ABSTRACT

Continued advances in next-generation sequencing (NGS) technology, such as capacity, speed, and reduced cost per sequenced base, revolutionize personalized medicine and bring genomics into routine clinical practice. Nevertheless, NGS is still under rapid development, and the variability of sequencing protocols and validation procedures is one of its current bottlenecks.

This thesis aimed to study the influence of different sample sources and NGS protocols (NGS library construction-sequencing-data analysis) on the accuracy of NGS analysis in diagnostic applications. In the first study, performed during the COVID-19 pandemic outbreak, we developed NGS protocols suitable for a whole genome analysis of the SARS-CoV-2 virus. Subsequently, in the second study, we examined the suitability of human saliva-derived gDNA for genomic/genetic analysis of selected variant types compared to traditional blood-derived gDNA using validated NGS protocol and statistical comparison of the obtained data.

Whole genome analysis of the SARS-CoV-2 genome was performed using two captured-based approaches and one amplicon-based approach to study the quality and effectivity of the respective NGS protocols. Synthetic controls were employed to verify the accuracy and specificity of the developed NGS protocols. We proved that real-time quantitative PCR-based quantitation of viral load was the right tool since subsequent sample plexing utilizing cycle threshold values resulted in sequencing data with required coverage uniformity between different samples. We found the capture-based NGS protocol the most suitable for whole genome analysis of the SARS-CoV-2 genome.

In the main study, we analyzed whether human saliva may serve as an equal source of gDNA to blood for single nucleotide (SNV) and small insertion and deletions (small-indels) variant analysis. We designed and validated entire NGS protocols for whole exome (WES) and whole genome (WGS) analysis employing the Coriel NA12878 standard sample and the latest human reference genome GRCh38. Consequently, we analyzed NGS data from 10 paired blood-saliva samples obtained by engaging the same Coriel NA12787 NGS protocol, using statistical analysis tools on the F1 score and other selected sequencing parameters. For the WES protocol, the median value of F1 score for ten paired blood-saliva samples was 0.9858 for SNVs and 0.9076 for small-indels. For the WGS protocol, the median value of F1 was 0.9761 for SNVs and 0.9511 for small-indels. The study's comprehensive results demonstrated a high level of concordance between blood and saliva samples compared to Coriel standard results for F1 scores in the case of SNV and small-indels and for both the WES and WGS NGS protocols, respectively.

These studies advanced our understanding of genome sequencing of samples of a different origin and proved saliva as a suitable source of genomic/genetic data comparable to blood. These findings affect further genomic/genetic research and NGS clinical applications.