The Million Veteran Program: Transforming Genomic Research into the Clinic

Ronald M. Przygodzki, MD

Director, Genomic Medicine Implementation
Office of Research and Development
US Department of Veterans Affairs
Washington, DC
Population Research in the VA Healthcare System

- VA ideal setting for nested large-scale population research
  - Stable and altruistic Veteran population of 9 million using the system each year
  - Integrated electronic medical record; data reaching back as far as 20 years; access to CMS and NDI data
  - Intramural research program with diverse clinician and doctoral expertise
  - >70% of investigators are clinicians linking research with Veterans’ needs
  - Strong academic affiliate relationships
  - An example of team science and big data within VA infrastructure, involving administrative, technical, ethical and scientific challenges.
Goal of VA Genomics Program

• To develop and provide standardized databases of medical and genomic data that will improve our understanding of disease
• Build out of recruitment and collection standards
• Analytical standards for genomic and phenotypic data
• Human subjects and data protection
• Discovery and repurposing of all data
• Implementation of tests that have clinical utility and clinical validity

• Hope is to provide a pseudo-wet bench for any researcher post merit and IRB review
Million Veteran Program (MVP)

- Enroll up to one million users of the VHA Health System into an observational mega-cohort
  - Standardization and uniformity of processes
  - Provide Informed Consent/HIPAA authorization
  - Blood collection for storage in biorepository for future research
  - Collect self-reported health and lifestyle information
  - Access to electronic medical record
  - Ability to recontact participants
Million Veteran Program: A mega-biobank to study genetic influences on health and disease

John Michael Gaziano\textsuperscript{a,b,1}, John Concato\textsuperscript{c,d,*,1}, Mary Brophy\textsuperscript{a,e}, Louis Fiore\textsuperscript{a,e}, Saiju Pyarajan\textsuperscript{a}, James Breeling\textsuperscript{a}, Stacey Whitbourne\textsuperscript{a}, Jennifer Deen\textsuperscript{a}, Colleen Shannon\textsuperscript{a}, Donald Humphries\textsuperscript{a}, Peter Guarino\textsuperscript{c,d}, Mihaela Aslan\textsuperscript{c,d}, Daniel Anderson\textsuperscript{c}, Rene LaFleur\textsuperscript{c}, Timothy Hammond\textsuperscript{f}, Kendra Schaa\textsuperscript{f,2}, Jennifer Moser\textsuperscript{f}, Grant Huang\textsuperscript{f}, Sumitra Muralidhar\textsuperscript{f}, Ronald Przygodezki\textsuperscript{f}, Timothy J. O’Leary\textsuperscript{f}

\textsuperscript{a}Massachusetts Area Veterans Epidemiology Research and Information Center (MAVERIC), VA Cooperative Studies Program, VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130, USA
\textsuperscript{b}Department of Internal Medicine, Harvard Medical School, Boston, MA, USA
\textsuperscript{c}Clinical Epidemiology Research Center (CERC), VA Cooperative Studies Program, VA Connecticut Healthcare System, 950 Campbell Avenue, 151B, West Haven, CT 06516, USA
\textsuperscript{d}Department of Internal Medicine, Yale University School of Medicine, New Haven, CT, USA
\textsuperscript{e}Department of Internal Medicine, Boston University School of Medicine, Boston, MA, USA
\textsuperscript{f}Office of Research and Development, Veterans Health Administration, 810 Vermont Avenue N.W., 10P9, Washington, DC 20420, USA

Accepted 22 September 2015; Published online 9 October 2015
Putting MVP in Context

• Major biobanks in the world
  – UK Biobank (500K)
  – Kadoori Biobank (>510K)
  – Kaiser Permanante (>270K)
  – MVP (1 million)- current >650,000
  – NIH PMI Cohort- aiming for 1 million

• Stable and altruistic patient population

“Knowing that I would be helping other GI’s is the reason I am part of the Million Veteran Program.”

Mons S. Sjaastad - U.S. Army - Korean War Era
VA Connecticut Healthcare System
MVP Sites – Recruitment from Sea to Shining Sea

- = Actively Recruiting
- = Closed to Recruitment
## MVP Recruitment to Date

<table>
<thead>
<tr>
<th>Category</th>
<th>Figure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Invitation mailings sent</td>
<td>&gt; 4.2 Million</td>
</tr>
<tr>
<td>Optout</td>
<td>13.8%</td>
</tr>
<tr>
<td>Completed Baseline Surveys</td>
<td>657,378</td>
</tr>
<tr>
<td>Consented Veterans</td>
<td>&gt; 656,000</td>
</tr>
<tr>
<td>Call volume</td>
<td>&gt; 500,000</td>
</tr>
</tbody>
</table>

*as of April 4, 2018*
What do MVP Participants Look Like?

Data from Enrolled Participants (N=586,427)

**Gender**
- Male, 90%
- Female, 9%
- Missing, 1%

**Race**
- White, 72%
- Black, 18.5%
- Asian, 3%
- American Indian, 3%
- Other, 2%
- Multiracial, 1%
- Missing, 6.5%

**Age**
- 18 to 29: 2.8%
- 30 to 39: 6.1%
- 40 to 49: 8.8%
- 50 to 59: 18.5%
- 60 to 69: 37.5%
- 70 to 79: 17.4%
- 80 to 89: 8.0%
- 90 to 99: 1.0%
- 100+: 0.0%
- Missing: 0.0%

**Service Era**
- Sept 2001 or later: 11.6%
- Aug 1990 to Aug 2001: 22.9%
- May 1975 to July 1990: 23.1%
- Aug 1964 to Apr 1975: 48.6%
- Feb 1955 to Jul 1964: 10.7%
- Jul 1950 to Jan 1955: 7.1%
- Jan 1947 to Jun 1950: 1.0%
- Dec 1941 to Dec 1946: 3.1%
- Nov 1941 or earlier: 0.1%
Evolving MVP Genetic Projects and Scientific Organization

SNP Array (750K) imputed to 1000G QC & Imputation on ~400K samples

WES >35000

WGS 2000, planned 25000 FY18/19
- “Alpha” Projects
- “Beta” Projects
- “Gamma” Projects (just begun): Suicide, DM2 Complications, Breast CA, Tinnitus, Osteoarthritis, Transcriptome/Epigenome

Core Working Groups:
- Genomic WG
- Phenomic WG
- Statistical Genetics WG

Disease Area Working Groups:
- Substance Abuse
- Cardiovascular Disease
- Metabolic Disease
- Lipids
- Chronic Kidney Disease
- Blood Pressure/HTN
- Eye Disease

Planned New Working Groups:
- Omics
- Pharmacogenetics
- PheWAS, PheMR
- Actionable Genome
- And Other...

VA ORD

MVP Executive Committee

VETERANS HEALTH ADMINISTRATION
Current Research Activities

- Collaborative Science
  - 20+ projects
  - 100+ researchers
  - 30+ VA and academic affiliates
- Current projects span various topics including cardiovascular risks, mental health, suicide, breast cancer and PTSD
- Partnership with DOE to leverage computing infrastructure and expertise
  - includes 3 collaborative projects addressing suicide, prostate cancer, and heart disease
### Initial MVP Genomics Projects and Design Features

#### VA Priority Phenotypes
- Schizophrenia/BPD
- Hypertension/BP
- PTSD
- Substance abuse
- Coronary heart disease
- Peripheral arterial disease
- Chronic kidney disease
- Lipid subfractions

#### Unique MVP Design Features
- Case-cohort design
- Multiethnic (AA, HA)
- CVAS and RVAS
- Prospective, longitudinal, repeated phenotypes
- Detailed pharmacy data
- Survey & EHR data

**MVP Science 2017**
- >22 Abstracts At ASHG
- ~20 Abstracts Other Meetings 2017
Poster 1963: 101 novel loci and novel associations with gene expression detected in transethnic genome-wide study of estimated glomerular filtration rate in over 270,000 participants: The Million Veteran Program. Adriana Hung et al on behalf of MVP

- Total N=56,253 AA and N=216,585 EA veterans
- SNPs in 157 loci with significance threshold of $p=5\times10^{-8}$ explaining 4.3% of phenotypic variance
- Strongest signal replicating the previously detected association at UMOD/PDILT
- 811 significant gene expression/tissue/phenotype associations
Poster 2666: African ancestry genome-wide and transcriptome-wide association study of blood pressure detects 9 novel loci in a large cohort from the Million Veteran Program.

Jacklyn Hellwege et al on behalf of MVP

Results for Pulse Pressure

- Total N>95,000 African Americans
- Identified >50 novel loci for systolic, diastolic, and pulse pressures
- Predicted gene expression pinpoints the functional gene targets at known and novel GWAS loci across 44 tissues
- 811 significant gene expression/tissue/phenotype associations
# Ongoing Scientific Projects

<table>
<thead>
<tr>
<th>Project Level</th>
<th>Topic</th>
<th>VA &amp; University Affiliations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha test</td>
<td>Schizophrenia and Bipolar Disorder (CSP 572)</td>
<td>Miami, Bronx, West Haven, Boston VAs; University of Miami; Mt. Sinai School of Medicine; Yale; Harvard; Stanford</td>
</tr>
<tr>
<td>Alpha test</td>
<td>PTSD in Veterans (CSP 575B)</td>
<td>San Diego, West Haven, Boston VAs; UC San Diego; Yale</td>
</tr>
<tr>
<td>Alpha test</td>
<td>Gulf War Illness (CSP2006)</td>
<td>Durham, West Haven, Boston VAs; Duke, Yale</td>
</tr>
<tr>
<td>Beta test</td>
<td>Cardiovascular Disease</td>
<td>Atlanta, Boston VAs; Emory; Harvard; Boston University</td>
</tr>
<tr>
<td>Beta test</td>
<td>Chronic Kidney Disease</td>
<td>Nashville, Memphis VAs; Vanderbilt; University of Tennessee Health Science Center</td>
</tr>
<tr>
<td>Beta test</td>
<td>Cardio-metabolic Disease</td>
<td>Palo Alto, Philadelphia, Albany, Boston, Phoenix VAs; Univ of Pennsylvania; Stanford; Albany Medical College; Univ of Massachusetts; Harvard; Arizona State University</td>
</tr>
<tr>
<td>Beta test</td>
<td>Multi-substance Abuse Disorders</td>
<td>West Haven, Philadelphia VAs; Yale; University of Pennsylvania</td>
</tr>
<tr>
<td>Beta test</td>
<td>Age-related Macular degeneration</td>
<td>Cleveland, Buffalo VAs; Case Western Reserve</td>
</tr>
</tbody>
</table>
Phenotyping - Key Activities

• Priority Phenotype Validation
• Development of Phenomics Annotation Library (Knowledge Base)
• Designing and coordinating phenomics data computing environment
• Cleaning, validation and organization of Survey Data
• Chart review tool modules
• Data query and reporting system
• Core tables development and versioning implementation

• Next generation phenotyping – robust, scalable and efficient
  - Development and implementation of high-throughput phenotyping testing and application
### Corporate Data Warehouse
#### Sample Data Facts

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lab Results</td>
<td>7.7B</td>
</tr>
<tr>
<td>Clinical Orders</td>
<td>4.5B</td>
</tr>
<tr>
<td>Immunizations</td>
<td>71 M</td>
</tr>
<tr>
<td>Appointments</td>
<td>1.4B</td>
</tr>
<tr>
<td>Pharmacy Fills</td>
<td>2.2B</td>
</tr>
<tr>
<td>Clinical Notes</td>
<td>3.2B</td>
</tr>
<tr>
<td>Health Factors</td>
<td>2.2B</td>
</tr>
<tr>
<td>Encounters</td>
<td>2.4 B</td>
</tr>
<tr>
<td>Radiology Proc</td>
<td>202 M</td>
</tr>
<tr>
<td>Vital Signs</td>
<td>3.3B</td>
</tr>
<tr>
<td>Consults</td>
<td>315 M</td>
</tr>
<tr>
<td>Admissions</td>
<td>17 M</td>
</tr>
<tr>
<td>Surgeries</td>
<td>14 M</td>
</tr>
<tr>
<td>Oncology</td>
<td>1.3 M</td>
</tr>
</tbody>
</table>

*Patients: 22 M*
GenISIS Secure Computing Environment

- Provides a secure computing environment with analytical tools
- Allows users to submit and manage their HPC jobs from a web dashboard
- Users authenticated by their VA network accounts
- Default 5 TB space allocated to each project, plus additional scratch space
- Analytic software available (PLINK, R/Bioconductor, SAS, Matlab, Perl/Bioperl, JMP Genomics and others) and will be updated based on user needs
VINCI: Resource for Clinical Data

- Veterans Informatics and Computing Infrastructure (VINCI) securely hosts select data from national VA databases
- Provides data to credentialed VA investigators with appropriate approvals
- Data updated nightly for many clinical domains
- Provides services and tools for data provisioning, curating, NLP, analytics and data services, annotation and chart review, feasibility determination, and application development
  - Established an enclave with the MVP crosswalk within VINCI
Partnership with Department of Energy

• MVP-CHAMPION & Big Data Initiative
  - Leverage DOE’s high performance computing capabilities and VA’s big data (CDW and MVP) and expertise to foster big data science, and expand data access
  - Secretary Perry enrolled in MVP

• Early projects focusing on:
  - Suicide prevention
  - Prostate cancer
  - Cardiovascular disease
“Big Data” Vision
National Healthcare User Facility
(a.k.a CERN for Healthcare)

Commercial Environment
- Infrastructure as a Service
  - Data Storage
  - Disaster Recovery
- Platform as a Service
  - Model Processing
- Software as a Service
  - Commercial Models

VA Computing Environment
- Data assimilation, aggregation & storage
- Data exploration & visualization
- Data curation
- Self-service tools
- Automated study mart development
- Minimal COTS tools

Commercial Cloud
(Azure, AWS)

DoE
(ORNL, ANL, ETC.)

VA

DOE Secure PHI enclave
- Heterogenous data: Phenotype, Genotype, Images
- Genetic data pipeline (e.g. Broad pipeline)
- Open Source Lab w/ Synthetic Data
- Advanced Computing Environment
- Data Store R&D
- Phenotype Registry
- Heterogenous Data Study Marts
- Model Building & Refinement
- “Data Science” Academy (Seminars, Conferences, Hackathons)

Data partners contribute data and expertise.

NDI
DoD
CMS
Partnership with Department of Defense

• Memorandum of Agreement to enroll separated and active duty members from the Millennium Cohort Study (MCS) in MVP beginning in FY17
  - Preliminary work: establish agreement; address regulatory issues; identify co-enrollees; prepare mailings; etc.
  - **Phase 1: recruit MCS participants who use VHA**
  - Phase 2: recruit MCS participants *not* enrolled in VHA
  - Phase 3: recruit “non-MCS” active duty at MVP sites
  - Phase 4: recruit “non-MCS” active duty at DoD-based sites

https://www.millenniumcohort.org
MVP-MCS Status Update: Phase 1

N=201,620
Total MCS Population

N=123,605
VHA enrollees

Already contacted by MVP N=40,230

5,389 enrolled
3,584 upcoming appointment
2,086 do not contact
29,171 no response

N=83,375
# not contacted by MVP
57,282 within 150 miles
Precision Medicine in Mental Health Care
PRIME Care 
(*P*recision *m*edicine *I*n *M*Ental *h*ealth*)

- Initiative led by GMI, HSR&D and QUERI, PI David Oslin, MD
- Pragmatic trial performing pharmacogenomic testing in Veterans with major depressive disorder (n=2000)
  - Assessment of clinical validity, utility and effect on treatment choice by clinician through impact on Veteran treatment response and outcomes
  - Return of results to clinician – and – patient
  - Education
- Study involves diverse scientific expertise in psychiatry, primary care, genetics, genomics medicine, pharmacy, biostatistics, and data analytics
- Several companies have introduced PGx batteries, but FDA does not approve/endorse the actual tests
Support for PRIME

• Depression is one of the world’s great public health problems
• At least 1 in 7 Veterans is currently suffering from a depressive disorder
• Untreated/poorly treated depression is implicated in 75% of suicides and amplifies the burden of all common chronic medical illnesses
• Only about 1/3rd remit with the first course of medication and 1/3rd will remain depressed despite multiple treatment trials
• Growing use of genetic testing for therapeutic decision making without clear evidence
If you’re in traffic, what do you do?
Points to ponder with PRIME

- Pt. prescribed multiple meds - nonadherence (side effects)
  - Found to be CYP2D6 *5/*5 and CYP2C19 *2/*2 (both poor metabolizers)
  - Prescribed fluvoxamine (1° 1A2/2D6, 2° 2C9/3A4) BUT not w/SSRI (inhibitory)
- Pt. prescribed clozapine or olanzapine
  - Nonsmoker - effective
  - Smoker and *1F/1F homozygote or *1F/*1D – likely little drug effect
- Pt prescribed bupropion
  - Metabolized 1° by 2B6 to hydroxybupropion
  - Metabolized 2° by 2D6
    - If poor metabolizer (2 inactive *3-8, 11, 15 – or 1 inactive/1 deficient *2B/D, 9, 10, 17, 41) accumulates and leads to seizures
Providers

- I feel well-informed about the role of PGx testing in choosing a psychotropic medication (1 strongly agree to 5 strongly disagree) Mean=2.6 (1.2)
- PGx testing will be beneficial to my patients with MDD Mean=2.2 (0.9)
- I am confident in my ability to effectively treat MDD Mean=1.4 (0.6)
“I want to thank you all for a job well done. Maybe I can help in some way to help other Veterans in the Million Veteran Program.”

–Bubba, MVP Veteran Participant
Thank you!

ronald.przygodzki@va.gov

@VAResearch

#MillionVeteranProgram

www.research.va.gov/MVP