**Summary of Steering Committee Meeting: September 2024**

September 25 - September 26, Zoom & In-Person

|  |  |
| --- | --- |
| **eMERGE Day 1: Wednesday, September 25, 2024** | |
| **Time** | **Event** |
| 9:00-9:20 AM | NHGRI Program Official Report | Robb Rowley (NIH/NHGRI) |
| 9:20-9:35 AM | Announcements, opening remarks | Rex Chisholm (SC Chair, Northwestern) |
| 9:35-10:35 AM | Publication panel | Josh Peterson (CC) |
| 10:55-11:40 AM | Measuring success: Primary and Secondary hypotheses | Dave Veenstra (UW) & Nita Limdi (UAB) |
| 11:40 AM-12:00 PM | Health Care Provider survey preliminary analysis | Ingrid Holm (BCH) & Georgia Wiesner (VUMC) |
| 12:40-1:40 PM | Workgroup breakout session one   * Outcomes (Provider Uptake & Outcomes, Phenotyping, & QA/QC Taskforce) * ELSI |
| 1:40-2:00 PM | Final Clinical Operations Update | Katie Larkin (Broad) & Ed Esplin (Invitae) |
| 2:00-3:00 PM | Workgroup breakout session two   * Return Close Out (CARE) * Data Utilization (GRID) |
| 3:30-4:30 PM | Panel on GIRA return update and progress | Hana Bangash (Mayo) |
| 4:30-4:35 PM | Closing remarks | Rex Chisholm (SC Chair, Northwestern) |

|  |  |
| --- | --- |
| **eMERGE Day 2: Thursday, September 26, 2024** | |
| **Time** | **Event** |
| 9:00 - 9:15 AM | Opening remarks & comments from ESP chair | Robb Rowley (NIH/NHGRI) & Dan Rader (University of Pennsylvania) |
| 9:15 - 9:35 AM | Network overview: Priorities, goals, progress and ESP recommendations | Rex Chisholm (SC Chair, Northwestern) |
| 9:35 - 9:55 AM | Network Manuscript Updates| Josh Peterson (CC) |
| 9:55 - 10:15 AM | Comprehensive Risk Assessment & Return (CARE) | Gail Jarvik (UW), Iftikhar Kullo (Mayo), & Cindy Prows (CCHMC) |
| 10:15 - 10:40 AM | Provider Uptake & Outcomes | Nita Limdi (UAB) & Dave Veenstra (UW) |
| 11:00 - 11:25 AM | Phenotyping and Outcomes data collection |Shawn Murphy (MGB) & Wei-Qi Wei (VUMC) |
| 11:25 - 11:40 AM | QA/QC Task Force | Jennifer Pacheco (NU) & Lisa Martin (CCHMC) |
| 12:25 - 12:45 PM | Genomic Risk Innovation and Discovery (GRID) | Adam Gordon (NU) & Matt Lebo (MGB) |
| 12:45 - 1:00 PM | CIRT | Emma Perez (MGB) & Bob Freimuth (Mayo) |
| 1:30 - 2:00 PM | Input/Feedback from the ESP, general discussion |
| 2:00 - 2:05 PM | Closing remarks | Rex Chisholm (SC Chair, Northwestern) |

**eMERGE Day 1: Wednesday, September 25, 2024 | Slides**

1. **NHGRI Program Official Report | Robb Rowley (NIH/NHGRI) | Slides**
   1. Accomplishments
      1. We have overcome several challenges including validating PRS in diverse populations and capturing data form 10 different sites.
      2. We have recruited over 25,000 participants and returned results to 19,600 participants.
      3. Today’s goal is to identify gaps and define analysis plans including Network publications.
   2. New Staff Changes were shared.
      1. Program Analyst: Jessica Reinach, B.S.
      2. Extramural Program Director, Division of Genomic Medicine: Mollie Minear, Ph.D.
      3. Program Management Fellows: Rachel Nusbaum, M.S., C.G.C. and Nicole Thompson, M.S. L.G.C.
      4. Information was shared about the NIH-ACMG Fellowship in Genomic Medicine Program Management. Due Date: December 6, 2024
   3. There are several funding announcements:
      1. Advancing Genomic Medicine Research (AGMR): Applications on implementation of genomic information and technologies in clinical care.
      2. Population Genomic Screening in Primary Care: Implementation and evidence generation pilot program of population genomic screening for common, actionable genomic conditions predominantly in the primary care setting.
      3. Small Business Program: NHGRI encourages researchers who are interested in commercializing their science to visit the Small Business Program webpage to learn more.
      4. Other funding opportunities: www.genome.gov/research-funding/Funding-Opportunities
      5. The NIH is issuing an RFI on Re-envisioning Postdoctoral Research Training.
      6. NIH Cloud Lab provides eligible researchers with credits for cloud services and access to curated bioinformatics tutorials and support.
      7. The Genomic Medicine Meeting will be held on Dec 12-13, 2024. The goal of the meeting is to identify needs, opportunities, and challenges for applying a patient’s genomic information in the diagnosis, prevention, and treatment of infectious diseases.
   4. Future of eMERGE
      1. One-Year Extension, FY25
         1. Submit your draft budget by October 15, 2024.
         2. Carryover requests can be submitted as needed.
         3. NHGRI will review all budget requests and provide final feedback by January 15, 2025.
         4. Instruct your AORs NOT to automatically extend your grants.
      2. eMERGE Four-Year Limited Competition and Single Source
         1. Timeline would be in winter 2025 and due spring 2025
         2. Limited Competition and Single Source only limits who can apply but does not guarantee funding – MUST pass peer review
2. **Announcements, opening remarks | Rex Chisholm (SC Chair, Northwestern) | Slides**
   1. Goals for the meeting include: Work on Network outcomes and developing publications.
   2. Recruitment has exceeded the goal of 25,000 enrollees.
   3. Return goal of 76.5% currently completed.
   4. Data utilization tools, i2B2 and Tanagra, have been demonstrated to the Network.
   5. Sites should not let any preliminary analysis influence or alter plans already established.
   6. Lessons learned from EHR data pulls should be shared across groups.
   7. Preliminary post-RoR survey data shows 59% of adults have completed the survey and 21% pediatric participants completed their version.
      1. 59% of the adult surveys were completed. Of these, almost 59% of completed surveys were not at high risk.
      2. 21% of pediatric surveys were completed.
   8. Preliminary EHR outcomes data pull
      1. Prevalence of health service outcomes after GIRA return
         1. 17.48% breast MRIs for high risk had breast MRIs.
         2. 42.49% prostate testing for high risk had screening completed.
         3. 45.67% high risk tested for type 2 diabetes had A1cs tested.
         4. 18.9% high risk tested for type 1 diabetes (peds) were tested for T1D.
         5. 33.3% high risk tested for type 2 diabetes (peds) were tested for T2D.
      2. Sites should focus on obtaining 100% of surveys completed now that recruitment is over.
   9. Network goals for year 5.
      1. Complete returns by the end of October 2024.
      2. Complete assessment and QC of outcomes data/variables needed by end of October 2024.
      3. Finalize data dictionary for EHR outcomes, including ‘order’ vs. ‘resulted’ values by end of December 2024.
      4. Interim data refresh in OMOP of EHR outcomes data by the end of February 2025.
      5. Final multisample VCF and BAMs on AnVIL by the end of February 2025.
      6. Execute demonstration projects on AnVIL by February 2025.
      7. Currently, 19,134 GIRA returns have been completed.
         1. An estimate of more than 21,000 completed returns using current projections.
      8. eMERGE will need to be registered in dbGaP by the first of 2025.
      9. Final genomic information and EHR data uploaded to AnVIL/dbGaP but the end of February 2026.
   10. Upcoming challenges include:
       1. Data will need to be censored if all returns are not completed by the end of October.
       2. Final EHR outcomes data dictionary needs to be approved by the end of 2024.
       3. Data harmonization for AnVIL will be a challenge, MeTree data harmonization for example.
       4. Waiver for consent for participants that turn 18 after study tasks are completed might be difficult because of the need to look at genetic data.
          1. A loss of pediatric outcomes data will likely occur if the waiver is denied.
          2. A shortened consent form might be possible if the waiver is denied but there is a chance that as loss of pediatric data may occur.
       5. AnVIL data are in one workspace for eMERGE seq, GIRA data, GWAS and sites will be asked to set up their own workspaces that point to these data.
3. **Publication panel | Josh Peterson (CC) |Slides** 
   1. Workgroups presented their in progress and planned publications.
   2. CARE
      1. Return of Genome Informed Risk Assessment for Common Conditions to 2X,XXX Adults and Children - NT510, led by Lucinda Lawson, Cindy Prows (CCHMC), Gail Jarvik (UW), Iftikhar Kullo (Mayo)
         1. Goal: To assess the prevalence of high genomic risk for 11 common conditions in a diverse cohort and implications for RoR in health care systems including infrastructure, effort and workforce needed.
         2. Knowledge gap: The prevalence of high genomic risk informed by polygenic risk, family history, Mendelian, or integrated risk for common conditions in a diverse population and the effort needed for returning genomic risk results is unknown.
         3. Scope: Prevalence of high genomic risk; GIRA results returned; factors associated with success/failure rates of conducting “high touch” returns for high risk GIRAs due to high PRS, P/LP monogenic variant, high integrated score; lessons learned and workforce implications.
         4. Timeline: A draft is expected in March 2025, with manuscript submission in April 2025.
         5. Data will include GIRAs returned (not high, high, carrier for AR conditions); high disease risks and sources of high risk; high risk return methods; demographics; sites’ high risk return sequence; #/type incidental results; personnel involved in RoR at various sites.
         6. The manuscript will include all phenotypes and all ages (adult and pediatric).
         7. Dependencies include the completion of the RoR dataset and several papers that are in progress that will include information relevant to this manuscript (Study design manuscript, GIRA development manuscript, Integration of GIRA into the EHR manuscript).
            1. If the relevant manuscripts are not yet complete, the group will rely on heavy use of tables and supplemental materials for detailed ‘Methods’ descriptions rather than citing the relevant papers.
         8. There are no expected barriers to publication.
      2. Additional CARE Manuscripts
         1. NT426 Genetic counselors experiences, perspectives returning PRS results (Sabrina Suckiel & Noura Abul-Husn)
         2. NT483 Complications of covering preventive care costs in research studies (Laura Rasmussen-Torvik)
         3. NT507 Challenging RoR cases (Matt Lebo)
         4. NT511 Parent perceptions of learning their children’s GIRA results (Jasmine Purcel & John Lynch)
   3. ELSI
      1. The ELSI group has been working on the sharing of racial/ethnic data (led by Anna Lewis & Maya Sabatello).
         1. Maya reported to the PI Workgroup on “Decision points for sharing race and/or ethnicity info” with recommendations to ensure ancestry accuracy and protect privacy for participants when cells are small.
      2. The group is also working on an addendum to the eMERGE Publication policy, including discussion regarding data sharing.
      3. ELSI Manuscripts in Progress
         1. Medical care costs (NT483)
         2. Data sharing tradeoffs (NT499)
      4. Planned ELSI- focused Manuscripts
         1. Missingness in survey data (with other workgroups)
         2. Genetic knowledge and expectations pre- and post-RoR
         3. Healthy behaviors pre- and post-RoR
         4. RoR experiences, including how participants want to receive results and barriers
         5. Peds vs Adults (cohort differences)
         6. Transgender participants
         7. Disability workgroup
      5. Concept Sheets in Preparation
         1. eMERGE-4 Participants’ Views on Return of Results (Ingrid Holm & Maya Sabatello)
            1. This manuscript will explore the changes in expectations and understanding of genetic testing and results pre-and post-RoR and will involve collaboration with the Outcomes Workgroup.
         2. Behavioral impacts of GIRA results: Experiences of eMERGE-4 Participants (Ingrid Holm & Maya Sabatello)
            1. This manuscript will study the impact of RoR on behavioral changes and will involve collaboration with the Outcomes and CARE Workgroups.
         3. Participants with disabilities - RoR cases (Maya Sabatello)
            1. This manuscript will be a collaboration between the RoR Roundtable group and the ELSI group.
      6. There will be discussion as to whether behavioral outcomes should be divided into pediatric and adult groups. Behavioral changes from an ELSI perspective focus on whether participants make changes, not whether a test or procedure was performed.
   4. Provider Uptake and Outcomes
      1. Study design and analysis framework for evaluating the impact of genome informed risk assessments for 10 diseases in clinical care: The Electronic Medical Records and Genomics (eMERGE) study - NT466.1
         1. Goal: This is a study design manuscript that will highlight how we recruited and total numbers, including projected numbers of high risk by disease. This is where the RoR will take over. Subsequent manuscripts will refer to these two manuscripts.
         2. Knowledge gap: How to evaluate impact of genome informed risk assessments on provider and patient behavior?
         3. Scope: This paper will describe the participant recruitment, follow-up, outcomes, and analysis framework.
         4. Timeline: A first draft is expected by the end of 2024 with submission in spring 2025.
         5. Data will include overall participants screened and enrolled in the eMERGE IV cohort.
         6. All phenotypes and all age ranges will be included.
         7. There are no expected dependencies or barriers to publication.
      2. The impact of genome informed risk assessments for 10 diseases in clinical care: The Electronic Medical Records and Genomics (eMERGE) study
         1. There are two manuscripts planned (Adult and Peds) that will be written in parallel. The MCSs have not yet been submitted.
         2. Goal: Report the influence of genomic risk assessment on uptake of medical management recommendations for common chronic diseases in 25,000 primary care patients.
         3. Knowledge gap: Assess the impact of returning genomic risk assessment on provider and patient behavior.
         4. Scope: Compare overall adoption of healthcare recommendations across conditions by providers (order) and patients (completion of order) in high risk (vs. not risk) participants.
         5. Timeline: The plan is to submit the manuscripts in Fall 2025, assuming all of the data freezes happen as planned.
         6. Data will include overall participants receiving GIRA in the e4 cohort.
         7. All phenotypes and all age ranges will be included.
         8. There are no expected dependencies or barriers to publication.
      3. Additional Outcomes manuscripts
         1. Provider post RoR survey and interview related manuscripts in collaboration with ELSI.
         2. Patient post RoR survey related manuscripts.
         3. Condition specific manuscripts, including CHD, CKD, Prostate Cancer, Breast Cancer, T2D.
   5. QA/QC Task Force
      1. Utility of structured EHR data for capturing referrals and encounters: Findings from eMERGE 4 Manual Chart Review (Jen Pacheco & Lisa Martin), MCS to be submitted
         1. Goal: To evaluate the quality of EHR algorithms to extract information about referrals and encounters for outcomes analyses. It is critical to evaluate whether algorithms are capturing the accurate information.
         2. Knowledge gap: While EHR data has been used previously using long term data to capture diagnoses, the use of EHR data to capture short term changes in patient management due to risk identification has not been evaluated.
         3. Scope: To describe the process of manual chart review and describe the agreement between the algorithms, to capture the data defined by the Outcomes workgroup needed for outcomes analyses, and manual chart review for a subset of the participants who have received GIRAs.
         4. Timeline: Manual chart review is anticipated to be completed by Spring 2025. Manuscript anticipated submission date is late Fall 2025.
         5. Data will include the results from the manual chart reviews implemented to evaluate data quality and the metrics about the chart review itself.
         6. Dependencies include manual chart review completion and outcomes description.
         7. No overlap with other manuscripts is anticipated.
      2. Lessons Learned manuscript
         1. Tentative title: “Multisite QA/QC: Successes, Challenges, and Opportunities”
         2. Tentative goal: Focus on processes implemented successfully in eMERGE that are unique to a multi-site study as well as what could be done differently.
         3. Tentative scope: Task force components (scope, purpose, member composition and representation), expectations for multi-site studies, expectations vs reality, cross-workgroup communication and engagement.
         4. There is potential overlap that needs to be discussed with the Phenotyping Workgroup.
      3. Non-random missingness (NT516, Elisabeth Rosenthal)
         1. Goals:
            1. Understand pattern of variable missingness in order to reduce or acknowledge potential bias in studies based on emerge IV data.
            2. Determine what variables may be associated with missingness.
            3. Describe the different methods each site used to fill in missing data or reduce missing data.
         2. Knowledge gap: When missingness is not-at-random, data analysis may be biased and results may be confounded by the missingness, limiting interpretation**.**
         3. Scope: The project will start with variables associated with outcomes. There is potential to look at completeness of data provided.
         4. The manuscript timeline is dependent on survey completion.
         5. Data:
            1. From the CDP: Survey, EHR, family history, withdrawal, loss-to-follow-up such as no sample due to no spit/blood kit return, failed genotyping.
            2. From R4 or from individual sites: number of contacts for each survey, failures at Invitae.
         6. Adult and pediatric populations will be included.
         7. The QA/QC Taskforce is discussing potential overlap with the Disabilities group manuscript and with the writing group for NT509.
   6. Clinical Implementation and Return Technologies (CIRT)
      1. Integrating Polygenic Risk Scores into the Electronic Health Record: An eMERGE Network Perspective (Bob Freimuth &Nephi Walton)
         1. This is a perspective piece and is not using eMERGE data.
         2. Goal: To describe what makes polygenic risk scores unique when considering integration of structured results into the electronic health record to inform clinical care.
         3. Knowledge gap: There are no existing papers about standards for clinical PRS broadly.
         4. Scope: To outline required elements for successful integration of PRS to best facilitate personalized medicine at the point of care, including regulatory, technical and clinical workflows. It will focus on the necessary components to contextualize and integrate risk into the EHR.
         5. Timeline: The first draft is expected by November 2024 with submission in December 2024.
         6. There are no expected dependencies or barriers to publication.
      2. eMERGE IV: Experience integrating genomic-based risk assessment reports into the medical record (Emma Perez & Bob Freimuth)
         1. Goal: To describe barriers to EHR precision medicine implementation studies to inform governing bodies.
         2. Knowledge gap: There is limited real-world experience of integrating clinical grade results into the EHR in a research setting.
         3. Scope: The manuscript will involve review of eMERGE workflows by utilizing existing eMERGE pubs, outline site workflows and highlight similarities and differences focused on integration processes, provider notification, and CDS.
         4. Timeline:
            1. Sites have been contributing to a Google Sheet outlining EHR workflow. Sites should continue to keep this up to date.
            2. The first draft is expected by December 2024 with submission in January 2025.
         5. There are no expected dependencies or barriers to publication.
            1. There is potential overlap with CARE and Outcomes manuscripts but it is expected that manuscripts will be able to be kept distinct.
   7. Education and RoR Subgroup
      1. Evaluating monogenic risk for common chronic conditions in eMERGE IV (Jessica Denton)
         1. A draft concept sheet has been circulated to the RoR Subgroup for input.
         2. Goals:
            1. Report on the prevalence of P/LP variants within the eMERGE IV cohort.
            2. Analyze the distribution of monogenic results by demographics such as self-reported race.
            3. Assess the concordance between monogenic results and patient-reported family health history and personal medical history.
      2. RoR Cases Potential Manuscripts
         1. Edge Cases - Challenging and interesting cases in returning Genetic Informed Risk Assessments to patients (Matt Lebo, Emily Miller, Emma Perez & Beth Karlson)
            1. The writing group is developing a REDCap instrument for sites to report these edge cases, as well as other interesting or unique cases that could contribute to additional case series.
         2. Pediatric cases (John Connolly & Shannon Terek)
         3. Participants with disabilities (Maya Sabatello)
   8. Phenotyping
      1. Use of EHR to capture outcomes in eMERGE (NT504, Kavi Wagholikar)
         * 1. The use of the EHR to capture the outcomes of an interventional study is a fairly novel concept and our experience implementing EHR-based outcomes is valuable.
      2. Utilizing Large Language Models to Facilitate Manual Chart Review Across Medical Centers (NT490, Wei-Qi Wei)
         1. It is known that manual chart review is difficult and time-consuming.
         2. This manuscript explores leveraging large language models (LLMs) to improve the efficiency of manual chart review.
4. **Measuring success: Primary and Secondary hypotheses | Dave Veenstra (UW) & Nita Limdi (UAB)**
   1. The cross condition analysis will include 6 and 12 month analyses of provider and patient actions, intervention versus control comparison groups, and provider and patient outcomes.
   2. Adult and pediatrics cross-condition outcomes will have separate analysis and papers.
      1. There is some question about keeping these separate, especially because the study was not designed to be solely for pediatrics in which the study would have been different.
      2. An argument for two separate papers is that the pediatric population may be overlooked in such a large paper.
   3. The study objective is to determine if providing a GIRA will influence recommended clinical action taken by the provider and patient.
   4. There will be three control groups: PRS/integrated score with only patients below the high risk threshold, PRS/integrated score plus monogenic with patients not high for genomic risk, and PRS/integrated risk plus family history with patients not high risk.
   5. The data analysis framework will include looking at the entire cohort including prevalent disease and then excluding prevalent disease.
   6. 6,200 patients are expected to be high risk.
   7. Clinical clinician orders will be looked at and compared across high risk and not high risk groups.
   8. In the fall of 2025, there should be 1 year of follow up data on the entire cohort and the 1 year extension will allow for more follow up analyses.
   9. EMR based algorithms to identify prevalent disease are being developed.
   10. The plan is to let the data guide the analyses, especially with regard to the 3 interventions/control groups.
   11. A question to look into is if high prs for breast cancer also enhances risk for ovarian cancer.
   12. It may be helpful to revisit the outcomes of interest to simplify/narrow down multiple outcomes for one condition.
       1. It would be interesting to compare uptake of multiple outcomes for one condition as well.
   13. For the PRS analysis, monogenics will not be included (other than for breast cancer which uses an integrated score).
   14. The RDD design will be used for each condition looking at the primary outcome. An analysis not using the RDD design to compare all high PRS versus not high PRS as well being that it creates pseudo randomization.
5. **Health Care Provider survey preliminary analysis | Ingrid Holm (BCH) & Georgia Wiesner (VUMC) | Slides**
   1. The goal of the HCP survey and interview is to gather opinions on the GIRA from providers.
   2. The interview will focus on different topics than the survey.
   3. Recruitment will conclude by November 30, 2024.
   4. Response completion date will be December 31, 2024.
   5. There have been 812 survey invitations sent with 151 responses so far.
      1. There were 40% of the respondents that were high risk for more than one condition.
   6. More than 80% of providers remembered receiving a GIRA and more than 90% reviewed the GIRA.
   7. HCPs were not confident in their PRS knowledge but felt somewhat comfortable making decisions based on the report.
   8. There was some trauma experienced from the sIRB on the HCP survey which caused some early HCPs to not receive the survey.
   9. HCP interviews will be conducted for HCPs who received one or more adult GIRAs.
   10. The HCP interview was developed to catch more information that the survey missed.
6. **Workgroup breakout session one**
   1. **Outcomes (Provider Uptake & Outcomes, Phenotyping, & QA/QC Taskforce)**
      1. eMERGE Instrument QC | Elisabeth Rosenthal (UW)
         1. The QA/QC group has been looking for missingness in variables that were flagged as needed for outcomes analysis by the phenotyping leads looking at people with a high risk GIRA.
         2. There are some implausible values in variables used for BMI. How the data was extracted through the API made a small difference. There is less missingness in the height and the weight variables than in the BMI calculation.
         3. Only 1 phenotype requested personal health history data - prostate cancer. There were 140 female participants who checked the box for having prostate cancer.
         4. There are some improbable and inconsistent values in vigorous activity variables. It is recommended to make a copy of days, hours, and minutes to create a new VAC and VAC\_UPDATE following a set of rules. There should also be a cutoff for hours and minutes per day.
         5. The order in which the questions are answered can cause issues. For example, some participants who reported vigorous activity have missing calculated VACs when they should have one and vice versa. It is recommended that depending on what people clicked on, the VAC should have recalculated VACs.
         6. The smoking related variables have outliers and implausible values. A recommendation is to remove these values and create an age cutoff (some have begun smoking at age 1).
      2. Logistics of Manual Chart Review
         1. There is an interim data pull for 6 months coming up in February 2025 followed by a dbGaP submission.
         2. Sites will be doing manual chart reviews for the referral and order data that will be pulled.
      3. Referral/Order Data Plans
         1. The phenotyping workgroup has been working on lab orders versus lab results and how to pull those from site EHRs.
         2. There are no standard codes that represent orders across hospitals which makes this collection challenging.
         3. The group has been sharing what order data looks like at different sites.
         4. All of the text descriptors to find this data would result in a very large data dictionary, which may end up being the move but instead the group is coming up with a keyword list so sites can find this data within notes.
         5. The list of outcomes that need order data is being compiled in this spreadsheet.
         6. Josh Cortopassi frim UAB did a deeper dive on how to pull orders from the EHR. Slides on what that looked like can be found here. It is encouraged for sites to look at the order data to make sure dates correspond with GIRA generation/distribution.
   2. **ELSI**
      1. There are several concept sheets in preparation.
         1. Participant views on RoR
         2. Behavioral impacts of GIRA results: Experiences of eMERGE Participants.
         3. Participants with disabilities - RoR cases
      2. Participant views of RoR MCS is very similar to a paper in development from Mt Sinai.
      3. Thought should be given to whether or not ELSI papers should be included with larger manuscripts or separate ELSI papers.
         1. Discussion settled on ELSI should be involved in larger papers with some thought on smaller stand alone ELSI papers to be considered, provider/participant perceptions on surveys for example.
      4. Condition leads will likely be writing about behavior impacts in their condition specific papers, which may conflict with an ELSI paper (MCS #2).
      5. A paper that contains behavioral aspects found in the post-RoR survey might be a stand alone paper.
      6. Factors that predict the behavioral actions that participants take could be looked at as outcomes in a manuscript.
         1. Challenges and interpreting results that were received from participants could have strong interest.
         2. Items that suggest some level of individual challenge in understanding the results could be used.
         3. Participants that reported high levels of confidence in interpreting results prior to receiving them, then reporting a lower level of confidence after receiving results is another possible topic to consider.
      7. An edge case paper is being thought through currently.
         1. Disabilities is one subtopic that had some discussion.
         2. Transgender individuals is another potential topic.
      8. Writing groups could be formed to not only tackle the larger papers, but to also start thinking through smaller and more specific papers, such as how to write about monogenic results and race/ethnicity.
7. **Final Clinical Operations Update | Katie Larkin (Broad) & Ed Esplin (Invitae)**
   1. Broad shared a status report.
      1. All samples have been received.
         1. 23,070 reports have been issued with 1,022 samples pending reports. Approximately 50 will be new fails/TNPs.
         2. Projected completion date is October 15.
         3. 21,326 samples are currently staged in AnVIL.
         4. 13,502 blood samples and 11,520 saliva samples have been processed.
         5. Reach out to Broad if site numbers of samples appear incorrect.
      2. PRS generation status was shared.
         1. 143,501 PRS have been generated for 22,101 individuals.
         2. 5,116 High risk results have been returned to 4,564 individuals.
      3. Next steps were shared for Broad.
         1. Broad will work on tying out all PRS samples.
         2. They will upload remaining Array data and Invitae data to AnVIL.
         3. A few eMERGE participants have requested genotyping data and Broad is looking into this request.
         4. Broad is moving from clinical GDA array to the Blended Genome Exome platform.
   2. Invitae shared a status report.
      1. 20,095 orders have been placed and 19,606 reports have been released
         1. Reach out to Invitae if site numbers of samples appear incorrect.
      2. Results were shared by overall status and by each gene.
         1. 462 results were positive (2.36%) and 19,121 (97.52%) were negative with a few carrier (n=14) and indeterminate (n=9) results.
         2. The most prevalent finding was LDLR (n=112) and BRCA2 (n=102).
         3. For TP53 results, some may be mosaic results that require confirmation.
         4. By clinical area, the breakdown of monogenic results shows risk as dyslipidemia (45%) Breast/Prostate Cancer (46%), Colorectal Cancer (19%).
         5. Failure rates were shared and were relatively low (n=301 1.51%)
      3. Invitae shared that they have information about self-reported ancestry and monogenic results from other patients who have had a panel test.
8. **Workgroup breakout session two**
   1. **Return Close Out (CARE)**
      1. The group reviewed manuscript progress and plans.
         1. Network CARE manuscript: This manuscript is focused on returns across the network and includes descriptive data and factors that influence whether or not we are able to do a high risk return as intended.
         2. Sabrina and Nora: Focused on people returning high risk GIRA due to high PRS, information collected via survey.
         3. Edge case manuscripts involve interesting cases without any outcomes data
      2. Planned and ongoing data collection beyond R4 includes:
         1. Survey of study professionals/staff returning high risk GIRA results, high risk due to high PRS/high BOADICEA
         2. Edge cases.
         3. Site return workflow for high risk.
         4. The group should think about whether there is any other data that will be needed for return of results focused manuscripts. Edge case manuscripts might include chart review.
         5. The manuscript process should involve discussion about any new data that will be needed beyond what sites are already collecting, and how that will be obtained.
      3. Manuscripts can be tackled from various lenses. For instance, outcomes can be viewed from a provider action standpoint as well as a behavioral lens.
      4. There is a difference in how we consider High PRS in the framework of an integrated score. This affects how results are returned.
         1. For CHD the integrated score is based on the PCE.
            1. PRS High and PCE Not High = High Touch Return
            2. PRS Not High and PCE High = Low Touch Return
         2. For Breast Cancer the integrated score is based on BOADICEA.
            1. PRS High and BOADICEA Not High = Low Touch Return. However, some sites are still offering HTR in these instances
            2. PRS Not High and BOADICEA High = High Touch Return
      5. The group reviewed the eMERGE Outcomes Lexicon from a CARE perspective.
         1. Incidental Findings vs Secondary Findings
            1. The American College of Medical Genetics (ACMG) has used both terms, but more recently is using Secondary Findings.
            2. Participants and providers often feel that the term Secondary Findings minimizes the findings.
            3. Mosaic and chromosome anomalies would be Incidental Findings.

ACMG has never classified chromosomal changes as secondary findings. They have limited it to sequencing changes.

The only incidental finding so far has been the Kleinfelter syndrome, loss of x chromosome.

* + - 1. The term Challenging Cases will be used instead of Edge Cases.
      2. In defining personnel returning results, the group agreed to use Genetic counselor vs. Non-genetic counselor.
      3. Mode of Return would be placement in the EHR. The mode of communication would be virtual, telephone or in person.
      4. The Date of Return is the placement in the EHR.
      5. Study flow diagrams will be harmonized.
    1. The group discussed GIRA update considerations.
       1. Invitae reports may be reissued due to changes in variant interpretation.
       2. Broad reports will not be changed as PRS will not be recalculated
       3. Invitae and Broad reports are clinical reports and amended clinical reports need to be placed in participant’s EHR.
          1. Participant and participant’s provider should be notified of change
       4. GIRAs are a research product that contain clinical reports as well as many other components/information. They are not accepted at all sites as a clinical report.
       5. Number of amended Invitae reports due to variant reinterpretation expected to be low through April 30, 2025.
          1. Sites should make sure that participants are aware of the Invitae patient portal, allowing them access to changes in their report after the study ends.
       6. In the case of a change in variant interpretation, the updated Invitae report should be placed in the EHR, but GIRAs should not be regenerated.
  1. **Data Utilization (GRID)**
     1. The group reviewed the eMERGE\_Consortium\_GIRA workspace on AnVIL, including the file on family relationships (CDP\_family\_relationships\_v2.csv).
     2. MeTree will be exporting their data and providing it to eMERGE. On the October 14th GRID workgroup call, MeTree representatives will be present to discuss the export. It is unknown where the family relationships file comes from. It is unknown what the ‘auto\_relationship\_link’ column signifies. Some records are ‘TRUE’ and some are ‘NA’.
     3. There are very well established standards for pedigree storage, which should be used for long term storage on AnVIL. The PLINK format was suggested. It was proposed to not include the family relationships file information in a larger analysis file, since a repeating instrument is atypical.
     4. The group recommends converting the CDP\_family\_relationships\_v2.csv file into a standard format for the relationship. They will generate family relationships in a format used standardly in genetic analysis.
     5. It was proposed to run an IBD on the entire dataset using AnVIL.
        1. It was suggested to measure admix estimates.
        2. ACTION ITEM: The group will investigate the availability of a WDL to run IBD or another relatedness measure in AnVIL on the eMERGE 4 cohort.
     6. At MGB, they have two rows for each person, one of which is the relative. For example, one row will be the mother, with the value as their daughter’s eMERGE ID. Another row would be the daughter, with the value as their mother’s eMERGE ID.
     7. The MeTree json files are located at each site and not centralized. There is a lot of information about disease and relatives in MeTree. It will be very interesting to overlay an IBD analysis atop the pedigree. There is a group that uses IBD data to automatically reconstruct pedigrees at a cohort scale. If and when we calculate IBD and admixture, it is important to account for the presence of short term relatedness among participants with similar ancestry.
     8. The format must be able to capture relationships and phenotype. Pedigrees will be differing sizes, as participants could report many or few conditions, with or without ages at onset.
        1. The PED file format was suggested, and a covar file with the phenotypes could link back to the eMERGE ID.
     9. Not everyone with an eMERGE ID will be in the pedigree. Pedigree members will have MeTree IDs, which is how they are represented in each json. It was proposed that the MeTree ID represent the individual ID, and the eMERGE ID be used as a family ID.
     10. Many sites have recruited full families, which is important to capture. We will ask the MeTree representatives if the MeTree IDs are unique. It was proposed to convert the family relationships file on AnVIL to a fam file to achieve a single output. It was proposed to limit the first analysis to participants we have genetic data on.
     11. There may be issues with parents getting confused when completing the surveys for their children.
     12. Consolidating the rescue survey and the MeTree results may be difficult because it may be unknown which one was ultimately used for GIRA generation. We can see when a participant has both results present. When the rescue survey disagreed with MeTree, the high risk result was used to generate GIRA.
     13. MeTree was always used to calculate BOADICEA. Sites handled the rescue survey-to-MeTree for BOADICEA differently. Some sites only entered rescue survey information into MeTree if it would have caused a high risk BOADICEA result.
     14. The group recommends that for the derived analysis dataset, one flag should be used for positive family history, and a separate flag to show if MeTree was present, and a separate flag to show if the rescue survey was present.
     15. It will be important to note in the README file that the FHH data is a scarce dataset. Until we see the batch MeTree data, we do not know how scarce the dataset is. There are areas in the pedigree data that should have QC, however it is unknown the reach and content until we receive the batch MeTree data.
     16. The group does not want to retain the repeating family relationships instrument in the derived analysis ready data set.
     17. The CDP data has been updated to have an “NA” instead of a blank cell. There are still differences where some variables use ‘true’ and others use ‘1’. We should make this consistent.
     18. The group reviewed the CDP DD. There are variables in the DD that do not appear in the CDP. These should be removed. The raw output includes some of the formatting code, which makes it difficult to interpret.
     19. The Invitae results instrument is not a repeating instrument. The results are pushed automatically from Invitae. This may result in a situation where in the CDP, the Invitae results instrument and GIRA review instrument look different from what was returned to the participant. The CARE workgroup will be examining this as well, so it is important to coordinate with them.
     20. The CARE workgroup came to consensus earlier this month that there will be no GIRA re-generation.
     21. The concern is that the CDP data will contain the reclassifications, even if the new GIRA is not generated. To detect the reclassifications, the GIRA timestamp would be earlier than the Invitae timestamp. We should use the timestamps to manually identify the records impacted so far. There will likely be a very small number who have had variant reclassification. We need to determine how to retroactively identify these participants.
         1. It was proposed to turn off Invitae updates being pushed once the last GIRAs are generated.
         2. There is an example of an LDLR reclassification that ClinVar and Invitae were still calling a VUS, but Invitae upgraded it internally.
         3. There are potentially participants with downgraded monogenic results. It was suggested to have a relational table to track the reclassification results.
     22. The CDP should be locked once the last GIRA is generated, aside from the post-RoR survey. As we move forward, we will have to decide what qualifies as ‘locked’. It could be the last person completing the survey or a specific date. Variant reclassifications can change outcomes. For the final cohort, the Invitae result that was present at the time of GIRA generation should be used.
     23. Potential ELSI implications were brought up. Having all participants sign up for the Invitae portal may help with this.
     24. Not all sites have been filling out the family relationships instrument.
         1. It was suggested if the same address or email address is used, the study team can look into those participants and ensure they are linked if they are related.

1. **Panel on GIRA return update and progress | Hana Bangash (Mayo) | Slides**
   1. Each site reported on their GIRA return progress.
   2. CCHMC
      1. CCHMC consented 894 participants (856 active) of their goal of 800.
      2. Of active participants 652 have a sample (545 pediatric, 107 adult). This is 81.5% of the goal.
      3. 650 samples were shipped to the Broad and 104 samples were shipped to Invitae. This is 99.7% of the goal.
         1. Some Broad results are still pending.
      4. 526 results have been returned to participants. This is 81% of the goal.
         1. 454 Peds results have been returned, with 62 being high risk.
         2. 89 Adult results have been returned, with 32 being high risk.
         3. An automated GIRA upload process has been instituted.
      5. CCHMC expects to meet the October RoR deadline.
   3. CHOP
      1. CHOP consented 3519 participants (3426 active) of their goal of 3250.
      2. Of active participants 3199 have a sample (3139 pediatric, 60 adult). This is 98.43% of the goal.
      3. 3133 samples were shipped to the Broad and 59 samples were shipped to Invitae. This is 96.4% of the goal.
         1. Some Broad results are still pending.
      4. 2681 results have been returned to participants. This is 82.5% of the goal.
         1. 2621 Peds results have been returned, with 369 being high risk.
         2. 60 Adult results have been returned, with 12 being high risk.
      5. CHOP expects to return 3125 results by the October RoR deadline.
   4. Columbia
      1. Columbia consented 2889 participants (2831 active) of their goal of 2600.
      2. Of active participants 2601 have a sample (202 pediatric, 2399 adult). This is 100% of the goal.
      3. 2601 samples were shipped to the Broad and 2397 samples were shipped to Invitae. This is 100% of the goal.
      4. 2253 results have been returned to participants. This is 86.7% of the goal.
         1. 179 Peds results have been returned, with 13 being high risk.
         2. 2010 Adult results have been returned, with 661 being high risk.
   5. MGB
      1. MGB consented 3044 participants (2856 active) of their goal of 2500.
      2. Of active participants 2515 have a sample (36 pediatric, 2479 adult). This is 100% of the goal.
      3. 2515 samples were shipped to the Broad and 2447 samples were shipped to Invitae. This is 99% of the goal.
      4. 2200 results have been returned to participants. This is 87.7% of the goal.
         1. No Peds results have been returned so far.
         2. 2200 Adult results have been returned, with 455 being high risk.
   6. Mayo
      1. Mayo consented 2955 participants (2766 active) of their goal of 2750.
      2. Of active participants 2731 have a sample (155 pediatric, 2576 adult). This is 99.3% of the goal.
      3. 2724 samples were shipped to the Broad and 2574 samples were shipped to Invitae. This is 99.9% of the goal.
         1. Mayo is waiting for 238 results from the Broad.
      4. 1780 results have been returned to participants. This is 65.3% of the goal.
         1. No Peds results have been returned so far.
         2. 1780 Adult results have been returned, with 341 being high risk.
   7. Mt. Sinai
      1. Mt. Sinai consented 2801 participants (2621 active) of their goal of 2675.
      2. Of active participants 2621 have a sample (20 pediatric, 2601 adult). This is 98% of the goal.
      3. 2649 samples were shipped to the Broad and 2619 samples were shipped to Invitae. This is 100% of the goal.
      4. 2423 results have been returned to participants. This is 92% of the goal.
         1. 3 Peds results have been returned, with 0 being high risk.
         2. 2420 Adult results have been returned, with 532 being high risk.
   8. Northwestern
      1. Northwestern consented 2797 participants (2492 active) of their goal of 2500.
      2. Of active participants 2466 have a sample (12 pediatric, 2454 adult). This is 98.7% of the goal.
      3. 2455 samples were shipped to the Broad and 2381 samples were shipped to Invitae. This is 99.6% of the goal.
      4. 1772 results have been returned to participants. This is 71.9% of the goal.
         1. 7 Peds results have been returned, with 1 being high risk.
         2. 1765 Adult results have been returned, with 330 being high risk.
   9. UAB
      1. UAB consented 3047 participants (2970 active) of their goal of 2918.
      2. Of active participants 2928 have a sample (56 pediatric, 2872 adult). This is 100.3% of the goal.
      3. 2928 samples were shipped to the Broad and 2872 samples were shipped to Invitae. This is 99.9% of the goal.
      4. 2733 results have been returned to participants. This is 93.3% of the goal.
         1. No Peds results have been returned so far.
         2. 2733 Adult results have been returned, with 669 being high risk.
   10. UW
       1. UW consented 3025 participants (2590 active) of their goal of 2500.
       2. Of active participants 2590 have a sample (146 pediatric, 2443 adult). This is 103.6% of the goal.
       3. 2590 samples were shipped to the Broad and 2443 samples were shipped to Invitae. This is 100% of the goal.
          1. 109 results are still pending from the Broad.
          2. 5 results are still pending from Invitae.
       4. 2211 results have been returned to participants. This is 85.8% of the goal.
          1. 83 Peds results have been returned, with 3 being high risk.
          2. 2128 Adult results have been returned, with 472 being high risk.
       5. The biggest challenge at UW has been getting participants scheduled for return.
   11. VUMC
       1. VUMC consented 2684 participants (2125 active) of their goal of 1898.
       2. Of active participants 2054 have a sample (82 pediatric, 1968 adult). This is 82% of the goal.
       3. 2050 samples were shipped to the Broad and 1949 samples were shipped to Invitae. This is 80% of the goal.
       4. 1607 results have been returned to participants. This is 64% of the goal.
          1. 63 Peds results have been returned.
          2. 1544 Adult results have been returned, with 355 being high risk.
   12. DECISION: All GIRAs should be placed in the EHR by October 31st even if a return appointment has not been scheduled.
2. **Closing remarks | Rex Chisholm (SC Chair, Northwestern)** 
   1. Northwestern has free CME/CNE modules about cardiogenomics in clinical practice.

**eMERGE Day 2: Thursday, September 26, 2024 | Slides**

1. **Opening remarks & comments from ESP chair | Robb Rowley (NIH/NHGRI) & Dan Rader (University of Pennsylvania)**
   1. ESP is impressed with the amount of progress that has been made since April 2024.
2. **Network overview: Priorities, goals, progress and ESP recommendations | Rex Chisholm (SC Chair, Northwestern) | Slides**
   1. Significant progress has been made with recruitment, sample processing, and return of results in the last 6 months.
   2. Total network recruitment breakdown:
      1. Total: 26,849 enrolled
      2. Pediatric participants: 4897 (18%)
      3. Racial & ethnic minorities: 12,864 (48%)
      4. Female at birth: 17,446 (65%)
   3. There are currently a total of 25,047 active participants.
      1. A total of 19,868 samples have been sent to Invitae and 24,257 samples sent to Broad.
      2. Roughly 77% of GIRAs have been returned as of September, 18 2024.
      3. Genomic DNA submitted had a low failure rate with saliva collections having the highest failure rate.
   4. Current GIRA return projections estimate 21,669 returns being completed by the October 31 deadline.
      1. Sites have been repurposing staff to focus on returns now that recruitment is over.
      2. The timing of when the GIRA is uploaded to the EHR has been modified to upload any GIRAs they have to the EHR and not wait for an appointment to be scheduled.
         1. This is anticipated to improve the return curve to allow the network to meet the deadline.
   5. Other accomplishments since April 2024 include the HCP survey starting, AnVIL GIRA data is available to the network, and a Clinical Data Platform has been released to the network for discovery research.
   6. eMERGE will have an extension year to allow for data analysis of EHR data. Final outcomes analysis is planned to be completed by April 2026.
   7. The breakdown of GIRA returns is as follows:
      1. Adult: 15,844 (82.8%)
      2. Children: 3,290 (17.2%)
      3. Female: 12,306 (64.3%)
      4. High risk: 6,462 (33.8%) \*higher than originally anticipated
   8. Monogenic results accounted for 2% of high risk returns while PRS results accounted for 26% of returns.
      1. Family history accounted for the majority of high risk returns.
   9. Current goals to be completed by October 2024 include completing assessment and QC of outcomes data and variables needed and to finalize the network publication framework.
   10. The data dictionary is to be completed by the end of 2024.
   11. eMERGE GIRA EHR data refresh and the final multisample VCF and BAMS on AnVIL are to be completed by February 2025.
   12. The extension year (May 2025-April 2026) will finalize all data refreshes, QA/QC of study data, complete site and network outcomes analysis, and publish network-wide manuscripts.
   13. Some challenges that remain include returning results by the October deadline, finalizing the EHR outcomes data dictionary, and AnVIL data analysis and demonstration projects.
   14. UW had a strong focus on Asian ancestry, which accounted for the majority of those enrolled in eMERGE.
3. **Network Manuscript Updates| Josh Peterson (CC) | Slides** 
   1. The goal is to publish impactful lessons and results from eMERGE to guide genomic test integration into clinical practice.
   2. A Network-wide publication refers to a publication that includes data from more than one site or reports design or conduct of the study across sites white single-site publications include data from a single site.
   3. The eMERGE Publication Hub was developed at the Coordinating Center to track publications across the Network. Concept sheets are authored within the Hub and individuals can sign up to contribute to the publication.
   4. Throughout the course of the Network there has been a transition to more Network-wide publications and fewer single-site publications.
   5. Network publication highlights from the current phase include development of materials, metadata, ELSI perspectives, development of PRS from an ELSI perspective and integration into large scale study, the Marker paper’s description of study details.
   6. There has been an increase in the submission of concept sheets, particularly over this summer.
   7. An internal tracker was developed to assist in the planning of publications.
      1. This tracker identifies planned manuscripts and outlines their scope, data needs, and author leads and helps to avoid overlap as well as identify gaps in scope.
   8. The Publication Framework for Network results analyses involves 3 dimensions:
      1. Phase of Study (Recruitment & Study Design, Genomics and PRS, Return of Results, Data & Phenotyping, Outcomes)
      2. Population (All Participants, Adults, Children)
      3. Condition
   9. Publications, submitted concept sheets, and planned concept sheets were reviewed in terms of Network-wide publications and Phenotype specific publications.
      1. There is more work to be done in regards to Outcomes.
   10. Areas for further discussion as we enter the next phase include:
       1. The timeline for development (concept to MCS and MCS to publication).
       2. The potential to combine or split manuscripts, particularly concerning age group.
       3. Reconciliation of potential areas of overlap and create complementary manuscripts.
       4. Increasing the cadre of authors contributing to manuscripts.
       5. Opportunities afforded by increased data availability
          1. Genotype, sequence, and eMERGE Clinical Data Platform derived data is expected to be available in the first half of 2025.
          2. Core EHR data expected to be available in 2025, with the complete set available in early 2026.
       6. The ESP suggests that the main Outcomes paper should include adults and pediatrics together unless there is good reason not to since the study was designed as a holistic study that includes both adults and pediatrics.
          1. There could be separate papers as well, but there should be one overall, ‘big picture’ paper including both adults and pediatrics.
4. **Comprehensive Risk Assessment & Return (CARE) | Gail Jarvik (UW), Iftikhar Kullo (Mayo), & Cindy Prows (CCHMC) | Slides** 
   1. The CARE workgroup is working on a Network CARE manuscript.
      1. The aims include:
         1. Highlighting prevalence of high genomic risk for 11 common conditions in a diverse cohort.
         2. Demonstrating the logistical needs of returning results in a large cohort, including infrastructure and workforce.
         3. Descriptive statistics of GIRAs returned.
      2. It is important to highlight that individuals would not have been identified if the genetic profiling was not done.
   2. There are two conditions that have integrated scores - Congenital Heart Disease (CHD) and Breast Cancer.
      1. There is a difference in how we consider High PRS in the framework of an integrated score. This affects how results are returned.
      2. For CHD the integrated score is based on the PCE.
         1. PRS High and PCE Not High = High Touch Return.
         2. PRS Not High and PCE High = Low Touch Return.
      3. For Breast Cancer the integrated score is based on BOADICEA.
         1. PRS High and BOADICEA Not High = Low Touch Return. However, some sites are still offering HTR in these instances.
         2. PRS Not High and BOADICEA High = High Touch Return.
   3. The CARE workgroup is involved in creating a lexicon to standardize terms.
      1. Unexpected findings will be called incidental findings.
      2. The term Challenging Cases will be used instead of Edge Cases.
      3. A description of who is returning results will be decided on.
      4. Study flow diagrams will be harmonized.
   4. GIRA update considerations were discussed.
      1. Invitae reports may be reissued due to changes in variant interpretation.
      2. Broad reports will not be changed as PRS will not be recalculated.
      3. Invitae and Broad reports are clinical reports and amended clinical reports need to be placed in participant’s EHR and the provider notified of the change.
      4. The number of amended Invitae reports due to variant reinterpretation expected to be low.
      5. Sites should work to make sure participants are aware of the option to sign up for Invitae patient portal access.
   5. ESP Discussion/Commentary
      1. There is a need to document study challenges in the interest of the research enterprise and ultimately healthcare. It is important to note commonalities as well as differences.
      2. Noting the methods used to get GIRA into the EHR is useful (including the difficulty and uniformity. Within the eMERGE Network there is consistency in where the GIRA is deposited in the EHR across sites. Some sites use a manual process for upload while others use an automated process.
      3. The Network should consider the impact of using a frozen PRS.
5. **Provider Uptake & Outcomes | Nita Limdi (UAB) & Dave Veenstra (UW)** 
   1. The outcomes workgroup has been working across workgroups to finalize plans and groups for cross condition analysis.
   2. Adult and pediatric cross condition outcomes are separate analysis/papers.
   3. The study objective is to determine if providing a GIRA will influence recommended clinical action taken by providers and patents.
   4. Outcomes including provider (orders) and participant (completing a recommended action) actions will be analyzed at 6 and 12 months.
   5. Breast cancer is using an integrated score with a lifetime risk over 25% for monogenic, PRS, personal health history, and family history among other factors.
   6. Returning risk and recommending actions are based on 3 pieces of information including for the high risk integrated score, patients below the high risk threshold in addition to all patients not high for genomic risk, and all patients not high for GIRA (PRS, monogenic, and family history).
   7. The data analysis framework for outcomes analysis will include a regression discontinuity design and logistic regression to adjust for clinical/demographic/monogenic/family history in PRS.
   8. The number of GIRAs expected is around 24,500. Currently 20,089 GIRAs have been generated and 18,8365 have been returned. The observed percentage of high risk GIRA are in line with what was expected.
   9. What is currently needed includes identification of EMR based algorithms for identifying prevalent disease and identification or development/validation of algorithms for identifying incident disease.
   10. The date of the GIRA return is going to be time 0 and orders placed after that EMR will indicate time. For example, if an order was placed 30 days after the GIRA return, the time to order will be 30 days. The day the order was completed will be tracked as well.
6. **Phenotyping and Outcomes data collection |Shawn Murphy (MGB) & Wei-Qi Wei (VUMC)**
   1. The phenotyping workgroup has been focusing on making sure what is being pulled from the EHR is what is needed for analysis. Short term outcomes included process variables reflecting clinicians responses to GIRA/PRS and long term outcomes included reflect alterations in patient care and patient diagnoses.
   2. Based on ESP recommendations, key healthcare actions for each condition were defined by condition groups. Healthcare actions include providers ordering recommendations and patients completing the recommendation.
   3. The group is working on pulling referral data which comes from free text in the EHR. Similarly, order data can be pulled using keywords if codes are not applicable. Manual review of the orders will need to be done to determine keywords to standardize the data pull across sites and also check the data pulled for quality control.
      1. Even where standard ordering codes exist, they are not used which is important to note.
      2. Depending on the site, if an order is completed at a third party location, sometimes it is fed directly to the site EHR and sometimes not (and it may not be in the tables sites are looking in for order data). This is where the use of large language models may be very beneficial.
7. **QA/QC Task Force | Jennifer Pacheco (NU) & Lisa Martin (CCHMC) | Slides**
   1. In October, the QA/QC Task Force will launch the beta test manual chart review of orders (including referrals) and encounters. We are asking sites to follow guidelines when selecting their 30 subjects.
      1. Selection criteria includes participants at risk for at least one high risk condition. It is preferred to have the majority of the selection not have prevalent disease.
   2. We will use REDCap forms for chart abstractions of relevant encounters and orders. We had initially created forms for encounters and referrals, however with the recent focus on orders we will be updating the referral form to be for orders, which will include referrals.
      1. ACTION ITEM: The QA/QC Task Force will update the referral manual chart review form to include orders before **October 2nd**.
   3. The encounters manual chart review uses encounters to include any contact with the participant, including MyChart messaging, phone calls, telemedicine, etc.
   4. It is important to capture if the diagnosis or high risk condition was discussed, if it was coded, and if so, how it was coded.
   5. The forms include a large free text field for reviewers to copy and paste in the entire note. This information can be used for later reference. In the future, this can be de-identified and used for LLM training.
   6. Identifiable data will not be shared between sites. We have developed instructions to accompany the chart review. There will be 30 subjects reviewed at each site for the beta test. UAB and CCHMC will have one month for the beta test. After the beta test, the task force will take time to review and make any changes, including to the number of charts or what is being asked to capture. In January 2025, the Task Force will ask all sites to perform the manual chart review over a few months.
   7. After the beta phase, we will ask sites to include a few subjects just below the high risk threshold. We aim to only review subjects with at least one encounter. If possible, it would be preferred to have participants who have had an encounter with the provider that was listed on their baseline survey. Not all conditions have orders or referrals in their outcomes.
   8. At the time of the network-wide launch, we will ask sites to pull referral and other data to compare what the chart reviewers are finding on the front end to what informaticists can find on the back end.
   9. Jen Pacheco is leading the MCS to publish this work. This work can contribute to standards such as OMOP, and provide guidance on how orders and referrals can be stored.
   10. The QA/QC Task Force is also looking at missingness assessments, which can inform QC guideline development. There is missing or implausible data in the CDP.
   11. The work is currently concentrating on the baseline and pre-RoR surveys, and clinical data elements pulled from the EHR that are used in GIRA generation. It is focusing on variables necessary for outcomes analysis. We are trying to share best practices for obtaining the data and finding participants with missing data. One challenge is how to code for participants who are missing data due to being lost to follow up. There is also a plan to publish on this work.
   12. The Task Force will work with GRID to implement the QC guidelines and metrics on the analysis ready CDP dataset.
   13. There is also a plan to publish a QC-focused lessons learned manuscript. The missingness assessment focuses on survey data completed by participants and is separate from the manual chart review. Many survey variables are used in outcomes measurements, so this will inform how missingness affects outcomes. Sites are able to choose who will conduct the manual chart review.
   14. The 30 charts per site was decided by examining distributions of probability of identifying systemic errors. It is also important to consider feasibility and availability from the sites.
   15. Sites are being asked to look anywhere in the chart for the manual chart review for any interaction we can find on the front end of the EHR. This includes inpatient and outpatient encounters. Clinics and specializations do not necessarily show up in the main EHR chart. Each site may have different structures. If a participant had care outside the site system, it would not necessarily be identifiable.
   16. There are automated ways to check if units are correct for the variable.
   17. Preliminary study results show that even when people provide a standard code, it is only 40% correct. People don’t match orders to the correct LOINC code. Chart reviewers are asked to record if they see the condition mentioned, and what codes are seen (type and actual code).
   18. If the additional funding for eMERGE is gained, additional chart reviews could be performed, which would benefit the LLM work.
   19. By the end of winter/beginning of spring 2025, we will be able to compare the charts to what data the informaticists have found, with the hope of being able to publish the paper in the fall of 2025.
   20. There was a suggestion to focus on the lower end of ages for pediatric participants, because the older the child the more prevalent disease we would see.
   21. It was suggested to record what kind of provider places the order.
       1. Some sites have study staff placing the orders in order for the study to pay for the orders.
       2. The review form does ask what type of provider the encounter was with.
   22. The attribution of who places the order is not the primary focus. The study has been designed to compare the number of orders and referrals among participants who received a high risk GIRA to those who did not.
8. **Genomic Risk Innovation and Discovery (GRID) | Adam Gordon (NU) & Matt Lebo (MGB) | Slides**
   1. Adam Gordon presented the GRID report out. eMERGE is a resource rich in the types of data being collected on participants.
   2. A goal for GRID is to ensure this data is in a usable format by the network and external individuals. One GRID initiative is to build out the toolset on AnVIL.
   3. We have a demonstration project to establish best practices and workflows. This is the VTE PRS, which has been led by the Columbia team. The MGB team worked with the Columbia team to modify the eMERGE workflow WDLs. These are currently being tested and the next step is to replicate and validate with eMERGE 4 data.
   4. We are also building out workflows for best practices for PRS assessment.
   5. There is currently eMERGE data in AnVIL. There is genomic and phenotype legacy data. If eMERGE members have questions on how to access and/or use legacy data on AnVIL, please contact the GRID workgroup.
   6. There are two workspaces with the eMERGE 4 data currently on AnVIL. These include PRS arrays, CDP data, and EHR data.
   7. In the future, AnVIL will contain the final Broad multisample, Invitae data, the CDP-derived analysis ready files, and the MeTree batch export.
   8. eMERGE has access to two new tools, i2b2 and Tanagra. Tanagra provides record counting access of eMERGE data for feasibility studies.
   9. I2b2 provides cohort building functionality for multiple eMERGE data types. I2b2 is part of the AnVIL Clinical Resource, which Matt Lebo co-leads. I2b2 is still in development.
      1. Currently, the ontology is based on the R4 instruments, which is not the most intuitive for analysis. The need to create a more user-friendly and intuitive ontology has motivated the work to create the CDP-derived analysis dataset.
   10. GRID has been focusing meetings on reviewing the CDP instruments and making decisions on any consolidation, translation into standard formats, and documentation. This will also be used to update the i2b2 ontology. It is important to extract the rich data from the CDP used in analysis in a way that is understandable, and to document any oddities or important notes.
   11. Adam presented three main recommendations for the CDP-derived analysis-ready dataset, which involve family relationships, FHH risk, and variant reclassification.
       1. Performing an AnVIL based IBD analysis may assist in capturing family relationships that were not manually captured in the R4/CDP family relationships instrument.
       2. FHH risk can be a combination of two sources, the rescue survey or MeTree. The group wants to consolidate the two risks for easier access and analysis.
       3. Due to the API connection with Invitae, Invitae variant reclassifications are pushed automatically into R4. The Invitae instrument is not a repeating instrument. Because of this, when a new Invitae result is pushed to R4, it automatically overwrites the previous Invitae result. There is potential to perform research on reclassification in the long term.
   12. The workgroup aims to alternate between working sessions/data workshops and regular workgroup meetings each month. On October 14th, there will be a MeTree presentation on what the batch data will look like. GRID will work on the initial QC and filtering of the data. There will be challenges in harmonization.
   13. MeTree data is very unique to this network, and can be a rich source of data especially overlaid with phenotyping and genotyping information. The GRID workgroup will work to QC the MeTree batch export in order to make it as usable as possible.
   14. The data dictionary for the CDP currently lists very high level details of each variable and instrument. A goal of the GRID workgroup is to expand on it, and make it more accessible to individuals not involved in building the GIRA. R4 was not designed with analysis in mind. There is effort toward translating it into a more easily shareable format. The CDP DD contains calculated data.
   15. The Phenotyping workgroup is working on the EHR level data dictionary.
   16. The GRID workgroup has not yet considered using a commercial product to map the CDP data into a standardized format. The EHR data will use OMOP and other standardized forms.
   17. The GRID workgroup is trying to develop WDLs on AnVIL that can be used more broadly across the network. The WDL used in the VTE PRS can be used for other PRSs to test against the eMERGE 4 dataset as a validation cohort. The first part is generating the PRS and performing any ancestry adjustments. The next part will be evaluation of the PRSs. A second use case of analysis on AnVIL is determining pedigree status from genomic data. Ideally, this WDL would be run on AnVIL.
   18. AnVIL is viewed as a home for network-wide, standardized workflows. As condition groups develop methods for outcomes analysis, it would be beneficial for those to be adapted into WDLs. This way, they can be run consistently on AnVIL.
   19. As a network, the decision was made to not regenerate and re-return a GIRA even if a variant is reclassified. The reclassifications are initiated by Invitae and not by the network. The update is pushed to the Invitae patient portal if the participant has signed up. The update is pushed through the Invitae API into REDCap.
   20. When the reclassifications occur, sites are not alerted. There are less than ten variant reclassifications per site. We do not have an idea of first degree relatedness in genotyped participants yet.
       1. There are three different sources for family relationships, which are the family relationships instrument, MeTree, and genetic data.
       2. Not everyone in MeTree has genotyping data, only those who are enrolled in eMERGE.
   21. It is important for the network to be cautious regarding the variant reclassification if the basis of reclassification is unknown.
       1. ClinVar lags several months behind clinical labs in terms of reclassification.
   22. GRID is hoping to determine how to best capture the family history data in a standard format.
   23. MeTree and the rescue survey contain a limited amount of information that varies depending on the participant. MeTree and the rescue survey were limited to the eMERGE conditions. However, initially the MeTree portal had additional options beyond the eMERGE conditions. This changed as recruitment moved forward.
       1. Because of the differing amounts of information, the GRID workgroup is considering generating a smaller derived set of data.
   24. eMERGE will receive full sequencing for the genes that were reported on in the study.
   25. ACTION ITEM: Invitae will follow up with information on reclassification, and on how many participants signed up for the patient portal.
   26. ACTION ITEM: Invitae will follow up to confirm whether the BAM file being provided to eMERGE is on the gene panel only or on the full assay.
9. **CIRT | Emma Perez (MGB) & Bob Freimuth (Mayo) | Slides**
   1. The workgroup is focusing on manuscript development.
      1. Integrating Polygenic Risk Scores into the Electronic Health Record: An eMERGE Network Perspective
         1. Contextualizing PRS in patient care is a large piece of this manuscript.
         2. Clarity of regulating bodies is another focus.
      2. eMERGE IV: Experience integrating genomic-based risk assessment reports into the medical record.
         1. A focus of this manuscript will outline the main workflow differences among sites.
         2. GIRA location and timing of uploads differed between sites and will also be outlined in this paper.
         3. Clarity on the 21st Century Cures Act and how to interpret is another focus of this paper.
   2. Versioning of the PRS, if any updates have been made, will prevent losing data from earlier in the study.
      1. Recalculating of the PRS is not done in this version of eMERGE.
      2. PRS versioning should be addressed in future versions.
      3. PRS cannot be created unless it is in the PGS catalog. eMERGE is listed as research PRS.
   3. Placement of the manuscript and key words for the eMERGE IV: Experience manuscript will be very important beyond genomics due to the 21st Century Cures Act and policy interpretations between sites.
10. **Input/Feedback from the ESP, general discussion | Slides**
    1. The ESP was very positive about the progress that the Network has made. This is a very important project in genomic medicine.
    2. CARE
       1. When Invitae reclassifies a variant the result is overwritten in REDCap. The old result can still be seen in the GIRA report and is also logged in REDCap.
          1. The Network should consider developing a way to maintain a more clear record of previous results.
          2. The ESP suggests the Network be careful about the process of variant reclassification and how it is handled from a research standpoint and from a clinical standpoint.
          3. Providers are notified of updated Invitae reports. GIRAs are not being updated.
          4. Sites should encourage participants to sign up for the Invitae patient portal.
          5. DECISION: Reclassification of Invitae variants should be returned as defined by each site's clinical workflow for return of monogenic results.
             1. Internal eMERGE discussion of reclassifications prior to return can be reviewed as needed for guidance.
       2. There could be research value in redoing the PRS and making comparisons.
    3. Outcomes
       1. Social determinants of health surveys can generate very valuable information.
       2. The Network is encouraged to look at how language and demographics are used in publications and if there can be consistency.
          1. There is currently a group looking at population labels and descriptors, including consistency of terms.
       3. The ESP suggested the value in the potential for a monogenic only analysis (when someone is labeled high risk based on monogenic alone).
    4. QA/QC
       1. The ESP feels the n of 30 per site regarding manual chart review is a little low and encourages the Network to think of it as an iterative process and perhaps do more manual chart review based on what is seen.
       2. Data that is labeled even with standard codes is notoriously poor quality (wrong units of measure, values outside of reasonable range).
          1. The Network is encouraged to look at ways to make sure data being used is as high quality as possible.
    5. GRID
       1. The plan is to incorporate as much legacy data as we can into the i2b2 build.
          1. Standardized algorithms have already been run on EHR data.
       2. The initial push is to get e4 data and get it published on AnVIL, then get legacy data.
       3. The intent is to have AnVIL be used for longitudinal data.
11. **Closing remarks | Rex Chisholm (SC Chair, Northwestern)** 
    1. Thank you to the eMERGE consortium for great advice over the years. Responses to the ESP suggestions will be put into writing and shared.
    2. Thank you to the CC for planning and executing the steering committee meetings.
    3. We look forward to a productive next few months. The next meeting will be in February in Bethesda.

**Action Items:**

1. Invitae
   1. Invitae will follow up with information on variant reclassification, and on how many participants have signed up for the patient portal.
   2. Invitae will follow up to confirm whether the whole BAM file is being provided to eMERGE or just the targeted genes that are being sequenced.
2. CC
   1. The CC will investigate in R4 how to capture and document variant reclassifications from Invitae.
3. QA/QC
   1. The QA/QC Task Force will update the referral manual chart review form to include orders before **October 2nd**.
4. Outcomes/Phenotyping
   1. Alanna DiVietro (CC) will put all condition orders of interest into one list and circulate to the outcomes group for confirmation. The list will then be brought to the phenotyping workgroup to begin pulling sample data.
5. GRID
   1. The GRID workgroup will investigate the availability of a WDL to run IBD or another relatedness measure in AnVIL on the eMERGE 4 cohort.
6. Network
   1. Re-review the publication plan for outcomes analysis with regards to including pediatric data in the main outcomes manuscript.
   2. Review and adjust timeline as needed for EHR data pulls to maximize most impactful data now the network will continue for multiple years.
   3. Sites should encourage participants to sign up for the Invitae participant portal if they have not already.
   4. The Network should re-review the publication plan, confirm authorship and scope overlap, and include manuscript concept sheets that focus on the process and details of the network in addition to the results.

**Decisions Made:**

1. All GIRAs should be placed in the EHR by October 31st even if a return appointment has not been scheduled.
2. Reclassification of Invitae variants should be returned as defined by each site's clinical workflow for return of monogenic results.
   1. Internal eMERGE discussion of reclassifications prior to return can be reviewed as needed for guidance.

**Official ESP Recommendations:**

Meeting Summary

eMERGE Network – External Scientific Panel (ESP) and Steering Committee

*Executive Session – 09/26/2024*

|  |  |  |  |
| --- | --- | --- | --- |
| **ESP** | **Dan Rader,** University of Pennsylvania **– Chair**  **Kimberly Doheny\*,** Johns Hopkins University  **Stanley Huff**, Intermountain Health  **Janina Jeff**, Illumina  **Brendan Lee\***, Baylor College of Medicine  **Lisa Parker**, University of Pittsburgh  **John Witte\***, Stanford University | **NHGRI** | **Teri Manolio**  **Jessica Reinach**  **Robb Rowley**  **Rene Sterling**  **Nephi Walton** |

\*Attended virtually

The ESP met with NHGRI program staff during the executive sessions of the eMERGE Steering Committee/ESP Meeting held on September 25-26, 2024. The ESP was extremely impressed with the progress the Network has made, including the Network’s thoughtful discussion defining the research questions, discussing data challenges, and refining and expanding manuscript topics. The ESP provided observations and recommendations about several aspects of the research to help move efforts forward. Initial feedback is elaborated upon in the summary below.

**The Network should define the process for relaying reclassified monogenic variants to its participants.**

The ESP discussed the importance of informing participants of a reclassified monogenic variant. The clinical importance and potential significant change in risk result must not be overlooked. While the Network has agreed not to update and re-return GIRA, it is still important to inform the participants of new risk if it arises. The Network sites should ensure that there is a transfer of responsibility for following up results from the physician of record to a participants’ health care provider. Leveraging existing infrastructure for genetic testing at each of the sites could be used to ensure appropriate follow up. The ESP also emphasized the importance of making sure that the Invitae results are versioned in R4 and the patient’s record; both the original monogenic report and the reclassified versions should be available for viewing and pulling for data analyses.

**The Network should work on a paper that addresses outcomes among adults and children.**

The ESP recommended that the Network publish a high-level outcomes paper including both adults and children. Since the study was designed to cover the lifespan, publishing high-level papers with adults separate from children could invite criticism regarding study design and implementation for child research participants. In addition to supporting the publication of papers focused on adults only, the ESP acknowledged the importance of pediatric PRS and the need for publication of papers focused on pediatric outcomes.

**The Network should address how best to publish findings on ELSI topics, whether in ELSI focused manuscripts or in manuscripts where ELSI is embedded in or providing an analytical lens on other topics.**

The ESP noted the importance of distinguishing when ELSI research findings, scholarship, or perspectives should be incorporated into other work group manuscripts, and when there should be standalone ELSI manuscripts. Manuscripts that use a multidisciplinary perspective inclusive of ELSI, genetics, genomics, health services research, data science, and/or other relevant disciplines are encouraged, where appropriate.

**The Network should consider how to maximize use of available social determinants of health data and use standardized language when discussing population descriptors.**

The ESP emphasized the importance of including the SDOH data available to the Network in outcomes analysis. In some cases, SDOH may be more accurate measures of a phenomenon of interest than measures of race, ethnicity, or genealogical ancestry. When using population descriptors pertaining to race, ethnicity, or genetic ancestry in analyses, the Network should consider standardizations such as the recently published NASEM report on the use population descriptors in genomics research, along with published guidelines and recommendations (e.g., Feero et al 2024). The specific research questions posed should guide whether and which SDOH and population descriptors are used in a given analysis. research, along with published guidelines and recommendations (e.g., Feero et al 2024).

The ESP suggested creating a list of terms, definitions, and appropriate uses in the context of eMERGE, and making it available to the entire Network. The list could be used for people analyzing the data and writing papers to ensure various descriptors are consistently used and interpreted across the Network.

**The Network should make sure to emphasize manuscripts detailing consortium processes in addition to its emphasis on disseminating results.**

The ESP recognized the unique perspective of the eMERGE cohort on harmonizing a genomic cohort study across ten clinical sites utilizing somewhat varied approaches. The ESP acknowledged the Network’s extensive experience with overcoming challenges such as extracting data from various health systems, compiling a wide range of data, returning GIRA to patients, incorporating the GIRA into the electronic health record (EHR), and collecting outcomes using EHRs and surveys. Recognizing the significant efforts involved, the ESP believes it is crucial to capture and share valuable lessons learned with the scientific, clinical, and ELSI research communities, as well as health policy and health systems management audiences. The Network is encouraged to provide detailed insights regarding study design and implementation to enhance researcher understanding of the challenges and requirements tied to genomic medicine research and practice.

**The Network should consider increasing manual chart review, implementing a semi-automated process, and sharing lessons learned about using EHR for outcome analysis.**

The ESP emphasized the need for more extensive and comprehensive chart review, suggesting the planned 30 manual chart reviews per site may not be sufficient and that a tiered approach, improving processes and efficiency with each set of reviews might be useful. They recommend supplementing manual chart reviews with an automated process. Validating the use of EHRs to capture outcome data would provide valuable insights not only for genomic medicine but for the broader clinical and ELSI research community. Sharing eMERGE’s approach could facilitate the use of EHR data for outcome capture across various settings.

**The Network should continue to work with AnVIL to establish a process to transform data into a common data model for analysis.**

The ESP recommends the Network continue to work with the AnVIL team to develop standardized processes and procedures to deposit the eMERGE data into a common data model. The ESP noted that many investigators, especially in multisite studies, spend the majority of their time capturing and harmonizing data with little time remaining for data analysis. eMERGE has an opportunity to reverse these time allocations by helping AnVIL develop an approach to capture and then ingest the data into a cloud-based platform, accompanied by a host of analytical tools. These efforts will not only help the Network but could also result in the development of robust tools and processes becoming available to future investigators to do similar work on the AnVIL.

**The Network should examine site differences in return of results procedures and share lessons learned about the challenges with implementing GIRA.**

The Network noted that the GIRA is not integrated into the EHR consistently across sites, and at some sites it was only entered into the media page. Differences also existed in what sites offered clinical decision support and which did not. It is important to address these differences as a group, both in research and for clinical care of patients. The ESP also recommends publishing on the ELSI of these differences and challenges and on ways that future groups might integrate genomics more thoroughly into the EHR.

ESP Recommendations for the Network:

1. The Network should define the process for relaying reclassified monogenic variants to its participants and assure the original and updated information is available in the research dataset.

2. The Network should work on an overall high-level paper that addresses outcomes among both adults and children.

3. The Network should address how best to publish findings on ELSI topics, e.g., in ELSI focused manuscripts or in manuscripts where ELSI is embedded and providing an analytical lens on other topics.

4. The Network should consider how to maximize use of available social determinants of health data and use standardized language when discussing population descriptors.

5. The Network should make sure to emphasize manuscripts detailing consortium processes in addition to its emphasis on disseminating results.

6. The Network should consider increasing manual chart review, implementing a semi-automated process, and sharing lessons learned about using EHR for outcome analysis.

7. The Network should continue to work with AnVIL to establish a process to transform data into a common data model for analysis.

8. The Network should examine site differences in return of results procedures and share lessons learned about the challenges with implementing GIRA.