Improving the Efficiency of Genome Variants Detection by the Parallelization of Its Computer Process

Cendes, Lucas

The analysis for genome variants is a very intensive computer process. It requires a very large amount of memory and processing power. As a result of that, it can take up to one full day for it to finish the analysis on a single genome. To solve this problem, a tool was developed, with the main intention of decreasing the processing time. The script works by parallelizing the process, by dividing it in several smaller pieces, and distributing it among several processors cores. As stated before, this tool is expected to significantly reduce the processing time of the analysis for genome variants. To accomplish this, the script was written to be interpreted by the software that runs the genome analysis. This was later run on a computer with 8 processors, with 6 cores each. It was run a total of 30 times, being composed of 5 trials, each one of it having the execution being run on 1, 2, 4, 8, 16 and 32 threads of execution. Based on the time of each execution, the efficiency and the speedup of the tool were calculated. The analysis process, running on a single thread took approximately 22 hours to be completed, while the one running on 32 threads took 3 hours. These results validated the hypothesis, and allowed more genome analyses to be made, decreasing the patients waiting time and allowing more exams to be processed in a considerably shorter amount of time.

Awards Won:
National Institute on Drug Abuse, National Institutes of Health &amp the Friends of NIDA: Second Award of $1,500