emerge network
ELECTRONIC MEDICAL RECORDS & GENOMICS

Summary of Steering Committee Meeting: Summer 2018

June 25-26th, 2018 in Cincinnati, OH

TABLE OF CONTENTS

Day 1: Monday, June 25th, 2018

- **NHGRI program official report |** Rongling Li ***(NIH/NHGRI)***
- **Announcements, opening remarks |** Rex Chisholm ***(SC Chair, Northwestern)***
- **Genomic data update |** David Crosslin ***(CC/UW)***
- **Clinical sequencing centers update |** Richard Gibbs *(BCM)*; Heidi Rehm ***(Partners/Broad)***
- **Science presentation: The adverse drug effect recognize (ADER): A decision support tool for recognizing potential adverse drug reactions |** Joshua Smith *(VUMC)*
- **Panel: Phenotyping best practices |** George Hripcsak *(Columbia)*; Peggy Peissig *(Marshfield)*; Luke Rasmussen *(Northwestern)*
- **Network discussion: Process for sharing phenotype data**
- **Science presentation: Transcription factors operate across disease loci, with EBNA2 implicated in autoimmunity |** John Harley *(CCHMC)*
- **Outcomes processes summary |** James Ralston *(KPW)*; Alanna Rahm *(Geisinger)*
- **Science presentation: An update on the return of actionable variants empiricals (RAVE) study at Mayo: Challenges identified & lessons learned |** David Kochran *(Mayo)*
- **Panel: Findings from re-review of medical records |** Margaret Harr *(CHOP)*; Hila Milo Rasouly *(Columbia)*

Day 2: Friday, January 26th, 2018

- **Genomics workgroup progress report |** Megan Roy-Puckelwartz *(Northwestern)*; Patrick Sleiman *(CHOP)*; David Crosslin *(CC/UW)*
- **Clinical annotation workgroup progress report |** Gail Jarvik *(KPW/UW)*; Heidi Rehm *(Partners/Broad)*
- **Phenotyping & OMOP workgroup progress report |** Peggy Peissig *(Marshfield)*; George Hripcsak *(Columbia)*
- **EHR integration workgroup progress report |** Sandy Aronson *(Harvard)*; Luke Rasmussen *(Northwestern)*
- **PGx workgroup progress report |** Laura Rasmussen-Torvik *(Northwestern)*; Cindy Prows *(CCHMC)*
- **ROR/ELSI workgroup progress report |** Ingrid Holm *(BCH)*; Iftikhar Kullo *(Mayo)*
- **Outcomes workgroup progress report |** Josh Peterson *(VUMC/CC)*; Hakon Hakonarson *(CHOP)*; Marc Williams *(Geisinger)*
- **Network discussion: Network goals & milestones for year 4 |** NHGRI eMERGE Team

Meeting Action Items

Presentation slides (login required)

DAY 1: Monday

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

- James Battey retired as Director, National Institute on Deafness and Communication Disorders (NIDCD), the Acting Director is currently Judith Cooper. Richard Nakamura retired as Director, NIH Center for Scientific Review, the Acting Director is Noni Byrnes.
- NIH Inclusion Management: NIH eRA launched the new Human Subjects System (HSS) on June 9, 2018 to replace the Inclusion Management System (IMS). More information can be found here: https://era.nih.gov/hss_overview.cfm
- All of Us Research Program intends to solicit applications for large genome centers. These centers will generate genotyping and whole-genome sequence data, develop workflows for high-confidence genomic variant calling, and securely process and transmit data. The funding announcement was published on May 23, application due on July 12, with a start date in September 2018. This provides an opportunity for eMERGE to contribute to the AoU's effort.
    - AoU Research Program Genome Centers (OT2); ID Number OT-PM-18-002; https://allofus.nih.gov/funding/current-funding-opportunities
- In February, Eric Green, the NHGRI director, announced the beginning of a new round of strategic planning for NHGRI, to establish a '2020 vision for genomics' by publishing a new strategic plan in late 2020. The NHGRI hosted a "virtual" town hall on May 4, 2018 to begin establishment of the 2020 vision for genomics. More details are located here: https://www.genome.gov/27570372/
- 'Unlocking Life's Code Exhibition' is making stops across North America to bring awareness surrounding personalized medicine: https://unlockinglifescode.org/traveling-exhibit
- FY18 budget for NHGRI is $556 million dollars which is a 5.4% increase from FY17.
- Update on Year 5 extension supplement
    - Council approved the proposal for a one-year extension of eMERGE Phase III, however FY18 funds will not be appropriated to this endeavor, fiscal year 2019 budget will determine award amount.
        - Expect to award 70-80% of requested amount
    - Supplement work will start on April 1, 2019 and end on March 31, 2020. After "official" end date, sites will likely be allowed to continue using leftover funds via a No Cost Extension (NCE).
- NHGRI request feedback on the following timelines:
    - All Return of Results (RoR) and baseline surveys are expected to be completed in March 2019.
    - Six-month follow-up surveys completed by September 2019; All 12-month follow-up surveys completed by March 2020.
    - All phenotyping will be completed by February 2019; and data deposited into dbGaP by June 2019. Additional data from OMOP and Natural Language Processing (NLP) phenotyping should be deposited into dbGaP by March 2020.
- During the above timeframe, NHGRI expects the Network to disseminate research results and lessons learned by March 2020.
- As a reminder, the NHGRI proposed 7 areas that the network needs to focus on in the one-year extension and asked the network workgroups to establish Network-wide milestones in Year 5. In one of the 7 areas, "Expand the understanding of penetrance and the impact a variant has on clinical outcomes, using a variety of approaches including cascade testing and family history analyses", the NHGRI proposed to include Genomics along with Clinical Annotation workgroups to establish milestones and timeline.
- A complete list areas for Year Five Milestones can be found here.
- The NHGRI would like to continue including panel discussions on lessons learned in eMERGE III.
    - Winter 2018: Return of Results - confirmed, completed
    - Summer 2018: Phenotyping - confirmed, in progress
    - Fall 2018: EHRI, CDS, at-risk family members - confirmed

    - Winter 2019: Outcomes (joint with CSER) - confirmed
    - Summer 2019: Penetrance - confirmed
    - Fall 2019: Genomics - TBD
    - Winter 2020: Cost-effectiveness analysis - TBD
    - Summer 2020: All aspects of genomic medicine - TBD
- Goals for this meeting:
    - Share results of ongoing scientific projects
    - Discuss and agree upon the network-wide milestones and timeline for remaining years
    - Refine site-specific milestones and timeline
    - Establish sites' collaborations based on the similarities of site-specific objectives and milestones
    - Report on workgroups' activities, challenges, lessons learned, research results and newly developed tools/methods (if any)

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

- Upcoming Steering Committee meeting dates:
    - Fall 2018 Steering Committee/ESP Meeting: October 25-26, 2018 in Bethesda, MD
    - Winter 2019 Steering Committee Meeting joint with CSER: January 17-18, 2019 in Bethesda. MD
    - Summer 2019 Steering Committee Meeting: June 20-21, 2019 in Seattle, WA
- Developments since January 2018
    - Feedback from ESP:
        - The ROR workgroup should publish their initial findings to the genomics community to create awareness.
        - Explore collaborations with the Undiagnosed Disease Network (UDN) on topics related to variant classification and Variants of Unknown Significance (VUS).
        - Document reasons for the delay between receiving clinical reports from sequencing centers and returning these results to the patients.
        - Establish outlines and goals for the final two years of eMERGE.
    - eMERGE III extended for a fifth year.
        - Workgroups to focus on seven Network wide domains: Penetrance, outcomes, NLP phenotyping, impact of return, familial implications of return, institutional impact of return, and disseminating lessons learned.
    - Sequencing and clinical reporting ahead of target, 15,754 samples for initial dbGaP submission.
        - Clinical reports set to be completed by October 2018
        - 25,029 of 25,206 sequenced
        - 15,376 clinical reports issued
    - Sixteen of 27 phenotypes deployed; 10 implemented
        - Record Counter refresh underway of common variables for all eMERGE datasets.
            - eMERGEseq, PGRNseq, GWAS, Exome Chip, WGS
            - Deposited into Record Counter Summer 2018
    - eMERGEseq data incorporated into SPHINX
        - Filter criteria by platform, updated Drug & Gene Pathways with automatic refreshes of data
- eMERGE has many Network projects in process and we should focus on turning those into publications over the next year.
- Going forward in years four and five the Network should focus on:
    - Workgroup deliverables and publishing lessons learned
    - Return of results and outcomes data analysis
    - Streamlining phenotyping with OMOP transition
    - Finalizing eMERGEseq dbGaP submission

- John Harley welcomed the group and thanked everyone who participated in the Reds-Cubs baseball game and reception at the art museum.

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

- GWAS imputation paper has been favorably reviewed at Genetic Epidemiology and will be published soon.
    - Ian is incorporating additional 5,000 samples from Harvard.
- Adam Gordon (KPW/UW) has submitted the PGRNseq paper.
- SPHINX now includes eMERGEseq data
    - Pathway and Drug interactions are continuously updated on a weekly basis, linking to both the PGRNseq and eMERGEseq data
    - Linked to GWAS catalog variants to view prior clinical associations
    - Filtering ability by data set has been incorporated
- Planning a U24 grant to enhance the SPHINX resource, including rebranding to *eMERGENT* "Electronic MEdical Records and GENomics Toolkit." David Crosslin will reach out to sites for involvement in the grant application process.

Clinical sequencing centers (CSG) update | Richard Gibbs *(BCM)* & Heidi Rehm *(Partners/Broad)*

- Partners/Broad
    - Clinical reports will be completely returned by September of 2018
    - Number of positive reports/inconclusive: Harvard (73), CCHMC (99), Geisinger (209), UW (58/175)
        - Projected to have ~734 positive reports
    - All sequencing is complete and all data is uploaded to DNAnexus Commons; all data is also accessible through the DCR (de-identified case repository) and all reports available on GeneInsight (identified). In addition, 22,095 classified variants were submitted to ClinVar on March 23, 2018, including all eMERGE variants reports as of February/March 2018.
- Baylor
    - Clinical reports will be completely returned by October 2018 with the majority of sequencing completed.
    - Current and ongoing activities:
        - ClinVar submissions are underway
        - DCR uploads are also progressing
        - DNAnexus Commons usage is getting some use from various projects
        - VUMC IDs: Have to update IDs midway through the analysis process has caused challenges, the change is currently being finalized.
        - Quality control of Coordinating Center pVCF - identified minor issues which being addressed now, but largely concurrent. This was due to the Baylor target being slightly larger than the CC had assumed. Baylor is working with Ian Stanaway to address these differences.
    - Activities for years four and five:
        - Reclassifying VUS - rules and results
        - Increased resolution and data integration with NIH Structural Variants (SV) https://www.ncbi.nlm.nih.gov/dbvar/content/human_hub/
            - Harmonize eMERGE Copy Number Variants (CNVs) with population SVs at the Baylor College of Medicine Human Genome Sequencing Center (BCM-HGSC)
            - Expect to have 80,000 WGS samples which will inform this effort greatly
        - Ongoing reporting with updates and alert, specifically surrounding suspicious VUSs (Baylor)
            - ReVU (Reanalysis of Variants and Updater)
                - Identifies internal and external updates; oversees new reports; alerts, versioning

        - FHIR (*Fast Healthcare Interoperability Resources)* standards integration in collaboration with EHRI workgroup.
            - Coordinate with multiple groups for adaptation of FHIR standards for structured data.
            - Development of FHIR compliant file structure usable in Year 5 to replace to the current .XML file.
        - Communication and deployment
            - eDAPER (eMERGE data analysis and report)
                - De-identified Baylor/P-B reports in a searchable interface
                - Variant curation data updated nightly
- CSG Proposed Projects
    - Project One: eMERGE III Variant Reanalysis using non-eMERGE Data
        - Approximately 4% of cases receive updates due to variant knowledge changes a year. (Partners/Broad)
        - Some updates have already happened if a variant is reanalyzed by a CSG during the course of eIII. This may be prompted by another eMERGE or non-eMERGE case seen in by CSG. Also prompted by LMM's participation in ClinGen-led variant discrepancy resolution efforts for labs that submit their data to ClinVar.
        - CSGs will use ClinVar as a repository to reassess variants from the eMERGEseq cohort, both for reported and unreported variants. Baylor and P/B will initially complete this independently and then they will harmonize before any reports are released.
    - Project Two: eMERGE III Variant Reanalysis of suspicious VUS using eMERGE EHR Data (VUSs that favor/lean towards LP).
        - Note: VUS-5s found in at least three people at LMM sites only.
    - Project Three: Development of FHIR Compliant Structure Usable in eMERGE Etension Year 5 to Replace the Current .XML File
        - This will be further discussed in the EHRI workgroup breakout session.
- Discussion
    - Non-indication based consensus returnable results analyses can determine the estimates of positive incidental findings.
    - P/B will examine why cardiac variants are represented over abundantly in the "LMM Stats on VUS-Leaning Likely Pathogenic" figure. Missense mutations may be a factor, indicative of challenges in variant interpretation. The CSGs account for this bias in the representation of this data.
    - Clinical Annotations workgroup planning a separate project on penetrance that is more related to pathogenicity, however the two are related.
        - It is important to not conflate pathogenicity with penetrance
        - This may also be confounded when there are multiple loci associated with gene-disease interactions

Science presentation: The Adverse Drug Effect Recognize (ADER): A decision support tool for recognizing potential adverse drug reactions | Joshua Smith *(VUMC)*

- Adverse Drug Reactions (ADR) cause issues with quality of life, increased healthcare costs, however it is difficult to recognize and diagnose.
- Goal: To identify potential ADRs from pre-admission medications in newly admitted inpatients and alert providers so they could take action if necessary.
- Developed a tool, based on NLP called ADER (Adverse Drug Effect Recognizer), which scans admission notes and identifies abnormal lab results from the 24 hours prior to admission, compares the list of Clinical Manifestations (CMs) such as diseases, symptoms, findings, etc. and medications to a set of know ADRs. ADER will then proceed to issues alerts.

- Group conducted a three-month study period on internal medicine interns and residents. Control analysis used historical admissions from a prior period.
- Examined antihypertensive medications as there are a wide variety in internal medicine with ADRs, narrowed list to 54 drugs with 499 drug-ADR pairs.
- Potential confounders include the fact that a patient's disease may be responsible for potential ADRs.
- During the pilot the group included three questions along with the alert to determine if the alert was useful. They received approximately a 55% response rate.
- There were 2300 in the intervention group, 2600 in control group. There were approximately 2.3 potential ADRs per alert.
- 55% of physicians would not (or did not) change the patient's therapy due to possible ADEs.
- 32% of physicians would (or did) change the patient's therapy due to possible ADEs.
- 13% of physicians were uncertain if they would change patient's therapy due to possible ADEs.
- The intervention group ordered significantly fewer suspected ADR causing medications over first 24 hours as well as during the total stay.
- Conclusions
    - When the physicians responded to the survey questions it was more likely to stop the order of ADR causing medications both in the hospital and at discharge.
    - ADER can recognize potential ADRs.
    - Providers who receive alerts are more likely to hold on ordering ADE causing pre-admissions medications during inpatient stay.
    - Providers who see and respond to ADER are more likely to hold on ordering or discontinue suspected ADR causing drugs at discharge.
    - Provider responses to survey questionnaires are encouraging.
- The group initiated this project when VUMC was using their previous EHR system, StarPanel, and is now working to incorporate ADER into Epic.

Panel: Phenotyping best practices | George Hripcsak *(Columbia)*, Peggy Peissig *(Marshfield)* & Luke Rasmussen *(Northwestern)*

- Overview of Phenotyping (Peggy Peissig).
    - Process
        - What the Network originally anticipated:
            - Groups would write pseudocode, SQL, and programs at each site and perhaps even adapt them for use at other sites.
            - Not sure if the "work" of creating the algorithm was in the implementation or the validation of components.
            - It is unclear if PPV would be portable across the sites.
        - What the Network discovered:
            - Algorithms as flowcharts may be best - direct SQL code does not port well
            - Local experts are needed to implement
            - The science of algorithm components is hard
            - A general algorithm contains:
                - Administrative data; Structured data (labs, etc); Medications; NLP
    - Methods
        - Billing codes, clinical notes, medications, and lab/test results
        - Complexity of phenotypes has increased over the years from rule-based phenotyping algorithms to include negation, NLP, temporal logic, environmental data (geocoding)
    - Data Representation

        - Adopting common data models like OMOP and thus converting data warehouses to same schema/vocabulary while preserving source information allows for easier sharing of data and phenotyping.
    - Portability
        - Logic (meaning) is written in Word and PDF and requires re-implementation
        - Code (execution) conducted in KNIME or PhEMA (see BPH use case)
    - Validation
        - It is important to validate an algorithm and improve the phenotype accuracy. This review falls on physicians or trained extractors and standard validation instruments and tools (i.e., PheKB, PhEMA).
    - Tools for algorithm sharing and collaboration
        - PheKB.org facilitates sharing of phenotype algorithms. Located within the Phenotype KnowledgeBase (PheKB) is the Phenotype Execution and Modeling Architecture (PhEMA) tool which standardizes the representation of phenotypes by leveraging existing national standards, and allows for modularity and flexibility in execution agnostic to tools (KNIME, 12b2, SQL).
    - Evolution of phenotyping is moving to high throughput with machine learning, harmonized common data models (CDM), standard and reusable definitions, phenotypes that are scalable, and computational validation.
- Challenges and Solutions (Luke Rasmussen)
    - Challenge: Organizational complexity
        - Productivity decreases with multiple phenotypers only working part time on multiple phenotyping projects, which leads to over saturation. Project management and consideration of the site's resources is key to success.
        - Delays also build on each other creating a compounding effect. Network requests and site needs pull time and effort away from phenotyping (e.g. data refreshes and paper analyses).
        - Suggested solutions:
            - Strong project management at both a network, workgroup, and site-level are important.
            - Continue upfront planning of known work.
            - Be flexible in times of need.
            - Conduct a retrospective analysis of planned vs actual implementation time (include unplanned events).
    - Challenge: Phenotype complexity
        - Contributing to this include:
            - Reinterpretation of logic for implementation which sometimes features ambiguity and vagueness that must be overcome. Number of elements in logic. Modalities of data: structured, NLP, other ancillary systems. Complexity of logic. Entire phenotype package: including the Data Dictionary (DD).
        - Suggested solutions: Gain a better understanding of the complexity to help future planning. Conduct a retrospective analysis of planned vs actual complexity.
    - Challenge: Phenotype Validation
        - Contributing elements to this include:
            - Can be comprised of two groups of reviewers: clinicians or trained abstractors.
            - Resourcing can be difficulty, specifically with respect to scheduling/availability and incentives.
        - Suggested solutions: Creating multiple validation forms which would be completed as needed. Improving tooling to streamline review, such as leveraging SMART-on-FHIR chart

review application. Conduct unit testing to assess whether your code is doing what you have asked it to do.
    - Challenge: Data Dictionaries
        - Number of elements per data dictionary. Each is its own "algorithm" (complexity varies), however sites need to determine what is actually needed, and be aware that additional data elements come at a cost to both time and money.
        - Suggested solutions: Leverage software such as eleMAP or data validation on PheKB. Reuse common variables. Stewardship is appropriate at all levels (PIs, Phenotyping Workgroup, individual sites), all groups should be informed of delays and issues quickly in order to move progress along.
- Best Practices in Data Harmonization (George Hripcsak)
    - Common vocabulary standards would avoid wasted effort.
        - Use LOINC (Logical Observation Identifiers Names and Codes) where possible to help limit challenges in lab values.
        - RxNorm for medications and standard mappings.
        - SNOMED for clinical input, NLP, and OHDSI maps diagnosis administrative codes.
    - Vocabulary translation studies
        - Vocabulary translation studies:
            - Translation ICD9-CM to SNOMED CT causes differences in cohorts (e.g. prevalence) but not differences in study association results. (*Evaluation of alternative standardized terminologies for medical conditions within a network of observational healthcare databases*)
            - Mapping among drug classification systems (ATC, ETC, NDF-RT) caused variation in opioid use statistics. (*Applying standardized drug terminologies to observational healthcare databases: a case study on opioid exposure*)
            - Several eMERGE vocabulary studies are under review.
                - Mapping the data from ICD9-CM and ICD10-CM to SNOMED CT caused cohort errors only up to ~.2%, as long as the new concept set was reviewed manually.
    - Collaborative phenotype development and need for tools in addition to standards.
    - Need for NLP, harmonizing on system or output.
- Discussion
    - For Columbia, breast cancer phenotype has been kept simple in lieu of CKD's complex phenotype construction. This is a consequence of trying to meet the requirements of the eIII phenotyping expected quantity. Unfortunately, OMOP does not get to the nuances of phenotypes that NLP can tease out.
    - A specific site is using a different manifestation of OMOP and the ODSHI tools will not run, the other two or three sites plan to implement, however have not completed the task to date.
    - Future of phenotyping: optimizing current phenotypes or building new ones that are already optimized.
        - Leveraging both options would be beneficial.
        - Building tools that would reduce the effort of sites running the phenotype algorithms, instead of forcing sites to re-interpret these.
        - Examining phenotyping across the world may lend information to our efforts. SNOMED is able to be used internationally.
    - Is eMERGE positioned to take advantage of 'Deep learning' or 'machine learning' and how much of that is dependent on structural data versus NLP. This concept is still in the initial stages and in the research realm.

        - Machine learning can sometimes give counterintuitive findings and is difficult to determine what should be relied upon, what defines cases and controls.
            - Clinician review/interpretation will help facilitate this.
    - Errors were coming from two ICD9 codes mapped to a single SNOMED CT code and a particular ICD code yet not another. The group discussed where is the differentiation between code not mapped to previous algorithm or truly incorrect mapping.
        - This occurs where there are two higher level ambiguous codes and a judgement call has to be made if they should be used or excluded, and the intent of the original author as well as their knowledge base. Leads into discussion surrounding *when genetic variant definitions becoming part of the phenotype definition*. This may be a good topic for eMERGE IV.
    - Geisinger has a phenotype they are developing based on the genetic data, associated with FH in pediatric population
- Note: Summaries of past 'lessons learned' panels can be found on the website here (login required): https://emerge.mc.vanderbilt.edu/lessons-learned-panels/

Network discussion: Process for sharing phenotype data

- A major challenge is that, in addition to the algorithm the sites are developing, they also have to implement other sites algorithms, submit to dbGaP, MCS requests, and internal site demands.
- One way to reduce diversions is to steer investigators to the CC who have some data already collected and available for distribution with an approved MCS.
- Common variables collected (Full list can be found on Phenotype workgroup webpage, login required)
    - Demographics; ICD; CPT; Phecodes; BMI; Labs (custom list); Medications (custom list)
- Suggestions to facilitate process:
    - If a site signs up for a MCS, it is assumed the site will send data for the study if needed.
    - The CC redesigned the MCS form to include a section that specifically indicates which Common Variables and 'other' data needed from the participating sites. This allows sites to easily filter requests to determine if they have phenotyping effort available to participate in the MCS.
        - Sites should ensure if they sign up for a MCS they can indeed produce the needed data
    - Site specific data retrieval: requesting site could develop an electronic algorithm from the OMOP CDM to extract the data.

Science presentation: Transcription factors operate across disease loci, with EBNA2 implicated in autoimmunity | John Harley *(CCHMC)*

- John presented a five-year project that began with the question of whether transcription factors are concentrated at the loci of phenotypes.
- Initial research found that auto-antibodies could be identified an average of five years before patients were sick.
- They examined the mechanism between Epstein Barr Virus (EBV) and Lupus
    - Assumed both genes and environment contribute to the origin of etiology of lupus
    - Observed the genetic identify changes in allelic frequency with no insight into mechanism. 90% of genetic loci are predicted to be regulatory. Transcription factors are regulators of gene expression.
    - Predicted if assumption and observation are true, then we predict that the transcription factors made by EBV are concentrated in the lupus genetic loci.
- To test the predictions, the team used The ChIP-Seq procedure (Chromatin ImmunoPrecipitation followed by next generation DNA Sequencing).
- Lupus risk loci variants strongly intersect the DNA immunoprecipitated by EBNA-2 from EBV.
- Future plans:
    - Create a Network wide project carried out at several sites as there are now more qualifying phenotypes available.

    - REDS-db (Regulatory Element DNA Sequence database) had 1,544 TF ChIP-seq datasets, now has 11,000 with 20,000 overall, they are being re-evaluated from the original data to explore peak read parameter.

Outcomes processes summary | James Ralston *(KPW)* & Alanna Rahm *(Geisinger)*

- KPW/UW's experience detailed by James Ralston:
    - KPW/UW has returned 61 results through the clinical genetics department (Kathy Leppig). First wave has either colon cancer or colon polyps, with the majority of ROR occurring this past fall.
    - Chart abstraction by experienced Research Specialist mostly straightforward for colorectal cancer outcomes. Abstraction for other outcomes and generic outcomes required more expertise.
    - Review is mostly complete, there was some variation due to difference between research specialist review and physician quality assurance. Each reviewer adds to the findings, so in general may be better to have two reviewers.
    - *Early lessons learned*: abstractors need broad familiarity with medical terminology, and an ability to recognize the variety of health conditions and tests associated with actionable variants. Two separate abstractions may be needed for some outcomes.
- Geisinger's experience detailed by Alanna Kulchak Rahm:
    - Background: Geisinger returned 178 reports that were issued from batch one samples (June 2017 - January 2018). ROR information is available. Also collecting patient reported outcomes at six-months via mail, as well as medical care outcomes at six-months via chart review.
    - ROR personnel contacted all 178 individuals. 160/178 were considered a successful contact (90%).
        - Letter is sent to patient and uncontactable patients' results are documented in EHR.
    - Most ROR was for familial hypercholesterolemia due to pre-selection bias at the site in their first batch. It is important to note that Geisinger had men with breast cancer susceptibility, however the outcomes forms were only designed for females.
    - Approximately 165 paper follow-up surveys have been sent out and 27 individuals have responded. Geisinger may try using phone calls to facilitate a greater response rate.
    - Genetic counselors were used for abstraction, some records took longer than others, and they are formulating lessons learned.

Science presentation: An update on the Return of Actionable Variants Empirical (RAVE) study at Mayo: Challenges identified & lessons learned | David Kochan *(Mayo)*

- David Kochan presented initial findings from the return of actionable variants empiricals (RAVE) study at Mayo.
- There were 3030 participants to assess clinical outcomes, codes & utilization of genomic return of results.
- 2538/3030 participants were selected for sequencing; 57% Female; Average age of 63 years.
- Participants were selected for Familial Hypercholesterolemia (FH) and hereditary Colon Cancer; returned variants in 68 genes and 14 SNVs.
- 2110 negative results have been returned in Minnesota, and 110 in the Arizona cohort. Only five participants with negative results have contacted the group with questions.
- 106 positive results have been received: 25 positive FH results, 16 positive hereditary breast and ovarian cancer, and 12 positive lynch syndrome. 64 positives have been returned to participants.
- Seven variants were designated as VUS, three were reclassified as likely pathogenic after discussion with CSG. The remaining four cases were not reclassified. Segregation analyses on these participants is forthcoming.
- Potential somatic mosaicism (72 year-old male with a history of CLL) was referred for a skin biopsy to confirm results.
- Lessons learned: logistical challenges
    - Results are placed in EHR after visit with genetic counselor.

        - Three FH positives were parsed early; results were disclosed directly by the primary care provider (PCP) to participant.
    - There were six non-responders or deliberate ignorance, after six months the records were placed into the medical record along with notification of the PCP.
    - Withdrawals: The postal service placed a negative result in wrong mailbox. One participant was concerned with "transparent envelope".
    - Phone call ROR capability was important for timely results as some people move south in the colder months (snowbirds).
- Discussion:
    - Geisinger has also encountered deceased patients and engaged ethics advisory groups. They decided it was important to develop a protocol to return results to families. 30-40% of patients appeared to have died as a result of the condition they were flagged for.
    - Northwestern also has snowbirds and those have simply relocated permanently. Could not deliver results depending on the state to which they moved due to genetic counselor licensing.
    - Regarding the 'nosy' neighbor and the responsibility or accountability of the researcher currently paper mail is the standard of clinical care.
        - Some sites have prioritized ROR via an enabled patient portal. However, this is limited to patients who have enabled their patient portal.
    - Utilizing community advisory boards and groups to shape return letters to maximize understanding may also help reduce confusion.

Panel: Findings from re-review of medical records | Margaret Harr *(CHOP)* & Hila Milo Rasouly *(Columbia)*

- Margaret Harr. *Screening for 22q11.2 Copy Number Variations in large datasets*
- Margaret detailed the screening for 22q11.2 copy number variations in large datasets.
- The methods of the study were to determine the prevalence of 22q11.2 copy number variants (CNVs) in 30,000 pediatric samples (Ilumina 550k/610k arrays with CHOP MRNs), categorized CNVs based on breakpoints, reviewed the EHR, and compared findings to 10,659 eMERGE-1 genotyped subjects.
- EHR Extracted Data Review and Manual Chart Review were performed.
- EHR Extraction: 31 patients had a code specific for 22q11.2 deletion and 12 patients had a code for DiGeorge syndrome.
  Manual review: 44 patients with 22q deletion documented in chart, 1 patient was unable to be determined, and 3 patients had no evidence of diagnosis.
- Comparison of two methods: Extracted data can inflate the phenotype with multiple codes for the same symptoms, primary versus secondary codes for the same syndrome, and inability to separate genetic codes from environmental codes. Manual chart review was time consuming, subjective, and can require multiple reviewers.
- All typical and nested 22q11.2 deletions are previously diagnosed.
    - Distal deletions have a milder phenotype and are therefore not as likely to be diagnosed.
    - Duplications are more likely to go undiagnosed and have a more variable phenotype.
    - Screening for 22q11.2 CNVs may be indicated in all patients with Goldenhar syndrome/hemifacial microsomia.
- Future Directions:
    - CAG to expand analysis to a larger cohort (additional array types).
    - Contact eMERGE sites for phenotype data of eMERGE-1 subjects with CNVs and expand analysis to Phase II and III datasets.
    - Additional screening of large unselected cohorts can help to refine the prevalence and phenotype of atypical 22q11.2 CNVs.
- Hila Milo Rasouly. *Potential for variant misclassification in diagnostic sequence interpretation for kidney and genitourinary disorders.*

- 14% of adults have chronic kidney disease, it is highly heterogeneous in terms of genetic and clinical causes.
- Examined how many pathogenic variants are identified in a 'control' population not known to have CKD
- 625 genes are associated with kidney phenotypes with variable modes of inheritance, 25 are part of the eMERGEseq panel. Group used ACMG guidelines to identify candidate pathogenic variants.
- Examined concordance with the reported phenotypes and examined variant minor allele frequency in large cohort.
- 22% of cohort of self-identified healthy individuals carry a candidate pathogenic variant, 94% of these carry a variant in a gene associated with a dominant disorder.
- Determined even with a stringent MAF cutoff 1.3% of cohort of self-identified healthy individuals that carry a candidate pathogenic variant.
- Examined genes associated with dominant kidney phenotypes with largest numbers of individuals carrying previously reported pathogenic variants. PTEN, TSC2, and HNF1B overlap with the eMERGEseq panel.
- The group had little information about the participants, and used ICD codes to identify previously reported disease in eight individuals.
- This work highlights the importance of the systematic re-assessment of the variants reported in ClinVar and a similar need for reassessment in HGMD to facilitate the use of those clinical databases and clarify their penetrance and clinical relevance.
- Implications for eMERGE:
    - High burden of candidate pathogenic variants requires laborious annotation process.
    - ClinVar updates are very important.
    - Variable penetrance impacts the capacity to re-classify variants.
    - The importance of the combination of phenotyping and clinical annotation in assessing pathogenicity to reduce the risk of removing real variants and identifying new causative variants.

DAY 2: Tuesday

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

- Genomic and phenotype data collection is far enough along for data-based studies to proceed.
    - ~83k subject genotypes imputed
    - ~15k subjects eMERGEseq data frozen for analysis
    - ~1800 WGS available
- Network-wide DNAnexus project: CNV/SV calling
    - Genome studio reports (BAFs and LRRs) available from Phase I data
    - No data from Phases II/III (~90k) has been shared
    - CHOP has called variants on the Phase I data and submitted "CNV Association of Diverse Clinical Phenotypes from eMERGE Reveals Novel Disease Biology Underlying Cardiovascular Disease" by Glessner et al. to International Journal of Cardiology on June 4th. Screened for 22q11 deletions/duplications
    - Approved concept sheets that need array CNV data: NT087, NT151, NT177, NT186, NT209
        - The Genomics Workgroup needs:
            - Genome studio reports with the BAFs and LRRs would be the most straightforward,
            - IDATS would allow more flexibility at the cost of more compute time.
- SPHINX Tool
    - eMERGEseq data has been incorporated into SPHINX, and it has been suggested cloud storage may be a possibility.
    - Including PheWAS data (case/control counts) creates a unique resource to the genetics community.

    - Privacy concerns have been discussed amongst the group with Brad Malin.
        - Risk is less to an individual and more to organization. The breach would be determining if a person is present in a dataset, not necessarily their identity or medical records. Sites may have told the IRB they would not reveal presence of a person in a given dataset.
    - Genomics Workgroup recommends inclusion of PheWAS data in SPHINX and will send a proposal to the PIs for inclusion of case/control counts. Each site should weigh the risks based on the wording of their IRBs.
- eMERGEseq dbGaP submission: Genomic data
    - Workgroup discussed which products should be uploaded to dbGaP.
        - Single VCFs for each subject as reported out by each sequencing center
        - Multisample pVCFs - called by CC using GATK best practices.
    - pVCFs
        - Missing 26 variants as compared to single sample VCFs, due to having the incorrect bed file to identify sequenced regions. This is an easy fix, but basic QC must be conducted before upload to dbGaP and Genomics Workgroup will make an SOP for variant quality which is to be implemented in future dbGaP submission of this data.
- Milestones/Timelines
    - The clinical annotation will work on the bulk of the penetrance data collection this fall, and once the six-month mark has passed will go back and complete data collection.
    - Workgroup needs PI approval for inclusion of PheWAS data in SPHINX. Workgroup will finalize their recommendation, which will detail how to address privacy concerns.
    - Survey and data release of eMERGE II/III reports, then SV and CNV calling on all datasets followed by PheWAS analysis.

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

- Penetrance is the proportion of those with P/LP variants of interest who express the phenotype of interest
    - May be multiple phenotypes of interest.
    - Age/sex dependent (pediatric/adult difference).
- Use existing Outcome abstraction forms, to be completed six months after variant return.
- The group ranked the outcomes form by priority and assigned the lead site.
- The outcomes forms will be used for the 'returned' results as well as the genes they are not returning but still examining the EMR for penetrance.
- There are still many 'unclaimed' genes/conditions to further investigate that were not immediately prioritized, however they are still on the group's radar. There are also several SNVs that are currently unassigned.
- VUS classification: there are around 180 variants that fall into this category. Workgroup discussed that in parallel to the penetrance questions, the group will use the same process to investigate VUS outcomes.
- Discussion: The co-chairs agree that additional functional assays should be worked into the discussion and may fall outside of the immediate expertise in the room, but is a good avenue to explore.

Phenotyping & OMOP workgroup progress report | Peggy Peissig *(Marshfield)* & George Hripcsak *(Columbia)*

- 75% completion of algorithm development (21/28); 64% validation completed (18/28); 61% have been implemented (17/28), and 36% have data collection complete (10/28).
- Common Variables proposed to be added to the twice yearly refreshes include White Blood Cell differentials (~8) as wells as Autoimmune disease labs (~75-100).
- Site-specific data requests were discussed and options to leverage moving forward. Sites were reminded that the CC houses "common variables" across sites, including: demographics, BMI, ICD, CPT, Phecode, Labs (limited), Meds (limited).
- The MCS form will now include a section that explicitly calls out data request for common variables which can be obtained from the CC as well as other data variables that would have to be requested from the sites

themselves. With this information, the group hopes it will be more clear what types of data are being requested and the work associated with the MCS for the specific site.

- Year 5 Network-milestone: NLP
    - Develop an NLP component for a max of five high priority phenotypes.
    - Fifth year supplement proposals included NLP
    - NLP must complement an existing phenotype
    - Must use regular expression or cTAKES
    - The workgroup will vote on five out of eight proposed phenotypes.
- All sites have implemented OMOP
    - 8/10 completed era tables; 9/10 started OHDSI stack; 5/10 have completed.
    - 7/10 have run T2D phenotype query; 5/10 have returned survey.
    - Next phenotype is ADHD, and likely the third phenotype will feature a DD of covariates (SQL-only).
    - Group will publish findings based upon surveys.
- Publications:
    - "LPA variants are associated with residual cardiovascular risk in patients receiving statins" Wei et al, published in Circulation on April 27, 2018.
    - "A case study evaluating the portability of an executable computable phenotype algorithm across multiple institutions and EHR environments" Pacheco et al, published in JAMIA.

EHR integration workgroup progress report | Sandy Aronson *(Harvard)* & Luke Rasmussen *(Northwestern)*

- The group focused on milestones three & seven.
    - Milestone Three: *Improve and/or standardize genomic clinical decision support (CDS) for return of clinically relevant genetic or incidental results directly to physicians, including assessment of impact on relevant process outcomes.*
    - Milestone 7: *Disseminate lessons learned on the various aspects of genomic medicine implementation by activities such as publishing articles that propose the key elements for effectively returning genomic results to providers and patients and comparing the impact different methods of RoR have on patient and physician care across all sites.*
- Workgroup paper, *Empowering genomic medicine by establishing critical sequencing result data flows: the eMERGE example*, was published in JAMIA.
- Luke Rasmussen (Northwestern) spoke at the AMIA 2018 panel Moving Genomics into the Clinic: Informatic approaches from eMERGE, ClinGen, HL7, and GA4GH.
- The eMERGE DocuBuild web application was featured in GenomeWeb's article, Researchers Develop Web App to Improve Curating Delivering Genomic Knowledge to Point of Care.
- Develop a FHIR compliant file structure usable in the eMERGE Extension Year 5 to replace the current .xml file.
    - This is important because in eMERGE-III the workgroup developed and published a new file format for transferring clinical genetic results.
    - By developing a FHIR compliant standard the group can both reduce this fragmentation and ensure that eMERGE cases care robustly transmittable.
    - This project entails working with the HL7 Genomics workgroup to determine the best way to develop a FHIR representation of eMERGE clinical results.
    - If resources are adequate to represent eMERGE transactions then the group will develop eMERGE XML to FHIR Resource Migration Guide.
    - If resources are inadequate to represent eMERGE transactions the group will identify gaps and work with HL7 to evolve FHIR Resources.
        - If efforts to evolve FHIR resources are unsuccessful, the next step will be to develop FHIR Profile for eMERGE and Migration Guide for this profile.

        - If efforts to evolve FHIR resources are successful, the group will develop eMERGE XML to FHIR Resource Migration Guide.
    - The workgroup is NOT recommending that eMERGE Network migrate to new standard.
- The EHRI group catalogued the Year 5 site specific activities, to drive the Network wide projects forward on Clinical Decision Support (CDS).
    - EHRI Overview of Survey Results: nine sites are on EPIC, four have or are transitioning EHRs, nine are using structured variants and three using ancillary systems.
    - EHRI CDS Survey: Five sites are doing pre-test alerts, ten are doing post-test alaerts and eight are doing PGx alerts.
    - Planned survey: Where and how each type of data are available within a site's EHR (i.e. PDF reports, structured overall results, structured variant level results).
- Year Four Milestones
    - Continue integration and implementation efforts at sites.
    - Internal dissemination and tracking within the workgroup, including individual site presentations to collect data and create lessons learned.
- Year Five Milestones
    - Site-specific projects
    - Network MCS (Rasmussen) - eMERGE Genomics CDS
        - Examples of best practices today, what is achievable in the near term, what are the gaps/what needs to be addressed.
    - Network MCS (Overby-Taylor) Phenotype enabled CDS. Incorporates phenotyping algorithms into clinical decision support tools and determine impact on clinical diagnosis and treatment.
- As a reminder, Marc Williams is using ACT sheets for the ACMG actionable genes for Tier 1 conditions (FH, breast cancer, and Lynch Syndrome) and working with Columbia to use these sheets to create CDS (NT274, Harmonization of Genomic Medicine Outcomes: Comparison of eMERGE Outcomes to ClinGen Actionability Working Group Evidence-based Summaries).

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

- Update: expanding the eII PGx database
    - Common variables now being collected for PGx participants at regular intervals.
    - Update of comprehensive drug collection discussion delayed until full OMOP implementation.
- Recent PGx workgroup challenge: indeterminate results because of no information about phasing
    - Issue One: CYP2C19, the translation table associates both *1/*4B and *4A/*17 to heterozygosity for rs28399504 and rs12248560 (Both are assigned the same phenotype).
        - The group decision: CYP2C19*1/*4B or CYP2C10*4A/*17.
    - Issue Two: TMPT, the translation table associated both *1/*3A and *3B/*3C with heterozygosity for rs1142345 and rs1800460 (The phenotypes are not the same. *1/*3A is an intermediate metabolizer and *3B/*3C is a poor metabolizer).
    - Because of the rarity of *3B/*3C, the group consensus: *1/*3A, intermediate metabolizer, cannot rule out poor metabolizer, phenotypic testing is available to distinguish between these alleles.
    - Heidi Rehm suggested the group could calculate the likelihood of compound vs configuration for allelic frequencies.
        - CYP2C19 change resulted in 14 reissued reports; TMPT resulted in 304 reissued reports.
- Six papers published, 11 papers in progress
- Role of PGx workgroup in Year Five
    - Can be involved with CDS integration, specifically EPIC integration questions.
    - Several sites are intending to return some or all of the PGx related variants on the platform. Only CHOP has begun to return the PGx results.
    - Outcomes related to PGx may also be an avenue to explore.

    - Identify and conduct Network projects using existing data in SPHINX or the potential new eMERGENT tool.
- Discussion:
    - Opportunity: EPIC-18 update will have more capability to store PGx variants and structured data, also will feature some CDS incorporated. It is not required for organizations to have Epic Beaker, the Laboratory Information System (LIS). The institutions can import that information directly into the EHR. The genomics model however is not included as a standard in EPIC 2018, it should be added on for an additional charge.
    - EPIC genomics group is looking for partners to provide input regarding how to store genomic information (knowledge management). eMERGE should open a dialog to discuss how our resources can support their group.
    - Be cognizant of the FHIR standards for migrating xml tool. There will be navigation from this data to new FHIR standard and how EPIC absorbs this data.
    - CONSENSUS: PIs agree it is very important to continue PGx work throughout the duration of the eMERGE Network. PGx can focus on PGRNseq for discovery research and the eMERGEseq panel for clinical related research
    - eIII outcomes forms are to be completed on subjects who have actionable results. This would not apply to the PGx participants. The group can continue to collect PGx outcomes data the way it has previously, using ICD codes.
- Opportunity: CPIC informatics group has overlap with the eMERGE PGx/EHRI groups
    - Marc volunteers Nephi Walton (Geisinger) to function as a formal liaison between the two groups. Group will also reach out to Bob Freimuth (Mayo).

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

- The group was involved with milestones two, five, and six.
- Milestone Two: *Determine the impact of ROR on patients' immediate outcomes*
    - Participant surveys to coordinate across sites, CP surveys and interviews (via RO1), coordinating with Outcomes workgroup, and produce lessons learned and challenge summaries.
        - Currently challenges include deceased, and not traceable, or non-engaged participants. How patients and providers respond, views on family sharing, and moving from a minor to becoming an adult.
- Milestone Five: *Explore challenges involved in identifying at risk family members.*
    - Participant surveys
    - More in-depth efforts at several sites, and deceased patients.
- Milestone Six: *Estimate the institutional impact of ROR*
    - No current projects, but intertwined with many other goals of the workgroup.
    - The group plans to develop a guide for IRB language, consent procedures, and ROR process as well as estimate the economic impact of ROR.
- The workgroup has several publications that are currently being worked on and several planned publications, including the ROR process and content comparison of letters sent to participants across sites.
- Returning genomic results to eMERGE participants: The who, what, where, and how of disclosure (Georgia Wiesner & Kathy Leppig)
    - Return of results process begins when the sequencing results are submitted to the clinical site.
    - Three steps: Returned to participant, returned to health care provider, and integrated into EHR.
    - There is significant variability between sites in terms in what is being returned, and the methods.
    - Lessons learned:
        - Variability in ROR across all sites and adjustments made during the process are used to accommodate institutional and participant feedback.
        - Most sites (7/10) disclose results to participant as initial step.

        - This approach differs from standard lab practices for test reporting of EHR first, then to HCP, then to patient.
        - Impact of results may differ for participants in different age groups.
    - Next steps:
        - Finalize implementation science review including correct errors in tables, draft with comments, and complete manuscript.
        - Use as framework for comparing final results of ROR across eMERGE sites.
        - Development of guide for return of results for genomic results.

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

- The Outcome instruments (18 in REDCap) have been the focus of the workgroup. There are 1410 fields and each site has begun implementation. An initial debugging cycle is underway.
- The forms have to count for the 'many to many' relationship between gene variants and disease.
- The generic outcomes form will also be included as an optional feature for each participant.
- As the forms are applied in actual practice there are issues that have to be addressed and solutions incorporated to the ongoing data collection.
- A Network wide MCS (NT296) is planned for this work, Collection and Analysis of Large-Scale Outcome Measures following Targeted Next Generation Sequencing).
- Year Five Milestones
    - Complete six-month Outcomes by summer of 2019, and deliver an interim analysis by June 2019 and submission of MCS by winter 2020.
    - Complete 12-month Outcomes assessment by Spring 2020.
    - Estimate the institutional impact of RoR and Outcomes.
        - Economic measures such as billing, visit volume, counseling time.
        - Examine chart reviews and complete surveys.
- Outcomes Harmonization (Marc Williams)
    - There is a lack of standardization of outcomes harmonization processes across projects and sites. These data are critical to leverage in genomic medicine research to inform current and future efforts.
    - Action taken by eMERGE: Geisinger compared eMERGE outcomes pairs (gene/condition/outcome) to ClinGen actionability workgroup scored outcomes, and then expanded to evidence reviews. The evidence reviews enhanced comparisons as they added additional, useful information on the Outcome pairs. Reviewed by senior author of MCS for fidelity (Marc Williams).
    - Results: Overall, 12 disorders were scored; three eMERGE gene/variant disorder pairs with outcomes do not have an actionability score or evidence review summary; five eMERGE outcomes had equivalent definitions; four had differences.
        - Difference example: Colorectal cancer featured grouped/coupled gene pairings (lynch syndrome + FAP), whereas AWG (actionability working group) decoupled colorectal cancer gene pairings into lynch syndrome genes and FAP.
    - MCS has been submitted as a special issue of Healthcare and is available as a pre-print.
    - Proposal to eMERGE: ClinGen AWG O/I pairs (scored and reviewed) to be used as starting point when available. This will facilitate creating a standardized process; evidence review is performed lowering reliance on personal opinion; and if one is not available, eMERGE can communicate with AWG to have the outcomes pair prioritized.
    - Data capture in REDCap can become complex and may impact collection.
        - Outcomes forms are intended to be used by a site-specific person familiar with outcomes abstraction from EMR for the conditions of interest.

        - Feedback from KPW/UW's initial experience: It is recommended that each site create a quality assurance process for their site-specific team conducting the outcomes abstraction and form completion.
    - All forms have been reviewed by the Outcomes Workgroup and are posted online.
    - A google document to catalogue information on specific fields and provide tips for completion of the Outcomes forms will be circulated to the group.

Network discussion: Network goals & milestones for year Four | NHGRI eMERGE Team

- Recommendation to examine the MCS timeline/milestones and contact authors to determine where they are on their MCS. The CC uses a tracking grid and most authors keep the document up to date. Reasons for delays should be documented on the tracking log and brought up on workgroup meetings in order to facilitate movement.
- Rongling reminded the Network to publish on current work even if not all data are collected.
- Milestone One was high priority across multiple sites based on the supplement request. Clinical annotation workgroup has developed a plan to address and will reconcile timeline with Genomics.
    - Concerns raised about cascade testing and family history. VUMC is not capturing these data so the group will need to assess which sites can actually participate.
    - There is overlap between milestone one and five, and the network should harmonize efforts.
- Milestone two has six and twelve-month hard deadlines for their timeline.
- Milestone three: To harmonize with FHIR (Year five) and the other activity. There will be a MCS focusing on the milestone, and the group will gather data over the next year.
- Milestone four: Phenotyping workgroup will recommend phenotypes for NPL by end of July. Timeline for execution of incorporating NLP into phenotypes will be devised post-selection.
- Milestone five: The ROR group has projects associated with the group to collect this data, timeline is pending. The cascade screening data for penetrance analysis may be difficult as it takes time to find the proband, bring them in, and contact family members, and may not be as feasible.
- Milestone six: Confirm sites are in agreement with the proposed table, timeline is pending.
    - Economic impact will be difficult to capture across all sites, so it may not make sense to explore this opportunity. However, it is suggested to Incorporate implementation science framework to help gather this information.
- Milestone seven will be tracked by publication tracking sheet and MCS process.
- Network-wide milestones should be reviewed closely and content harmonized where appropriate.
- The CC will assist workgroup by created a GANTT chart after milestone timelines have been finalized to ensure timelines are met.
- NHGRI is working with sites individually for their specific Year five milestones. Once the milestones are agreed upon they will be incorporated into the grants management system and be signed by the institutions business office.
- The milestones for year five will be incorporated into the notice of award for each site.

Action Items

PIs/Sites/Co-chairs

- All workgroups will finalize their Year 5 milestones and timelines and return to the CC by Monday July 16th for distribution and discussion on the PI call, July 19th including addressing challenges and obstacles for addressing these items.
- The PIs will confirm and provide any feedback on the list of proposed panels at the future face to face meeting to the CC by Monday July 16th.
- The PIs should provide feedback on the proposed additions to the Common Variables to the CC by Monday July 9th and inform the CC if their site approves of the additional laboratory values.
- PIs will finalize site-specific milestones and timelines and report these back to the NHGRI.

- Please review the linked publication tracking log and edit as needed. Site leads should document reasons for delays in order to facilitate progress, these will be discussed at the Thursday July 19th PI call.

Coordinating Center

- The CC will circulate the newly redesigned MCS sheet which incorporates check boxes for common variable and 'other' data requests**.**
- The CC will produce a GANTT chart for the seven milestones across the Network once the timelines have been established by workgroup for circulation to the NHGRI and the Network.
- The CC will collect and circulate a list of standard definitions of commonly used terms for return of results (e.g. positive, pathogenic, negative, non-pathogenic, eDAPER, eCAP tool, ect…).
- CC will investigate 'impact tracking' to capture the utility of eMERGE products through tracking number of downloads of a publication, including on social media (Altmetric).
- The CC will recirculate the publication tracking log and discuss at the next PI call any discrepancies in publication milestones and update.

Phenotyping

- The Phenotyping Workgroup will create a summary of lessons learned panel for circulation to the Network.
- The Phenotyping Workgroup should submit a MCS based off the lessons learned for publication.

Clinical Annotation

- The clinical annotation group will finalize the prioritization of outcomes and assignment to each site for the upcoming penetrance analysis.

EHRI

- The workgroup will Create EHRI lessons learned panel for October 2018 Steering Committee meeting.
- The EHRI group will submit the two newly proposed MCS documents by Luke Rasmussen and Casey Overby-Taylor to the CC for circulation to the PI group.

Genomics

- The Genomics workgroup will create a formalized recommendation by Monday July 16th for the incorporation of PheWAS case/control counts into SPHINX for the PIs to vote on during the July 19th PI call.
- The Genomics workgroup will work with Clinical Annotation to finalize timelines for penetrance analysis.
- The workgroup will provide feedback on renaming of the expanded SPHINX resource.
- The Genomics Workgroup will create a REDCap survey to inventory CNV genomic data and determine if sites are willing to share, the data are accessible, and contact personnel.
- The Genomics Workgroup will create standard operating procedures for the QC of the VCF files by September 1, 2018.

PGx

- PGx group will create a formal relationship with the CPIC Informatics group.
- The PGx group will continue to meet during year five and provide report-outs in the Genomics and EHRI group as needed.

ROR/ELSI

- The ROR group will request updates on quantity of negative and positive return of results at each site monthly to ensure leadership and the steering committee is informed of progress.
- Sites should review the draft table slide for the ROR process project and correct any inconsistencies.

Outcomes

- The Outcomes group will submit a MCS to the CC for circulation examining the Outcomes process (NT296, Collection and Analysis of Large-Scale Outcome Measures following Targeted Next Generation Sequencing).
- The outcomes group will disseminate a shared google document guide providing information on specific fields and provide tips for completion of the Outcomes forms.