Budget Justification - Yale University

PERSONNEL:

Mark Gerstein, Ph.D. PI (.45 summer months). Dr. Gerstein 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 be directly supervising Dr. Harmanci and Dr. Navarro on this project.

Dr. Arif Harmanci, Ph.D., Assoc. Research Scientist (12 calendar months). Dr. Harmanci has extensive experience with bioinformatic approaches to genome-wide analysis and a strong background in scientific computation. As part of his PhD thesis, he developed advanced methods for RNA secondary structure prediction. In the Gerstein laboratory, he has developed new algorithms to identify transcription factor binding peaks from ChIP-Seq data. He is currently working on transcriptome analysis of several RNA-Seq datasets that include the Geuvadis dataset (RNA-Seq on 500 individuals). He will work on the analysis proposed in the grant under the direction of Dr Gerstein. He will supervise the involved laboratories during the integration of structural variation detection methods and also integrate the of RNA-seq dataset into the functional prioritization of SVs.

Fábio Navarro, Ph.D., Postdoctoral Assoc. (7 calendar months years 1 and 3, 8 calendar months year 2). Dr. Navarro has a strong background in scientific computation and biochemistry. He graduated from Universidade Federal de São Carlos in Computer Engineering and obtained his Ph.D. at the Biochemistry program from the Universidade de São Paulo. His background is development of software solutions to biological problems involving analysis and mining of genomic and trascriptomic data. He used to work with the description of genetic variation in human populations and human diseases such as cancer and Xeroderma Pigmentosa. He is now working with the expression profile of repetitive elements in the healthy tissues and investigating the functional impact of point mutation and structural variation in the human genome. Dr. Navarro will integrate and further develop tools to detect structural variations based on whole genome sequencing data and will extend the methods to prioritize SVs based on protein coding annotation, non-coding regulatory regions and eQTL analysis.

Fringe Benefits:

Fringe benefits are calculated at 31.1% for the PI, ARS and PDA.

EQUIPMENT:

In year 1, we need a Dell Poweredge R815 server with 160GB of memory and four AMD Opteron processors. This will be used for processing our data and digital visualization. This is needed to complete the proposed research and will solely benefit this project.

TRAVEL:

We are budgeting incremental funds for airfare, lodging and meal expenses to attend a scientific meeting annually that benefit 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, covering such expenses as: diskettes, tapes, and other miscellaneous computer parts (e.g. replacing worn out surge suppressors), 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.

Indirect costs are calculated at Yale's federally negotiated rate of 67.5% of modified total direct costs. DHHS agreement dated 02/23/2016.