

## **New strategies using unmapped reads to characterize viral sequences in different human populations.**

Mio Shibata

Kyoto–McGill International Collaborative School in Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan

Unit of Statistical Genetics, Center for Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan

Department of Human Genetics, McGill University, Montréal, QC H3A 1C7, Canada

Most studies using human whole-genome sequencing have focused on reads mapped to the reference genome. However, not all reads map to this reference and some studies have shown that 0.13% of the unmapped reads have similarities to viral sequences. Indeed, some viruses, such as HHV6, are integrated into human genomes during evolution and inherited from generation to generation, but these polymorphic sequences might be missing from the reference. Moreover, other viruses, like HIV and HTLV-1, are associated with active infections and could be associated with new integration sites. Some viruses' detection implies viremia, such as the Human herpes virus family, such as EBV, HHV7 or potential pathogenesis such as hepatitis with Anello virus. The mechanisms of interaction between various viruses and humans have not been completely revealed yet. Moreover, host genetic factors contributing to immune response factors remain to be determined.

In our study, whole-genome sequences were extracted from blood samples obtained from two cohorts of different ancestry: 1,830 samples of European ancestry and 1,322 samples of Japanese origin. Both cohorts represent populational samples and were not selected for a specific disease. We segregated the unmapped reads from these samples after aligning to human genome reference and identified the viral reads using the Refseq virus database (version 91). As the pilot study using a small sample size, we found that 0.018% and 0.008% of the unmapped reads on average showed similarity to viruses in European cohorts and Japanese cohorts, respectively. In these cohorts so far, we found that the prevalence of EBV was 24.0 % in Europe, 15.5 % in Japanese, respectively, and that of HHV6A/B was 2 % in Europe and 0.5 % in Japan. Moreover, some samples were found to have viral reads from multiple viruses. The unmapped reads rooted from viral reads in human samples would make it possible to further understand the characteristic of the population and the interaction between the host, especially humans, and viruses.