Using High Performance Computing To Create And, Freely Deliver, The Asian Genomic Database Necessary For Precision Medicine In This Population

SS Jamuar[1], JD Picker[1]
[1]Global Gene Corporation Pte Ltd, Singapore

Introduction: Precision medicine refers to tailoring medical management of an individual based on one's characteristics, with genomics data being most commonly used. Hence, delivery of precision medicine requires robust population specific reference genome datasets. However, the majority of publicly available genomic databases are biased towards the Caucasian population and do not reflect the diversity that exists within the Asian populations. For example, in a recent analysis by the Exome Aggregation Consortium group at Broad Institute, only 9 out of 192 variants, previously called as pathogenic, were truly pathogenic, while over 160 variants were population specific polymorphisms, and hence, likely benign. Furthermore on comparing the standard datasets to bushmen in the KB1 African genome analysis, it was found that there was an increased frequency of sequence variation between them, and over 47% of variants identified were novel, affecting over 7700 genes; indicating the scale of population diversity.

Aim: To build reference data sets for Asian populations and to make these freely available to researchers and to the public.

Method/ Results: We aggregated genomic datasets, generated either internally or by collaborators, of cohorts of individuals of South Asian ethnicity. Using high performance computing, we combined these datasets and performed bioinformatics analysis to convert the raw sequence data in FASTQ format to jointly-called gVCF format. In collaboration with Global Alliance for Genomics and Health, this data was then uploaded to create the Beacon for South Asian population (ggcINDIA: https://beacon-network.org/#/search?beacon=238669).

Conclusion: The Beacon is a freely available resource and allows researchers and/or public to query the presence or absence of a given variant detected in their own discovery cohort, and allows for filtering of variants for rarity, specifically within the South Asian population. Over time, as more data is generated and aggregated, the ggcINDIA beacon will provide precise genomic data that is critical to deliver precision medicine within Asia.