WWU MÜNSTER
Institut für Medizinische Informatik
IMI
medizinische fakultät
Westfälische Wilhelms-Universität Münster

Titel:

SimFFPE: An Artifact Chimeric NGS Read Simulator Specifically Designed for FFPE samples

Abstract:

Technical ease and low storage costs make formalin-fixed paraffin-embedded (FFPE) tissue processing a common method for long term tissue specimen storage. However, formalin-fixation can result in fragmented, degraded, protein cross-linked DNA, introducing false positive results to next generation sequencing (NGS) data analysis. Studies on FFPE artifacts have been focusing on noise in single nucleotide variant calling, with little attention paid to structural variation (SV). We found that FFPE samples are enriched with artifact chimeric reads that would result in a large number of false positive SV calls. To evaluate and improve the performance on SV calling in FFPE samples, ground-truth data set is needed. However, publicly available real-world FFPE data sets with matched fresh frozen samples are scarce, and there is no experimental validation on SV candidates. Due to the lack of real-world ground-truth data sets, optimization of SV calling can be much easier if simulated data with the properties of FFPE samples is available. Here we present SimFFPE, a NGS read simulator that can simulate artifact chimeric reads in FFPE samples. To our knowledge, this is the first NGS read simulator designed specifically for FFPE samples. With SimFFPE, we were able to compare the performance of different SV callers, and further improve the performance by developing a tool to remove these artifact chimeric reads in FFPE samples.