**Variant detection model with enhanced robustness and accuracy**

**for low-depth targeted next-generation sequencing data**

Massively parallel sequencing data has been generated by Next-generation sequencing (NGS) technology for single nucleotide variants (SNVs) identification among related populations. To address the detection of SNVs, a variety of methods are being under-represented. However, by the reason of the error rate of the clonal heterogeneity and the limitation of the algorithms, identifying the true variants at minor allele frequencies with the very low sequencing read depth remains challenging.

We proposed a novel empirical Bayesian model to call variants accurately. Sensitivity of this hierarchical model is analyzed by the improper prior, information prior and non-information prior on the synthetic sequence data with varying read depth and a range of allele frequencies. Our model with information prior (log-normal prior) and non-information prior (Jeffreys prior) performs a highly accuracy (96%) when applied to a know 0.1% minor allele frequency with very low read depth (39). The Jeffreys prior, as a typical and influential one, also shows 0 false discovery rate (FDR) to a known 0.1% minor allele frequency event, which is much better than using improper prior. For further validation, our statistical model with Jeffreys prior was applied on the yeast sequence data which also shows a highly specificity and sensitivity over a wide range of read depth. Thus our model has a high robustness and accuracy for low read depth and minor allele frequencies.