|  |  |  |
| --- | --- | --- |
| **eMERGE Network: Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* | NT336 | |
| **Submission Date** | 3/11/2019 | |
| **Project Title** | Detection of Mosaic Variants in the eMERGE and MyCode Projects. | |
| **Tentative Lead Investigator** *(first author)* | Sara Kalla | |
| **Tentative Senior Author**  *(last author)* | Eric Venner | |
| **All Other Authors** | Joshua Traynelis, Richard A. Gibbs, Georgia Wiesner, Marc S. Williams, Melissa Kelly | |
| **Sites Participating** | eMERGE sites who submitted sample data for sequencing on the eMERGE 109 gene panel. | |
| **Background / Significance** | Post-zygotic mutations (PZM) can lead to distinct populations of cells within an individual, and these ‘mosaic’ variants can play a contributing role in human diseases such as rare-neurodevelopmental disorders and cancers.  However, the detection of mosaic variants remains challenging[1] and their contributions to disease pathogenicity are not fully understood. Studies have estimated that 7.5% of de novo mutations occurring in Austism Spectrum Disorder are PZMs[2] and blood-specific clonal mutations have been identified at a greater rate of 5-6% in older individuals [3], but the rate of mosaic variants in the general population is not clear. Methods for detection of mosaic variants have utilized the expected ratio from heterozygous mutations and measure potential deviations in the VAF [4].  Clonal hematopoiesis (CH) has been found to be much more common in the elderly than previously realized. This detected form of CH has been associated with increased mortality, smoking and different phenotypic characteristics. Several studies have also shown that somatic mutations of certain genes can be associated with CH [4]. It was proposed that neutral drift plays a contributing role by acting on a smaller population of Hematopoietic Stem Cells (HSCs) [4,5]. Studies have also linked CH (or CHIP) with myelodysplastic discorders and leukemias [5].  For the eMERGE project, we developed a targeted gene panel of 109 genes. This panel was sequenced to higher coverage than a standard genome or exome, which presents an opportunity to examine the detection of mosaic variants while leveraging population data, and estimate the prevalence of mosaic variants across a larger population.  The MyCode Community Health Initiative is a large clinical research project at Geisinger. Participants consent to linkage of electronic health record data to genomic information that includes high density genotyping and exome sequencing. The consent permits broad research use and, where applicable, return of results deemed to be medically significant. The Geisinger eMERGE participants are taken from the larger MyCode population. The entire MyCode cohort can be used for collaborative research following appropriate protocol review and approval.  References:  1. Dou, Y., Gold, H. D., Luquette, L. J., & Park, P. J. (2018). Detecting Somatic Mutations in Normal Cells. http://doi.org/10.1016/j.tig.2018.04.003  2. Lim, E. T., Uddin, M., Rubeis, S. De, Chan, Y., Kamumbu, A. S., Zhang, X., … Walsh, C. A. (2017). Rates, distribution and implications of postzygotic mosaic mutations in autism spectrum disorder. Nature Neuroscience, 20(9). http://doi.org/10.1038/nn.4598  3. Steensma, D. P., Bejar, R., Jaiswal, S., Lindsley, R. C., Sekeres, M. A., Hasserjian, R. P., & Ebert, B. L. (2015). Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. Blood, 126(1), 9–16. http://doi.org/10.1182/blood-2015-03-631747  4. Zink, F., Stacey, S. N., Norddahl, G. L., Frigge, M. L., Magnusson, O. T., Jonsdottir, I., … Stefansson, K. (2017). Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. Blood, 130(6), 742–752. http://doi.org/10.1182/blood-2017-02-769869  5. Genovese, G., Kähler, A. K., Handsaker, R. E., Lindberg, J., Rose, S. A., Bakhoum, S. F., … McCarroll, S. A. (2014). Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. The New England Journal of Medicine, 371, 2477–2487. http://doi.org/10.1056/NEJMoa1409405 | |
| **Outline of Project** | 1. Review literature for prevalence of mosaicism and known mosaic variant rates for specific genes.  2. Describe current laboratory methodology for determining mosaicism.  3. Assess frequency of mosaic variants in eMERGE & MyCode samples. Further refine detection method using population data.  4. Use existing Sanger data to assess accuracy of putative mosaic calls  5. Compare to known mosaic rate and discuss implications for general population  6. Discuss health implications of mosaic variants in patient population.  7. Write manuscript | |
| **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  **Known blood dysplasias** |
| **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)* | |
| **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):  1. Read depth sequencing data of all samples sequenced during the eMERGE project.  2. Read depth sequencing data from the MyCode Samples. | |
| **Does project pertain to an existing eMERGE Phenotype?** | Yes, if so please list  No | |
| **Planned Statistical Analyses** | 1. Collect statistical metrics for refining the detection of mosaic variants, including statistics involving read depth and other sequencing metrics.  2. Calculate variant allele fraction and the deviation from an expected heterozygous allele fraction of 0.5.  3. Analyze the distribution of VAF (Variant Allele Fractions) and compare the distribution of VAF between genes expected to be involved in CH with the other genes in the eMERGE Panel.  4. Compare the distribution of VAF between samples based on age. | |
| **Ethical Considerations** | There are no physical risks involved. | |
| **Target Journal** | Nature Genetics in Medicine, BMC Genomics, PLOS Genetics, Journal of molecular diagnostics. | |
| **Milestones**  *(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | Approval: April 30, 2019  Project Duration: Through July 31, 2019  Draft Completion: September 30, 2019  Submission: October 31, 2019 | |

**\*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