Dan Xie, Ph. D

Associate Professor

College of Information Engineering

Hongshan District West Huangjia Lake No. 1

Hubei University of Chinese Medicine

Wuhan , Hubei, China

February 24, 2013

Dear Dr. Dan Xie:

I am excited about your project:” 面向肝病临床大数据的多元时序分析模型及其应用研究”。I will strongly support your research and allow you to use our computer sources for your researches. Our computer sources are described below.

The Human Genetics Center (HGC) has a series of interconnected microcomputer resources. All key personnel have late model mobile computers and/or desktop workstations. Researches in HGC operate multiple high-performance multi-CPU Intel workstations with 8GB memory. Two 4-CPU Sun Enterprise 450 servers with 8GB memory and two 2-CPU Sunfire 280 servers are dedicated and maintained for the Human Genetics Center faculty and staff. Additional UNIX resources include two V880, and 4-CPU Sun servers. Dr. Xiong’s lab has a DELL PRECISION T5400 sever with 2-CPU and 16GB memory, two DELL OPTIPLEX workstations with multi-core CPU and 8GB memory.

The analyses of exceptionally large amounts of genomic data require sufficient storage space, memory, and numerous multi-core processors. Thus, HGC recently upgraded our computational capacity by creating a computer cluster of the following components: one large server (24Gb memory, 4Tb hard drive) for data storage, data sharing and light data analysis; seven traditional servers (12Gb memory, 500Gb hard drive) for dedicated, computationally intensive analyses; two network storage servers (12 Tb) for data backup; and one GPU based workstation for simulation-based analyses.

Division of Biostatistics has a state-of-the-art Hewlett Packard high performance computing (HPC) cluster. The HPC cluster provides an environment that can accommodate many types of large-scale genetic and genomic computational tasks, as well as database management, with 1164 computing cores of 2.2GHz-2.6GHz CPUs across 35 HP ProLiant DL165 G6 (dual 6-cores) servers , 31 HP ProLiant DL165 G7 (dual 12-cores) servers (computing nodes), a HP ProLiant DL380 G7 server (master node) with dual 6-cores of 3.45GHz CPUs , and a HP ProLiant DL385 G7 server (master node) with dual 12-cores of 2.5GHz CPUs. A HP ProLiant DL385 G6 server (master node) with dual 6-cores of 2.6GHz CPUs as the front end, a Nvidia

Tesla S1070-500 System with 960 GPU computing cores, and a large scratch node memory and disk space. The proposed research will utilize this cluster to fully meet the statistical and computing needs for the project. In addition, the cluster is highly expandable and is ready to meet future computational demand.

The HPC cluster has a large node memory and storage space. The master node has 64GB RAM installed. Among all the computing nodes 58 nodes have 16GB RAM each, 8 nodes have 64GB RAM each,1 nodes has 192GB RAM,1 nodes has 96GB RAM, and the GPU node has extra 16GB dedicated shared memory. Each computing node comes with a 160GB hard disk for OS and applications. The master node comes with 2x160GB hard disks configured in RAID for system and applications. Five additional internal hard disks with 500GB, each configured with RAID 5, on the master node provide users with 2TB effective storage space. Additional three internal disk slots are available for future expansion. Recently a new HP StorageWorks MSA60 Array with 12x2TB disks and a new HP P2000 G3 MSA Array system with 12x2TB disks have been added to the cluster installed.

The HPC cluster is managed by HP's Cluster Management Utility (CMU) which administers and monitors the system. The open source resource manager Torque is used for the batch system administration. All computing nodes and the master node in the HPC cluster use SUSE Linux Enterprise Server 11 as the operating system, supporting standard open source GNU compilers and Java SDK, as well as a wide range of statistical and scientific computing softwares, such as R, PLINK, FBAT, MATLAB and the Message Passing Interface (MPI) with R support (Rmpi) for large-scale parallel computing. All the computing nodes are connected and managed in a secured and dedicated private network with Gigabit connection. The master node is in the network security zoon 40 behind the network firewall. It is the frontend of the cluster that only authorized users can access. Users' data can be efficiently and securely transferred by sftp within the UTSPH network. For outside users, the standard VPN connection is required for system access and data transfer. In summary,this cluster will offer the computational capability for carrying out the proposed large-scale genomic and epigenomic data analysis.

Sincerely,

Momiao Xiong

Professor

University of Texas School of Public Health

1200 Herman Pressler Dr.

Houston, TX77030