|  |  |  |
| --- | --- | --- |
| **eMERGE Network: Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* | NT357 | |
| **Submission Date** | September 10, 2019 | |
| **Project Title** | Lessons from the eMERGE Network: Balancing genomic discovery and implementation science | |
| **Authorship approach** | In addition to the overall authorship list we will include lead and senior authors for major subgroupings in a similar fashion to the Harmonization paper that was recently published <https://doi.org/10.1016/j.ajhg.2019.07.018>.  *Contacts: Drs. Jackson & Crosslin* | |
| **Tentative Lead Investigator** *(first author)* | The eMERGE Consortium,  CC representative, Genomics WG representative, | |
| **Tentative Senior Author**  *(last author)* | The eMERGE Consortium,  CC representative, Genomics WG representative, | |
| **All Other Authors** | The eMERGE consortium, all other relevant members. | |
| **Sites Participating** | All eMERGE sites | |
| **Background / Significance** | This manuscript is intended to build on previous Network wide manuscripts [Kho et al 2011](https://www.ncbi.nlm.nih.gov/pubmed/21508311), [Pathak et al 2011](https://www.ncbi.nlm.nih.gov/pubmed/21597104), and [McCarty (phenotype) et al 2011](https://www.ncbi.nlm.nih.gov/pubmed/21269473); Phase II [Gottesman et al 2013](https://www.ncbi.nlm.nih.gov/pubmed/23743551) and [Crawford et al 2014](https://www.ncbi.nlm.nih.gov/pubmed/24987407), while describing the lessons we learned from our work on discovery based genomics data sets. The manuscript will recap lessons from Phase I and II, then focus on how discovery and implementation research are being produced off our newly created datasets, lessons learned from those endeavors, and how lessons around our continued production of large discovery based datasets (GWAS) for research. The manuscript will also describe the Network including sites, goals, workgroups, and generally discuss how this phase focused on both implementation and discovery-based research. We will be referencing current and past lessons learned manuscripts; however, the aim is not to overlap or impinge on manuscripts and lessons learned manuscripts already in process or being planned by the various workgroups. | |
| **Outline of Project** | History of Network to date   * Phase I & II - hypothesis driven - secondary use of data for EMR data with genomics   + Referencing past lessons learned articles ([McCarty 2011](https://www.ncbi.nlm.nih.gov/pubmed/21269473); [Pathak et al 2011](https://www.ncbi.nlm.nih.gov/pubmed/21597104); [Kho et al 2011](https://www.ncbi.nlm.nih.gov/pubmed/21508311); [Gottesman et al 2013](https://www.ncbi.nlm.nih.gov/pubmed/23743551), [Crawford et al 2014](https://www.ncbi.nlm.nih.gov/pubmed/24987407);[Crosslin 2015](https://www.ncbi.nlm.nih.gov/pubmed/26221186))   + Evolution of security   + Link to discovery in eII & how PGRNseq supplement shaped implementation & ROR   + Impact on translational medicine to date   + High level goals of eIII & study design & transitioning from Phase II, including adding diversity age (peds & adults) and race/ethnicity (Arizona & Meharry) * Phase III: Implementation with discovery   + Description of current Network, sites, goals, & workgroups     - Experiments of nature - site specific protocols shape the cohort   + Creation of new data sets for phenotypic discovery (phenotype and dataset utilization references)     - Enhancement of phenotypes focusing on implementation     - Utilization of eMERGEseq for discovery in addition to implementation (reference [Harmonization](https://www.sciencedirect.com/science/article/pii/S0002929719303015?via%3Dihub) paper 2019).     - Paving the way for multiple types of research off one dataset.     - Briefly mention ROR & Outcomes references/publications   + Network operations     - Publication policy, data sharing, equity of access     - Phenotypic data:Common data models ([OMOP](https://www.ncbi.nlm.nih.gov/pubmed/31325501) paper), variables, dbGaP submissions,     - Collaborations with external groups * Lessons learned   + Genomics lessons learned     - Balancing discovery research with implementation, site enrollment bias, diversity.       * Parallel retrospective array data collection, imputation, and aggregation in addition to prospective sequencing efforts to support discovery based research     - Logic to data freezes and set release to Network, both phenotypic & genomic data     - Resource requirements for large datasets (future of cloud computing)   + Links to other lessons learned papers from eIII     - ([Antommaria et al 2018](https://www.ncbi.nlm.nih.gov/pubmed/30240342); [Fossey et al 2018](https://www.ncbi.nlm.nih.gov/pubmed/29301385); [eMERGE Consortium 2019](https://www.sciencedirect.com/science/article/pii/S0002929719303015?via%3Dihub); [Williams 2018](https://www.ncbi.nlm.nih.gov/pubmed/?term=Harmonizing+Outcomes+for+Genomic+Medicine%3A+Comparison+of+eMERGE+Outcomes+to+ClinGen+Outcome%2FIntervention+Pairs); [Hripscak 2019](https://www.ncbi.nlm.nih.gov/pubmed/?term=Facilitating+phenotype+transfer+using+a+common+data+model.), [Rohrer Vitek 2017](https://www.ncbi.nlm.nih.gov/pubmed/28639489); [Herr 2019](https://www.ncbi.nlm.nih.gov/pubmed/30590574)) and others * Future efforts   + AnVIL and cloud computing   + High level penetrance?   + Collaborations & shaping practice for other networks and clinical research   + Implications for future rounds   + High level eIV where the network is going - RFA on PRS/GRA and implementation     - Science of implementation | |
| **Desired Data - Common Variables\***  *(Available from the CC)* | Demographics  ICD9/10 codes  CPT codes  Phecodes  BMI | Common Variable Labs  Common Variable Meds  Other: Case/Control status on Phase I and Phase II phenotypes  **NONE** |
| **Other Desired Data *(Available from participating sites)*** | *Please specifically list out any data elements that participating sites would collect or extract from clinical or other sources for this project (i.e. not common variables above)*  **NONE** | |
| **Desired Genetic Data** | 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): **NONE** | |
| **Does project pertain to an existing eMERGE Phenotype?** | Yes, if so please list  No | |
| **Planned Statistical Analyses** | **NONE** | |
| **Ethical Considerations** | **NONE** | |
| **Target Journal** | Genetics in Medicine  Science of Translational Medicine  BMC Med Genomics | |
| **Milestones**  *(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | MCS submitted & accepted - September 2019  Draft of MCS to Network - November 2019  Submission to journal - January 2020 | |

**\*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