Introduction

The Million Veteran Program (MVP) is a national genetics research initiative of the Department of Veteran Affairs Office of Research and Development (ORD). The purpose of MVP is to create a longitudinal cohort of at least one million Veterans to understand better how genes, lifestyle behaviors, and military exposures impact health and illness, and to bring personalized medicine to VA health care. Veterans who volunteer to join MVP provide access to their medical records, complete a baseline survey, an optional lifestyle survey, and provide a blood specimen. Currently, MVP has over 930,000 Veterans participating.

MVP data access is open to all VA researchers through Merit Awards. Since 2015, VA has funded over 70 VA research projects using MVP data. These projects represent VA investigators from all four ORD research services and the Cooperative Studies Program. The project teams have helped MVP test the computational infrastructure, and the regulatory and administrative frameworks around data access in addition to increasing the scientific knowledge regarding the genetics of various chronic illnesses and traits impacting US Veterans.

This guidance document outlines 1) how VA investigators can apply for VA Merit Awards that include aim(s) requiring MVP data across the four ORD research services and 2) the updated data types that will be available to use for VA Merit awards starting in the Spring 2023 review cycle.

Available MVP data and Computing Environment:

All data analysis using MVP data will take place behind the VA firewall in the Genomic Information System for Integrative Sciences (GENISIS) research environment. GENISIS serves as the MVP informatics and computing platform. In addition, GENISIS also supports recruitment and enrollment module, and the Biorepository Laboratory Information Management System (LIMS). The MVP data resources include electronic health records extracted from the VA Corporate Data Warehouse (CDW) through the Veterans Informatics and Computing Infrastructure (VINCI), curated self-reported survey data and genomic data, and a High-Performance Computing (HPC) cluster with analytical tools. Researchers will have access to data, computing resources, and analytical tools within this secure study-specific study mart in GENISIS. The Data and Computational Sciences (DACS) Core team builds and manages the protected data and computing infrastructure for MVP behind the VA firewall.
Researchers will also have access to the Centralized Interactive Phenomics Resource (CIPHER), the VA Phenotype library with a catalog of phenotype descriptions and associated metadata.

The following updated MVP data will be available for approved projects starting in the Spring 2023 Merit Cycle:

1) **Electronic Health Records (EHR) data** on ~930,000 MVP participants, extracted from the CDW through the Veterans Informatics and Computing Infrastructure (VINCI).

2) **Data from two surveys completed by the MVP participants.** The first is a “Baseline Survey” on ~75% of the participants, focusing on demographic characteristics, medical and family history, health status, and lifestyle habits. The second is an optional “Lifestyle Survey” designed to gather detailed military and environmental exposure, dietary habits, and other behavior data, available on ~60% of the MVP participants.

3) **Genotype data on ~650,000 MVP participants** generated by a custom designed Axiom genotyping array designed to maximize genomic coverage of common and rare SNPs as well as markers with clinical significance. There are ~800,000 SNPs on the MVP Axiom array. More information can be found in GenHub resource.

**Data coming in FY2023—can be requested starting in Spring 2023 Merit Round**

4) **MVP Nutrition data**— Dietary energy and nutrient intake data were derived from food frequency questionnaire (FFQ) data collected on the MVP Lifestyle Survey. This included frequency of 61 food items in addition to questions about added sugar, fried food consumption, and 21 dietary supplements. More information can be found at [https://vhacdwdwhweb100.vha.med.va.gov/phenotype/index.php/Nutrient-Level_Data_(MVP_Core_Data)]

5) **MVP Whole Genome Sequence (WGS) data**—WGS data from over 100,000 MVP participants will available in FY23. The WGS data has been generated on the short read Illumina platform with 30X depth depth of coverage.

6) **Methylation data** – Methylation data on ~40,000 MVP participants will also be available in FY23. The methylation data is generated on the Illumina EPIC array with ~850,000 probes.

7) **COVID-19 Survey data** —surveys from ~255,000 MVP participants on symptoms, diagnosis, hospitalization, behavioral and psychosocial factors physical and psychological impacts of the COVID pandemic--- available for COVID-related aims only.

**Notes on available data and analyses:**

- Applicants may not request access to MVP biospecimens or re-contact of MVP participants as part of the proposal.
- MVP sample collection and processing is not currently CLIA certified, and resulting data cannot be used to make clinical decisions.
- Interested applicants should note that MVP is an iterative data resource for VA and the broader research community.
  - Tools and final summary data sets remain inside the GENESIS computing environment for future query and research by other researchers following an initial embargo period.
Any results and summary statistics that are published from funded awards are required to be submitted to the GenHub results browser/Genisis as a condition for accessing MVP data.

Summary results and statistics from MVP data analyses will be deposited into dbGaP for access by the broad scientific community.

Any new phenotypes created and validated as a result of an MVP Merit Award will be required to be submitted to the VA Phenotype library, CIPHER (the Centralized Interactive Phenomics Resource) which can be found at this link Centralized Interactive Phenomics Resource (CIPHER) - VA Phenomics Library.

Applicant requirements:

1. The applicant PD/PI and or MPI (if applicable) should be VA employee(s) and should meet eligibility requirements of the Service to which they are applying. All other named personnel on the application should meet the eligibility requirements of the service to which they are applying.

2. Any person on the application requiring access to the MVP data must have a VA appointment OR a without compensation (WOC) VA appointment.

Notes for applicants:

- Collaborative research teams with expertise in genomics (if applicable), phenomics and data science are encouraged.

- Career Development Applicants for all four services can apply for MVP data access, but this guidance document does not apply. Please refer to documentation and guidance located here: RFAs and Program Announcements (va.gov)

Request for Application (RFA information)

Applicants can apply to any relevant RFA across any ORD research services. Applicants must follow all the rules and requirements of the specific RFA to which they are applying. Budgets, project duration, and eligibility are Service and RFA specific.

Briefly, steps involved in applying to access to MVP data for VA Merit Projects are described below:

Step 1: Registration in GenHub and Preparatory to Research Access:

- Applicants interested in the opportunity to access MVP data for Merit projects should register for an account in GenHub where they can view aggregate data from the MVP cohort to help prepare their letters of intent (LOI), if required and full applications and full applications. Please refer to https://GenHub.va.gov

- Explore MVP aggregate data with the cohort builder and results explorer tools in GenHub.

- Ensure that the MVP cohort numbers and data are available to support PD/PI’s research questions. Ex. You can check how many of your cohort of interest have associated Genotype, WGS, nutrition data.
Note: BLR&D/CSR&D RFAs other than the parent Merit Award RFA may have LOI requirements.

Step 2: Submission of Proposals (or LOI if required)

- Investigator groups should submit a full proposal according to the dates and requirements of the service and RFA to which they are applying. The corresponding peer-review process will occur using existing ORD-based scientific and administrative review mechanisms. Specifically, the process for preparation, submission, and review of applications for peer-review will utilize existing RFAs and service processes.
- Approval of proposals and final funding decision will be communicated to the PD/PIs via the usual service-specific mechanisms.

General Questions may be directed to MVPLOI@va.gov.

Service or RFA specific questions should go to the contact for the appropriate research services here RFAs and Program Announcements (va.gov)