|  |
| --- |
| **eMERGE Network: Manuscript Concept Sheet** |
| **Reference Number** *(to be assigned by CC)* | NT443 |
| **Submission Date** | 2/17/2022 |
| **Project Title** | Clinical implementation of polygenic risk scores for 10 conditions |
| **Tentative Lead Investigator** *(first author)* | Niall Lennon |
| **Tentative Lead Investigator Email Address** | nlennon@broadinstitute.org |
| **Tentative Senior Author** *(last author)* | Eimear E. Kenny |
| **All Other Authors**  | Patrick Sleiman, Maegan Harden, Megan Roy-Puckelwartz, other co-chairs, leadership, PIs; members of PRS/Genotyping wg interested in joining the writing team.  |
| **Sites Participating** | eMERGE IV sites |
| **Background / Significance** | Recent accelerations in the genomics field mean that polygenic risk scores (PRS) are available for a wide array of traits and conditions, offering many potential applications to preventative medicine. However, a clear picture has yet to emerge on how to PRS may be integrated with traditional risk information, and of the clinical utility of the return of PRS to patients. There is also a growing concern that clinical use of PRS could contribute to health disparities due to the poorer performance of PRS in non-European ancestry individuals. Evidence to guide the implementation of PRS in clinical care is lacking, and most studies so far have focused on applications for specific conditions. The eMERGE genomics risk assessment and management network was funded in summer 2020 with the goal of integrating PRS in clinical care and to measure health outcomes in diverse populations. Over the next 24 months, the Network will recruit 25,000 children and adults of diverse ancestry, prospectively calculate their genomic risk for selected conditions, return risk estimates, deliver management recommendations to participants and providers in clinical care settings, and measure outcomes. This paper will describe our experiences selecting PRS for 10 conditions, optimizing those scores for portability in diverse populations, validating them in a CLIA/CAP laboratory setting, and implementing them as clinical PRS tests. We will discuss the trade-off’s in robustness and accuracy, limitations of data access, and considerations for implementation in a clinical setting.  |
| **Outline of Project** | * Describe the landscape audit of PRS for 17 conditions deemed of interest to the eMERGE Network (i.e. SNP-based h^2, PRS availability, phenotype definition, clinical actionability/utility, FH vs PRS risk, balance between pediatric and adult conditions, published/validated vs emerging PRS)
* Describe PRS grid/evaluation tool and considerations informing the trade-off of PRS robustness vs accuracy, including; multiple metrics (AUC, OR, PPV/NPV etc), evidence for performance in African American and Hispanic/Latino populations, prevalence in general populations, PRS in the context of traditional risk factors (how to handle covariates or if integrated clinical risk equation exists), and other factors impacting selection of final 10 conditions
* Strategies to expand portability of PRS to diverse populations (i.e. trans-ancestry algorithm vs fine-mapping, discovery/validation in more populations). Focus on evaluation/optimize models for diverse populations, rather than develop new scores.
* Portability to pediatric populations (BMI, Asthma, T1D, T2D), accounting for differences in phenotyping, ethical concerns.
* Describe implementation pipeline, and how missingness/technical artifacts are accounted for, impact of choice of array/imputation panel during implementation (data access), balancing cost vs technical accuracy
* Observing differences in variance in PRS distribution (particularly at tails) and appropriately adjusting for genetic ancestry. Also, choice of ref panel and statistical adjustment, issues of data access etc
* CLIA- and NYS-test validation framework, what documentation is necessary, what are the considerations developing PRS reports (i.e. population group labels, regulatory compliance)

Figures/Tables include:PRS GridSchema of selection processStatistical adjustment for genetic ancestry |
| **Desired Data - Common Variables\*** *(Available from the CC)* | [x] Demographics [ ] ICD9/10 codes[ ] CPT codes[ ] Phecodes[ ] BMI | [ ] Common Variable Labs[ ] Common Variable Meds[ ]  Geocoding 2015 ACS variables’[ ] Other: Case/Control status on Phase I and Phase II phenotypes |
| **Other Desired Data *(Available from participating sites)*** | We have all of the data needed for this project.  |
| **Desired Genetic Data** | [x] eMERGE I-III Merged set (HRC imputed, GWAS)[ ] eMERGE PGx/PGRNseq data set [ ] eMERGEseq data set (Phase III)[ ] eMERGE Whole Genome sequencing data set[ ] eMERGE Exome chip data set[ ] eMERGE Whole Exome sequencing data set[ ] Other (not listed above): |
| **Does project pertain to an existing eMERGE Phenotype?** | [x] Yes, if so please list BMI/Obesity, AFib, BC, PC, CHD, HC, T2D, T1D, Asthma, CKD[ ] No |
| **Planned Statistical Analyses** |  |
| **Ethical Considerations** | None |
| **Target Journal** | GIM or ASHG and/or special issues |
| **Milestones***(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | * Complete analyses – April 2022
* Draft manuscript to share with coauthors – May 2022
* Submit manuscript – June 2022
 |

**\*Common Variables available across all datasets:**

* Demographics: sex, year of birth, decade of birth, race, ethnicity
* Codes: (repeated values & age at event): ICD, CPT, Phecodes
* BMI: (repeated value & age at event) height, weight, BMI
* Labs: (lab name, repeated lab value & age at event) Serum total cholesterol, LDL, HDL, Triglycerides, Glucose fasting/non-fasting/unknown, & White Blood Cell count
* Medications: (medication name, repeated, & age at event) Cerivastatin sodium, Rosuvastatin, Simvastatin, Fluvastatin, Pravastatin, Lovastatin, Atorvastatin, & Pitavastatin
* Other: Case/Control status on Phase I and Phase II phenotype: only on GWAS dataset participants