**Summary of Steering Committee Meeting: August 2020**

August 20-21 via Zoom

**eMERGE Day 1: Thursday, August 20th, 2020**

* [Welcome and Meeting Goals| Rex Chisholm (SC Chair, Northwestern) & Robb Rowley (NIH/NHGRI)](#io0zj93l19ig)
* [Major workgroup decision points | Workgroups](#t95hb73dhb09)
  + [PRS validation & evaluation | Eimear Kenny (Mt. Sinai); Patrick Sleiman (CHOP))](#940ig66dydz8)
  + [Provider uptake & outcomes | Noura Abul-Husn (Mt. Sinai); Nita Limdi (UAB)](#qfjns97cvx4u)
  + [Genotyping | Meg Roy-Puckelwartz (NU); Niall Lennon (Broad/CC)](#q0bif9k10nxk)
  + [Recruitment, Retention, IRB| Digna Velez-Edwards (VUMC); Ingrid Holm (BCH); Wendy Chung (Columbia)](#8bxg0sy3jt2i)
  + [EHR workflow & infrastructure | Luke Rasmussen (NU); Robert Freimuth (Mayo)](#2zvet7w535t)
  + [Comprehensive risk assessment & return | Gail Jarvik (UW); Iftikhar Kullo (Mayo), Cindy Prows (CCHMC)](#xvxapfgzeq0j)
  + [Phenotyping | Chunhua Weng (Columbia); Wei-Qi Wei (VUMC)](#epcx1pflzcof)
* [Evidence Review of conditions | Sites](#ipkbfhwri57)
  + [CCHMC](#erix0j8durc7), [CHOP](#ol2ylofsjprn), [Columbia](#d9jaok1z6xeq) , [Mt.Sinai](#kqgplb8xbvyz), [MGB](#bhkwr09vegy1), [Mayo](#fkq4stmnzqdq), [NU](#7ss23kipwmhp), [UAB](#y83tab70usgu), [UW](#gdlwjdj3bcp3), [VUMC](#nnkkym2typ2l)

**eMERGE Day 2: Friday, August 21st, 2020**

* [Targeted Sequencing in Prospective Cohort| Rex Chisholm (NU)](#cl9g3kj6t128)
* Workgroup Report Outs
  + [Genotyping | Meg Roy-Puckelwartz (NU); Niall Lennon (Broad/CC)](#aa7syshydfmq)
  + [PRS validation & evaluation | Eimear Kenny (Mt. Sinai); Patrick Sleiman (Chop)](#uhge2jf0oeat)
  + [Phenotyping | Chunhua Weng (Columbia); Wei-Qi Wei (VUMC)](#s9qzhhk4vbj7)
* Workgroup Report Outs
  + [EHR workflow & infrastructure | Luke Rasmussen (NU); Robert Freimuth (Mayo)](#378wumhg34bq)
  + [Comprehensive risk assessment & return | Gail Jarvik (UW); Iftikhar Kullo (Mayo), Cindy Prows (CCHMC)](#xvxapfgzeq0j)
  + [Provider uptake & outcomes | Noura Abul-Husn (Mt. Sinai); Nita Limdi (UAB)](#efqy3idigpx6)
* [Workgroup report out & sIRB framework| Digna Velez-Edwards (VUMC); Ingrid Holm (BCH); Wendy Chung (Columbia)](#wuwg00eqqmvf)
* [Initial PRS condition ranking | Rex Chisholm](#huqu1312qq7f)
* [Closing remarks | Rex Chisholm (SC Chair, Northwestern)](#6s88rzwimjdz)

[**ACTION ITEMS**](#oltxui9zt3vj)

**eMERGE Day 1: THURSDAY August 20th, 2020**

* **Welcome and Meeting Goals| Rex Chisholm (SC Chair, Northwestern) & Robb Rowley (NIH/NHGRI) |Robb** & **Rex**
  + Program Official Report: Robb Rowley welcomed the Network to eMERGE. The Network was introduced to the NHGRI team and ESP members. The first ESP meeting is tentatively scheduled for December 2020. Special thanks were given to the coordinators of the CC.
  + The progress of eMERGE: Phase 1 of eMERGE involved addressing data privacy issues and demonstrating the feasibility of using EMR for genomic research. In Phase 2, GWAS were performed and PheWAS methods were developed. Phase 3 focused on integrating sequencing results into the electronic medical record and understanding the outcomes. The goal of Phase 4 is to study how to combine genomics, clinical covariates, and family history to understand a patient's risk. The risk estimate and a management plan will be returned to 25,000 participants while measuring the impact the approach has on clinicians and patients.
  + Current eMERGE Data: eMERGE has a total of 157,480 phenotype & genotype data available to the Network. The centrally collected phenotypes (demographics, codes, DMI, lab values, medications, and geocoding data) are refreshed yearly. Currently the data is not contained in a Observational Medical Outcomes Partnership (OMOP) data model. The goal is the data in the OMOP data model by [date].
  + The August 2020 Steering committee meeting's goals are to review each condition's evidence and select at least ten conditions to include in the Networks set for PRS validation, review the network decisions that need to be made, and to identify working group decision points that need to be addressed to allow the Network to proceed with eMERGE goals. The Network will also discuss targeted sequencing for the 25,000 participant cohort and establish network needs for PRS score development, as well as establish sIRB framework needs, and identify the recruitment population criteria.
* **Major workgroup decision points | Workgroups** 
  + *PRS Validation & Evaluation Workgroup* 
    - Each site used the Evidence Review Framework to evaluate the suitability for implementation of their two assigned phenotypes. The Framework included analytic viability, feasibility, actionability, and translatability. The Framework was used to capture the available data including whether there already exists a polygenic risk score (PRS).
      * Feasibility: To examine if the PRS is able to be implemented in our prospective population. This will examine existing polygenic risk scores and racial diversity. It also includes information on SNP heritability and the prevalence of a disease.
      * Actionability: If return of a PRS would necessitate a change in care. This assumes a score has already been generated and examines its usefulness within clinical care.
      * Translatability: Ability to put into clinical practice. Examines other known predictors of risk and how public health is being impacted as well as the feasibility and cost of screening.
    - Each site presented their analysis in order to help finalize the list of conditions the PRS Validation and Evaluation WG would start investigating to determine if ready for implementation. To make this happen, a process needs to be in place to evaluate the scientific validity of existing scores, remembering that not all scores are equal, and the ability to generate new PRS scores for the network.
      * The workgroup would like to develop a standardized way to validate each score and evaluate predictive accuracy not only for individual scores but across scores. The WG recognizes the importance of assessing PRS in diverse populations while considering availability of data.
      * The workgroups will work together to evaluate clinical utility. The PRS workgroup will take a deeper dive into technical considerations for implementing polygenic risk scores while also focusing on making decisions to move a condition forward.
    - If stratification occurs, there might be a loss of power but increased precision or visa versa. Drawing upon the literature and more validated polygenic risk scores within different populations may inform if the PRS is utilizable in multiple diverse populations. Stratification will be phenotype dependent to a certain degree and transethnic scores that work across populations will be favored as much as possible.
  + *Provider Uptake & Outcomes Workgroup* 
    - Participant baseline measures will be captured by a single harmonized core survey including factors such as medical history, family history, socioeconomic status, geographic location, environmental factors, and prior genetic testing.
    - Provider knowledge and attitudes will be captured using a core survey that will have elements from previously used surveys from eMERGE, CSER, and other consortiums as well other validated tools and references. The core survey will be finalized by the workgroup before recruitment. Provider knowledge and attitudes towards polygenic risk scores and genetic risk assessment will be captured through a separate survey.
    - Proposed framework (study design) was presented.
      * The group expects to have around 7500 patients identified as high risk. Risk reduction recommendations are expected to be 50-90% implemented.
    - **ACTION ITEM:** The outcomes workgroup will collaborate with other workgroups on the study design.
  + *Genotyping*
    - The Genotyping workgroup has identified initial questions that need to be answered within the group and the network.
    - There was a concern raised in the first Genotyping workgroup meeting regarding whether Low Pass Whole Genome Sequencing would be a better option than the Genotyping Array. In making this decision, consideration should be given to method availability, lead time, validation status, relative performance, compatibility with aims, and cost.
    - Process Inputs and Outputs
      * **ACTION ITEM:** Sites whose institution requires an MTA to transfer genomic DNA should notify the Meagan Harden (mharden@broadinstitute.org) at the Broad.
        + Sites should be aware that the Broad does not require an MTA to receive the material.
      * Both blood and saliva samples are currently validated using the Global Diversity Array.
      * Sites will send extracted DNA directly to the Broad, the GDA is not validated by specific kit type, sites can use any method to collect/extract saliva as long as it is in a CLIA setting.
      * When determining the PRS methods, it needs to be determined what form these will take and the covariates that are expected to be incorporated.
      * Process Outputs
        + The workgroup needs to determine what files are returned to each site and where they should be delivered – AnVIL, R4 etc..
        + Do sites expect a PDF report?
        + What deliverables fall under the CLIA umbrella? The GRA ‘report’ is not an anticipated deliverable from the Broad CLIA pipeline.
        + The Broad currently plans to deliver to the sites raw genotyping data, initial Virtual Card File(VCF), and Polygenic Risk Score (PRS) outputs. There will also be pre-imputation cleaning dine as part of the genotyping workgroup pipeline.
      * The group will need to determine whether or not the genome build is important to sites and the implications for monogenic annotation.
      * The workgroup will need to determine whether or not TOPMED or other reference panels are important, and the measures sites are using to evaluate the impact of various reference panels.
      * Network needs to establish the logistics of sending samples, one at a time or in batches. Lab prefers batches. It should also be determined whether or not the cadence of the sample return will impact the plans of returning results at other sites.
      * The Broad is not NY certified as of yet, but is currently in the process of becoming certified. If certification is not received in time, then the Broad will submit a request for exemption. The certification process has been delayed due to the COVID-19 pandemic.
  + *Recruitment, Retention, sIRB, & ELSI* 
    - sIRB
      * Protocol decisions need to be made including how to develop the single Institutional Review Board (sIRB) protocol and integrate input from other workgroups into protocol decisions. This workgroup will work on the process on how all of the information from these workgroups will be written into a single protocol.
    - ELSI
      * Site-specific ELSI projects: Several sites are conducting similar projects. The working group has identified similar ELSI projects with plans to combine them to make the findings more robust. Since most ELSI projects will be completed in the first year, findings from these projects might be used to inform the Network study.
      * ELSI for the main project: The main ELSI project that came out of the workgroup kickoff call is Return of Results. Multiple questions regarding return of results will need to be addressed across the main project.
    - Common Metrics
      * Data collection elements need to be harmonized for recruitment for the common protocol. Drafting common participants survey instruments will also need to be harmonized across sites. Discussions with the Provider Uptake and Outcomes workgroup will help with determining measures and getting those into the protocol. Another strategy regarding surveys is to build on survey development for Health Care Providers and participants in eMERGE III.
    - Confidence in results for ethnicity and age, reinterpreting results and the need for giving revised reports, and timing of return of results are very difficult issues that need to be solved for the protocol.
    - The role of ELSI needs to be collaborative with all the other workgroups in order to anticipate as many possible issues ahead of time.
  + *EHR workflow & infrastructure* 
    - The work surrounding the definition of standards (bulleted below) for representation and transmission of data and reports needs to start as soon as possible.
      * Reports: genetic, clinical, FH, confirmatory monogenic
      * PRS: calculation, object structure, metadata
      * Data flows: between sites, Broad, R4, AnVIL
    - The timeline for defining data standards needs to be accelerated as all other work streams depend on knowing these standards. Workflows for data collection and return of results are going to be impacted by decisions by the sIRB.
    - Data collection can either be a balance between uniformity of process between sites or focus on the touchpoints at each site (interoperability). An understanding of what needs to be collected and shared across the site will impact the outcomes data.
    - **ACTION ITEM:** Common terminology/nomenclature should be established across the network. Example: PRS vs GRA.
    - The scope of responsibility between sites/systems needs to be established for data flow. Establishment of Service Level Agreements needs to be upfront if needed to bring clinical systems together.
    - Flow of clinically-returnable information will need to be determined.
    - Dissemination and promotion of outcomes (traditional and non-traditional modes) can be used to report metrics and other outcomes. For example traditional modes (publications and conferences) or non-traditional modes (open source, white papers, implementation guides).
    - The Network should consider mechanisms to put into place to continually ask new questions throughout the project. Including concerns about data flows , decisions made regarding data exchange need to be highly collaborative, accelerating the timeline for defining data standards.
  + *Comprehensive risk assessment & return (CARE group)* 
    - The workgroup co-chairs presented the workgroup’s major decision points and cross collaboration with other workgroups to accomplish them.
    - The PRS validation pipeline for diverse ancestries will be available in collaboration with the PRS and Genotyping workgroups. The PRS will be efficiently calculated for minorities and customize reports based on score. To model risk, the workgroup is considering including clinical risk equations, environmental variables, family history through MeTree, and monogenic risk. Determining how to integrate these risks and return to participants, providers, and EHRs needs to be determined.
    - The Genotyping workgroup will assist in the decision making of the risk calculations and imputation based on the Global Diversity Array (GDA).
    - The workgroup will work with the Recruitment, Retention, and ELSI workgroup to assess the ability of sites to adapt sIRB protocol and understand what participants would prefer to receive in their reports. Reports will need to be adapted for groups such as children and minorities.
    - The workgroup will decide if clinical equations will be extracted from the EHR or through a combination of surveys to be administered at the time of enrollment. Participant surveys will supplement family history information from the EHR.
    - The workgroup will collaborate with the EHR Integration workgroup to mine risk estimates from EHR and integrate decision support for providers. A plan will be developed for high risk participants through CDS creation for each condition and placement in the AnVIL. The group will work to define high polygenic risk scores (PRS) and comprehensive risk score (CRS) and if there is a cut off by trait.
    - The comprehensive risk estimates will be provided to the Provider Uptake & Outcomes for outcomes assessment.
  + *Phenotyping* 
    - The phenotyping workgroup will evaluate if there are existing algorithms for each PRS condition, if the phenotypes target the same specific condition as the PRS, and how case/control status is captured, either in a binary fashion, or with a timing component which may be relevant when examining onset of disease.
    - Collaboration with other workgroups will be key in order to accomplish goals.
      * Questions that the phenotyping group believe to be critical at this stage:
        + Do we have existing phenotyping algorithms for each PRS condition? If not, how challenging is it to create one within a short time frame?
        + How will phenotyping algorithms/approaches facilitate identifying clinical risk factors?
        + How are we using electronic phenotyping to define outcomes?
        + What is the timeline of providing phenotyping status for PRS?
        + How should we work with other workgroups to prioritize and customize phenotype status (covariates)? Without the bandwidth to work on 10-20 phenotypes simultaneously, it is important to work together to figure out the best way to move forward.
    - Typically phenotype data are submitted through PheKB which has a basic, initial quality control process. In this new phase of eMERGE, it will be important to establish if sites will upload their data directly to AnVIL and if so, how the quality control will be done.
    - For each phenotype, the workgroup needs to identify the relevant phenotypic traits and outcome measures for each condition.
    - It was suggested that for each case and control status, age is included. The phenotyping, genotyping, and PRS workgroups should collaborate to decide on the timestamp for controls.
* **Evidence Review of conditions | Sites** 
  + *CCHMC*
    - Asthma
      * Rationale: Asthma is the most common chronic disease in children and is the third leading cause of hospitalization in children. Asthma is also four times more common in African American children. Asthma is a heritable condition and the likelihood of a child having asthma increases if both parents have asthma. Actionability will include educational material, trigger identification, and avoidance, environmental modifications, establishing PCP, early treatment with controllers, education about quick relief –vs.- controller medications.
      * Sample Size and Data: 17 of the 87 GWAS studies were done in African American children. The eMERGE III GWAS in African American pediatrics was 1,573 and 3,995 controls. The goal for CCHMC was a low-read sequencing of 18,000 African American samples. Of these samples, about 5,200 samples in African American pediatrics and about 10,000 controls.
    - Atopic Dermatitis
      * Rationale: Atopic Dermatitis occurs in 10-20% of children, and several clinical measures can be taken early on after diagnosis. This condition is very common in asthmatic children and 1.6 times more common in African American children. Actionability will include bathing using mild or non-soap cleansers, using emollients in the neonatal period, education to enable early recognition, and to treat acute flares. This condition is highly heritable, and risk increases based on family history and gestational diabetes in mothers. Rare variants include loss-of-function mutations in the gene that encodes the epidermal structural protein filaggrin (FLG) (OR=6). There are 11 published GWAS studies and 111 variants with genome-wide significance.
      * Sample size and data: The eMERGE III GWAS in African American pediatrics has a sample size of 1244 cases and 2795 controls. CCHMC has the goal of having an independent dataset in the CCHMC biobank.
    - Systemic Lupus Erythematosus
      * Rationale: This is a prototype autoimmune disease common in the African American population. African Americans are three and a half times as likely to have this condition than Europeans. It is ten times more common in women than men with the highest deaths in women ages 25 to 35. There have been 187 published and 60 unpublished works regarding heritability. There are 314 SLE publications by Bahram Namjou, Ken Kaufman, or John Harley. Actionability includes teaching symptom and signs assessment, discouraging sunbathing, monitoring serology, cytopenias, & kidney function, examining for arthritis, rashes, oral ulcers, alopecia, and early treatment with less toxic medications than chemotherapy.
      * Sample Size and Data: There have been publications done in every ancestry. There are 124 genomic studies, with the largest study, including 200,000 participants. The eMERGE III GWAS African American sample size. In the eMERGE III GWAS, the African American sample size was 363 cases and 10,000 controls.
      * Independent datasets: CCHMC biobank has a goal of 18,000 in African Americans. The unpublished African American GWAS in African Americans includes 1,300 cases and 7,000 controls. There are also additional collaborative cohorts.
      * Summary GWAS files: GWAS files are available in the GWAS-atlas, GWAS-catalog, UK biobank. More than five summary GWAS statistics & ImmunoChip studies are also available.
  + *CHOP* 
    - Obesity/BMI
      * Rationale: Obesity is highly polygenic, with an estimated heritability explained 23.4% from the largest meta-analysis to date. There have been at least seven large scale studies published using PRS derived from GWAS obesity. A minimum of 2 loci were used to predict obesity in the studies with a maximum of 97. The AUCs ranged from 0.52 (two loci FTO & MC4R in African ancestry) to 0.7, not including traditional risk factors in the score.
      * In a study conducted by Khera et al. in April 2020, PRS’s from the UK Biobank were used all with European ancestry. The top 10% highest PRS scores were considered. Individuals considered to be in the high PRS group average were found to have a 2.9 kg/m2 higher BMI and 8.0 kilograms higher than noncarriers.
      * The CAG pilot polygenic BMI PRS with IHCC used similar association statistics but included different ancestry specific LD reference panels for CEU, AA, HSP/LTN, ASN, and Trans-Ethnic (TE). Scores were generated, and AUC was tested on the top 1% of BMI distribution. There were three separate Asian cohorts, a Scandinavian cohort, 2 European ancestry cohorts, and one Brazilian cohort. Trans-ethnic scores outperform population-specific scores in non-European cohorts with similar predictive values. In this study, CHOP observed a 1% BMI of 0.76 in the Shanghai cohort compared with 0.75 in MoBa. European ancestry weights outperform TE in European cohorts (NHS1, NH2, MoBa, CHOP EA). Scores can be improved, incorporating additional ancestry specific GWAS summary stats.
    - Crohn’s Disease (CD)
      * Rationale: Crohn’s Disease is the Fourth highest SNP heritability of any phenotype. It is often classified as an autoimmune disease. However, unlike ulcerative colitis, which has a strong HLA/MHC genetic signal. CD has not shown an association with this locus. Crohn’s Disease has a high rate of under-diagnosis in African Americans and other minorities and children. Many patients originally responsive to anti-TNFa biologics (e.g., Remicade, Humera) become resistant and undergo 2-3 different therapies, often with significant side effects. PRS will allow for earlier diagnosis and reduce the time from presentation to effective therapies. PRS may also aid identification of risk of response deviation and help guide alternative therapies. All published PRS scores have been tested in CEU.
      * Trans-ethnic summary stats are available from IIBDC meta-analysis on immunochip. CHOP also has summary statistics available for the African American population.
    - Type 1 Diabetes Mellitus
      * Rationale: PRS scores are highly discriminatory in Type 1 Diabetes. Distinguishing patients with monogenic diabetes from Type 1 Diabetes is important for treatment (i.e., sulfonylurea vs. insulin). Genetic etiology of Maturity-onset diabetes of the young (MODY)and NDM is unknown in 20% of cases. Additionally, gene panel screening does not pick up all cases.
      * Data: There are four PRS developed to classify Type 1 Diabetes vs. healthy individuals, there is also one PRS Type 1 Diabetes vs. Type 2 Diabetes, and three PRS Type 1 Diabetes vs. monogenic diabetes / MODY.
  + *Columbia*
    - Breast Cancer
      * Rationale: Columbia University has extensive experience in the Breast Cancer Association Consortium (BCAC). It has an NCI-funded Breast Cancer Family Registry to assess breast cancer risks. Columbia recently submitted the breast cancer PRS paper and breast cancer penetrance paper based on the eMERGE III dataset analysis. PRS scores can increase accuracy in minority populations.
      * Data: The breast cancer Phenotype was validated in eMERGE III. Columbia expects to have an immense amount of European based cases and controls. The largest GWAS dataset contains 100,000 cases and 120,000 controls. SNP based heritability around 8.8%. The mortality rate of breast cancer is found to be higher in the African American population.
    - CKD
      * Rationale: Columbia has an active research program in the genetics of kidney disease and has validated the Chronic Kidney Disease (CKD) phenotype in eMERGE III. In terms of datasets, serum Cr is available for most biobanks linked to EHR. Additional access to the Columbia CKD Biobank and several kidney transplant cohorts are available. CKD has a high public health impact on many racial disparities. There are also strong clinical predictors of risk with CKD.
      * Data: The PRS for CKD will require a new design, optimization, and tailoring by ancestry. At least 10% of US adults have CKD.
      * Actionability includes early prevention and intervention, screening serum Cr and U microalbumin, low salt diet, smoking cessation, avoidance of Nonsteroidal anti-inflammatory drugs (NSAIDs) and Proton Pump Inhibitors (PPIs), blood pressure control, renin-angiotensin system (RAS) blockers (proteinuria) sodium-glucose cotransporter-2 inhibitors.
  + *Mt. Sinai*
    - Hypercholesterolemia (HC)
      * Rationale: Hypercholesterolemia accounts for 20-30% of clinical family history. This condition can be applied to both the pediatric and adult populations. The heritability is about 20-90%.
      * Data: There have been several large GWAS of LDL-C published, including GLGC GWAS of ~1.6M trans-ethnic participants. Summary statistics from unpublished GLGC HRC/1KGP3 trans-ancestry imputation analysis will be made available.
      * Actionability/Translatability: Clinical scores are based on the pooled cohort equation for 10-year ASCVD risk. Interventions include a lipid panel, statin therapy (pediatric/adult, and the coronary artery calcium score (adult). There is a high prevalence and disproportionate burden of Hypercholesterolemia and Coronary Artery Disease in African Americans.
      * There is a concern that Hypercholesterolemia overlaps with CAD PRS (and the outcome is the same). LDL-C is a continuous trait, less heterogeneous, and high powered.
      * Common SNPs include APOE, which has a strong effect size. There are 100-200 associated loci. Rare variants include FH genes LDLR, PCSK9, and APOB. Mount Sinai has site expertise ROR in diverse populations.
    - Non-Alcoholic Fatty Liver Disease (NAFLD)
      * Rationale: NAFLD has a 25% prevalence due to rising obesity rates and is the leading cause of end-stage liver disease. NAFLD e phenotype accurately predicts disease risk. Genetic risk scores predict NAFLD and improve risk prediction in combination with clinical factors. The timing of a NAFLD publication is a factor that should be considered. There are increasing rates of NAFLD in pediatrics. Interventions include fibroscan as well as lifestyle interventions.
      * Data: There is concern that there are no valid NAFLD PRS. There is a large GWAS study that is unpublished through the GOLD consortium.
      * Clinical factors currently used to find underlying fatty liver disease are non-specific. They do not progress the advancement of NAFLD to late-stage liver disease. There is an ultrasound-based way of finding the presence of advanced liver disease
      * Mount Sinai has expertise in liver disease genomic discovery expertise.
  + *Mass General Brigham* 
    - Depression
      * Rationale: Depression has a large GWAS available for constructing depression PRS, comparable heritability, major source of morbidity and mortality, and actionable strategies for prevention and intervention. Depression poses the unique opportunity to include a psychiatric phenotype, psychiatric PRS are already being returned to individuals in commercial platforms and it is important to evaluate how health systems and clinicians can return in a more informative manner.
      * Data: The concerns are that the diversity of GWAS ancestry is limited, screening can be associated with stigma, and PRS effect sizes are modest (OR 2-3.5) though comparable with PRS of other common diseases.
    - T2D
      * Rationale: Type 2 Diabetes PRS are one of the best studied instances of chronic disease genetic prediction with large published data sets available for construction of ancestry specific and multi ancestry Type 2 Diabetes PRS.
      * Data: Concerns of Type 2 Diabetes PRS are that studies to date despite their deficiencies increasingly point to “no utility” of providing Type 2 Diabetes PRS and genetic counseling to participants to change Type 2 Diabetes related behaviors or other outcomes.
  + *Mayo*
    - Abdominal Aortic Aneurysm
      * Rationale: Abdominal Aortic Aneurysm is associated with significant cardiovascular mortality and morbidity with around 40% of individuals with AAA diagnosed late in their disease course. There are existing guidelines from the U.S. Preventive Services Task Force (USPSTF) to recommend single abdominal USG screening for men 65-75 years with a history of smoking. In collaboration with the Million Veteran Program (MVP), Mayo derived a 29 SNV PRS (OR 1SD, Top 5% versus remaining around 1.5) independent of family history of Abdominal Aortic Aneurysm and conventional risk factors. The use of PRS can potentially extend current screening guidelines to other individuals who may otherwise not be candidates.
      * Data: There is access to several necessary datasets including summary statistics from large GWAS meta-analysis available, collaboration with AAA Genetics Consortium for a larger meta-analysis, access to UK Biobank summary-level data from external collaborators such as MVP, and individual level data at Mayo Clinical Vascular Disease Biorepository.
    - Coronary Heart Disease
      * Rationale: Coronary Heart Disease is the leading cause of death in the US with nearly 18.2 million adults. The current risk prediction tools have suboptimal accuracy with PRS OR per 1SD around 1.5 and top 5% versus remaining about 3.3. The comprehensive risk estimate combines PRE, family history, PRS, and monogenic risk. The PRS actionability consists of statin therapy, lifestyle modification, prescription medications, and screening tests (e.g. stress test, CAC).
      * Data: There is access to several necessary datasets including summary statistics from large GWAS meta-analysis available, eMERGE III adult cohort, current access to UK Biobank and multiple prospective cohorts from dbGaP such as ARIC, MESA, and JHS; and collaborating with the Million Veterans Program investigators for PRS validation.
  + *Northwestern* 
    - Atrial Fibrillation (AF)
      * Rationale: Afib is associated with significant mortality and morbidity with an increased risk for stroke, heart failure, dementia and death. 1 in 5 ischemic strokes are attributable to cardio embolic events from AF. There is a high prevalence of AF in both whites (15%) and blacks (11%), especially over 65 years of age.
        + In AF, the absolute risk of developing stroke, HF, CAD was greater for African American populations which makes looking at that population specifically important. The SNP based heritability is 22% which is pretty high compared to some other diseases being considered for this round of eMERGE.
      * Pros of the AF Phenotype
        + There is an existing validated PRS which will save the workgroups time. The screening strategies range from minimally invasive (example: history of palpitations) to wearables (example: Apple watch) to more costly devices (example: loop recorders). Management should reduce an individual’s stroke rate, so this phenotype has high clinical utility.
        + Both a pro and a con of AF is that there is no clinical validation of PRS in non-Europeans ancestry which means there is a high eMERGE impact.
      * Cons of the AF Phenotype
        + There are limited GWAS in other ancestries for PRS generation with the exception of 1 multiethnic GWAS with about 9,000 African American/Afro Caribbean, 500,000 European, 5,000 Hispanics, and 36,000 East Asian subjects. This phenotype cannot include LOF TTN(Titin loss-of-function variants, which are significantly associated with AF (odds ratio of 5.35 after removal of HF subjects) which is especially relevant for younger onset AF cases.
    - Prostate Cancer
      * Rationale: Prostate Cancer is the 2nd most common cancer in US males with significant population disparities in incidence and mortality as well as testing uptake. As one of the more heritable cancers, genetic variation explains about 34.4% of the familial relative risk. Surveillance is done via PSA screening which reduces mortality but over screening of low-risk individuals is a concern.
      * Pros of the Prostate Cancer Phenotype
        + There are existing large, multi-ethnic GWAS and risk scores validated retrospectively via RCT and there are currently available Prostate Cancer polygenic risk scores.
        + Low risk, accessible intervention rooted in current guidelines, so no new standards need to be written. There is also an opportunity to improve and encourage testing in high-risk individuals with the return of genetic test results.
        + There has been a lot of clinical validation of scores in cohorts including 7-year studies of cancer incidents and PSA uptake but they have all been in exclusively European ancestry cohorts so there is a high opportunity for eMERGE impact for populations in other ancestries.
      * Cons of Prostate Cancer Phenotype
        + There is reduced power given that this is a male only phenotype.
  + *UAB* 
    - Ischemic Stroke
      * Rationale: Ischemic Stroke is the 5th leading cause of death and leading cause of long-term disability in the United States. African Americans shoulder a greater burden but are not well represented in stroke PRS. On average, the heritability for Ischemic Stroke is around 38%.
      * It is believed that Ischemic stroke would be a good phenotype due to the low amount of data in minorities and the need for validation in these groups.
      * Data: UAB has one of the larger Stroke cohorts with 10,800 African Americans that are already genotyped.
    - Rheumatoid Arthritis (RA)
      * Rationale: Rheumatoid Arthritis is one of the most common autoimmune diseases in adults affecting 1.3-1.5 million people with an estimated over $30 billion a year in health care costs. There are risk reduction measures such as not smoking and screening at risk individuals using a patient-recorded arthralgia inventory.
      * Pros of the Rheumatoid Arthritis Phenotype: There have been some GWAS studies, one recently with a group including both eMERGE data and data from Partners combining five traits including RA.
      * Cons of the Rheumatoid Arthritis Phenotype: There is not a lot of data in African Americans however the CLEAR registry has 1,000 patients and 540 controls of African American individuals that are very well phenotyped.
  + *University of Washington*
    - Colorectal Cancer
      * Rationale: Colorectal cancer is the 2nd leading cause of cancer death. There are highly efficient low-risk screenings (colonoscopy, blood in stool) but about ⅓ eligible people do not get the screening. The population with the highest risk for colorectal cancer is African Americans and this population also has the lowest screening rates. Early onset colorectal cancer (age less than 50 years) increased in the last decade, but this population is not generally eligible for screening. Heritability is about 33%.
      * Data: The best PRS is in European populations. PRS could be used to identify people with similar risk that could identify ~⅔ of all early onset colorectal cancer if screening would be implemented. UW has sizable sample sizes of non-European colorectal cancer cases and controls available for PRS development.
    - Primary Open Angle Glaucoma
      * Rationale: There is a known algorithm developed and implemented in eMERGE II. The Women's Health Initiative and UKBiobank can be used to increase case numbers, but control size is not known. Screening is low risk and readily available. Glaucoma is the second highest cause of blindness and the leading cause of blindness in African populations.
      * Data: There are three GWAS Summary datasets available that include a large number of minorities represented to develop a PRS for both European and non-Europeans.
  + *VUMC*
    - Age-related Macular Degeneration
      * Rationale: Age-related refers to older ages, typically with an onset in 60s and is progressive. It is the leading cause of blindness in older adults. It is highly heritable with a 30-50% CHIP heritability. The highest prevalence in Europeans, second highest prevalence among Asians, lowest prevalence among African Americans.
      * Data: There is a large GWAS of over 20,000 cases and controls among Europeans. Standard clinical models based on risk factors. The phenotype algorithm was developed in eMERGE in 2013 by Marshfield but has never been implemented. Northwestern published an EHR phenotype that they were able to validate.
      * Concerns/Challenges
        + Macular degeneration is much more common in Europeans than in African ancestries.
        + Ophthalmologists are interested in a PRS that identifies people at risk of progression from early stage disease.
    - Bone Mineral Density
      * Rationale: Low bone mass affects 43 million people in the US and associated with a greater risk of fractures. Dual-energy X-ray Absorptiometry (DXA) derived BMD and femoral neck geometry are heritability ranges around 50-80%. There is a known racial disparity in fracture risks with White and Hispanic at the highest risk. Fractures and related conditions can be measured using code-based phenotyping for controls, but a DXA is needed for bone mineral density scores for cases.
      * Data: There are about 20 GWAS published on bone mineral density with the majority in European ancestry populations (the largest comes from the UK Biobank with the most subjects, around 425,000 participants). There is a single published PRS for bone mineral density from the UKBiobank but it has not been independently validated in other cohorts. Potential data resources include eMERGE, BioVU, and UKBiobank.
      * A major strength of this trait is that primary interventions are possible in pediatric and young adult populations focusing on increasing calcium intake and implementing weight-bearing exercises.
      * Major concerns include no validated EHR based phenotype and scores would require extensive phenotyping and GWAS primarily from European ancestry.
    - Hypertension
      * Rationale: Hypertension is a complex trait with many causes, one being genetic exposures.
      * The prevalence of hypertension is quite high in the US with an overall prevalence of 45%. There are racial disparities with hypertension with around 50% of African American having hypertension. There are also racial disparities regarding treatments to control hypertension. There are lower rates of control in Asian populations. Only about 24% of people with hypertension are able to control their hypertension adequately with medications. About 50% of adults with hypertension are either not prescribed or taking medications to control the condition. There are roughly 1300 deaths per day in the US. VUMC has experience working with blood pressure traits using BioVU and Million Veterans Program.
      * Data: VUMC currently has unpublished studies ongoing in about 1 million white individuals. There are about 140,000 African Americans, 28,000 Hispanics, and 130,000 Asians with hypertension in additional databases that are available. These provide substantial resources to pursue PRS validation and evaluation.
      * A concern is that predicting hypertension with a PRS may not be more useful than measuring blood pressure in individuals. PRS is a more stable measure of hypertension liability.
      * Opportunities: Current blood pressure GWAS are pretty large. Blood pressure PRS might be useful in adding predictive performance where blood pressure is in the causal path like stroke. There is good representation among the major global population.

**eMERGE Day 2: FRIDAY August 21st, 2020**

* **Targeted sequencing in prospective cohort | Rex Chisholm(NIH/NHGRI)** 
  + The CC has proposed the use of the Global Diversity Array. The Broad is CLIA/CAP certified. The Broad will generate PRS calculations for all sites. Rapid genotyping will need to be done, and as such, there is a rapid timeline for the first year of eMERGE. The Broad will indicate potential hits on ACMG monogenic targets for follow up sequencing. A third-party vendor will complete follow-up confirmatory sequencing.
  + CCHMC believes that low pass sequencing might be superior to the Global Diversity Array (GDA) because of the ability to find mitochondrial changes and other changes. However, the GDA will be utilized for this project. The Network needs to consider how generalizable results will be with low pass sequencing.
  + *Proposal: There are two options for targeted sequencing*
    - Option 1: Conduct targeted sequencing across the whole prospective cohort for rare variation in monogenic genes (25,000 participants).
      * Determine what to sequence: Tier One only, ACMG, additional genes of interest.
      * This option would include a possible collaboration with Invitae for no cost sequencing.
    - Option 2: Conduct targeted sequencing for specific participants at risk where GRA would be strongly influenced by monogenic sequencing.
      * Participants' sequencing will happen at higher clinical, family history risk for monogenic Breast Cancer, or other Tier 1 conditions.
      * This option would also include sequencing the indicated hits off the array of monogenic variants.
  + Considerations for targeted sequencing of all participants
    - Address additional monogenic return, calling, processing effort
      * Which data to receive back, consideration of negative or VUS results, obligations of return to participants and providers, ACMG conditions not included in our PRS conditions, PGx, etc.
      * The network should also consider integrating data into GRA for all conditions, the complexity of consent, the complexity of EHR integration, and additional costs for genetic counselors/return sites.
      * Invitae: CLIA cap lab, standard work, and reports. No additional costs
      * If collaborating with a for-profit company, privacy, and ethical issues related to downstream data use and data sharing should be considered.
    - Targeted Sequencing
      * Returning PGx and all ACMG could be cost-prohibitive. It may be useful to look at the condition list and determine what genes are needed in the panel.
    - Invitae wants to learn with eMERGE and be involved in discussions. It will need to be considered how education of participants would be conducted, information shared with others, and how providers will react to the shared information. The information that Invitae is willing to share should also be determined.
    - Material Transfer Agreements (MTAs) may be needed with Invitae as well.
    - The group seemed the most interested in option one. It should be tailored to the conditions being studied in eMERGE. Invitae has several large panels and can customize to eMERGE needs.
    - Ethical issues regarding "masking" results should be considered as well as privacy risks. The network should also consider issues with recruitment due to eMERGE working with a for-profit company.
    - Participants have been known to opt-out of receiving new information and their DNA being used in the future.
    - Network will need to consider what is included in the DUA with Invitae.
    - In eMERGE III, a lot of time was lost due to multiple changes made throughout the course of the project, once options are weighed the Network should make a decision regarding targeted sequencing.
  + The data flow between the sites, AnVIL, and Invitae needs to be considered.
  + Invitae can process samples as they are acquired instead of batching if that is useful to the study design. Sites will need to determine how results from both Invitae and the Broad will be compiled. The Network needs to determine if adding targeted genes is a logical decision and is appropriate for a genomic screening risk program.
  + The CC does have some funds available to do a subset of targeted validation sequencing.
  + Helix and Embry have been contacted regarding a no-cost package. Helix seemed particularly interested but they mentioned reaching out to participants which may not be in the best interest of the network. There are also issues regarding contracts with ten different institutions. There could be more companies willing to collaborate with eMERGE for a fee.
  + Invitae contacted the NHGRI regarding working with programs that focus on minority involvement.
  + **ACTION ITEM:** The Network will work with Invitae to determine the flexibility allowed for returning specific subset of results to sites, and what data Invitae would want to receive, to determine if a partnership is beneficial.
  + **ACTION ITEM:** The Network will need to determine how to integrate results from multiple risk report streams, including Broad, FHH, Clinical, and a third-party vendor for monogenic risk.
* **Workgroup Report Outs** 
  + **Genotyping | Meg Roy-Puckelwartz (NU); Niall Lennon (Broad/CC)** 
    - Charter
      * The Genotyping workgroup charter will include information regarding the genotyping reports being returned to the CC, the genotyping sign-off and approval, genotyping validations, and the genotyping platform's validation and launch.
    - Process Inputs
      * Some sites may require the use of an MTA when shipping samples to the Broad. This needs to be determined soon.
      * Sites will need to elect a point of contact for site logistics and sample submissions.
    - **ACTION ITEM:** The CC will collect information regarding points of contact, sample types, order providers, and whether or not samples will be shipped from other enrollment sites and circulate it to the Genotyping workgroup co-chairs.
    - Global Diversity Array
      * Both blood and saliva extracted DNA have been validated on the Global Diversity Array. Site can opt for blood or saliva on the Genotyping workgroup spreadsheet.
    - The Genotyping workgroup will need to collaborate frequently with the PRS workgroup to understand the technical components of the PRS score development. The workflow for PRS development will also need to be confirmed.
    - Technical deliverables
      * The genotyping pipeline will need to be defined. The Broad can incorporate “flags” to identify pathogenic variants.
      * The Broad has assumed that the intended audience for PRS results is researchers/geneticists at sites and not participants.
      * Results are expected to be delivered via the AnVIL.
    - Maximizing the utility and multi-ethnic performance
      * The current Genome build is 37. There will need to be an imputation reference panel included. The PRS and GWAS pipelines at the Broad were built on build 37, in addition to the majority of the PRS calculations published, moving to genome build 38 will come with additional validation issues and should be thought through.
      * Work should be done to identify measurable metrics to answer the major decision points questions. This will be the initial focus of this workgroup.
      * The Broad needs guidance from the PRS group if covariates are incorporated as well as the Phenotyping workgroup.
    - Data delivery cadence
      * Timing of return:
        + Arrays will be run on samples as they arrive on a rolling basis. PRS scores will also be generated on a rolling basis once validated.
  + **PRS validation & evaluation | Eimear Kenny (Mt. Sinai); Patrick Sleiman (Chop)** 
    - The workgroup has approximately 40 members with about half being statistical genetics experts and the other half disease experts with population geneticists included as well. Many have expertise doing GWAS in eMERGE and also in other large-scale consortia. There is also expertise in PRS for conditions, building multiethnic PRS methods, and clinical expertise which are all very valuable.
    - Charter Development
      * It is important to evaluate from literature review/own subject area expertise the scientific validity of extant scores and generating new scores.
      * The workgroup wants to assess what datasets are accessible inside and outside the network and their quality/usability.
      * The PRS workgroup will focus on application and will develop PRS pipelines (model, calibration, etc) while revising existing PRS for diverse populations.
      * The workgroup will need guidance from other groups to develop a framework to help decide if a PRS is suitable and will develop a standardized way to validate PRS and evaluate predictive accuracy.
      * The group will need to discuss with genotyping and phenotyping workgroups how to incorporate non-genetic covariates into PRS models.
      * One of the big considerations for the PRS group that will need guidance from other groups is how to develop a framework to weigh evidence for moving any PRS forward for clinical application.
        + The ClinGen workgroup has built out a framework to evaluate polygenic risk scores, focused more on the framework than the implementation itself, and it will be useful for the groups to work together.
    - Scientific Considerations
      * Standardized Metrics
        + There is currently a lack of standardized metrics in existing literature and the group needs to understand the exact comparisons to, for example, make r2 comparable, among other statistical metrics.
        + The group wants to develop metrics for standardizing what metrics across conditions, age groups, gender groups, etc.
      * Optimized Number of SNPs
        + Part of the work group's job will be to unearth the optimal number of SNPs to be included across conditions, which can be highly variable.
      * Multi-ethnic/Population Specific Scores
        + This is an active conversation within the group thinking about optimizing scores across populations that jointly consider environmental contributions and genetic architecture.
        + Scores may be different depending on the condition and need to be tested while using the framework to decide which is better for the program.
      * Demonstrate Evidence for Decisions
        + It is important to develop a framework to evaluate the totality of the evidence for every PRS-condition pair to determine if they are suitable for clinical applications.
    - Logistical Considerations
      * Covariates
        + For polygenic risk scores, any non-genetic covariates used in GWAS should be used in the PRS application model (age, sex, genetic ancestry PCs, etc.).
        + The group needs a logistical pipeline in place if whether or not covariates are included, they can be executed.
      * Timing and Deliverables
        + The group needs guidance to prioritize efforts and deliver in sync with other workgroups.
        + There is always going to be “bigger and better” GWAS in the works so it will be important to draw a line as opposed to waiting for more data.
        + The group is starting initial validations in September once conditions have been decided upon.
      * Validation Dataset
        + An independent dataset of samples is needed to validate PRS scores, ideally multiethnic scores.
      * Evidence Tracking and Versioning
        + This is a multistep process on handling several approaches with many different methods and choices. It will be important to go back to information years later if needed.
    - Immediate Next Steps
      * **ACTION ITEM:** The group needs to develop a set of guidelines to do comprehensive literature reviews for selected conditions.
      * **ACTION ITEM:** Standardizing effect size and prediction accuracy measures will enable comparison across conditions and populations.
      * **ACTION ITEM:** Additionally, gathering information about details and timing of emerging GWAS/PRS that might be available to the program will be important to know.
      * **ACTION ITEM:** The group will work with the CC to develop a grid of case/control or quantitative trait counts that exist within eMERGE III data.
    - It is important to keep in mind that other groups have data that hasn’t been shared with eMERGE and there may be a need to reach into those to validate polygenic risk scores.
    - Regarding the engagement of the Polygenic Risk Score Diversity Consortium, the timing will likely be a bit far out for direct impact for the group’s work in year one. The consortium is focused on method development particularly for multiethnic populations and it may be possible to integrate the information along the way if not in year one.
  + **Phenotyping | Chunhua Weng (Columbia); Wei-Qi Wei (VUMC)** 
    - Charter Development
      * The core functions of the workgroup are EHR based phenotyping and advancing the science of phenotype algorithm development. Instead of limiting to binary case and control methods, more time will be spent extracting continuous features, calculating probability, and defining time events while moving closer to PRS validation. Additionally, phenotyping will be conducted not only in retrospective analysis (PRS validation and clinical risk assessment) but also in prospective analysis (outcomes).
      * Converting time independent phenotype algorithms to time dependent algorithms will be another focus of the workgroup.
    - Goals of the Workgroup
      * Cross-collaboration with other workgroups and outside stakeholders.
      * The group will align existing algorithms with cohort defined conditions required for PRS validation.
      * The group plans on adopting popular standards like OMOP, knowledge engineering, and machine learning. Sharing algorithms, datasets, and best practices with the public is also a priority.
    - eMERGE Network Milestones/Timeline
      * Since the phenotyping status needs to be delivered before PRS validation can take place, the group feels pressure to tackle their responsibilities as soon as possible. However, there may be gaps between existing algorithms and the features used in the GWAS.
      * Once data is delivered to the PRS group, the group needs to consider how to extract clinical risk factors and build a model for disease prediction.
    - Interactions with the Network
      * The group acknowledges that working with other workgroups is key, especially need PRS group’s immediate help with prioritizing and identifying phenotypes.
      * The phenotyping group also needs to work with the CARE group to identify clinical risk factors that will be needed for extraction and analysis.
      * With the recruitment and protocol workgroup, the group will determine what data needs to be collected from patients directly as well as the data gaps in the EHR.
      * In addition to workgroup collaboration, it will be necessary to work with the CC to refresh data periodically and with the AnVIL team on how to deliver data
    - Phenotype Data Delivery
      * Data was previously uploaded to PheKB which has a beneficial quality check process. The CC is currently looking into utilizing GitHub for data collection as well as looking into quality control tools available on The AnVIL.
    - Phenotype Gap Analysis: Fitness and Robustness
      * For each chosen phenotype, it is important to ask:
        + Is there an existing algorithm for phenotype (on PheKB)?
        + Do phenotypes need to be upgraded?
        + What is the gap between algorithms and criteria used in large GWAS?
      * For the 20 pre-selected conditions, 16 of them have published phenotype algorithms and 15 of those are on PheKB which means they were cross validated by two additional sites.
      * The 4 phenotypes without published algorithms are bone mineral density (may require natural language processing (obesity/BMI, prostate cancer, and hypercholesterolemia (can repurpose familial hypercholesterolemia in PheKB).
    - Implementation Challenges
      * Among the 16 existing algorithms, there are 3 from eMERGE I (T2DM, Crohn’s AAA), 4 from eMERGE II (HTN, Asthma, Atopic Dermatitis, Glaucoma), and 8 from eMERGE III. Some algorithms from eMERGE I and II may need updating or modification. Eighteen algorithms need updating and 5 already have an OMOP query for implementation. Additionally, 6 algorithms require NLP, 11 have case/control data available, and 13 collect relevant covariates.
    - A phenotype algorithm network-wide implementation readiness table can be found [here](https://docs.google.com/document/d/1R040MuWOWO5WSvop0H_WzvmQ3Yy1WYufAUDosUaiZAU/edit?usp=sharing) to be used when determining which conditions may require more work to validate.
    - **Action Items**
      * The phenotyping workgroup will collaborate with other groups to identify covariates and risk factors for the selected conditions over the next month.
      * The phenotyping workgroup will evaluate existing algorithms and their fitness and robustness. The group will also prioritize algorithm development by complexity and urgency in the next quarter.
      * The phenotyping workgroup will establish best practices, especially for OMOP and NLP work.
      * The phenotyping workgroup will work with the CC to standardize the process to collect requests and deliver results (via PheKB and/vs AnVIL).
* **Workgroup Report Outs**
  + **EHR workflow & infrastructure | Luke Rasmussen (NU); Robert Freimuth (Mayo)** 
    - Prospective GRA data flow
      * EHRI are primarily involved in the Risk Report, how AnVIL and the R4 Portal will interface, and how Sites will utilize the R4 Portal will work together.
      * EHRI will also need to be involved in discussions for data standardization and the communication mechanisms that occur between these entities.
      * EHRI will also be involved with capturing EHR data needed for outcomes.
    - Recruitment, Results, and Risk Reduction (R4 ) Portal
      * Workflow considerations will need to take place, but EHRI is not concerned with individual workflows but how data is structured and represented.
      * EHRI also understands that sites may diverge slightly in terms of sending and receiving data and utilizing the workflow in practice.
    - eMERGE Network milestones
      * The work tasked to EHRI is heavily front loaded in the first 2 years.
        + This work is directly related to enabling the transmission of data between sites.
      * Milestones are mostly in year two.
        + The EHRI work needs to start now to accomplish milestones by year two.
        + EHRI needs to work with other workgroups immediately on what data elements that will be exchanged.
    - Recommendations from eMERGE III
      * An early start on determining infrastructure and how standards can best fit in necessary due to the aggressive timeline.
    - Major areas of work
      * **ACTION ITEM:** EHRI co-chairs examine sites ability to automatically transfer data for outcomes analysis from the EHR.
      * Clinical and research context of the workflow will need to be continuously clarified.
      * The major underlying theme for these areas of work is data and interoperability standards.
    - EHR Workflow and Infrastructure Workgroup Charter
      * Network: Coordination among stakeholders will need to be pragmatic and innovative.
      * Standards: Develop and/or adopt existing standards for data interoperability across a heterogeneous Network.
      * Community- Build and foster relationships, not only in this Network but with others that are doing similar work.
    - Activities and Proposed Timeline (many of these steps will be done in parallel)
      * Timeline of activities in 2020 include:
        + Identify key collaborators across workgroups
        + Refine scope of primary EHRI activities
        + Determine Minimal Viable Product (MVP)

The MVP is the version of a product that provides the basic features needed for consumers in the early stages of the product.

EHRI will be looking for sites and workgroups to establish boundaries around what is already known (genotyping data, PRS).

Working on the known parts (genotyping, PRS) now will form the core part of the MVP.

* + - * + Define requirements for MVP
        + Document anticipated clinical workflows, data flows, and system architecture at sites
        + Document data flows and system architecture among sites/network
      * Timeline of activities in 2021 include:
        + Identify standards needed for MVP (Q1)
        + Establish change management process (Q1)

This process will minimize the impact to sites when changes need to be made.

* + - * + Adopt, adapt, and/or develop standards
      * Timeline of activities in 2022 and ongoing include:
        + Implementation
        + Feedback to and collaborate with SDOs
        + Refine requirements as needed
        + Refine workflows, data flows, and system architecture as needed
    - EHRI Stakeholder Collaborations
      * Initial questions involve the entirety of the workgroups.
      * ELSI components regarding the implications of returning or not returning data need to be included.
  + **Comprehensive risk assessment & return | Gail Jarvik (UW); Iftikhar Kullo (Mayo), Cindy Prows (CCHMC)** 
    - The major areas of work for the workgroup are family history data collection, genomic risk assessment (GRA) clinical data element integration, GRA development harmonization, confirming monogenic disease risk integration, adopt validated PRS into GRA, return and management of risk recommendations, return of risk profiles (PRS, Mendelian, family history, clinical risk) and timing, and AnVIL cloud based tools (SMART on FHIR), ANVIL FHIR to LIMS systems for GRA.
    - To complete the genomic risk assessment, the group will develop a single linear equation of covariates (e.g. age, sex, family history, BMI) and genetic variants.
      * Different PRSs will have different covariates and different collinearity between covariates and SNVs. The group will have to decide a weight for each covariate and each risk factor. The risk to be returned could be the linear combination of covariates and PRS, but genotyping centers only return a PRS report. How to return the risks will need to be addressed.
    - The group will decide if PRS should be the best PRS for each ancestry or a single PRS that best captures all ancestries.
    - The data needed for a comprehensive risk estimate are a clinical risk equation, family history (MeTree), the baseline survey created from the final phenotypes, EHR extraction of risk factors, rare variant assessment.
      * For pediatric sites enrolling newborns and infants it is unlikely that there will be sufficient EHR risk factors for extractions.
      * Rare variant assessment will be needed for breast cancer, colon cancer, and family hypercholesterolemia. Pediatric sites will decide if they should purposely avoid sequencing genes for adult onset Mendelian risk in pediatric participants.
      * The data will best estimate relative risk, but absolute risk is more useful.
    - For the return of results, the group will need to decide how the reports are generated, if low risk results will be returned, the method of return (e.g. by EHR, medical doctor, genetic counselor, study coordinator), and if Mendelian results will be returned to deceased participants.
    - Clinically tested positive monogenic risk for selected phenotypes of eMERGE III participants GRS will be excluded. Participants with prior testing for a relevant condition with a negative result will not be excluded as they may have a high PRS. There is a concern over biases of excluding only those with positive results and the group expects to identify some previously unknown Mendelians. Population-based positive Mendelian results will not be excluded from the study but rather tracked for outcome work. Those with recognized clinical diagnosis of one or more of the chosen phenotypes will be excluded.
    - Cindy presented the pediatric site enrollment and return considerations.
      * CCHMC would like to enroll mother and newborn dyads. Parents will provide permission for adolescent mothers and mothers will give permission for newborns. Adolescent mothers will need to re-consent at age 18. Newborn’s cord blood will be collected if possible. The dyads allow for a convenient time for low resourced families to participate in the initial study visit. CCHMC will follow guidance for post-partum patients and 0-3 year well child clinical follow ups which will provide several opportunities for return of result and capture of clinical outcomes.
        + CHOP is considering other age groups and possible dyads and VUMC has proposed to enroll 200 children in addition to their adult cohort.
    - All newborns will have had prior metabolic and genetic screening so this should not be used as exclusion criteria.
      * The workgroup will decide whether to withhold adult onset conditions from pediatric participants. The timing of return will include consideration of childhood growth and development as well as post-partum timing. Pediatric sites and Ingrid Holm will prepare recommendations for pediatric enrollment and return.
    - The tentative workgroup timeline includes phenotype list finalization and mock reports for each phenotype in year 1. Year 2 and onwards PRSs finalized for each phenotype and strength of association reviewed to decide how to incorporate into the risk report.
    - **ACTION ITEM:** The workgroup will create a subgroup focused on the return of results including low risk results.
    - Reports will contain wording about what is being tested for accepted conditions with those at high risk to be highlighted. It will be a hard decision to make on the definition of high and low risk cutoffs. Report will provide information on what makes participants at risk.
  + **Provider uptake & outcomes | Noura Abul-Husn (Mt. Sinai); Nita Limdi (UAB)** 
    - The workgroup goals are to collaborate with other workgroups on the study design, harmonize a baseline participant survey, engage providers at baseline to assess knowledge and attitudes towards PRS, and establish short-term, intermediate-term, and long-term outcomes for both participants and providers.
    - The components needed for the study design will be the hypothesis, expectations and study design.
      * The group will decide how to define the value that PRS adds to the clinical risk prediction for common diseases. This should be viewed as a prospective cohort study.
    - The workgroup has begun collating measures and survey sources to assist with the participant baseline measures.
      * Group has begun to collate measures and survey sources to collect factors such as socioeconomic status, physical activity, environmental measures, geocoding, health insurance, and health literacy. MacArthur Network on SES and Health is an 11 question on socio-demographic questionnaire. Stanford has a brief physical activity survey. SF-36 (or SF-12), Euro-Quality of Life (QoL) 5D, PROMIS are a few quality of life surveys that could be included.
      * Baseline surveys will be done at enrollment and the group is deciding how they will be administered either remotely or with a coordinator. An outcomes survey will be completed at a later time point or will rely on EMR data. The survey will incorporate eMERGE-3 participant and provider surveys and other validated surveys (e.g. PhenX Toolkit, HINTS, FACTOR, MI-GENE trial, CSER).
    - Outcomes analysis will include short term (e.g. patient and provider understanding and intention to follow up), intermediate term (e.g. patient engagement and adherence to follow up), and long term (e.g. patient uptake of intervention and clinical outcomes).
    - Outcomes will be dependent on the conditions assigned and the polygenic risk scores.
* **Workgroup report out & sIRB framework| Digna Velez-Edwards (VUMC); Ingrid Holm (BCH); Wendy Chung (Columbia)** 
  + Initial goal is to have the sIRB submitted in the next three months.
  + **Action Item:** The CC will create a [decision log](https://docs.google.com/spreadsheets/d/1BPJ9gG2n2LE7S2_fugDsb5cGSTSVEnVyAx6KqqtrIG0/edit#gid=480715247) to track Network decisions.
  + The main project ELSI should address the issues with equity and inclusivity due to participant focus.
  + sIRB
    - Inclusion/Exclusion: Several issues will need to be defined around age limits (lower and upper), genetic testing naive, and if participants need to have a designated PCP, among others. Harmonization of recruitment methods as much as possible will need to be prioritized.
    - Use of social media, online recruitment, emails/paper/flyers, and clinics for recruitment can be inventoried across the Network.
    - Participant education methods are another area that needs considerable discussion.
    - The Informed Consent method(s) should involve several modalities; electronic, interactive, A/V (English and Spanish).
    - In order to try to make sample collection as easy for participants as possible, information regarding sample type will need to be collected from the sites.
    - Return
      * Decisions on the following topics need to be finalized before the sIRB can be written:
        + Sequence analysis in addition to genotyping, polygenic risk/continuous risk/integrated risk, dichotomous or continuous, can results be tailored for children for certain types of results, return of results for PRS scores not currently validated for some races/ethnicities, if a participant “opt out” of some results and choose to get others, if the Network decide to provide updated results as they are generated, and if PRS be valuable for all groups.
      * Engagement works best when the time gap isn’t too large from enrollment and when results are returned. Multiple timepoints due to staggered returns or updates puts a burden on study teams. Addressing deceased participants will be needed.
      * Other considerations on return of results include: Who will return the results, what are the methods of return, are returns based on risk associated, are there novel technologies for return.
    - sIRB Data to be collected include:
      * Demographics, EHR: baseline, family history from MeTree, post-result, baseline questionnaires, participant surveys, Health Care Provider surveys.
      * The group will collaborate with Provider Uptake and Outcomes workgroup should be coordinated to help address some of these issues.
  + ELSI- year 1 - Site specific
    - In order to maximize usefulness to the network, harmonization of ELSI projects across sites could help to provide stronger evidence by sharing info or study materials and determining common questions across sites. The sites should consider the need to coordinate ELSI projects as soon as possible.
  + ELSI Network study
    - Work with the other workgroups to identify important questions regarding equity and responsibility. Take into consideration what data elements to collect to answer Network-wide ELSI questions. The questions addressed need to be considered in collaboration with the other workgroups.
* **Initial PRS condition ranking | Rex Chisholm** 
  + Based on the evidence, the review results from the network indicated good coverage across all conditions. The goal from the RFA was to have 15 phenotypes in which PRS scores can be developed across the network.
  + Sites provided feedback on several of the lower ranked conditions.
    - Type 1 Diabetes Mellitus is one of the few phenotypes with a good genetic risk score that works across multiple ethnicities.
    - Hypertension could be a condition relevant to the pediatric population for determining risk as an adult. It is very easy to measure, highly prevalent, and has a major public health impact.
    - Depression is a condition that could have a major public health impact.
    - The group raised concerns regarding how genetic risk could aid in the treatment of Asthma. The group suggested that more aggressive treatment therapies could be used with those participants with a high PRS. Higher PRS may also indicate more advanced disease. It should be considered whether or not treatment will prevent chronic lung disease.
  + Sites suggested that separate rankings be used for adults and pediatrics based on the condition.
  + The network decided to do a second round of evidence review and to consider the multi-ethnic component in each condition.
  + Leadership also decided to keep all ranked conditions except Hypertension, Age-Related Macular Degeneration, Crohn's Disease, Depression, Rheumatoid Arthritis, Asthma, and Atopic Dermatitis.
  + After the discussion of conditions, sites were asked to indicate their new 1st through 4th phenotype choices and submit them to the CC. Sites were reminded that if there is a condition that they would like to do and they are willing to do the work, then there's a possibility that this condition can be used.
  + Sites were reminded that by ranking a condition, they agree to be responsible for PRS validation and future GRA development as needed.

**ACTION ITEMS**

**Network/CC**

* The CC will create a [decision log](https://docs.google.com/spreadsheets/d/1BPJ9gG2n2LE7S2_fugDsb5cGSTSVEnVyAx6KqqtrIG0/edit#gid=480715247) to track Network decisions.
* The Network will work with Invitae to determine the flexibility allowed for returning specific subset of results to sites, and what data Invitae would want to receive, to determine if a partnership is beneficial for both parties.
* The Network will need to determine how to integrate results from multiple risk report streams, including Broad, FHH, Clinical, and a third-party vendor for monogenic risk.

**R2, ELSI, sIRB**

* Workgroup co-chairs (Digna Velez-Edwards, Ingrid Holm, Wendy Chung) will form an ELSI subgroup to work on harmonizing ELSI projects to identify a common Network ELSI project.

**EHRI**

* EHRI co-chairs (Luke Rasmussen and Bob Freimuth) will assemble a method to capture [outcome data](https://drive.google.com/drive/u/1/folders/1VdKfuFLMyXCE5n9XzxFdX-dSSWcPdBB6) by 9/24/2020.

**Genotyping**

* The CC will collect information regarding points of contact, sample types, order providers, and whether or not samples will be shipped from other locations and circulate it to the Genotyping workgroup co-chairs.
* Sites whose institution requires an MTA to transfer genomic DNA should notify the Meagan Harden (mharden@broadinstitute.org) at the Broad.

**Phenotyping**

* The phenotyping workgroup will collaborate with other groups to identify covariates and risk factors for the selected conditions over the next month.
* The phenotyping workgroup will evaluate existing algorithms and their fitness and robustness. The group will also prioritize algorithm development by complexity and urgency over the next quarter.
* The phenotyping workgroup will establish best practices documents for things such as OMOP and NLP work.
* The phenotyping workgroup will work with the CC to standardize the process to collect requests and deliver results (via PheKB and/vs AnVIL).

**CARE**

* The CARE workgroup will create a subgroup focused on the return of results including low risk results.

**Provider uptake & Outcomes**

* The Provider uptake & Outcomes workgroup will collaborate with other workgroups on the study design.