**Yale University**

Mark Gerstein, Ph.D. (0.45 summer month) is the Albert Williams Professor of Biomedical Informatics. His lab (http://gersteinlab.org) was one of the first to perform integrated data mining on functional genomics data and to do genome-wide surveys. His tools for analyzing motions and packing are widely used. Most recently, he has designed and developed a wide array of databases and computational tools to mine genome data in humans, as well as in many other organisms. He has worked extensively in the 1000 genomes project in the SV and FIG groups. He also worked in the ENCODE pilot project and currently works extensively in the ENCODE and modENCODE production projects. He is also a co-PI in DOE KBase and the leader of the Data Analysis Center for the NIH exRNA consortium. In these roles Dr. Gerstein has designed and developed a wide array of databases and computational tools to mine genomic data in humans as well as in many other organisms. He will lead the whole project, and will focus on the aim 1 to develop a comprehensive prioritization model for both coding and noncoding variants; he alo need coordinate with other investigators to achieve the goal described in this grant.

Shaoke Lou, Postdoctoral Associate (12 calendar months). has multidisciplinary background with a B.S. in Chemistry, an M.S. in Biochemistry and Molecular Biology, and a PhD in Computer Science and Engineering. He earned his PhD from the Chinese University of Hong Kong in 2012. During his PhD study, he mainly works on development of read-mapping algorithm, next-generation sequencing data analysis, data mining and statistics modelling. He has broad research interests and profiles, and is expertize in gene regulatory network study, disease association study and genomic variation prioritization. He also involved in the investigation and data analysis in model organisms: angiogenesis activities of Traditional Chinese Medicine in Zebrafish, regulation of chromatin structure factors(CapH2) in fly and circadian rhythmic study in Mouse and Neurospora et cetera. He will develop a model for coding and coding variants using molecular features extraction and do further iteratively parameter turning based on experimental evaluation results.

Jonathan Warrell, Associate Research Scientist (6 calendar months, year 1, 10 calendar months years 2-4, 9 calendar months year 5). has strong background in machine learning. He received a Ph.D. in Music Theory and Analysis from Kings College London in 2006 followed by a M.Sc. in Computer Science from University College of Londong in 2007. He subsequently did three postdoctoral fellowships focusing on the subject of machine learning and and using the method in network analysis in the field of bioinformatics. He will work on statistical modelling using cohort data,which will consider the recurrence of genotype compared the disease patient with normal person.

**Fringe Benefits:**

Fringe benefits are calculated at the rate of 32.3% for the PI, Associate Research Scientist and PDA according to University guidelines.

**Equipment:**

The Dell PowerEdge R815 is a server machine that will be used as a virtual machine hypervisor to support the project databases and web servers. It will play an integral part in our nascent cloud infrastructure used for the project.

**Supplies**:

We are budgeting an incremental amount of supplies for the individuals named above. This supplies budget will be used to cover computer supplies for them. It will cover such expenses as: diskettes and backup tapes, software upgrades, web hosting and "cloud computing" fees, and reprint charges. These items are needed to complete the proposed research and will solely benefit this project.

**Travel:**

As this is a collaborative project, we are budgeting considerable funds for travel between sites. Here we are requesting incremental funds for each of the FTEs for airfare, lodging and meal expenses to attend scientific meetings annually that benefit the project. In particular, the travel will include 1 trip per year to a scientific meeting of choice (e.g. ISMB).

**Indirect Costs**

Indirect costs are calculated at Yale’s federally negotiated rate of 67.5% of modified total direct costs. DHHS agreement dated 02/16/2017.