**APPENDIX 2: External Manuscript Concept Sheet**

|  |  |  |
| --- | --- | --- |
| **eMERGE Network: External Collaborator Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* | NT460 | |
| **Submission Date** | Sept. 22, 2022 | |
| **Project Title** | **Genotyping of surgical patients: A machine-learning tool for a future precision medicine approach for prevention of postoperative complications.** | |
| **Tentative Lead Investigator** *(first author)* | Mathias Christensen | |
| **Tentative Lead Investigator Email** *(first author)* | [mathias.aagaard.christensen@regionh.dk](mailto:mathias.aagaard.christensen@regionh.dk) | |
| **Tentative Senior Author**  *(last author)* | Martin Sillesen | |
| **eMERGE Site Sponsor & Contact** | Northwestern Memorial Hospital (Northwestern University)  Contact: Jennifer Pacheco | |
| **All Other Authors** | Kaleem Ahmed, Hasan Alam, Jennifer Pacheco, Megan Roy-Puckelwartz, and any others involved from any other participating eMERGE site. | |
| **Sites Participating** | Northwestern Memorial Hospital (Northwestern University); Copenhagen University Hospital, and any other interested eMERGE sites. | |
| **Background / Significance** | Results from genome wide association studies (GWAS) have indicated that commonly occurring single nucleotide polymorphisms (SNPs) regulate the risk of surgically relevant aspects, including common infections, the dynamics of both the innate and adaptive immune response, the risk of thrombosis and potentially also the risk of organ failure. While studies have identified SNPs associated with AE’s in pre-defined genes following specific surgical procedures, GWAS-type studies on mixed surgical cohorts are lacking. Potentially, such studies could identify novel biomarkers and treatment targets of surgical AE’s and could thus add significant value to surgical prediction models.  As multiple different immunological pathways interact in defining the postoperative phenotype, surgical AE’s are per definition a complex disease. As such, commonly used GWAS methodologies may, however, not be optimally suited for detecting relevant SNPs in this setting. Furthermore, the role of other methodologies such as the polygenic risk score, is not clearly established.  To this end, our group has recently in collaboration with the Department of computer science at the University of Copenhagen, started work on exploring the role of machine learning, including deep neural networks (DNN), in identifying both the phenotype and relevant SNPs for complex surgical diseases. | |
| **Outline of Project** | The overall aim of the proposed study is to investigate whether genotyping of surgical patients can be useful as a precision medicine tool towards risk stratification of AEs, either alone or in combination with artificial intelligence enhanced stratification tools.   1. Gather other desired data from sites by fall 2022 2. Analyze data to assess for outcomes: We will use CPT codes to determine when certain surgeries were done, and then we will use ICD & phecodes, and labs, to determine if/when certain outcomes occurred after the surgeries, and also to assess co-morbid conditions. BMI and smoking status will be a covariates in our analyses as they are risk factors for some of the complications from surgeries. Anticoagulants, aspirin/NSAIDS and tamoxifen will also be covariates in our analyses, as these medications are important for the risk of some complications as well. 3. Perform GWAS on patients who had selected surgeries, comparing those with and without selected outcomes 4. Create PRSs 5. Use DNN to identify phenotype and relevant SNPs for complex surgical diseases 6. Draft manuscript and distribute to all co-authors for feedback 7. Submit manuscript | |
| **Desired Data - Common Variables\***  *(Available from the CC)* | * **Demographics** * ☐**ICD9/10 codes** * ☐**CPT codes** * ☐**Phecodes** * ☐**BMI** | * **Common Variable Labs** * **Common Variable Meds** * **Geocoding 2015 ACS variables** * **Other: Case/Control status** |
| **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)*  In addition to the above, we need the following lab results from admission: [BUN (blood urea nitrogen), Creatinine, Albumin, Bilirubin, SGOT (serum glutamic oxaloacetic transaminase), Alkaline phosphatase, , Platelet count, PTT (partial thromblastin time), PT (prothrombin time), INR (international normalized ratio), Sodium, Potassium, Hemoglobin.  smoking status, 30-day postop mortality  **Please note we’ll provide OMOP SQL where possible for sites to hopefully retrieve this data more easily.** | |
| **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** | The initial GWAS-analyses will be conducted with a mixed linear model (MLM)-based approach as fastGWA with a sparse GRM. A p-value of 5x10-8 will be considered statistically significant. Polygenic risk scoring will be performed according to previously published methods from our collaborators at the University of Copenhagen. For the DNN part, convolutional neural networks will be trained on the genotyping data using feature importance methodologies to identify the most relevant SNPs for outcome prediction.  In the final step of the projects, the different methodologies (GWAS, polygenic risk score and DNN) will be fitted to risk prediction models including relevant clinical confounder variables for the purpose of assessing AUROC performance metrics. | |
| **Ethical Considerations** | No direct interaction with patients will take place, and the proposed studies will only use data already present in approved biobanks (eMerge and NUgene). | |
| **Available Funding or Resources** | Access to NUgene and eMerge is funded by a grant from the Novo Nordisk Foundation to Martin Sillesen (Grant #NNF19OC0055183) | |
| **Target Journal** | To be decided.  Annals of Surgery  JAMA Surgery  British Journal of Surgery | |
| **Milestones*\*\****  *(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | We anticipate manuscript submissions within 1 year of approval of data access. Analysis is expected to one completed by Dec 31, 2022 and the manuscripts will be completed by summer 2023 as follows:   * + 1. ***06/01/2023:*** 1st manuscript draft sent to all authors for review     2. ***06/15/2023:*** Final draft sent to all authors for approval     3. ***06/25/2023:*** Submission to journal | |

**\*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

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