Title of the talk: Secondary Analysis of Genomes

Abstract: Once the raw data is obtained from the sequencing machine, it must go through primary analysis and secondary analysis before the tertiary analysis can start. Primary analysis is NGS platform dependent and allows little scope for hardware agnostic improvement in producing the final reads and base quality values. The secondary analysis is therefore the critical stage where one must focus to harvest the best and the most confident insights from the genomic data. Any error in this stage is carried forward to the tertiary stage leading to spurious inferences. Due to the large volume of the NGS data one encounters during the secondary analysis in terms of FASTQ files, most users tend to treat this stage as a black box and focus on single or multi-sample variant calling using the SAM and BAM outputs obtained from the aligners. In the talk I will review some popular aligners and compare them with the aligner that we have developed in our group to bring out the possibilities of improvement in the genomic data harvesting and its impact on biological inference.