|  |
| --- |
| **eMERGE Network: Manuscript Concept Sheet** |
| **Reference Number** *(to be assigned by CC)* | NT337 |
| **Submission Date** | 3/14/19 |
| **Project Title** | Gender Genotype Analysis for Clinical Quality Assessment |
| **Tentative Lead Investigator** *(first author)* | Jianhong Hu |
| **Tentative Senior Author** *(last author)* | Donna Muzny |
| **All Other Authors**  | Viktoriya Korchina, Eric Venner, Mullai Murugan, Richard Gibbs, David Murdock, Eric Boerwinkle, Alyssa Macbeth, Steven Harrison, Heidi Rehm, Hana Zouk, Niall Lennon |
| **Sites Participating** | Baylor College of Medicine, Partners/Broad, Affiliated eMERGE sites (Vanderbilt, CHOP, Columbia, Meharry, Mayo, Marshfield),  |
| **Background / Significance** | The tracking of sample-assigned gender by genotype is a precise and accurate measure of process fidelity that is underrecognized and underutilized. |
| **Outline of Project** | Genotyping of key sites in samples from eMERGE sites is routinely carried out as an independent process control step, aside from the examination of records and sequence data from other tests. Inconsistencies of gender assignment are able to be identified and tracking these events monitors and improves process control.The Human Genome Sequencing Center has provided clinical testing for 14,500 of the 25,000 eMERGE III samples and applied gender consistency tests for each. Initial inconsistencies have been re-tested and reexamined at the testing laboratory and the clinical collection sites have been surveyed for possible explanations. Approximately ½ of the inconsistencies have been resolved as sample tracking errors at collection, while the remainder have been revealed to be due to uncommunicated bone marrow transplant or transgender status. |
| **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)*** | Communications regarding gender status inconsistencies.  |
| **Desired Genetic Data** | [ ] eMERGE I-III Merged set (HRC imputed, GWAS)[ ] eMERGE PGx/PGRNseq data set [x] 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 Gender [ ] No |
| **Planned Statistical Analyses** | N/A |
| **Ethical Considerations** | N/A |
| **Target Journal** | Genome Medicine |
| **Milestones***(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | 90% of data are already collected. Manuscript to be developed by May 31st, Submitted June 1. |

**\*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