OOKAMI PROJECT APPLICATION

Date: May 10, 2021

Project Title: Microbial marker identification as a model for early Alzheimer's diagnosis

Usage:

☒ Testbed

☐ Production

Principal Investigator: Weg Ongkeko

University/Company/Institute: University of California, San Diego Department of Surgery

Mailing address including country:

Phone number: (858) 552-8585 ext 7165

Email: rongkeko@health.ucsd.edu

Names & Email of initial project users:

Abby Lee acl008@ucsd.edu;
Aditi Gnanasekar <aditi.gnanam@gmail.com>,
Anjali Iyangar <aiyangar@ucsd.edu>,
Daniel John <d1john@ucsd.edu>,
Emilue Truong <emilie.truong4@gmail.com>,
Grant Castaneda <gecastan@ucsd.edu>,
Griffith Nguyen <gcnguyen@ucsd.edu>,
Ishan Ranjan <iranjan@ucsd.edu>,
Jaideep Chakladar <jchaklad@ucsd.edu>,
Jeffrey Liu <jcliu@ucsd.edu>,
Jiayan Ma <jim095@ucsd.edu>,
Jingyue Xu <j1xu@ucsd.edu>,
Joseph Tsai <jctsai@ucsd.edu>,
Kypros Dereschuk <kderesch@ucsd.edu>,
Lauren Apostol <laaposto@ucsd.edu>,
Lindsay Wong <lmw010@ucsd.edu>,
Matthew Uzelac <muzelac@ucsd.edu>,
Neil Shende <nshende@ucsd.edu>,
Shruti Magesh <smagesh@ucsd.edu>,
Wei Tse Li <wtl008@ucsd.edu>,
Yuxiang Li <yul055@ucsd.edu>

Usage Description:

Alzheimer's disease (AD) is one of the most common diseases in elderly people, with most people developing AD as they age. AD is currently understood to be irreversible, and diagnosis inevitably leads to limited interventions that only control AD symptoms. Therefore, a recent trend in AD research is the identification of biological markers that can be used for

early diagnosis of AD. A compelling new player in the study of neurodegenerative disorders is the microbiome. Dysbiosis of the gut microbiome specifically has been correlated to a wide range of disorders, leading to extensive research of the so-called gut-brain axis. However, recent findings have shown that the brain cavity, previously thought to be sterile, may have its own microbiome by way of the cerebrospinal fluid (CSF). In this study, we propose an analysis of CSF sequencing data of AD and healthy patients in order to identify biomarkers in the CSF microbiome that are indicative of AD development. RNA-sequencing data from AD and normal patients will be analyzed from the NIA Genetics of Alzheimer's Disease Data Storage site (NIAGADS). We plan to use the Ookami system to test the software used for patient sequencing sample analysis and compare methods by which to obtain microbial reads using sequence alignment methods, including Pathoscope2, Kaken2, and Qiime2. Once we have determined a proper analysis pipeline, we will use the system to outline the performance of the pipeline using multiple patient samples in order to understand the computational demand. Finally, we will test the use of TensorFlow or other available machine learning software on a select number of patient samples to determine how to use these software, the computational demands of applying them to patient samples, and how these software scale to fit a larger dataset. We hope to determine how the A64FX affects the speed at which our analyses are performed and also how the memory demands of our jobs are solved by the A64FX's high-bandwidth memory. This would allow us to scale our project to the A64FX with the goal of applying for a Production project in the future.

Computational Resources:

Total node hours per year: 15,000

Size (nodes) and duration (hours) for a typical batch job: 1 node, 24 hours per job Disk

space (home, project, scratch): home: 30 GB; project: 8 TB; scratch: 12 TB

Personnel Resources (assistance in porting/tuning, or training for your users**):**
Assistence with porting/tuning and troubleshooting errors related to system-sepcific architecture

Required software:
R, python3, python2, anaconda3, gnu_parallel, TensorFlow

If your research is supported by US federal agencies:

Agency: Not applicable

Grant number(s): Not Applicable

Production projects:

Production projects should provide an additional 1-2 pages of documentation about how
(a) the code has been tuned to perform well on A64FX (ideally including benchmark data comparing performance with other architectures such as x86 or GPUs)
(b) it can make effective use of the key A64FX architectural features (notably SVE, the high-

bandwidth memory, and NUMA characteristics)
(c) it can accomplish the scientific objectives within the available 32 Gbyte memory per node