|  |
| --- |
| **Nt3eMERGE Network: Manuscript Concept Sheet** |
| **Reference Number** *(to be assigned by CC)* | NT394 |
| **Submission Date** | June 3, 2020 |
| **Project Title** | Neptune: An environment for the delivery of precision medicine |
| **Tentative Lead Investigator** *(first author)* | Eric Venner |
| **Tentative Lead Investigator Email Address** | venner@bcm.edu |
| **Tentative Senior Author** *(last author)* | Richard Gibbs |
| **All Other Authors**  | Victoria Yi, Mullai Murugan, other authors from sites who designed and implemented components of Neptune |
| **Sites Participating** | BCM HGSC and all other interested eMERGE sites |
| **Background / Significance** | The delivery of precision medicine in a clinical setting demands high-throughput clinical reporting, but these activities are currently limited by an expensive manual review process, challenges handling protected health information (PHI), and managing report customizability. Further challenges, including a low rate of actionable findings, unfamiliarity of providers with genetic data and the reluctance of payers to support precision medicine demand more data be collected to make the case. Many national, large-scale clinical sequencing projects have stepped up to fill this need, including eMERGE, All of Us, and CSER as well as a large number of private and regional initiatives.In support of these high-throughput clinical sequencing projects, we have developed Neptune, a pipeline that enables users to identify and report known disease-causing variants in gene sets of interest, to gather curations for potential novel pathogenic variants from an external review system, and to share VIP variants as desired with collaborators and clinical partners. An expert reviewer, upon receiving quality-controlled sample data, is presented with a highly filtered list of genomic variants to review to interpret according to ACMG guidelines [ref]. Once variant review is complete, reports are automatically generated and made available for approval by a laboratory director. Our system relies on the HIPAA-compliant DNANexus environment to store all PHI data and can operate as a hybrid system, with non-PHI elements running outside of DNAnexus.The key features of Neptune are to: 1) take as input genomic data and compare against a ‘VIP database’ of known genetic variation, marking known variants with previously-curated data and selecting novel genomic variants for review 2) to combine data from diverse sources including sample metadata from a LIMS and variant information from the VIP database and output data in a structured report file 3) convert that structured data into a customizable human-readable report, 4) enable corrected and updated reports 4) Handle the reanalysis and re-interpretation of genomic data over time. |
| **Outline of Project** | In this paper, we describe the infrastructure we built to support eMERGE III and other clinical reporting projects. Specifically, this paper will cover:* The approach for variant filtering and annotation with previous variant curation data
* Reporting of copy-number variation
* Site-dependant pharmacogenomics reporting
* Configurable / customizeable report templates
* Structured data format conversions
* Polygenic risk score reporting
* Report tracking and verification with ARBOR
* Tools for reanalysis and variant watchlists
* A case study describing EMERGE III reporting
* A case study describing reporting for the BCM HeartCare test
 |
| **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 |
| **Other Desired Data *(Available from participating sites)*** | Summary data from eMERGE III clinical reporting efforts, including variant review statistics, reanalysis summaries and other data pertaining to developing and issuing customized reports. |
| **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): |
| **Does project pertain to an existing eMERGE Phenotype?** | ☐Yes, if so please list: . 🗹 No |
| **Planned Statistical Analyses** |  |
| **Ethical Considerations** | None. |
| **Target Journal** | JAMIA |
| **Milestones***(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | March - May 2020: sites complete implementation of phenotypes not already implementedMarch - May 2020: sites assess how they developed, tested, trained, validated &/or implemented the NLP/ML algorithms using the defined metricsMay 31, 2020 – 1st draft with lessons from primary & secondary sites on development & validation, and initial lessons from all sites on implementationsJune 4-7: 1st full draft (~1 week after sites complete implementations)June 21, 2020: 2nd / final draftJune 30, 2020: submit to journal |

***\*\**** *This section should include the timeline for completion of project, including: approval, project duration, first and second draft of the paper and submission.*

**References**

Denny JC, Spickard III A, Johnson KB, Peterson NB, Peterson JF, Miller RA. Evaluation of a method to identify and categorize section headers in clinical documents. Journal of the American Medical Informatics Association. 2009 Nov 1;16(6):806-15.