****

**Summary of eMERGE Phase III Closeout Call**

June 4, 2020

**Meeting Agenda**

* [NHGRI program official report | Robb Rowley (NIH/NHGRI)](#kix.y1zr56k7uzxd)
* [Meeting Goals *&* Announcements | Rex Chisholm (SC Chair, Northwestern)](#ef6u1jtvdn4t)
* [CSG Update | Hana Zouk (Partners/Broad), Eric Venner (Baylor), & Mullai Murugan](#mim8jbmj81qj)
* [Phenotyping workgroup natural language processing update & eMERGE III final deliverables |Wei-Qi Wei (VUMC) & Chunhua Weng (Columbia)](#rlt2znq5h5j3)
* [Outcomes workgroup outcomes analyses & eMERGE III final deliverables | Josh Peterson (VUMC), John Connolly (CHOP), & Christin Hoell (Northwestern)](#w1qb0870yv83)
* [ROR/ELSI workgroup eMERGE III final deliverables | Ingrid Holm (BCH) & Iftikhar Kullo (Mayo)](#8yrs4wigu76b)
* [EHRI workgroup eMERGE III final deliverables | Sandy Aronson (Partners/Broad) & Casey Overby Taylor (JHU/Geisinger)](#ok977aqiuhu5)
* [PGx workgroup eMERGE III final deliverables | Cindy Prows (CCHMC) & Laura Rasmussen-Torvik (Northwestern)](#9a48vdpnnzt7)
* [Genomics workgroup eMERGE III final deliverables | David Crosslin (KP/UW), Megan Roy-Puckelwartz (Northwestern), Patrick Sleiman (CHOP)](#y1ik6irhxqmu)
* [Clinical Annotation workgroup eMERGE III final deliverables | Gail Jarvik (KP/UW) & Heidi Rehm (Partners/Broad)](#6z3wys7ebtwl)
* [Overall lessons learned from eMERGE III and recommendations for eMERGE IV](#e9069wehqky9)
* [Closing Remarks | Rex Chisholm (SC Chair, Northwestern)](#8kitg6eaz736)
* **NHGRI program official report | Robb Rowley (NIH/NHGRI)**
  + NIH COVID-19 Information
    - Changes to policies are available on the website for investigators. Please notify program staff with any issues so they can help resolve.
    - NIH is regularly issuing news releases regarding COVID-19 findings which is also available on the website.
    - The NHGRI issued a notice of special interest and received some excellent requests that are currently being reviewed. If it was submitted as a competitive revision, the review is July 13, 2020.
  + ‘Genomics2020’ Strategic Planning
    - The institute’s new strategic plan will be published in October 2020. A draft version was made available for comments last month and NHGRI is now actively synthesizing the feedback which will be submitted for publication in early summer.
    - The NIH-ACMG Fellowship in Genomic Medicine Program Management is a collaboration between NIH and the American College of Medical Genetics and Genomics (ACMG) and applications are due December 1, 2020. The two-year fellowship is now expanded to include genetic counselors, nurse practitioners, and physicians’ assistants.
  + FOAs: PRS Methods and Analysis for Populations of Diverse Ancestry
    - This is looking at new ways of developing analytical validity to tests and it is a plan to collaborate with this in eMERGE IV.
    - The LOI is due 30 days prior to application due date of June 23, 2020 and the earliest start date is June 2021. More information for study sites can be found [here](https://grants.nih.gov/grants/guide/rfa-files/RFA-HG-20-001.html).
  + Research Performance Progress Report (RPPR)
    - The RPPRs are due within 120 days of project end dates.
      * Applicants that applied as Type 1 for the next phase of eMERGE or did not apply for the next phase will need to submit a final RPPR.
      * Applicants that applied as Type 2 will need to submit an interim RPPR and if not funded for the next phase of eMERGE, that could be converted into a final RPPR.
  + NHGRI would like to thank all eMERGE investigators and their great contributions over the years.
* **Meeting Goals *&* Announcements | Rex Chisholm (SC Chair, Northwestern)**
  + The goal of this final meeting is to review the final workgroup accomplishments and discuss how our work has affected the research, infrastructure, and the genomics field in general.
  + A Reminder of Network Accomplishments
    - The specific aims of eMERGE III were to sequence and assess clinically relevant genes, assess the phenotypic implications of variants, integrate genetic variants into EMRs for clinical care, and create community resources all which were completed.
      * A huge milestone includes the six eMERGE datasets with over 158,000 samples on AnVIL and dbGaP as well as lessons learned publications across all 7 workgroups.
    - As of May 2020, there were 863 network and site specific projects with 683 of those published.
    - eMERGE dbGaP submissions have been widely used with 1,473 external downloads as of May 2020.
* **CSG Update | Hana Zouk (Partners/Broad), Eric Venner (Baylor), & Mullai Murugan (Baylor)**
  + Hana Zouk (Partners/Broad) and Eric Venner (Baylor) provided a broad update on the reassessment work at Partners/Broad and Baylor.
    - The goal of Partners/Broad re-assessment of variation identified in eMERGE participants is to create a framework to strategically identify, prioritize, and re-assess variants most likely to undergo reclassification in the eMERGE dataset.
      * 1,995 prioritized unique variants identified in post-filtration. 67 participants were affected by 45 variant reclassifications including 37 upgraded and 8 downgraded. There were 32 novel or amended reports, however not all were issued or amended due to site reporting preferences.
    - The goal of Partners/Broad variant reanalysis of VUS-leaning pathogenic variants using eMERGE EHR data was to identify variants with higher probability of reclassification from VUS-LP to LP/P and integrate phenotype data from EHRs for improved interpretation.
      * Primary review of phenotypic data from 161 participants for 104 unique variants did not identify any variants that would warrant an upgraded classification. For most variants in genes associated with more complex phenotypes (e.g. cardio, FH, CRC) EHR data was limited.
    - At Baylor there were two capstone reanalysis projects (CLinVar re-analysis and VUS-Leans Pathogenic) following on 6,473 variant reviews and 20,585 issued reports.
    - The Baylor ClinVar re-analysis project was completed to update reports when variant classifications change using REVU to compare the latest ClinVar to local data to identify upgrades and downgrades. 26 variants were identified with a new P/LP classification in ClinVar and 86 variants identified with new VUS/LB/B assertion resulting in 5 reclassified variants and 5 updated reports.
    - The Baylor VUS-Leans Pathogenic project included 83 variants. Clinical sites were contacted with variant lists leading to 4 updated reports due to the new clinical information.
  + All DNA Commons projects were decommissioned between May 11-20, 2020. PHI Commons projects are retained with restricted access.
  + All GeneInsight user interfaces: De-identified case repository (DCR), Lab and Clinic Interfaces have been sunsetted as of May 15, 2020.
  + Mullai Murugan (Baylor) provided a FHIR/EHR update including FHIR accomplishments, remaining work, and future considerations.
    - Accomplishments include the development of a computable and standardized clinical reporting specification for eMERGE with HL7 FHIR, development of a base eMERGE FHIR spec on HL7 Clinical Genomics Implementation Guide, POC implementation pilot to generate eMERGE reports with FHIR, and NU/JHU feasibility testing of FHIR enabled ingestion and CDS with EHRs.
    - There is remaining work in the completion and publication of readthedocs eMERGE FHIR Spec documentation and genomic considerations for FHIR manuscript.
    - The group will consider adoption and direction, variant representation, participation in clinical genomics, models of EHR interoperability, diversity of tech landscape, and additional use cases.
  + With regards to how much are the recommended FHIR specifications contributed being used:
    - They are being used quite a bit with 22 major ones incorporated in specifications.
    - The group is continuing to work on variant representation with Bob Freeman of HL7 workgroup.
* **Phenotyping workgroup natural language processing update & eMERGE III final deliverables |Wei-Qi Wei (VUMC) & Chunhua Weng (Columbia)**
  + Phenotyping Timeline of 25 Algorithms (not including 5 NLP algorithms)
    - By May 2020, the workgroup has successfully developed, validated, and implemented 25 algorithms.
      * The average PPV from primary sites was 0.95 for cases and 0.98 for controls.
      * The average PPV from secondary sites was 0.91 for cases and 0.95 for controls.
      * Once the work group became familiar with the process after the first year, it took the group about 5 years to implement all 25 algorithms indicating that they can manage 5 to 6 high quality algorithms per year.
  + The phenotyping milestones included converting the EHR data of the eMERGE cohort to OMOP significantly improving the reusability of SQL scripts. Additionally, milestones included the identification of a common set of variables for future study use as well as phecode expansion into ICD 9 and 10 versions.
  + NLP Subgroup Charter
    - The NLP subgroup began working together in 2019 to develop, validate, and implement NLP components for five high priority phenotypes while identifying and summarizing the challenges to develop suggestions for future work.
    - After working together for a year, all five NLP algorithms were developed and implemented with three of them having been released to the network.
      * It was anticipated early on that each site would not have sufficient resources to maintain 5 different NLP pipelines so a survey revealed the two most popular pipelines: cTAKES and MetaMap.
    - CRS - Geisinger & Northwestern
      * Since the PPV of the original algorithm remained low, it was explored whether or not extra data would help improve PPV.
      * It took Geisinger 12 months to develop the algorithm and Northwestern five months to validate.
      * The result from Geisinger indicated very slight improvement in both PPV and sensitivity compared with the previous algorithm using structured data but a large improvement in specificity did occur.
    - ECG Traits - VUMC & Mayo
      * The goal of this algorithm was to identify 15 distinct ECG sub-traits that were unidentifiable simply using structured EHR data.
      * It took VUMC nine months to develop the algorithm and Mayo seven months to validate (six months for controls).
      * The average PPV was .97 (except for Afib which was around .80).
      * The algorithm could not find sufficient Brugada cases from ECG reports.
    - Lupus - Northwestern & VUMC
      * The goal of creating a Lupus algorithm was to improve sensitivity. It did improve from 79% to 91%.
      * It took Northwestern 12 months to develop the algorithm and VUMC 11 months to validate.
      * Validation indicated better performance in two of the three sub-phenotypes.
    - ACO - Harvard & KPW/UW
      * The goal of the ACO algorithm was to improve sensitivity.
      * It took Harvard six months to develop and KPW one month to validate.
      * The workgroup was very happy to see a substantial improvement in sensitivity (38% to 54%).
    - FH - Mayo & Geisinger
      * The primary goal of this algorithm was to get more information from family history while improving PPV.
      * Mayo spent 12 months creating the algorithm with an additional two months converting the code to cTAKES. It took Geisinger six months to validate.
      * The algorithm did not perform better when validated due to differing progress notes at the two sites. Additionally, switching pipelines indicated the default medication modules need improvement in cTAKES.
    - Conclusions
      * NLP does improve phenotyping performance, especially sensitivity.
      * Results suggest the feasibility of developing portable NLP algorithms with very consistent performance.
        + To do so, extra resources are required. For example, an NLP algorithm cannot run against an EHR database so a separate server is required. Additionally, running an NLP algorithm requires specific code development.
    - Lessons Learned
      * The biggest difference between NLP and structured data is privacy protection. Using notes requires special permission and sites need to have separate servers.
      * The output from the NLP pipeline may not be accurate. For example negation errors, missing abbreviations, and misspellings are all common issues.
      * Codes of NLP algorithms are different from SQL scripts and could be considered intellectual property of sites.
      * Different sites may have different naming systems for their notes creating more challenges.
      * NLP may not be able to generalize to rare phenotypes.
    - Suggestions for Future NLP Experiments
      * Starting with semi-structured notes may improve success, for example using problem lists, medication lists, or drug allergies.
      * A vital step is formalizing the documentation strategy and standardizing protocol for NLP validation and implementation.
      * Enhancing code modularization and improving communication between sites are also suggestions for future NLP experiments.
    - Wei-Qi and Chunhua acknowledge all participation of the eMERGE phenotyping group and thank all involved.
    - What are some recommendations for choosing to use NLP for future phenotyping and determining ahead of time if NLP is needed?
      * Starting with reliable and structured data is a start. If that does not work, for example if you want to improve sensitivity, you would definitely want to use NLP. Semi-structured data like problem or medication lists are easy to start with and allow you to avoid issues like negation errors.
      * It is the hope that after this pilot study, more institutions can implement NLP work, especially after the OMOP NLP table model is ready.
* **Outcomes workgroup outcomes analyses & eMERGE III final deliverables | Josh Peterson (VUMC), John Connolly (CHOP), & Christin Hoell (Northwestern)** 
  + Josh Peterson (VUMC, Workgroup Co-Chair) presented the Outcomes workgroup accomplishments on Network milestones and contributions made.
    - In order to access the impact of return of genetic results on outcomes, 104 data collection instruments were developed with abstraction guides. After data discrepancies were identified, a final (third) data freeze was distributed in March. The workgroup refreshed post-ROR CPT and ICD data for all sites.
    - Cost analysis associated with NT296 post-ROR diagnostic tests and procedures has helped estimate the institutional impact of ROR.
    - The outcomes forms were modified to also collect data on results not returned to help estimate penetrance.
    - The outcomes workgroup has defined and harmonized measurable health services and clinical outcomes across a broad range of genomic medicine scenarios.
  + The Outcomes workgroup has six primary manuscripts in progress focused on the outcomes data collection instruments.
    - NT242, Clinical outcomes after screening for cardiomyopathy genes led by Christin Hoell (NU)
    - NT245, Penetrance, cancer types, and outcomes of cancers associated with germline mutations in hereditary breast cancer genes and the impact of return of results of mutations for hereditary breast cancer on medical utilization and health outcomes led by Katherine Crew (Columbia)
    - NT376, Heterozygous FH in the eMERGE Network: Penetrance, Outcomes and Cardiovascular Risk led by Ozan Dikilitas (Mayo)
    - NT365, The Association between Variants in Ion Channel Genes and Arrhythmia Phenotypes led by Andrew Glazer (VUMC)
    - NT209, 22Q11.2 Deletion Syndrome, Leveraging Copy Number Variation to Examine Health Outcomes led by Patrick Sleiman (CHOP)
    - NT296, Impact of Targeted Next Generation Sequencing on Diagnostic Testing and Screening PRactices led by Josh Peterson (VUMC)
  + Hakon Hakonarson (CHOP) presented 22q Deletion/Duplication syndrome preliminary outcomes analyses findings.
    - SNP-array data from 10,659 adult participants was collected in eMERGE-I. SNP-array data from over 30,000 participants from CAG was collected in eMERGE II and III.
    - CNVs were classified according to breakpoints and size consistent with the regions of segmental duplications LCR22A through LCR22H. Previously unknown CNVs resulted in clinical feature reviewing to identify findings consistent with 22q11.2 deletion or duplication.
    - Rates of 22q11.2 CNVs in CAG cohort pediatric patients were 1/394 with CNV called in 22q11.2, 1/625 Deletions, and 1/1071 Duplications.
    - Rates of 22q11.2 CNVs in eMERGE adult patients were 1/592 with a CNV call in 22q11.2, 1/2664 Deletions, and 1/761 Duplications.
    - Out of the 48 Deletions: 31 had a EHR code for 22q11.2, 3 diagnoses not indicated in EHR, 12 had a code for DiGeorge, one in text only with no code, and one unable to be determined.
    - Out of the 28 Duplications: 25 had a diagnosis not indicated in EHR, two patients with known 22q duplications, and one patient with a code for chromosome 22 microdeletion.
    - In conclusion, pathogenic CNVs in 22q11.2 region are often missed, data lend support to national efforts of newborn screening, and four COMT1 SNPs included on the eMERGEseq panel were sufficient to identify undiagnosed patients with 22q11.2 deletions and duplications.
    - Ongoing work includes working with sites for patient record confirmation, return of results ongoing at CHOP of previously undiagnosed patients, and assessment of major outcomes between known and unknown cases.
  + Christin Hoell (NU) presented Cardiomyopathy outcomes analyses preliminary findings.
    - Out of 138 cardiomyopathy records, 112 were appropriate for outcomes analyses, 13 were pediatric and 11 were not returned. Out of the 112 for outcomes analyses, 86 had no previous diagnosis, 15 had previous clinical diagnosis, and 11 had previous genetic diagnosis.
    - Of the 86 participants with no prior diagnosis, seven were diagnosed with cardiomyopathy post-ROR.
    - 90% had assessment of heart function and 43% had assessment for arrhythmia post-ROR.
    - Initial outcomes due to ROR intervention can be assessed by 0-6 month forms alone with most tests performed or diagnoses made 0-6 months post-ROR.
* **MeTree Supplement Update: Exploring the implementation of family health history into EHR | Georgia Wiesner (VUMC)**
  + The overall goal of the MeTree supplement is to promote the adoption of family health history (FHH) in genomic medicine by leveraging strengths of the eMERGE Network and Duke’s MeTree FHH collection tool.
  + The group has developed a semi-structured interview guide, developed key informant list for each site and specific key informants for each location with video explanation for key informants and IRB review by all sites that consent is not needed; and guided by the CFIR, conducting semi-structured interviews with key stakeholder sites.
  + 57 interviews have been completed and transcribed, structural coding is complete, and final review by the entire team is planned by June 2020. An abstract of results to be submitted to ASHG 2020 conference.
  + Next steps for the supplement are to perform a multi-domain technical assessment of the requirements for SMART-on-FHIR technology for integration of FHH and submission of manuscripts by summer 2020.
* **ROR/ELSI workgroup eMERGE III final deliverables | Ingrid Holm (BCH) & Iftikhar Kullo (Mayo)**
  + Ingrid and Iftikhar presented the ROR/ELSI workgroup accomplishments, remaining work, and lessons learned from workgroup milestones.
  + To access the institutional impact of ROR, the ROR workgroup has manuscripts submitted or published (NT273, NT277, NT300, NT323) and in development (NT322 and NT332). The lessons learned from collecting the institutional impact has been to coordinate the return of result process across sites prior to initiation.
  + To collect the impact of return of results on participant outcomes, the ROR workgroup has coordinated the Participant Survey across sites and reconciled the data dictionary for most questions. The related manuscripts (NT363, NT373, and NT389) are in progress.
  + Participant surveys and healthcare provider surveys have assisted with the collection of patient challenges in communicating results with family members.
  + The collection of participant survey and cascade screening data at pediatric sites had contributed to the return of results in the pediatric setting. The group is considering a lessons learned manuscript on the challenges of return of results in the pediatric setting.
  + The ROR/ELSI workgroup has been an overarching impact on the field by providing guidance for investigators in navigating IRB for ROR projects, impacting providers and patients with comprehension and personal value, giving insight of how participants react to neutral results, instructions for how to be prepared for challenges encountered in ROR (e.g. deceased participants, non-responders, previously testing, mismatches, and reinterpretations), and providing help for ROR in low resource settings (Meharry and Mayo Mountain Park Health Center).
  + The workgroup’s recommendations for eMERGE IV are to develop efficient and scalable means for ROR including innovations such as chatbots, specialized nurses, and videos, use existing ROR protocols for non-responders and deceased participants, and include provider education and engagement and familial communication.
* **EHRI workgroup eMERGE III final deliverables | Sandy Aronson (Partners/Broad) & Casey Overby Taylor (JHU/Geisinger)**
  + The EHRI Workgroup charter had three key areas of interest, which were engineering, science, and community.
    - The EHRI Workgroup engineering goal was to establish, document, and seek to continuously improve process flows for delivery of eMERGE reports and data.
    - The EHRI Workgroup goal for science was to experiment with innovative approaches that go beyond core requirements and to evaluate their effectiveness.
    - The EHRI Workgroup's goal for the community was to liaise with other groups, engage collaborative projects, and disseminate learning and best practices.
  + EHRI accomplishments
    - Milestone 3: Improve and/or standardize genomic clinical decision support (CDS) for the return of clinically relevant genetic or incidental results directly to physicians.
      * The EHRI Workgroup proved that it is possible to establish and operate a multi-lab to a multi-site network capable of electronically transmitting structured genetic results intended for clinical use.
      * The group was also able to develop an eMERGE XML result transfer format.
        + The goal was for the LMM at Baylor to not only share information but ingesting information as knowledge into common data points.
        + Each site had its own set of IT systems that had evolved based on each process, and the goal was to harmonize information across sites. For this to be accomplished, clinical and IT teams at sites worked together and set up a common interface format.
    - The process of getting files from the labs to the providers in a clinically validated form is also challenging as security must be managed, and the question of data custody had to be addressed throughout the process.
    - The sites have to digest the data and then stand up clinical decision support to meet their ultimate objectives.
  + Eight sites have reported that XML reports are being ingested into the EHR ecosystem.
  + Twelve sites have reported that the EHR is being used for the return of genomic results.
  + Two sites have reported that results are captured in the EHR. Four sites have reported that structured results are captured in Ancillary' omic only. Four sites reported that structured results are being captured in both systems. Two sites report that structured results are not being captured in either system.
  + There are eight sites using the XML format. The EHR is being used for all twelve sites being surveyed for the return of genetic results.
  + The workgroup has worked on two network-wide projects, and several lessons learned panels.
  + The group has also disseminated several open-sourced specifications and supporting code in GitHub.
  + Specifications include results-schema and FHIR specification. In terms of support/utility, there has been GeneInsight-XML-Parser, GeneInsight-XSLT, FHIR-implementation, FHIRGenomicsImporter, and FHIRGenomicsProxy.
  + Remaining work
    - For milestone 3, several manuscripts are still in progress that are scheduled to be submitted in June 2020.
      * [NT213: Info button Genomic Medicine Initiatives Survey](https://emerge-network.org/wp-content/uploads/2017/02/NT213-Overby-Genomic-Medicine-Initatives-Survey.doc)
      * [NT319: Use of Info buttons to Find Answers to Clinician's Questions in Clinical Genomics](https://emerge-network.org/wp-content/uploads/2019/01/NT319-Watkins-Use-of-infobuttons-to-find-answers-to-clinicans-questions-in-clinical-genomics.docx)
      * [NT380: Genomic Considerations for FHIR: Lessons from the eMERGE Implementation](https://emerge-network.org/wp-content/uploads/2020/03/NT380-Supplemental-analysis-to-%E2%80%98Genome-wide-Modeling-of-Polygenic-Risk-Score.docx)
    - For milestone 7, there is an ACI Open case series in progress listed under [NT352](https://emerge-network.org/wp-content/uploads/2020/01/NT352-Taylor-Lessons-from-eMERGE-on-readiness-for-genomic-cot-tracked-REVISED-1.docx).
      * There are 5 case reports from 5 eMERGE sites that have been submitted and are under review.
  + Beyond eMERGE III: Future Considerations
    - Vendors are adopting FHIR, and the approaches that are being explored in leveraging FHIR could be used in the future.
    - Consensus on variant representation is needed in future work.
    - If FHIR is used in phase IV, participation in the clinical genomics group is needed for HL7.
    - There is some diversity in how servers are implemented so that reports from Baylor can be loaded and accepted. There needs to be a plan for diversity in technology.
  + The Ancillary Omics system model was the main model that was used for integration. Other models can be used, such as SMART on FHIR.
  + Lessons learned
    - Clinical-grade cross-institutional interface projects require significant effort.
    - Clinical and laboratory sites are extremely heterogeneous (data production, format, and structure of reports, interpretation standards, exchange of data between systems, how it is stored in the EHR, how it is consumed and used for CDSS).
    - IT processes and clinical processes are symbiotic (when you are designing IT processes, you are also designing clinical processes).
    - Cross-site differences need to be assessed early to get networks off the ground.
    - Some genomics functionality may not be practical to implement in real-world clinical systems.
    - Forums to share lessons learned as implementations occur can be very helpful.
  + Recommendations for eMERGE IV
    - The network should start early on designing network topology and inter-institutional interfaces.
    - It is important to recognize that cloud infrastructure and traditional infrastructure are different - start early on designing for this as well.
    - The network should start early working on FHIR standard-based germline variant communication.
    - It is important to start early on engaging cross-industry stakeholders who will influence the rate of FHIR adoption.
    - The network should start early on thinking about how eMERGE internal network/app development could be used to encourage commercial players to leverage standard interfaces.
  + The EHRI workgroup established a result transmission network that was used to develop standards and enable a clinical decision support system (CDSS).
* **PGx workgroup eMERGE III final deliverables | Cindy Prows (CCHMC) & Laura Rasmussen-Torvik (Northwestern)**
  + Accomplishments
    - The PGx workgroup tracked PGx return during eMERGE III.
    - The workgroup established a relationship with CPIC, shared lessons learned at the SC meetings, and learned what the consortium's need was, which was to generate additional association information to be used in guideline statements.
    - The workgroup was able to explore potential collaborations with IGNITE.
    - The workgroup also examined the feasibility of additional eMERGE III PGx outcomes or SPHINx PGx projects
    - The workgroup determined that more data was needed in SPHINx in regards to the medications and the timing of medications.
    - The workgroup was able to successfully contribute lessons learned in combination with other workgroups.
    - The workgroup has been tracking the PGx, ROR, and there have been four PGx ROR returned to the EHR with reports containing information about star allele phenotypes and predicted phenotypes. Geisinger and Vanderbilt have completed the PGx ROR.
  + Advocacy for eMERGE IV
    - The PGx workgroup reports that decision making has to happen in the beginning.
    - PGx phenotypes also need to be prioritized early.
    - Outcomes from PGx need to be looked at for the extraction of drug data, including dosage and timing.
* **Genomics workgroup eMERGE III final deliverables | David Crosslin (KP/UW), Megan Roy-Puckelwartz (Northwestern), Patrick Sleiman (CHOP)**
  + Genomics workgroup accomplishments includes the publishing of details of the genotyping and imputation of 84,000 eMERGE subjects led by Ian Stanaway, guidance to the eMERGE CC regarding genetic data activities in order to produce four large multiple use discovery-based datasets, supplement-funded efforts to link Geocoding data to eMERGE participants led by Patrick Sleiman, establishment of focus groups for SPHINX PheWAS data integration, assistance to VUMC with implementation of alternate methods of assessing penetrance and risk scores using network wide data, imputation of HLA region in 105,000, and active contribution to lessons learned manuscript (NT357).
  + The Genomics workgroup has made an impact on the field by generating one of the largest genetic datasets with moderate diversity linked to EHR for deep phenotyping, development of common phenotypes to assist in covariate adjustment in regression models, and understanding the framework of discovery in a data set enriched for European-ancestry while striving for more diversity at a network level.
  + The group will continue efforts to conduct PheWAS on Structure Variant/Copy Number Variant datasets.
  + UW plans to submit a R01 to impute X chromosome in the eMERGE cohort. The group has submitted funding for the Electronic Medical Records and Genomics Toolkit (eMERGENT) and hopes to hear back soon. There is also future work planned for the imputation of the killer-cell immunoglobulin-like receptor (KIR) region and continuing the assessment of this and HLA into models.
* **Clinical Annotation workgroup eMERGE III final deliverables | Gail Jarvik (KP/UW) & Heidi Rehm (Partners/Broad)**
  + Clinical Annotation milestone 1 was to conduct penetrance analysis in conditions with sufficient data and assess the impact on clinical outcomes; describe lessons learned.
  + The group was able to:
    - Develop outcomes forms for ten disease areas, and there was insufficient data for four disease areas.
    - There was analysis performed for six disease areas with sufficient data, including arrhythmias, aortic disease, cardiomyopathy, FH, breast cancer, and colon cancer.
  + Remaining Work:
    - The Clinical annotation workgroup needs to finalize penetrance work on top conditions.
    - The work on penetrance publications is still ongoing.
  + Significant Accomplishments from Phase III
    - The group was able to apply the ClinGen approach to gene-disease validity assessment to all genes on the eMERGE gene panel (including SNP genes), defining each associated condition and the strength of evidence.
    - The Clinical Annotation workgroup also developed consistency in variant interpretation approaches.
    - The group developed a consensus on the most common clinically reportable variants in the eMERGE panel and whether to recommend a return to patients based on actionability assessments The group also developed a re-analysis approaches to support report updating and to determine the utility of studying VUSs that favor pathogenicity.
    - Lastly, the workgroup was able to explore the rate of secondary findings (SFs) in the eMERGE cohort and the penetrance of phenotypes associated with those findings.
  + Clinical Annotation Workgroup Influences
    - The workgroup was able to:
      * Develop a better understanding of the rates of SFs in biobank populations. This helps research programs estimate resources required for adding SFs return to genomic sequencing programs and informs consideration of screening.
      * Develop a pre-reporting interlaboratory harmonization approach to improve consistency in interpretation across clinical laboratories. This approach is now being applied in the All of Us Research program.
    - The Clinical Annotation Workgroup choices of additional non-ACMG IF to return were considered by others consortia and by ACMG.
* **Overall lessons learned from eMERGE III and recommendations for eMERGE IV**
  + The question raised early on in eMERGE is whether or not biobank samples linked to the EHR could obtain meaningful data. The network has demonstrated that this is possible.
  + Data in the EHR is valuable for doing research, and eMERGE has proven this to be correct.
  + eMERGE has also shown how this data can be reused in significant ways.
  + The evolution of emerge
    - The eMERGE network since 2007 has been able to establish discovery-based GWAS/PheWAS analyses with biobank samples & EHR records as well as develop sophisticated electronic phenotyping of clinical disease.
    - The eMERGE network has new large scale data sharing methods that are available across sites and phases.
    - There have been clinical applications of genomics developed.
    - The network has completed recruitment and return of PGx & ACMG sequencing results.
    - The network has successfully integrated genetics in EHRs and clinical practice.
    - The network has been able to influence genomic guidelines and their impact on implementation.
  + Leveraging EHRs for genomic research
    - The sites have been able to enroll and return results for genomic testing of large cohorts.
    - There has been the development of novel, complex algorithms for disease prediction.
    - The network has been able to integrate family history and clinical data to assess the overall risk of disease.
    - The network has been able to return results, integrate them, and use them for clinical implementation.
  + At the end of eMERGE III, the network determined how to return results to participants and integrate them in EHR, and use those results for clinical decision support (CDS).