



Contribution ID: 69

Type: **Invited talk (plenary/keynote)**

H3ABioNet –enabling bioinformatics and big data research in Africa

Wednesday, 6 December 2017 11:00 (30 minutes)

Bioinformatics, or computational biology, is the application of computing to the analysis and interpretation of biological data. With the emergence and reducing cost of new laboratory technologies for high throughput data generation, African scientists are increasingly generating bigger and more complex datasets. This has led to significant challenges in all aspects of data management, including transfer, storage, analysis and interpretation. One of the initiatives which aims to address these challenges is H3ABioNet, a Pan-African bioinformatics network established to build the capacity for large scale genomics research on the continent. A major project within H3ABioNet has been the design of a new genotyping array for African populations, which identifies genetic differences between individuals that may be causative of specific diseases. The design of the array required significant computing hardware at several different stages due to the size and complexity of the data and algorithms. The project incorporated over 4000 human genome sequences, with collectively over 100 million genetics changes (Single nucleotide variants), that needed to be analysed by multiple different software tools, with millions of statistical tests. Through necessity, the H3Africa array design project utilized the CHPC resources for data transfer, storage and processing. This talk will outline the need for high performance computing in bioinformatics, using the array design project as an example, and will also discuss the challenges in big data in the biomedical sciences in Africa and some of the solutions implemented through H3ABioNet.

HPC content

The analysis of large biomedical datasets is increasingly requiring high performance computing. Here I describe a project within H3ABioNet where the CHPC was used in both the design and evaluation of a new genotyping array.

Primary author: Prof. MULDER, Nicola (University of Cape Town)

Co-authors: Dr CHOUDHURY, Ananyo (Wits University); Mr MEINTJES, Ayton (University of Cape Town); Dr CHIMUSA, Emile (University of Cape Town); Mr BOTHA, Gerrit (University of Cape Town); Mr MBIYAVANGA, Mamana (University of Cape Town); Prof. HAZELHURST, Scorr (Wits university)

Presenter: Prof. MULDER, Nicola (University of Cape Town)

Session Classification: BioInformatics

Track Classification: Bioinformatics and Biological Sciences