Analyzing the SARS-CoV-2 genome with target enriched RNA-seq

Genomics

Sarah Trusiak (sarah.trusiak@roche.com), Roche Sequencing and Life Science, **Jonathan Nowacki**, Roche Sequencing and Life Science, **Ranjit Kumar**, Roche Sequencing and Life Science, **Spencer Debenport**, Roche Sequencing and Life Science, **Rachel Kasinskas,** Roche Sequencing and Life Science

Since the emergence of the SARS-CoV-2 and its resulting disease, COVID-19, over 30 million people have been infected worldwide and over 900,000 people have died. Many research programs have pivoted to studying the virus and understanding its health impacts and disease complications to develop treatments. RNA sequencing (RNA-seq) is a high-throughput method that enables research into the host's transcriptional response to viral infection and the RNA genome of the virus itself; this is critical for understanding the impact of new viral mutations and emerging strains. Samples from infected hosts typically contain a single SARS-CoV-2 strain, while environmental samples like soil and sewage often contain a mixture of strains. When studying the RNA genome of viruses in mixed-RNA samples, such as total RNA from a human host, it is necessary to enrich for the viral RNA molecules above the much-more-abundant background RNA. Hybridization-based target enrichment (TE) using well-designed probes complementary to viral sequences has the potential to isolate and greatly enrich for viral reads of interest, and enable the capture of diverse viral strain mutations with a single TE panel, providing insights into the distribution and evolution of the virus. To enable hybridization capture RNA-seq for SARS-CoV-2 genome sequencing, we have developed a KAPA COVID-19 TE panel and workflow. Our probe panel covers >99.7% of 184 publicly available SARS-CoV-2 sequences (NCBI). Using this panel, we have developed a new target-enriched RNA-seq workflow that incorporates the KAPA COVID-19 TE panel into a modified version of the HyperCap Workflow v3 and includes the KAPA RNA HyperPrep Kit. In order to determine the lowest viral load that yields full coverage of the SARS-CoV-2 genome using the KAPA COVID-19 TE workflow, we tested varying levels of viral genome copies in different amounts of human total RNA. We then compared the performance of the KAPA COVID-19 TE panel workflow to panels from three different suppliers to identify differences in minimum viral load required for full genome coverage, as well as the ability to detect strain mutations. We show here that the KAPA COVID-19 TE panel detects mutations from six different strains of SARS-CoV-2 within a single target enrichment reaction. We conclude that this panel, when used with the modified HyperCap Workflow v3 adapted for RNA-seq, is a powerful tool for studying the SARS-CoV-2 genome in both infected host samples and samples containing mixed viral strains, and for distinguishing between multiple strains in a single sample.