Abstract ID:- 161

Abstract Topic:- Evolutionary and population genetics

Abstract Title:- Genome reference wars: T2T-CHM13 vs. GRCh38 for Next-Generation Sequencing analysis.

Presenting author name :- Jyoti Sharma

Presenting author institute:- Indian Institute of Technology Jodhpur

Co-author name:- Pankaj Yadav

Co-author institute:-Indian Institute of Technology Jodhpur

Aims:-The human genome has been studied for decades, and its application as a reference genome has increased since the release of "The Human Genome Project." The latest reference genome available from the Genome Reference Consortium (i.e., GRCh38) includes alternate haplotypes to represent population diversity. Yet, 151 Mbp of gap sequences are distributed throughout the genome in GRCh38. Recently, a Telomere-to-Telomere (T2T) Consortium has released a complete human genome sequence, including gapless assemblies for all chromosomes. This new genome release provides a comprehensive and accurate reference assembly, which can help in understanding the functions of novel genes and the genetic basis of diseases across the population.

Methods:- In our work, we analysed next-generation sequencing (NGS) data using the T2T genome as a reference to evaluate its potential advantages. Using the GRCh38 and T2T genomes, we compared the alignment of NGS data of diverse populations obtained from the 1000 Genomes Project Consortium. We examined six chromosomes (e.g., 9, 13, 15, 20, 21, and 22) to identify additional variants present in the T2T alignment, albeit absent in the GRCh38 alignments.

Results:- Our results revealed that the T2T alignment identified a higher number of genetic variants than the GRCh38 alignment, including increased single-nucleotide polymorphisms (SNPs), multi-nucleotide polymorphisms, and insertions. For instance, gene CDKN2B-AS1 shows the insertion of eight bases when aligned with the T2T genome as compared to the alignment with the GRCh38 genome.

Conclusions:- These findings suggest that the T2T genome can improve the accuracy and sensitivity of NGS data analysis, particularly in detecting population-specific genetic variants. T2T-CHM13 genome enables the detection of previously undetected frame-shift mutations that were missed when using the GRCh38 genome as a reference.

Keywords:- Human genome reference, T2T-CHM13, GRCh38, Next-Generation Sequencing analysis