U.S. Department of Health and Human Services
U.S. Food and Drug Administration

2015 Science Writers Symposium

Lab Tour: HIVE: High-performance Integrated Virtual Environment

Vahan Simonyan, Ph.D., Lead scientist, HIVE

Center for Biologics Evaluation and Research

High-performance computation infrastructure performing NGS bioinformatics computations that are massively parallel (executed on multiple computers simultaneously)

The High-performance Integrated Virtual Environment (HIVE) is a distributed storage and computer environment designed primarily to handle Next-Generation Sequencing (NGS) data. This multicomponent cloud infrastructure provides secure Web access for authorized users to deposit, retrieve, annotate, and compute on NGS data, and to analyze the outcomes using Web interface visual environments appropriately built in collaboration with research and regulatory scientists and other end users.

Sequencing even a single person's genome requires collecting a vast amount of data. HIVE was designed by a collaboration between the FDA and academia to provide analysis and storage support to accumulate such genetic information. It is also capable of performing quality control checks to ensure that the sequences are accurate and valid and to make a variety of computations that allow for the identification of biologically valuable information and to present the data in easily analyzable visualizations.

Maxi-hive: To support long-term storage and large-scale computations for regulatory submissions for NGS and provide standardization portal, White Oak /CDRH

Storage: ~2 Petabytes (comparable to 1 million HD movies), metal + SunGrid

CPU: 1500 cores, extensible to 3000–5000

Network: 10Gb ⇒ Internet2, 40Gb ⇒ Infiniband

Mini-hive: Research and scientific NGS portal with cutting edge production quality tools, White Oak/CBER server room

Storage: ~500 Terabytes, metal

CPU: ~350 cores

Network: wan 1Gb, lan 40 GB

Questions? Contact FDA's Office of Media Affairs at 301-796-4540 or fdaoma@fda.hhs.gov