Million Veteran Program: What is it & how will it be developed

John B. Harley, MD, PhD
Staff Physician, US Department of Veterans Affairs Medical Center, Cincinnati
Director, Center for Autoimmune Genomics and Etiology, Cincinnati Children’s Hospital Medical Center
Professor of Pediatrics & Medicine, University of Cincinnati

Genomics and Ethics Conference 2015
10:40 AM, March 13, 2015, Kingsgate Marriott, Cincinnati
Nesting Population Research in the VA Healthcare System

VA ideal setting for nested large-scale population research
– Stable and willing veteran population of 8 million using the system each year
– Outstanding electronic medical record; fully integrated; data from 2 decades; access to CMS and NDI data
– Research infrastructure with diverse expertise
– Prototypes for health system based research
  – Pragmatic clinical trials of HCTZ vs chlorthalidone
  – Million Veteran Program
Large-Scale Biobanks

Europe
- Icelandic Biobank and deCODE
- UK Biobank
  - Banco Nacional de ADN [Spain]
  - GenomEUtwin
  - Finnish biobank
  - Swedish biobank
  - German biobank, KORA
  - UK DNA Banking Network & British biobank
  - Estonian biobank:
    - Family-based collections [Nordic]
    - Generation Scotland
    - HUNT (cardiovascular)& Biohealth [Norway]
    - EPIC, European (cancer)
    - Danubian Biobank Consortium
    - GATiB Genome Austria Tissue Bank
    - Biobank Hungary

North America
- Vanderbilt University BioVU
  - Canadian Consortium [Canada]
  - dbGaP, NIH [US]
  - National Children's Study [US]
  - Marshfield Clinic [US]
  - National Health and Nutrition Examinations Surveys [US]
  - Kaiser Permanente Northern CA [US]
  - Howard University African Diaspora [US]
  - Mayo Clinic
  - ACS
Million Veteran Program (MVP)

Enroll up to one million veterans into an observational mega-cohort

- Blood collection for biorepository for future research
- Collect self-reported health and lifestyle information
- Access to electronic medical record
- Ability to recontact participants
Distribution of MVP Sites

- Denver: 4,202
- New England Consortium
  - White River Junction, VT
  - Northampton, MA
  - Bedford, MA
  - Manchester, NH
  - Togus, ME

Legend:
- Red Star: Actively Recruiting
- Black Circle: Closed to Recruitment

Veterans Health Administration
MVP Organizational Structure

Million Veteran Program (MVP): A Partnership with Veterans

VA Central Office / Office of Research and Development
- Genomic Medicine Program
- GMPAC
- Cooperative Studies Program
- Genetics CSSEC
- VA Central IRB
- VA Communications
- MVP Executive Committee

MAYERIC
- Recruitment Center
- Project Management
- GenESIS
- VA Central Biorepository

CERC
- West Haven CSP Coordinating Center
- Admin Center
- Scanning Center

SMART

Albuquerque CRPCC

Study Planning
- ERIC’s
- CSP CC’s

Participating VAMC
- Local Site Investigator
- Study Visit: A) ICP/HIPAA B) Blood draw

Outside Vendor
- Mailing Center
- Scanning Center

Veteran

Data Sources: VA, Medicare, etc.

Genomic Analysis
- PAL
- Keil Lab
- Other laboratories

Investigators
- VA
- Non-VA
MVP Recruitment and Enrollment

• Invitational Mailing/Appointment Mailing
  – Invitation letter, Baseline Survey, MVP Brochure
  – Appointment letter, Informed consent language
• Walk-in recruitment Study visit
  – Informed consent/HIPAA
  – Blood collection
• Thank You Mailing
  – Thank you letter, Lifestyle Survey
• Specimens sent daily
Automated Specimen Processing
VA Central Biorepository
<table>
<thead>
<tr>
<th>MVP Recruitment to Date (February 20150)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Invitation mailings sent</td>
</tr>
<tr>
<td>Expressed interest by mail</td>
</tr>
<tr>
<td>Optout</td>
</tr>
<tr>
<td>Completed Baseline Surveys</td>
</tr>
<tr>
<td>Consented Veterans</td>
</tr>
<tr>
<td>Specimens in Lab</td>
</tr>
<tr>
<td>Unscheduled (proportion)</td>
</tr>
<tr>
<td>Upcoming appointments</td>
</tr>
<tr>
<td>Call volume</td>
</tr>
</tbody>
</table>
Current Lab Activities (February 2015)

- Receiving and Processing - 400-600 per day
- Shipping Samples for Sequencing and Genotyping:

<table>
<thead>
<tr>
<th>Assay Type</th>
<th>Shipments to-date</th>
<th>Targeted Shipments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole Genome sequencing</td>
<td>1886</td>
<td>1370 + 516</td>
</tr>
<tr>
<td>Whole Exome sequencing</td>
<td>24260</td>
<td>24126</td>
</tr>
<tr>
<td>SNP Genotyping</td>
<td>206,603</td>
<td>~200,000</td>
</tr>
</tbody>
</table>
Axiom MVP Biobank Array

23K eQTLs Markers
- Selected from 1000 Genomes database, NCBI/NHGRI eQTL database and Axiom Genomic Database
- ~7K tiled and ~23K pair-wise tagged

70K Novel Exome/LOF Variants
- New Exome/LoF Content from 26K Exome Sequencing initiative
- High Confidence LOFs
- No Singletons
- Known disease-causing mutations
- Potential Splice Variants
- 30K INDEL/45K SNPS

264K cSNPs & InDels Variants
- 197K non-synonymous
- 18K InDels
- 15K compatibility SNPs
  GWAS, ESP, HLA, Fingerprint, mtDNA, Y chr, miRNA targets, AIMs

400K multiethnic Grid
- polymorphic CEU, YRI, LAT
- Impute v2-based Selection

Imputation GWAS grid

2K variants
selected from:
- Pharmaadme.org
- PharmaGKB.org

ADME Content

VA Custom Disease Specific Variants

New Exome/LoF Content

Exome Content
Million Veteran Program (MVP) Data Universe

- Self-reported MVP surveys
- VA - Clinical VINCI, VIReC
- Non-VA NDI, CMS, etc.
- Molecular Data
- Biospecimen

MVP Participant
## Other Data Sources

### MVP Data
- **Self-Reported Survey Data:**
  - Lifestyle Survey Data (Personal Information, Well-Being, Activity, Health, Military Experience, Dietary Intake, Medication, Habits)
  - Baseline Survey (Health, Military Experiences, family medical history)
- **Genetic Data**
  - Genotype data
  - Sequence data

### Other Data
- **VA Healthcare System Data**
- **Other Data**
  - National Death Index (NDI)
  - Centers for Medicare and Medicaid Services (CMS)
  - State Mortality Data
VA Data Sources

- Corporate Data Warehouse Databases
- National Patient Care Databases
- Vital Status
- Decision Support System
- National Data Extract
- Beneficiary Identification Records Locator (BIRLS) death file
- New England VISN-1 Pharmacy files
- Outpatient Clinic File (OPC)
- Patient Treatment File (PTF)
- Inpatient and Outpatient Hospitalizations

- Clinic Inpatient and Outpatient Visits
- Diagnosis (ICD-9) codes
- Procedure (CPT) codes
- Pharmacy data and laboratory data

- Pharmacy Benefit Management (PBM) system database
- OEF/OIF and OND Roster
- VA Clinical Assessment Reporting and Tracking (CART)
- Veterans Affairs Surgical Quality Improvement Program (VASQIP)
- Veterans Affairs Central Cancer Registry (VACCR)
MVP Phenotyping Activities

Core Variables
- Demographics
  - Age, Sex, Race
- Laboratory values
  - Total cholesterol
  - HDL, LDL
  - Albumin
  - Serum creatinine
  - Triglycerides
- Medications
- Other characteristics
  - Blood pressure
  - Height/weight/BMI
  - Smoking
  - Alcohol consumption
  - Combat exposure

Complex Phenotypes
- Myocardial infarction (MI)
- Stroke
- Unstable angina, revascular tx
- Acute congestive heart failure
- Death from cardiovascular disease
- Vascular procedure
- Posttraumatic stress disorder (PTSD)
- Schizophrenia
- Bipolar disorder
- Traumatic brain injury
- Depression
- Vascular dementia
- Cognitive impairment
- Type 2 diabetes mellitus
- Creatinine trajectory
- Glucose trajectory

Algorithm Development
Validation Methods
3-Tier 7-Step Phenotyping Process

Tier I Algorithm (T1A)

Step 1: Define initial working algorithm (T1A)
Step 2: Create study cohort and apply T1A
Step 3: Create Annotation Data Set
Step 4: Create Phenomic Database through Data Processing Pipelines
Step 5: Derive T2A
Step 6: Evaluate T2A to formulate T3A
Step 7: Develop probabilistic model and assign caseness

Deposit resulting algorithms to a central Phenotype Library
Ongoing alpha and beta test projects

- Core MVP descriptive projects
  - Characterization of the cohort
  - Race: EHR v. self report v. genes
- Alpha Test projects
  - Schizophrenia/bipolar: Cases collected separately; controls from MVP
  - PTSD, Post Traumatic Stress Disorder
- Beta test RFA
  - LOIs
  - Full proposals submitted this week, due March 11.
    Ours is on **Chronic Obstructive Pulmonary Disease (COPD)**
- Projects in planning
Acknowledgments

VA Central Office
Timothy O’Leary, M.D., Ph.D.; Sumrita Muralidhar, Ph.D.; Ronald Przygodzk, Ph.D.; Grant Huang, Ph.D., M.P.H.; James Breeling, M.D.; Jennifer Moser, Ph.D.; Kendra Schau, Kristina Hill
MVP Steering Committee and Subcommittees
Steering Committee: John Conato, M.D., M.S., M.P.H., Co-Chair; J. Michael Gaziano, M.D., M.P.H., Co-Chair; Rayan Al Jundi, M.D.; James L. Breeling, M.D.; Louis Fiore, M.D., M.P.H., ex-officio; John B. Harley, M.D., Ph.D.; Elizabeth Hauser, M.D., Grant Huang, M.P.H., Ph.D. Ex-Officio; Phil Tsao, Ph.D., Janet D. Crow, ORD Liaison; Philip D. Harvey, Ph.D.; Timothy J. O’Leary, M.D., Ph.D., Ex-Officio; Stephen G. Wynn, M.D., Ph.D.; Peter W. F. Wilson, M.D.
Access Policy Subcommittee: John, Conato, M.D., M.S., M.P.H., (Chair); Sumrita Muralidhar, Ph.D. (VACO Liaison); Michael Hart, M.D.; Fred Wright, M.D.; John (Tom) Callaghan M.D., Ph.D.; J. Michael Gaziano, M.D., M.P.H., Saju Pyarajan, Ph.D.; Carl Grunfeld, M.D., Ph.D.
Chemical Analysis Subcomittee: Philip Tsao, Ph.D., (Chair); Steven Schichman, M.D., Ph.D.; Saju Pyarajan, Ph.D.; Chris Corless, M.D., Ph.D.; Robert Lundsten; Jenny Moser, Ph.D.; Sunit Ahuja, M.D.
Recruitment Subcommitee: Rayan Al Jundi, M.D., (Chair); Jennifer Deen, Hermes Florez, M.D., Ph.D., M.P.H.; Amy Kilbourne, Ph.D., M.P.H.; Frank Lederle, M.D.; Kendra Schau, ScM, C.G.C. (ORD Liaison); Ralph Heussen; Stephen Herrig.
Phenotyping and Epidemiology Subcommitee: J. Michael Gaziano, M.D., M.P.H., (Chair); Jenny Moser, Ph.D.; Seth Eisen, M.D., M.Sc; David Gagnon, M.D., Ph.D.; Saju Pyarajan, Ph.D.; Jonathan Nebecker, M.D., M.S.; Denise Hynes, Ph.D., M.P.H., R.N.; Matthew Samore, M.D.
Boston Coordinating Center (MAYVRC)
J. Michael Gaziano, M.D., M.P.H.; Stacey Whittome, Ph.D., Jennifer Deen; Colleen Shannon, M.P.H.; Kelly Cho, Ph.D.; Xuan-Mai Nguyen, Ph.D.; David Gagnon, M.D., Ph.D.; Donald Pratt; Lindsay Hansen, M.P.A.; Keri Hannagan, M.P.H.; Derrick Morin
West Haven Coordinating Center (CERC)
John Conato, M.D.; M.P.H.; Mihaela Aslan, Ph.D.; Peter Guarino, Ph.D.; Daniel Anderson; Rene LelaFleur; Lesley Mancini, M.B.A.; John Russo; Cindy Cushing; Vales Jean-Paul; Fred Sayward; Louis Piscitelli; Doni Cavaliere
VA Central Biorepository - Boston
Mary Brophy, M.P.H.; Donald Humphries, Ph.D.; Robert Lundsten; Geetha Sugumaran, Ph.D.; Deborah Govan; Christine Govan
Genomic Information System for Integrated Sciences (GeniSIS) - Boston
Louis Fiore, M.D., M.P.H.; Saju Pyarajan, Ph.D.; John Gargas; Tony Chen, Ph.D.; Edmund Peirce; Paul Hsieh, Ph.D.; Luis Selva, Ph.D.; Beth Katcher; Robert Hall; M.P.H.; Weijia Mo; Felicia Fitzmeyer; Anahit Aleksanyan
MVP Information Center – Canandaigua, NY
Annie Correa; Melissa Juhl; George Barzec III; Liam Cerveney; Anne Van Patten; Jonathan Martinez; Eileen Loh-Fontier; Jessica Marino

MVP Local Site Investigators
Thank You!
Building the Plane as You Fly It
Million Veteran Program

J. Michael Gaziano, M.D., M.P.H.

Scientific Director, MAVERIC, VA Boston Healthcare System
Chief, Division of Aging, Brigham and Women’s Hospital
Professor of Medicine, Harvard Medical School

February 11, 2015