ABSTRACT

The continued success of antiretroviral programmes in the treatment of HIV is dependent on access to a cost-effective HIV drug resistance test (HIV-DRT). HIV-DRT involves sequencing a fragment of the HIV genome and characterising the presence/absence of mutations that confer resistance to one or more drugs. HIV-DRT using conventional DNA sequencing is prohibitively expensive (~US$150 per patient) for routine use in resource-limited settings such as many African countries. While the advent of ultra deep pyrosequencing (UDPS) approaches have considerably reduced (3-5 fold reduction) the cost of generating the sequence data, there has been an even more significant increase in the volume of data generated and the complexity involved in its analysis. In order to address this issue we have developed Seq2Res, a computational pipeline for HIV drug resistance test from UDPS genotypic data.

We have developed QTrim, software that undertakes high throughput quality trimming of UDPS sequencing data to ensure that subsequently analysed data is of high quality. The comparison of QTrim to other widely used tools showed that it is equivalent to the next best method at trimming good quality data but outperforms all methods at trimming poor quality data. Further, we have developed, and evaluated, a computational approach for the analysis of UDPS sequence data generated using the novel Primer ID that enables the generation of a consensus sequence from all sequence reads originating from the same viral template, thus reducing the presence of PCR and sequencing induced errors in the dataset as well as reducing. We see that while the Primer ID approach does undoubtedly reduce the prevalence of PCR and sequencing induced errors, it artificially reduces the diversity of the subsequently analysed data due to the large volume of data that is discarded as a result of there being an insufficient number of sequences for consensus sequence generation.

We validated the sensitivity of the Seq2Res pipeline using two real biological datasets from the Stanford HIV Database and five simulated datasets The Seq2Res results correlated fully with that of the Stanford database as well as identifying a drug resistance mutations (DRM) that had been incorrectly interpreted by the Stanford approach. Further, the analysis of the simulated datasets showed that Seq2Res is capable of accurately identifying DRMs at all prevalence levels down to at least 1% of the sequence data generated from a viral population.

Finally, we applied Seq2Res to UDPS resistance data generated from as many as 641 individuals as part of the CIPRA-SA study to evaluate the effectiveness of UDPS HIV drug resistance genotyping in resource limited settings with a high burden of HIV infections. We find that, despite the FLX coverage being almost three times as much as that of the Junior platform, resistance genotyping results are directly comparable between both of the approaches at a range of prevalence levels to as low as 1%. Further, we find no significant difference between UDPS sequencing and the “gold-standard” Sanger based approach, thus indicating that pooling as many as 48 patient’s data and sequencing using the Roche/454 Junior platform is a viable approach for HIV drug resistance genotyping. Further, we explored the presence of resistant minor variants in individual’s viral populations and find that the identification of minor resistant variants in individuals exposed to nevirapine through PMTCT correlates with the time since exposure.

We conclude that HIV resistance genotyping is now a viable prospect for resource limited setting with a high burden of HIV infections and that UDPS approaches are at least as sensitive as the currently used Sanger-based sequencing approaches. Further, the development of Seq2Res has provided a sensitive, easy to use and scalable technology that facilitates the routine use of UDPS for HIV drug resistance genotyping.