the potential to new harmonization and interpretation.
    - Variant reinterpretation and the impact of this on patient ROR is a real-world problem clinicians experience. However, it is the clinician's responsibility to openly communicate all information to patients on the forefront prior to the event of a reinterpretation.
- After the information is being exported and interpreted from PB, it is being sent to Baylor.
- Novel variants are classified by both P/B and Baylor for any variants coming off the eMERGEseq platform.

**Clinical Annotation: Workgroup progress & future directions** | Gail Jarvik *(KPW/UW)* & Heidi Rehm *(Partners/Broad)*

- Penetrance goals include collaboration to evaluate penetrance of P/LP variants in gene-disease pairs in participants that are not ascertained for a relevant phenotype. This will be done by aligning chart review for penetrance with Outcomes using the same forms. Individual sites are responsible for chart review abstraction form. The timeline will be from January 2019-December 2019.
- Actionable incidental findings in eMERGE III (3.12% positive, n= 13, 979); 2.55% ACMG (n= 356); 0.56% non-ACMG (n=80). An oral presentation on this topic, Frequency and features of incidental findings across 12,702 eMERGE network participants, was given at ASHG 2018 by Adam Gordon (Northwestern).
- Workgroup Project: eMERGE III Variant Reanalysis using new eMERGE data addresses the notion of reanalysis and reinterpretation.
    - Reclassifications of 1.2% (n= 7/596) of variants are affecting 16 patients (2 LP to P, 1 VUS to LP, 3 LP to VUS, and 1 P to VUS)
    - For variants that are not seen again by Baylor or P/B LMM or involved in ClinGen discrepancy resolution efforts, updates may not be prompted.
- Opportunity for eMERGE in determining best practices for delivering and managing updated genetic testing results:
    - The "*ACMG clinical laboratory standards for next-generation sequencing*" published in *Genetics in Medicine* article (Rehm et al, 2013) focuses on updating reports unsustainable without payment, labs need to document policies, changes, and encouraged to develop innovative approaches; and suggest physicians to inquire about updates for VUS and LP and labs amend reports for pathogenic or benign changes.
- Examples of laboratory management of knowledge updates include issue amended reports, allowing direct access to laboratory database (e.g. Emory), regularly depositing variants into ClinVar (recommended by the draft ACMG guideline), and delivering automated knowledge updates on reported variants to physician (e.g. LMM/GeneInsight).
- Patient registry management of updates includes GenomeConnect (ClinGen patient registry), which alerts patients if a lab's interpretation in ClinVar has been updated.
- Workgroup Project: eMERGE III variant reanalysis of suspicious VUSs using eMERGE EHR data (VUSs that favor/lean towards LP).

    - Members of expert panels have contributed unpublished data for gene and variant curation. The internal data impacted rule application or strength for 25 variants. Six sources contributed to over 420 probands and over 75 segregations overall. 12 of the 25 (48%) changed classification.
- Future Efforts:
    - Collaborate to evaluate penetrance of P/LP variants in gene-disease pairs in participants not ascertained for a relevant phenotypes and gather additional clinical data to inform the pathogenicity classification of reported (ex. pathogenicity) and non-reported (ex. VUSs) variants.
    - Define types of variants that are worth further investigation by eMERGE sites using EHR data. This is done a few ways: define variants with a higher probability of shifting from VUS to LP/P, define phenotypes associated with eMERGEseq panel genes, and define availability of follow-up functional assays.
    - Mayo has conducted a pilot study of giving polygenic risk scores in the MyGenome study. They are considering expanding that pilot in order to incorporate an omnigenomic model that uses millions of SNPs.
- The ESP suggested that eMERGE should consider returning their knowledge gained to ACMG/AMP/CAP.
    - There have been efforts between the three groups (ACMG/AMP/CAP). ACMG had public sessions and sent out drafts prior to release, along with posting for member comments. However AMP's workgroup efforts are entirely closed until their project is published. This is a bureaucratic barrier to entry that must be overcome should this relationship come to fruition.
- ClinGen is forming a committee to begin investigating of polygenic risk scores. eMERGE members are strongly encouraged to consider joining.
- ACTION ITEM: eMERGE investigators interested in joining ClinGen's polygenic risk score investigation committee should contact Heidi Rehm (HREHM@mgh.harvard.edu).

**Outcomes: Workgroup progress & future directions** | Hakon Hakonarson *(CHOP)*, Josh Peterson *(VUMC/CC)*, & Marc Williams *(Geisinger)*

- There are 19 phenotype-outcome linked instruments on REDCap, and currently there are 133 records with six-month follow-up: Geisinger (n= 40), KPW/UW (n= 30), and NU (n= 53).
- Additional changes or adaptations consist of adding site-specific reporting tools within REDCap, standardizing editing rights with new centralized approach and the amount of data / type of data acquired on deceased patients, assessing the version history of changes and log of changes, and the training version of the Outcomes data collection tool.
    - ACTION ITEM: Geisinger will send the CC a list of any issues that require REDCap coding changes for the Outcomes forms.
    - ACTION ITEM: Northwestern will review the REDCap form designed to capture male breast cancer susceptibility and provide feedback.
- Migrating all the sites onto the CC hosted REDCap will facilitate form updates and changes going forward.
    - ACTION ITEM: The CC will initiate next steps regarding the migration of existing site Outcomes data into the centralized CC REDCap instance.
- The Clinical Annotations and Outcomes workgroup are collaborating to use the Outcomes forms to estimate penetrance. These forms are designed for prospective data capture with an emphasis of impact of intervention. The penetrance calculation will require age of onset of genomic medicine diagnoses found in retrospective data.

    - ACTION ITEM: Outcomes forms will be distributed via Google Docs by the CC in order to facilitate editing of Workgroup documents. (Due: 11/16/2018)
- Year 5 Plan include estimating the penetrance of genes and gene sets by leveraging Outcomes data.
    - The goal is to complete a 6-month Outcomes assessment, with an interim analysis by June 2019 SC meeting, and anticipated completion by Summer 2019.
    - Outcomes Workgroup intends to prepare and submit manuscript by Winter 2020. Publications will be split by sets of phenotypes.
- The Clinical Annotation group assigned timeframes to specific Outcomes forms and any changes should be documented in the next two weeks by contacting the CC. Each form will have an abstraction guide that is in progress, when ambiguity is discovered, the guides will be amended. It also allows sites to add notes and definitions for specific data elements.
- The group should plan to conduct a quantitative analysis of the findings, and impact of negative results should also be assessed if possible.
    - The participant survey is going to examine negative and positive result participants.
    - Mayo has conducted 87 open-ended interviews and 1,500 surveys with participants with negative results to determine how well participants understand the limitations and plans. CCHMC has conducted interviews asking participants what results they want to receive. HCP surveys has been sent to physicians about negative results.
- If there is anything at the chart review level that is helpful to understanding the negative result the data should be included.

**Phenotyping & OMOP: Workgroup progress & future directions** – George Hripcsak *(Columbia)* & Peggy Peissig *(Marshfield)*

- The Phenotyping Workgroup has completed 13 electronic phenotype algorithm implementations across the Network. There are two still in development at the primary site, and six in secondary validation.
- The Phenotyping Common Variables subgroup has defined a list of variables collected by all sites across all datasets. This is housed at the CC.
- The OMOP subgroup populated data in OMOP CDM v5.2 and converted terminology for drugs and diagnoses, labs are in process.
    - Half of the participating eMERGE sites successfully installed the full OHDSI stack; however, hurdles noted include the substantial burden of verifying security requirements. A major challenge encountered by the OMOP group has been converting complex phenotype definitions and covariate list to each institution's local model.
    - There has been a quick turn-around in testing OMOP phenotypes. Testing has revealed bugs in new and original algorithms, highlighting differences in interpretation between sites. Two phenotypes have been tested: Type 2 Diabetes and ADHD.
    - Type 2 Diabetes resulted in a large drop in implementation time (median 3-5 hours), encouraged cross-site debugging in real time, and the ability to identify bugs during the OMOP data conversion. Several bugs were identified with the OMOP data conversion during implementation. Additionally as most sites stored lab codes via strings instead of in a standardized coded fashion such as LOINC, it made it difficult to track down and convert labs.
- The major Year 5 goal is to develop a natural language processing (NLP) component for a maximum of five high-priority phenotypes.

    - cTAKES & Metamap pipelines were chosen for implementation, some in combination with Regex. All sites will use OMOP to store NLP results, with help from OMOP group. A timeline for implementation has been established.
    - A third site will be used for validation of NLP algorithms, in addition to original primary and secondary sites. This will be based on volunteers to validate each additional phenotype.
    - Next steps include deciding the order in which to do phenotypes, and establishing metrics for evaluating portability and the complexity of NLP phenotyping algorithms, in a post-hoc evaluation.
    - NLP Phenotypes include:
        - ACO - asthma COPD (Harvard/KPW): capture PFT results to define asthma and COPD.
        - Long QT/arrhythmia (VUMC/Mayo): capture additional information for phenotyping and analyses
        - Chronic rhinosinusitis (Geisinger/NU): needed to improve algorithm PPV.
        - Familial hypercholesterolemia (Mayo/Geisinger): needed to improve algorithm PPV.
        - Systemic lupus erythematosus (NU/VUMC): needed to power lupus subtype analyses.
- The ESP noted that LOINC has developed a guidance document of how to weight LOINC codes, mappings, and categorizations. The file is called 'Groups.'

**Panel: Lessons learned summary - ROR/ELSI, Phenotyping, EHRI** | Ingrid Holm *(BCH)*, Iftikhar Kullo *(Mayo)*, Peggy Peissig *(Marshfield)*, George Hripcsak *(Columbia)*, Casey Overby *(Geisinger/JHU)*, & Sandy Aronson *(Harvard)*

- **ROR/ELSI**
    - There are a variety of processes for return at each site which allows for experiments of nature to inform difficulties and resolutions. Study design differences include randomization, cohort, genotype, age of study population, choice of return on secondary findings, whether to return negative results, and the variation in timing of placement of results in EHR.
    - ROR process also varies by site. These process differences include the ability to assessing the impact of ROR on healthcare providers across sites, which is a challenging as sites have different processes for who returns the results and to whom results are sent. In addition, coordinating the participant survey across sites in challenging, given the different populations at each site and the different site priorities regarding their focus of research. IRBs vary significantly in their requirements, processes, and views towards ROR from genomic sequencing.
    - The ROR process is also dependent on institutional cultures and priorities and thus is difficult to create standard guideline and practice.
    - Unexpected delays in ROR may include difficulties in contacting participants and in integrating the results in to the EHR prior to the ROR.
    - Site-specific delays, including participants who are lost to follow-up, were enrolled as children and need to be re-consented as adults, difficulties with IRB approvals, and cultural competency in order to connect with participants also add to increased turn-around times.
    - Differences in the ROR process across sites allow for overarching analysis methods to affect patient comprehension, engagement, and outcomes and natural experiment to study the differences and how they impact outcomes.
- **Phenotyping**
    - The Phenotyping process consists of knowledge engineering, data extraction, analysis, sample, review & classification, and adjusting precision.

- The Phenotyping methods used included billing codes (structured data queries), clinical notes (natural language processing or NLP), lab & test results, structured and semi-structured data queries (medications/ePrescribing and structured data queries.).
- In order to diversify the source of information within Phenotyping, the Workgroup had the goal to facilitate the adoption of a common data model (OMOP), which converts data warehouses to the same schema/vocabulary while preserving source information. A main challenge to OMOP is that there is a learning curve and ramp-up time associated with the OMOP work.
- NLP is also necessary for phenotyping accurately, and the group is focusing on this in Year 5.
- Secondary review of the phenotype is a huge learning point in development of accurate algorithms. It helps the site learn about their own EHR, and validation determines the universality of the algorithm. Even though only one site actually conducts secondary validation, multiple sites implement the algorithm and with each implementation changes and iterations are documented, making the algorithm itself more universal.
- Future planned approaches consist of having a better working knowledge regarding how to help future planning and a retrospective analysis of planned vs. actual complexity.
- There is a need to collaborate phenotype development and need for tools (ex. PheMAP) in addition to standards. The workgroup has a vision for sites working together on a phenotype with a quick turnaround. The concept of innovation is fundamental to the Phenotyping process, as the introduction of machine learning could be integrated.

- EHR Integration
  - As data standardization and harmonization are moving to HL7 standards, there is a need for tracking, dashboard, and analysis tools. This factors into managing of variant reclassification and reporting.
  - Tight collaboration between clinicians, laboratorians, and IT professionals are needed to embed clinical decisions support into clinical workflows accurately. Genetic Aware Clinical Decision Support is dependent on access to structured data and knowledge, which is highly fractured across the healthcare system and sites are now actively engaged with these issues.
  - For EHR Genomic data integration with genomic indicators in EPIC, there is a need to establish a vetting process for variants, as well as a need to define what to display with that variant, and to understand how physicians will react/use the variant.
  - The workflow of the project team should be considered in this implementation process, as many people are involved, there is often asynchronous communication, and it is 'hand-off.' There are also changes in personnel that are impacted by EHR transition.
  - The data flow, conversion from XML to HTML or extracting PDFs, as well as user interfaces (REDCap, NYP, eMERGE research tools), influence integration.
  - It is feasible to build a unified clinical Network linking heterogenous laboratories and provider system in the context of a NIH consortium. However, in order to do this, investigators must: establish core infrastructure, determine common denominator of tests, ingestion capabilities, format for knowledge and data transmissions, interface systems in ways they were not intended to be interfaced, establish report flow, establish variant interpretation flow, satisfy laboratory, site, and vendor legal requirements

    - Agreeing on a format and building the eMERGE Network required manpower and numerous steps. These steps include; interfacing clinical systems in ways they were never designed to be interfaced, establishing report flow, establishing variant interpretation knowledge flow, getting vendors to invest heavily in supporting this process, satisfying all laboratory, site and vendor privacy, security, and legal, validating transmission capability, operating the Network, developing mechanisms for enhancing structure, and responding to feedback.
  - The ESP believes there are exciting lessons learned within the EHRI workgroup, utilizing the WordCloud produced at the meeting would be a good reference point to begin categorizing challenges.
  - Two papers could be developed. The first could be for individual sites lessons learned and the second for informing how consortia can establish EHRI.
  - Vendor engagement is important, in this case DNAnexus and GeneInsight, but it would apply globally as well. There were discrepancies between what the commitment was when applying for the grant and what was ultimately needed. However, vendor engagement is a resource intensive process, similar to DNAnexus and GeneInsight.
  - Global Alliance is evolving in this direction but is perhaps naive in regard to return of results.

**Genomics: Workgroup progress, future directions & data set updates** | David Crosslin *(UW/CC)*, Megan Roy-Puckelwartz *(Northwestern)* & Patrick Sleiman *(CHOP)*

- The quantity of the genetic data as well as the excellent phenotyping really sets the Network apart. The Network has built a variety of tools to help utilize these resources.
- "*The eMERGE genotype set of 83,717 subjects imputed to 30 million variants genome wide and association with the herpes zoster medical record phenotype*" has been published in *Genetic Epidemiology* (Stanaway, 2018).
- Harvard has submitted an additional 15k samples which has been merged in to the eI-III Merged Imputed dataset, bringing the total to ~99,000. An additional 20,000 samples will be submitted over Year 5.
- Year 5 Milestone Activities includes structural variant calling and BEAGLE phasing and imputation
  - The group will impute common structural variation using the 1000 genomes. Initial runs are complete, QC is underway. There are multiple structural variation (SV) efforts ongoing in eMERGE.
  - CHOP conducted a survey of the sites to determine the availability and ability to access genotype array intensity values for CNV analysis.
- The eMERGEseq set was frozen at 15,000 samples to begin setting up the analysis and QC pipeline, as well as complete an initial submission to dbGaP.
- Genomics workgroup Year 5 aims are to develop methods to estimate penetrance. CHOP is seeking to assess penetrance through both a family-based approach and a population-based approach, and VUMC is developing Phenotype Risk Scores to gauge penetrance.
- The group plans to be reimagining the eMERGE SPHINX tool to eMERGENT (Electronic Medical Records and GENomics Toolkit), in order to include all eMERGE data within this tool and streamline utilization. eMERGE has richer phenotype data and diversity than the UK Biobank, however the UK Biobank has a good toolkit to utilize the data. A user interface to maximize usage both internally and externally is needed. There will be a user-centered approach to design of this tool, engaging with stakeholders and focus groups to determine requirements. Ultimately, a paper on lessons learned in regard to informatics tool building is anticipated based on eMERGE's experience.

- The PheWAS community tool utilizes PheKB allows researchers with published PheWAS studies to integrate their data into the PheWAS catalog itself. The tool is in the beta version, but it is accepting submissions.
- Currently, the individual investigators are pushing incorporation of findings that require mechanistic follow-up forward in the Network. However, there is not an overarching programmatic method. The UDN has a good model as they fund both the functional groups along with the sequencing groups.
- The Network needs a plan for exposing eMERGENT to the appropriate audiences. Social media, being nimble, cross consortia demonstration, presenting at scientific meeting are all options.

**PGx: Workgroup progress & future directions** | Cindy Prows *(CCHMC)* & Laura Rasmussen-Torvik *(Northwestern)*

- Workgroup accomplishments include publications, how the group has expanded the data collection for PGx, and the eIII result return update.
- The analysis of eIII PGx data includes ~25,000 genotyped on eMERGESeq which includes selected PGx SNPs, for which the return plans vary by site, VUMC has initiated returns.
- The eII PGx work is being expanded. Common phenotypic variables are now being collected for PGx participants at regular intervals. The comprehensive drug collection discussion has been delayed until full OMOP implementation, and there are inevitable limitations that will remain even after OMOP implementation is complete.
- Year tasks include CPIC subgroup participation and involvement with the CPIC API. There will be a call or with CPIC/eMERGE to discuss ideas for collaboration, areas of overlap, and feedback about CPIC guideline resources.
    - ACTION ITEM: If individuals are interested in joining the CPIC subgroup they should contact the PGx co-chairs Laura Rasmussen-Torvik and Cindy Prows
    - ACTION ITEM: An API representative will be invited to join on a future PGx call to discuss harmonization efforts.
- Other Year 5 tasks include discussing where to emphasize and focus collaboration, such as in the lessons learned for implementation across multiple sites.
- The PGx workgroup is considering collaborating with ROR/ELSI workgroup on a lessons learned paper or an individual manuscript on return of PGx results.
    - There is a challenge regarding there being very little bandwidth for novel PGx data collection (quantitative and qualitative) across sites. The most powerful products from the PGx incorporate data across sites. A possible solution regarding this issue is for connections to journals and editors interested in cross-site commentaries to support this novel PGx data collection.
    - The Workgroup is investigating the possibility of releasing a lessons learned deliverable in the form of a workshop or presentation rather than a manuscript.
- In the case of Thiopurine methyltransferase (TPMT), the translation table associates both *1/*3A and *3B/*3C with heterozygosity for rs1142345 and rs1800460. The phenotypes are not the same. For example:
    - *1/*3A is an intermediate metabolizer, which translates to a a risk of overdosing and toxicity if it really is a poor metabolizer.
    - *3B/*3C is a poor metabolizer, which translates to a risk of ineffectual dosing for serious illness if it really is an intermediate metabolizer

- Marc Williams is on the ESP for CPIC. There will be a larger in-person meeting in June 2019 and they seek participation for the planning committee. Potentially an eMERGE representative could participate and have an eMERGE presence at the meeting itself.
- The SNP list could also be discussed by CPIC. As many as 320 people from around the world have joined CPIC and that resource would be good to utilize when thinking through the PGx questions.
- ROR lessons learned manuscript will include the return of PGx results, comparing to standard return of result methods, including data supporting those claims.
- eMERGE II PGx was returned at most sites but eMERGE III is only being returned at only a handful. The Network should examine any differences and determine how PGx on a genomics report with and without CDS affects clinical care.
- There was a suggestion for the PGx workgroup to create a PGx outcomes form.

**ROR/ELSI: Workgroup progress & future directions** | Ingrid Holm *(BCH)* & Iftikhar Kullo *(Mayo)*

- Site progress in returning of results to participants, both positive and negative, is progressing.
- The Year 5 Milestone #2 is to determine the impact of return of genetic results (ROR) on patients' immediate outcomes, six months and/or 12 months after ROR, which includes identifying outcomes related to processes of care, clinical utility, family utility, provider utility, and patients' psychosocial factors.
    - This milestone is being partially addressed via the Health Care Provider Survey (HCP), an R01-funded three-year initiative. Surveys and interviews of HCPs are carried out through an across-site R01 to assess the impact of ROR (positive and negative) on HCP across sites. This work led to the creation of surveys as well as the development of cognitive interviews that were piloted and conducted.
    - Participant surveys will be used to assess the impact of ROR (positive and negative) on participants. In addition, a participant survey subgroup is established to discuss the aims of this initiative.
    - A data collection tools will be developed that can be implemented across sites.
    - The ROR workgroup will collaborate with the Outcomes workgroup to develop lessons learned publications based on sites returning results in different ways.
- The group assess if and how participants with positive results inform their family members and informing them of their potential risk as well as collect the responses of the family members in Year 5.
    - Family implications of ROR on family members will be evaluated.
    - In addition, the workgroup will examine HIPPA issues around relatives and deceased patients. Members of the ROR-ELSI group are working with the Genomics and Population Health Action Collaborative (GPHAC) Cascade Screening working group, to address these HIPAA issues.
- During Year 5 the group will also estimate the institutional impact of ROR.
    - The group plans to develop guidelines for IRB language and procedures for consenting patients to undergo sequencing and to receive results, and for institutions returning results to patients.
    - Return of results processes and outcomes will be described at each site. Georgia Weisner (VUMC) and Kathy Leppig (*KPW/UW*) will co-leading this effort.
    - Institutional impact data will be collected, and interviews of investigators will be conducted at all sites. There is an MCSN (T273) that describe plans for ROR at all eMERGE sites prior to the release of results from the sequencing laboratories. Once results have been returned, the final processes

and outcomes at each site will be described, and the the impact of the ROR processes on participants across the sites will be compared.

- Goals include developing guidelines for institutions that are returning genomic results to patients. The return of results processes and outcomes will be described at each site. George Wiesner (*VUMC*) and Kathy Leppig (*KPW/UW*) co-lead this.
- Ongoing projects on ROR/ELSI include participant surveys, R01, NT273, NT277, NT300, and the CASH-study.
- The ESP noted it would be detrimental if there was not enough time to assess the impact of return of reclassified variant results to patients.

**EHR Integration: Workgroup progress & future direction** | Sandy Aronson (*Harvard*) & Casey Overby Taylor *(Geisinger/JHU)*

- The group surveyed the sites to determine the requirements and needs for EHRI. Many of these sites conduct pre-and-post testing for EHRI and CDS.
- Findings of this survey demonstrated the need for a reporting standard, in regard to structure, content, vocabulary, and ontologies, as well as easier integration and interoperability and communication. GI XML format was identified.
- Research projects are underway to target CDS and EHR integration across the sites.
- There was a panel at the AMIA 2018 informatics summit which led to media coverage of the mentioned developed tools.
- Ongoing collaborations are with IGNITE Clinical Informatics Working Group (NT301), the eMERGE ROR WG, which has a survey on preferences for research updates, and with ClinGen EHR Workgroup on evaluation of genomic infobutton.
- Research projects underway with approved concept sheets: NT265, NT270, NT272, NT294, NT301, NT303.
    - ACTION ITEM: Those with EHRI publication updates should email these directly to Casey Overby Taylor <cot@jhu.edu> and Sandy Aronson <SARONSON@PARTNERS.ORG>
- Future collaborations are in progress for the EHRI workgroup:
    - The EHRI workgroup intends to brainstorm with PGx working group to collaborate and to consolidate lessons from CDS implementation, which is led by Bob Freimuth.
    - There is a MeTree Family History collection tool, which is led by Andrea Ramirez. The group will take volunteers to review/beta test implementation guide (March-May 2019 timeline).
    - Knowledge sharing activities will be implemented by the Workgroup.
    - Development of FHIR Compliant File Structure Usable in eMERGE IV to replace the current .xml file an initiative.
        - The ultimate output will be an implementation guide. The sequencing centers are working together on the specification and will engage the workgroup for feedback. Specification will be available for eMERGE IV and others supporting these types of transmissions.
- Currently no major EHR vendors support inbound messages, specifically genetics at this time. The EHR itself may not be the ideal place to manage genomic information, there may need to be an ancillary system.
- Baylor is going to regenerate the reports in the new FHIR standard, however the sites are not obligated to implement them during eMERGE III. GeneInsight has not committed to regenerating reports with FHIR standards.

- Ancillary systems to manage genetic data should be considered. The ESP mentioned that this may not be a practical long-term solution across the Network and that the EHRI systems are like “working economies”.
- Sandy believes that things become more scalable by operating under distributed scalable solutions that feature optimized interoperability. In this way, all sites are not at the mercy of a single management system under one company.
    - ACTION ITEM: EHRI Workgroup should describe and codify the standards that are required to be present during integration.

**Network discussion: Input/Feedback from the ESP** | *ESP Members & NHGRI Team*

- Longitudinal follow up regarding the impact of reclassified variant result return is important and the Network should try to incorporate time for understanding the impact on participants, perhaps by obtaining external funding.
- By building a career through this Network, junior investigators learn what questions to struggle with and how to build innovative strategies to address these. The Network continues the tradition of helping junior investigators in their careers through eMERGE.
- Efforts that have been made to collect data across sites consistently on a variety of the ‘softer’ elements such as ROR/ELSI outcomes surveys.
- The Network has generated useful lessons learned to apply to other networks including All of Us Research Program with respect to sequencing.
- Group has been active working with outside stakeholders (FDA, CPIC, IRB), implement locally but disseminate globally and this is essential to not just the Network but also the field.
- The ESP is pleased with the efforts made regarding gathering lessons learned.
- The ESP appreciate that the recommendations were incorporated and addressed so quickly.

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

- It is beneficial to the consortium to collect its progress and present to the ESP. The Network really appreciates the feedback and recommendations.

**ACTION ITEMS:**

**CC**

- The CC will produce an infographic for the phenotypes that have been generated on the genetic data

**NHGRI**

- eMERGE NHGRI Program Staff, in coordination with the CC, will invite AnVIL team to present to the Network at an upcoming SC meeting or PI meeting.

**PI/Network**

- eMERGE investigators interested in joining ClinGen’s polygenic risk score investigation committee should contact Heidi Rehm (HREHM@ mgh.harvard.edu).

**EHRI**

- Those with EHRI publication updates should email these directly to Casey Overby Taylor <cot@ jhu.edu> and Sandy Aronson <SARONSON@ PARTNERS.ORG>

- EHRI Workgroup should describe and codify the standards that are required to be present during integration.

PGx

- If individuals are interested in joining the CPIC subgroup they should contact the PGx-co-chairs Laura Rasmussen-Torvik and Cindy Prows.
- An API representative will be invited to join on a future PGx call to discuss harmonization efforts.

Outcomes

- Geisinger will send the CC a list of any issues that require REDCap coding changes for the Outcomes forms.
- Northwestern will review the REDCap form designed to capture male breast cancer susceptibility and provide feedback.
- The CC will initiate next steps regarding the migration of existing site Outcomes data into the centralized CC REDCap instance.
- Outcomes forms will be distributed via Google Docs by the CC in order to facilitate editing of Workgroup documents. (Due: 11/16/2018)

**ESP EXECUTIVE SESSION NOTES & RECOMMENDATIONS:**

- The ESP met with members of the NHGRI Program Staff in an Executive Session before and after the October 26, 2018 ESP meeting. Since some ESP members did not attend the first day of the SC meeting held on October 25, 2018, Rongling provided a summary of the Day 1 activities.
- Overall, the ESP was impressed with eMERGE's progress and appreciated their efficient and timely responses to their previous recommendations. The ESP was also pleased with the EHR integration (EHRI) and Phenotyping workgroups' engagement with external groups, such as Health Level Seven International (HL7) and Observational Health Data Sciences and Informatics (OHDSI) collaborative, to improve Phenotyping and EHR integration. They found the lessons learned panel extremely useful. They complimented the Network for its continued ability to collaborate and integrate their work across the multiple sites.
- A key recommendation the ESP had regarding the lessons learned panel was that the Network should consider publishing these findings in a single special issue journal. A single issue will be more convenient for the community to access, compared to publishing the lessons learned papers in multiple journals or in the same journal at different times.
- It was noted that eMERGE has around 60 in-process manuscripts that need to be published with the understanding that most of the delay was due to the unavailability of sequence results. The ESP recommended that the Network should prioritize publishing papers that are already in the pipeline, and that the Network-wide papers should be submitted for publication before the end of phase 3.
- The ESP appreciated the CSGs efforts to develop an automated pipeline for the re-evaluation of reported variants, and the Network's focus on variant re-classification. They agreed that this is an important topic and the findings will be useful for the genomic medicine community. They encouraged the ROR workgroup to study physicians' responses to reclassification of variants. They also recommended that the EHRI workgroup investigate the effectiveness of clinical decision support (CDS) in relaying this information to patients and health care providers.

- The ESP understood that the findings from the ROR and Outcomes workgroups are not yet available due to delays in completing sequencing and generating clinical reports. Once ROR is complete, eMERGE will have 4 years' worth of data that will need to be analyzed. The ESP encouraged the sites to apply for additional, investigator-initiated funding to ensure that they have resources beyond the 1.5 year eMERGE extension to analyze and publish results.
- The ESP emphasized that the main goal of the next 1.5 years should be focused on understanding ROR and the effects of potential misconceptions, such as participants resuming risky behavior after learning of their low genetic risk. The ESP recommended that the ROR and Outcomes workgroups should work together to identify patients' misunderstandings about genomic results, and offer solutions that can address these issues. There are several studies that report on the challenges and ELSI issues associated with ROR, but there are a limited number of papers that offer solutions or strategies to overcome these difficulties.
- The ESP was concerned about the lack of consistency across the Network's outcomes forms. They agreed with the Outcomes workgroup's effort to centralize the surveys. The ESP recommended that this occur as soon as possible to help ensure a consistent dataset across all the sites. They also noted that individual sites had concerns with various aspects of the outcomes forms, such as the inability to add comments to entries. Overall, the ESP recommended creating a process that enables the ability to quickly adopt sites' changes. This should be done soon because sites are starting to collect ROR data.
- There was some concern that eMERGE tools are not being adequately disseminated or publicized. For example, the Sequence and Phenotype Integration Exchange (SPHINX) is an extremely useful tool, but it is not well-known outside of the eMERGE community. The ESP recommended linking SPHINX and other eMERGE resources to more well-known databases, such as the Pharmacogenomics Knowledgebase (PharmGKB). In general, the ESP recommended that the eMERGE website should be more user-friendly and include a toolbox that is easily accessible, similar to IGNITE's SPARK toolbox.
- The ESP emphasized that eMERGE should continue to investigate ways to collaborate with the *All of Us* (AOU) program. For example, the eMERGE Network can educate AOU on the challenges of returning negative sequencing and PGx results directly to participants without genetic counseling.
- There was also a concern about the 15-50% response rate to the ROR outcome surveys. The ESP recommended that the sites develop a plan to optimize participant response to these surveys, through EHR reminders, medical appointments, etc., as this information will be important for the scientific and clinical communities.

<u>There are nine recommendations from the ESP given to investigators:</u>

1. The Network should publish all their lessons learned as a special issue in a relevant journal.
2. The Network should prioritize publishing papers that are already in the pipeline with the priority to the Network-wide papers. The goal for publishing the network-wide paper(s) should be before the end of phase three.
3. The ROR workgroup should study physicians' responses to variant reclassification. The EHRI workgroup should continue to investigate the effectiveness of clinical decision support (CDS) in relaying this information to patients and health care providers.
4. The sites are encouraged to apply for additional, investigator-initiated sources of funding to help complete outcomes studies beyond the the eMERGE extension period.

5. The ROR and Outcomes workgroups should focus on understanding the misconceptions surrounding the return of genomic results on patient care, and identify sustainable solutions or strategies that can address these misconceptions.
6. The Outcomes workgroup should focus on being more consistent in all the Outcomes forms and resolving inconsistencies before sites start collecting more outcomes data.
7. The eMERGE website should be more user-friendly and include a toolbox that is easily accessible, similar to IGNITE's SPARK toolbox. To help disseminate eMERGE's tools, ESP recommended linking SPHINX and other eMERGE resources to more well-known databases, such as the Pharmacogenomics Knowledgebase (PharmGKB).
8. The Network should help AOU understand the impact of returning negative sequencing and PGx results directly to participants without genetic counseling.
9. The sites should have a plan to optimize the response rate of patient surveys through EHR reminders, medical appointments, etc., to improve the quality of the results. The results of this study will be important for the scientific and clinical communities.