



Contribution ID: 92

Type: **Talk**

Challenges and Opportunities for African Genomics Research

Wednesday, 5 December 2018 11:20 (15 minutes)

For the promise of genomics to drive precision medicine to be realized, numerous human genome sequence samples and high depth coverage and from a variety of populations is required. Through initiatives like H3Africa, population scale genomics is being applied to help characterize diverse African populations that have currently not been featured in the genomics revolution. African genomes represent an uncharacterized portion of human genomics and may hold opportunities for novel discoveries.

Genomics data however is big data, the collection, transfer, storage and analyses of genomics data poses challenges to the infrastructure needed to manage such data. Key to human genomics data is governance policies that control the access of sensitive genetic data and achieving a balance between not being a barrier to access data that drives scientific discoveries, and at the same time protect the individuals that provide data. Associated with human genomics data is the rich diversity of meta-data that provide contextual information of the data. Some of the meta-data can be shared, other meta-data that might possibly make a sample identifiable will need to be under controlled access. Meta-data standards that are interoperable are required for the sharing of data and also for any meta-analysis to be conducted and form the core of good data management when dealing with copious volumes of data. The adoption of FAIR principals (Findable, Accessible, Interoperable and Reusable) is becoming prevalent across all scientific disciplines and in some cases mandated by different funding agencies. The technological adoption needed to turn FAIR principals into reality is non-trivial, but achievable and should be incorporated within good data management practices.

For genomics data to be translated into actionable outcomes with impacts, intensive computational processing and analyses of the data needs to be undertaken to convert information into knowledge. Traditionally in bioinformatics numerous software packages, versions and dependencies are needed to undertake an analyses which can run into 100s or 1000s of computational hours making reproducibility of the science a key issue. Development of computational workflows and containerized environments with user defined software stacks which can be hosted on various high performance computing systems enables the sharing of software environments, versions, and computational workflows leading to reproducible science.

Presenter Biography

Sumir Panji obtained his PhD in Bioinformatics from the University of the Western Cape as part of the Stanford South Africa Biomedical Informatics (SSABMI) programme where he developed computational and analyses pipelines to determine the intersection between bacterial virulence and positive selection in Professor Winston Hide's laboratory. He completed his postdoctoral studies in Professor Alan Christoffels' laboratory at the South African National Bioinformatics Institute (SANBI) where his focus was on genome assembly, annotation, data mining, large scale statistical analysis of genomics data and development of various computational pipelines and analyses workflows for a myriad of genomics' data types. Dr Panji's main interests are in creating and implementing computational and analyses workflows, statistical analysis of biomedical data, biological algorithms, high performance computing and the overall application of bioinformatics and genomics methods to better understand complex biological systems. Dr. Panji is currently a bioinformatician within the H3ABioNet consortium working on the implementation and interpretation of bioinformatics solutions and new technologies such as Machine Learning to diverse biological problems and actively involved in

providing bioinformatics support to the H3Africa projects and helping to develop bioinformatics capacity on the African continent.

Primary author: Dr PANJL, Sumir (UCT Computational Biology Division / H3ABioNet)

Co-authors: MULDER, Nicola (University of Cape Town); Mr MEINTJES, Ayton (UCT Computational Biology Division / H3ABioNet); Mr PARKER, Ziyaad (UCT Computational Biology Division / H3ABioNet); Mr BOTHA, Gerrit (UCT Computational Biology Division / H3ABioNet); Mr MASLAMONEY, Suresh (UCT Computational Biology Division / H3ABioNet)

Presenter: Dr PANJL, Sumir (UCT Computational Biology Division / H3ABioNet)

Session Classification: DIRISA

Track Classification: DIRISA